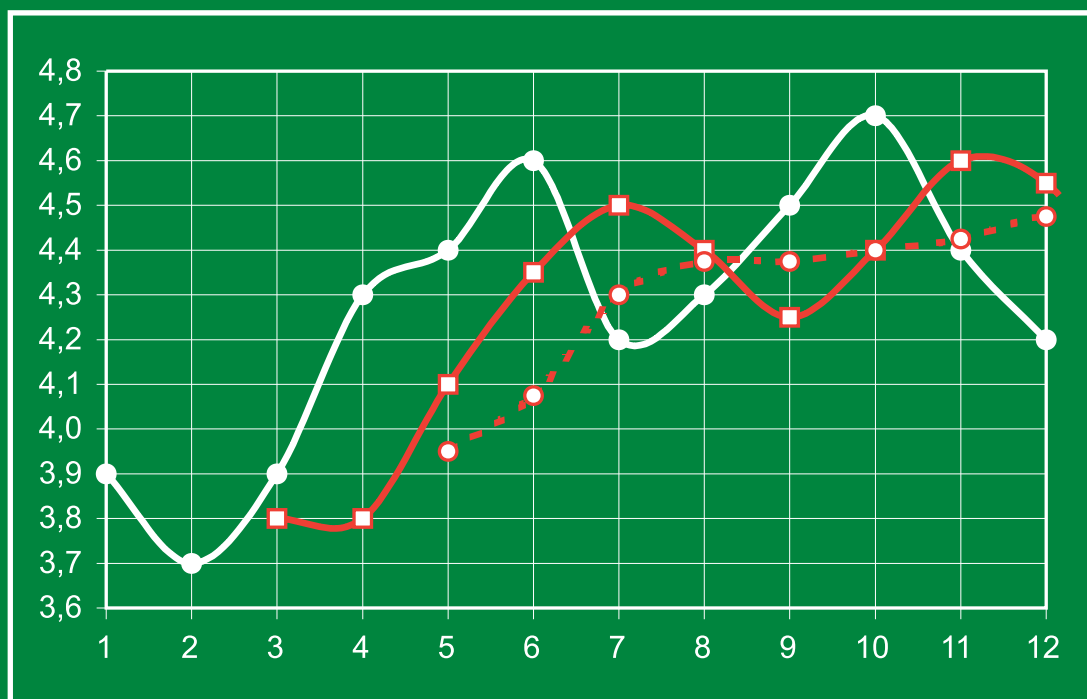


Marian Duczmal
Władysław Wornalkiewicz

WSTĘP DO EKONOMETRII I BADAŃ OPERACYJNYCH

Zbiór przykładów z zastosowaniem mikrokomputera



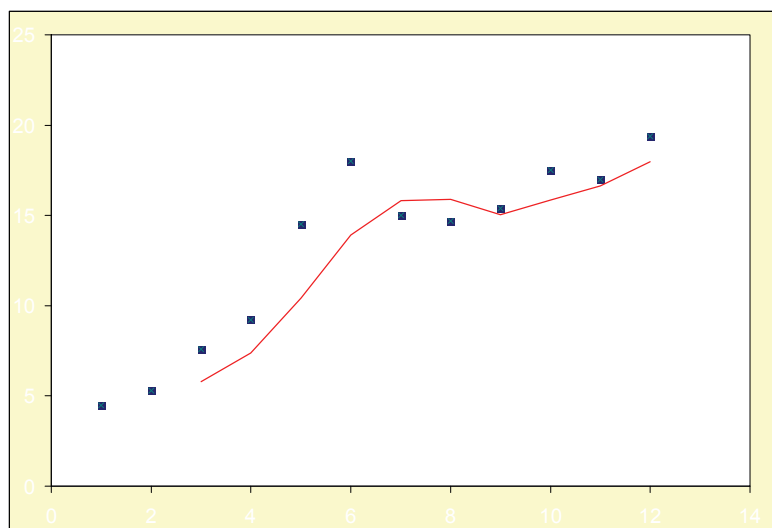
**WSTĘP DO EKONOMETRII
I BADAŃ OPERACYJNYCH**

Wyższa Szkoła Zarządzania i Administracji w Opolu

Marian Duczmal
Władysław Wornalkiewicz

WSTĘP DO EKONOMETRII I BADAŃ OPERACYJNYCH

Zbiór przykładów z zastosowaniem mikrokomputera



Opole 2010

Recenzent
prof. dr hab. Ryszard Broszkiewicz

Redakcja
Violetta Sawicka

Korekta
Agata Miemus
Violetta Sawicka

Redakcja techniczna
Agata Miemus
Andrzej Pasierbiński

ISBN 978-83-88980-94-7
ISBN 978-83-62105-41-0

Seria MONOGRAFIE I OPRACOWANIA
pod redakcją prof. dr. hab. **Mariana Duczmala**

WYDAWNICTWA
WYŻSZEJ SZKOŁY ZARZĄDZANIA I ADMINISTRACJI W OPOLU

WYDAWNICTWO INSTYTUT ŚLĄSKI Sp. z o.o.
Opole, ul. Piastowska 17, tel. (77) 4540 123
e-mail: wydawnictwo@is.opole.pl
Nakład 250 egz. Objętość 39,46 ark. wyd., 37,38 ark. druk.

Spis treści

Wstęp	11
C z ę ś ć 1	
Zagadnienia wstępne i wybrane modele ekonometryczne	
1.1. Wprowadzenie	17
1.1.1. Pojęcia „ekonometria” i „model ekonometryczny”	17
1.1.2. Struktura modelu ekonometrycznego	17
1.1.3. Kryteria klasyfikacji modeli ekonometrycznych	18
1.1.4. Etapy budowy modelu ekonometrycznego	21
1.1.5. Testy oceniające dopasowanie modelu do danych empirycznych	23
1.1.5.1. Istotność współczynnika korelacji	23
1.1.5.2. Istotność współczynnika korelacji wielorakiej	24
1.1.5.3. Istotność parametrów strukturalnych	25
1.1.6. Podstawowe testy statystyczne do sprawdzania własności odchyłeń losowych	26
1.1.6.1. Losowość	26
1.1.6.2. Normalność rozkładu	27
1.1.6.3. Nieobciążoność	30
1.1.6.4. Autokorelacja	31
1.1.6.5. Stałość (jednorodność) wariancji składnika losowego (oparta na teście Fishera-Snedecora)	33
1.1.7. Pytania	34
1.2. Oprogramowanie mikrokomputerowe do ekonometrii	36
1.2.1. Uwagi wstępne	36
1.2.2. Programy komputerowe	37
1.2.2.1. GRETL	37
1.2.2.2. EViews	39
1.2.2.3. WinQSB	42
1.2.2.4. Excel	44
1.2.3. Informacje o innych programach	46
1.2.4. O oprogramowaniu modeli symulacyjnych	49
1.2.4.1. TESS-2	52
1.2.4.2. Program: Zarządzanie kapitałem	52
Literatura pomocnicza do podrozdziału 1.2	53
1.3. Miary rozkładu jednej zmiennej (przykłady w Excelu)	55
1.3.1. Przedziały obserwacji	55
1.3.2. Miary przeciętne (położenia)	57
1.3.2.1. Średnie	57
1.3.2.2. Przeciętne pozycyjne	62
1.3.2.3. Wykres pudełkowy oraz histogram	66
1.3.3. Miary rozproszenia (zmienności)	68
1.3.4. Miary asymetrii (skośności)	74
1.3.5. Pytania	77
1.3.6. Zadania	77

1.4. Rozkład wielu zmiennych	78
1.4.1. Sposób zapisu	78
1.4.2. Normalizacja	83
1.4.2.1. Standaryzacja	83
1.4.2.2. Unitaryzacja	84
1.4.2.3. Odległość	86
1.4.2.4. Klasyfikacja obserwacji	88
1.4.2.5. Porządkowanie obserwacji	94
1.4.3. Pytania	100
1.4.4. Zadania	101
1.5. Regresja i transformacja liniowa	103
1.5.1. Idea metody najmniejszych kwadratów	103
1.5.2. Określenie regresji względem jednej zmiennej	106
1.5.3. Określenie wielkości statystycznych	109
1.5.4. Pytania	115
1.5.5. Zadania	116
1.5.6. Funkcje trendu	117
1.5.6.1. Trend liniowy	117
1.5.6.2. Trend pełzający	119
1.5.6.3. Trend potęgowy	121
1.5.6.4. Trend wykładniczy	124
1.5.6.5. Trend logarytmiczny	127
1.5.6.6. Trend hiperboliczny	128
1.5.6.7. Trend logistyczny	131
1.5.6.8. Pytania	134
1.5.6.9. Zadania	134
1.6. Wybór kombinacji zmiennych objaśniających	137
1.6.1. Uwagi wstępne	137
1.6.2. Metoda wskaźników wektora zmiennych objaśniających	138
1.6.3. Bazowanie na mierze błędów <i>ex ante</i>	139
1.6.4. Analiza korelacji	141
1.6.5. Metoda pojemności nośników informacji	142
1.6.6. Metoda grafowa i inne	142
1.6.7. Testy porównywania modeli	147
1.6.8. Metoda regresji krokowej oraz minimalizacji resztowej sumy kwadratów	148
1.6.9. Metoda Hellwiga	149
1.6.10. Określenie zmiennych objaśniających z zastosowaniem Excela	151
1.6.11. Wybór zmiennych do modelu regresji z czterema zmiennymi objaśniającymi	160
1.7. Zastosowanie programów Excel oraz GRETL do weryfikacji modeli ekonometrycznych	164
1.7.1. Modele jednorównaniowe	164
1.7.2. Zastosowanie funkcji Excela	168

1.7.3. Dalsze badania modelu ekonometrycznego (w Excelu)	178
1.7.4. Zastosowanie programu GRETL	189
1.8. Zależność między jedną zmienną objaśnianą a wieloma zmiennymi objaśniającymi	199
1.8.1. Dobór modelu regresji poprzez obserwację wykresów korelacyjnych	200
1.8.2. Mierniki dopasowania modelu	204
1.8.3. Pytania i zadania	206
1.8.4. Zadania	206
1.9. Zagadnienie modeli wielorównaniowych z zastosowaniem programu GRETL	210
1.9.1. Model wielorównaniowy	210
1.9.2. Model prosty	211
1.9.3. Model rekurencyjny	212
1.9.4. Przykład modelu rekurencyjnego	212
1.9.5. Model o równaniach łącznie współzależnych	216
1.9.6. Przykład zastosowania podwójnej metody najmniejszych kwadratów (2MNK) w Excelu	219
1.9.7. Przykład zastosowania metody 2MNK w programie GRETL	224
1.10. Modele stacjonarnych szeregów czasowych	230
1.10.1. Badanie stacjonarności przy określonych warunkach	230
1.10.2. Badania rzędu p autoregresji	236
1.10.3. Badania rzędu średniej ruchomej	238
1.10.4. Pytania	244
1.10.5. Zadania	244
1.11. Modele zmiennych dychotomicznych	246
1.11.1. Idea uogólnionej metody najmniejszych kwadratów	246
1.11.2. Liniowy model prawdopodobieństwa (LPM)	247
1.11.3. Model logitowy (logit)	248
1.11.4. Model probitowy (probit)	249
1.11.5. Zastosowanie programu Excel do modelowania zmiennych dychotomicznych	249
1.11.6. Pytania	260
1.11.7. Zadania	260
1.12. Zastosowania modelu regresji oraz modeli opartych na szeregach czasowych w prognozowaniu (użycie Excela)	262
1.12.1. Błąd prognozy <i>ex ante</i> oraz badanie aktualności modelu	266
1.12.2. Pytania	271
1.12.3. Zadania	271
1.12.4. Budowa i weryfikacja modelu regresji	272
1.12.5. Przykłady różne z zakresu kilku zmiennych objaśniających	288
1.12.6. Przykłady na zastosowanie testów porównywania modeli	291
1.12.7. Pytania (do 1.12.4–1.12.6)	293

1.12.8. Zadania (do 1.12.4–1.12.6)	294
1.12.9. Modele oparte na szeregach czasowych	296
1.12.9.1. Modele prostych średnich ruchomych	298
1.12.9.2. Modele ważonych średnich ruchomych	300
1.12.9.3. Model prostego wyrównywania wykładniczego	302
1.12.9.4. Model adaptacyjny wyrównywania wykładniczego	303
1.12.9.5. Modele podwójnych średnich ruchomych	305
1.12.9.6. Model Browna	309
1.12.9.7. Pytania	311
1.12.9.8. Zadania	312
1.12.10. Klasyczny model autoregresyjny (zastosowanie programu WinQSB)	315
1.12.10.1. Pytania	327
1.12.10.2. Zadania	327
1.13. Wyrównywanie trendu z zastosowaniem modelu Holta-Wintersa	329
1.13.1. Wprowadzenie składowej sezonowej do modelu Holta-Wintersa (z użyciem Excela)	332
1.13.2. Przykłady doboru najlepszych parametrów modelu (zastosowanie modułu <i>FC</i> programu WinQSB)	339
1.13.2.1. Podwójne wyrównywanie wykładnicze z trendem	339
1.13.2.2. Model Holta-Wintersa z sezonowością multiplikatywną	342
1.13.3. Pytania	344
1.13.4. Zadania	345
1.14. Model trendu, sezonowości i autoregresji (przykład w programie GRETL)	347
1.14.1. Wielomianowe modele trendu	347
1.14.2. Model Kleina ze zmiennymi sezonowymi	353
1.14.3. Przykład modelu autoregresyjnego	358

C z ę ś ć 2

Elementy badań operacyjnych

2.1. Zastosowanie programu WinQSB oraz rozszerzenia Excel-Solver	365
2.1.1. O pakiecie WinQSB w zakresie badań operacyjnych	365
2.1.2. Rozszerzenie arkusza kalkulacyjnego Excel-Solver	368
2.2. Analiza procesów decyzyjnych	372
2.2.1. Drzewo celów	372
2.2.2. Drzewo decyzyjne jako narzędzie do podejmowania decyzji w warunkach niepewności	376
2.2.3. Przedstawienie sytuacji decyzyjnej za pomocą macierzy	378
2.2.4. Reguły decyzyjne, gdy nieznanne jest prawdopodobieństwo	380
2.2.5. Bilans przepływów <i>input-output</i>	382
2.2.6. Pytania	386
2.2.7. Zadania	387

2.3. Rozwiązywanie programów liniowych	390
2.3.1. Idea algorytmu simpleks	390
2.3.2. Formułowanie zadań decyzyjnych	393
2.3.3. Zastosowanie modułu (LP) pakietu WinQSB	399
2.3.4. Metoda geometryczna rozwiązywania zadań programowania liniowego	405
2.3.5. Zastosowanie dodatku Excel-Solver do rozwiązania zadania decyzyjnego	409
2.3.6. Pytania	419
2.3.7. Zadania	420
2.4. Algorytm simpleks – sposób według elementu rozwiązyującego	425
2.4.1. Pytania	433
2.4.2. Zadania	434
2.5. Algorytm simpleks – zastosowanie rachunku macierzowego	435
2.5.1. Pytania	442
2.5.2. Zadania	442
2.6. Sztuczna baza danych	445
2.6.1. Pytania	452
2.6.2. Zadania	452
2.7. Analiza wrażliwości	455
2.7.1. Pytania	468
2.7.2. Zadania	468
2.8. Zadanie dualne	470
2.8.1. Pytania	477
2.8.2. Zadania	478
2.9. Zagadnienia nieliniowe	480
2.9.1. Metoda mnożników Lagrange’a	480
2.9.2. Wprowadzenie zmiennych nieistotnych	486
2.9.3. Pytania	490
2.9.4. Zadania	491
2.10. Metody rozwiązywania zagadnień transportowych	494
2.10.1. Zamknięte zagadnienie transportowe (ZTT) – zastosowanie WinQSB	494
2.10.2. Przykład zagadnienia transportowego rozwiązanego w dodatku Solver	500
2.10.3. Otwarte zagadnienie transportowe – OZT	501
2.10.4. Zamknięte zagadnienie transportowe – o niebilansujących się ilościach w klatkach zerowych	503
2.10.5. Pytania	520
2.10.6. Zadania	521

2.11. Zastosowanie WinQSB do programowania dynamicznego	523
2.11.1. Problem wyboru najkrótszej trasy (komiwojażera)	523
2.11.2. Załadunek	526
2.11.3. Zarządzanie produkcją i zapasami	527
2.11.4. Pytania	529
2.11.5. Zadania	529
2.12. Analiza przedsięwzięć	532
2.12.1. Metoda ścieżki krytycznej CPM	532
2.12.2. Wersja kosztowa CPM	534
2.12.3. Zadanie programowania liniowego według metody simpleks	335
2.12.4. Metoda PERT	542
2.12.5. Budowa sieci zależności programem Project	545
2.12.6. Zadania różne	547
2.12.7. Pytania	551
2.12.8. Zadania	552
Aneks	
Wyrażenia Excela	557
Zastosowane wzory (z podaniem podrozdziałów)	562
Tablice statystyczne	577
Bibliografia	592
Indeks rzeczowy	594

Wstęp

Podjęcie decyzji jest procesem złożonym. Składa się z etapów, w których nieodzowna jest umiejętność posługiwania się określonymi sformalizowanymi metodami ilościowymi. Zagadnienia z obszaru teorii podejmowania decyzji są dość starannie opracowane w wielu publikacjach. W wydaniach książkowych z zakresu statystyki, ekonometrii, prognozowania i symulacji czy też badań operacyjnych opisano metody postępowania oraz algorytmy częściowo poparte przykładami rozwiązań. Nieliczne i fragmentaryczne są jednak rozwiązania z zastosowaniem techniki mikrokomputerowej.

Niniejsze opracowanie, zdaniem autorów, będzie pomocne studentom w określaniu parametrów statystycznych, modelowaniu ekonometrycznym, wyznaczaniu prognoz gospodarczych, jak również znajdowaniu najlepszego wyniku funkcji celu przy podanych ograniczeniach zasobów. W pracy wykorzystuje się funkcje oraz moduły programowe pakietów programowych, a zwłaszcza:

- wyrażenia i funkcje elementarne arkusza kalkulacyjnego Excel,
- możliwości sporządzania statystycznej analizy danych źródłowych, bazując na rozszerzeniu programowym (Excel-Dodatki),
- określania prognoz oraz błędów ich oszacowania z zastosowaniem programu WinQSB,
- poszukiwania minimalnej lub maksymalnej wartości funkcji celu, mając do dyspozycji rozszerzenie programowe (Excel-Solver), czy też moduł (*Programowanie liniowe i całkowitoliczbowe* w pakiecie WinQSB),
- do estymacji oraz weryfikacji modeli ekonometrycznych w programie GRETL.

W rozdziale pierwszym *Zagadnienia wstępne i wybrane modele ekonometryczne* przedstawiono stosowane w literaturze programy komputerowe oraz wskazano na ich dostępność w internecie. Zaprezentowano przykłady w arkuszu kalkulacyjnym Excel zawierającym miary rozkładu jednej zmiennej, rozkład wielu zmiennych, z zakresu regresji i transformacji liniowej, a także dotyczące problematyki doboru zmiennych objaśniających. Oprócz wykorzystania Excela pokazano możliwość zastosowania programu GRETL do weryfikacji modelu ekonometrycznego. GRETL występuje również w przykładach modeli

wielorównaniowych. Opracowanie obejmuje przykłady modeli stacjonarnych szeregów czasowych, zmiennych dychotomicznych z zastosowaniem trendu, sezonowości i autoregresji.

W rozdziale drugim *Elementy badań operacyjnych* oparto się głównie na programie WinQSB oraz wspomagająco na arkuszu kalkulacyjnym Excel. Rozdział ten obejmuje tematy:

- analiza procesów decyzyjnych,
- rozwiązywanie zadań z programowania liniowego z zastosowaniem metody simpleks oraz programowania nieliniowego,
- zagadnienie transportowe,
- metody programowania dynamicznego,
- analiza przedsięwzięć.

Omawiając temat, zaprezentowano także przykład z wykorzystaniem programu Project, który służy do tworzenia sieci zależności złożonych przedsięwzięć inwestycyjnych lub modernizacyjnych.

Biorąc pod uwagę cel tej pracy – łatwość samodzielnego uczenia się na podstawie przykładów – starano się omijać opisy teoretyczne metod, do których Czytelnik może sięgnąć w dostępnej literaturze, podanej w zakończeniu opracowania. Pozostawiono jedynie niezbędne, krótkie wprowadzenia.

Impulsem do opracowania tak ujętej tematyki było widoczne zapotrzebowanie samych studentów studiów zaocznych kierunku zarządzanie, z uwzględnieniem ich uwag dotyczących przystępności pozycji literaturowych, prowadzonych wykładów oraz ćwiczeń. Łącząc pracę zawodową z nauką, chcą mieć przystępny materiał do studiowania, ułożony, w przybliżeniu, w bloki tematyczne spotkań okresowych. Wychodząc naprzeciw ich życzeniom, w sposób niemal „rzemieślniczy” pokazano możliwości zastosowania wybranych metod modelowania ekonometrycznego, a także z zakresu badań operacyjnych z użyciem dostępnych narzędzi programistycznych. Dla wygody studiujących w *Aneksie* zestawiono wzory, których opisy i zastosowanie znajdują się w niniejszej pracy, oraz dołączono tablice statystyczne.

Zastrzeżenia Czytelnika może budzić ujęcie naukowe, a także obszerności poruszonej problematyki z zastosowaniem programów mikrokomputerowych. Zdajemy sobie również sprawę z możliwości wystąpienia jeszcze licznych usterek merytorycznych. Dlatego będziemy bardzo wdzięczni tym Czytelnikom, których uwagi przyczynią się do poprawienia i wzbogacenia niniejszego opracowania z zachowaniem jednak celu głównego, tj. przystępności.

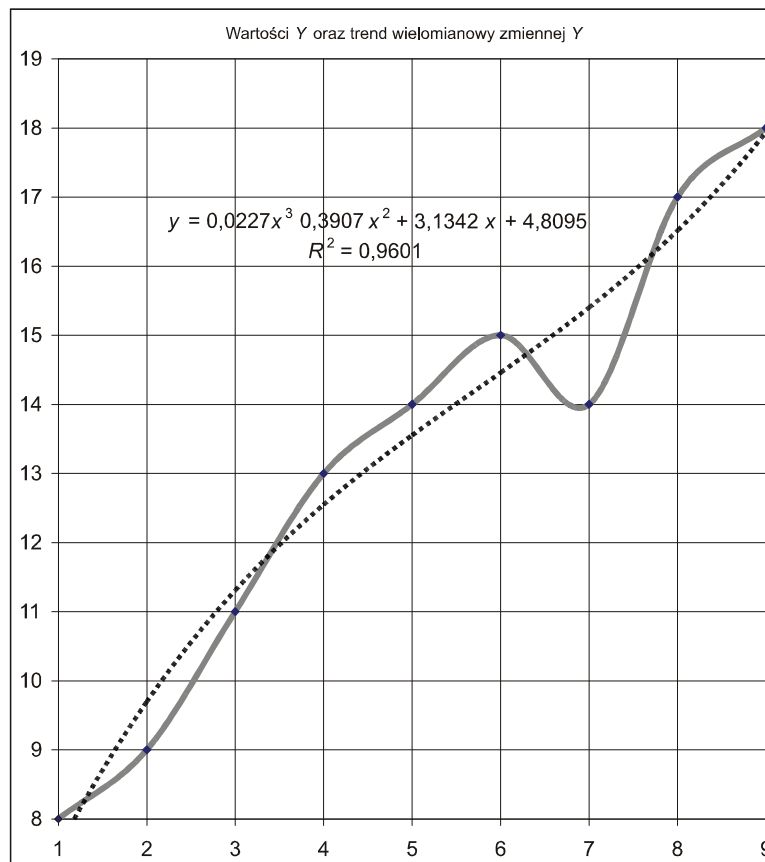
Wyrażamy podziękowania Panu Profesorowi Ryszardowi Broszkiewiczowi, bez którego inicjatywy i życzliwych spostrzeżeń nie byłoby tej pracy.

Szczególnie dziękujemy Pani Profesor Krystynie Hanusik za wnikliwą merytoryczną ocenę pracy, dzięki której można było dokonać istotnych poprawek oraz zmian w strukturze opracowania.

Autorzy

C z ę ś ć 1

Zagadnienia wstępne i wybrane modele ekonometryczne



1.1. Wprowadzenie

1.1.1. Pojęcia „ekonometria” i „model ekonometryczny”

Ekonometria to nauka o mierzeniu zależności między zjawiskami ekonomicznymi, przyrodniczymi, technicznymi, demograficznymi i socjologicznymi w celach poznawczych, symulacyjnych¹. Sprawne studiowanie ekonometrii wymaga znajomości podstaw makro- i mikroekonomii, statystyki oraz matematyki. Ekonometria bada przede wszystkim związki o charakterze ilościowym, występujące pomiędzy elementami zjawisk ekonomicznych, za pomocą metod statystyczno-matematycznych. Twórcami nowoczesnej ekonometrii są R. Frisch oraz J. Tinbergen – laureaci Nagrody Nobla z ekonomii.

Obeenie kładzie się nacisk na modelowy aspekt ekonometrii, przy czym:

- modelem ekonometrycznym nazywamy opis interesującego nas fragmentu ekonomicznej rzeczywistości, uwzględniający tylko istotne jej elementy. Charakterystyczną jego cechą jest wyodrębnienie z obiektywnie istniejącego systemu zasadniczych relacji występujących w badanym fragmencie rzeczywistości; zewnętrzną formą modelu ekonometrycznego jest układ równań;
- model ekonometryczny przedstawia zasadnicze powiązania występujące pomiędzy rozpatrywanymi zjawiskami ekonomicznymi.

Na potrzeby ekonometrii opracowano różnorodne oprogramowanie. W niezbędnym zakresie do celów dydaktycznych studentów studiów zaocznych kierunku zarządzanie wydaje się wystarczający zbiór narzędzi oferowany przez funkcje i opcje arkusza kalkulacyjnego Excel, pakiet WinQSB oraz program GRETL, który jest bezpłatnie dostępny w internecie.

1.1.2. Struktura modelu ekonometrycznego

Przez model ekonometryczny badanych zjawisk i procesów ekonomiczno-społecznych rozumiemy zapisaną w języku matematycznym formalną konstrukcję o określonej strukturze, przedstawiającą powiązania oraz zależności występujące między zjawiskami i procesami modelowanego systemu mierzalnej rzeczywistości, z dokładnością potrzebną w procesie wnioskowania. Ogólna postać modelu liniowego jest następująca:

¹ J. Dziechciarz (red.), *Ekonometria. Metody, przykłady, zadania*, Akademia Ekonomiczna, Wrocław 2003, s. 10.

$$\begin{aligned}
X_1 &= \sum_{j \neq 1}^m b_{1j} X_j + c_1, \\
X_2 &= \sum_{j \neq 1}^m b_{2j} X_j + c_2, \\
&\dots\dots\dots \\
X_h &= \sum_{j \neq 1}^m b_{hj} X_m + c_h,
\end{aligned}
\tag{1.1.1}$$

gdzie:

X_1, \dots, X_h – po lewej stronie – zmienne objaśniane (endogeniczne),

$X_j (X_{h+1}, \dots, X_m)$ po prawej stronie – zmienne objaśniające (egzogeniczne).

Przyjmijmy, że zmienne oznaczamy dużymi literami a wartości zmiennych poszczególnych obserwacji literami małymi. Zmienne endogeniczne i zmienne egzogeniczne mogą występować bez opóźnień czasowych lub mogą być zmiennymi opóźnionymi w czasie, tzn. opisującymi poziom zjawisk lub procesów we wcześniejszych okresach w stosunku do bieżącego.

Zmienne endogeniczne odnoszące się do tego samego okresu, dla którego równania modelu opisują badane zależności, nazywamy zmiennymi łącznie współzależnymi. Zmienne endogeniczne opóźnione w czasie i wszystkie egzogeniczne nazywamy zmiennymi z góry ustalonymi.

b_{ij} – parametry strukturalne modelu, które charakteryzują wewnętrzne powiązania między zmiennymi objaśnianymi i objaśniającymi;

c_j – elementy losowe, czyli reszty między wartościami z obserwacji a obliczonymi teoretycznie z funkcji danego modelu.

Elementy losowe wynikają z niemożności uwzględnienia w modelu wszystkich zmiennych wpływających na zmienną zależną, ponadto zjawiska ekonomiczno-społeczne zależą od czynników niezdeterminowanych. Istotne znaczenie ma także jakość materiałów empirycznych, na podstawie których model jest szacowany.

1.1.3. Kryteria klasyfikacji modeli ekonometrycznych

Podano przykładowo kryteria modeli: przyczynowo-skutkowych, symptomatycznych, tendencji rozwojowej. Zależność przyczynowo-skutkową między zmiennymi rozpoznaje się na drodze analizy jakościowej, prowadzącej do wykrycia wewnętrznej istoty badanego zjawiska i określenia jego logicz-

nych związków z innymi procesami. W modelach symptomatycznych nie występuje możliwość uwzględnienia zmiennych objaśniających, będących przyczyną kształtowania się zmiennej objaśnianej. Wówczas w rozpatrywanej relacji modelu zamiast zmiennej objaśniającej bezpośrednio oddziałującej na zmienną objaśnianą wprowadza się inną zmienną, przyczynowo nieoddziałującą na zmienną objaśnianą, skorelowaną jednak z rozpatrywaną zmienną objaśniającą. W modelach tendencji rozwojowej zakłada się, że prawidłowości dotyczące zmiennej endogenicznej są funkcją czasu t . Modele te służą do badania tendencji rozwojowej w określonym horyzoncie czasowym, wahań cyklicznych, a także wyjaśniania wahań przypadkowych zmiennej endogenicznej.

Z punktu widzenia występowania zmiennych w czasie modele ekonometryczne dzielimy na statyczne i dynamiczne. W modelu statycznym występujące w nim zmienne endogeniczne odnoszą się do tego samego okresu lub momentu. W modelu tym nie uwzględnia się ani zmiennej czasowej, ani określonej funkcji zmiennej reprezentującej trendy czasowe. Natomiast w modelach dynamicznych zmienne endogeniczne występują z opóźnieniami czasowymi.

Kryteria powiązań istniejących między zmiennymi endogenicznymi pozwalają wyodrębnić trzy klasy modeli: prosty, rekurencyjny, o równaniach współzależnych. W modelach prostych w roli zmiennych objaśniających występują tylko zmienne z góry ustalone. W modelach rekurencyjnych mogą istnieć tylko jednostronne zależności między nieopóźnionymi w czasie zmiennymi endogenicznymi. Jeśli chodzi o modele o równaniach współzależnych, to w ich konstrukcji nie istnieją żadne ograniczenia co do rodzaju, liczby i rozkładu powiązań między zmiennymi endogenicznymi.

W badaniach zjawisk i procesów ekonomiczno-społecznych za pomocą modeli ekonometrycznych równań opisowych wyróżniamy etapy:

- określenia celu badań modelowych wyodrębnionych procesów i zjawisk ekonomicznych;
- sformułowania modelowego systemu i otoczenia systemu badanych zjawisk i procesów gospodarczo-społecznych oraz wyszczególnienia elementów tego systemu i jego otoczenia;
- wyszczególnienia oraz doboru zmiennych, które będą występowały w modelu;
- wyboru klasy modelu ekonometrycznego i postaci analitycznej;
- estymacji, czyli oszacowania parametrów strukturalnych;
- weryfikacji modelu w celu sprawdzenia, czy określony model zapewnia pożądaną dokładność wyników;

- zastosowania modelu, np. w prognozowaniu zjawisk społeczno-gospodarczych.

Przykład

Model oparty na danych umownych, zwany modelem nierównowagi, w którym badane rynki są zbilansowane, czyli zmienna objaśniana X (podaż) jest równa zmiennej objaśnianej Z (popyt), a zmienne te są funkcją jednej zmiennej objaśniającej C (ceny). Zmniejszenie podaży z X_1 na X_2 uruchamia mechanizm powodujący spadek ceny i ustalenie jej na niższym poziomie równowagi.

Tabela 1.1.1

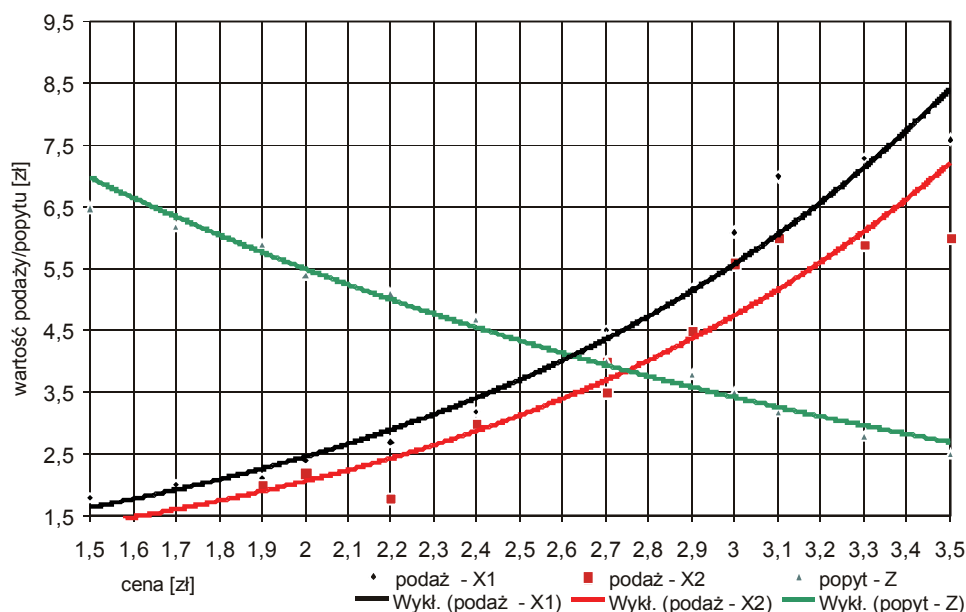
Dane umowne do wykresu zależności podaży i popytu od ceny

Podaż X_1	Popyt Z	Cena C	Podaż X_2
1,8	6,5	1,5	
2	6,2	1,7	
2,1	5,9	1,9	2
2,4	5,4	2	2,2
2,7	5,1	2,2	1,8
3,2	4,7	2,4	3
4	4,4	2,7	3,5
4,5	4	2,7	4
5,2	3,8	2,9	4,5
6,1	3,5	3	5,6
7	3,2	3,1	6
7,3	2,8	3,3	5,9
7,6	2,5	3,5	6

Źródło: Opracowanie własne w Excelu.

Do wystąpienia zakłóceń wystarczy zablokowanie jednego z mechanizmów regulujących, np. wprowadzenie cen minimalnych. Jeżeli rzeczywista cena będzie niższa od ceny równowagi, popyt będzie wyższy od podaży; gdy cena jest wyższa niż cena równowagi, podaż będzie wyższa od popytu.

Model zakładający możliwość wystąpienia niezbilansowania jest nazywany modelem nierównowagi, co wyklucza bezpośrednią obserwację popytu i podaży.



Źródło: Opracowanie własne w Excelu.

Wykres 1.1.1. Zależność podaży i popytu od ceny

1.1.4. Etapy budowy modelu ekonometrycznego

Badanie zależności między zjawiskami ekonomicznymi za pomocą modelu ekonometrycznego jest procesem wieloetapowym i wymaga²:

1. Ustalenia badanego zjawiska, co jest jednoznaczne z określeniem zjawiska wyjaśnianego przez model. W pierwszym etapie należy wybrać zmienne objaśniające (zob. rozdz. 1.6 niniejszego opracowania) spośród wielu czynników wpływających na zmienną objaśnianą.

2. Wyboru postaci analitycznej modelu, a więc określonej postaci matematycznej funkcji opisującej zależność zmiennej objaśnianej od zmiennych objaśniających.

3. Szacowania (estymacji) parametrów modelu, które sprowadza się do wyznaczenia wartości ocen poszczególnych parametrów.

4. Weryfikacji modelu, której celem jest sprawdzenie, czy zbudowany model dobrze opisuje badaną rzeczywistość ekonomiczną.

5. Wnioskowania na podstawie modelu, a więc jego praktyczne wykorzystanie. Może to być:

² E. N o w a k, *Zarys metod ekonometrii*, Wydawnictwo Naukowe PWN, Warszawa 2007, s. 8.

- analiza ekonomiczna,
- prognozowanie.

Przyjęcie określonej procedury badawczej i zastosowanie określonej metody badania zależy od typu modelu. Jak już wspomniano, pod względem możliwości poznawczych modele dzielimy na: przyczynowo-skutkowe, symptomatyczne, autoregresyjne, tendencji rozwojowej. Liczba równań modelu (jednorównaniowe, wielorównaniowe) odgrywa również rolę w procedurze jego formułowania. Istotne znaczenie ma także postać analityczna modelu (liniowe, nieliniowe).

W modelu ekonometrycznym wyróżniamy:

- parametry strukturalne, np. w modelu (1.1.2) są nimi $\alpha_0, \alpha_1, \dots, \alpha_k$,
- parametry struktury stochastycznej – rozkładu odchyłeń losowych (wartość oczekiwana, wariancja odchyłeń losowych, współczynniki autokorelacji odchyłeń).

W modelu liniowym w postaci ogólnej zmienna objaśniana Y jest liniową funkcją zmiennych objaśniających X i odchylenia losowego ε :

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + \varepsilon. \quad (1.1.2)$$

Gdy pominiemy ε , otrzymamy równanie na wartość modelową zmiennej objaśnianej Y :

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k. \quad (1.1.2)$$

Modele nieliniowe mogą być:

- nieliniowe względem zmiennych objaśniających i jednocześnie liniowe względem parametrów strukturalnych,
- nieliniowe względem zmiennych objaśniających oraz nieliniowe względem parametrów strukturalnych.

Modelom jednorównaniowym liniowym poświęcono rozdz. 1.7 i 1.8, a elementom budowy wielorównaniowych modeli liniowych rozdz. 1.9 niniejszej pracy. Przejścia z jednorównaniowego modelu nieliniowego w postać liniową dokonujemy poprzez transformację (zob. rozdz. 1.5).

W zależności od rodzaju danych statystycznych modele mogą być budowane na podstawie:

- danych przekrojowych, mających postać szeregów dotyczących obiektów ekonomicznych w jednej ustalonej jednostce czasu,
- danych czasowych, mających postać szeregów dynamicznych dotyczących jednego obiektu ekonomicznego w kolejnych jednostkach określonego przedziału czasu.

Przedstawione testy znajdują zastosowanie w przykładach zamieszczonych w rozdz. 1.7 oraz 1.8.

1.1.5. Testy oceniające dopasowanie modelu do danych empirycznych³

Weryfikując liniowy model ekonometryczny, można zbadać trzy własności:

- a) stopień zgodności modelu z danymi empirycznymi; do podstawowych stosowanych miar należą: odchylenie standardowe reszt, współczynnik zmienności losowej, współczynnik zbieżności oraz współczynnik determinacji (zostaną omówione w rozdz. 1.8),
- b) jakość ocen parametrów strukturalnych,
- c) rozkład odchyleń losowych.

1.1.5.1. Istotność współczynnika korelacji

Współczynnik korelacji obliczany jest na podstawie danych z próby⁴. Empiryczny (próbkowy) dla N obserwacji współczynnik korelacji R_{xy} , określający sposób oszacowania współczynnika korelacji dla całej populacji r_{xy} , obliczany jest według wzoru:

$$R_{xy} = \frac{\sum_{n=1}^N (y_n - \bar{y})(x_n - \bar{x})}{N \cdot S(y) \cdot S(x)}, \quad (1.1.4)$$

gdzie:

\bar{y} – średnia wartość y ,

\bar{x} – średnia wartość x ,

$S(y)$ – empiryczne odchylenie standardowe zmiennej Y ,

$S(x)$ – empiryczne odchylenie zmiennej X ,

n – numer kolejnej obserwacji.

Istotność współczynnika korelacji sprawdzamy poprzez weryfikację hipotez.

³ Ibidem, rozdz. *Weryfikacja modeli liniowych*.

⁴ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003, rozdz. 1.3: *Ocena istotności współczynnika korelacji*.

Hipoteza zerowa (nie istnieje zależność pomiędzy badanymi zmiennymi):

$$H_0 : r_{xy} = 0,$$

a hipoteza alternatywna:

$$H_1 : r_{xy} \neq 0.$$

Hipotezę weryfikujemy poprzez porównanie statystyki t_{obl} obliczonej z próby:

$$t_{obl} = R_{xy} \sqrt{\frac{N-2}{1-R_{xy}^2}} \quad (1.1.5)$$

z wartością teoretyczną (krytyczną), którą odczytujemy z tablicy rozkładu t -Studenta dla danego poziomu istotności α (np. 0,05) i wówczas oznaczamy jako t_α . Następnie przeprowadzamy badanie, czy:

$$|t_{obl}| \geq |t_\alpha|.$$

Jeżeli tak, hipotezę H_0 odrzucamy na rzecz hipotezy alternatywnej H_1 , ponieważ istnieje zależność pomiędzy badanymi zmiennymi Y oraz X w populacji generalnej.

1.1.5.2. Istotność współczynnika korelacji wielorakiej

Badanie to przeprowadzamy, aby stwierdzić, czy dopasowanie modelu do danych empirycznych jest dostatecznie duże. Weryfikujemy hipotezę zerową o istotności współczynnika korelacji wielorakiej, który jest pierwiastkiem kwadratowym współczynnika determinacji (badanie to nie jest konieczne, gdy występuje bardzo wysoki poziom współczynnika korelacji wielorakiej):

$$H_0 : (R = 0),$$

wobec hipotezy alternatywnej:

$$H_1 : (R \neq 0).$$

Sprawdzianem tej hipotezy jest statystyka:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k}, \quad (1.1.6)$$

gdzie:

k – ilość zmiennych objaśniających (określana przez niektórych autorów jako m); występujący w tym wzorze współczynnik determinacji ma postać:

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad (1.1.7)$$

gdzie:

y_t – zaobserwowana wartość empiryczna zmiennej objaśnianej,
 \hat{y}_t – wartość teoretyczna (modelowa) zmiennej objaśnianej.

Dopasowanie modelu do danych jest tym lepsze, im współczynnik determinacji jest bliższy 1.

Podana statystyka ma rozkład F (Fishera-Snedecora) o $m_1 = r_1 = k$ oraz $m_2 = r_2 = n - k - 1$ stopniach swobody. Z zamieszczonej w *Aneksie* tabl. V dla przyjętego poziomu istotności α oraz m_1 i m_2 stopni swobody odczytujemy wartość krytyczną F^* . Następnie przeprowadzamy badanie, czy:

$$F \leq F^*.$$

Jeżeli tak, to nie ma podstaw do odrzucenia hipotezy zerowej; współczynnik korelacji wielorakiej jest nieistotnie różny od zera, a dopasowanie modelu do danych jest zbyt słabe. Jeżeli:

$$F > F^*,$$

wówczas hipotezę zerową należy odrzucić na rzecz hipotezy alternatywnej; współczynnik korelacji wielorakiej jest istotny, a stopień dopasowania modelu do danych jest dostatecznie wysoki.

1.1.5.3. Istotność parametrów strukturalnych

Badanie istotności parametrów strukturalnych $\alpha_1, \alpha_2, \dots, \alpha_k$ liniowego modelu ekonometrycznego ma na celu sprawdzenie, czy zmienne objaśniające istotnie wpływają na zmienną objaśnianą. Weryfikuje się hipotezę zerową

$$H_0 : (a_i = 0)$$

wobec hipotezy alternatywnej:

$$H_1 : (a_i \neq 0).$$

Sprawdzianem tej hipotezy jest statystyka określana dla każdego $i = 1, 2, \dots, k$:

$$I_i = \frac{|a_i|}{S(a_i)}, \quad (1.1.8)$$

gdzie:

a_i – wartość oceny parametru strukturalnego α_i (po oszacowaniu parametrów strukturalnych liniowego modelu ekonometrycznego klasyczną metodą najmniejszych kwadratów KMNK, co omówiono w rozdz. 1.5),

$S(a_i)$ – standardowy błąd szacunku tego parametru (jako pierwiastek kwadratowy z wartości na głównej przekątnej macierzy wariancji i kowariancji ocen parametrów strukturalnych).

Do określenia wartości krytycznej I^* korzystamy z zamieszczonej w *Aneksie* tabl. I., gdzie P – poziom istotności α , n' – ilość stopni swobody obliczana jako $n - k - 1$. Następnie przeprowadzamy badanie, czy:

$$I_i \leq I^*.$$

Jeżeli tak, to nie ma podstaw do odrzucenia hipotezy zerowej; parametr strukturalny różni się nieistotnie od zera, a zmienna objaśniająca X_i nie wpływa w istotny sposób na zmienną objaśnianą Y . Jeżeli:

$$I_i > I^*,$$

wówczas hipotezę zerową odrzucamy na rzecz hipotezy alternatywnej; parametr a_i różni się w sposób istotny od zera, a zmienna objaśniająca X_i wpływa w sposób istotny na zmienną objaśnianą Y .

1.1.6. Podstawowe testy statystyczne do sprawdzania własności odchyłeń losowych⁵

Przed szacowaniem parametrów modelu ekonometrycznego przyjmujemy określone założenia dotyczące odchyłeń losowych ε . Po doborze postaci analitycznej modelu oraz zestawieniu zmiennych objaśniających należy dokonać analizy własności odchyłeń losowych, by ocenić trafność modelu. W tym celu przeprowadzamy badania: losowości, normalności rozkładu, nieobciążoności, autokorelacji, stałości wariancji. Weryfikację hipotez przeprowadza się na podstawie ciągu reszt e_t , będących oszacowaniem odchyłeń losowych modelu.

1.1.6.1. Losowość

Weryfikując hipotezę o losowości rozkładu odchyłeń losowych, oceniamy trafność doboru postaci analitycznej modelu. Do weryfikacji hipotezy zerowej:

$$H_0 : [\hat{Y} = f(X_1, X_2, \dots, X_k)]$$

⁵ E. N o w a k, *Zarys metod ekonometrii...*, op.cit., s. 8.

wobec hipotezy alternatywnej:

$$H_1 : [\hat{Y} \neq f(X_1, X_2, \dots, X_k)]$$

służy test serii. Punktem wyjścia jest ciąg reszt uszeregowanych według kolejności jednostek czasu, gdy model jest budowany na podstawie danych dynamicznych. Gdy model jest budowany na podstawie danych przekrojowych, reszty porządkuje się według rosnących wartości zmiennej objaśniającej, jeżeli jest jedna zmienna objaśniająca, lub według rosnących wartości wybranej zmiennej objaśniającej, w sytuacji modelu z wieloma zmiennymi objaśniającymi. Następnie oblicza się liczbę serii S reszt modelu. Serią jest każdy podciąg reszt złożonych wyłącznie z reszt dodatnich lub ujemnych. Z tablic testu serii (tabl. III w *Aneksie*) dla danej liczby reszt dodatnich n_1 , liczby reszt ujemnych n_2 oraz przyjętego poziomu istotności α (tj. dla $\alpha/2$ i $1 - \alpha/2$) odczytuje się dwie krytyczne liczby serii S_1^* oraz S_2^* . Dokonujemy badania. Jeżeli:

$$S_1^* < S < S_2^*,$$

to nie ma podstaw do odrzucenia hipotezy H_0 i oznacza to, że rozkład odchyleń jest losowy, a postać analityczna modelu została dobrana prawidłowo. Jeżeli:

$$S \leq S_1^* \text{ lub } S \geq S_2^*,$$

hipotezę H_0 należy odrzucić i przyjąć hipotezę H_1 . W tym wypadku rozkład odchyleń jest nielosowy, a postać modelu została dobrana nietrafnie.

1.1.6.2. Normalność rozkładu

Badanie normalności rozkładu odchyleń losowych sprowadza się do weryfikacji hipotezy, że dystrybuanta odchyleń $F(\varepsilon)$ jest równa dystrybuancie rozkładu normalnego $F_N(\varepsilon)$. Weryfikujemy więc hipotezę zerową:

$$H_0 : [F(\varepsilon) = F_N(\varepsilon)],$$

wobec hipotezy alternatywnej:

$$H_1 : [F(\varepsilon) \neq F_N(\varepsilon)].$$

Do weryfikacji hipotezy o normalności rozkładu służy m.in. test zgodności Hellwiga, a procedura testu jest następująca:

1. Przeprowadzenie standaryzacji reszt e_t , dla $t = 1, 2, \dots, n$:

$$u_t = \frac{e_t - \bar{e}}{\hat{S}_e}, \quad (1.1.9)$$

gdzie:

\bar{e} – średnia arytmetyczna reszt, która w modelu liniowym szacowanym według klasycznej metody najmniejszych kwadratów równa się zero,

\hat{S}_e – odchylenie standardowe reszt obliczane według wzoru:

$$\hat{S}_e = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n}}. \quad (1.1.10)$$

2. Porządkowanie zestandaryzowanych reszt według wartości malejących:

$$e_{(1)} \leq e_{(2)} \leq \dots e_{(n)}.$$

3. Odczytanie z tabl. VI (w *Aneksie*) wartości dystrybuanty rozkładu normalnego:

$$\Phi(u_{(i)}) = P(u < u_{(i)}).$$

Dla ujemnych wartości zestandaryzowanych reszt obliczamy dystrybuantę jako:

$$\Phi(u_{(i)}) = 1 - \Phi(u_{(i)}).$$

4. Wyznaczenie tzw. cel I_i , którymi są przedziały liczbowe o rozpiętości $1/n$ powstałe po podzieleniu odcinka $[0, 1]$ na n równych części.

5. Przyporządkowanie wartości dystrybuanty $\Phi(u_{(i)})$ odpowiednim celom i określenie liczby cel pustych K .

6. Odczytanie z tabl. VII w *Aneksie* dla danej liczby obserwacji n przyjętego poziomu istotności α wartości krytycznych oraz K_1 i K_2 .

7. Przeprowadzenie badania. Jeżeli:

$$K_1 \leq K \leq K_2,$$

to nie ma podstaw do odrzucenia hipotezy H_0 ; odchylenia losowe mają zatem rozkład normalny. Jeżeli:

$$K < K_1 \text{ lub } K > K_2,$$

wówczas hipotezę zerową H_0 odrzucamy na rzecz hipotezy alternatywnej H_1 ; odchylenia losowe nie mają wtedy rozkładu normalnego.

Badanie normalności rozkładu odchyłeń losowych możemy także przeprowadzić, korzystając z testu Jarque'a-Bery, który opisany jest według formuły⁶:

$$JBT = \frac{N}{6} \left[\beta_1 + \frac{1}{4}(\beta_2 - 3)^2 \right], \quad (1.1.11)$$

⁶ M. Osińska (red.), *Ekonometria współczesna*, TNOiK, Toruń 2007.

gdzie:

N – liczność próby,

β_1 – kwadrat współczynnika skośności (S^2), przy czym S obliczamy według wzoru:

$$S = \sqrt{\beta_1} = \frac{1}{N} \sum_{i=1}^N \frac{e_i^3}{S^3(e_i)}, \quad (1.1.12)$$

gdzie:

e_i – reszty modelu,

$S(e_i)$ – standardowy błąd reszt obliczony według wzoru:

$$S(e_i) = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}, \quad (1.1.13)$$

β_2 – współczynnik kurtozy (określany również jako K):

$$\beta_2 = \frac{1}{N} \sum_{i=1}^N \frac{e_i^4}{S^4(e_i)}. \quad (1.1.14)$$

We wzorach na współczynnik skośności oraz współczynnik kurtozy przyjęto wartość średnią reszt modelu równą zero.

Badanie polega na porównaniu rozkładu składnika losowego z próby z postulowanym teoretycznym rozkładem normalnym.

Hipoteza zerowa (rozkład składnika losowego jest rozkładem normalnym):

$$H_0 : F(e_i) = F_N(e_i).$$

Hipoteza alternatywna (rozkład składnika losowego nie jest rozkładem normalnym):

$$H_1 : F(e_i) \neq F_N(e_i).$$

Rozkład statystyki JBT jest zbieżny do rozkładu $\chi_a^2(2)$ przy określonym poziomie istotności α oraz dwóch stopniach swobody.

Sprawdzamy, czy:

$$JBT < \chi_a^2(2).$$

Jeżeli tak, to nie ma podstaw do odrzucenia hipotezy H_0 ; rozkład składnika losowego jest normalny. Jeżeli:

$$JBT \geq \chi_a^2(2),$$

wówczas odrzucamy H_0 na rzecz H_1 .

Nieco inna formuła testu Jarque'a-Bery, różniąca się liczbą szacowanych parametrów k , występuje w publikacji M. Lipiec-Zajchowskiej⁷:

$$JBT = \frac{N-k}{6} \left[S^2 + \frac{1}{4}(K-3)^2 \right], \quad (1.1.15)$$

gdzie:

N – liczność próby,

S – współczynnik skośności,

K – współczynnik kurtozy.

1.1.6.3. Nieobciążoność

Dla modeli nieliniowych, ze względu na parametry strukturalne, przeprowadzamy badanie nieobciążoności odchyłeń losowych. Sprowadza się ono do zweryfikowania hipotezy zerowej:

$$H_0 : [E(\varepsilon) = 0],$$

wobec hipotezy alternatywnej:

$$H_1 : [E(\varepsilon) \neq 0].$$

Sprawdzianem jest statystyka:

$$I = \frac{|\bar{e}| \sqrt{n-1}}{\hat{S}_e}, \quad (1.1.16)$$

gdzie:

\bar{e} – średnia arytmetyczna reszt, która w modelu liniowym szacowanym według klasycznej metody najmniejszych kwadratów równa się zero,

\hat{S}_e – odchylenie standardowe reszt obliczane według wzoru (1.1.10).

Z tablicy I dla $m = n - 1$ stopni swobody (oznaczonej jako n) oraz dla przyjętego poziomu istotności α odczytujemy wartość krytyczną I^* . Przeprowadzamy badanie. Jeżeli:

$$I \leq I^*,$$

to nie ma podstaw do odrzucenia hipotezy H_0 ; wartość oczekiwana odchyłeń losowych jest nieistotnie różna od zera i odchylenia te są nieobciążone. Jeżeli:

$$I > I^*,$$

⁷ M. Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 2. *Ekonomia...*, op.cit.

wtedy hipotezę zerową H_0 odrzucamy na rzecz hipotezy alternatywnej H_1 ; wartość oczekiwana odchyłeń losowych jest istotnie różna od zera i odchylenia losowe są obciążone.

1.1.6.4. Autokorelacja

Celem badania autokorelacji odchyłeń losowych jest sprawdzenie jednego z założeń klasycznej metody najmniejszych kwadratów (KMNK), tj. czy autokorelacja odchyłeń losowych nie występuje⁸. Miarą siły i kierunku autokorelacji odchyłeń losowych ε_t z okresu t i odchyłeń losowych $\varepsilon_{t-\tau}$ z okresu $t-\tau$ jest współczynnik korelacji:

$$\rho_\tau = \rho(\varepsilon_t, \varepsilon_{t-\tau}),$$

nazywany współczynnikiem autokorelacji reszt rzędu τ . Oszacowaniem tego współczynnika jest współczynnik autokorelacji reszt ε_t i $\varepsilon_{t-\tau}$ określony wzorem:

$$r_\tau = \frac{\sum_{t=\tau+1}^n (e_t - \bar{e}_t)(e_{t-\tau} - \bar{e}_{t-\tau})}{\sqrt{\sum_{t=\tau+1}^n (e_t - \bar{e}_t)^2 \sum_{t=\tau+1}^n (e_{t-\tau} - \bar{e}_{t-\tau})^2}}, \quad (1.1.17)$$

gdzie:

$\bar{e}_t, \bar{e}_{t-\tau}$ – średnie arytmetyczne odpowiednich ciągów reszt dla $n - \tau$ obserwacji.

Ponieważ wartość oczekiwana składnika losowego wynosi zero (a w przypadku reszt otrzymanych dla modelu szacowanego KMNK średnia reszt także wynosi zero), współczynnik autokorelacji pierwszego rzędu w próbie r_1 możemy określić ze wzoru:

$$r_1 = \frac{\sum_{t=2}^n e_t \bar{e}_{t-1}}{\sqrt{\sum_{t=2}^n e_t^2 \sum_{t=2}^n e_{t-1}^2}} \approx \frac{\sum_{t=2}^n e_t \bar{e}_{t-1}}{\sum_{t=2}^n e_t^2}. \quad (1.1.18)$$

Dotyczy to dużej próby, dla której $e_t^2 \cong e_{t-1}^2$,

gdzie:

e_t – reszty z okresu bieżącego,

e_{t-1} – reszty opóźnione o jeden okres.

⁸ D. Strahl D. (red.), *Modelowanie ekonometryczne z Excelem*, Akademia Ekonomiczna, Wrocław 2002.

Do weryfikacji hipotezy zerowej:

$$H_0 : (\rho_1 = 0),$$

o braku autokorelacji odchyłek losowych ε_t i ε_{t-1} , wobec hipotezy alternatywnej:

$$H_1 : (\rho_1 \neq 0),$$

służy test Durбина-Watsona, a sprawdzianem tej hipotezy w przypadku autokorelacji dodatniej ($r_1 > 0$) jest statystyka:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad (1.1.19)$$

Test Durбина-Watsona służy do badania autokorelacji pierwszego rzędu przy założeniach:

- model musi zawierać wyraz wolny,
- składnik losowy powinien mieć rozkład normalny,
- w roli zmiennej objaśniającej nie może występować zmienna endogeniczna opóźniona w czasie.

W tablicy IV (Durбина-Watsona) w *Aneksie* dla przyjętego poziomu istotności α i danej liczby obserwacji n oraz liczby zmiennych objaśniających k – bez stałej c (w niektórych publikacjach k określane jest jako m) – odczytuje się dwie wartości krytyczne: d_L i d_U . Następnie przeprowadzamy badanie, czy:

$$d \geq d_U.$$

Jeżeli tak, to nie ma podstaw do odrzucenia hipotezy zerowej; współczynnik autokorelacji ρ_1 jest nieistotny (nie występuje autokorelacja pierwszego rzędu składnika resztowego). Jeżeli:

$$d \leq d_L,$$

hipotezę H_0 należy odrzucić na rzecz hipotezy alternatywnej H_1 ; współczynnik autokorelacji ρ_1 jest istotny (co oznacza występowanie w modelu autokorelacji pierwszego rzędu składnika losowego). Jeżeli:

$$d_L \leq d \leq d_U,$$

wówczas przy użyciu testu Durбина-Watsona nie można podjąć żadnej decyzji i należy zastosować inny test (np. na istotność współczynnika korelacji).

W przypadku autokorelacji ujemnej ($r_1 < 0$) oblicza się wartość statystyki $d' = 4 - d$, a postępowanie dalsze jest analogicznie do korelacji dodatniej.

Za pomocą testu istotności współczynnika autokorelacji możemy badać istotność autokorelacji odchyłeń losowych dowolnego rzędu, a test ten umożliwia zweryfikowanie hipotezy zerowej:

$$H_0 : (\rho_\tau = 0),$$

wobec hipotezy alternatywnej:

$$H_1 : (\rho_\tau \neq 0).$$

Sprawdzianem tej hipotezy jest statystyka:

$$I_\tau = \frac{|r_\tau| \sqrt{n - \tau - 2}}{\sqrt{1 - r_\tau^2}}. \quad (1.1.20)$$

Na podstawie tablicy I testu *t*-Studenta (*Aneks*) w przyjętym poziomie istotności α oraz $m = n - \tau - 2$ stopnie swobody, gdzie np. $\tau = 1$, odczytuje się wartości krytyczne I^* . Następnie przeprowadzamy badanie, czy:

$$I_\tau \leq I^*.$$

Jeżeli tak, to nie ma podstaw do odrzucenia hipotezy H_0 ; współczynnik autokorelacji ρ_τ jest nieistotny. Jeżeli:

$$I_\tau > I^*,$$

hipotezę H_0 należy odrzucić na rzecz hipotezy H_1 ; współczynnik autokorelacji ρ_τ jest istotny.

1.1.6.5. Stałość (jednorodność) wariancji składnika losowego (oparta na teście Fishera-Snedecora)

Rozpatrzmy dwa podzbiory obserwacji, co do których istnieje przypuszczenie, że wariancja odchyłeń jest najmniejsza i największa. W pierwszym podzbiory o liczbie obserwacji n_1 są obserwacje $t = 1, 2, \dots, n_1$, a w drugim, n_2 , są obserwacje $n - n_2 + 1, n - n_2 + 2, \dots, n$.

Do zweryfikowania hipotezy zerowej o równości wariancji odchyłeń losowych obu podzbiorów:

$$H_0 : (\sigma_{\varepsilon,1}^2 = \sigma_{\varepsilon,2}^2),$$

wobec hipotezy alternatywnej:

$$H_1 : (\sigma_{\varepsilon,1}^2 < \sigma_{\varepsilon,2}^2),$$

wykorzystuje się test F (Fishera-Snedocora), którego sprawdzianem jest statystyka:

$$F = \frac{S_{e,2}^2}{S_{e,1}^2}, \quad (1.1.21.)$$

przy czym wariancja resztowa n_1 pierwszych obserwacji określona jest według wzoru:

$$S_{e,1}^2 = \frac{\sum_{t=1}^n (e_t - \bar{e}_1)^2}{n_1 - k - 1}, \quad (1.1.22)$$

a wariancja resztowa n_2 ostatnich obserwacji według:

$$S_{e,2}^2 = \frac{\sum_{t=1}^n (e_t - \bar{e}_2)^2}{n_2 - k - 1},$$

gdzie:

\bar{e}_1, \bar{e}_2 – średnie arytmetyczne odpowiednich ciągów reszt.

W tablicy V (*Aneks*) w przyjętym poziomie istotności α oraz $m_1 = r_1 = n_2 - k - 1$ i $m_2 = r_2 = n_1 - k - 1$ stopni swobody odczytujemy wartość krytyczną F^* . Następnie przeprowadzamy badanie, czy:

$$F \leq F^*.$$

Jeżeli tak, to nie ma podstaw do odrzucenia hipotezy zerowej H_0 ; stwierdza się, że wariancje z podzbiorów nieistotnie się od siebie różnią; wariancja składnika losowego jest jednorodna (stała w czasie). Jeżeli:

$$F \geq F^*,$$

wówczas hipotezę zerową należy odrzucić na rzecz hipotezy alternatywnej; wariancje z prób (podzbiorów) istotnie się od siebie różnią (wariancja składnika losowego jest niejednorodna); w miarę upływu czasu wariancja odchyleń losowych wzrasta.

Statystyka F wymaga, aby w liczniku znalazła się większa wartość estymatora składnika losowego spośród $S_{e,1}^2$ i $S_{e,2}^2$.

1.1.7. Pytania

1. Podaj definicję ekonometrii.
2. Co nazywamy modelem ekonometrycznym?

3. Wymień klasy modeli ekonometrycznych.
4. Czym charakteryzują się modele symptomatyczne?
5. Wymień etapy budowy modelu ekonometrycznego.
6. Wymień testy oceny dopasowania modelu do danych empirycznych.
7. Wymień testy do sprawdzania własności odchyleń losowych.
8. Podaj procedurę postępowania przy badaniu losowości rozkładu odchyleń losowych.
9. Do czego służy test zgodności Hellwiga?

1.2. Oprogramowanie mikrokomputerowe do ekonometrii

1.2.1. Uwagi wstępne

Ekonometria jako odrębna dyscyplina naukowa rozwija się coraz prężniej. Powstają nowe metody i modele. Doskonalony jest aparat teoretyczny do estymacji i weryfikacji modeli. Na naszym rynku wydawniczym pojawiają się publikacje sugerujące stosowanie nowoczesnej techniki obliczeniowej. Obszerny dorobek z zakresu oprogramowania (znany już w wiodących gospodarczo krajach) przenika do wydawanych u nas pozycji książkowych i coraz szerzej stosowany jest na uczelniach ekonomicznych. Nieliczne są jednak publikacje poświęcone w całości rozwiązywaniu problemów z zastosowaniem określonego pakietu komputerowego. Często występuje tylko nawiązanie do programów lub zastosowania podane są fragmentarycznie w formie dołączonego aneksu.

G.S. Maddala⁹ w swoim podręczniku do ekonometrii stwierdza, że nie ma jednak pakietu, który by objął wszystkie tematy omówione w jego pracy. Wiele jest programów o różnorodnych symbolach, które w swej funkcjonalności nieznacznie różnią się między sobą oraz dotyczą metod i testów podstawowych.

Autorzy chcieliby, aby niniejszy materiał był zainicjowaniem czegoś w rodzaju przewodnika po publikacjach i pakietach programowych, zarówno tych bezpłatnych, ogólnie dostępnych, jak i zamkniętych, wymagających indywidualnych licencji. Jest to bardzo obszerna tematyka znajdująca swoje fragmentaryczne odzwierciedlenie w internecie w przykładach zastosowań. Wystarczy sięgnąć do strony [www: \[feweb.vu.nl/economet/riclinks/software/software.html\]](http://feweb.vu.nl/economet/riclinks/software/software.html), aby się przekonać o ogromnym dorobku oprogramowania rozwijanego w miarę postępu w zakresie systemów operacyjnych, programów narzędziowych i sieciowych. Strona ta zawiera linki do dystrybutorów określonego oprogramowania. W niniejszym opracowaniu dokonano przeglądu dostępnych pozycji literaturowych pod kątem zastosowania elektronicznej techniki obliczeniowej. Dalsze penetrowanie możliwości zainstalowania pro-

⁹ G.S. Maddala, *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2006.

gramów, jak i rozpoznania ich funkcjonalności pozostawia się studiującym, tak by ekonometria nie kojarzyła się tylko z rozważaniami teoretycznymi i mnóstwem wzorów.

1.2.2. Programy komputerowe

Omówienie w niniejszym materiale obszarów zastosowania całej palety programów komputerowych przeznaczonych do ekonometrii nie jest możliwe. Ponadto nie wszystkie stosowane są w naszym kraju. Z tego względu oparto się na programach przykładowych, powszechnie dostępnych.

1.2.2.1. GRETL

Oprogramowanie GRETL (*GNU Regression, econometric and Time-series Library*) opracowane jest w wersji polskiej. Dostępne jest na stronie internetowej [http://www \[gretl.sourceforge.net\]](http://www.gretl.sourceforge.net) lub [\[kufel.torun.pl\]](http://kufel.torun.pl). Zawiera podstawowe tematy z ekonometrii oraz prognozowania i symulacji. Publikacja T. Kufla obejmuje szereg tematów dotyczących możliwości zastosowania menu oprogramowania GRETL-a na bazie plików danych oferowanych w internecie przez tegoż autora¹⁰. Oprogramowanie wydane jest bezpłatnie przez Fundację Wolnego Oprogramowania (*Free Software Foundation*). Jego kod źródłowy może być przez użytkowników modyfikowany, jednak ryzyko skuteczności działania funkcji ponosi sam użytkownik. W trakcie pisania tej książki lansowana była wersja GRETL-a nr 1.6.5. Procedurę instalacji pliku [gretl.install.exe] oraz plików z danymi pochodzącymi z podręczników ekonometrii podano w publikacji T. Kufla¹¹. Menu programu obejmuje:

- plik – operacje na plikach,
- narzędzia – funkcje odczytu wartości krytycznych oraz prezentacji graficznej rozkładów,
- dane – operacje umożliwiające edycję zmiennych na potrzeby modelu,
- widok – zestaw funkcji obejmujących m.in. budowę wykresów, szacowanie statystyk,
- dodawanie zmiennych – tworzenie nowych zmiennych na bazie istniejących,
- próba – określenie zakresu obserwacji z pliku danych,
- zmienna – udostępnianie statystyk opisowych, rozkładów i ich zgodności,

¹⁰ T. K u f e l, *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2007.

¹¹ Ibidem, rozdz. 1.2.

– model – obszerny zestaw metod estymacji modeli ekonometrycznych, szeregów czasowych, nieliniowych, równań współzależnych, panelowych.

Oprogramowanie GRETL-a ma dostęp do serwera [ricardo.ecn.wfu.edu] poprzez menu: (*Plik/Pliki baz danych/Instalowanie baz danych*)¹². Aby pracować z GRETL-em, pozyskany plik należy zapisać na dysku własnego komputera. W ramach programu GRETL występują przykłady modeli ekonometrycznych, a część z nich opatrzona jest zestawem poleceń-komend. Wbudowano również tablice statystyczne rozkładów: normalnego, *t*-Studenta, χ^2 , *F*, Durбина-Watsona. Ułatwieniem jest kalkulator testów statystycznych pozwalający wyznaczyć wielkość statystyki testowej i empiryczne poziomy istotności do weryfikacji modelu. Wprowadzane informacje o parametrach kalkulatora testów są wartościami liczbowymi.

Oprogramowanie GRETL-a w zakresie ekonometrycznych modeli do danych przekrojowych umożliwia zarówno dobór zmiennych, jak i estymację parametrów klasyczną metodą najmniejszych kwadratów (KMNK) oraz przeprowadzenie etapów weryfikacji modelu. W doborze zmiennych objaśniających stosuje się funkcje wyznaczania macierzy korelacji. Oprócz KMNK w programie wykorzystuje się także inne metody estymacji, będące modyfikacją metody najmniejszych kwadratów. Bardzo pomocne są informacje z procesu estymacji oraz specjalne funkcje GRETL-a do weryfikacji modelu, służące do oceny:

- istotności parametrów strukturalnych,
- stopnia dopasowania modelu,
- normalności rozkładu reszt,
- jednorodności wariancji reszt,
- liniowości postaci analitycznej,
- współliniowości zmiennych objaśniających,
- wpływowych obserwacji.

Pomocną cechą programu GRETL jest zapisanie wyników modelowania, grafiki i wykonywanych komend w formie pliku sesji. Analizę ekonometryczną procesów ekonomicznych ułatwiają uzyskane z tego programu charakterystyki, w tym funkcja autokorelacji i korelacji cząstkowej. W pracy T. Kufla¹³ na uwagę zasługuje nawiązanie do krótkiego wyjaśnienia teoretycznego podstawowych modeli procesów ekonomicznych z przedstawieniem na konkretnym pliku danych odpowiedniej funkcji GRETL-a. Daje to możliwość zrozumienia raportów programu z realizacji budowy modeli: wielomianowych trendu, wahań sezonowych, autoregresyjnych.

¹² Ibidem, rozdz. 2.10.

¹³ Ibidem.

Na podstawie modelu trendu i sezonowości przy skorzystaniu z funkcji (*Analiza/Prognoza*) możemy wyznaczyć prognozy i oszacować błędy predykcji każdego prognozowanego okresu. Gdy nie są spełnione założenia o stosowaniu KMNK, tj. występuje autokorelacja reszt lub heteroskedastyczność wariancji (niestalność), to wówczas stosowana jest uogólniona metoda najmniejszych kwadratów (UMNK). W warunkach autokorelacji reszt program GRETL proponuje w ramach modeli szeregów czasowych metody: Cochrane-Orcutta; Hirdertha-Lu; Prais-Winstena.

Istnienie odstających (nietypowych, wpływowych) obserwacji jest przyczyną występowania heteroskedastyczności wariancji resztowej. Niespełnienie warunku stałości wariancji oceniamy, stosując KMNK i test White'a, posługując się funkcją (*Test/Test heteroskedastyczności*) GRETL-a. Gdy wynik testu wskaże na istotność zmienności wariancji, wówczas, w celu lepszej estymacji parametrów, stosujemy UMNK. W wersji polskiej nr 1.6.5 można rozwiązywać również modele nieliniowe, a w ich ramach: logitowe, probitowe, tobitowe (określone na podstawie modeli logitowych i probitowych).

Godna uwagi jest możliwość określania wielowymiarowych modeli ekonometrycznych, które charakteryzują wiele zjawisk ekonomicznych. Do estymacji równań tych modeli służy podwójna metoda najmniejszych kwadratów (2MNK), wywoływana z funkcji (*Modele/Inne liniowe modele/Podwójna metoda najmniejszych kwadratów*).

Niekiedy brakuje teoretycznych podstaw do wskazywania współzależności procesów, a więc rodzą się wątpliwości co do celowości modelowania strukturalnego. W związku z tą niepewnością opracowano modele VAR, czyli wielorównaniowe, w których zestaw zmiennych objaśniających obejmuje procesy opóźnione w czasie. Modele te były doskonalone poprzez wprowadzenie trendu i sezonowości. Do oszacowania modelu VAR korzystamy z menu (*Model/Modele szeregów czasowych/Model wektorowej autoregresji*).

Z danymi panelowymi, np. przestrzennie-czasowymi, mamy do czynienia, gdy obserwujemy dane na kilku obiektach, obszarach i okresach czasowych. T. Kufel jako przykład podaje „[...] badanie ekonomicznej efektywności ludności [...], stopy bezrobocia w powiatach w Polsce w latach 1998–2002” oraz rozwiązuje ten przykład w GRETL-u¹⁴.

1.2.2.2. EViews

Statystyczny pakiet komputerowy *Econometric Views* (EViews) posłużył M. Lipiec-Zajchowskiej do przeprowadzenia analiz, których wyniki są w for-

¹⁴ Ibidem, rozdz. 12.

macie tego pakietu¹⁵. W aneksie tej publikacji zawarto opisy komputerowych programów Excel oraz EViews. Wyniki analiz dyskryminacyjnych otrzymano, stosując pakiet *Statistica*. EViews posłużył do budowy liniowych modeli regresji z jedną zmienną objaśniającą (niezależną). W ramach tej czynności oszacowano współczynniki korelacji, estymacji parametrów modelu, otrzymując podstawowe statystyki opisujące dopasowania modelu do danych empirycznych; uzyskano m.in.:

- współczynniki przy zmiennych objaśniających,
- błędy standardowe oszacowania parametrów,
- wartości testu *t*-Studenta,
- współczynnik determinacji, skorygowany współczynnik determinacji,
- odchylenie standardowe składnika resztowego,
- wartość testu Durбина-Watsona,
- średnią zmienną zależną, odchylenie standardowe zmiennej zależnej,
- wartość testu Fishera, poziom istotności statystyki *F*.

W wymienionej publikacji użyto EViews również do oszacowania parametrów liniowego dwuwymiarowego modelu ekonometrycznego, uzyskując także statystykę dopasowania modelu. Program zastosowano także do wielowymiarowych modeli regresji, gdy liczba zmiennych objaśniających jest większa niż dwie. W publikacji M. Lipiec-Zajchowskiej¹⁶ spotykamy przykład oszacowania z użyciem EViews nieliniowego wielowymiarowego modelu regresji z opóźnioną zmienną endogeniczną Y_{t-1} . We wspomnianej publikacji, oprócz zastosowania pakietu do modeli przyczynowo-skutkowych, spotykamy również zastosowanie EViews do modeli rozwoju zjawiska w czasie. Interesujący jest przykład doboru funkcji trendu spośród funkcji liniowej, potęgowej oraz wykładniczej. Gdy występuje sezonowość, w kwartałach możemy wprowadzić zmienne zero-jedynkowe i wówczas estymujemy parametry modelu obejmującego czas oraz zmienne. Zastosowanie to dotyczy zarówno modeli multiplikatywnych, jak i addytywnych. Zjawiska w czasie opisywane są w celach prognozowania za pomocą modeli:

- prostych średnich ruchomych, np. dwu- lub czterookresowych,
- ważonych średnich ruchomych,
- prostego wyrównywania wykładniczego z parametrem wyrównującym α ,
- adaptacyjnego wyrównywania wykładniczego z sygnałem adaptacyjnym TST_{α} ,

¹⁵ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria...*, op.cit.

¹⁶ Ibidem, rozdz. 1.14.2.

- podwójnych średnich ruchomych,
- podwójnego wyrównywania wykładniczego – model Browna,
- Holta-Wintersa (bez sezonowości, z sezonowością multiplikatywną lub addytywną).

Omawiany pakiet umożliwia zbudowanie modeli wyrównywania wykładniczego (prostego, podwójnego) oraz Holta-Wintersa. Szybka realizacja komputerowa daje szansę wyboru adekwatnego modelu, jak i najlepszych parametrów z punktu widzenia miary błędu wygasłych już prognoz, jaką jest *RMSE* (pierwiastek średniego kwadratu błędu).

Modele autoregresyjne, stosowane do prognoz krótkoterminowych, wykorzystujemy wtedy, gdy utrudnione jest zebranie danych dotyczących zmiennych objaśniających. W klasycznych modelach autoregresyjnych, np. gdy zmienne objaśniające stanowią zmienne opóźnione o 1, 2, 3 okresy, estymujemy parametry w EViews, stosując KMNK. Warto dodać, że, oprócz estymacji parametrów modelu, możemy, używając tego pakietu, przeprowadzić (podobnie jak w programie GRETL) badania:

- istotności parametrów (test *t*-Studenta),
- autokorelacji (test LM),
- normalności reszt (test JB),
- stałości wariancji (test White'a).

Na podstawie analizy wyników testów decydujemy o odrzuceniu lub nie modelu oszacowanego na podstawie KMNK, biorąc pod uwagę kryteria założeń tej metody.

Gdy mamy do czynienia ze stacjonarnym szeregiem czasowym, który charakteryzuje się stałą średnią i wariancją, i kowariancją zależną od rzędu opóźnienia p zmiennej objaśnianej Y , a nie od czasu t , to budujemy modele Boxa i Jenkinsa. Do badania stacjonarności szeregu stosujemy funkcję autokorelacji lub rozszerzony test Dickey-Fullera zawarty w programie EViews. Jeśli szereg jest niestacjonarny, to, na podstawie różnic rzędu d zmiennej objaśnianej Y , budowane są modele średnich ruchomych, a także modele mieszane. Po usunięciu, przykładowo, z modelu mieszanego nieistotnych zmiennych otrzymujemy poprawny statystycznie model, nadający się do praktycznego stosowania.

Odrębną grupę stanowią zastosowania EViews do estymacji parametrów modeli ze zmiennymi dychotomicznymi, gdy Y ma wartość zero-jedynkową oraz występuje zestaw cech jakościowych X . Poszukujemy związku między X a prawdopodobieństwem realizacji określonego zdarzenia wyrażonego przez Y . W zależności od formuły wyrażającej tę zależność występują modele, któ-

rych parametry możemy oszacować, również używając programu EViews (niestety, nie jest ogólnie dostępny w internecie):

- liniowy model prawdopodobieństwa (LPM),
- logit (logarytm naturalny współczynnika szansy),
- probit, wykorzystujący dystrybuantę rozkładu normalnego.

Procedura zastosowania pakietu EViews w wersji 3.11 z przykładami ekranów menu znajduje się w 1 rozdziale *Aneksu*¹⁷. Możemy dowiedzieć się, jak: wprowadzać dane do modelu regresji, wykonać odpowiednie wykresy, przeprowadzić operacje na szeregach danych, dokonać estymacji parametrów, sporządzić wykres rozrzutu reszt oraz otrzymać miary statystyki opisowej. EViews jest też pomocny w modelach trendu. Możemy prognozować, bazując na modelu autoregresyjnym czy też wyrównywania wykładniczego. Opcje programu (*Histogram i statystyki opisowe/Szereg rozdzielczy*) dają nam materiał do podstawowej analizy statystycznej. Warto jeszcze zaznaczyć, że pracując w systemie operacyjnym Windows, mamy możliwość dokonania importu/eksportu plików między EViews a arkuszem kalkulacyjnym Excel lub edytorem tekstów Word. Zastosowanie omawianego pakietu spotykamy również u E. Lipiec-Zajchowskiej¹⁸. W *Aneksie* tegoż tomu podano sposób korzystania z menu do pozyskania statystyki opisowej oraz następujących wykresów¹⁹:

- liniowego, liniowego warstwowego,
- słupkowego, słupkowego warstwowego, liniowo-słupkowego,
- korelacyjnego,
- kołowego.

W ramach statystyki opisowej otrzymujemy określenie: średniej wartości zmiennej, mediany, wartości maksymalnej oraz minimalnej, odchylenia standardowego, współczynnika skośności i kurtozy, testu Jarque'a-Bery.

1.2.2.3. WinQSB

Pakiet *Windows Quantitative System for Business* (WinQSB) służy do rozwiązywania zadań z programowania matematycznego. Pliki tego programu oraz jego opis uzyskujemy w internecie przez wywołanie hasła WinQSB w domenie [//www.onet.pl]²⁰. Spośród różnych możliwości wejścia wybieramy *Zakład Metod Ilościowych w Zarządzaniu*. Opis instalujemy na własnym

¹⁷ Ibidem.

¹⁸ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 1: *Statystyka*, Wydawnictwo C.H. Beck, Warszawa 2003.

¹⁹ Ibidem, *Aneks*, rozdz. 1.

²⁰ Dostęp w internecie może ulec zmianie. Podana procedura dotyczy okresu pisania niniejszego opracowania.

dysku jako plik [winqsb.doc] o pojemności 30 Mb (34 strony formatu A4), a w formie skompresowanej 1,8 Mb. Forma programu zależy od źródła pozyskania, np. zapisany jest na trzech plikach: [winqsb(2).exe], [winqsb.r00], [winqsb.r01]. Pliki programu instalujemy również na dysku własnego komputera. Opis programu (*Rozwiązywanie zadań za pomocą pakietu WinQSB*) możemy pobrać również z publicznej domeny o symbolu dostępu [miz.marc.pl]. Wyszczególnione są w nim funkcje i elementy pozycji menu, mianowicie:

- *File* (plik): nowy problem, otwarcie istniejącego problemu, zapisywanie problemu, wyjście z programu,
- *Edit* (edycja): wycięcie, kopiowanie, wklejenie,
- *Format*: format zapisu liczbowego, czcionka, wyrównanie do lewej, wyśrodkowanie, wyrównanie do prawej, wysokość wiersza, szerokość kolumny,
- *Solve and Analyze* (rozwiązanie i analiza): rozwiązanie, rozwiązanie krok po kroku, metoda graficzna, analiza parametryczna,
- *Results*: wyniki, grafy,
- *Utilities* (pomocnicze): kalkulator, zegar, pomoc.

Zamieszczone przykłady dydaktyczne dotyczą zasadniczo programowania liniowego i całkowitoliczbowego oraz zagadnienia transportowego z zastosowaniem algorytmu simpleks. Pakiet WinQSB obejmuje przede wszystkim zagadnienia z badań operacyjnych. Natomiast z obszaru ekonometrii na uwagę zasługuje moduł *FC – Forecasting (Prognozowanie i liniowa regresja)* umożliwiający estymację modeli regresji i modeli trendu. Występuje więc analogia w zakresie rodzaju modeli do pakietu EViews. Korzystając z modułu *FC*, otrzymujemy raporty z oszacowaniem parametrów, wygasłymi oraz przewidywanymi na żądanie określonymi prognozami w zakresie modeli:

- regresji (jedno- i wielowymiarowej); autoregresji,
- średnich (prostych i ruchomych),
- ważonych średnich ruchomych, podwójnych średnich ruchomych,
- prostego oraz adaptacyjnego wyrównywania wykładniczego,
- Browna (podwójnego wyrównywania wykładniczego).

Ponadto możemy określić prognozy poprzez wyrównywanie trendu z zastosowaniem modeli Holta-Wintersa. Moduł *FC* umożliwia nam określenie parametrów wyrównywania lub doboru najlepszej ich wartości z punktu minimalizacji *RMSE* wygasłych już prognoz (opcja: *Search the best*). Właśnie wymieniona opcja jest szczególnie przydatna, gdy określamy prognozy według modelu Holta-Wintersa z sezonowością multiplikatywną lub addytywną. Musimy wtedy rozpatrzyć modele z wieloma kombinacjami parametrów wyrównywania: α – zmiennej, β – trendu, γ – sezonowości, z krokiem co 0,01 z przedziału 0–1, co jest możliwe tylko przy realizacji komputerowej.

Opis pakietu programowego WinQSB wspomagającego podejmowanie decyzji znajdujemy także w *Aneksie* t. 3 publikacji pod redakcją naukową M. Lipiec-Zajchowskiej²¹. Obejmuje on m.in. moduły:

- prognozowanie liniowe i całkowitoliczbowe,
- programowanie dynamiczne,
- problem wyboru najkrótszej trasy,
- problem załadunku,
- sterowanie produkcją i zapasami,
- drzewo decyzyjne,
- analiza przedsięwzięć (metoda ścieżki krytycznej CPM, metoda PERT).

Zaprezentowane przykłady pochodzą głównie z obszaru badań operacyjnych.

1.2.2.4. Excel

Jest powszechnym i uniwersalnym narzędziem wspomagającym analizy statystyczne, estymację oraz weryfikację modeli ekonometrii. Posługuje się arkuszem kalkulacyjnym tworzącym skoroszyt. Znając wzory na poszczególne statystyki, testy czy równania modeli, możemy przy pomocy wyrażeń napisać potrzebne formuły obliczeniowe. Ponadto Excel ma wbudowane funkcje statystyczne i matematyczne. Do grupy statystycznych należą funkcje elementarne:

- WSP.KORELACJI,
- ŚREDNIA; NORMALIZUJ,
- WARIANCJA; WARIANCJA.POPUL; KOWARIANCJA,
- NACHYLENIE; ODCIĘTA,
- ODCH.STANDARDOWE; ODCH.STANDARDOWE.POPUL,

a także funkcje:

- R.KWADRAT (współczynnik determinacji),
- REGLINP (regresji liniowej),
- REGEXPP (regresji wykładniczej),
- REGBŁSTD (błędu standardowego reszt).

Istnieje również możliwość korzystania z opcji (*Excel/Analiza danych/Histogram*) w celu sporządzenia wykresów oraz tabel częstości występowania wartości danych w określonych przedziałach. Funkcje matematyczne, w szczególności z zakresu rachunku macierzowego, umożliwiają estymację parametrów równań liniowych, potęgowych, wykładniczych, wielowymia-

²¹ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003.

rowych. Należą do nich: TRANSPONUJ (transpozycja macierzy), MACIERZ.ILOCZYN, MACIERZ.ODW (macierz odwrotna).

Interesującym segmentem tego programu, niezbędnym w ekonometrii i statystyce opisowej, jest moduł (*Dodatki*). Między innymi dzięki niemu Excel znalazł zastosowanie w dydaktyce przedmiotu ekonometria, stąd fragmenty jego użycia spotykamy w różnych pozycjach literaturowych²². Analizie ekonometryczno-statystycznej z zastosowaniem zbioru procedur modułu (*Dodatki*) arkusza kalkulacyjnego Microsoft Excel poświęcona jest część trzecia publikacji²³. Przedstawiono w niej zastosowania opcji: korelacja, kowariancja, statystyka opisowa i regresja.

W Excelu ustawiamy się na aktywnej komórce, korzystamy z menu (*Wstaw/Funkcje/Statystyczne*) i wybieramy nazwę funkcji (*Kwartył*). Możemy też wybrać ścieżkę (*Narzędzia/Dodatki/Analiza danych/Statystyka opisowa*). Jeśli mamy wprowadzone dane do określonej kolumny, to w oknie dialogowym wpisujemy lub wybieramy m.in.:

- zakres wejściowy komórek zawierających dane,
- grupowanie według wierszy/kolumn,
- poziom ufności dla średniej,
- zakres wyjściowy (adres lewej górnej komórki tabeli wyników),
- statystyki podsumowujące: średnią, błąd standardowy, medianę, dominantę, odchylenie standardowe, wariancję, współczynnik kurtozy, skośność, zakres, minimum, maksimum.

W niniejszym opracowaniu przyjęto, że modele użyte do prognozowania i symulacji są wytworem ekonometrii, stąd wzięto też pod uwagę literaturę z tego obszaru. W rozdziale piątym pracy pod redakcją B. Radzikowskiej²⁴ podano przykłady zastosowania arkusza kalkulacyjnego Excel w ustalaniu prognoz. Określono regułę prognozowania oraz ocenę dopuszczalności prognozy, a funkcje Excela zastosowano do:

- współczynnika zmienności; współczynnika korelacji,
- oszacowania parametrów wielowymiarowego liniowego modelu regresji,
- statystyki regresji i analizy wariancji,
- określenia formuł modelu Holta i jego ekstrapolacji w celu wyznaczenia prognoz,

²² M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 1: *Statystyka...*, op.cit.; K. Jajuga K. (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002.

²³ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne...*, op.cit.

²⁴ B. Radzikowska (red.), *Metody prognozowania. Zbiór zadań*, Akademia Ekonomiczna, Wrocław 2004.

– danych analizy harmonicznej, współczynnika determinacji; odchylenia standardowego.

Realizacja komputerowa umożliwia dobieranie parametrów wygładzania, aby zminimalizować miarę *RMSE*. W omawianej publikacji zademonstrowano wygodną w tym zakresie funkcję *SUMA.XMY.2*, która oblicza sumę kwadratów różnic pomiędzy wartościami w dwóch zakresach komórek. Skorzystano także z Excela do obliczania błędu względnego prognozy, przewidywanej zmiennej objaśniającej (zależnej), stosując funkcję logarymiczną trendu. Sformułowano wyrażenia Excela na określenie błędów *ex ante* wyznaczanych prognoz.

Z zastosowaniem arkusza kalkulacyjnego Excel do budowy modeli prognostycznych, wyznaczania prognoz oraz oceny ich jakości spotykamy się też w pracy pod redakcją M. Cieślak²⁵. Omówiono w niej, na przykładzie, kolejne kroki postępowania (budowę i ekstrapolację, dobór parametrów, dopuszczalność prognozy) w modelu wygładzania wykładniczego. Zaprezentowano procedurę zbudowania prognoz punktowych i przedziałowych z zastosowaniem modelu analitycznego bazującego na szeregu czasowym i funkcji *REGLINP* Excela. W zaprezentowanych przykładach wyniki podano za pomocą programu Excel.

1.2.3. Informacje o innych programach

Tylko niektóre programy komputerowe zastosowane w ekonometrii, statystyce czy też prognozowaniu i symulacji doczekały się na rynku polskim odrębnych publikacji. Najczęściej książki z ekonometrii przytaczają i rozwijają sekwencje opisów teoretycznych z innych prac, wzmiankują o możliwości zastosowania informatyki we wspomaganii budowy i weryfikacji modeli. Przeglądając dostępną literaturę, znaleziono nawiązania do stosowania w różnych obszarach oprogramowania:

- EView 5 – program do analizy szeregów czasowych i modeli VAR,
- R-CRAN – program statystyczno-ekonometryczny, w tym rozwiązywanie szeregów czasowych,
- GAUSS – program macierzowy do obliczeń statystycznych,
- MATLAB – laboratorium macierzowe,
- Scilab – program typu MATLAB lub GASS,
- Stat – wyznaczanie podstawowych statystyk oraz prognoz według Holta,

²⁵ M. Cieślak (red.), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa 2005.

– DEMS – estymacja i prognozowanie (wykorzystujący prostotę i elastyczność Excela).

W przedmowie do wydania drugiego publikacji *Ekonometria*²⁶ czytamy: „Mamy obecnie do dyspozycji wiele programów komputerowych, z których można wybierać do woli. Ponadto zmieniają się one tak szybko, że ich listy należałoby bardzo często uzupełniać i rozszerzać”. G.S. Maddala w *Ekonometrii (Dodatek D: Programy komputerowe)* wymienia programy dostępne w końcu lat dziewięćdziesiątych. Od tego czasu opracowano ich nowe wersje, uwzględniające postęp w zakresie systemów operacyjnych. Oprócz programu SAS (którym posłużył się G.S. Maddala w dyskusji na temat współczynnika determinacji) są i inne, które ten wybitny ekonometryk uznał za powszechnie stosowane, a należą do nich:

- SHAZAM; MICROFIT 4.0; MICRO-TSP,
- RATS, CATS w pakiecie RATS do analizy kointegracji,
- LIMDEP (do analizy modeli z ograniczoną zmienną zależną, a także do modeli: logit, probit, tobit).

Według G.S. Maddali, niepokojące jest jednak to, że różne programy dają różne odpowiedzi na to samo pytanie. Autor sprawdził to na przykładzie estymacji parametrów modeli logitowych i probitowych, używając programów SAS oraz LIMDEP. Po analizie zaproponował dla prostych modeli logitowych i probitowych program SAS, a dla skomplikowanych LIMDEP. W zakresie podstawowych modeli z niedużymi bazami zaleca SHAZAM oraz MICROFIT.

Z rozpoznania w internecie wynika, że wyższe uczelnie ekonomiczne w coraz szerszym zakresie w nauczaniu statystyki, ekonometrii, jak również prognozowania i symulacji prowadzą ćwiczenia z tych przedmiotów w laboratoriach. Do modelowania zjawisk gospodarczych stosowane są programy: Statistica, EViews, EasyReg, GRETL oraz inne pracujące w środowisku Windows, Linux lub w systemach sieciowych uczelni. Zdaniem niektórych autorów, do prowadzenia ćwiczeń laboratoryjnych wystarczający jest jednak zestaw narzędziowy oferowany przez arkusz kalkulacyjny Excel²⁷.

Także firmy internetowe oferujące swoje usługi w rozwiązywaniu zadań z ekonometrii służą szeroką gamą oprogramowania, mianowicie programami: Statistica, Excel, SPSS, EViews, Winstat, Microfit, GRETL, GiveWin, G, PDG, Statgraf. Do zrealizowania zamówienia, opracowania i sprawdzenia modelu można podać własne dane modelowe, jak też wskazać ich źródło, np. Główny Urząd Statystyczny. W doborze zmiennych objaśniających stosowane

²⁶ G.S. Maddala, *Ekonometria...*, op.cit.

²⁷ J. Dziechciarz (red.), *Ekonometria. Metody...*, op.cit.

są przez te firmy takie metody, jak: Hellwiga (pojemności informacji), grafo-
wa, CIV, współczynnika korelacji, współczynnika zmienności, test *t*-Studenta.
Estymacja parametrów modelu przeprowadzana jest najczęściej metodami:
klasyczną metodą najmniejszych kwadratów, największej wiarygodności,
Gaussa-Seidela, Newtona-Raphsona. W ofercie firm usługowych jest wyko-
nywanie modeli: liniowych, nieliniowych, wielorównaniowych (rekurencyj-
nych, o równaniach współzależnych), VAR i innych o złożonej postaci anali-
tycznej. Jak już wspomniano, usługa zawiera weryfikację modeli, a w jej ra-
mach badanie: autokorelacji, stałości wariancji (heteroskedastyczności), nor-
malności reszt, współczynnika determinacji, odchylenia standardowego, prze-
prowadzenie testu *t*-Studenta, testu serii, określenie symetrii i losowość skład-
nika losowego (reszt), a także ocena stabilności parametrów. Wykonane opra-
cowania obejmują opisy, wzory oraz sformułowane wnioski.

Na stronie [<http://feweb.vu.nl/econometriclinks/software.html>] występuje
bardzo obszerny indeks aktualizowanych informacji o programach z ekono-
metrii. Pozostawia się Czytelnikowi zapoznanie z dokładnymi informacjami
na określonej stronie WWW, podającej autorów, wersje, linki do stron dystry-
butorów.

Ze względu na ramy niniejszego materiału podano tylko mały fragment
z tekstu w internecie, obejmujący symbole programów i to tylko w zakresie
liter A–G:

- A+, ACML, ADMB, APL, AIMMS, ALOGIT, Alyuda, AMOS, AMPL, Apophenia, Arc, AREMOS, AutoBox, Autometrics, AutoSignal;
- B34S, BACC, BATS, BETA, BIOGEME, BMDP, Brodgar, BUGS, BV4;
- CalculationCenter 3, CART 5.0, Census X12, Caterpillar-SSA 3.30, CPLEX 10, Confort 2.0, CVar Expert 1.7;
- DataDesk, Dataplore, Dataplot, DATAVIEW, DEA-Solver, DEMETRA, DYALOG, DYNARE NEW,
- EasyFit, EasyReg, EcoWin, ECTS, EQS, Eviews UPDATED, Excel UPDATED, Excel 2007, EXPO;
- FAME, ForecastPro, Fortran, FreeFore, FSQP;
- GAMS, G@RCH UPDATED, GAUSS, GAUSSX, GiveWin, Gem-pack, GeoDa, Genstat, GLIM, GLIMMIX, GQOPT, graphpad, Gnuplot, GSL, GRETL, (w ramach GRETL – wersja GRETL 1.7.4 (*Gnu Regression, econometric and Time-series Library*)).

1.2.4. O oprogramowaniu modeli symulacyjnych

Symulacja to utworzenie ogniwa pośredniego między realizmem a idealizmem. Burzliwy rozwój informatyki, zastosowań ekonometrii, badań operacyjnych oraz osiągnięcia w matematyce i komunikacji elektronicznej są bodźcem do rozwoju oprogramowania do symulowania w wirtualnej rzeczywistości funkcjonalności obiektów gospodarczych, bankowych, logistycznych i innych.

Z punktu widzenia symulacji model jest zbiorem obiektów abstrakcyjnych, relacji pomiędzy tymi obiektami. Cechuje go homeomorfizm, czyli analogiczne zachowania jak obiektu rzeczywistego, zatem badania symulacyjne wymagają odpowiednich metod budowy, weryfikacji, walidacji (ustalenia stopnia trafności) modelu. Jeśli mówimy obecnie o symulacji, to jest to przede wszystkim symulacja komputerowa bazująca na odpowiednich modelach matematycznych. Na przykład są to układy równań różniczkowych zachowania się zawieszona w technice nowo zaprojektowanym samochodzie jeżdżącym po różnych nawierzchniach. W kreowaniu modelu symulacyjnego wyróżniamy:

- grupowe budowanie modelu konceptualnego; przygotowanie danych źródłowych,
- wyodrębnienie elementarnych podsystemów; budowę prostych modeli (z reguły nieliniowych),
- formułowanie „modelu całościowego” do rozwiązania analitycznego, w który następnie, do rozwiązania matematycznego skomplikowanych równań z reguły nieliniowych, wprowadza się z konieczności uproszczenia.

Stosowanie modelu symulacyjnego ma następujące zalety:

- umożliwia przyspieszenie, spowolnienie czasu, by lepiej poznać zjawisko,
- można stosować wiele kryteriów jednocześnie, nawet w warunkach ekstremalnych,
- przy obecnym aparacie informatycznym jest możliwa budowa skomplikowanych modeli bez dużej wiedzy matematycznej,
- koszt budowy modelu symulacyjnego jest mniejszy niż koszt budowy obiektu.

Są również wady:

- wyniki badań są zasadne w określonych warunkach,
- do sformułowania wniosku konieczna jest wielokrotność badań, co wiąże się z długim czasem ich prowadzenia,
- w aparacie matematycznym występują ograniczenia, jeżeli chcemy znaleźć rozwiązanie optymalne,
- może wystąpić łatwość nadużywania, czyli przeceniania „dobroci” modelu tylko ze względu na jego efektywny interfejs.

W przypadku symulacji systemów społeczno-gospodarczych stosowane są różne techniki matematyczne i informatyczne modelowania. Niektóre techniki są łączone, tworząc hybrydy (krzyżówki), a jako przykłady można wymienić:

- zastosowanie symulacji i optymalizacji w doborze struktury modelu,
- planowanie eksperymentu w układzie symulacja–wspomaganie decyzji (poszukiwanie optymalnych wartości zmiennych decyzyjnych w modelu przedsiębiorstwa).

W ujęciu branżowym można wymienić np.:

- modelowanie symulacyjne w usprawnieniu obsługi klienta w banku (zdefiniowanie problemu, określenie etapów budowy modelu, propozycja zmiany systemu obsługi klienta),

- analizę zyskowności produktów ubezpieczeń majątkowych i osobowych,

- analizę symulacyjną efektywności inwestycji długookresowej.

Jednak największe trudności w algorytmizacji systemów gospodarczych występują w odwzorowaniu czynnika ludzkiego, czyli możliwości poznawczych i emocji samych decydentów²⁸. W tym względzie wykonuje się obserwacje ich zachowań przez „wciągnięcie” ich do wirtualnej rywalizacji w grze. Przykładem jest gra symulacyjna ze zwycięzcą, a w niej:

- występuje prowadzący, który ilościowo i jakościowo analizuje oraz ocenia dany przebieg gry,

- biorą udział uczestnicy – specjaliści danej branży,

- ustalone są reguły funkcjonowania firmy (obiektu) zakodowane w grze,

- wyłania się konkluzja, uczestnik, może nim być również student kierunku zarządzanie, który chce pogłębić swoją wiedzę z zakresu planowania produkcji oraz poznania metod i narzędzi informatycznych, które ten proces ułatwiają.

Warto zauważyć, że bankructwo wirtualnej firmy spowodowane niesprawnością w grze nie pozbawia pracy i majątku, ale wyzwala w uczestnikach prawdziwe emocje: gniew, wstyd, radość, dumę, współczucie dla przegranych. Taka więc gra tworzy minisystem społeczny, bowiem długotrwałe przebywanie w rzeczywistości alternatywnej (wirtualny rynek, wirtualny zarząd) ma wpływ na funkcjonowanie uczestników. Gry symulacyjne kreują cechy menedżerskie – dają poczucie władzy, uczą odpowiedzialności. W trakcie gry podejmuje się właściwe, zdaniem użytkowników, decyzje i ponosi ich konsekwencje. Ekonomiczne gry symulacyjne pozwalają eksperymentować na

²⁸ A. Balcerak, W. Kwaśnicki (red.), *Symulacja systemów społeczno-gospodarczych*, Instytut Organizacji i Zarządzania Politechniki Wrocławskiej, Wrocław 2005.

modelach wielu zmiennych decyzyjnych. Symulacja pomaga w rozpoznaniu i wykorzystaniu własnych zdolności i predyspozycji. Doskonalenie swoich cech menedżerskich pozwala unikać eksperymentowania na żywych organizmach firm. W internecie oferowane są różne kursy z tego zakresu.

W tej pracy nawiązano do niektórych modeli symulacyjnych w zakresie zarządzania produkcją, które dość wcześnie, bo już w latach siedemdziesiątych, pojawiły się w wiodących gospodarczo krajach i stosowane są również w dydaktyce.

T a b e l a 1.2.1

Opis modeli symulacyjnych

Nazwa modelu/gry	Opis
Joblot	Działanie wydziału produkcyjnego w zadanym rynku, konkurencji, procesu technologicznego
PROSIM	Gra niekonkurencyjna – koncentracja na minimalizacji kosztu. Przedmiotem jest produkcja wyrobów opisanych według BOM, proces zakupu surowców, planowania i sprzedaży wyrobów. Wyniki w formie raportów z całego procesu produkcyjnego oraz rozliczeniowego kosztów
OPT	System produkcyjny. Jakość zarządzania produkcją oceniana jest na podstawie zysku
CAMP	Praca fabryki – harmonogramowanie procesu produkcyjnego
DECIDE-P/OM	Niekonkurencyjna gra kierownicza. Wyniki mają postać rachunku wyników produkcji i rozliczeń finansowych
SIMAN-CINEMA	Animacja przemieszczania się części i formowania kolejek w strukturze technologicznej i przedmiotowej produkcji
The Executive Game	Gra dla menedżerów. Zmienne: wielkość produkcji, zakup materiałów, wydatki na badania
TRAIN-F	Gra – dobór parametrów planowania i sterowania produkcją
FMS Design Game	Gra do nauczenia projektowania elastycznych systemów produkcyjnych (ESP)
TEES-2	Symulacyjna gra decyzyjna zarządzania produkcją
MRP-SIM	Gra oparta na arkuszu kalkulacyjnym w celu zrozumienia planowania i sterowania produkcją

Ź r ó d ł o: A. Balcerak, W. Kwaśnicki (red.), *Symulacja systemów społeczno-gospodarczych*, Instytut Organizacji i Zarządzania Politechniki Wrocławskiej, Wrocław 2005, tab. 1.

Przedstawię teraz dwa z programów symulacyjnych – wymieniony w tab. 1.2.1 TEES-2 oraz Zarządzanie kapitałem.

1.2.4.1. TESS-2

TESS-2 opracowano z myślą o doskonaleniu umiejętności zarządzania firmą. Uczestnicy przygotowują biznesplan oraz realizację w warunkach zmieniającego się rynku i ostrej konkurencji. TESS-2 posiada moduł zarządzania finansami i kontroli wyników na podstawie wskaźników finansowych. Gra opiera się na kompleksowej symulacji decyzyjnej w obszarze zarządzania przedsiębiorstwem. Każda z grup prowadzi fikcyjną firmę oferującą jeden produkt, i działającą na jednym rynku. Symulacja obejmuje od 12 do 20 kwartałów rozliczeniowych. Konkurencja jest bezpośrednia i ostra, a wynik jednego zespołu wpływa bezpośrednio na wynik pozostałych. Wyniki gry ocenia się na podstawie parametrów finansowych, takich jak: zysk, udział w rynku, wartość likwidacyjna firmy, wysokość wypłaconych dywidend. Po zakończeniu gry ogłaszany jest ranking poszczególnych zespołów.

1.2.4.2. Program: Zarządzanie kapitałem

Program jest modelem uniwersalnego banku komercyjnego. Uczestnicy gry działają w zespołach tworzących minizarządy banków komercyjnych. Zespoły konkurują ze sobą na rynkach finansowych, przeprowadzając operacje: finansowe, depozytowo-kredytowe, giełdowe, walutowe i inne. Zadaniem każdego zespołu jest najefektywniejsze prowadzenie instytucji finansowej w ciągu 3–4 lat. Można powiedzieć, że gra jest:

- przewodnikiem po rynkach finansowych (depozytowym, kredytowym, papierów wartościowych, walutowym, międzybankowym, finansowym ochrony środowiska) i ich mechanizmach,
- kształceniem umiejętności oceny zdolności kredytowej firmy i osoby fizycznej oraz oceny ryzyka,
- instrumentem rozwijania umiejętności współdziałania w grupie i zespołowego podejmowania decyzji.

Podobnie możemy prowadzić symulację komputerową na modelu ekonometrycznym, np. w celu udzielenia odpowiedzi na pytania:

- jakie byłyby wartości zmiennych endogenicznych, gdyby zmienne egzogeniczne przyjęły określone wartości?
- jakie powinny być wartości zmiennych endogenicznych, by uzyskać pożądane wartości zmiennych egzogenicznych?

W takiej symulacji powinniśmy mieć możliwość wymiany zmiennych, parametrów, właściwości składnika losowego. Wynik symulacji, czyli warianty rozwoju obiektu opisywanego przez model, może być traktowany jako prognozy mające wpływ na zmianę cen oraz rozwój przedsiębiorstwa.

Na zakończenie niniejszego przeglądu autorzy wyrażają nadzieję, że studiujący, zachęcony gamą programów oraz możliwościami zastosowania informatyki w ekonometrii, będzie penetrował dalej istniejącą literaturę przedmiotu, jak i rozpoznawał pakiety w celu uzyskania perfekcji w budowaniu modeli nie tylko gospodarczych.

Literatura pomocnicza do podrozdziału 1.2

Bartosiewicz S. (red.), *Ekonometria wspomagana mikrokomputerem*, Akademia Ekonomiczna, Wrocław 1991.

Bartosiewicz S., *Ekonometria. Technologia ekonometrycznego przetwarzania informacji*, PWE, Warszawa 1989.

Cieślak M. (red.), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa 2005.

Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania*, Akademia Ekonomiczna, Wrocław 2003.

Gajda J., *Ekonometria. Wykład i łatwe obliczenia w programie komputerowym*, Wydawnictwo C.H. Beck, Warszawa 2004.

Jajuga K. (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002.

Koślacz P. (red.), *Nowoczesne pakiety komputerowe w podejmowaniu decyzji*, Uniwersytet Warszawski, Warszawa 1998.

Kowerski M., *Ekonometria. Przewodnik po wykładach*, Wyższa Szkoła Zarządzania i Administracji w Zamościu, Zamość 1998.

Krokosz-Krynke Z., *Symulacja w nauczaniu zarządzania produkcją*, [w:] Balcerak A., Kwaśnicki W. (red.) *Symulacja systemów społeczno-gospodarczych*, Instytut Organizacji i Zarządzania Politechniki Wrocławskiej, Wrocław 2005.

Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2007.

Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 1: *Statystyka*, Wydawnictwo C.H. Beck, Warszawa 2003.

Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003.

Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck Warszawa 2003.

Lipiec-Zajchowska M., Sikora E., Zajchowski J., *Modele podejmowania decyzji w pakietach komputerowych*, Uniwersytet Warszawski, Warszawa 1995.

Maddala G.S., *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2006.

Nowa E., *Zarys metod ekonometrycznych. Zbiór zadań*, Wydawnictwo Naukowe PWN, Warszawa 1990.

- Osińska M. (red.), *Ekonometria współczesna*, TNOiK, „Dom Organizatora”, Toruń 2007.
- Kukuła K. (red), *Badania operacyjne w przykładach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa 2004.
- Sikora W. (red), *Przykłady i zadania z badań operacyjnych i ekonometrii*, Akademia Ekonomiczna, Poznań 2003.
- Czyżewski R., Hundert M., Klósk a R., *Wybrane zagadnienia z prognozowania*, ECONOMICUS, Szczecin 2006.
- Radzikowska B., *Metody prognozowania. Zbiór zadań*, Akademia Ekonomiczna, Wrocław 2004.
- Skrzypek J., Szubra M., *TESS-2. Symulacyjna gra decyzyjna*, Poltex, Warszawa 1996.
- Strahl D. (red.), *Modelowanie ekonometryczne z Excelem*, Akademia Ekonomiczna, Wrocław 2004.

1.3. Miary rozkładu jednej zmiennej (przykłady w Excelu)

1.3.1. Przedziały obserwacji

Obrazem zmiennej jest rozkład, który można przedstawić jako szereg statystyczny. Jeśli mamy wszystkie obserwacje, to taki szereg nazywamy szczegółowym n obserwacji. Zmienne oznaczamy dużymi literami, np. X , a kolejne obserwacje tej zmiennej jako x_1, x_2, \dots, x_n ²⁹. Gdy jest duża liczba obserwacji, tworzy się szereg rozdzielczy o K przedziałach.

Środek przedziału oznaczamy przez \dot{x}_i , a liczebność danego przedziału przez n_i , przy czym $i = 1, 2, \dots, K$.

P r z y k ł a d

Dla zobrazowania przedziałów podano dane przedstawiające odczytane z licznika Geigera ilości cząstek emitowanych przez substancję radioaktywną w 100 kolejnych 40-sekundowych okresach.

T a b e l a 1.3.1

Ilość emitowanych cząstek w 100 czterdziestosekundowych okresach

23	20	16	18	30	22	26	15	13	18
14	17	11	37	21	16	10	20	22	25
19	19	19	20	12	23	24	17	18	16
27	16	28	26	15	29	19	28	22	17
12	24	21	22	20	15	18	16	23	24
15	24	28	19	24	22	17	19	8	18
17	18	23	21	25	19	20	22	21	21
16	20	19	11	23	17	23	13	17	26
26	14	15	16	27	18	21	24	33	20
21	27	18	22	17	20	14	21	22	19

Ź r ó d ł o: Opracowanie własne na podstawie: J. F r e u n d, *Podstawy nowoczesnej statystyki*, PWE, Warszawa 1971, s. 22.

Średnia ze wszystkich obserwacji wynosi 20,04; wartość minimalna 8, a wartość maksymalna 37. Przyjęto siedem klas danych pogrupowanych po pięć liczb.

²⁹ K. Jajuga (red.), *Ekonometria. Metody...*, op.cit.

Tabela 1.3.2

Liczność klas

Przedział	Ilość wystąpień
5–9	1
10–14	10
15–19	37
20–24	37
25–29	12
30–34	2
35–39	1

Źródło: Opracowanie własne na podstawie:
J. F r e u n d, *Podstawy nowoczesnej statystyki*,
PWE, Warszawa 1971, s. 23.

Po uporządkowaniu w kolumnach i wierszach narastająco, korzystając z narzędzia Excela (*Excel/Dane/Sortuj*) – bez nazwy kolumny (przy czym dla wierszy: sortowanie od lewej do prawej) – otrzymamy tabelę 1.3.3. W zakończeniu tabeli czcionką pogrubioną zaznaczono licznosc klas. Wartość najmniejszą i największą w każdej klasie nazywamy granicą klasy.

Tabela 1.3.3

Uporządkowane narastająco dane odczytane z licznika Geigera

										5–9	10–14	15–19	20–24	25–29	30–34	35–39	Spraw- dzenie
8	10	11	11	12	12	13	14	15	16	1	7	2	0	0	0	0	10
13	14	14	15	15	15	16	16	16	17	0	3	7	0	0	0	0	10
15	16	16	17	17	17	17	17	18	18	0	0	10	0	0	0	0	10
16	17	18	18	18	18	18	18	19	20	0	0	9	1	0	0	0	10
17	19	19	19	19	19	19	20	21	21	0	0	7	3	0	0	0	10
19	19	20	20	20	20	20	21	22	23	0	0	2	8	0	0	0	10
20	21	21	21	21	21	22	22	22	24	0	0	0	10	0	0	0	10
22	22	22	22	23	23	23	24	24	25	0	0	0	10	0	0	0	10
23	23	24	24	24	25	26	26	27	28	0	0	0	5	5	0	0	10
26	26	27	27	28	28	29	30	33	37	0	0	0	0	7	2	1	10
										1	10	37	37	12	2	1	100

Źródło: Opracowanie własne.

Do scharakteryzowania rozkładu zmiennej stosujemy miary położenia, zmienności oraz skośności (asymetrii).

1.3.2. Miary przeciętne (położenia)

Rozróżniamy średnie (arytmetyczna, harmoniczna, geometryczna) oraz przeciętne pozycyjne (wartość środkowa – mediana, kwartyle) miary położenia.

1.3.2.1. Średnie

Często stosowana średnia arytmetyczna, np. zmiennej X , jest sumą wartości wszystkich obserwacji podzieloną przez jej liczebność.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}. \quad (1.3.1)$$

Średnia arytmetyczna ma kilka własności³⁰:

a) suma odchyleń od wartości średniej równa się zero:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0,$$

b) suma kwadratów odchyleń wartości zmiennej od średniej arytmetycznej jest najmniejsza,

c) jeżeli do wszystkich wartości zmiennej X dodamy stałą c , to wówczas wartość średnia wyrażona jest wzorem:

$$\overline{X + c} = \bar{X} + c,$$

d) iloczyn stałej i zmiennej:

$$\overline{X \cdot c} = \bar{X} \cdot c,$$

e) jeżeli jedna zmienna jest sumą dwóch innych zmiennych (Y , Z), to średnia arytmetyczna jest sumą średnich arytmetycznych wspomnianych zmiennych:

$$\bar{Y} = \bar{X} + \bar{Z}.$$

³⁰ Ibidem.

Przykład

Dla $n = 10$ obserwacji x_i pierwszej kolumny tabeli 1.3.1 oraz własności c), d), e).
Przyjęto $c = 2$.

Tabela 1.3.4

Obliczenia średnich według własności „c” i „d”

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$x_i + c$	$x_i \cdot c$
23	4	16	25	46
14	-5	25	16	28
19	0	0	21	38
27	8	64	29	54
12	-7	49	14	24
15	-4	16	17	30
17	-2	4	19	34
16	-3	9	18	32
26	7	49	28	52
21	2	4	23	42
19		236	21	38

gdzie:

19 – średnia arytmetyczna zmiennej X obliczona wyrażeniem Excela,

np. =SUMA(A2:A11)/10,

236 – suma kwadratów odchyłeń,

21 – średnia z $(x_i + c)$,

38 – średnia iloczynu $(x_i \cdot c)$.

Źródło: Opracowanie własne; x_i według tabeli 1.3.1.

Tabela 1.3.5

Obliczenie średniej według własności „e”

x_i	z_i	$y_i = x_i + z_i$
23	20	43
14	17	31
19	19	38
27	16	43
12	24	36
15	24	39
17	18	35
16	20	36
26	14	40
21	27	48
19	19,9	38,9

Źródło: Opracowanie własne; dane x_i oraz z_i kol. 1 i 2 tab. 1.3.1.

Jeżeli wartości zmiennej pogrupowane są w szereg rozdzielczy, co prezentuje tab. 1.3.2, to średnia arytmetyczna, zwana ważoną, obliczana jest według wzoru:

$$\bar{x} = \frac{\sum_{i=1}^K n_i \cdot \dot{x}_i}{n}, \quad (1.3.2)$$

gdzie:

\dot{x}_i – środek i -tego przedziału,
 n_i – liczność i -tego przedziału,
 K – liczba przedziałów

Przykład

Dla $n = 100$ obserwacji według tab. 1.3.1.

Tabela 1.3.6

Obliczenia pomocnicze średniej ważonej

Przedział	n_i	\dot{x}_i	$n_i \cdot \dot{x}_i$
5–9	1	7	7
10–14	10	12	120
15–19	37	17	629
20–24	37	22	814
25–29	12	27	324
30–34	2	32	64
35–39	1	37	37
Suma			1 995

Źródło: Opracowanie własne.

Do obliczenia sumy można zastosować wyrażenie Excela np.:
 =SUMA(D129:D135).

Po podstawieniu do wzoru (1.3.2) otrzymamy wynik 19,95.

Warto zauważyć, że średnia arytmetyczna z danych szeregu szczegółowego nieuporządkowanego (tab. 1.3.1) wynosi 20,04, zatem różnica między tymi dwoma średnimi jest mała i wynosi:

$$20,04 - 19,95 = 0,09.$$

Przykładowe wyrażenie Excela na obliczenie tej różnicy: =E29 – E138.

Kolejną stosowaną miarą położenia jest średnia harmoniczna H_x , stosowana wówczas, gdy wartości jednostek podawane są w formie odwrotności.

Dla szeregu szczegółowego przy x_i różnym od zera średnią harmoniczną obliczamy ze wzoru:

$$H_x = \frac{n}{\sum_{i=1}^n (1/x_i)}, \quad (1.3.3)$$

natomiast średnią harmoniczną ważoną według wzoru:

$$H_x = \frac{n}{\sum_{i=1}^k (n_i / \dot{x}_i)}. \quad (1.3.4)$$

Przykład

Tabela 1.3.7

Obliczenia pomocnicze średniej harmonicznej

Przedział	n_i	\dot{x}_i	n_i / \dot{x}_i
5–9	1	7	0,14285714
10–14	10	12	0,83333333
15–19	37	17	2,17647059
20–24	37	22	1,68181818
25–29	12	27	0,44444444
30–34	2	32	0,0625
35–39	1	37	0,02702703

Źródło: Opracowanie własne.

Suma n_i / \dot{x}_i wynosi około 5,37 i po podstawieniu do wzoru (1.3.4) otrzymamy $H_x = 18,86$.

Gdy porównamy średnią ważoną oraz średnią harmoniczną ważoną obliczone dla szeregu rozdzielczego według tab. 1.3.3, to uzyskamy różnicę: $19,95 - 18,63 = 1,32$.

Przejdźmy teraz do obliczania średniej geometrycznej G_x , bazując na danych z tab. 1.3.7. Średnia ta stosowana jest między innymi do rozkładów, w których zachodzą znaczne różnice między wartościami. Dla szeregu szczegółowego jest ona pierwiastkiem n -tego stopnia z iloczynu n obserwacji zmiennej X , przy czym $i = 1, 2, \dots, n$.

$$G_x = \sqrt[n]{\prod_{i=1}^n x_i}, \quad (1.3.5)$$

natomiast dla szeregu rozdzielczego obliczamy ją z zależności:

$$G_x = \sqrt[n]{\dot{x}_1^{n_1} \cdot \dot{x}_2^{n_2} \dots \dot{x}_k^{n_k}}. \quad (1.3.6)$$

Przykład

Opierając się na tab. 1.3.7, obliczamy G_x :

Tabela 1.3.8

Obliczenia pomocnicze średniej geometrycznej

n_i	\dot{x}_i	$\dot{x}_i^{n_i}$
1	7	7
10	12	6,1917E+10
37	17	3,3621E+45
37	22	4,6735E+49
12	27	1,5009E+17
2	32	1024
1	37	37

Źródło: Opracowanie własne.

Po przemnożeniu w Excelu wartości $\dot{x}_i^{n_i}$ otrzymamy wynik około $3,9E + 128$, gdzie E podstawa logarytmu naturalnego.

Przykładowe wyrażenie Excela do otrzymania powyższego iloczynu:
=D194·D195·D196·D197·D198·D199·D200.

Suma n_i , jak wiadomo, równa się 100 obserwacjom. Obliczmy zatem pierwiastek setnego stopnia z otrzymanego iloczynu (możemy skorzystać również z wyrażenia Excela), np. =E201^0,01,

gdzie:

$e = E \approx 2,718282$.

$$G_x = \sqrt[100]{3,9 \cdot 2,718282^{128}} = 19,31.$$

Porównanie wartości różnych średnich ważonych dla szeregu rozdzielczego przedstawia się następująco:

$$H_x = 18,63 \leq G_x = 19,31 \leq \bar{x} = 19,95.$$

Średnia geometryczna ma tę właściwość, że jeśli jedna zmienna jest iloczynem dwóch innych zmiennych, to jej średnia geometryczna jest iloczynem średnich geometrycznych zmiennych:

$$G_{xy} = G_x G_y. \quad (1.3.7)$$

Przykład

Tabela 1.3.9

Obliczenia pomocnicze iloczynu średnich geometrycznych

x_i	z_i	$y_i = x_i z_i$
23	20	460
14	17	238
19	19	361
27	16	432
12	24	288
15	24	360
17	18	306
16	20	320
26	14	364
21	27	567
18,3865704	19,53696	359,217642

Źródło: Opracowanie własne; dane x_i oraz z_i z tab. 1.3.1; w wierszu dolnym obliczone średnie.

Korzystając z wyrażenia Excela, obliczamy średnie geometryczne zmiennej X oraz Z , jak również sumy iloczynów $y_i = x_i z_i$, bazując na wyrażeniu np.:

=ŚREDNIA.GEOMETRYCZNA(C218:C227).

Taki sam wynik otrzymamy, jeśli przemnożymy średnie geometryczne podane w tab. 1.3.9 zmiennej X oraz zmiennej Z .

1.3.2.2. Przeciętne pozycyjne

Drugą grupę miar położenia stanowią przeciętne pozycyjne, do wyznaczenia których stosujemy szereg szczegółowy uporządkowany. Rozpocznijmy od mediany, której liczba jednostek mających wartość nie większą od mediany jest równa liczbie jednostek mających wartość nie mniejszą od niej.

W przypadku nieparzystej liczby obserwacji medianą jest obserwacja o numerze $[(n + 1)/2]$:

$$M_x = x_{[(n+1)/2]}.$$

Jeśli liczba obserwacji jest parzysta, mediana jest średnią arytmetyczną dwóch wartości leżących w środku uporządkowanego szeregu.

$$M_x = \frac{x_{n/2} + x_{n/2+1}}{2}. \quad (1.3.8)$$

Wyjaśnijmy to na przykładzie, w którym zmienną X podano jako szereg uporządkowany narastający dziesięciu obserwacji.

T a b e l a 1.3.10

Szereg uporządkowany

x_i	n
12	1
14	2
15	3
16	4
17	5
19	6
21	7
23	8
26	9
27	10

Ź r ó ł o: Opracowanie własne; x_i z tab. 1.3.1.

$$M_x = \frac{17+19}{2} = 18.$$

Do szeregu rozdzielczego stosujemy wzór³¹:

$$M_x = x_0 + h \cdot \frac{(n+1)/2 - n_s}{n_0}, \quad (1.3.9)$$

gdzie:

n – ogólna liczba obserwacji,

n_s – liczba obserwacji w szeregu skumulowanym w przedziale poprzedzającym przedział zawierający medianę,

x_0 – dolna granica przedziału zawierającego medianę,

h – rozpiętość przedziału zawierającego medianę,

n_0 – liczba obserwacji w przedziale z medianą.

³¹ Ibidem, s. 16.

Przykład

Na podstawie danych szeregu rozdzielczego promieniowania – tab. 1.3.3.

Tabela 1.3.11

Liczebność skumulowana cząstek

Przedział	Ilość cząstek	n_i	Liczebność skumulowana
0–4	<5	0	0
5–9	<10	1	1
10–14	<15	10	11
15–19	<20	37	48
20–24	<25	37	85
25–29	<30	12	97
30–34	<35	2	99
35–39	<40	1	100

Źródło: Opracowanie własne według tabeli 1.3.2.

W tabeli: $n_s = 48$, $x_0 = 20$, $h = 5$, $n_0 = 37$, $n = 100$ (liczba obserwacji)

$$M_x = 20 + 5 \cdot \frac{(100 + 1) / 2 - 48}{37} = 20,338.$$

Dla porównania w szeregu szczegółowym (tab. 1.3.6) średnia ważona wynosiła 19,95.

Kolejne miary pozycyjne to kwartyle. Dzielą one szereg szczegółowy uporządkowany względem liczebności na cztery równe części. Na przykład kwartył pierwszy jest to taka wartość zmiennej, która dzieli szereg uporządkowany tak, że 1/4 jednostek ma od niej nie większe, a 3/4 nie mniejsze wartości.

Przykład

Podział szeregu uporządkowanego promieniowania cząstek na kwartyle dla $n = 100$ obserwacji.

Tabela 1.3.12

Szereg szczegółowy uporządkowany narastająco

x_i	n	x_i	n	x_i	n	x_i	n
8	1	17	26	20	51	23	76
10	2	17	27	20	52	23	77
11	3	17	28	20	53	23	78
11	4	17	29	20	54	24	79
12	5	17	30	20	55	24	80
12	6	17	31	20	56	24	81
13	7	18	32	21	57	24	82
13	8	18	33	21	58	24	83
14	9	18	34	21	59	24	84
14	10	18	35	21	60	25	85
14	11	18	36	21	61	25	86
15	12	18	37	21	62	26	87
15	13	18	38	21	63	26	88
15	14	18	39	21	64	26	89
15	15	19	40	22	65	26	90
15	16	19	41	22	66	27	91
16	17	19	42	22	67	27	92
16	18	19	43	22	68	27	93
16	19	19	44	22	69	28	94
16	20	19	45	22	70	28	95
16	21	19	46	22	71	28	96
16	22	19	47	22	72	29	97
16	23	19	48	23	73	30	98
17	24	20	49	23	74	33	99
17	25	20	50	23	75	37	100

Źródło: Opracowanie własne według tab. 1.3.3; dane posortowane narastająco.

Wartość kwartyli zmiennej x_i przy obserwacjach n_i :25 – kwartył 1 $Q_1 - 17$,50 – kwartył 2 $Q_2 - 20$,75 – kwartył 3 $Q_3 - 23$,100 – kwartył 4 $Q_4 - 37$.

Dla szeregu rozdzielczego kwartył pierwszy obliczmy według wzoru:

$$Q_x = x_0 + h \cdot \frac{n/4 - n_s}{n_0}, \quad (1.3.10)$$

gdzie:

n – liczba obserwacji,

n_s – liczba obserwacji w szeregu skumulowanym, w przedziale poprzedzającym przedział zawierający kwartył 1,

x_0 – dolna granica przedziału zawierającego kwartył 1,

h – rozpiętość przedziału zawierającego kwartył 1,

n_0 – liczba obserwacji w przedziale z kwartyłem,

Dla obliczenia Q_3 zamiast $n/4$ wstawiamy $3n/4$.

P r z y k ł a d

Obliczamy kwartył pierwszy, drugi (jako mediana) i kwartył trzeci szeregu rozdzielczego, mając dane z tabeli 1.3.11 oraz 1.3.12:

kwartył 1: $n_s = 11$, $x_0 = 15$, $h = 5$, $n_0 = 37$,

kwartył 2: $x_{50} = 20$, $x_{51} = 20$,

kwartył 3: $n_s = 48$, $x_0 = 20$, $h = 5$, $n_0 = 37$.

$$Q_1 = 15 + 5 \cdot \frac{100/4 - 11}{37} = 16,03,$$

$$Q_2 = (20 + 20) / 2 = 20,$$

$$Q_3 = 20 + 5 \cdot \frac{3 \cdot 100/4 - 48}{37} = 23,65.$$

1.3.2.3. Wykres pudełkowy oraz histogram³²

Wykres pudełkowy składa się z prostokąta rozciągniętego na poziomie kwartyli i przedzielonego na poziomie mediany. Do niego dołączone są dwa odcinki. Sposób ich określenia omówiony zostanie na przykładzie danych z tabeli 1.3.12 i wyników wcześniejszego przykładu:

gdzie:

$$x_{\min} = 8, x_{\max} = 37, Q_x = (Q_3 - Q_1) / 2 = (23,65 - 16,03) / 2 = 3,81.$$

Sprawdzamy warunki:

$$x_{\min} > Q_1 - 3Q_3 = 16,03 - 3 \cdot 3,81 = 4,6 \text{ (warunek jest spełniony).}$$

Rysujemy odcinek od pudełka do x_{\min} .

³² Ibidem, s. 22.

Podobnie rysuje się drugi odcinek łączący kwartył trzeci z wartością największą $x_{\max} = 37$. Jeżeli:

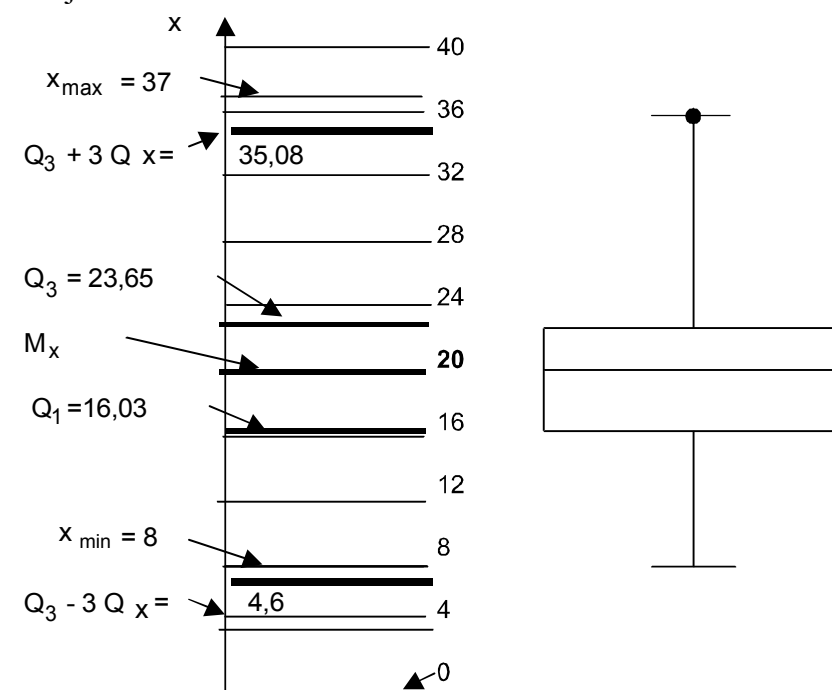
$$Q_3 + 3Q_x = 23,65 + 3 \cdot 3,81 = 35,08,$$

to warunek nie jest spełniony i dlatego drugi odcinek łączymy z wyliczoną wartością.

Ponadto zaznacza się wartości nietypowe, jeśli występują. Obserwacja i -ta uznawana jest za nietypową, jeśli zachodzi jeden z dwóch warunków:

$$x_i < Q_1 - 3Q_x; \quad x_i > Q_3 + 3Q_x.$$

Warunek pierwszy nie występuje w naszym przykładzie, bo $x_1 = 8$, natomiast warunek drugi występuje, gdyż $x_{100} = 37$ i jest to nietypowa wartość obserwacji.



Źródło: Opracowanie własne na podstawie: K. Jajuga (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002, rys. 1.5.

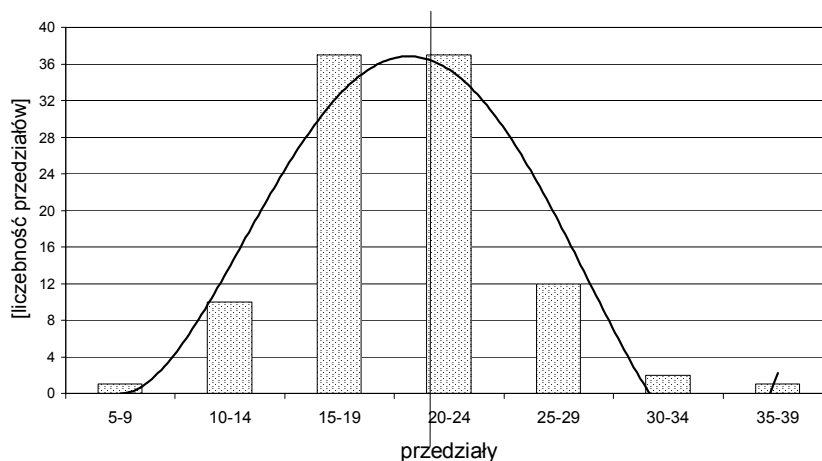
Wykres 1.3.1 Wykres pudełkowy szeregu rozdzielczego

Opierając się na prezentacji graficznej, wyciągamy wnioski o:

– wartościach przeciętnych (środek pudełka – mediana; brzegi pudełka – kwartyle),

- zmienności (długość pudełka),
- skośności (położenie linii podziału pudełka w stosunku do brzegów pudełka – w rozkładach o asymetrii lewostronnej linia podziału jest bliżej wierzchołka pudełka).

Przedstawmy teraz rozkład jednej zmiennej przy pomocy histogramu (wykres 1.3.2) w układzie współrzędnych prostokątnych na przykładzie danych z tabeli 1.3.2. Pamiętamy z wcześniejszego przykładu, że średnia arytmetyczna z kolejnych 100 obserwacji wynosi 20,04 (tab. 1.3.1). Na wykresie 1.3.2 zaznaczono to linią pionową. Na osi rzędnych zaznaczono liczebności przedziałów poszczególnych prostokątów. Otrzymaliśmy wykres o asymetrii prawostronnej.



Źródło: Opracowanie własne wykonane w Excelu.

Wykres 1.3.2. Histogram

1.3.3. Miary rozproszenia (zmienności)

Po omówieniu miar położenia przejdźmy do miar zmienności, które służą do scharakteryzowania stopnia zróżnicowania wartości badanej zmiennej. Miary zmienności dzielimy na absolutne i stosunkowe. Miarami absolutnymi są: rozstęp, odchylenie standardowe i wariancja oraz odchylenia przeciętne i ćwiartkowe.

Przykładem miary zmienności stosunkowej jest współczynnik zmienności V_x .

Prezentację rozkładu dwóch zmiennych X_1 oraz X_2 o podanej liczności n_i występowania danych przedstawia tab. 1.3.13.

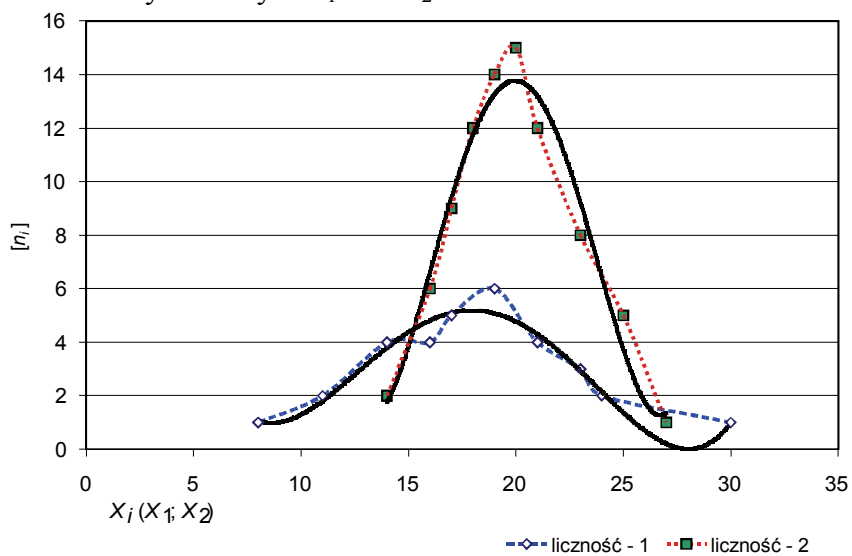
Tabela 1.3.13

Liczność występowania wartości zmiennych

X_1	n_1	X_2	n_2
8	1	14	2
11	2	16	6
14	4	17	9
16	4	18	12
17	5	19	14
19	6	20	15
21	4	21	12
23	3	23	8
24	2	25	5
30	1	27	1
18,3		20	Średnia
6,53		4,08	Odchylenie standardowe

Źródło: Opracowanie własne; dane umowne.

Na wykresie 1.3.3 większemu skupieniu odpowiada krzywa o wyraźnie zaznaczonym maksimum (dla licznosci n_2 zmiennej X_2). Wykres pokazuje również trendy zmiennych X_1 oraz X_2 .



Źródło: Opracowanie własne w Excelu według danych tab. 1.3.13.

Wykres 1.3.3. Histogramy zmiennych X_1 oraz X_2

Obszar zmienności (rozstęp) należy do miar zmienności absolutnych. Wykorzystane zostaną tu wzory zamieszczone w publikacji K. Jajugi³³.

Rozstęp określamy wzorem:

$$O_x = x_{\max} - x_{\min}.$$

P r z y k ł a d

Na podstawie danych o liczności promieniowania dla 100 obserwacji zmiennej X obliczamy rozstęp O_x . Według tabeli 1.3.12 szeregu uporządkowanego obserwacji $x_{\min} = 8$, $x_{\max} = 37$, rozstęp $O_x = 37 - 8 = 29$.

Teraz uwagę skupimy na odchyleniach. Rozróżniamy odchylenia ćwiartkowe oraz przeciętne:

a) Ćwiartkowe, bazujące na miarach pozycyjnych, czyli połowa różnicy między trzecim a pierwszym kwartylem:

$$Q_x = (Q_3 - Q_1) / 2.$$

P r z y k ł a d

Mając dane z wcześniejszego obliczenia $Q_3 = 23,65$ oraz $Q_1 = 16,03$ określamy odchylenie ćwiartkowe:

$$Q_x = (23,65 - 16,03) / 2 = 3,81.$$

Odchylenie ćwiartkowe jest średnią odległością mediany $M_x = 20$ od kwartyli Q_1 oraz Q_3 .

b) Przeciętne, które jest średnią arytmetyczną bezwzględnych wartości odchyień od średniej arytmetycznej z wszystkich obserwacji danej zmiennej.

$$D_x = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}. \quad (1.3.11)$$

P r z y k ł a d

Obliczmy D_x szeregu szczegółowego według danych x_i z promieniowania cząstek (określone w przykładzie danych tabeli 1.3.1, gdzie średnia arytmetyczna $= 20,04$, a ilość obserwacji $n = 100$). W tabeli 1.3.14 obok kolumny wartości x_i podano kolumnę z wartością $|x_i - \bar{x}|$, obliczoną przykładowo jako: $|23 - 20,04| = 2,96$.

³³ Ibidem, rozdz. 1.3.

Tabela 1.3.14

Obliczenia pomocnicze w Excelu do określenia przeciętnej

Lp.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	T	U
1	23	2,96	20	0,04	16	4,04	18	2,04	30	9,96	22	1,96	26	5,96	15	5,04	13	7,04	18	2,04
2	14	6,04	17	3,04	11	9,04	37	16,96	21	0,96	16	4,04	10	10,04	20	0,04	22	1,96	25	4,96
3	19	1,04	19	1,04	19	1,04	20	0,04	12	8,04	23	2,96	24	3,96	17	3,04	18	2,04	16	4,04
4	27	6,96	16	4,04	28	7,96	26	5,96	15	5,04	29	8,96	19	1,04	28	7,96	22	1,96	17	3,04
5	12	8,04	24	3,96	21	0,96	22	1,96	20	0,04	15	5,04	18	2,04	16	4,04	23	2,96	24	3,96
6	15	5,04	24	3,96	28	7,96	19	1,04	24	3,96	22	1,96	17	3,04	19	1,04	8	12,04	18	2,04
7	17	3,04	18	2,04	23	2,96	21	0,96	25	4,96	19	1,04	20	0,04	22	1,96	21	0,96	21	0,96
8	16	4,04	20	0,04	19	1,04	11	9,04	23	2,96	17	3,04	23	2,96	13	7,04	17	3,04	26	5,96
9	26	5,96	14	6,04	15	5,04	16	4,04	27	6,96	18	2,04	21	0,96	24	3,96	33	12,96	20	0,04
10	21	0,96	27	6,96	18	2,04	22	1,96	17	3,04	20	0,04	14	6,04	21	0,96	22	1,96	19	1,04
11	20,04	44,08		31,16		42,08		44,00		45,92		31,08		36,08		35,08		46,92		28,08

Źródło: Opracowanie własne w Excelu.

W komórce A11 arkusza kalkulacyjnego programu Excel znajduje się wartość średniej arytmetycznej – 20,04. Suma wszystkich sum kolumnowych wynosi 384,48. Po podzieleniu przez 100 otrzymujemy $D_x = 3,84$. Aby obliczyć kolejne wartości w kolumnach, np. 2,96, zastosowano wyrażenie Excela:

=MODUŁ.LICZBY(A1-\$A\$11).

Dla szeregu rozdzielczego wzór jest następujący:

$$D_x = \frac{\sum_{i=1}^K |\dot{x}_i - \bar{x}| \cdot n_i}{n} \quad (1.3.12)$$

Przykład

Tabela 1.3.15

Obliczenie pomocnicze przy określaniu przeciętnej szeregu rozdzielczego

Przedział	n_i	\dot{x}_i	$ \dot{x}_i - \bar{x} \cdot n_i$	$(\dot{x}_i - \bar{x})^2 \cdot n_i$	$\dot{x}_i - \bar{x}$	$ \dot{x}_i - \bar{x} $
5–9	1	7	13,04	170,0416	-13,04	13,04
10–14	10	12	80,4	646,416	-8,04	8,04
15–19	37	17	112,48	341,9392	-3,04	3,04
20–24	37	22	72,52	142,1392	1,96	1,96
25–29	12	27	83,52	581,2992	6,96	6,96
30–34	2	32	23,92	286,0832	11,96	11,96
35–39	1	37	16,96	287,6416	16,96	16,96

Źródło: Opracowanie własne w Excelu.

Suma wartości w kolumnie $|\dot{x}_i - \bar{x}| \cdot n_i$ wynosi 402,84. Dzieląc przez 100, otrzymamy $D_x \approx 4,03$.

Ponadto do miar zmienności należy S_x^2 , czyli wariancja populacji, którą dla szeregu szczegółowego obliczamy ze wzoru:

$$S_x^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n} . \quad (1.3.13)$$

P r z y k ł a d

Dla danych promieniowania cząstek obliczono wariancję, podnosząc do kwadratu w Excelu wartości obliczeniowe z tab. 1.3.14, np. stosując wyrażenie w Excelu =2,96^2, co daje wynik 8,76.

T a b e l a 1.3.16

Kwadraty odchyłeń od średniej arytmetycznej

8,76	0,00	16,32	4,16	99,20	3,84	35,52	25,40	49,56	4,16
36,48	9,24	81,72	287,64	0,92	16,32	100,80	0,00	3,84	24,60
1,08	1,08	1,08	0,00	64,64	8,76	15,68	9,24	4,16	16,32
48,44	16,32	63,36	35,52	25,40	80,28	1,08	63,36	3,84	9,24
64,64	15,68	0,92	3,84	0,00	25,40	4,16	16,32	8,76	15,68
25,40	15,68	63,36	1,08	15,68	3,84	9,24	1,08	144,96	4,16
9,24	4,16	8,76	0,92	24,60	1,08	0,00	3,84	0,92	0,92
16,32	0,00	1,08	81,72	8,76	9,24	8,76	49,56	9,24	35,52
35,52	36,48	25,40	16,32	48,44	4,16	0,92	15,68	167,96	0,00
0,92	48,44	4,16	3,84	9,24	0,00	36,48	0,92	3,84	1,08
246,82	147,10	266,18	435,06	296,90	152,94	212,66	185,42	397,10	111,70

Ź r ó d ł o: Opracowanie własne w Excelu.

Ostatni wiersz stanowią sumy kolumnowe, a suma z tych sum (2451,84) podzielona przez 100 daje $S_x^2 \approx 24,52$.

Obliczenie wariancji populacji szeregu rozdzielczego podanego w tab. 1.3.15 przeprowadzamy według wzoru:

$$S_x^2 = \frac{\sum_{i=1}^K (\dot{x}_i - \bar{x})^2 \cdot n_i}{n} . \quad (1.3.14)$$

Czasami w obliczeniach szeregu rozdzielczego stosowana jest wariancja próby:

$$S_x^2 = \frac{\sum_{i=1}^K (\dot{x}_i - \bar{x})^2 \cdot n_i}{n-1}.$$

P r z y k ł a d

Wykorzystamy dane w kolumnie $(\dot{x}_i - \bar{x})^2 \cdot n_i$ tab. 1.3.15. Suma wyników w tej kolumnie daje 2455,56, co po podzieleniu przez 100 wynosi $S_x^2 = 24,56$. Gdy podzielimy przez $(n-1) = 99$, wówczas $S_x^2 = 24,8$.

Do określenia miary rozproszenia używa się często odchylenia standardowego S_x :

$$S_x = \sqrt{S_x^2}.$$

Zatem odchylenia standardowe dla przykładu szeregu szczegółowego (tabela 1.3.16) oraz rozdzielczego dla populacji n wynoszą:

$$S_x = \sqrt{24,52} = 4,95; \quad S_x = \sqrt{24,56} = 4,96.$$

Natomiast przyjmując, że obserwacje z promieniowania stanowią próbę, to dla szeregu rozdzielczego:

$$S_x = \sqrt{24,8} = 4,98.$$

Współczynnik zmienności V_x stanowi iloraz odchylenia standardowego do wartości średniej zmiennej:

$$V_x = \frac{S_x}{\bar{x}}. \quad (1.3.15)$$

P r z y k ł a d

Dane jak dla szeregu rozdzielczego próby promieniowania cząstek:
 $S_x = 4,98$; $\bar{x} = 20,04$.

$$V_x = \frac{4,98}{20,04} = 0,25.$$

Natomiast względem miar pozycyjnych:

$$V_x = \frac{Q_x}{M_x}. \quad (1.3.16)$$

Przykładowo dla danych z promieniowania cząstek:

$$V_x = \frac{3,81}{20} = 0,19.$$

Zamiast obliczać tradycyjnie miary rozkładu jednej zmiennej, można posłużyć się programem Excel i skorzystać z opcji (*Analiza danych/Statystyka opisowa*).

P r z y k ł a d

Opcja (*Analiza danych/Statystyka opisowa*) umożliwia obliczenie szeregu miar, w tym już opisanych (tab. 1.3.17 *czcionka pogrubiona*). Niektóre miary z tej opcji obliczane są dla próby, tj. $n-1$ obserwacji. Wyniki dla zmiennej X promieniowania cząstek z licznika Geigera przy $n-1 = 100 - 1 = 99$ obserwacji podaje tabela 1.3.17.

T a b e l a 1.3.17

Statystyka opisowa

Średnia (dla $n = 100$)	20,04	Minimum	8
Mediana	20	Maksimum	37
Odchylenie standardowe ($n-1$)	4,98	Suma (z elementów)	2004
Wariancja próbki ($n-1$)	24,8	Licznik	100
Kurtoza ($n-1$)	0,8	Najw. (2)	33
Skośność	0,4	Najm. (2)	10
Zakres (rozstęp)	29		

Ź r ó d ł o: Opracowanie własne w Excelu; opcja (*Analiza danych/Statystyka opisowa*).

Wymieniona w tabeli 1.3.17 kurtoza oraz skośność zostaną omówiona w dalszej części opracowania. Najw. (2) jest to przedostatnia z największych wartości obserwacji, a najm. (2) to odpowiednio najmniejsza.

1.3.4. Miary asymetrii (skośności)

Miary asymetrii (skośności) rozkładu jednej zmiennej służą do badania, czy obserwacje są równomiernie rozłożone wokół wartości przeciętnej. Zamieszczone w tej części wzory pochodzą z publikacji K. Jajugi³⁴:

a) Klasyczny współczynnik asymetrii, który dla szeregu szczegółowego wyznaczany jest według wzoru:

³⁴ Ibidem, rozdz. 1.4.

$$A_s = \frac{(1/n) \cdot \sum_{i=1}^n (x_i - \bar{x})^3}{s_x^3}. \quad (1.3.17)$$

Dla rozkładów prawostronnie asymetrycznych A_s przyjmuje wartości dodatnie, natomiast dla lewostronnie asymetrycznych jest wielkością ujemną.

b) Pozycyjny współczynnik asymetrii bazujący na kwartylach Q_1 i Q_3 oraz medianie M_x .

$$A_Q = \frac{Q_3 - 2 \cdot M_x + Q_1}{Q_3 - Q_1}. \quad (1.3.18)$$

A_Q przyjmuje wartości w przedziale $(-1, 1)$, natomiast A_s z przedziału $(-\infty, \infty)$.

Przykład

Obliczmy klasyczny współczynnik asymetrii obserwacji ilości promieniowanych cząstek, mając dane $\bar{x} = 20,04$, ale najpierw wyznaczamy:

$$\sum_{i=1}^n (x_i - \bar{x})^3,$$

które wynosi 4935,77. Dla $n = 100$ i wcześniej obliczonego $S_x = 4,95$:

$$A_s = \frac{(1/100) \cdot 4935,77}{4,95^3} = 0,77.$$

Zatem jest to rozkład prawostronny (prawie symetryczny) obliczony dla szeregu szczegółowego. A teraz określimy pozycyjny współczynnik asymetrii na podstawie kwartyli z obserwacji promieniowania cząstek na podstawie szeregu szczegółowego oraz rozdzielczego. Szereg szczegółowy: $Q_1 = 17$; $M_x = 20$; $Q_3 = 23$:

$$A_Q = \frac{23 - 2 \cdot 20 + 17}{23 - 17} = 0.$$

Według tej miary rozkład zbliżony jest do symetrycznego.

Szereg rozdzielczy bazuje na wcześniejszych obliczeniach: $Q_1 = 16,03$; $M_x = 20$; $Q_3 = 23,65$:

$$A_Q = \frac{23,65 - 2 \cdot 20 + 16,03}{23,65 - 16,03} = -0,04.$$

Ponieważ natężenie asymetrii wynosi blisko zero, świadczy to o rozkładzie zbliżonym do symetrycznego, ze skłonnością do asymetrii lewostronnej. Posługując się modułem (*Excel/Analiza danych*), z danych przykładu promie-

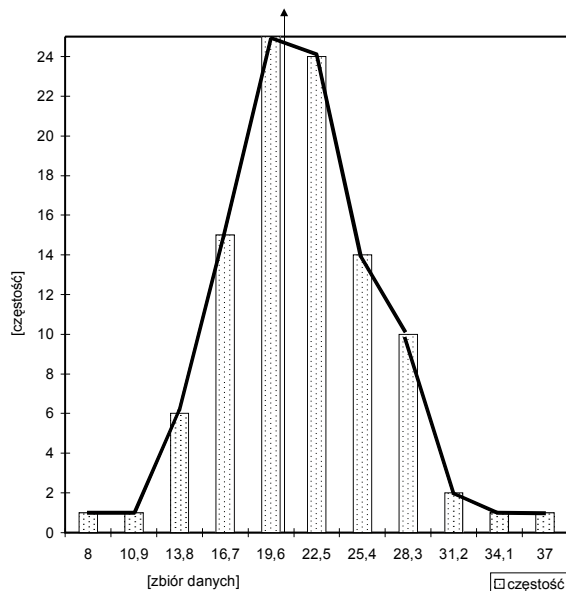
niowania cząstek elementarnych możemy otrzymać tabelę z zakresem przedziałów wartości co 2,9 oraz obraz graficzny w postaci histogramu. Zaprezentowano to w tabeli 1.3.18 oraz na wykresie 1.3.4.

Tabela 1.3.18

Przedziały częstości

Zbiór danych	Częstość	Zakres
8	1	
10,9	1	2,9
13,8	6	2,9
16,7	15	2,9
19,6	25	2,9
22,5	24	2,9
25,4	14	2,9
28,3	10	2,9
31,2	2	2,9
34,1	1	2,9
Więcej	1	

Źródło: Opracowanie własne w Excelu; opcja (*Analiza danych*).



Źródło: Opracowanie własne w Excelu;
opcja (*Analiza danych*).

Wykres 1.3.4. Histogram

1.3.5. Pytania

1. Podaj przykład szeregu szczegółowego uporządkowanego i na jego podstawie określ szereg rozdzielczy zmiennej X .
2. Wymień własności średniej arytmetycznej.
3. Podaj oraz wyjaśnij wzory na średnią harmoniczną ważoną oraz średnią geometryczną ważoną.
4. Co to jest kwartył trzeci?
5. Podaj procedurę sporządzenia wykresu pudełkowego na przykładzie szeregu zmiennej X $\{1, 4, 6, 6, 8, 10, 12, 12, 14, 6, 10, 10, 22, 16, 22, 23, 16, 25, 25, 15, 22, 15, 15, 19, 17, 18, 17, 17, 20, 25\}$.
6. Jak wyznaczamy współczynnik zmienności przykładowej zmiennej X o wartościach podanych w pytaniu piątym?
7. Podaj i wyjaśnij miary asymetrii.

1.3.6. Zadania

1. Oblicz średnią arytmetyczną szeregu $\{22, 16, 23, 29, 15, 22, 19, 17, 18, 20\}$.
2. Oblicz średnią arytmetyczną ważoną szeregu $\{1, 4, 6, 6, 8, 10, 12, 12, 14, 6, 10, 10, 22, 16, 22, 23, 16, 25, 25, 15, 22, 15, 15, 19, 17, 18, 17, 17, 20, 25\}$.
3. Uporządkuj narastająco szereg szczegółowy podany w zadaniu drugim, oblicz medianę oraz określ kwartył pierwszy i kwartył trzeci.
4. Sporządź wykres pudełkowy dla szeregu rozdzielczego, utworzonego z szeregu szczegółowego wymienionego w zadaniu drugim.
5. Oblicz wariancję oraz odchylenie standardowe szeregu szczegółowego podanego w zadaniu drugim.
6. Wyznacz klasyczny współczynnik asymetrii dla szeregu zadania drugiego.

1.4. Rozkład wielu zmiennych

1.4.1. Sposób zapisu

Zjawiska ekonomiczne opisywane są za pomocą co najmniej dwóch zmiennych. Zjawiskiem złożonym jest np. poziom życia określony przez m zmiennych objaśniających tworzących wektor $X = [X_1 X_2 \dots X_m]$. Charakterystyki, analizy oraz wzory zamieszczone w niniejszym podrozdziale zaczerpnięto z publikacji K. Jajugi³⁵.

Przyjmujemy możliwość zapisu rozkładu wielu zmiennych w postaci macierzy n obserwacji m -wymiarowych.

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nm} \end{bmatrix}$$

x_{ij} – wartość j -tej zmiennej dla i -tej obserwacji.

Zależności między parami zmiennych obrazujemy przy pomocy wykresu punktowego (1.4.1), określanego również jako wykres rozrzutu lub korelacyjny.

Przykład

Sporządzamy przykładowy wykres punktowy zależności między dochodem ludności a spożyciem alkoholu. W szerszym zakresie można podjąć badanie *ex post* poziomu życia mieszkańców, np. w latach 1970–1985, biorąc pod uwagę cztery zmienne objaśniające:

- X_1 – spożycie alkoholu w litrach spirytusu/osobę,
- X_2 – dochody ludności w bln zł,
- X_3 – liczba przestępstw w tys.,
- X_4 – liczba rozwodów na 100 osób.

³⁵ Ibidem, rozdz. 2.

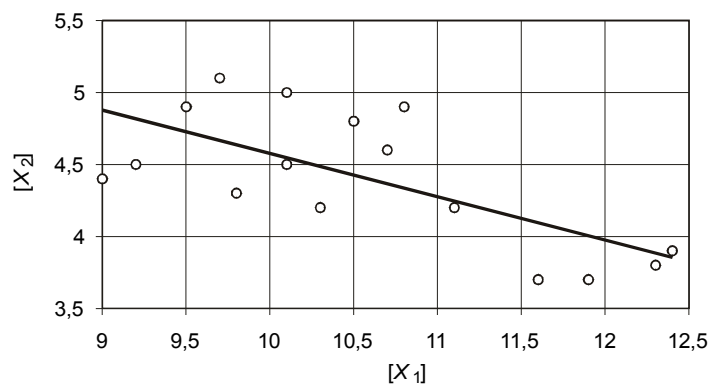
Tabela 1.4.1

Wartości zmiennych

n	X_1	X_2	X_3	X_4
1	9	4,4	424	1,1
2	9,2	4,5	414	1,1
3	9,5	4,9	432	1,1
4	9,7	5,1	414	1,2
5	10,1	5	410	1,3
6	10,8	4,9	394	1,3
7	10,5	4,8	375	1,2
8	10,7	4,6	366	1,2
9	11,1	4,2	348	1,1
10	10,3	4,2	333	1,1
11	9,8	4,3	338	1,1
12	10,1	4,5	380	1,1
13	11,6	3,7	436	1,3
14	11,9	3,7	466	1,4
15	12,3	3,8	539	1,4
16	12,4	3,9	569	1,5

Źródło: Opracowanie własne na podstawie: K. Jajuga (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002, cz. 2 (tab. 6.6).

Na zamieszczonym wykresie oprócz poszczególnych wartości obserwacji zaznaczono również regresję liniową. Widzimy, że gdy dochód maleje, wzrasta spożycie alkoholu.



Źródło: Opracowanie własne w Excelu.

Wykres 1.4.1. Zależności zmiennych X_1 oraz X_2 ; wykres punktowy

Przykład

Na podstawie tab. 1.4.1 określamy macierz symetryczną współczynników kowariancji będącą przykładem charakterystyki rozkładu m -wymiarowego. Kowariancja między zmiennymi X_j a X_l określana jest ze wzoru:

$$\text{cov}_{jl} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l), \quad (1.4.1)$$

gdzie:
 $j, l = 1, 2, \dots, m$.

Tabela 1.4.2

Macierz współczynników kowariancji

Zmienne	X_2	X_1	X_3	X_4
X_2	0,216625			
X_1	-13,292500	4355,983333		
X_3	-0,338417	40,801667	1,125167	
X_4	-0,026125	6,955833	0,116083	0,017625

Źródło: Opracowanie własne w Excelu.

Na głównej przekątnej macierzy kowariancji występują wariancje zmiennych X_j . Macierz kowariancji jest bazą do określenia macierzy korelacji \mathbf{R} .

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1l} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2l} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{j1} & r_{j2} & \dots & r_{jl} & \dots & r_{jm} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{ml} & \dots & 1 \end{bmatrix}$$

Elementy r_{jl} – współczynniki korelacji (zawarte w przedziale $[-1; 1]$) między zmiennymi X_j oraz X_l tej macierzy wyznaczamy dwoma sposobami, korzystając ze wzoru (1.4.2):

$$r_{jl} = \frac{\text{cov}_{jl}}{\sqrt{\text{var}_{jj} \cdot \text{var}_{ll}}}, \quad (1.4.2)$$

którego rozwinięciem jest wzór:

$$r_{jl} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \cdot \sum_{i=1}^n (x_{il} - \bar{x}_l)^2}},$$

gdzie:

$$j, l = 1, 2, \dots, m; S_{jj}^2 = \text{var}_{jj}, S_{ll}^2 = \text{var}_{ll}$$

Przykład

Korzystając z rozwinięcia wzoru 1.4.2, wyznaczamy współczynniki korelacji zmiennych dotyczących poziomu życia.

Tabela 1.4.3

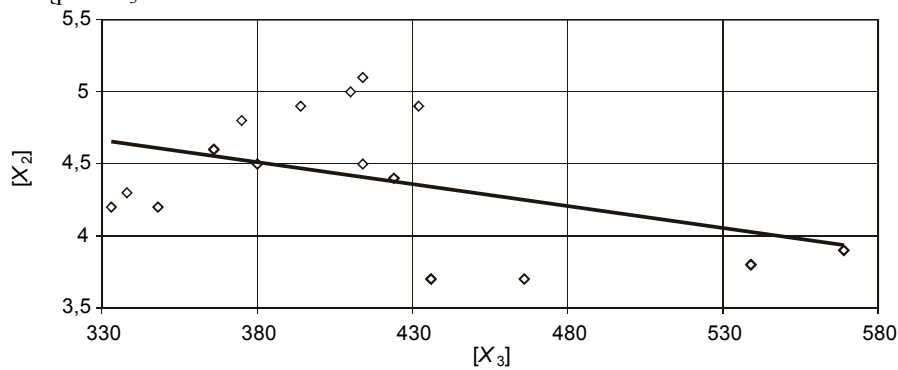
Współczynniki korelacji

Zmienne	Spożycie alkoholu (X_1)	Dochód (X_2)	Liczba przestępstw (X_3)	Liczba rozwodów (X_4)
Spożycie alkoholu	1			
Dochód	-0,649734	1		
Liczba przestępstw	0,425310	-0,336668	1	
Liczba rozwodów	0,769799	-0,327913	0,684701	1

Źródło: Opracowanie własne w Excelu.

Przykład

Sporządźmy jeszcze wykres punktowy zależności między dochodem X_2 a liczbą przestępstw X_3 .



Źródło: Opracowanie własne w Excelu

Wykres 1.4.2. Zależność między dochodem a liczbą przestępstw

Podobnie jak spożycie alkoholu, wraz ze spadkiem dochodu wzrasta liczba przestępstw.

Ponadto charakterystyką rozkładu m -wymiarowego jest wektor miar położenia, którym najczęściej są średnie arytmetyczne z obserwacji zmiennych:

$$\bar{x} = [\bar{x}_1 \bar{x}_2 \dots \bar{x}_j \dots \bar{x}_m].$$

Mając macierz współczynników korelacji r_{jl} między zmiennymi X_j i X_l , możemy określić syntetyczną miarę współzależności liniowej M_R wszystkich składowych wektora $[X_1 X_2 \dots X_m]$, która wyraża się wzorem:

$$M_R = \sqrt{\frac{2}{m \cdot (m-1)} \sum_{j=2}^m \sum_{l < j} r_{jl}^2}, \quad (1.4.3)$$

gdzie:

r_{ij} – współczynnik korelacji między zmiennymi X_j i X_i ,

m – wymiar macierzy korelacji.

P r z y k ł a d

Obliczamy M_R na podstawie współczynników macierzy korelacji z tab. 1.4.3, sporządzając najpierw tablicę pomocniczą kwadratów współczynników korelacji części dolnej macierzy.

T a b e l a 1.4.4

Kwadraty współczynników korelacji

Zmienne	Spożycie alkoholu (X_1)	Dochód (X_2)	Liczba przestępstw (X_3)	Liczba rozwodów (X_4)
Spożycie alkoholu	1			
Dochód	0,422154	1		
Liczba przestępstw	0,180889	0,113345	1	
Liczba rozwodów	0,592591	0,107527	0,468815	1
Σ	1,195634	0,220872	0,468815	1,885321

Ź r ó ł o: Opracowanie własne w Excelu.

$$M_R = \sqrt{\frac{2}{4 \cdot (4-1)} \cdot 1,885321} = 0,56.$$

Miara ta jest przeciętną współczynników korelacji i przyjmuje wartości z przedziału $[<0;1>]$.

1.4.2. Normalizacja

Normalizację stosujemy w celu doprowadzenia do bezwymiarowości zmiennych, a ponadto do zbliżonych rzędów wielkości obserwacji.

1.4.2.1. Standaryzacja

Jednym ze sposobów normalizacji określonej zmiennej X_j dla j -tej obserwacji jest standaryzacja zmiennej realizowana za pomocą wzoru (1.4.4), przy czym $i = 1, 2, \dots, n; j = 1, 2, \dots, m$:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (1.4.4)$$

gdzie:

\bar{x}_j – średnia arytmetyczna zmiennej X_j ,
 s_j – odchylenie standardowe zmiennej X_j .

Przykład

Przeprowadzimy standaryzację zmiennych X_j poziomu życia na próbie 16 obserwacji zamieszczonych w tab. 1.4.1. Ponieważ odchylenie standardowe funkcji Excela obliczane jest odrębnie dla próby oraz populacji, w przykładach zastosowano funkcję ODCH.STANDARDOWE ($n = 15$). Pozostawia się czytelnikowi wykonanie przykładu, stosując funkcję ODCH.STANDARD.POPUL ($n = 16$).

Tabela 1.4.5

Odchylenie standardowe próby ($n-1$)

n	X_1	X_2	X_3	X_4
1	9	4,4	424	1,1
2	9,2	4,5	414	1,1
3	9,5	4,9	432	1,1
4	9,7	5,1	414	1,2
5	10,1	5	410	1,3
6	10,8	4,9	394	1,3
7	10,5	4,8	375	1,2
8	10,7	4,6	366	1,2
9	11,1	4,2	348	1,1
10	10,3	4,2	333	1,1
11	9,8	4,3	338	1,1
12	10,1	4,5	380	1,1
13	11,6	3,7	436	1,3
14	11,9	3,7	466	1,4
15	12,3	3,8	539	1,4
16	12,4	3,9	569	1,5
Średnia	10,563	4,406	414,875	1,219
Odchylenie standardowe	1,061	0,465	66,000	0,133

Źródło: Opracowanie własne, wartości zmiennych według tab. 1.4.1.

Tabela 1.4.6

Zestandardyzowane zmienne „ X_1 ”, „ X_2 ”, „ X_3 ”, „ X_4 ”

$i = n$	Z_1	Z_2	Z_3	Z_4
1	-1,473	-0,013	0,138	-0,894
2	-1,284	0,201	-0,013	-0,894
3	-1,002	1,061	0,259	-0,894
4	-0,813	1,491	-0,013	-0,141
5	-0,436	1,276	-0,074	0,612
6	0,224	1,061	-0,316	0,612
7	-0,059	0,846	-0,604	-0,141
8	0,130	0,416	-0,741	-0,141
9	0,507	-0,443	-1,013	-0,894
10	-0,247	-0,443	-1,241	-0,894
11	-0,719	-0,228	-1,165	-0,894
12	-0,436	0,201	-0,528	-0,894
13	0,978	-1,517	0,320	0,612
14	1,261	-1,517	0,775	1,365
15	1,638	-1,303	1,881	1,365
16	1,732	-1,088	2,335	2,118

Źródło: Opracowanie własne według funkcji Excela.

1.4.2.2. Unitaryzacja

Kolejnym sposobem normalizacji zmiennych jest unitaryzacja przeprowadzona przy wykorzystaniu zależności:

$$z_{ij} = \frac{x_{ij} - \min x_{ij}}{O_j}, \quad (1.4.5)$$

gdzie:

 $i = 1, 2, \dots, n,$ $j = 1, 2, \dots, m,$ O_j – rozstęp zmiennej X_j .Średnia arytmetyczna zmiennej Z_j :

$$\bar{z}_j = \frac{\bar{x}_j - \min x_{ij}}{O_j}. \quad (1.4.6)$$

Przykład

Oblicz średnie i przeprowadź unitaryzację wartości zmiennych poziomu życia.

Tabela 1.4.7

Obliczenia pomocnicze do unitaryzacji zmiennych poziomu życia

n	X_1	X_2	X_3	X_4
1	9	4,4	424	1,1
2	9,2	4,5	414	1,1
3	9,5	4,9	432	1,1
4	9,7	5,1	414	1,2
5	10,1	5	410	1,3
6	10,8	4,9	394	1,3
7	10,5	4,8	375	1,2
8	10,7	4,6	366	1,2
9	11,1	4,2	348	1,1
10	10,3	4,2	333	1,1
11	9,8	4,3	338	1,1
12	10,1	4,5	380	1,1
13	11,6	3,7	436	1,3
14	11,9	3,7	466	1,4
15	12,3	3,8	539	1,4
16	12,4	3,9	569	1,5
$\min x_{ij}$	9	3,7	333	1,1
$\max x_{ij}$	12,4	5,1	569	1,5
Rozstęp ($\max x_{ij} - \min x_{ij}$)	3,4	1,4	236	0,4
Średnia	10,563	4,406	414,875	1,219

Źródło: Opracowanie własne; wartości zmiennych według tab. 1.4.1.

Tabela 1.4.8

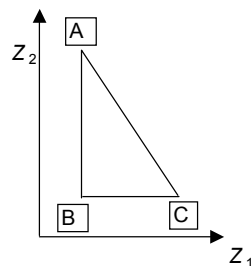
Zunitaryzowane zmienne „ X_1 ”, „ X_2 ”, „ X_3 ”, „ X_4 ”

i	Z_1	Z_2	Z_3	Z_4
1	0,000	0,368	0,385	0,000
2	0,052	0,420	0,343	0,000
3	0,130	0,630	0,419	0,000
4	0,181	0,735	0,343	0,143
5	0,285	0,683	0,326	0,287
6	0,466	0,630	0,258	0,287
7	0,389	0,578	0,178	0,143
8	0,440	0,473	0,140	0,143
9	0,544	0,263	0,063	0,000
10	0,337	0,263	0,000	0,000
11	0,207	0,315	0,021	0,000
12	0,285	0,420	0,199	0,000
13	0,674	0,000	0,436	0,287
14	0,751	0,000	0,563	0,430
15	0,855	0,053	0,872	0,430
16	0,881	0,105	0,999	0,574

Źródło: Opracowanie własne według funkcji Excela.

1.4.2.3. Odległość

Możemy wzajemnie porównywać obserwacje zmiennych poprzez podobieństwo, którego miarą jest odległość euklidesowa lub miejska. Interpretację tych odległości przedstawiono na wykresie.



Źródło: Opracowanie własne na podstawie: K. Jajuga (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002, rys. 2.5.

Wykres 1.4.3. Interpretacja geometryczna odległości euklidesowej i miejskiej

Dwa punkty obserwacji A i C określone przez zmienne znormalizowane Z_1 i Z_2 wyznaczają odległość euklidesową – odcinek AC, a miejska równa się sumie odcinków AB + BC.

P r z y k ł a d

Obliczamy odległość euklidesową oraz miejską między obserwacjami 5. i 10. (tab. 1.4.8) dwóch zmiennych zunitaryzowanych Z_1 oraz Z_2 na podstawie danych o poziomie życia, korzystając ze wzorów (1.4.7) i (1.4.8) i wyrażeń Excela zapisanych dla przykładowych komórek adresowych:

a) Odległość euklidesowa:

$$d_{ii} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{lj})^2}. \quad (1.4.7)$$

Wyrażenie Excela na przykładowych adresach komórek jest następujące:

=PIERWIASTEK((A402-A407)^2)+(B402-B407)^2) → 0,423.

b) Odległość miejska:

$$d_{ii} = \sum_{j=1}^m |z_{ij} - z_{lj}|. \quad (1.4.8)$$

Wyrażenie Excela:

=SUMA(MODUŁ.LICZBY(A402-A407)) i (MODUŁ.LICZBY(B402-B407)) → 0,472.

Zatem odległość euklidesowa jest mniejsza od odległości miejskiej.

T a b e l a 1.4.9

Dane tabeli 1.4.8 z podaniem obserwacji oraz komórek Excela

n	Kol. A (j = 1)	Kol. B (j = 2)	Wiersz
	Z ₁	Z ₂	
1	0,000	0,368	398
2	0,052	0,420	399
3	0,130	0,630	400
4	0,181	0,735	401
5	0,285	0,683	402
6	0,466	0,630	403
7	0,389	0,578	404
8	0,440	0,473	405
9	0,544	0,263	406
10	0,337	0,263	407
11	0,207	0,315	408
12	0,285	0,420	409
13	0,674	0,000	410
14	0,751	0,000	411
15	0,855	0,053	412
16	0,881	0,105	413

Ź r ó d ł o: Opracowanie własne.

Jeśli wykonamy obliczenia określonych odległości, jak w przykładzie, dla wszystkich par obserwacji zmiennych X_1 oraz X_2 , to otrzymamy macierz kwadratową.

$$D = \begin{bmatrix} 0 & d_{12} & \dots & d_{1l} & \dots & d_{1n} \\ d_{21} & 0 & \dots & d_{2l} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{il} & d_{i2} & \dots & d_{il} & \dots & d_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{n1} & d_{n2} & \dots & d_{nl} & \dots & 0 \end{bmatrix}$$

Macierz ta jest symetryczna, a na głównej przekątnej (od lewego górnego rogu do prawego dolnego rogu) występują zera.

P r z y k ł a d

Korzystając z wyrażień Excela, określamy macierz odległości miejskich między zmiennymi zunitaryzowanymi Z_1 oraz Z_2 dla 16 obserwacji.

Tabela 1.4.10

Macierz odległości miejskich między zmiennymi „ Z_1 ” i „ Z_2 ”

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0															
2	0,104	0														
3	0,392	0,288	0													
4	0,549	0,445	0,157	0												
5	0,600	0,496	0,208	0,156	0											
6	0,729	0,625	0,337	0,390	0,234	0										
7	0,599	0,494	0,312	0,365	0,209	0,130	0									
8	0,545	0,441	0,468	0,522	0,365	0,183	0,157	0								
9	0,649	0,650	0,782	0,835	0,679	0,445	0,471	0,314	0							
10	0,442	0,443	0,575	0,628	0,472	0,497	0,367	0,314	0,207	0						
11	0,260	0,260	0,393	0,446	0,445	0,574	0,444	0,391	0,389	0,182	0					
12	0,338	0,233	0,365	0,419	0,263	0,391	0,261	0,208	0,417	0,209	0,183	0				
13	1,041	1,042	1,174	1,227	1,071	0,837	0,863	0,706	0,392	0,599	0,781	0,809	0			
14	1,119	1,120	1,252	1,305	1,149	0,915	0,940	0,783	0,470	0,677	0,859	0,886	0,078	0		
15	1,170	1,171	1,303	1,356	1,200	0,966	0,991	0,835	0,521	0,728	0,910	0,938	0,234	0,156	0	
16	1,143	1,144	1,276	1,330	1,174	0,940	0,965	0,808	0,494	0,702	0,884	0,911	0,312	0,235	0,078	0

Źródło: Opracowanie własne w Excelu.

1.4.2.4. Klasyfikacja obserwacji

W m -wymiarowym układzie współrzędnych mogą istnieć klasy informacji. Stosowana jest metoda środków ciężkości bazująca na minimalizacji funkcji zapisanych na dwa sposoby:

$$C_k = \sum_{i=1}^K \sum_{i \in C_k} \sum_{j=1}^m (z_{ij} - v_{kj})^2 \quad \text{lub} \quad C_k = \sum_{i=1}^K \sum_{i \in C_k} (d_{ij})^2,$$

gdzie:

C_k – i -ta obserwacja należy do k -tej klasy,

v_{kj} – j -ta składowa wektora miar położenia dla obserwacji należących do k -tej klasy.

Minimalizowana funkcja jest zsumowaniem odległości między i -tą obserwacją a wektorem miar położenia danej klasy. Wektorem miar położenia jest wektor średnich obliczony dla obserwacji należących do k -tej klasy.

W metodzie iteracyjnej środków ciężkości rozpoczynamy od ustalenia liczby klas, a następnie:

- sporządzamy klasyfikację początkową (wstępną),
- obliczamy środki ciężkości,
- obliczamy odległości obserwacji od środków ciężkości.

Sposób klasyfikacji początkowej zaprezentowany zostanie na przykładzie zmiennych zestandaryzowanych Z_1 i Z_2 . W przykładzie tym wyodrębniono trzy klasy i przydzielono obserwacje do tych klas, posługując się informacją z wykresu punktowego (wykres 1.4.4) układu dwuwymiarowego Z_1 i Z_2 .

P r z y k ł a d

Sporządzamy klasyfikację początkową zmiennych zestandaryzowanych Z_1 oraz Z_2 , a następnie dzielimy na klasy obserwacji.

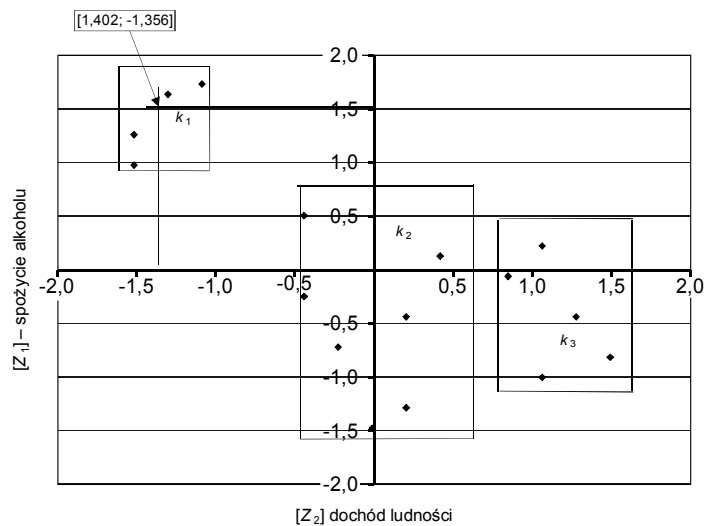
T a b e l a 1.4.11

Przydzielenie obserwacji do trzech klas

i	Z_1	Z_2	k_1	k_3	k_2
1	-1,473	-0,013			
2	-1,284	0,201			
3	-1,002	1,061		k_3	
4	-0,813	1,491		k_3	
5	-0,436	1,276		k_3	
6	0,224	1,061		k_3	
7	-0,059	0,846		k_3	
8	0,130	0,416			k_2
9	0,507	-0,443			k_2
10	-0,247	-0,443			k_2
11	-0,719	-0,228			k_2
12	-0,436	0,201			k_2
13	0,978	-1,517	k_1		
14	1,261	-1,517	k_1		
15	1,638	-1,303	k_1		
16	1,732	-1,088	k_1		

Ź r ó d ł o: Opracowanie własne.

Wstępnie ustalono trzy klasy obserwacji: k_1 , k_2 , k_3 . Na przykład do klasy pierwszej k_1 należą obserwacje: 13, 14, 15, 16.



Ź r ó d ł o: Opracowanie własne.

Wykres 1.4.4. Wyodrębnienie klas obserwacji

Dokonyamy teraz przykładowego obliczenie wektora miar położenia klasy k_1 , posługując się wzorem:

$$v_{kj} = \frac{1}{n_k} \sum_{i \in C_k} z_{ij}. \quad (1.4.9)$$

T a b e l a 1.4.12

Obliczenie wektora miar położenia dla obserwacji klasy K_1

	z_{i1}	z_{i2}
	0,978	-1,517
	1,261	-1,517
	1,638	-1,303
	1,732	-1,088
Σ	5,609	-5,425
v_{kj}	1,402	-1,356

Ź r ó d ł o: Opracowanie własne.

Współrzędne tego wektora są środkami ciężkości klasy k_1 . Analogicznie można obliczyć środki ciężkości klasy k_2 oraz k_3 . Kolejno oblicza się odległości i -tej obserwacji od środków ciężkości poszczególnych wcześniej ustalonych klas.

Mając macierz odległości między poszczególnymi obserwacjami, np. odległości miejskich, możemy skorzystać z innej metody klasyfikacji, tzw. **taksonomii wrocławskiej**. W metodzie tej w pierwszym kroku poszukuje się najbliższego sąsiada każdej obserwacji. Najbliższym sąsiadem i -tej obserwacji jest ta, której odległość od i -tej obserwacji jest najmniejsza.

Przykład

Określamy minimalne odległości miejskie między obserwacjami 1–16 zmiennych zunitaryzowanych Z_1 oraz Z_2 , opierając się na macierzy (tab. 1.4.10).

Tabela 1.4.13

Minimalne odległości miejskie między obserwacjami

Z1	Z2	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	s_i
0,000	0,368	1	0,000	0,104	0,392	0,549	0,600	0,729	0,599	0,545	0,649	0,442	0,260	0,338	1,041	1,119	1,170	1,143	1
0,052	0,420	2	0,104	0,000	0,288	0,445	0,496	0,625	0,494	0,441	0,650	0,443	0,260	0,233	1,042	1,120	1,171	1,144	1
0,130	0,630	3	0,392	0,288	0,000	0,157	0,208	0,337	0,312	0,468	0,782	0,575	0,393	0,365	1,174	1,252	1,303	1,276	2
0,181	0,735	4	0,549	0,445	0,157	0,000	0,156	0,390	0,365	0,522	0,835	0,628	0,446	0,419	1,227	1,305	1,356	1,330	2
0,285	0,683	5	0,600	0,496	0,208	0,156	0,000	0,234	0,209	0,365	0,679	0,472	0,445	0,263	1,071	1,149	1,200	1,174	2
0,466	0,630	6	0,729	0,625	0,337	0,390	0,234	0,000	0,130	0,183	0,445	0,497	0,574	0,391	0,837	0,915	0,966	0,940	3
0,389	0,578	7	0,599	0,494	0,312	0,365	0,209	0,130	0,000	0,157	0,471	0,367	0,444	0,261	0,863	0,940	0,991	0,965	3
0,440	0,473	8	0,545	0,441	0,468	0,522	0,365	0,183	0,157	0,000	0,314	0,314	0,391	0,208	0,706	0,783	0,835	0,808	3
0,544	0,263	9	0,649	0,650	0,782	0,835	0,679	0,445	0,471	0,314	0,000	0,207	0,389	0,417	0,392	0,470	0,521	0,494	4
0,337	0,263	10	0,442	0,443	0,575	0,628	0,472	0,497	0,367	0,314	0,207	0,000	0,182	0,209	0,599	0,677	0,728	0,702	4
0,207	0,315	11	0,260	0,260	0,393	0,446	0,445	0,574	0,444	0,391	0,389	0,182	0,000	0,183	0,781	0,859	0,910	0,884	4
0,285	0,420	12	0,338	0,233	0,365	0,419	0,263	0,391	0,261	0,208	0,417	0,209	0,183	0,000	0,809	0,886	0,938	0,911	4
0,674	0,000	13	1,041	1,042	1,174	1,227	1,071	0,837	0,863	0,706	0,392	0,599	0,781	0,809	0,000	0,078	0,234	0,312	5
0,751	0,000	14	1,119	1,120	1,252	1,305	1,149	0,915	0,940	0,783	0,470	0,677	0,859	0,886	0,078	0,000	0,156	0,235	5
0,855	0,053	15	1,170	1,171	1,303	1,356	1,200	0,966	0,991	0,835	0,521	0,728	0,910	0,938	0,234	0,156	0,000	0,078	6
0,881	0,105	16	1,143	1,144	1,276	1,330	1,174	0,940	0,965	0,808	0,494	0,702	0,884	0,911	0,312	0,235	0,078	0,000	6
Min:			0,104	0,104	0,157	0,156	0,156	0,130	0,130	0,157	0,207	0,182	0,182	0,183	0,078	0,078	0,078	0,078	

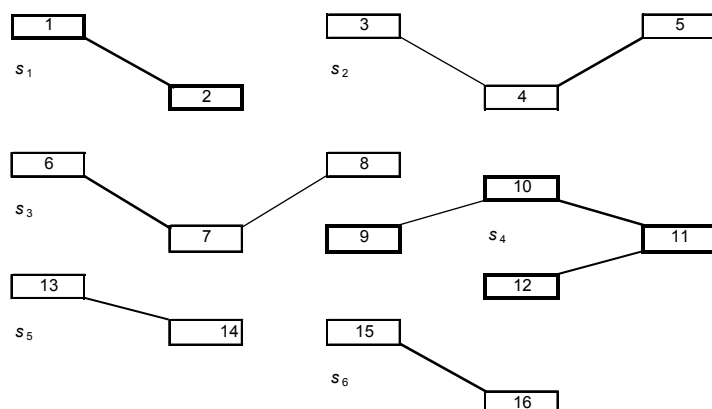
Źródło: Opracowanie własne w Excelu.

W tabeli 1.4.13 podano również kolumnę s_i podziału na skupienia obserwacji, o czym powiemy w dalszej części opracowania. Algorytm taksonomii wrocławskiej można realizować również graficznie, bazując na grafie składającym się z wierzchołków i wiązań. Trzy kolejne etapy tej metody przedstawione zostaną na podstawie informacji z tab. 1.4.13 o minimalnych odległościach miejskich między obserwacjami.

Przykład

Wyłonienie skupień 16 obserwacji o wzajemnych odległościach podanych w tab. 1.4.13.

W grafie wierzchołkami są obserwacje przedstawione np. w postaci prostokątów z numerami obserwacji. Jest n wierzchołków. Łączymy każdą obserwację z najbliższym sąsiadem. Na przykład pierwszą z drugą, gdyż odległość jest minimalna i wynosi 0,104 (tab. 1.4.13). Niektóre połączenia powracają do obserwacji wcześniejszej, przerywając w ten sposób dendryt (graf), np. 1–2, 2–1. W ten sposób otrzymujemy skupienia pierwszego rzędu.



Źródło: Opracowanie własne.

Wykres 1.4.5. Skupienia pierwszego rzędu

W kolejnym etapie wyłaniamy skupienia drugiego rzędu na podstawie skupień pierwszego rzędu. W tym celu dla każdego skupienia pierwszego rzędu s_i wyłaniamy najbliższych sąsiadów p_j w innych skupieniach. Odległość między skupieniami to minimalna odległość między obserwacjami należącymi do tych skupień³⁶. W przykładzie jest sześć skupień pierwszego rzędu. Skupienia łączą się z najbliższym sąsiadem i w ten sposób powstają skupienia drugiego rzędu. Powtarzając ten etap, wszystkie skupienia utworzą spójny graf. Na przykład w określeniu najmniejszej odległości między s_1 a s_2 rozpatrujemy relacje obserwacji, biorąc dane z tab. 1.4.13:

- 1–3: 0,392,
- 1–4: 0,549,
- 1–5: 0,600,
- 2–3: **0,288**,
- 2–4: 0,445,
- 2–5: 0,496.

Odległość minimalna to zatem 0,288. W analogiczny sposób możemy określić odległości s_1 względem dalszych skupień, tj. s_2, s_3, s_4, s_5, s_6 .

³⁶ Ibidem.

Tabela 1.4.14

*Minimalne odległości skupienia „s₁”
do innych skupień*

Skupienie	Odległość
s ₁	0
s ₂	0,288
s ₃	0,441
s ₄	0,233
s ₅	1,041
s ₆	1,143
Min	0,233

Źródło: Obliczenia własne w Excelu.

Analizując poszczególne relacje odległości obserwacji przykładowego skupienia pierwszego z obserwacjami skupień drugiego do szóstego, okazuje się, że najbliższym sąsiadem jest skupienie s₄ o obserwacjach: 9, 10, 11, 12.

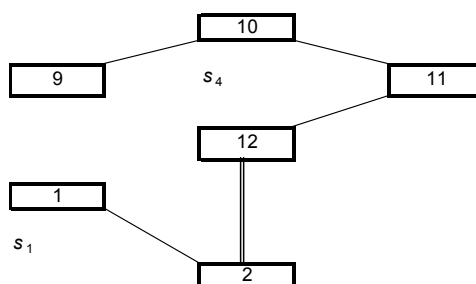
Tabela 1.4.15

Relacje odległości między „s₁” i „s₄”

Obserwacje	9	10	11	12
1	0,649	0,442	0,2605	0,338
2	0,650	0,443	0,2605	0,233

Źródło: Obliczenia własne.

Aby lepiej rozróżnić odległości, zwiększono precyzję wartości wyników obliczeń wykonywanych w Excelu dla relacji: 1–11, 2–11 w stosunku do wartości podanych w tab. 1.4.13. Najbliższe są obserwacje 2–12, które łączymy wiązadłem, tworząc w ten sposób pierwsze skupienie drugiego rzędu.



Źródło: Opracowanie własne.

Wykres 1.4.6. Przykład skupienia drugiego rzędu

Podobnie postępujemy, rozpatrując kolejne skupienia w relacji z pozostałymi. Szukamy dalej skupień trzeciego rzędu. Postępując dalej, utworzymy graf spójny. Pozostaje nam etap trzeci, czyli podział grafu spójnego. Gdy liczba klas jest z góry ustalona, z grafu usuwa się wiązadła o długości d_i większej od odległości krytycznej d^* ustalonej według wzoru Hellwiga:

$$d^* = \bar{d} + k \cdot s_d, \quad (1.4.10)$$

gdzie³⁷:

\bar{d} – średnia arytmetyczna wszystkich odległości z grafu spójnego,

s_d – odchylenie standardowe z powyższych odległości,

k – stała przyjmowana z podziału (0–2), mająca wpływ na liczbę klas.

Bez odniesienia do wzoru Hellwiga można, nawiązując do wcześniejszej klasyfikacji według metody środków ciężkości, podzielić zbiór 16 obserwacji na trzy klasy. Wówczas usuwamy ($K - 1 = 3 - 1 = 2$) wiązadła w grafie spójnym, lecz wprawdzie należałoby zakończyć żmudne prace nad jego tworzeniem.

1.4.2.5. Porządkowanie obserwacji

Jedną z metod jest porządkowanie liniowe zbioru informacji. Zmienne opisujące dane zjawisko, np. poziom życia ludności określonego regionu, mogą mieć charakter stymulant, destymulant czy też nominant. Porządkujemy je od „najlepszej” do „najgorszej”. W przypadku stymulanta jest to wzrost zmiennej X_1 – dochód; wywołuje on wzrost poziomu zjawiska złożonego, jakim jest poziom życia.

P r z y k ł a d

Uporządkujemy zaokrąglone wartości obserwacji z tab. 1.4.16 zmiennych Z_1 – spożycie alkoholu (destymulanta) oraz Z_2 – dochód (stymulanta), przyjmując określone kryteria: najlepsza minimum Z_1 oraz maksimum Z_2 , najgorsza maksimum Z_1 oraz minimum Z_2 .

³⁷ Ibidem.

Tabela 1.4.16

Uporządkowanie wartości obserwacji

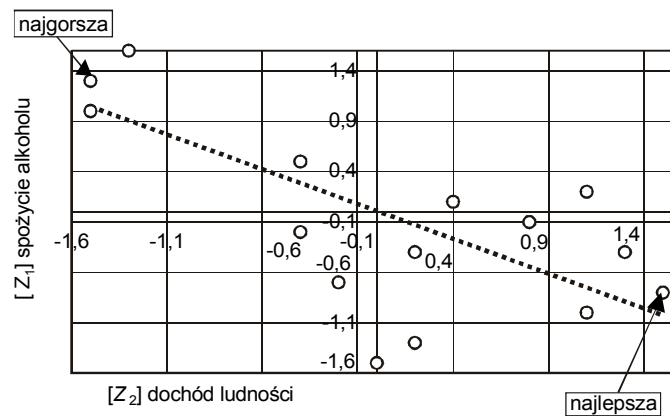
Z_1	Z_2	Kolejność według kryterium	n
-1,5	0,0		1
-1,3	0,2		2
-1,0	1,1		3
-0,8	1,5	1	4
-0,4	1,3		5
0,2	1,1		6
-0,1	0,8		7
0,1	0,4		8
0,5	-0,4		9
-0,2	-0,4		10
-0,7	-0,2		11
-0,4	0,2		12
1,0	-1,5		13
1,3	-1,5	16	14
1,6	-1,3		15
1,7	-1,1		16
-1,5	1,5	min/max	najlepsza
1,7	-1,5	max/min	najgorsza

Źródło: Opracowanie własne.

Kolejność wyboru 1 – najlepsza (obserwacja 4) oraz 16 – najgorsza (obserwacja 14) przede wszystkim z uwagi na stymulantę: dochód ludności.

Przykład

Przedstawmy zmienne Z_1 oraz Z_2 na wykresie punktowym.



Źródło: Opracowanie własne.

Wykres 1.4.7. Wybór obserwacji na wykresie punktowym zmiennych Z_1 i Z_2

Kolejnym sposobem porządkowania zbioru obserwacji jest **metoda wzorca rozwoju**, w której nominanty sprowadzamy do stymulant lub destymulant. W zastosowaniu tej metody wyróżniamy trzy następujące etapy³⁸:

a) Wyznaczenie wzorca rozwoju z_0 o najlepszych wartościach dla każdej zmiennej:

$$z_0 = [z_{01}, z_{02} \dots z_{0j} \dots z_{0m}],$$

przy czym dla obserwacji i -tych:

$$z_{0j} = \max z_{ij}, \text{ gdy zmienna } z_j \text{ jest stymulantą,}$$

$$z_{0j} = \min z_{ij}, \text{ gdy zmienna } z_j \text{ jest destymulantą,}$$

a także antywzorca z_{-0} o najgorszych wartościach każdej zmiennej.

$$z_{-0} = [z_{-01}, z_{-02} \dots z_{-0j} \dots z_{-0m}]$$

Na przykład dla danych zestandaryzowanych Z_1 oraz Z_2 (tab. 1.4.6) określono wzorzec i antywzorzec rozwoju.

$$z_0 = [z_{01}, z_{02}]$$

$$z_0 = [-1,473 \quad 1,491],$$

$$z_{-0} = [z_{-01}, z_{-02}]$$

$$z_{-0} = [1,732 \quad -1,517];$$

b) Określenie odległości euklidesowej obserwacji i -tej od wzorca rozwoju, stosując wzór:

$$d_{i0} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j})^2}. \quad (1.4.11)$$

P r z y k ł a d

Określmy podobieństwo odległości obserwacji zmiennych zestandaryzowanych Z_1 oraz Z_2 , korzystając z przykładowego wyrażenia Excela dla obserwacji pierwszej:

$$=PIERWIASTEK(SUMA((A131-\$A\$147)^2;(B131-\$B\$147)^2)).$$

³⁸ Ibidem, s. 38.

Tabela 1.4.17

Odległości euklidesowe obserwacji
zmiennych „ Z_1 ” i „ Z_2 ”

Z_1	Z_2	d_{i0}
-1,473	-0,013	1,5
-1,284	0,201	1,3
-1,002	1,061	1,5
-0,813	1,491	1,7
-0,436	1,276	1,3
0,224	1,061	1,1
-0,059	0,846	0,8
0,130	0,416	0,4
0,507	-0,443	0,7
-0,247	-0,443	0,5
-0,719	-0,228	0,8
-0,436	0,201	0,5
0,978	-1,517	1,8
1,261	-1,517	2,0
1,638	-1,303	2,1
1,732	-1,088	2,0
-1,473	1,491	z_{0j}
1,732	-1,517	z_{-0j}
-3,205	3,008	

Źródło: Opracowanie własne.

W tabeli 1.4.17 z_{0j} to wzorzec, a z_{-0j} antywzorzec. Na końcu tabeli podano również różnicę między wzorcem a antywzorcem:

$$-1,473 - 1,732 = -3,205; 1,491 - (-1,517) = 3,008.$$

a) Obliczenie dla każdej obserwacji **miary rozwoju**:

$$m_i = 1 - \frac{d_{i0}}{d_0}. \quad (1.4.12)$$

Wartość d_0 , czyli odległość między wzorcem a antywzorcem, obliczamy z zależności:

$$d_0 = \sqrt{\sum_{j=1}^m (z_{0j} - z_{-0j})^2}.$$

P r z y k ł a d

Zastosowano wyrażenie Excela, przykładowo:
=PIERWIASTEK(SUMA((A149^2);(B149^2)))

A149 i B149 to komórki zapisu odpowiednio wartości -3,205 oraz 3,008 (tab. 1.4.17).

W tabeli 1.4.18 najwyższą miarę rozwoju ma obserwacja trzecia, a najniższą piętnasta.

Tabela 1.4.18

Miary rozwoju

d_{i0}	m_i	n	m_i	odp. n
1,5	0,658	1	0,049	15
1,3	0,704	2	0,064	16
0,6	0,855	3	0,075	14
0,7	0,850	4	0,117	13
1,1	0,759	5	0,370	9
1,8	0,602	6	0,479	10
1,6	0,646	7	0,561	8
1,9	0,561	8	0,573	11
2,8	0,370	9	0,602	6
2,3	0,479	10	0,624	12
1,9	0,573	11	0,646	7
1,7	0,624	12	0,658	1
3,9	0,117	13	0,704	2
4,1	0,075	14	0,759	5
4,2	0,049	15	0,850	4
4,1	0,064	16	0,855	3

Źródło: Opracowanie własne.

W metodzie sum standaryzowanych doprowadzenie zmiennych do stymulant następuje poprzez pomnożenie destymulanty przez (-1). Procedura obejmuje dwa etapy:

- a) Wyznaczenie sumy wartości zmiennych³⁹:

$$p_i = \sum_{j=1}^m z_{ij}$$

Aby podkreślić znaczenie jednej lub wielu zmiennych na poziomie zjawiska złożonego, wyznacza się ważoną sumę wartości zmiennych:

$$p_i = \sum_{j=1}^m z_{ij} \cdot w_j,$$

gdzie:

w_j – waga j -tej zmiennej, przy warunku: $\sum w_j = 1$.

P r z y k ł a d

Porządkujemy 16 obserwacji – od najlepszej do najgorszej ze względu na poziom zjawiska złożonego, jakim jest poziom życia, korzystając ze zmiennych zestandaryzowanych Z_1 – spożycie alkoholu oraz Z_2 – dochód.

³⁹ Ibidem, s. 40.

Tabela 1.4.19

Określenie priorytetu obserwacji

Destymulanta Z_1	Stymulanta Z_2	Zamiana na stymulantę [$Z_1 \cdot (-1)$]	i	$p_i = \Sigma[(z_2; z_1 \cdot (-1))]$	$p_i = \Sigma[(z_2 \cdot w_2; z_1 \cdot (-1) \cdot w_1)]$	Posortowanie według p_i (bez wag)	i
-1,473	-0,013	1,473	1	1,460	0,284	2,304	4
-1,284	0,201	1,284	2	1,486	0,418	2,063	3
-1,002	1,061	1,002	3	2,063	1,049	1,712	5
-0,813	1,491	0,813	4	2,304	1,355	1,486	2
-0,436	1,276	0,436	5	1,712	1,108	1,460	1
0,224	1,061	-0,224	6	0,837	0,804	0,905	7
-0,059	0,846	0,059	7	0,905	0,689	0,837	6
0,130	0,416	-0,130	8	0,287	0,307	0,637	12
0,507	-0,443	-0,507	9	-0,950	-0,456	0,491	11
-0,247	-0,443	0,247	10	-0,196	-0,305	0,287	8
-0,719	-0,228	0,719	11	0,491	-0,039	-0,196	10
-0,436	0,201	0,436	12	0,637	0,248	-0,950	9
0,978	-1,517	-0,978	13	-2,496	-1,410	-2,496	13
1,261	-1,517	-1,261	14	-2,778	-1,466	-2,778	14
1,638	-1,303	-1,638	15	-2,941	-1,370	-2,820	16
1,732	-1,088	-1,732	16	-2,820	-1,217	-2,941	15
Wzorzec z_{0j}	1,491	1,473	max	2,304	1,355		
Antywzorzec z_{-0j}	-1,517	-1,732	min	-2,941	-1,466		

Źródło: Opracowanie własne.

Tabela 1.4.19 uwzględnia również sytuację zastosowania wag $w_1 = 0,2$ oraz $w_2 = 0,8$.

Otrzymany wynik obserwacji bez wag: najlepsza 4, a najgorsza – 15. W sytuacji określonych przez badacza przykładowych wag $w_1 = 0,2$; $w_2 = 0,8$: najlepsza również – 4, a najgorsza – 14.

b) Obliczenie dla każdej i -tej obserwacji miary rozwoju według innej zależności niż we wcześniejszej metodzie wzorca rozwoju:

$$m_i = \frac{p_i - p_0}{p_0 - p_{-0}}, \quad (1.4.13)$$

$p_0 = \sum z_{0j} w_j$; $p_{-0} = \sum z_{-0j} w_j$ w obliczeniach dla zmiennych o indeksach $j = 1 - m$.

Przykład

Określamy miary rozwoju obserwacji dwóch zmiennych poziomu życia, biorąc pod uwagę zamieszczone w tab. 1.4.19 wartości wzorca i antywzorca oraz wagi równe 1:

$$p_0 = \sum z_{0j} = 1,491 + 1,473 = 2,964; \quad p_{\cdot 0} = \sum z_{\cdot 0j} = -1,517 + (-1,732) = -2,49;$$

$$p_0 - p_{\cdot 0} = 2,964 - (-2,49) = 5,454.$$

Przykładowe wyrażenie w Excelu zastosowane w tab. 1.4.20 do obliczenia miary rozwoju $m_i = (B92-C88)/(C88-C90)$, przy czym p_0 oraz $p_{\cdot 0}$ zapisano odpowiednio w komórkach C88 oraz C90 arkusza kalkulacyjnego. Dla obserwacji pierwszej: $m_1 = (1,460 - 2,964)/5,454 = -0,274$

Tabela 1.4.20

Miary rozwoju

Wiersz	Kol. B	m_i	i
91	p_i		
92	1,460	-0,242	1
93	1,486	-0,238	2
94	2,063	-0,145	3
95	2,304	-0,106	4
96	1,712	-0,201	5
97	0,837	-0,342	6
98	0,905	-0,331	7
99	0,287	-0,431	8
100	-0,950	-0,630	9
101	-0,196	-0,508	10
102	0,491	-0,398	11
103	0,637	-0,374	12
104	-2,496	-0,879	13
105	-2,778	-0,924	14
106	-2,941	-0,950	15
107	-2,820	-0,931	16
	max	-0,106	najlepsza
	min	-0,950	najgorsza

Źródło: Opracowanie własne w Excelu.

Zatem według metody sum standaryzowanych bez uwzględnienia wag zmiennych najlepsza jest obserwacja 4, a najgorsza 15.

1.4.3. Pytania

1. Podaj i objaśnij wzór na syntetyczną miarę współzależności liniowej.
2. Czym różni się standaryzacja od unitaryzacji?
3. Wyjaśnij pojęcia „odległość euklidesowa” oraz „odległość miejska” pary obserwacji zmiennych X_3 i X_4 .
4. Wymień i scharakteryzuj metody klasyfikacji obserwacji.

5. Objaśnij pojęcia stosowane w taksonometrii wrocławskiej: „skupienia pierwszego rzędu”, „skupienia drugiego rzędu”, „graf spójny”, „wiązańdo”, „odległość krytyczna d^* ”.

6. Wyszczególnij i objaśnij metody porządkowania obserwacji.

1.4.4. Zadania

1. Dana jest macierz symetryczna kwadratów współczynników korelacji. Oblicz syntetyczną miarę współzależności liniowej M_R .

Tabela 1.4.21

Kwadraty współczynników korelacji

Zmienne	Dochód X_2	Liczba przestępstw X_3	Liczba rozwodów X_4
Dochód	1		
Liczba przestępstw	0,11	1	
Liczba rozwodów	0,10	0,47	1

Źródło: Opracowanie własne; dane zaokrąglone do dwóch miejsc po przecinku, według tabeli 1.4.4.

2. Przeprowadź standaryzację zmiennej X_3 (liczba przestępstw) dla 10 obserwacji: 424, 414, 432, 414, 410, 394, 375, 366, 348, 333.

3. W tabeli 1.4.22 dane są szeregi 10 obserwacji zmiennych X_1 oraz X_3 . Opracuj macierz kwadratową odległości euklidesowych wszystkich 10 obserwacji.

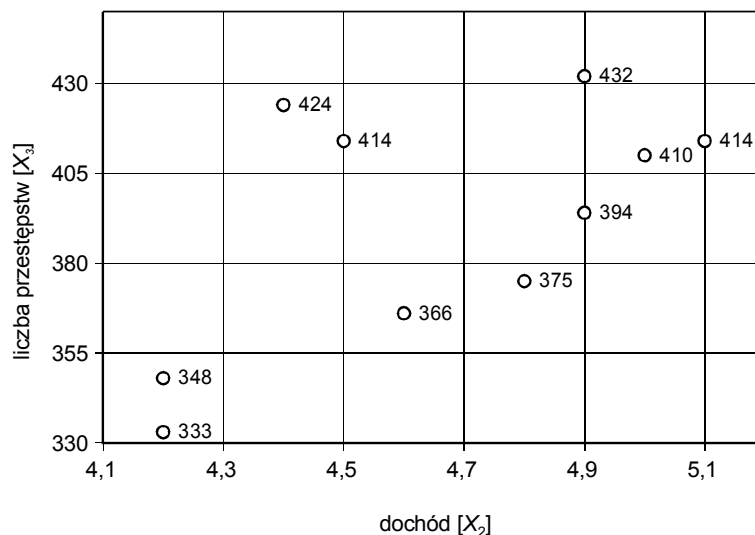
Tabela 1.4.22

Wartości obserwacji „ X_1 ”, „ X_2 ”, „ X_3 ”

Spożycie alkoholu w l/osobę (X_1)	Dochód w mln zł (X_2)	Liczba przestępstw (X_3)
9	4,4	424
9,2	4,5	414
9,5	4,9	432
9,7	5,1	414
10,1	5	410
10,8	4,9	394
10,5	4,8	375
10,7	4,6	366
11,1	4,2	348
10,3	4,2	333

Źródło: Opracowanie własne według danych tab. 1.4.5.

4. W podanym wykresie 1.4.8 wyodrębnij wstępnie klasy obserwacji oraz oblicz środki ciężkości poszczególnych klas.



Ź r ó d ł o: Opracowanie własne.

Wykres 1.4.8. Wykres punktowy zmiennych X_2 i X_3

5. Dane są przykładowe odległości grafu spójnego między obserwacjami:
 1–2: 0,104; 2–3: 0,233; 3–4: 0,183; 4–5: 0,182; 5–6: 0,234.
 Oblicz odległość krytyczną d^* , przyjmując stałą $k = 0,1$.

6. Przyjmując, że dochód X_2 jest stymulantą, a liczba przestępstw X_3 de-
 stymulantą, wyznacz wzorzec rozwoju danych z zadania czwartego.

1.5. Regresja i transformacja liniowa

1.5.1. Idea metody najmniejszych kwadratów

Metoda najmniejszych kwadratów, zwana również klasyczną metodą najmniejszych kwadratów (KMNK), według Maddali⁴⁰ jest nadal „mechanizmem napędowym współczesnej analizy statystycznej”. Oparto na niej różne procedury, a idea sprowadza się do takiego wyznaczenia wartości ocen a_0, a_1, \dots, a_k parametrów strukturalnych $\alpha_0, \alpha_1, \dots, \alpha_k$ liniowego modelu ekonometrycznego, aby suma kwadratów odchyleń zaobserwowanych wartości zmiennej objaśnianej od jej wartości teoretycznych obliczonych z modelu była najmniejsza⁴¹. Postać liniowego modelu jest następująca:

$$Y = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k + \varepsilon.$$

Oszacowaniem ε są reszty. Oznaczmy przez e_t resztę dla $t = 1, 2, \dots, n$, stanowiącą różnicę między wartością zaobserwowaną a obliczoną z modelu po oszacowaniu parametrów:

$$e_t = y_t - \hat{y}_t.$$

Warunek minimalizacji sumy kwadratów odchyleń oraz równanie liniowego modelu ekonometrycznego są następujące:

$$S = \sum_{t=1}^n e_t^2 \rightarrow \min,$$

$$\hat{y}_t = a_0 + a_1 x_{t1} + \dots + a_k x_{tk}.$$

Klasyczna metoda najmniejszych kwadratów oparta jest na następujących założeniach⁴²:

- szacowany model jest modelem liniowym,
- zmienne objaśniające są wielkościami nielosowymi o ustalonych elementach,
- nie występuje zjawisko współliniowości zmiennych objaśniających,

⁴⁰ G.S. Maddala, *Ekonometria...*, op.cit.

⁴¹ Opis KMNK zob.: E. Nowak, *Zarys metod ekonometrii...*, op.cit.

⁴² Ibidem.

– składnik losowy ε ma wartość oczekiwaną równą zero i stałą skończoną wariancję,

– nie występuje zjawisko autokorelacji składnika losowego, czyli zależności składnika losowego w różnych jednostkach czasu.

Gdy liniowy model ekonometryczny ma jedną zmienną objaśniającą, możemy go zapisać w postaci:

$$Y = \beta + \alpha X + \varepsilon.$$

Określamy funkcję S na podstawie warunku minimalizacji sumy kwadratów reszt, przyjmując wartości ocen b i a parametrów strukturalnych β i α :

$$S = \sum_{t=1}^n (y_t - b - ax_t)^2 \rightarrow \min.$$

Jeśli wyznaczymy pochodne cząstkowe tej funkcji względem a oraz b i przyrównamy je do zera, to otrzymujemy układ równań normalnych, których rozwiązanie doprowadza do otrzymania wzorów na oceny parametrów a oraz b :

$$a = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad (1.5.1)$$

$$b = \bar{y} - a\bar{x}, \quad (1.5.2)$$

gdzie:

\bar{y} oraz \bar{x} – średnie arytmetyczne zmiennych Y i X .

Obliczanie średniej wartości obserwacji zmiennych oraz inne obliczenia pomocnicze możemy wykonać w Excelu, według funkcji zamieszczonych w tab. 1.5.1.

Tabela 1.5.1

Wybrane funkcje Excela przydatne w modelowaniu ekonometrycznym

Lp.	Składnia funkcji	Nazwa
1	ŚREDNIA(liczba1;liczba2;...)	Średnia arytmetyczna
2	ODCH.ŚREDNIE(liczba1;liczba2;...)	Odchylenie przeciętne
3	WARIANCJA(liczba1;liczba2;...)	Wariancja próby (dla $n-1$)
4	ODCH. STANDARDOWE(liczba1; liczba2;...)	Odchylenie standardowe próby (dla $n-1$)
5	NORMALIZUJ(x;średnia;odchylenie_std)	Wartość znormalizowana z rozkładu

Lp.	Składnia funkcji	Nazwa
6	KOWARIANCJA(tabela1;tabela2;...)	Średnia z iloczynów odchyłeń
7	MEDIANA(liczba1;liczba2;...)	Mediana
8	SKOŚNOŚĆ(liczba1;liczba2;...)	Skośność rozkładu prawdopodobieństwa
9	KURTOZA(liczba1;liczba2;...)	Kurtoza (względna szczyto- wość/płaskość rozkładu)
10	R.KWADRAT(znane_y;znane_x)	Współczynnik determinacji
11	REGBLSTD(znane_y;znane_x)	Błąd standardowy regresji (reszt)
12	ROZKŁAD.NORMALNY(x;średnia;odchylenie_std; skumulowany)	Skumulowany rozkład normalny
13	ROZKŁAD.NORMALNY.ODW(prawdopodobieństwo;średnia;odchylenie_std)	Odwrotność skumulowanego rozkładu normalnego
14	ROZKŁAD.T(x;stopnie_swobody;strony)	Rozkład <i>t</i> -Studenta
15	ROZKŁAD.T.ODW(prawdopodobieństwo; stopnie_swobody)	Odwrotność rozkładu <i>t</i> -Studenta
16	ROZKŁAD.CHI(x;stopnie_swobody)	Jednostronne prawdopodobieństwo rozkładu χ^2
17	ROZKŁAD.CHI.ODW(prawdopodobieństwo; stopnie_swobody)	Odwrotność jednostronnego prawdopodobieństwa rozkładu χ^2
18	ROZKŁAD.F(x;stopnie_swobody1;stopnie_swobody_2)	Rozkład <i>F</i> prawdopodobieństwa
19	ROZKŁAD.F.ODW(prawdopodobieństwo; stopnie_swobody1;...)	Odwrotność rozkładu <i>F</i> prawdopodobieństwa
20	WSP.KORELACJI(tablica1;tablica2;...)	Współczynnik korelacji
21	REGLINP(znane_y;znane_x;stała;statystyczny)	Tabela opisująca regresję liniową według klasycznej metody najmniejszych kwadratów (KMNK)
22	REGLINW(znane_y;znane_x,nowe_x,stała)	Określenie wartości zmiennej objaśnianej dla regresji liniowej według KMNK i podanego <i>x</i> .
23	REGEXPP(znane_y;znane_x;stała;statystyczny)	Tabela opisująca regresję nieliniową (wykładniczą)
24	REGEXPW(znane_y;znane_x,nowe_x,stała)	Określenie wartości zmiennej objaśnianej dla regresji nieliniowej – wykładniczej i podanego <i>x</i>
25	TRANSPONUJ(tablica)	Transpozycja zakresu danych
26	WYZNACZNIK.MACIERZY(tablica)	Wyznacznik macierzy
27	MACIERZ.ILOCZYN(tablica1;tablica2)	Iloczyn dwóch macierzy
28	MACIERZ.ODW(tablica)	Macierz odwrotna

Lp.	Składnia funkcji	Nazwa
29	SUMA(liczba1;liczba2;...)	Suma liczb
30	SUMA.ILOCZYNÓW(tablica1;tablica2;tablica3;...)	Suma iloczynów odpowiadających sobie składników w tablicach
31	MODUŁ.LICZBY(liczba)	Wartość bezwzględna liczby

Ź r ó d ł o: Opracowanie własne na podstawie: Opis funkcji, Microsoft Excel. Arkusz kalkulacyjny z grafiką prezentacyjną i bazą danych, wersja 4.0 dla Windows TM, Microsoft Corporation 1993.

1.5.2. Określenie regresji względem jednej zmiennej

Do opracowania tego zagadnienia zastosowano rozszerzenie programowe (*Excel/Dodatki/Analiza danych*). Jako dane wykorzystano zmienne: liczba przestępstw Y oraz spożycie alkoholu X .

Tabela 1.5.2

Dane tabeli 1.4.1: Wartości zmiennych

Liczba przestępstw (Y_i)	Spożycie alkoholu (X_i)
424	9
414	9,2
432	9,5
414	9,7
410	10,1
394	10,8
375	10,5
366	10,7
348	11,1
333	10,3
338	9,8
380	10,1
436	11,6
466	11,9
539	12,3
569	12,4

Ź r ó d ł o: Opracowanie własne.

Zastosowanie rozszerzenia (*Excel/Dodatki/Analiza danych*) umożliwia uzyskanie raportów z zakresu statystyki opisowej, oszacowania parametrów, standaryzacji składników resztowych.

Statystyka regresji

Tabela 1.5.3

Statystyka regresji dla $Y = f(x)$

Wielokrotność R	0,582809161	współczynnik korelacji wielorakiej
R^2	0,339666519	współczynnik determinacji
Dopasowany R^2	0,292499841	skorygowany współczynnik determinacji
Błąd standardowy S_{y-x}	55,5144927	standardowy błąd prognozowanej wartości y dla zadanego x
Obserwacje	16	liczba obserwacji

Źródło: Obliczenia własne w Excelu według: K. Jajuga (red.), *Ekonometria. Metody, analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002 (*Dodatki*, tab. 2.5).

Identyczny wynik otrzymujemy przy użyciu funkcji REGBLSTD do regresji zmiennych Y oraz X według wzoru:

$$S_{y-x} = \sqrt{\left[\frac{1}{n(n-2)} \right] \left[n \sum y^2 - (\sum y)^2 - \frac{[n \sum xy - (\sum x)(\sum y)]^2}{n \sum x^2 - (\sum x)^2} \right]} \quad (1.5.3)$$

Parametry modelu regresji

Tabela 1.5.4

Określenie parametrów modelu liniowego

Parametry	Wartości parametrów linii regresji (a_i)	Błąd standardowy (S_i)	$t_{obl} = (a_i/S_i)$
b	31,84942971	143,404404	0,222095
m	36,26277589	13,51301681	2,683544

a_i (a_0, a_1) – oceny parametrów regresji liniowej jednej zmiennej,

s_i (s_1, s_2) – średni błąd szacunku parametrów odpowiednio a_0 i a_1 ,

t_{obl} – statystyka empiryczna testu t -Studenta.

Źródło: Opracowanie własne w Excelu na podstawie: K. Jajuga (red.), *Ekonometria. Metody, analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002, tab. 2.7.

Wiersz b odnosi się do wartości, w której linia regresji przecina oś zmiennej zależnej, czyli do wyrazu wolnego a_0 równania regresji. Wiersz m odnosi się do wartości nachylenia linii regresji, czyli do parametru stojącego przy zmiennej objaśniającej x . Na podstawie danych z tab. 1.5.4 model regresji liniowej po zaokrągleniu parametrów jest następujący:

$$\hat{Y} = mx + b = a_1x + a \cdot 36,26 + 31,85.$$

Standaryzowane składniki resztowe

Program Excel oblicza reszty, uwzględniając wartości obserwacji zmiennych Y i X . Przyjmujemy założenie, że średnia reszt równa się zero, a odchylenie standardowe reszt jest odchyleniem standardowym próby $n - 1 = 15$. Po obliczeniu odchylenia standardowego według funkcji Excela podanej w tab. 1.5.1 poz. 4 uzyskamy wynik 53,6321.

Do standaryzacji reszt stosujemy wzór 1.4.4. Program Excel podaje nam jednak gotowy raport (tab. 1.5.5).

Tabela 1.5.5

Obliczenie standaryzowanych składników resztowych

n	Obserwowana liczba przestępstw	Modelowa liczba przestępstw	Składniki resztowe	Standaryzowane składniki resztowe
1	424	358,2144	65,785587	1,226609
2	414	365,4670	48,533032	0,904925
3	432	376,3458	55,654199	1,037703
4	414	383,5984	30,401644	0,566855
5	410	398,1035	11,896534	0,221817
6	394	423,4874	-29,487409	-0,549809
7	375	412,6086	-37,608577	-0,701233
8	366	419,8611	-53,861132	-1,004271
9	348	434,3662	-86,366242	-1,610346
10	333	405,3560	-72,356021	-1,349118
11	338	387,2246	-49,224633	-0,917820
12	380	398,1035	-18,103466	-0,337549
13	436	452,4976	-16,497630	-0,307607
14	466	463,3765	2,623537	0,048917
15	539	477,8816	61,118427	1,139587
16	569	481,5079	87,492149	1,631339

Źródło: Opracowanie własne w Excelu.

1.5.3. Określenie wielkości statystycznych

Niniejszy podrozdział nawiązuje do obliczeń wykonanych według wzorów, z użyciem funkcji elementarnych Excela, lecz tym razem zastosowano opcję (*Excel/Analiza danych/Korelacja*), która przyspiesza proces obliczeniowy.

Wartości zmiennych objaśniających

Przypomnijmy sobie zmienne objaśniające poziom życia w Polsce w latach 1970–1985:

- realne dochody ludności w cenach z roku 1982 w bln zł (X_2),
- liczba przestępstw wykrytych i zakończonych w postępowaniu przygotowawczym w tys. zł (X_3),
- spożycie alkoholu w przeliczeniu na spożycie czystego spirytusu w litrach/osobę (X_1),
- liczba rozwodów przypadających na 100 osób (X_4).

Tabela 1.5.6

Dane tabeli 1.4.1 w układzie kolumn arkusza kalkulacyjnego

Rok	Dochód (X_2)	Liczba przestępstw (X_1)	Spożycie alkoholu (X_3)	Liczba rozwodów (X_4)
1970	4,4	424	9,0	1,1
1971	4,5	414	9,2	1,1
1972	4,9	432	9,5	1,1
1973	5,1	414	9,7	1,2
1974	5,0	410	10,1	1,3
1975	4,9	394	10,8	1,3
1976	4,8	375	10,5	1,2
1977	4,6	366	10,7	1,2
1978	4,2	348	11,1	1,1
1979	4,2	333	10,3	1,1
1980	4,3	338	9,8	1,1
1981	4,5	380	10,1	1,1
1982	3,7	436	11,6	1,3
1983	3,7	466	11,9	1,4
1984	3,8	539	12,3	1,4
1985	3,9	569	12,4	1,5

Źródło: Opracowanie własne.

Macierz symetryczna współczynników korelacji (pomiędzy czterema zmiennymi objaśniającymi)

Procedura postępowania:

- zapisujemy dane np. w obszarze A1:E17 arkusza kalkulacyjnego Excel,
- wywołujemy opcję (*Analiza danych/Korelacja*),
- zaznaczamy zakres wejściowy danych B1:E17 z podaniem, że pierwszym wierszem są tytuły kolumn zmiennych,
- zaznaczamy wyjście na nowy arkusz danego skoroszytu,
- naciskamy klawisz *OK* i uzyskujemy tabelę wzajemnych korelacji między zmiennymi.

Tabela 1.5.7

Współczynniki korelacji między zmiennymi objaśniającymi

Zmienne	Dochód	Liczba przestępstw	Spożycie alkoholu	Liczba rozwodów
Dochód	1			
Liczba przestępstw	-0,336668	1		
Spożycie alkoholu	-0,649734	0,425310	1	
Liczba rozwodów	-0,327913	0,684701	0,769799	1

Źródło: Obliczenia własne.

Otrzymaliśmy wyniki identyczne jak przy obliczeniu z zastosowaniem funkcji elementarnych Excela.

Macierz symetryczna kowariancji

Obliczenia wykonano z użyciem opcji (*Excel/Analiza danych/Kowariancja*).

Tabela 1.5.8

Współczynniki kowariancji między zmiennymi objaśniającymi

Zmienne	Dochód	Liczba przestępstw	Spożycie alkoholu	Liczba rozwodów
Dochód	0,216625			
Liczba przestępstw	-13,292500	4355,983333		
Spożycie alkoholu	-0,338417	40,801667	1,125167	
Liczba rozwodów	-0,026125	6,955833	0,116083	0,017625

Źródło: Obliczenia własne.

Gdy porównamy wartości z tab. 1.4.2, zauważamy identyczność wyników. W tabeli 1.5.8 elementami głównej przekątnej są wariancje S^2 poszczególnych zmiennych. Przy komputerowym obliczaniu ustalana jest wariancja próbki bazująca na wzorze:

$$S^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{X})^2}{n-1}, \quad (1.5.4)$$

gdzie:

n – ilość obserwacji = 16,

x_{ij} – kolejne wartości j -tej zmiennej objaśniającej w obserwacji i -tej,

\bar{x} – średnia wartość danej zmiennej objaśniającej.

Wielkości statystyczne zmiennych objaśniających

Wielkości statystyczne można wygenerować przez opcję (*Excel/Analiza danych/Statystyka opisowa*). Procedura postępowania jest następująca:

– wprowadzamy zakres wejściowy obszaru adresów np. B1:E17 zmiennych podanych w tabeli 1.5.6 z zaznaczeniem, że pierwszym wierszem są opisy zmiennych,

– określamy wyjście na nowy arkusz i naciskamy klawisz *OK*.

W wyniku działania otrzymujemy raporty ze statystyką opisową do poszczególnych czterech zmiennych objaśniających (tab. 1.5.9 i tab. 1.5.10).

Tabela 1.5.9

Wielkości statystyczne zmiennych objaśniających „ X_1 ” i „ X_3 ”

Miary statystyczne	Dochód	Liczba przestępstw
Średnia	4,406250	414,875000
Błąd standardowy	0,116357	16,499968
Mediana	4,45	412
Tryb (modalna)	4,5	414
Odchylenie standardowe	0,465430	65,999874
Wariancja próbki	0,216625	4355,983333
Kurtoza	-1,155682	1,114396
Skośność (asymetria)	-0,167708	1,109983
Zakres (rozstęp)	1,4	236
Minimum	3,7	333
Maksimum	5,1	569
Suma (liczebności)	70,5	6638
Licznik	16	16

Źródło: Opracowanie własne na podstawie: K. Jajuga (red.), *Ekonometria. Metody, analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002 (*Dodatki*, tab. 2.3).

Tabela 1.5.10

Wielkości statystyczne zmiennych objaśniających X_2 i X_4

Miary statystyczne	Spożycie alkoholu	Liczba rozwodów
Średnia	10,562500	1,218750
Błąd standardowy	0,265185	0,033190
Mediana	10,4	1,2
Tryb (modalna)	10,1	1,1
Odchylenie standardowe	1,060739	0,132759
Wariancja próbki	1,125167	0,017625
Kurtoza	-0,826135	-0,521123
Skośność (asymetria)	0,401254	0,782244
Zakres (rozstęp)	3,4	0,4
Minimum	9	1,1
Maksimum	12,4	1,5
Suma (liczebności)	169	19,5
Licznik	16	16

Źródło: Opracowanie własne.

Wykonanie histogramów zmiennych

Korzystamy z opcji (*Excel/Analiza danych/Histogram*). Procedura postępowania jest następująca:

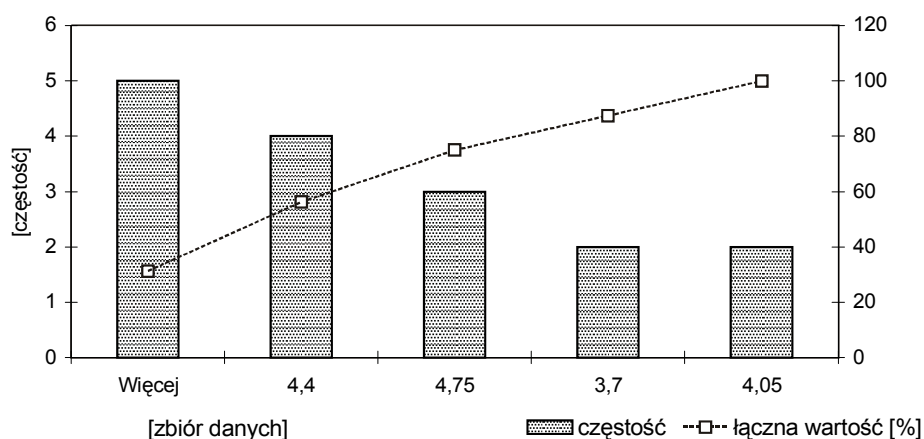
- dla danej zmiennej podajemy zakres komórek, np. dla „Dochodu” B1:B17 z podaniem występowania tytułu w pierwszym wierszu,
- określamy wyjście na nowy arkusz danego skoroszytu,
- podajemy następujące dodatkowe informacje do określenia przez program częstości i narysowania wykresu: posortowanie do wykresu Pareto, podanie łącznego udziału procentowego, podanie na wyjściu wykresu,
- po naciśnięciu klawisza *OK* uzyskamy tabelę częstości oraz wykresy,
- z użyciem kolejnych zmiennych postępujemy analogicznie.

Tabela 1.5.11

Częstość występowania wartości zmiennej „ X_1 ”

Zbiór danych (koszyk)	Częstość (w przedziale)	Łączna wartość [%]	Zbiór danych (koszyk)	Częstość (malejąco)	Łączna wartość [%] (narastająca)
3,7	2	12,50	Więcej	5	31,25
4,05	2	25,00	4,4	4	56,25
4,4	4	50,00	4,75	3	75,00
4,75	3	68,75	3,7	2	87,50
Więcej	5	100,00	4,05	2	100,00

Źródło: Opracowanie własne.



Źródło: Opracowanie własne.

Wykres 1.5.1. Histogram zmiennej X_1 „Dochód”

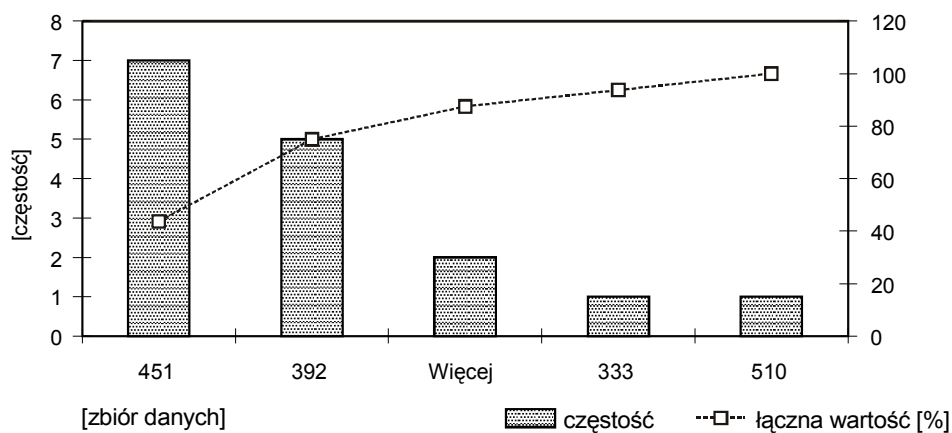
W tabeli 1.5.11 łączna wartość procentowa narastająco, np. 31,25%, jest różnicą $100 - 68,75\%$ według częstości w przedziale. Druga część tabeli stanowi posortowanie malejące według częstości występowania w danym przedziale wartości.

Tabela 1.5.12

Częstość występowania wartości zmiennej „ X_3 ”

Zbiór danych (koszyk)	Częstość	Łączna wartość	Zbiór danych (koszyk)	Częstość (malejąco)	Łączna wartość [%] narastająca
333	1	6,25	451	7	43,75
392	5	37,50	392	5	75,00
451	7	81,25	Więcej	2	87,50
510	1	87,50	333	1	93,75
Więcej	2	100,00	510	1	100,00

Źródło: Opracowanie własne.



Źródło: Opracowanie własne.

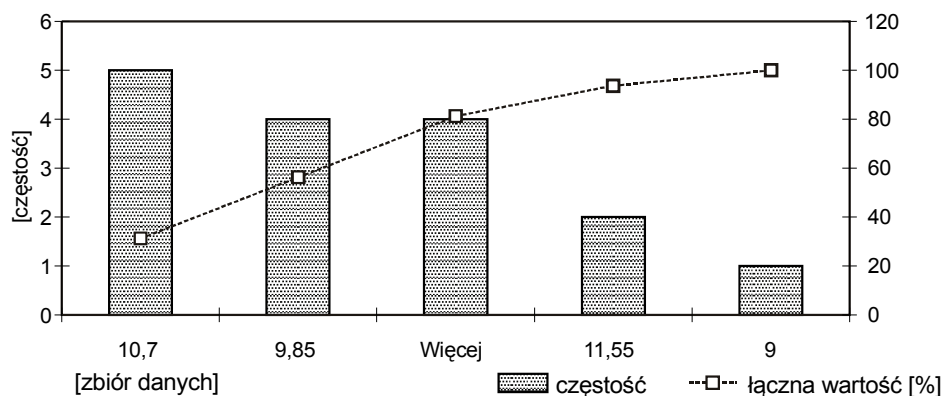
Wykres 1.5.2. Histogram zmiennej X_3 „Liczba przestępstw”

Tabela 1.5.13

Częstość występowania wartości zmiennej „ X_2 ”

Zbiór danych (koszyk)	Częstość	Łączna wartość [%]	Zbiór danych (koszyk)	Częstość (malejąco)	Łączna wartość [%] (narastająco)
9	1	6,25	10,7	5	31,25
9,85	4	31,25	9,85	4	56,25
10,7	5	62,50	Więcej	4	81,25
11,55	2	75,00	11,55	2	93,75
Więcej	4	100,00	9	1	100,00

Źródło: Opracowanie własne.



Źródło: Opracowanie własne.

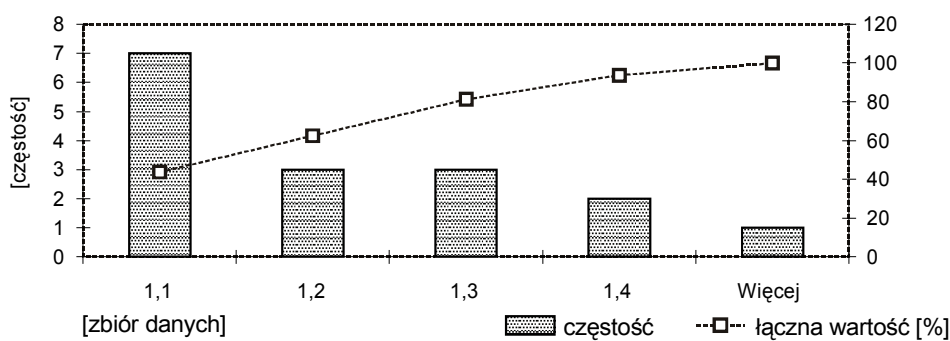
Wykres 1.5.3. Histogram zmiennej X_2 „Spożycie alkoholu”

Tabela 1.5.14

Częstość występowania wartości zmiennej X_4

Zbiór danych (koszyk)	Częstość (w przedziale)	Łączna wartość [%]	Zbiór danych (koszyk)	Częstość (malejąco)	Łączna wartość [%] (narastająca)
1,1	7	43,75	1,1	7	43,75
1,2	3	62,50	1,2	3	62,50
1,3	3	81,25	1,3	3	81,25
1,4	2	93,75	1,4	2	93,75
Więcej	1	100,00	Więcej	1	100,00

Źródło: Opracowanie własne.



Źródło: Opracowanie własne.

Wykres 1.5.4. Histogram zmiennej X_4 „Liczba rozwodów”

1.5.4. Pytania

Podaj procedurę zastosowania przykładowej funkcji statystycznej WSP. KORELACJI programu Excel, mając dane szeregi zmiennych Y oraz X , których wartości umowne zapisane są odpowiednio w komórkach A2:A11 i B2:B11.

Tabela 1.5.15

Szeregi zmiennych Y i X

Y	X
5	1
4	1,5
4,5	2,2
6	3
5,4	3,5
6	4
6,2	4,3
7	5
7,5	6
8	7

Źródło: Opracowanie własne;
dane umowne.

- Wyjaśnij symbole a_0 i a_1 w funkcji regresji liniowej, $\hat{y} = a_0 + a_1x$.
- Podaj różnicę między funkcjami ODCH.STANDARDOWE a ODCH.STAND.POPULACJI.
- Dla szeregów danych z pytania pierwszego napisz wyrażenie w Excelu do obliczenia odchylenia standardowego.
- Przyjmując, że szeregi podane w pytaniu pierwszym mają współczynnik korelacji 0,943, napisz wyrażenie Excela do obliczenia skorygowanego (dopasowanego) współczynnika determinacji bazując na wzorze:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k} \cdot (1 - R^2).$$

- Podaj przykładową składnię funkcji REGBŁSTD do zakresów komórek zmiennej Y oraz zmiennej X z pytania pierwszego.

1.5.5. Zadania

- Przyjmując szeregi zmiennych Y oraz X z pytania pierwszego, zastosuj odpowiednie funkcje Excela do obliczenia:
 - współczynnika korelacji R ,
 - wariancji zmiennych Y (S_y^2) i X (S_x^2) dla próby ($n-1$),
 - odchylenia standardowego zmiennych Y (S_y) oraz X (S_x) dla próby ($n-1$).

2. Przyjmując, że między zmienną zależną Y a zmienną niezależną X występuje regresja liniowa, określ parametry modelu $\hat{y} = a_1x + a_0$, posługując się funkcją REGLINP.

3. Mając model regresji liniowej jako wynik zadania drugiego, oblicz dla kolejnych obserwacji zmiennych z pytania pierwszego standaryzowane składniki resztowe, posługując się tylko wyrażeniami podstawowymi Excela. Wymaga to uprzednio obliczenia wartości średniej reszt e , jak też ich odchylenia standardowego S_e .

4. Zastosuj opcje (*Excel/Analiza danych/Histogram*) do przedstawienia częstości oraz krzywej Pareto (histogramu w postaci narastających częstości) zmiennej X .

1.5.6. Funkcje trendu

Szereg czasowy jest to ciąg obserwacji zjawiska w kolejnych jednostkach czasu. Kolejne jednostki czasu t określamy jako $1, 2, \dots, n$, natomiast obserwację zmiennej objaśnianej jako y_t . Aby zaobserwować tendencję zmian określonego zjawiska wyrażonego przez zmienną Y , zachodzi potrzeba wyrównania szeregu czasowego.

Model szeregu czasowego jest następujący:

$$Y = f(t) + \varepsilon,$$

gdzie:

$f(t)$ – funkcja trendu,

ε – odchylenie losowe.

Gdy oszacujemy parametry funkcji trendu, otrzymamy równanie trendu:

$$\hat{Y} = a_0 + a_1t. \quad (1.5.5)$$

W zastosowaniach występują trendy: liniowy, wykładniczy, potęgowy, logistyczny, hiperboliczny oraz pełzający. W dalszej części przedstawiono postaci analityczne funkcji trendu oraz sposoby oszacowania parametrów⁴³.

1.5.6.1. Trend liniowy

W równaniu trendu liniowego a_0, a_1 stanowią oceny parametrów funkcji trendu. Parametry te szacuje się klasyczną metodą najmniejszych kwadratów, posługując się wzorami (1.5.1) oraz (1.5.2). Do realizacji przykładu przypomnijmy sobie te wzory zamieniając x zmienną czasową t :

⁴³ K. Jajuga (red.), *Ekonometria. Metody...*, op.cit, rozdz. 3.

$$a_1 = \frac{\sum_{t=1}^n (y_t - \bar{y}) \cdot (t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2}, \quad a_0 = \bar{y} - a_1 \cdot \bar{t}.$$

Uwaga:

Ze względu na możliwości edycyjne w Excelu wartości średnie oraz inne oznaczenia, które wystąpią w dalszej części opracowania, wpisano obok zmiennej zamiast nad zmienną.

P r z y k ł a d

Określmy trend liniowy posługując się danymi umownymi określającymi inflację z przedziału dwunastu lat, mając dane obserwacje zmiennej objaśnianej Y .

T a b e l a 1.5.16

Obliczenia pomocnicze do oszacowania parametrów modelu zmiennej „ Y ”

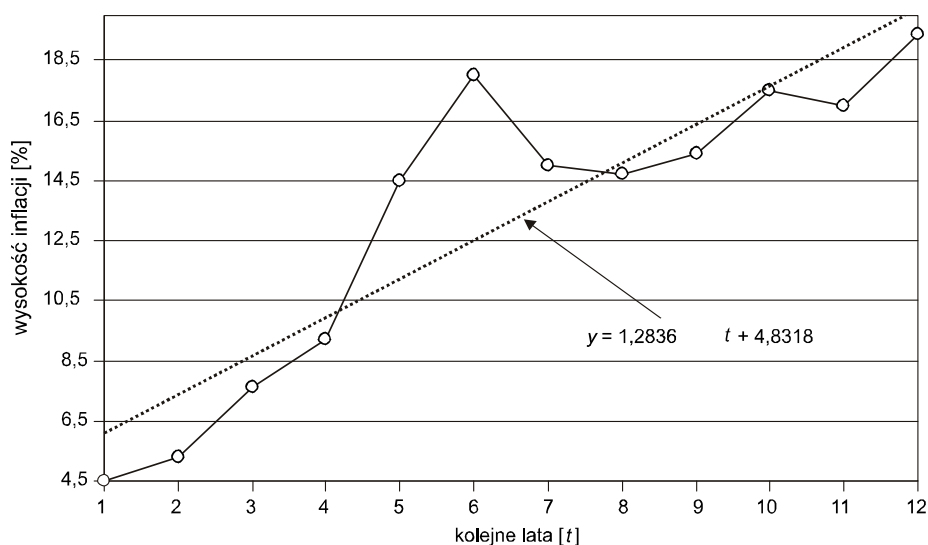
t	y_t	$y_t - y^-$	$t - t^-$	$(t - t^-)^2$	$(y_t - y^-)(t - t^-)$
1	4,5	-8,7	-5,5	30,3	47,7
2	5,3	-7,9	-4,5	20,3	35,4
3	7,6	-5,6	-3,5	12,3	19,5
4	9,2	-4,0	-2,5	6,3	9,9
5	14,5	1,3	-1,5	2,3	-2,0
6	18	4,8	-0,5	0,3	-2,4
7	15	1,8	0,5	0,3	0,9
8	14,7	1,5	1,5	2,3	2,3
9	15,4	2,2	2,5	6,3	5,6
10	17,5	4,3	3,5	12,3	15,1
11	17	3,8	4,5	20,3	17,2
12	19,4	6,2	5,5	30,3	34,2
6,5	13,2		Σ	143,0	183,6

Ź r ó d ł o: Opracowanie własne w Excelu, gdzie: $y^- \rightarrow \bar{y}$, $t^- \rightarrow \bar{t}$.

Obliczenie parametrów i określenie modelu:

$$a_1 = \frac{183,6}{143} = 1,2839; \quad a_0 = 13,2 - 1,2839 \cdot 6,5 = 4,8546;$$

$$\hat{Y} = 4,8546 + 1,2839t.$$



Ź r ó d ł o: Opracowanie własne w Excelu.

Wykres 1.5.5. Trend liniowy. Inflacja

Różnica w wartościach parametrów modelu oszacowanych analitycznie oraz podanych w formie funkcji trendu na wykresie wynika z dużej precyzji obliczeń komputerowych.

1.5.6.2. Trend pełzający

Założmy, że mamy szereg czasowy y_1, y_2, \dots, y_n zmiennej Y w przedziale czasowym $t = 1-n$. Określamy długość segmentu $k \leq n$, a następnie obliczamy trendy dla k -elementowych podszeregów czasowych, których jest $n-k-1$. W kolejności określamy wartości modelowe według funkcji trendu liniowego zmiennej Y dla obserwacji podszeregów. Mając te obliczenia, określamy średnie wartości na podstawie \hat{y}_t kolejnych obserwacji. Podaną procedurę zdemonstrujemy na przykładzie.

Przykład

Określmy podszeregi czasowe oraz parametry trendów liniowych segmentów trzyelementowych.

Określenie podszeregów czasowych $n - k - 1 = 12 - 3 + 1 = 10$

Tabela 1.5.17

Podszeregi czasowe trzelementowe

t	y_t	Podszeregi (ciągi)			
1	4,5	4,5	5,3	7,6	1
2	5,3	5,3	7,6	9,2	2
3	7,6	7,6	9,2	9,2	3
4	9,2	9,2	9,2	14,5	4
5	9,2	9,2	14,5	18	5
6	14,5	14,5	18	14,7	6
7	18	18	14,7	15,4	7
8	14,7	14,7	15,4	17,5	8
9	15,4	15,4	17,5	17	9
10	17,5	17,5	17	19,4	10
11	17				
12	19,4				

Źródło: Opracowanie własne.

Oszacowanie parametrów

Na podstawie pomocniczej tab. 1.5.17 szacujemy parametry trendów liniowych poszczególnych segmentów według KMNK. Na przykład obserwacje 1–3 pierwszego segmentu według zależności:

$$a_1^{(1)} = \frac{\sum_{t=1}^{1+k-1} (y_t - \bar{y}) \cdot (t - \bar{t})}{\sum_{t=1}^{1+k-1} (t - \bar{t})^2}; \quad a_0^{(1)} = \bar{y} - a_1^{(1)} \cdot \bar{t}.$$

Dla pozostałych dziewięciu segmentów obliczenia będą analogiczne.

Wartości średnie \bar{y} oraz \bar{t} obejmują trzy elementy danego podszeregu. Ze względu na znaczną pracochłonność obliczenia poszczególnych parametrów a_0 i a_1 kolejnych podszeregów czasowych można zastosować funkcję REGLINP Excela. Rezultat tego w formie raportu podaje tab. 1.5.18.

Tabela 1.5.18

Obliczenia pomocnicze do określenia wartości modelowych trendu pełzającego

Segment	a_0	a_1	1	2	3	4	5	6	7	8	9	10	11	12
1	2,7	1,55	4,25	5,80	7,35									
2	1,5167	1,95	$u = 1$	5,42	7,37	9,32								
3	5,4667	0,8		$u = 2$	7,87	8,67	9,47							
4	-2,2833	2,65			$u = 3$	8,32	10,97	13,62						
5	-12,5	4,4				$u = 3$	9,50	13,90	18,30					
6	15,033	0,1						15,63	15,73	15,83				
7	26,433	-1,3							17,33	16,03	14,73			
8	3,2667	1,4								14,47	15,87	17,27		
9	8,6333	0,8									15,83	16,63	17,43	
10	7,5167	0,95										17,02	17,97	18,92
	Średnia		4,25	5,61	7,53	8,77	9,98	14,38	17,12	15,44	15,48	16,97	17,70	18,92

Źródło: Opracowanie własne w Excelu.

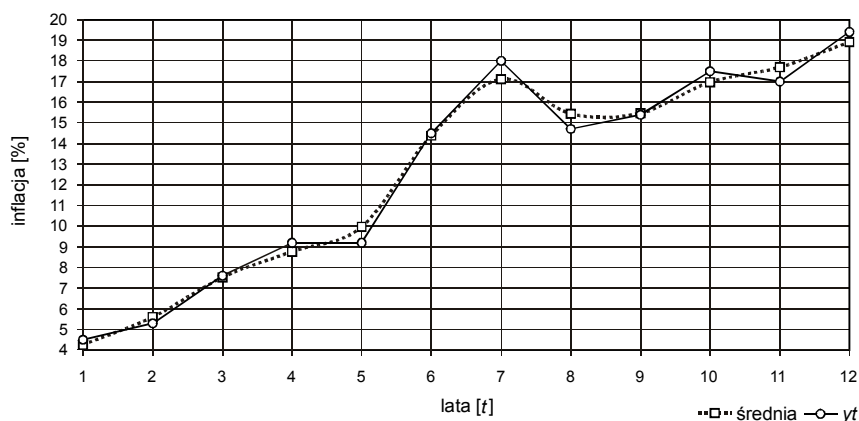
Mając równania trendów segmentowych, obliczono modelowe (teoretyczne) wartości zmiennej Y . W kolejności, biorąc pod uwagę ilości określonych wartości teoretycznych, dla poszczególnych jednostek czasu (okresów 1–12) obliczono średnie arytmetyczne tych wartości.

W tabeli 1.5.18 u oznacza ilość elementów do obliczenia średniej. Wartość teoretyczną pierwszej obserwacji pierwszego segmentu obliczono następująco:

$$\hat{y}_1 = 2,7 + 1,55t = 2,7 + 1,55 \cdot 1 = 4,25.$$

P r z y k ł a d

Porównać dane empiryczne i teoretyczne (powyższych średnich) poszczególnych obserwacji.



Ź r ó d ł o: Opracowanie własne, gdzie: $yt \rightarrow y_t$.

Wykres 1.5.6. Trend pełzający

Poszczególne wartości średnich trendu pełzającego połączone linią wygładzoną. Zwróćmy uwagę na dobre dopasowanie danych z obserwacji i modelowych.

1.5.6.3. Trend potęgowy

Aby zastosować KMNK do trendów nieliniowych, musimy dokonać transformacji liniowej, przekształcając odpowiednio funkcję danego trendu do postaci trendu liniowego. Trend potęgowy określamy równaniem:

$$\hat{Y} = a_0 t^{a_1}. \quad (1.5.6)$$

Przykład

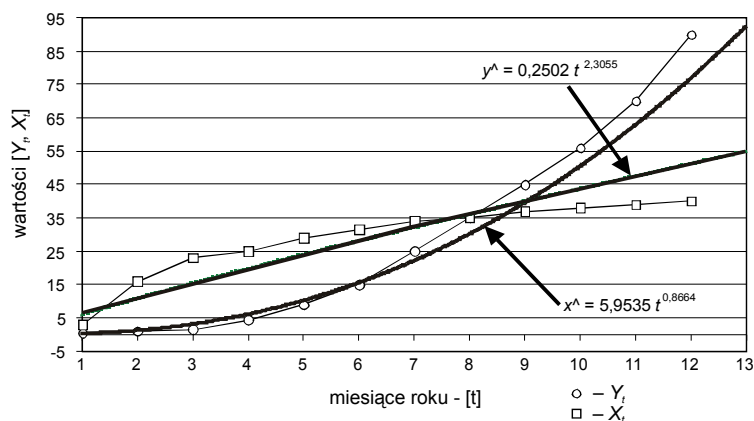
Dla zmiennych Y_t i X_t o trendzie potęgowym podanych w tab. 1.5.19 sporządzamy w Excelu wykres.

Tabela 1.5.19

Dane umowne do trendu potęgowego zmiennych „Y” i „X”

t	y_t	x_t
1	0,5	3,0
2	1	16,0
3	1,5	23,0
4	4,4	25,0
5	9	29,0
6	15	31,5
7	25	34,0
8	35	35,0
9	45	37,0
10	56	38,0
11	70	39,0
12	90	40,0

Źródło: Opracowanie własne.



Źródło: Opracowanie własne, gdzie: $y^{\wedge} \rightarrow \hat{y}$, $x^{\wedge} \rightarrow \hat{x}$.

Wykres 1.5.7. Trend potęgowy zmiennych Y i X

Według wykresu \hat{Y} ma parametry: $a_0 = 0,2502$, $a_1 = 2,3055$, natomiast $\hat{X} - a_0 = 5,9535$, $a_1 = 0,8684$.

Uwaga: w przypadku $a_1 < 1$ Excel rysuje trend liniowy zamiast potęgowego.

Po obustronnym zlogarytmowaniu funkcji trendu potęgowego otrzymujemy:

$$\ln \hat{Y} = \ln a_0 + a_1 \ln t.$$

Jeśli podstawimy $\hat{Y}' = \ln \hat{Y}$, $a'_0 = \ln a_0$, $t' = \ln t$, to dokonamy transformacji liniowej:

$$\hat{Y}' = a'_0 + a_1 t'$$

Przykład

Obliczamy parametry a_0 oraz a_1 trendu potęgowego zmiennej Y , korzystając ze wzorów KMNK:

$$a_1 = \frac{\sum_{i=1}^n (y'_i - \bar{y}') \cdot (t' - \bar{t}')}{\sum_{i=1}^n (t' - \bar{t}')^2}, \quad a'_0 = \bar{y}' - a_1 \cdot \bar{t}'$$

Tabela 1.5.20

Obliczenia pomocnicze do oszacowania parametrów modelu zmiennej „Y”

t	y	x	y'	y' - y'	t'	t' - t'	(y' - y')(t' - t')	(t' - t') ²
1	0,5	3,0	-0,693	-3,148	0,000	-1,666	5,243	2,774
2	1	16,0	0,000	-2,454	0,693	-0,972	2,387	0,946
3	1,5	23,0	0,405	-2,049	1,099	-0,567	1,162	0,321
4	4,4	25,0	1,482	-0,973	1,386	-0,279	0,272	0,078
5	9	29,0	2,197	-0,257	1,609	-0,056	0,014	0,003
6	15	31,5	2,708	0,254	1,792	0,126	0,032	0,016
7	25	34,0	3,219	0,764	1,946	0,280	0,214	0,079
8	35	35,0	3,555	1,101	2,079	0,414	0,456	0,171
9	45	37,0	3,807	1,352	2,197	0,532	0,719	0,283
10	56	38,0	4,025	1,571	2,303	0,637	1,001	0,406
11	70	39,0	4,248	1,794	2,398	0,732	1,314	0,536
12	90	40,0	4,500	2,045	2,485	0,819	1,676	0,671
Średnia			2,454		1,666	Σ	14,488	6,284

Źródło: Opracowanie własne w Excelu gdzie: $y' \rightarrow \bar{y}'$; $t' \rightarrow \bar{t}'$.

$$a_1 = \frac{14,448}{6,284} \rightarrow 2,3055; \quad a'_0 = 2,454 - 2,3055 \cdot 1,666 \rightarrow -1,3856.$$

Wartość a_1 i a'_0 wynika z dokładności obliczeń w Excelu (patrz wykres 1.5.7 – model \hat{Y}).

$a'_0 = \ln a_0$; $a_0 = e^{a'_0} = 2,718282^{-1,3856} = 0,2502$ lub wyrażeniem Excela =EXP(-1,3856).

Przykład

Obliczamy w analogiczny sposób parametry a_0 i a_1 zmiennej X .

Tabela 1.5.21

Obliczenia pomocnicze do oszacowania parametrów modelu zmiennej „ X ”

t	x	x'	$x' - x''$	t'	$t' - t''$	$(x' - x'')(t' - t'')$	$(t' - t'')^2$
1	3	1,099	-2,129	0,000	-1,666	3,545	2,774
2	16	2,773	-0,455	0,693	-0,972	0,442	0,946
3	23	3,135	-0,092	1,099	-0,567	0,052	0,321
4	25	3,219	-0,008	1,386	-0,279	0,002	0,078
5	29	3,367	0,140	1,609	-0,056	-0,008	0,003
6	31,5	3,450	0,223	1,792	0,126	0,028	0,016
7	34	3,526	0,299	1,946	0,280	0,084	0,079
8	35	3,555	0,328	2,079	0,414	0,136	0,171
9	37	3,611	0,384	2,197	0,532	0,204	0,283
10	38	3,638	0,410	2,303	0,637	0,261	0,406
11	39	3,664	0,436	2,398	0,732	0,320	0,536
12	40	3,689	0,462	2,485	0,819	0,378	0,671
Średnia		3,227		1,666	Σ	5,445	6,284

Źródło: Opracowanie własne w Excelu, gdzie $x'' \rightarrow \bar{x}$; $t'' \rightarrow \bar{t}$.

$$a_1 = \frac{5,445}{6,284} = 0,8664; \quad a'_0 = 3,227 - 0,8664 \cdot 1,666 = 1,7884;$$

$$a'_0 = \ln a_0 \quad a_0 = e^{a'_0} = 2,718282^{1,7884} = 5,9535$$

lub wyrażeniem Excela =EXP(1,7840).

Zatem dla dwóch modeli wartości parametrów uzyskanych na wykresie, jak też obliczonych analitycznie, są identyczne.

1.5.6.4. Trend wykładniczy

Trend wykładniczy określony jest równaniem:

$$\hat{Y} = a_0 a_1^t. \quad (1.5.7)$$

W tabeli 1.5.22 podano szeregi czasowe zmiennych Y i X o przewidywanym trendzie wykładniczym.

Przykład

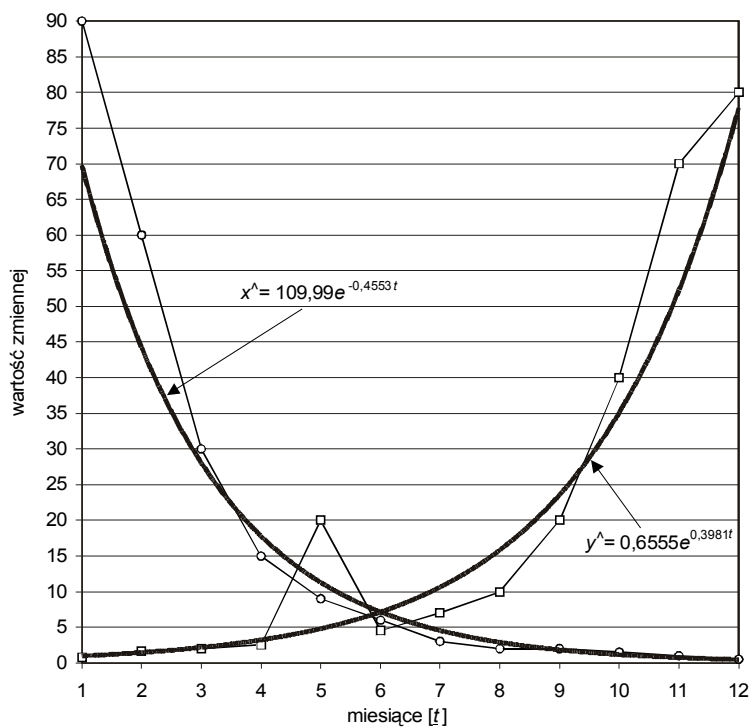
Sporządzamy w Excelu wykres zmiennych oraz uzyskujemy model \hat{X} oraz \hat{Y} .

Tabela 1.5.22

Dane modelowe do trendów
wykładniczych zmiennych „Y” i „X”

t	y	x
1	0,8	90,0
2	1,7	60,0
3	2	30,0
4	2,5	15,0
5	20	9,0
6	4,5	6,0
7	7	3,0
8	10	2,0
9	20	2,0
10	40	1,5
11	70	1,0
12	80	0,5

Źródło: Opracowanie własne
w Excelu.



Źródło: Opracowanie własne w Excelu, gdzie: $\hat{x} \rightarrow \hat{x}$; $\hat{y} \rightarrow \hat{y}$.
Wykres 1.5.8. Trend wykładniczy zmiennych Y i X

Według wykresu trend zmiennej X ma parametry $a_0 = 109,99$, $a_1 = e^{-0,4553} = 0,634 < 1$, a zmiennej Y $a_0 = 0,6555$, $a_1 = e^{0,3981} = 1,489 > 1$.

Po przekształceniu zapis funkcji na wykresie w postaci trendu wykładniczego dla zmiennych X oraz Y ma postać modeli:

$$\hat{X} = 109,99 \cdot 0,634^t; \quad \hat{Y} = 0,6555 \cdot 1,489^t.$$

Aby zastosować KMNK, sprowadzamy manualnie trend wykładniczy modelu zmiennej Y do postaci liniowej przez obustronne logarytmowanie.

$$\ln \hat{Y} = \ln a_0 + t \cdot \ln a_1.$$

Jeżeli $\ln \hat{Y} = \hat{Y}'$, $\ln a_0 = a'_0$, $\ln a_1 = a'_1$,
to otrzymamy funkcję liniową $\hat{Y}' = a'_0 + a'_1 t$.

Parametry a'_0 oraz a'_1 obliczamy jak w trendzie liniowym, używając elementów oznaczonych prim (').

$$a'_1 = \frac{\sum_{t=1}^n (y'_t - \bar{y}') \cdot (t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2}; \quad a'_0 = \bar{y}' - a'_1 \cdot \bar{t}.$$

Przykład

Oszacujmy parametry trendu wykładniczego zmiennej Y , wykorzystując powyższe sprowadzenie tego trendu do postaci trendu liniowego i stosując wzory KMNK.

Tabela 1.5.23

Obliczenia pomocnicze do oszacowania parametrów modelu zmiennej „Y”

t	y	y'	$(y' - \bar{y}')$	$(t - \bar{t})$	$(t - \bar{t})^2$	$(y' - \bar{y}') \cdot (t - \bar{t})$
1	0,8	-0,223	-2,388	-5,5	30,25	13,135
2	1,7	0,531	-1,634	-4,5	20,25	7,355
3	2	0,693	-1,472	-3,5	12,25	5,152
4	2,5	0,916	-1,249	-2,5	6,25	3,122
5	20	2,996	0,831	-1,5	2,25	-1,246
6	4,5	1,504	-0,661	-0,5	0,25	0,330
7	7	1,946	-0,219	0,5	0,25	-0,110
8	10	2,303	0,138	1,5	2,25	0,206
9	20	2,996	0,831	2,5	6,25	2,077
10	40	3,689	1,524	3,5	12,25	5,333
11	70	4,248	2,083	4,5	20,25	9,376
12	80	4,382	2,217	5,5	30,25	12,193
6,5	Srednia	2,1653		Σ	143	56,924

Źródło: Opracowanie własne w Excelu, gdzie $y'' \rightarrow \bar{y}'$.

$$a_1' = \frac{56,924}{143} = 0,3981, \quad a_0' = 2,165 - 0,3981 \cdot 6,5 = -0,42265.$$

$$a_0 = e^{(-0,42265)} = 2,718282^{(-0,42265)} \rightarrow 0,6555.$$

Otrzymany analitycznie wynik a_0' dla zmiennej Y jest zgodny z wykresem 1.5.8.

1.5.6.5. Trend logarytmiczny

Aby sprowadzić trend logarytmiczny wyrażony równaniem:

$$\hat{Y} = a_0 + a_1 \ln t \quad (1.5.8)$$

do postaci trendu liniowego, zastosujemy podstawienie:

$$t' = \ln t.$$

Przykład

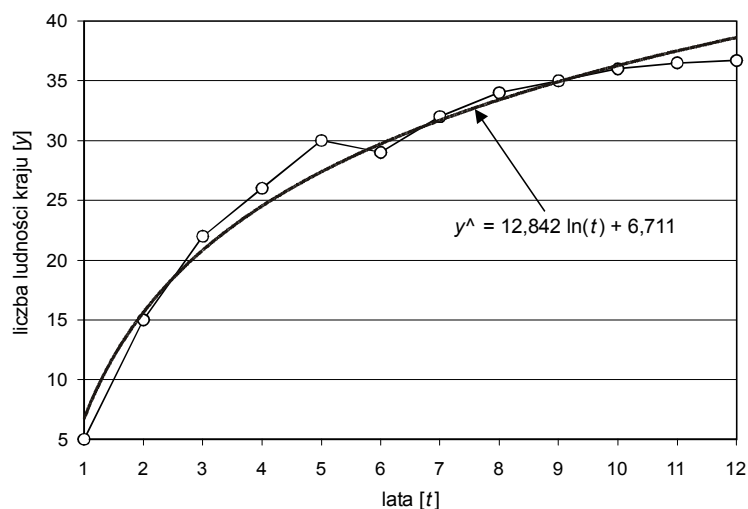
Dla danych umownych zmiennej Y (liczba ludności) w kolejnych latach, o trendzie logarytmicznym, sporządzamy wykres oraz określamy parametry równania trendu logarytmicznego. Wcześniej dokonamy transformacji liniowej w celu zastosowania KMNK.

Tabela 1.5.24

Obliczenia pomocnicze do oszacowania parametrów modelu zmiennej „Y”

t	y	t'	$(y - y^-)$	$(t' - t'^-)$	$(t' - t'^-)^2$	$(y - y^-)(t' - t'^-)$
1	5	0,0	-23,1	-1,7	2,8	38,5
2	15	0,7	-13,1	-1,0	0,9	12,7
3	22	1,1	-6,1	-0,6	0,3	3,5
4	26	1,4	-2,1	-0,3	0,1	0,6
5	30	1,6	1,9	-0,1	0,0	-0,1
6	29	1,8	0,9	0,1	0,0	0,1
7	32	1,9	3,9	0,3	0,1	1,1
8	34	2,1	5,9	0,4	0,2	2,4
9	35	2,2	6,9	0,5	0,3	3,7
10	36	2,3	7,9	0,6	0,4	5,0
11	36,5	2,4	8,4	0,7	0,5	6,2
12	36,7	2,5	8,6	0,8	0,7	7,0
Średnia	28,1	1,7		Σ	6,3	80,7

Źródło: Opracowanie własne w Excelu, gdzie $t'^- \rightarrow \bar{t}'$.



Ź r ó d ł o: Opracowanie własne w Excelu, gdzie $y^- \rightarrow \hat{y}$.

Wykres 1.5.9. Trend logarytmiczny zmiennej Y

Według wykresu $a_0 = 6,711$, $a_1 = 12,842$ obliczmy parametry a_0 oraz a_1 na podstawie wzorów KMNK:

$$a_1' = \frac{\sum_{t=1}^n (y - \bar{y}) \cdot (t' - \bar{t}')}{\sum_{t=1}^n (t' - \bar{t}')^2}, \quad a_0 = \bar{y} - a_1' \cdot \bar{t}'$$

$$a_1 = \frac{80,7}{6,3} = 12,810, \quad a_0 = 28,1 - 12,810 \cdot 1,7 = 6,323.$$

Występuje więc przybliżenie wyników wartości parametrów otrzymanych analitycznie i komputerowo w formie modelu trendu logarytmicznego podanego na wykresie 1.5.9.

1.5.6.6. Trend hiperboliczny

Trend ten ma dwie postacie:

$$\hat{Y} = a_0 + \frac{a_1}{t}, \quad (1.5.9)$$

$$\hat{Y} = \frac{a_0 \cdot t}{t + a_1}. \quad (1.5.10)$$

Przykład

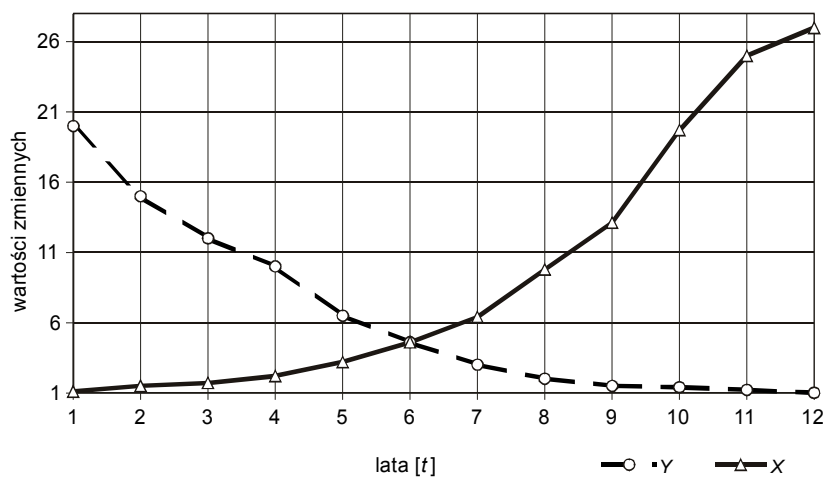
Dla podanych szeregów czasowych zmiennych Y oraz X wykonamy wykresy punktowe w Excelu.

Tabela 1.5.25

*Dane umowne
do trendu hiperbolicznego*

t	y	x
1	20,0	1,1
2	15,0	1,5
3	12,0	1,7
4	10,0	2,2
5	6,5	3,2
6	4,6	4,6
7	3,0	6,4
8	2,0	9,8
9	1,5	13,1
10	1,4	19,7
11	1,2	25,0
12	1,0	27,0

Źródło: Opracowanie własne w Excelu.



Źródło: Opracowanie własne.

Wykres 1.5.10. Trend hiperboliczny zmiennych Y i X

Program Excel nie rysuje oraz nie podaje funkcji trendu hiperbolicznego. Przyjmijmy podstawienie do pierwszej postaci funkcji trendu $t' = 1/t$, jest to następująca postać liniowa:

$$\hat{Y} = a_0 + a_1 t'$$

Obliczamy parametry według KMNK, posługując się wzorami:

$$a_1' = \frac{\sum_{i=1}^n (y - \bar{y}) \cdot (t' - \bar{t}')}{\sum_{i=1}^n (t' - \bar{t}')^2}, \quad a_0 = \bar{y} - a_1' \cdot \bar{t}'.$$

Przykład

Obliczamy parametry a_0 oraz a_1 dla funkcji trendu hiperbolicznego pierwszej postaci.

Tabela 1.5.26

Obliczenia pomocnicze do trendu hiperbolicznego zmiennej „Y”

t	y	t'	y - ȳ	t' - t̄'	(y - ȳ)(t' - t̄')	(t' - t̄') ²
1	20	1	13,48	0,741	9,997	0,550
2	15	0,5	8,48	0,241	2,048	0,058
3	12	0,333	5,48	0,075	0,410	0,006
4	10	0,25	3,48	-0,009	-0,030	0,000
5	6,5	0,2	-0,02	-0,059	0,001	0,003
6	4,6	0,167	-1,92	-0,092	0,176	0,008
7	3	0,143	-3,52	-0,116	0,407	0,013
8	2	0,125	-4,52	-0,134	0,603	0,018
9	1,5	0,111	-5,02	-0,147	0,740	0,022
10	1,4	0,1	-5,12	-0,159	0,812	0,025
11	1,2	0,091	-5,32	-0,168	0,892	0,028
12	1	0,083	-5,52	-0,175268	0,967	0,031
Średnia	6,52	0,259		Σ	17,022	0,762

Źródło: Opracowanie własne w Excelu, gdzie: $t^{-1} \rightarrow \bar{t}'$.

$$a_1 = \frac{17,022}{0,762} = 22,338, \quad a_0 = 6,52 - 22,324 \cdot 0,259 = 0,734,$$

$$\hat{Y} = 0,734 + 22,338/t, \quad a_1 > 0.$$

P r z y k ł a d

Wykonujemy transformację drugiej postaci funkcji trendu hiperbolicznego:

$$\hat{Y} = \frac{a_0 \cdot t}{t + a_1}.$$

Następnie wymagane jest odwrócenie tej funkcji i zastosowanie odpowiedniego podstawienia w celu przeprowadzenia transformacji liniowej:

$$\frac{1}{\hat{Y}} = \frac{t + a_1}{a_0 \cdot t} = \frac{1}{a_0} + \frac{a_1}{a_0} \cdot \frac{1}{t} = a'_0 + a'_1 \cdot t', \quad \hat{Y}' = a'_0 + a'_1 t'.$$

Parametry tej funkcji możemy obliczyć z kolejnych wzorów, a następnie określić a_0 oraz a_1 , wracając do podstawień:

$$a'_1 = \frac{\sum_{t'=1}^n (y' - \bar{y}') \cdot (t' - \bar{t}')}{\sum_{t'=1}^n (t' - \bar{t}')^2}, \quad a_0 = \bar{y}' - a'_1 \cdot \bar{t}'.$$

1.5.6.7. Trend logistyczny

Trend logistyczny ma postać:

$$\hat{Y} = \frac{a_0}{1 + a_1 \cdot e^{-t}}, \quad (1.5.11)$$

przy czym:

$$a_0 > 0, \quad a_1 > 1$$

P r z y k ł a d

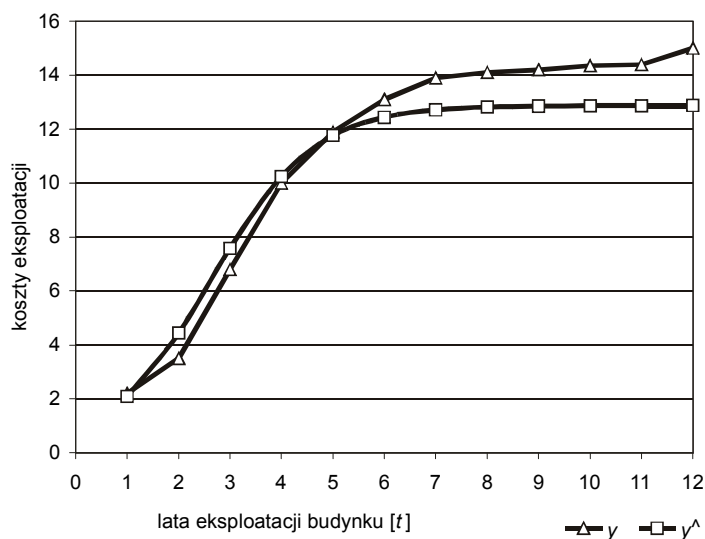
Dla zmiennej Y kosztów eksploatacji budynku (dane umowne) o trendzie logistycznym sporządzamy wykres w Excelu.

Tabela 1.5.27

Obliczenia pomocnicze do modelu trendu logistycznego zmiennej „Y”

Lp.	A	B	C	D	E	F	G	H
9	t	y	y' = 1/y	t' = e ^{-t}	y' - y ⁻	t' - t ⁻	(y' - y ⁻)(t' - t ⁻)	(t' - t ⁻) ²
10	1	2,2	0,455	0,36788	0,324	0,319	0,103	0,102
11	2	3,5	0,286	0,13534	0,155	0,087	0,013	0,008
12	3	6,8	0,147	0,04979	0,016	0,001	0,000	0,000
13	4	10	0,1	0,01832	-0,031	-0,030	0,001	0,001
14	5	11,9	0,084	0,00674	-0,047	-0,042	0,002	0,002
15	6	13,1	0,076	0,00248	-0,054	-0,046	0,002	0,002
16	7	13,9	0,072	0,00091	-0,059	-0,048	0,003	0,002
17	8	14,1	0,071	0,00034	-0,056	-0,048	0,003	0,002
18	9	14,2	0,070	0,00012	-0,061	-0,048	0,003	0,002
19	10	14,35	0,070	0,00005	-0,061	-0,048	0,003	0,002
20	11	14,39	0,069	0,00002	-0,061	-0,048	0,003	0,002
21	12	15	0,067	0,00001	-0,064	-0,048	0,003	0,002
22	Średnia		0,131	0,04850		Σ	0,140	0,128

Źródło: Opracowanie własne w Excelu z zaznaczeniem kolumn i wierszy arkusza kalkulacyjnego; dane umowne, gdzie: $y^- \rightarrow \bar{y}$; $t^- \rightarrow \bar{t}$.



Źródło: Opracowanie własne w Excelu, gdzie: $y^ \rightarrow \hat{y}$.

Wykres 1.5.11. Wartości empiryczne y i modelowe \hat{y} trendu logistycznego

Program Excel nie rysuje linii trendu logistycznego i dlatego do porównania wprowadzono na wykresie wyrównane wartości funkcji modelowej. W transformacji liniowej funkcji logistycznej odwracamy tę funkcję i dokonujemy podstawień:

$$\frac{1}{\hat{Y}} = \frac{1 + a_1 \cdot e^{-t}}{a_0}, \quad \hat{Y}' = \frac{1}{\hat{Y}}, \quad a'_0 = \frac{1}{a_0}, \quad a'_1 = \frac{a_1}{a_0}, \quad t' = e^{-t},$$

$$\hat{Y}' = a'_0 + a'_1 \cdot t'.$$

Obliczamy parametry, korzystając z wyników pośrednich podanych w tabeli 1.5.27.

$$a'_1 = \frac{\sum_{t'=1}^n (y' - \bar{y}') \cdot (t' - \bar{t}')}{\sum_{t'=1}^n (t' - \bar{t}')^2}, \quad a'_0 = \bar{y}' - a'_1 \cdot \bar{t}',$$

$$a'_1 = \frac{0,140}{0,128} = 1,09375 \approx 1,094, \quad a'_0 = 0,131 - 1,09375 \cdot 0,0485 = 0,07795 \approx 0,078,$$

$$\hat{Y}' = 0,078 + 1,094 \cdot t'.$$

Obliczamy parametry a_0 oraz a_1 trendu logistycznego:

$$a_0 = \frac{1}{a'_0} = \frac{1}{0,078} = 12,82, \quad a_1 = a'_1 \cdot a_0 = 1,094 \cdot 12,82 = 14,025,$$

$$\hat{Y} = \frac{12,82}{1 + 14,025 \cdot e^{-t}}.$$

P r z y k ł a d

Porównamy dane rzeczywiste z obserwacji zmiennej Y oraz obliczone na podstawie powyższego modelu reszty e (określane przez niektórych autorów też jako u).

T a b e l a 1.5.28

Reszty obserwacji modelu logistycznego zmiennej „Y”

t	y	y^{\wedge}	e
1	2,2	2,088	0,112
2	3,5	4,439	-0,939
3	6,8	7,577	-0,777
4	10	10,240	-0,240
5	11,9	11,761	0,139
6	13,1	12,440	0,660
7	13,9	12,711	1,189
8	14,1	12,813	1,287
9	14,2	12,851	1,349
10	14,35	12,865	1,485
11	14,39	12,870	1,520
12	15	12,872	2,128

Ź r ó d ł o: Opracowanie własne w Excelu, gdzie: $y^{\wedge} \rightarrow \hat{y}$.

Tabela 1.5.28 jest pomocniczą do wykonania wykresu 1.5.11.

Analogicznie do modeli trendów możemy budować modele regresji z jedną zmienną objaśnianą Y , przy czym zastępujemy wtedy czas t zmienną objaśniającą X .

1.5.6.8. Pytania

1. Wymień oraz wyjaśnij wzory trendów: liniowego, potęgowego, wykładniczego, logarytmicznego, hiperbolicznego, logistycznego.
2. Omów klasyczną metodę najmniejszych kwadratów (KMNK) do obliczenia parametrów a_0 oraz a_1 .
3. Podaj procedurę budowy trendu pełzającego.
4. Wykonaj transformację liniową trendów potęgowego oraz wykładniczego.
5. Podaj podstawienia do transformacji liniowej trendów logarytmicznego, hiperbolicznego, logistycznego.

1.5.6.9. Zadania

1. Oblicz analitycznie parametry a_0 oraz a_1 funkcji trendu liniowego zmiennej Y (powierzchnia zasiewów zboża w tys. ha), stosując KMNK na podstawie danych w tabeli 1.5.29.

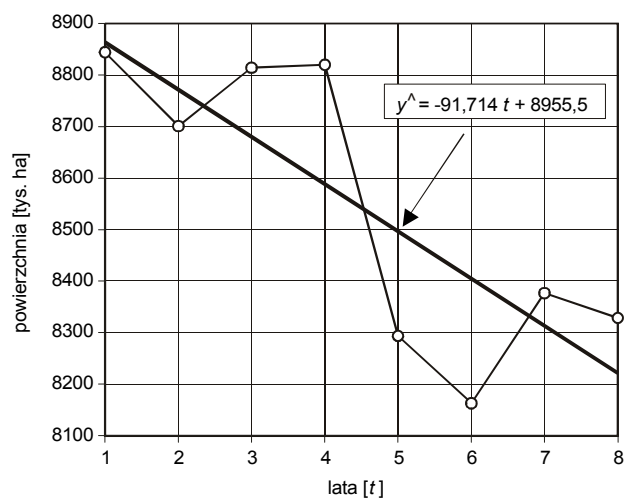
T a b e l a 1.5.29

Szereg czasowy zmiennej „Y”

Rok	Zasiewy [tys. ha]
1998	8 844
1999	8 701
2000	8 814
2001	8 820
2002	8 294
2003	8 163
2004	8 377
2005	8 329

Ź r ó d ł o: Opracowanie własne na podstawie „Roczników Statystycznych” z lat 2000–2006.

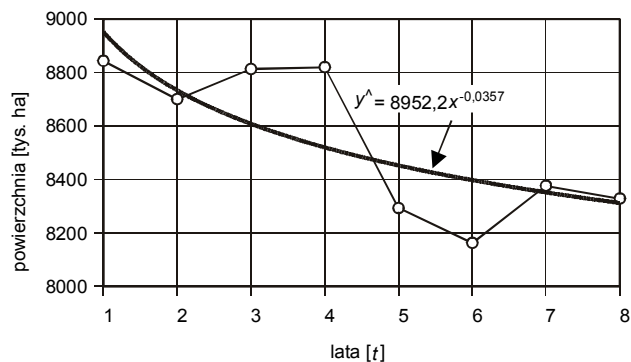
Jako odpowiedź podano wykres uzyskany z Excela.



Źródło: Opracowanie własne, gdzie: $y^ \rightarrow \hat{y}$.

Wykres 1.5.12. Trend liniowy zmiennej Y „Zasiewy”

2. Mając dane zadania pierwszego, oblicz parametry trendu potęgowego. Podpowiedzią jest linia trendu potęgowanego podana na wykresie 1.5.13.



Źródło: Opracowanie własne w Excelu, gdzie: $y^ \rightarrow \hat{y}$.

Wykres 1.5.13. Wykres trendu potęgowego zmiennej Y „Zasiewy”

3. Przyjmując, że zużycie energii elektrycznej podane w gigawatogodzinach (GWh) w gospodarstwach domowych (bez gospodarstw indywidualnych) ma rozkład wykładniczy, określ parametry a_0 oraz a_1 tego trendu, korzystając z danych tab. 1.5.30.

T a b e l a 1.5.30

Dane do trendu wykładniczego zmiennej „Y”

Rok	Energia [GWh]
1998	20 288
1999	20 800
2000	21 037
2001	21 375
2002	21 659
2003	22 052
2004	22 804
2005	26 565

Ź r ó d ł o: Opracowanie własne na podstawie „Roczników Statystycznych” z lat 2000–2006.

4. Korzystając z funkcji Excela w tworzeniu wykresów, a następnie doboru odpowiedniego trendu, wybierz taki, którego współczynnik determinacji jest najmniejszy dla danych statystycznych podanych w tab. 1.5.31.

T a b e l a 1.5.31

Dane do określenia modelu trendu zmiennej Y
[tys. szt. ciągników]

Rok	Ciągniki
1998	1 310,50
1999	1 305,50
2000	1 306,70
2001	1 308,50
2002	1 364,60
2003	1 371,30
2004	1 364,40
2005	1 437,20

Ź r ó d ł o: Opracowanie własne na podstawie „Roczników Statystycznych” z lat 2000–2006.

Sprawdzić słuszność podpowiedzi: spośród trendów liniowego, logarytmicznego, potęgowego, wykładniczego największy współczynnik determinacji $R^2 = 0,792$ ma trend wykładniczy.

1.6. Wybór kombinacji zmiennych objaśniających

1.6.1. Uwagi wstępne

Na początkowym etapie budowy liniowego modelu ekonometrycznego spotykamy się z wieloma przyczynami warunkującymi badane zjawisko. Rodzi się pytanie, jakimi zmiennymi ilościowymi je opisać oraz które z nich możemy uznać jako potencjalne zmienne objaśniające. Kryterium wstępnym klasyfikacji jest współczynnik zmienności lub rozstępu. Po tej selekcji otrzymujemy zestaw zmiennych, które mogą tworzyć wiele wzajemnych kombinacji. W kolejnym kroku poszukujemy metody, która wskazałaby nam kombinację zmiennych istotnych statystycznie, dającą dobre dopasowanie modelu do danych rzeczywistych. Aby ułatwić to zadanie, w niniejszym rozdziale zaprezentowano coś w rodzaju mikromonografii metod doboru zmiennych objaśniających. Przegląd obejmuje opublikowane metody w pozycjach książkowych. Oprócz najstarszych metod regresji krokowej dominują propozycje bazujące na analizie korelacji lub badaniu pojemności nośników informacji, jakimi są poszczególne kombinacje z zestawu potencjalnych zmiennych objaśniających. Rozpoczniemy od liniowego modelu ekonometrycznego, który określamy jako:

$$Y = f(X_1, X_2, \dots, X_n) + u; \quad Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n + u,$$

gdzie:

Y – zmienna objaśniana (endogeniczna),

X_1, X_2, \dots, X_n – zmienne objaśniające (egzogeniczne),

a_1, a_2, \dots, a_n – oszacowane na podstawie szeregów obserwacji parametry strukturalne modelu,

u oznaczamy także e – składnik losowy uwzględniający różnice, jakie wystąpią między wartościami z obserwacji a teoretycznymi wyznaczonymi przez model.

Procedura formułowania modelu ekonometrycznego sprowadza się zasadniczo do:

- określenia zmiennych objaśnianych i objaśniających,
- wyboru postaci analitycznej modelu i wskazania źródeł danych statystycznych,

- estymacji, czyli oszacowania parametrów na podstawie danych statystycznych, korzystając z określonej metody,
 - weryfikacji modelu.
- Zmienne objaśniające (egzogeniczne):
- muszą mieć wysoką zmienność, bo w przeciwnym wypadku są to zmienne *quasi*-stałe,
 - muszą być silnie skorelowane ze zmienną objaśnianą,
 - nie mogą być skorelowane ze sobą.

1.6.2. Metoda wskaźników wektora zmiennych objaśniających

Metody doboru zmiennych do modelu ekonometrycznego przedstawiają przede wszystkim relacje między zmiennymi w przeszłości. W modelu do celów prognostycznych na zmienną objaśnianą mają wpływ w przyszłości także inne zmienne objaśniające, np. mające inną siłę lub kierunek oddziaływania. M. Cieślak⁴⁴ proponuje do oceny znaczenia zbioru zmiennych stosowanie:

- metody wskaźników wektora zmiennych objaśniających,
- wyboru zmiennych na podstawie błędu prognoz *ex ante*.

W pierwszym sposobie wektor y przedstawia zbiór zaobserwowanych w próbie wartości zmiennej objaśnianej (prognozowanej), a macierz X zbiór zmiennych objaśniających X_1, X_2, \dots, X_m .

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$$

W tej metodzie s pierwszych obserwacji jest mniejsze od n . Obliczamy współczynniki korelacji R_0^1 między zmienną y a każdymi ze zmiennych X_1, X_2, \dots, X_m oraz między potencjalnymi zmiennymi objaśniającymi R^1 dla s pierwszych obserwacji. Bazując na współczynnikach korelacji, obliczamy indywidualne pojemności nośników informacji h_{Li}^1 dla poszczególnych zmiennych X , biorąc pod uwagę L kombinacji, których jest $2^m - 1$, przy czym $i, j = 1, \dots, m$, np. dla trzech zmiennych $2^3 - 1 = 7$. Metoda ta wykazuje analogię do metody *a priori* Hellwiga (doboru zmiennych objaśniających do modelu ekonometrycznego).

⁴⁴ M. C i e ś l a k, *Prognozowanie gospodarcze...*, op.cit.

$$\mathbf{R}_0^1 = \begin{bmatrix} r_1^1 \\ r_2^1 \\ \dots \\ r_n^1 \end{bmatrix} \quad \mathbf{R}^1 = \begin{bmatrix} 1 & r_{12}^1 & \dots & r_{1m}^1 \\ r_{21}^1 & 1 & \dots & r_{2m}^1 \\ \dots & \dots & \dots & \dots \\ r_{m1}^1 & r_{m2}^1 & \dots & 1 \end{bmatrix} \quad h_{Li}^1 = \frac{(r_i^1)^2}{1 + \sum_{j \neq i} |r_{ij}^1|} \quad (1.6.1)$$

gdzie:

r_{ij} – współczynniki korelacji między zmiennymi dla s pierwszych obserwacji

Następnie obliczamy pojemności integralne dla kombinacji:

$$H_L^1 = \sum_{i=1}^{m_L} h_{Li}^1. \quad (1.6.2)$$

Podjęciem odmiennym w stosunku do znanej metody Hellwiga jest wprowadzenie „ruchomej próby”, czyli w kolejnym etapie dodanie następnej obserwacji ($s + 1$), a zrezygnowanie z pierwszej. Dla „nowej” drugiej próby obliczamy R_0^2 oraz macierz \mathbf{R}^2 i w kolejności H_L^2 dla wszystkich kombinacji.

Postępując tak dalej, otrzymujemy $(n - s + 1)$ pojemności integralnych dla $2^m - 1$ kombinacji, które możemy przedstawić w postaci macierzy:

$$\mathbf{H} = \begin{bmatrix} H_1^1 & H_2^1 & \dots & H_{2^m-1}^1 \\ H_1^2 & H_2^2 & \dots & H_{2^m-1}^2 \\ \dots & \dots & \dots & \dots \\ H_1^{n-s+1} & H_2^{n-s+1} & \dots & H_{2^m-1}^{n-s+1} \end{bmatrix}.$$

W tej macierzy poszczególne kolumny stanowią szeregi czasowe kombinacji zmiennych objaśniających. Wprowadzamy kryterium wyboru $G_L = H_L^0 W_L$, przy czym:

$$W_L = 1 + (n - s + 1)\beta_L, \quad H_L = \alpha_L + \beta_L t, \quad t = 1, \dots, n - s + 1.$$

gdzie:

β_L – współczynnik trendu pojemności integralnych poszczególnych kombinacji zmiennych X ,

H_L^0 – pojemność integralna L -tej kombinacji określona na podstawie n obserwacji.

Przyjmuje się jako optymalną kombinację spełniającą warunek: $G = \max_L G_L$

1.6.3. Bazowanie na mierze błędów *ex ante*

Kolejnym sposobem proponowanym w publikacji M. Cieślak⁴⁵ jest wybór zmiennych objaśniających bazujący na błędzie *ex ante*, który dla modelu

⁴⁵ Ibidem.

trendu liniowego obliczany jest według wzoru (1.6.3). Dla modeli nieliniowych wykonujemy wpraw transformację na model liniowy.

$$v_T = \left[\frac{(T - \bar{t})^2}{\sum_{t=1}^n (t - \bar{t})^2} + \frac{1}{n} + 1 \right]^{0,5} \cdot S_e, \quad (1.6.3)$$

gdzie:

T – okres prognozowany,

\bar{t} – średnia wartość zmiennej czasowej o n obserwacjach,

S_e – odchylenie standardowe reszt obliczane według wzoru (1.6.4):

$$S_e = \left[\frac{1}{n - m - 1} \sum_{t=1}^n (y_t - \hat{y})^2 \right]^{0,5}, \quad (1.6.4)$$

gdzie:

\hat{y} – wartość teoretyczna (prognozowana) na podstawie określonego modelu kombinacji zmiennych objaśniających.

Stosując miarę v_T postępowanie w zakresie doboru zmiennych objaśniających jest następujące:

– określenie błędu *ex ante* v_T^0 dla modelu obejmującego wszystkie zmienne,

– wyznaczenie błędów prognoz *ex ante* dla poszczególnych L kombinacji zmiennych,

– wybór kombinacji L zmiennych, dla której spełniany jest warunek:

$$v_T^0 - v_i^2 = \min [v_T^0 - v_i^L],$$

$$L = 1, \dots, 2^m - 1.$$

Jak widzimy, jest to metoda uciążliwa i wymaga parametrów modeli oraz wartości zmiennych dla okresu prognozy T .

Wobec pracochłonności metody *ex ante* przyjmujemy, że prognoza według danej kombinacji zmiennych X jest dopuszczalna dla $t > n$, jeśli alternatywnie spełnione są warunki:

$$v_T \leq \gamma_T^*; \quad \eta_T \leq \eta_T^*.$$

Wartości z gwiazdką zostają określone przez odbiorcę prognozy. O wyborze kombinacji może zdecydować minimalna wartość średniego bezwzględnego błędu prognozy *ex ante*, czyli v_T , lub średni procentowy względny błąd prognozy *ex ante* określony jako:

$$\eta_T = \frac{V_T}{y_T} \cdot 100 \quad (1.6.5)$$

gdzie:

y_T – wartość zmiennej objaśnianej ustalona po zakończeniu okresu prognozowania T .

1.6.4. Analiza korelacji

Z tematyką doboru zmiennych objaśniających do modelu ekonometrycznego spośród potencjalnych zmiennych $X_1, X_2, X_j, \dots, X_m$ spotykamy się również w publikacji K. Jajugi⁴⁶. Pierwszym krokiem jest zbadanie wystarczającego zróżnicowania danej zmiennej X_j . Miarą zmienności jest współczynnik zmienności $V_j = S_x / \bar{x}$, czyli stosunek odchylenia standardowego danej zmiennej do średniego poziomu wartości tej zmiennej. Wymagany jest jednak warunek, aby $V_j > V^*$, przy czym krytyczna wartość V^* zwykle przyjmowana jest jako 0,1.

W opracowaniu pod redakcją K. Jajugi proponuje się dwa sposoby wyboru zmiennych objaśniających do modelu ekonometrycznego: analizę korelacji oraz metodę pojemności nośników informacji.

Pierwsza stanowi analogię do metody wcześniej opisanej, proponowanej do celów prognostycznych przez M. Cieślak. Występuje tu macierz korelacji \mathbf{R} między zmiennymi objaśniającymi X_j oraz X_l (określona na bazie wszystkich n obserwacji próby) oraz wektor korelacji \mathbf{R}_0 .

$$R = \sqrt{1 - \frac{\det R^*}{\det R}} \quad (1.6.6)$$

$$\mathbf{R}^* = \begin{bmatrix} 1 & R_0^T \\ R_0 & R \end{bmatrix}$$

Budowana jest macierz rozszerzona \mathbf{R}^* oraz określony jest według wzoru (1.6.6) współczynnik korelacji wielokrotnej R , bazujący na wyznacznikach $\det R^*$ oraz $\det R$.

$$\mathbf{R}_0 = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_n \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix}$$

⁴⁶ K. Jajuga (red.), *Ekonometria. Metody...*, op.cit, rozdz. 4.2.

Współczynnik R kształtuje się w przedziale $[0;1]$. Gdy $R = 0$, to nie ma zależności liniowej, a gdy $R = 1$, to istnieje funkcyjna zależność liniowa między zmienną objaśnianą a określoną kombinacją zmiennych objaśniających. Wartość R stanowi podstawę doboru zmiennych w liniowym modelu ekonometrycznym. Spośród kombinacji zmiennych objaśniających do modelu przyjmuje się tę, dla której współczynnik korelacji wielorakiej ma wartość największą.

1.6.5. Metoda pojemności nośników informacji

Kolejny sposób na wybór potencjalnych zmiennych zaprezentowany w wymienionej publikacji pod redakcją K. Jajugi⁴⁷ to metoda pojemności nośników informacji, o której wspomniano wcześniej (patrz podrozdz. 1.6.2) przy omawianiu propozycji zawartych w publikacji M. Cieślak. Zmienne powinny być możliwie silnie skorelowane ze zmienną objaśnianą, natomiast słabo skorelowane ze sobą. Przyjmuje się, że nośnikiem informacji są wszystkie z m potencjalnych zmiennych objaśniających, dla których istnieje ($k = 2^m - 1$) kombinacji tych zmiennych o liczbie m_k .

Dla całej próby n obserwacji obliczamy indywidualne pojemności nośników informacji.

W kolejnym wzorze: j to numer zmiennej w rozpatrywanej kombinacji ($j = 1, 2, \dots, m_k$); r_j – wartość z wektora korelacji R_0 ; r_{lj} – wartość z macierzy korelacji \mathbf{R} . W kolejności oblicza się również pojemności integralne kombinacji nośników informacji dla $k = 1, 2, \dots, 2^m - 1$.

$$h_{Li}^1 = \frac{r_j^2}{\sum_{l=1}^{m_k} |r_{lj} I|}, \quad H_k = \sum_{j=1}^{m_k} h_{kj}.$$

H_k stanowi kryterium wyboru odpowiedniej kombinacji zmiennych objaśniających, przy czym wybieramy zestaw zmiennych, dla których H_k jest maksymalne.

1.6.6. Metoda grafowa i inne

Szersze zajęcie się tematem wyboru zmiennych do modelu ekonometrycznego spotykamy w pracy pod redakcją J. Dziechciarza⁴⁸. Oprócz klasycznego współczynnika zmienności V_j , o którym była mowa wcześniej, autor

⁴⁷ Ibidem.

⁴⁸ J. Dziechciarz (red.), *Ekonometria. Metody, przykłady...*, op.cit., rozdz. 2.

proponuje uznać daną zmienną jako potencjalną zmienną objaśniającą, i zabazować na współczynniku W_j opartym na rozstępie.

$$W_j = 2 \cdot \frac{\max_i x_{ij} - \min_i x_{ij}}{\max_i x_{ij} + \min_i x_{ij}}. \quad (1.6.7)$$

$$r_{jl} = \frac{\text{cov}_{jl}}{s_j s_l} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (x_{il} - \bar{x}_l)^2}}.$$

Jednak uważa to za działanie wstępne, wymagające dalszej analizy korelacji zmiennych. Posługuje się w tym względzie wymienionym współczynnikiem korelacji liniowej r_{jl} Pearsona do utworzenia macierzy korelacji \mathbf{R} składającej się z współczynników r_{jl} .

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1l} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2l} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{j1} & r_{j2} & \dots & r_{jl} & \dots & r_{jm} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{ml} & \dots & 1 \end{bmatrix}$$

Potrzeba nam jeszcze określić skorelowanie par dwóch zmiennych – określonej objaśniającej X_i oraz objaśnianej Y . Możemy arbitralnie obrać krytyczną wartość współczynnika korelacji lub uzależnić ją od liczności próby obserwacji. W tym drugim przypadku posługujemy się hipotezą $H_0: r_{xy} = 0$, czyli zakładamy, że zmienne X_i oraz Y nie są istotnie skorelowane i obliczamy statystykę:

$$t = \frac{|r_{xy}|}{\sqrt{1 - r_{xy}^2}} \cdot \sqrt{n - 2}. \quad (1.6.8)$$

Obliczoną wartość statystyki t porównujemy z t_α odczytanym z rozkładu t -Studenta. Na przykład dla poziomu istotności $\alpha = 0,05$ i ilości stopni swobody $iss = n - 2 = 28 - 2 = 26$ $t_{0,05} = 2,0555$. Gdy $t \geq t_\alpha$, to hipotezę zerową odrzucamy. Zdaniem wymienionego autora, gdy próba $n \geq 120$, należy zastosować rozkład normalny. Statystykę t stosujemy także dla par zmiennych objaśniających X_j oraz X_l . W ten sposób możemy wyeliminować nieistotnie statystycznie

skorelowane najpierw zmienne X_j ze zmienną objaśnianą Y , a następnie zbędne pary zmiennych objaśniających. Pozwoli to na wstępne sformułowanie liniowego modelu ekonometrycznego.

Alternatywnym podejściem jest oparcie się na współczynniku korelacji wielorakiej, który poznaliśmy (podrozdz. 1.6.4) przy propozycjach K. Jajugi. W tym względzie określamy dla zmiennych macierz korelacji \mathbf{R} , wektor korelacji \mathbf{R}_0 . Następnie budujemy macierz rozszerzoną \mathbf{R}^* oraz obliczamy współczynnik korelacji wielorakiej, bazując na wyznacznikach $\det \mathbf{R}^*$ oraz $\det \mathbf{R}$.

$$\mathbf{R}_0 = \begin{bmatrix} r_1 \\ r_1 \\ \dots \\ r_1 \end{bmatrix} \quad R = \sqrt{1 - \frac{\det R^*}{\det R}}.$$

Wartość tego współczynnika stanowi podstawę doboru jednakowo licznych kombinacji zmiennych objaśniających. Gdy mamy dwa zestawy potencjalnych zmiennych objaśniających, to przyjmujemy ten, którego współczynnik korelacji wielorakiej jest większy. Przykładowo: dla pięciu zmiennych X_j otrzymujemy 10 kombinacji dwuelementowych oraz 10 kombinacji trzelementowych. Dla poszczególnych kombinacji dwuelementowych obliczamy macierz \mathbf{R} , wektor \mathbf{R}_0 , macierz rozszerzoną \mathbf{R}^* , wyznaczniki macierzy \mathbf{R}^* oraz \mathbf{R} . Następnie określamy współczynniki korelacji wielorakiej R między Y a poszczególnymi kombinacjami dwuelementowymi np. dla $\{X_1, X_2\}$. Przyglądamy się wynikom i wybieramy dwuelementową kombinację zmiennych do modelu ekonometrycznego, której współczynnik R jest największy. Pozostaje nam jednak postąpić podobnie z kombinacjami trójelementowymi i dopiero, opierając się na wynikach opracowań, zdecydować o składzie zmiennych w modelu.

Nieco inną propozycję doboru zmiennych objaśniających do modelu ekonometrycznego znajdujemy w pracy J. Dziechciarza⁴⁹. Polega ona na określeniu krytycznej wartości współczynnika korelacji według wzoru:

$$r^* = \sqrt{\frac{t_\alpha^2}{t_\alpha^2 + n - 2}}, \quad (1.6.9)$$

gdzie:

t_α – statystyka wyznaczona na podstawie rozkładu t -Studenta.

⁴⁹ Ibidem, rozdz. 2.3.

Przykładowo dla $\alpha = 0,05$ i $iss = n - 2 = 28 - 2 = 26$, $t_\alpha = 2,0555$. Postępowanie w wyborze zmiennych objaśniających jest następujące:

a) sprawdzamy $|r_j| \leq r^*$ i eliminujemy zmienne X_j nieistotnie skorelowane z Y ,

b) z pozostałych zmiennych wybieramy X_h według kryterium:

$$|r_h| = \max_j |r_j|,$$

c) badamy relacje powyższej zmiennej z pozostałymi zmiennymi i eliminujemy te, w których $|r_{hi}| > r^*$, gdyż są one zbyt silnie skorelowane ze zmienną X_h .

Takie postępowanie prowadzimy do momentu wyczerpania się potencjalnych zmiennych objaśniających. Model ekonometryczny tworzy więc zmienna X_h oraz niewyeliminowane zmienne na etapie c).

W pracy J. Dziechciarza⁵⁰ spotykamy również omówienie doboru zmiennych objaśniających do modelu metodą wskaźników pojemności informacji. Oprócz nieco innych oznaczeń do obliczenia takich wielkości jak indywidualna oraz integralna pojemność nośników informacji, zaprezentowana procedura jest identyczna z już wcześniej omówioną.

Na uwagę zasługuje jednak **metoda grafowa** doboru zmiennych objaśniających. Rozpoczynamy od określenia wektora \mathbf{R}_0 oraz macierzy \mathbf{R} . Następnie wyznaczamy r^* , korzystając z wzoru (1.6.9) lub reguły minimaksowej $j \neq i$:

$$r^* = \min_i \max_j |r_{ij} - I|.$$

Jeżeli w metodzie grafowej $|r_{ij}| \leq r^*$, to w miejscu elementów macierzy \mathbf{R} spełniających ten warunek wpisujemy zera, tworząc w ten sposób macierz \mathbf{R}' . Na kolejnym etapie rysujemy graf, którego wierzchołkami są zmienne a wiązałkami niezerowe współczynniki korelacji. Do modelu wchodzi zmienne odosobnione (niezwiązane z innymi elementami grafu) oraz te zmienne, które mają najwięcej wzajemnych wiązań. Jeśli ilość wiązań jest taka sama, to wybieramy zmienną o największym r_j . Tę metodę zaprezentujemy na przykładzie⁵¹ pięciu zmiennych objaśniających: X_1, X_2, X_3, X_4, X_5 zbioru $n = 30$ obserwacji skorelowanych wzajemnie oraz skorelowanych ze zmienną objaśniającą Y .

Niech macierz symetryczna \mathbf{R} i wektor \mathbf{R}_0 o i wierszach i j kolumnach będą następujące:

⁵⁰ Ibidem.

⁵¹ Ibidem (analogia do przykładu 2.5).

$$\mathbf{R} = \begin{bmatrix} 1 & 0,8 & 0,3 & 0,6 & 0,1 \\ 0,8 & 1 & 0,4 & 0,9 & 0,6 \\ 0,3 & 0,4 & 1 & 0,2 & 0,5 \\ 0,6 & 0,9 & 0,2 & 1 & 0,7 \\ 0,1 & 0,6 & 0,5 & 0,7 & 1 \end{bmatrix} \quad \mathbf{R}_0 = \begin{bmatrix} 0,6 \\ 0,8 \\ 0,5 \\ 0,9 \\ 0,8 \end{bmatrix}$$

przy $\alpha = 0,05$, $iss = n - 2 = 30 - 2 = 28$, $t_\alpha = 2,0484$, to wartość r^* obliczamy ze wzoru 1.6.9:

$$r^* = \sqrt{\frac{2,0484^2}{2,0484^2 + 30 - 2}} = 0,36,$$

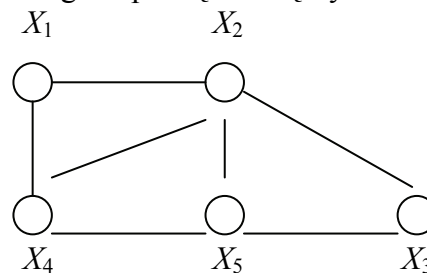
w sytuacji macierzy niesymetrycznej, gdy $j \neq i$, korzystamy z reguły minimaksowej:

$$r^* = \min_i \max_j |r_{ij}| = 0,5.$$

Dla $r^* = 0,36$ i $|r_{ij}| \leq r^*$, macierz \mathbf{R} przybiera postać:

$$\mathbf{R}' = \begin{bmatrix} 1 & 0,8 & 0 & 0,6 & 0 \\ 0,8 & 1 & 0,4 & 0,9 & 0,6 \\ 0 & 0,4 & 1 & 0 & 0,5 \\ 0,6 & 0,9 & 0 & 1 & 0,7 \\ 0 & 0,6 & 0,5 & 0,7 & 1 \end{bmatrix}$$

Macierz \mathbf{R}' zmiennych X_1 – X_5 o współczynnikach korelacji zerowych dla $|r_{ij}| \leq 0,36$ umożliwia zbudowanie grafu powiązań między zmiennymi.

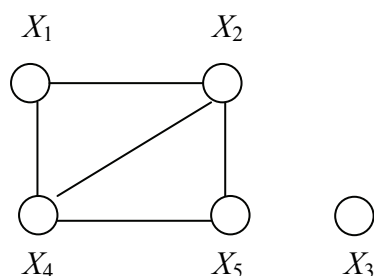


Zmienna X_2 ma cztery wiązadła, a zmienne X_4 i X_5 po trzy wiązadła, przy czym najsilniej skorelowana według \mathbf{R}_0 ze zmienną objaśnianą jest X_4 ($r_4 = 0,9$). Zatem model jest następujący:

$$\hat{Y} = a_0 + a_1 X_2 + a_2 X_4.$$

Teraz przykładowo zbudujmy graf, opierając się na $r^* = 0,5$ uzyskane z reguły minimaksowej. Kryterium $|r_{ij}| \leq 0,5$ przyjęto jako zero.

$$\mathbf{R}' = \begin{bmatrix} 1 & 0,8 & 0 & 0,6 & 0 \\ 0,8 & 1 & 0 & 0,9 & 0,6 \\ 0 & 0 & 1 & 0 & 0 \\ 0,6 & 0,9 & 0 & 1 & 0,7 \\ 0 & 0,6 & 0 & 0,7 & 1 \end{bmatrix}$$



Zmienna X_3 jest odosobniona, a więc wchodzi do modelu, natomiast zmienne X_2, X_4 mają po trzy wiązadła. Według \mathbf{R}_0 najsilniej skorelowana jest zmienna X_4 , tak więc liniowy model ekonometryczny ma postać:

$$\hat{Y} = a_0 + a_1 X_3 + a_2 X_4$$

1.6.7. Testy porównywania modeli

W literaturze spotykamy testy umożliwiające porównywanie modeli służących do prognozowania. Jak się wydaje, mogą one posłużyć również do wyboru wariantów według tego samego rodzaju liniowego modelu ekonometrycznego, utworzonych z różnych kombinacji zmiennych objaśniających. Gdy zastosujemy transformację liniową, to po estymacji parametrów modelu możemy zastosować testy bazujące na sumie kwadratów reszt również do modeli nieliniowych. Zamieszczone testy Akaike'a (AK) oraz Schwarza (SCH) dla n obserwacji stosowane są, gdy występują podobne wartości miary oceny dopasowania modeli do danych empirycznych oraz jeśli spełnione są testy istotności i stabilności parametrów⁵².

$$AK = n \ln(\Sigma z_n^2) + 2k. \quad (1.6.10)$$

⁵² M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria...*, op.cit.

$$SCH = n \ln(\sum z_n^2) + k \ln(n), \quad (1.6.11)$$

gdzie:

k – liczba parametrów modelu o n obserwacjach,

$\sum z_n^2$ – suma kwadratów reszt jako różnicy wartości z obserwacji y_n i wartości teoretycznej \hat{y}_n (obliczonej na podstawie równania modelu danej dla kombinacji zmiennych objaśniających).

$$z_n = y_n - \hat{y}_n.$$

1.6.8. Metoda regresji krokowej oraz minimalizacji resztowej sumy kwadratów

Tematyka doboru zmiennych objaśniających znajduje się również w *Ekonometrii* G.S. Maddali⁵³. Spośród k potencjalnych zmiennych objaśniających wybieramy określony podzbiór według określonego kryterium. Autor nadmienia, że w latach sześćdziesiątych powstawały metody regresji krokowej, które polegały na eliminowaniu zmiennych o statystykach t niższych od z góry ustalonych we wstępnym równaniu regresji o wszystkich k zmiennych objaśniających. Podejściem alternatywnym było rozpoczęcie procedury z pojedynczą zmienną o najwyższym r_j względem zmiennej objaśnianej Y . W kolejnych krokach wybierano do modelu zmienną o najwyższym współczynniku korelacji cząstkowej z Y . Metody te nazwane przez Maddalę „mechanicznymi” nie są obecnie stosowane, lecz wykorzystuje się do modeli regresji jedno z kryteriów statystycznych, które minimalizuje wyrażenia podane w tab. 1.6.1.

T a b e l a 1.6.1

Kryteria statystyczne minimalizujące dane wyrażenie

Kryterium	Wyrażenie
R^2 Thiela	$RSS_j / (n - k_j)$
S_p Hockinga	$RSS_j / [(n - k_j)(n - k_j - 1)]$
C_p Mallowsa	$RSS_j + 2 k_j \hat{\sigma}_m^2$
PC Amemiya	$RSS_j (n + k_j) / (n - k_j)$
AIC Akaike'a	$RSS_j \exp[2(k_j + 1)/n]$

Ź r ó d ł o: Opracowanie własne na podstawie: G.S. M a d d a l a, *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2006, tab. 12.3.

⁵³ G.S. M a d d a l a, *Ekonometria...*, op.cit, rozdz. 12.7.

Wyrażenia w tab. 1.6.1. opierają się na resztowej sumie kwadratów RSS modelu j (utworzonego np. na bazie określonej kombinacji zmiennych objaśniających). Określenie $\hat{\sigma}_m^2$ oznacza wariancję resztową $RSS_j/(n-k)$ w modelu o wszystkich k zmiennych objaśniających. Na podstawie wartości z minimalizowanych wyrażen określonego kryterium podanego w tabeli wybieramy model (kombinację zmiennych objaśniających) najlepiej dopasowany do wartości Y z n obserwacji, które stanowiły podstawę estymacji parametrów niezbędnej do obliczenia wartości modelowej zmiennej objaśnianej.

1.6.9. Metoda Hellwiga

Metoda pojemności nośników informacji, tzw. metoda Hellwiga, została wcześniej zasygnalizowana w podrozdziale 1.6.5. Postępowanie jest analogiczne do wcześniej omówionego, lecz przy realizacji przykładu zostało ponownie przytoczone.

a) Określenie macierzy współczynników korelacji pomiędzy zmiennymi objaśniającymi oraz wektora współczynników korelacji zmiennych objaśniających ze zmienną objaśnianą.

b) Ustalenie ilości kombinacji zmiennych objaśniających według wzoru ($L = 2^n - 1$) gdzie: n – ilość zmiennych objaśniających.

c) Obliczenie dla każdej kombinacji l i określonej zmiennej j indywidualnego wskaźnika pojemności informacji z zależności:

$$h_{ij} = \frac{r_j^2}{1 + \sum_1^{m_l} |r_{ij}|},$$

gdzie:

$l = 1, \dots, L$ – numery kombinacji,

$j = 1, \dots, m_j, m_l$ – liczba zmiennych w kombinacji l ,

r_{ij} – współczynnik korelacji między zmiennymi o indeksach i oraz j .

d) Obliczenie integralnego wskaźnika pojemności poszczególnych kombinacji:

$$H_l = \sum_{j=1}^{m_l} h_{ij}.$$

e) Wybranie kombinacji zmiennych, w których H_l jest największe:
 $H^* = \max \{H_l\}$

Przykład

Dla zmiennych objaśniających X_1, X_2, X_3, X_4 dany jest poniższy wektor i macierz korelacji o wartościach umownych. Ze względu na pracochłonność pełnej procedury obliczeń, w tym przykładzie określono tylko ilość kombinacji oraz, przykładowo, integralny wskaźnik pojemności informacji kombinacji piątej:

$$[\mathbf{r}_j] = [0,45 \quad -0,70 \quad 0,16 \quad 0,6]$$

$$[\mathbf{r}_{ij}] = \begin{bmatrix} 1 & -0,6 & 0,15 & 0,44 \\ -0,6 & 1 & -0,11 & -0,5 \\ 0,15 & -0,11 & 1 & -0,05 \\ 0,44 & -0,5 & -0,05 & 1 \end{bmatrix}$$

Liczba kombinacji czterech zmiennych objaśniających: $L = 2^n - 1 = 2^4 - 1 = 15$.

Tabela 1.6.2

Kombinacje zmiennych objaśniających

1: $\{x_1\}$	9: $\{x_2, x_4\}$
2: $\{x_2\}$	10: $\{x_3, x_4\}$
3: $\{x_3\}$	11: $\{x_1, x_2, x_3\}$
4: $\{x_4\}$	12: $\{x_1, x_2, x_4\}$
5: $\{x_1, x_2\}$	13: $\{x_1, x_3, x_4\}$
6: $\{x_1, x_3\}$	14: $\{x_2, x_3, x_4\}$
7: $\{x_1, x_4\}$	15: $\{x_1, x_2, x_3, x_4\}$
8: $\{x_2, x_3\}$	

Źródło: Opracowanie własne.

Indywidualne wskaźniki pojemności informacji przykładowo dla kombinacji piątej h_{51} i h_{52} :

$$h_{51} = \frac{(0,45)^2}{1 + 0,6} = 0,127,$$

$$h_{52} = \frac{(-0,7)^2}{1 + 0,6} = 0,306.$$

Integralny wskaźnik kombinacji piątej: $H_5 = 0,127 + 0,306 = 0,433$

Gdybyśmy rozpatrzyli wszystkie kombinacje, to wtedy możemy określić maksymalną wartość integralnej pojemności wskaźnika informacji H^* , której zestaw zmiennych stanowi najlepsze rozwiązanie modelu ekonometrycznego.

Mając określone zmienne objaśniające, zapisujemy postać analityczną modelu regresji liniowej:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n + \varepsilon$$

i przystępujemy do estymacji parametrów a_i modelu ekonometrycznego:

$$\hat{Y} = aX + a_0.$$

Znając powyższą procedurę metody pojemności informacji, możemy spróbować samodzielnie wybrać najlepszy zestaw spośród trzech zmiennych objaśniających X_1, X_2, X_3 , mając dany kolejny wektor i macierz korelacji:

$$[r_j] = [0,5 \quad -0,8 \quad 0,8] \quad [r_{ij}] = \begin{bmatrix} 1 & -0,6 & 0,5 \\ -0,6 & 1 & -0,5 \\ 0,5 & -0,5 & 1 \end{bmatrix}.$$

1.6.10. Określenie zmiennych objaśniających z zastosowaniem Excela

W konstruowaniu jednorównaniowych modeli zależności istotną rolę odgrywa właściwy spośród zasugerowanych dobor potencjalnych zmiennych objaśniających. W tym celu, jak już wspomniano, posługujemy się współczynnikiem zmienności, a następnie możemy zastosować np. analizę korelacji. Z tego względu zajmiemy się w kolejnym przykładzie zmierzaniem poziomu zróżnicowania obserwacji oraz określeniem współczynnika korelacji wielorakiej. Współczynnik ten jest miarą zależności między zmienną objaśnianą a liniową kombinacją zmiennych objaśniających.

We wstępnym formułowaniu zależności zmiennej Y od zmiennych X_m występuje przeważnie nadmierna ilość zmiennych objaśniających słabo związanych ze zmienną objaśnianą. Zmienne, których współczynnik zmienności $V_j < V^*$, uznaje się za mało zróżnicowane, eliminuje się ze zbioru potencjalnych zmiennych objaśniających.

Współczynnik zmienności jest ilorazem odchylenia standardowego do średniej wartości obserwacji danej zmiennej X_j (wzór 1.3.15). Aby przypomnieć, przytaczamy ten wzór i wzory pomocnicze:

$$V_x = \frac{S_x}{\bar{x}}, \quad S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}, \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

Obliczenie współczynników zmienności z zastosowaniem arkusza kalkulacyjnego Excel przedstawiono na przykładzie opisu wartości produkcji stolarki okiennej w mln zł w firmie *F*, przyjmując cztery zmienne objaśniające: X_1 – zatrudnienie w osobach, X_2 – wartość maszyn i urządzeń w tys. zł, X_3 – czas przestoju maszyn w dniach, X_4 – nakłady inwestycyjne. Dobierzmy zmienne, przyjmując wartość krytyczną $V^* = 12\%$.

Przykład

Obliczenie współczynników zmienności.

Tabela 1.6.3

Dane modelowe zmiennych „ X_1 ”, „ X_2 ”, „ X_3 ”

<i>i</i>	<i>Y</i>	X_1	X_2	X_3	X_4
1	12	8	7	12	10
2	10	6	8	14	12
3	18	11	10	16	11
4	16	10	12	18	14
5	12	8	8	18	10
6	14	10	8	18	12
7	18	11	12	20	13
8	20	12	16	24	12
9	20	12	16	26	12
10	22	14	18	26	10
S	162	102	115	192	116
Średnia	16,2	10,2	11,5	19,2	11,6
	\bar{Y}	\bar{X}_1	\bar{X}_2	\bar{X}_3	\bar{X}_4

Źródło: Opracowanie własne w Excelu, gdzie np. $y^- \rightarrow \bar{y}$.

Tabela 1.6.4

Obliczenia pomocnicze z zastosowaniem wyrażeń Excela

B	C	D	E	F	G	H	I	J
<i>i</i>	$(x_1 - x_1)$	$(x_1 - x_1)^2$	$(x_2 - x_2)$	$(x_2 - x_2)^2$	$(x_3 - x_3)$	$(x_3 - x_3)^2$	$(x_4 - x_4)$	$(x_4 - x_4)^2$
1	-2,2	4,84	-4,5	20,25	-7,2	51,84	-1,6	2,56
2	-4,2	17,64	-3,5	12,25	-5,2	27,04	0,4	0,16
3	0,8	0,64	-1,5	2,25	-3,2	10,24	-0,6	0,36
4	-0,2	0,04	0,5	0,25	-1,2	1,44	2,4	5,76
5	-2,2	4,84	-3,5	12,25	-1,2	1,44	-1,6	2,56
6	-0,2	0,04	-3,5	12,25	-1,2	1,44	0,4	0,16
7	0,8	0,64	0,5	0,25	0,8	0,64	1,4	1,96
8	1,8	3,24	4,5	20,25	4,8	23,04	0,4	0,16
9	1,8	3,24	4,5	20,25	6,8	46,24	0,4	0,16
10	3,8	14,44	6,5	42,25	6,8	46,24	-1,6	2,56
S		49,6		142,5		209,6		16,4

s_1	s_2	s_3	s_4
2,23	3,77	4,58	1,28
V1	V2	V3	V4
21,83	32,83	23,84	11,04
T	T	T	N

Przykłady wyrażeń:

=PIERWIASTEK(D28/10)

=(C31/D14)*100

=JEŻELI(C33>12;"T";"N")

Źródło: Opracowanie własne w Excelu, gdzie:

$x_3^- \rightarrow \bar{x}_1, x_2^- \rightarrow \bar{x}_2, x_3^- \rightarrow \bar{x}_3, x_4^- \rightarrow \bar{x}_4$.

Na końcu tabeli zastosowano funkcję badania logicznego – czy obliczony współczynnik zmienności jest większy od przyjętego 12%. Zatem potencjalną zmienną X_4 z punktu widzenia współczynnika zmienności możemy pominąć w konstruowaniu modelu ekonometrycznego. Zastosowanie współczynnika zmienności V_j jest tylko wstępną selekcją zmiennych objaśniających. Możemy przeprowadzić dalszy wybór zmiennych spośród X_1, X_2, X_3 , stosując **analizę korelacji**, która opiera się na współczynnikach korelacji między zmiennymi. W przykładzie zastosowano wzory z publikacji K. Jajugi⁵⁴. Wprowadźmy oznaczenia oraz przypomnijmy sobie wzory omówione wcześniej:

r_j – współczynnik korelacji między zmienną objaśnianą Y a zmienną objaśniającą X_j ,

r_{jl} – współczynnik korelacji między zmiennymi objaśniającymi X_j oraz X_l ,

\mathbf{R}_0 – wektor współczynników korelacji między zmienną objaśnianą Y a zmiennymi objaśniającymi X_j ,

\mathbf{R} – macierz korelacji między zmiennymi objaśniającymi,

\mathbf{R}^* – macierz rozszerzona,

R_k – współczynnik korelacji wielorakiej.

$$\mathbf{R}_0 = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_n \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix}, \quad \mathbf{R}^* = \begin{bmatrix} 1 & \mathbf{R}_0^T \\ \mathbf{R}_0 & \mathbf{R} \end{bmatrix}, \quad R_k = \sqrt{1 - \frac{\det \mathbf{R}^*}{\det \mathbf{R}}}$$

Przy dwóch zestawach potencjalnych zmiennych X_j wybiera się ten zestaw, którego współczynnik korelacji wielorakiej ma większą wartość. Bazując na tabeli 1.6.3 w zakresie zmiennych X_1, X_2, X_3 , obliczamy współczynniki korelacji otrzymania wektora \mathbf{R}_0 , macierzy \mathbf{R} , macierzy rozszerzonej \mathbf{R}^* i obliczenia współczynnika korelacji wielorakiej R_k :

Współczynniki korelacji między zmienną objaśnianą a kolejnymi zmiennymi objaśniającymi obliczamy według wzoru:

$$r_{xy} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1.6.12)$$

⁵⁴ K. Jajuga (red.), *Ekonometria. Metody...*, op.cit, rozdz. 4.2.

Przykład

Obliczenie współczynników korelacji między zmienną objaśnianą a zmiennymi objaśniającymi.

Tabela 1.6.5

Obliczenia pomocnicze w Excelu

i	Y	X_1	X_2	X_3	$y - \bar{y}$	$(y - \bar{y})^2$	$x_1 - \bar{x}_1$	$(x_1 - \bar{x}_1)^2$	$(x_1 - \bar{x}_1) \cdot (y - \bar{y})$
1	12	8	7	12	-4,2	17,64	-2,2	4,84	9,24
2	10	6	8	14	-6,2	38,44	-4,2	17,64	26,04
3	18	11	10	16	1,8	3,24	0,8	0,64	1,44
4	16	10	12	18	-0,2	0,04	-0,2	0,04	0,04
5	12	8	8	18	-4,2	17,64	-2,2	4,84	9,24
6	14	10	8	18	-2,2	4,84	-0,2	0,04	0,44
7	18	11	12	20	1,8	3,24	0,8	0,64	1,44
8	20	12	16	24	3,8	14,44	1,8	3,24	6,84
9	20	12	16	26	3,8	14,44	1,8	3,24	6,84
10	22	14	18	26	5,8	33,64	3,8	14,44	22,04
Σ	162	102	115	192		147,6		49,6	83,6
Średnia	16,2	10,2	11,5	19,2					

Tabela 1.6.5 cd.

$x_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$	$(x_2 - \bar{x}_2) \cdot (y - \bar{y})$	$x_3 - \bar{x}_3$	$(x_3 - \bar{x}_3)^2$	$(x_3 - \bar{x}_3) \cdot (y - \bar{y})$
-4,5	20,25	18,9	-7,2	51,84	30,24
-3,5	12,25	21,7	-5,2	27,04	32,24
-1,5	2,25	-2,7	-3,2	10,24	-5,76
0,5	0,25	-0,1	-1,2	1,44	0,24
-3,5	12,25	14,7	-1,2	1,44	5,04
-3,5	12,25	7,7	-1,2	1,44	2,64
0,5	0,25	0,9	0,8	0,64	1,44
4,5	20,25	17,1	4,8	23,04	18,24
4,5	20,25	17,1	6,8	46,24	25,84
6,5	42,25	37,7	6,8	46,24	39,44
	142,5	133		209,6	149,6

gdzie: $y^- \rightarrow \bar{y}$, $x_1^- \rightarrow \bar{x}_1$, $x_2^- \rightarrow \bar{x}_2$, $x_3^- \rightarrow \bar{x}_3$.

Źródło: Opracowanie własne w Excelu.

Wyniki obliczeń współczynników korelacji między zmienną Y a zmiennymi x_1 , x_2 , x_3 .

R_{yx_1}	R_{yx_2}	R_{yx_3}
0,98	0,92	0,85

W obliczeniu R_{yx} zastosowano wyrażenie Excela dla zmiennej X_1 :

$$=O48/PIERWIASTEK((N48*L48)),$$

przy czym adres O48 oznacza komórkę arkusza z wartością obliczoną dla:

$$\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y}),$$

a adresy N48 oraz L48 wartości komórek z obliczeniami:

$$\sum_{i=1}^n (y_i - \bar{y})^2, \quad \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2.$$

Współczynniki korelacji między zmiennymi objaśniającymi j -tą a l -tą:

$$r_{x_j x_l} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \cdot \sum_{i=1}^n (x_{il} - \bar{x}_l)^2}}.$$

Przykład

Obliczenie współczynników korelacji między zmiennymi objaśniającymi

Tabela 1.6.6

Obliczenia pomocnicze do wyznaczenia współczynników korelacji między „ X_1 ”, „ X_2 ”, „ X_3 ”

i	$x_1 - \bar{x}_1$	$(x_1 - \bar{x}_1)^2$	$x_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$	$x_3 - \bar{x}_3$	$(x_3 - \bar{x}_3)^2$	$(x_1 - \bar{x}_1) \cdot (x_2 - \bar{x}_2)$	$(x_1 - \bar{x}_1) \cdot (x_3 - \bar{x}_3)$	$(x_2 - \bar{x}_2) \cdot (x_3 - \bar{x}_3)$
1	-2,2	4,84	-4,5	20,25	-7,2	51,84	9,9	15,84	32,4
2	-4,2	17,64	-3,5	12,25	-5,2	27,04	14,7	21,84	18,2
3	0,8	0,64	-1,5	2,25	-3,2	10,24	-1,2	-2,56	4,8
4	-0,2	0,04	0,5	0,25	-1,2	1,44	-0,1	0,24	-0,6
5	-2,2	4,84	-3,5	12,25	-1,2	1,44	7,7	2,64	4,2
6	-0,2	0,04	-3,5	12,25	-1,2	1,44	0,7	0,24	4,2
7	0,8	0,64	0,5	0,25	0,8	0,64	0,4	0,64	0,4
8	1,8	3,24	4,5	20,25	4,8	23,04	8,1	8,64	21,6
9	1,8	3,24	4,5	20,25	6,8	46,24	8,1	12,24	30,6
10	3,8	14,44	6,5	42,25	6,8	46,24	24,7	25,84	44,2
Σ		49,6		142,5		209,6	73	85,6	160

$R_{x_1 x_2}$	$R_{x_1 x_3}$	$R_{x_2 x_3}$
0,87	0,84	0,93

Źródło: Opracowanie własne w Excelu, gdzie: $x_1^- \rightarrow \bar{x}_1$, $x_2^- \rightarrow \bar{x}_2$, $x_3^- \rightarrow \bar{x}_3$.

Na podstawie zamieszczonych obliczeń wektor, macierz korelacji oraz macierz rozszerzona są następujące:

$$\mathbf{R}_0 = \begin{bmatrix} 0,98 \\ 0,92 \\ 0,85 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0,87 & 0,84 \\ 0,87 & 1 & 0,93 \\ 0,84 & 0,93 & 1 \end{bmatrix}, \quad \mathbf{R}^* = \begin{bmatrix} 1 & 0,98 & 0,92 & 0,85 \\ 0,98 & 1 & 0,87 & 0,84 \\ 0,92 & 0,87 & 1 & 0,93 \\ 0,85 & 0,84 & 0,93 & 1 \end{bmatrix}$$

Przykład

Obliczenie wyznaczników

Wyznaczniki macierzy \mathbf{R}^* oraz macierzy \mathbf{R} z zastosowanie funkcji Excela.

WYZNACZNIK.MACIERZY(.....)

Tabela 1.6.7

Obliczenie w Excelu wyznaczników
macierzy \mathbf{R}^*

	C	D	E	F
55	1	0,98	0,92	0,85
56	0,98	1	0,87	0,84
57	0,92	0,87	1	0,93
58	0,85	0,84	0,93	1

0,00047 det \mathbf{R}^* =WYZNACZNIK.MACIERZY(C55:C58)

0,03189 det \mathbf{R} =WYZNACZNIK.MACIERZY(D56:F58)

Źródło: Opracowanie własne w Excelu.

Aby wyznaczyć det \mathbf{R} , można też do macierzy trzeciego stopnia zamiast funkcji WYZNACZNIK.MACIERZY(.....) zastosować regułę Sarrusa.

Tabela 1.6.8

Macierz \mathbf{R} i obliczenie wyznacznika „det \mathbf{R} ” w Excelu

1	0,87	0,84	1	0,87
0,87	1	0,93	0,87	1
0,84	0,93	1	0,84	0,93

Wyrażenie na obliczenie det (\mathbf{R}) operatorami Excela według reguły Sarrusa jest następujące:

$$=(1*1*1 + 0,87*0,93*0,84+0,84*0,87*0,93)-(0,84*1*0,84+1*0,93*0,93+0,87*0,87*1)$$

$$\det \mathbf{R} = 0,03189$$

Źródło: Opracowanie własne.

Natomiast obliczenie det \mathbf{R}^* dla macierzy czwartego stopnia wymaga wcześniejszego doprowadzenia macierzy do trzeciego stopnia, poprzez wyko-

rzystanie rozwinięcia Laplace'a względem wskazanego wiersza lub kolumny, lub zastosowanie algorytmu Chio⁵⁵.

Określenie współczynnika korelacji wielorakiej:

$$R_k = \sqrt{1 - \frac{\det R^*}{\det R}} = \sqrt{1 - \frac{0,00047}{0,03189}} = 0,9926.$$

Współczynnik ten wskazuje na duże natężenie zależności liniowej między zmienną objaśnianą Y a trzema zmiennymi objaśniającymi X_1, X_2, X_3 . Przyjęcie trzech zmiennych objaśniających X_1, X_2, X_3 jednocześnie w relacji ze zmienną objaśnianą Y nie daje odpowiedzi na pytanie, którą z nich wyeliminować. Dlatego współczynnik korelacji wielorakiej oblicza się kolejno według procedury podanej w przykładzie dla zestawów zmiennych objaśniających: $X_1X_2; X_1X_3; X_2X_3$. Jak już wspomniano, do modelu przyjmuje się ten zestaw, którego współczynnik korelacji wielorakiej ma wartość największą. Zostawmy na chwilę pracochłonną analizę korelacji, a spróbujmy teraz zastosować metodę pojemności nośników informacji (o której wspomniano wcześniej). Bazuje ona na wektorze i macierzy umownych trzech zmiennych objaśniających X_1, X_2, X_3 i procedura jest następująca:

a) Określenie możliwych kombinacji zmiennych objaśniających:

$$L = 2^n - 1 = 2^3 - 1 = 7$$

T a b e l a 1.6.9

Kombinacje zmiennych „ X_1 ”, „ X_2 ”, „ X_3 ”

Kombinacja k	Liczba zmiennych m_k
1: $\{X_1\}$	1
2: $\{X_2\}$	1
3: $\{X_3\}$	1
4: $\{X_1, X_2\}$	2
5: $\{X_1, X_3\}$	2
6: $\{X_2, X_3\}$	2
7: $\{X_1, X_2, X_3\}$	3

Ź r ó d ł o: Opracowanie własne.

b) Obliczenie indywidualnych pojemności nośników informacji h_{ij} każdej zmiennej w kombinacji k , a następnie integralnej pojemności nośników informacji H_k z wykorzystaniem wzorów:

⁵⁵ T. J u r l e w i c z, J. S k o c z y ł a s, *Algebra liniowa 1. Przykłady i zadania*, Oficyna Wydawnicza GiS, Wrocław 2003, s. 65, 72.

$$h_{ij} = \frac{r_j^2}{1 + \sum_{j=1}^{m_i} |r_{ij}|}, \quad H_l = \sum_{j=1}^{m_l} h_{ij}.$$

gdzie:

r_j – wartość z wektora \mathbf{R}_0

r_{ij} – wartość z macierzy \mathbf{R} ,

m_k – liczba zmiennych w k -tej kombinacji.

$$\mathbf{R}_0 = \begin{bmatrix} 0,98 \\ 0,92 \\ 0,85 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0,87 & 0,84 \\ 0,87 & 1 & 0,93 \\ 0,84 & 0,93 & 1 \end{bmatrix}.$$

Przykład

Wybór zestawu zmiennych objaśniających

Tabela 1.6.10

Obliczenie pojemności nośników informacji

k	Zmienne	r_j	r_j^2	$ r_{ij} $	$\sum r_{ij} $	h_{ki}	H_k
1	X_1	0,98	0,960	1	1	0,960	0,960
2	X_2	0,92	0,846	1	1	0,846	0,846
3	X_3	0,85	0,723	1	1	0,723	0,723
4	$X_1 X_2$	0,98	0,960	$ 1 + 0,87 $	1,87	0,513	
4	$X_1 X_2$	0,92	0,846	$ 1 + 0,87 $	1,87	0,452	0,965
5	$X_1 X_3$	0,98	0,960	$ 1 + 0,84 $	1,84	0,522	
5	$X_1 X_3$	0,85	0,723	$ 1 + 0,84 $	1,84	0,393	0,915
6	$X_2 X_3$	0,92	0,846	$ 1 + 0,93 $	1,93	0,438	
6	$X_2 X_3$	0,85	0,723	$ 1 + 0,93 $	1,93	0,375	0,813
7	$X_1 X_2 X_3$	0,98	0,960	$ 1 + 0,87 + 0,84 $	2,71	0,354	
7	$X_1 X_2 X_3$	0,92	0,846	$ 1 + 0,87 + 0,84 $	2,71	0,312	
7	$X_1 X_2 X_3$	0,85	0,723	$ 1 + 0,87 + 0,84 $	2,71	0,267	0,933

Źródło: Opracowanie własne.

Wskazanie zestawu zmiennych o największym H_k .

$H_{max} = \{0,960; 0,846; 0,723; 0,965; 0,915; 0,813; 0,933\} = 0,965$.

Model zatem jest następujący:

$$\hat{Y} = a_1 X_1 + a_2 X_2.$$

Powróćmy do wcześniejszego przykładu i sprawdźmy analizą korelacyjną zestawy zmiennych X_1X_2 , X_1X_3 , X_2X_3 . Interesuje nas z pewnością, czy kombinacja zmiennych X_1X_2 w tej metodzie również będzie najlepsza.

Zestaw X_1X_2 :

$$\mathbf{R}_0 = \begin{bmatrix} 0,98 \\ 0,92 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0,87 \\ 0,87 & 1 \end{bmatrix}, \quad \mathbf{R}^* = \begin{bmatrix} 1 & 0,98 & 0,92 \\ 0,98 & 1 & 0,87 \\ 0,92 & 0,87 & 1 \end{bmatrix}.$$

Zastosujmy funkcję Excela (=WYZNACZNIK.MACIERZY(...)). W nawiasie obszar adresowy macierzy \mathbf{R}^* lub \mathbf{R} .

$$\det \mathbf{R}^* = 0,0051, \det \mathbf{R} = 0,2431$$

$$R_k = \sqrt{1 - \frac{\det R^*}{\det R}} = \sqrt{1 - \frac{0,0051}{0,2431}} = 0,9895.$$

Zestaw X_1X_3 :

$$\mathbf{R}_0 = \begin{bmatrix} 0,98 \\ 0,85 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0,84 \\ 0,84 & 1 \end{bmatrix}, \quad R_k = \sqrt{1 - \frac{0,0109}{0,2944}} = 0,9813,$$

$$\det \mathbf{R}^* = 0,0109, \det \mathbf{R} = 0,2944.$$

Zestaw $X_2 X_3$:

$$\mathbf{R}_0 = \begin{bmatrix} 0,92 \\ 0,85 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0,93 \\ 0,93 & 1 \end{bmatrix},$$

$$\det \mathbf{R}^* = 0,0207, \det \mathbf{R} = 0,1351$$

$$R_k = \sqrt{1 - \frac{0,0207}{0,1351}} = 0,9202.$$

Z porównania par zmiennych najwyższy współczynnik korelacji wielorakiej ma zestaw X_1X_2 , dla którego $R_k = 0,9895$. Jest to zgodne z wynikiem metody pojemności nośników informacji, lecz nieco mniejsze od obliczonego na podstawie tabel 1.6.7 i 1.6.8 zestawu trzech zmiennych $X_1X_2 X_3$ o $R_k = 0,9926$.

Reasumując, przyjmijmy model liniowy o trzech zmiennych objaśniających:

$$\hat{Y} = a_1X_1 + a_2X_2 + a_3X_3.$$

1.6.11. Wybór zmiennych do modelu regresji z czterema zmiennymi objaśniającymi

Procedurę postępowania w zakresie wyboru zestawu zmiennych zamodelowano na danych umownych (tab. 1.6.11). Poszczególne etapy prac można wykonać w arkuszu kalkulacyjnym Excela.

Tabela 1.6.11

Dane umowne zmiennych „Y”, „X₁”, „X₂”, „X₃”, „X₄”

t	y	x ₁	x ₂	x ₃	x ₄
1	18,5	4,4	10,5	2,80	8
2	19,2	4,3	10,0	2,83	10
3	20,0	4,5	11,7	3,01	15
4	20,8	4,5	12,1	3,17	12
5	21,7	4,4	13,0	3,42	7
6	20,8	4,6	13,9	3,56	10
7	23,7	4,6	15,1	3,73	11
8	25,0	4,6	13,9	3,81	13
9	26,7	4,5	18,1	4,07	8
10	28,3	4,5	20,0	4,32	17

gdzie:

Y – produkcja przedsiębiorstwa w tys. szt.,

X₁ – zatrudnienie w tys. osób,

X₂ – moc zainstalowanych maszyn w KW,

X₃ – dostawa surowców w tys. ton,

X₄ – przestoje z powodu awarii w dniach.

Źródło: Opracowanie własne (kopia tabeli 1.12.14).

Założono, że model regresji liniowej wielowymiarowej tworzą zmienne w ogólnej postaci zlogarytmowanej:

$$\ln \hat{Y} = f(\ln X_1, \dots, \ln X_j).$$

Po uwzględnieniu parametrów model jest następujący:

$$\ln \hat{Y} = a_0 + a_1 \ln X_1 + \dots + a_n \ln X_n.$$

Procedura z podziałem wzorów tabel:

a) Określamy podstawowe miary statystyczne zmiennych Y, X₁, X₂, X₃, X₄ w postaci zlogarytmowanej.

Tabela 1.6.12

Rozplanowanie tabeli do miar zmiennych

ln ...	Średnia	Odchylenie standardowe	Maksimum	Minimum
ln Y				
ln X ₁				
ln X ₂				
ln X ₃				
ln X ₄				

Źródło: Opracowanie własne.

b) Wyznaczamy współczynniki kowariancji i korelacji pomiędzy logarytmem zmiennej objaśnianej a logarytmami zmiennych objaśniających.

Do obliczenia korelacji między zmienną objaśnianą Y a zmiennymi objaśniającymi X_j lub między zmiennymi objaśniającymi dla n = 10 obserwacji zastosujemy wzory, np.:

$$\text{cov}_{yx} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}), \quad r_{yx} = \frac{\text{cov}_{yx}}{\sqrt{\text{var}_y \cdot \text{var}_x}} = \frac{\text{cov}_{yx}}{S_y \cdot S_x},$$

przy czym wartości zmiennych zgodnie z wymogiem modelu podamy w postaci logarytmów.

Tabela 1.6.13

Rozplanowanie tabeli do kowariancji i wariancji

Lp.	Relacje	Kowariancja	Korelacja
1	ln Y, ln Y		
2	ln Y, ln X ₁		
3	ln Y, ln X ₂		
4	ln Y, ln X ₃		
5	ln Y, ln X ₄		
6	ln X ₁ , ln X ₁		
7	ln X ₁ , ln X ₂		
8	ln X ₁ , ln X ₃		
9	ln X ₁ , ln X ₄		
10	ln X ₂ , ln X ₂		
11	ln X ₂ , ln X ₃		
12	ln X ₂ , ln X ₄		
13	ln X ₃ , ln X ₃		
14	ln X ₃ , ln X ₄		
15	ln X ₄ , ln X ₄		

Źródło: Opracowanie własne.

c) Zastosowanie analizy korelacji w doborze zmiennych objaśniających, w której kroki postępowania są następujące:

– wyznaczamy wartość krytyczną współczynnika korelacji według wzoru (1.6.9):

$$r^* = \sqrt{t_\alpha^2 / (t_\alpha^2 + n - 2)}$$

($n = 10$, $iss = n - 2 = 8$, $P = \alpha = 0,01$ według tablicy wartości krytycznych rozkładu t -Studenta $t_\alpha = 3,3554$).

– eliminujemy zmienne objaśniające, których korelacja między $\ln Y$ a $\ln X_i$ jest statystycznie nieistotna, według zasady: $|r_{ij}| \leq r^*$,

– wybieramy spośród pozostałych korelacji $|r_{ij}|$ zmienną o korelacji maksymalnej między ($\ln Y$, $\ln X_i$) według kryterium: $|r_h| = \max |r_j|$ (*),

– eliminujemy zmienne o korelacji większej od krytycznej (względem zmiennej o korelacji maksymalnej, bazując na kryterium) $|r_{hi}| > r^*$,

– wybieramy następną zmienną objaśniającą przez powrót do etapu oznaczonego (*),

– gdy takiej nie ma, to formułujemy model.

d) Przeprowadzamy oszacowanie parametrów klasyczną metodą najmniejszych kwadratów dla wielu zmiennych, przykładowo stosując funkcję REGLINP Excela, przy czym wartości zmiennych występują w postaci logarytmów.

e) Określamy funkcję produkcji, jeżeli np. analiza współczynników korelacji wskaże zestaw X_1, X_2, X_3 .

$$\ln \hat{Y} = a_0 + a_1 \ln X_1 + a_2 \ln X_2 + a_3 X_3$$

po zamianie jej w postać funkcji wykładniczej:

$$\hat{Y} = e^{(a_0 + a_1 \ln X_1 + a_2 \ln X_2 + a_3 \ln X_3)}$$

P r z y k ł a d

Procedura dojścia do modelu funkcji produkcji bazująca na logarytmach zmiennych jest pracochłonna. Z tego względu zaprezentowano tabelę wynikową kowariancji i korelacji (tab. 1.6.14) oraz przeprowadzono analizę współczynników korelacji, korzystając z możliwości pisania wyrażeń w Excelu.

Tabela 1.6.14

Kowariancja i korelacja logarytmów zmiennych

Lp.	Relacje	Kowariancja	Korelacja
1	$\ln Y, \ln Y$	0,0185533	1
2	$\ln Y, \ln X_1$	0,0018058	0,6273325
3	$\ln Y, \ln X_2$	0,0274461	0,9586453
4	$\ln Y, \ln X_3$	0,0191587	0,9863685
5	$\ln Y, \ln X_4$	0,0311507	0,8414850
6	$\ln X_1, \ln X_1$	0,0004466	1
7	$\ln X_1, \ln X_2$	0,0025172	0,5666754
8	$\ln X_1, \ln X_3$	0,0019432	0,6448232
9	$\ln X_1, \ln X_4$	0,0027249	0,4744211
10	$\ln X_2, \ln X_2$	0,0441800	1
11	$\ln X_2, \ln X_3$	0,0290493	0,9691856
12	$\ln X_2, \ln X_4$	0,0448532	0,7851803
13	$\ln X_3, \ln X_3$	0,0203344	1
14	$\ln X_3, \ln X_4$	0,0308671	0,7851803
15	$\ln X_4, \ln X_4$	0,0738622	1

Źródło: Opracowanie własne.

Natomiast realizacja analizy współczynników korelacji według opisanej procedury była następująca:

– obliczenie wartości krytycznej współczynnika korelacji:

$$r^* = \sqrt{3,3554^2 / (3,3554^2 + 8 - 2)} = 0,76,$$

– wyeliminowanie zmiennej objaśniającej, dla której korelacja między $\ln Y$ a $\ln X_j$ jest statystycznie nieistotna według zasady:

$|r_{ij}| \leq r^*, |r_{ij} = 0,627|$, czyli jest mniejsza od r^* , dla zmiennej X_1 i ona podlega wyeliminowaniu,

– wybór spośród pozostałych zmiennej o korelacji maksymalnej między $(\ln Y, \ln X_j)$. Jest to zmienna X_3 o $r_3 \approx 0,986$.

– eliminacja zmiennych o korelacji względem X_3 większej od krytycznej. Są nimi zmienne X_4 o korelacji $r_3 \approx 0,785$ oraz X_2 o korelacji $r_2 \approx 0,969$,

– pozostała tylko zmienna X_3 i ona wchodzi do modelu regresji liniowej:

$$\ln \hat{y} = f(\ln X_3) = a_0 + a_1 \ln X_3.$$

Kończąc ten rozdział, trzeba zaznaczyć, że powstają coraz to nowsze statystyczne metody wyboru kombinacji zmiennych objaśniających. Zyskują one zaufanie ekonometryków. Mamy nadzieję, że zebrany materiał będzie pomocny w przyjęciu właściwej metody i przyczyni się do doskonalenia umiejętności budowy liniowych modeli ekonometrycznych, oczywiście po wcześniejszej analizie wyboru kombinacji zmiennych objaśniających.

1.7. Zastosowanie programów Excel oraz GRETL do weryfikacji modeli ekonometrycznych

1.7.1. Modele jednorównaniowe

Weryfikacja modeli ekonometrycznych obejmuje szereg pracochłonnych etapów. Zachodzi więc konieczność – zwłaszcza przy modelach o wielu zmiennych – wspomaganie komputerowego. Zaprezentowany materiał należy traktować jako rodzaj pomocy dla studiujących ekonometrię w zakresie poznania procedur zastosowania funkcjonalności arkusza kalkulacyjnego wraz z jego dodatkami oraz pakietu GRETL, który jest bezpłatnie dostępny w internecie.

Liniowy model ekonometryczny ma postać:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_j X_j + \dots + \alpha_m X_m + \varepsilon,$$

gdzie:

$\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m$ – parametry,

X_j – zmienna objaśniająca (niezależna), przy czym $j = 1, 2, \dots, m$,

Y – zmienna objaśniana (zależna),

ε – składnik losowy.

Po oszacowaniu, czyli estymacji parametrów, otrzymujemy wartości teoretyczne zmiennej objaśnianej i wówczas parametry α_j zastępujemy parametrami oszacowanymi:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + a_j X_j + \dots + a_m X_m.$$

Kolejne obserwacje zmiennej Y zapisujemy jako y_i ($i = 1, 2, \dots, n$), natomiast zmiennych X_j jako x_{ij} . Obserwacje zmiennej Y stanowią wektor kolumnowy, a wartości zmiennej X_j są macierzą o wymiarach ($n \times m$). Jedną z metod określenia wektora parametrów a przy pomocy klasycznej metody najmniejszych kwadratów jest posłużenie się równaniem:

$$a = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T y, \quad (1.7.1)$$

gdzie:

\mathbf{X}^T – transpozycja macierzy \mathbf{X} .

Podane w dalszej części tego tematu wzory bazują na opracowaniu K. Jajugi⁵⁶ oraz nawiązują do rozdziału 1.6. Dla przypomnienia: zależność między Y a X_j mierzona jest współczynnikiem korelacji wielorakiej, przyjmującym wartości z przedziału $[0,1]$. Współczynnik ten wyznaczamy na podstawie wyznacznika macierzy korelacji \mathbf{R} , obejmującej współczynniki korelacji między zmiennymi objaśniającymi X_j oraz macierzy rozszerzonej \mathbf{R}^* :

$$R_w = \sqrt{1 - \frac{\det R^*}{\det R}}, \quad \mathbf{R}^* = \begin{bmatrix} 1 & \mathbf{R}_0^T \\ \mathbf{R}_0 & \mathbf{R} \end{bmatrix}.$$

Wektor korelacji \mathbf{R}_0 obejmuje współczynniki między zmienną Y a zmiennymi X_j . Im wartość R_w jest bliższa jedności, tym natężenie zależności jest większe. Do obliczania R_w można skorzystać z funkcji Excela (WYZNACZNIK.MACIERZY(...)).

Modele jednorównaniowe obejmują:

- regresję jednej zmiennej z dopasowaniem funkcji liniowej lub nieliniowej (wykładniczej, potęgowej, logarytmicznej, logistycznej),
- trend liniowy lub nieliniowy (po wcześniejszej transformacji na postać liniową), gdy zmienną objaśniającą jest zmienna czasowa t ,
- regresję wielu zmiennych X_j , czyli tzw. regresję wieloraką.

Weryfikacja liniowych lub sprowadzalnych do liniowych modeli jednorównaniowych opiera się na resztach e_i , stanowiących różnicę między wartościami empirycznymi a modelowymi zmiennej objaśnianej, obliczoną dla poszczególnych obserwacji. Pozwala to, przy założeniu, że średnia reszt równa się zero, na określenie wariancji resztowej:

$$S_e^2 = \frac{\sum_{i=1}^n e_i^2}{n-k},$$

gdzie:

n – ilość obserwacji,

$k = m + 1$ ilość zmiennych, w tym tożsamościowo równa 1 przy parametrze a_0 .

Standardowy błąd oceny poszczególnych parametrów α_j określamy z zależności:

$$S(a_j) = \sqrt{V(a_j)}.$$

Wariancje parametrów $V(a_j)$ znajdują się na głównej przekątnej macierzy kowariancji wyznaczonej według:

⁵⁶ K. Jajuga (red.), *Ekonometria. Metody...*, op.cit., część I, rozdz. 4.

$$\mathbf{D}^2(\alpha) = S_e^2 (\mathbf{X}^T \mathbf{X})^{-1}. \quad (1.7.2)$$

Występuje tu zatem iloczyn wariancji resztowej i macierzy odwrotnej.

Istnieje szereg mierników dopasowania modelu do danych rzeczywistych (empirycznych) i należą do nich:

– współczynnik determinacji, stanowiący również kwadrat współczynnika korelacji wielorakiej, przyjmujący wartości z przedziału $[0, 1]$, który obliczamy ze wzoru:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

gdzie:

\bar{y} – średnia arytmetyczna zmiennej Y ,

R^2 – miara stopnia wyjaśnienia przez model zmiennej zależnej Y .

– alternatywnie wobec R^2 stosowany współczynnik zbieżności $\Phi^2 = 1 - R^2$, który wskazuje, jaką część zmiennej objaśnianej stanowi zmienność składnika losowego ε .

Współczynnik zbieżności oznaczany również jako Φ^2 można obliczyć bezpośrednio ze wzoru:

$$\Phi^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1.7.3)$$

– wymienione wcześniej odchylenie standardowe S_e , określające przeciętne różnice między y_i a \hat{y}_i .

– współczynnik wyrazistości wskazujący, jaką część wartości średniej zmiennej Y stanowi odchylenie standardowe reszt:

$$W = (S_e / \bar{y}) \cdot 100. \quad (1.7.4)$$

Wstępnie określony liniowy model jednorównaniowy może mieć część zmiennych objaśniających nieistotnych statystycznie w relacji do zmiennej Y . Z tego względu najpierw badamy istotność całego zestawu zmiennych X_j , posługując się testem parametrycznym F (wzór 1.1.6), którego hipoteza zerowa zakłada nieistotność wektora parametrów bez wyrazu wolnego:

$$F = \frac{(n-k)R^2}{m(1-R^2)},$$

gdzie:

$k = m + 1$,

m – ilość zmiennych objaśniających.

Jeśli $F > F^*$, to hipotezę zerową odrzucamy. Wartość F^* określamy na podstawie rozkładu F -Fishera dla poziomu istotności α i stopni swobody $n_1 = m$ oraz $n_2 = n - k = n - m - 1$.

Parametryczny test t -Studenta, odnoszący się do określonej zmiennej X_j , stosowany przeważnie dla $n < 30$ obserwacji przy poziomie istotności $\alpha = 0,05$ umożliwia badanie statystycznej istotności danej zmiennej objaśniającej na zmienną objaśnianą.

$$t_j = \frac{|a_j|}{S(a_j)}.$$

Hipoteza zerowa zakłada brak statystycznej istotności danej zmiennej objaśniającej na zmienną Y . Gdy $t_j > t^*$, to odrzucamy hipotezę zerową. Wartość krytyczną t^* oznaczaną również jako t_α wyznaczamy według tablicy rozkładu t -Studenta.

Dalsze badania składnika losowego (reszt) możemy prowadzić zarówno analitycznie, jak i poprzez analizę graficzną, określając:

- rozkład reszt (częstość występowania reszty danej wartości), co pozwala na sprawdzenie założenia o normalności rozkładu reszt, czyli średnia reszt równa się zero, a wynika to z KMNK,
- symetrię reszt, czyli równość częstości występowania reszt dodatnich i ujemnych (bez uwzględnienia reszt o wartości zero),
- losowość (nie może być zbyt długich lub krótkich serii reszt dodatnich lub ujemnych); założenie, że reszty jako obserwacje pojawiają się kolejno w momentach lub okresach, pozwala wyodrębnić kolejne grupy reszt o tendencji wzrostowej lub spadkowej,
- stałość/niestalałość wariancji resztowej w kolejnych przedziałach czasu; jeśli wariancja zwiększa się w czasie, to w celu „przebudowy” modelu należy stosować ważoną metodę najmniejszych kwadratów⁵⁷,
- brak zależności obserwacji od czasu; niespełnienie tego założenia może oznaczać wzajemny liniowy wpływ zmiennych objaśniających.

Jeśli mamy do czynienia z nieliniowym modelem ekonometrycznym z wieloma zmiennymi objaśniającymi, to, jak już wspomniano, przeprowa-

⁵⁷ Ibidem, rozdz. *Analiza reszt*.

dzamy transformację liniową, aby zastosować KMNK. W tym celu stosuje się podstawienia, odwrotności lub obustronne logarytmowanie danego równania nieliniowego. W modelach z tak oszacowanymi parametrami możemy stosować miary weryfikacji jak w modelach liniowych, przy czym w modelach nieliniowych stosujemy dodatkowo wskaźnik średniego poziomu reszt p :

$$p = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{\hat{y}_i}. \quad (1.7.5)$$

Wskaźnik p jest średnim poziomem wartości bezwzględnych reszt do wartości modelowej określonej po estymacji parametrów⁵⁸.

1.7.2. Zastosowanie funkcji Excela

Ogólnie dostępnym dla studiujących jak i praktyków narzędziem do komputerowego wspomaganie obliczeń z zakresu estymacji jak i weryfikacji jednorównaniowych modeli ekonometrycznych jest program Excel z dodatkami programowymi. Pokażmy zastosowanie jego funkcji na przykładzie fragmentu bazy danych dotyczącej stopy bezrobocia w 50 powiatach, pozyskanej z internetu na stronie [www \[kufel.torun.pl\]](http://www.kufel.torun.pl)⁵⁹. Zbiór „powiaty.gdt” obejmuje 380 rekordów (obserwacji) dla 42 zmiennych objaśniających i zmiennej objaśnianej „bezrobot”. Procedura wyłonienia podzbioru z zastosowaniem menu programu GRETL jest następująca:

- menu (*Próba/Zakres próby*), startowa obserwacja: 1, końcowa obserwacja: 50, *OK*,
- zaznaczenie w strukturze rekordu informacji kopiowanych przez wybór z wciśnięciem klawisza *Ctrl*,
- menu (*Dane/Pokaż wartości*) – pojawia się okno prezentacji danych,
- kopiowanie do bufora,
- wklejenie do arkusza kalkulacyjnego, zastąpienie kropek dziesiętnych przecinkami, wykonanie obramowań i sformatowanie komórek (ustawienie czcionki, umiejscowienia),
- menu Excela: (*Edycja/Kopiuj*),
- wstawienie do Worda (*Edycja/Wklej*) specjalnie jako tekst o formacie rtf.

Tabela 1.7.1 oprócz liczby porządkowej zawiera skróty 50 powiatów oraz wybrane informacje:

- bezrobot – stopa bezrobocia w powiatach,
- regon – liczba jednostek regon na 1 tys. mieszkańców,

⁵⁸ Ibidem, s. 67.

⁵⁹ T. K u f e l, *Ekonometria. Rozwiązywanie problemów...*, op.cit.

lud_prod – udział ludności w wieku produkcyjnym,
noclegi – liczba miejsc noclegowych na 1 tys. mieszkańców.
a ponadto kolumnę „bezrobot_t” z obliczonymi wartościami teoretycznymi
zmiennej objaśnianej „bezrobot_t” i kolumnę „reszta” stanowiąca różnicę war-
tości empirycznej i modelowej.

Do estymacji parametrów oraz krótkiej statystyki opisowej Excel ma
wbudowaną funkcję REGLINP, której przykładowa składnia wygląda tak:

=REGLINP(C3:C52;D3:F52;PRAWDA;PRAWDA)

gdzie:

C3:C52 – obszar zmiennej bezrobot,

D3:F52 – obszar zmiennych: regon, lud_prod, noclegi,

PRAWDA – odpowiednio w kolejności potwierdzenie występowania stałej, żądania staty-
styki opisowej.

T a b e l a 1.7.1

*Dane do oszacowania parametrów uproszczonego
modelu stopy bezrobocia w powiatach*

Lp.	obs	bezrobot	regon	lud_prod	noclegi	bezrobot_t	reszta
1	aleksand	0,23731146	88,77903924	0,61042606	3,88940192	0,177008855	0,060302605
2	augustow	0,1932204	72,64343743	0,57643522	6,21548128	0,1790033	0,0142171
3	bartoszy	0,31262608	54,58067822	0,6120593	6,64126293	0,278419895	0,034206185
4	belchato	0,20068024	78,64737232	0,63906123	2,99502763	0,226923911	-0,02624367
5	bedzinsk	0,17313645	97,30628446	0,64825293	1,87554627	0,198274027	-0,02513758
6	bialski	0,15619965	46,96730826	0,55321287	7,48601029	0,193017156	-0,03681751
7	Biała Po	0,17700602	83,87472787	0,6397655	0,78166515	0,177962109	-0,00095609
8	bialobrz	0,18572871	83,89281683	0,57444596	6,53555938	0,166993567	0,018735143
9	bialogar	0,31052263	85,1998515	0,61458978	5,69648971	0,225101535	0,085421095
10	bialosto	0,13364342	64,52498231	0,58589608	6,35716246	0,210257678	-0,07661426
11	Białysto	0,1159479	117,2178271	0,65325462	0,26840109	0,150446916	-0,03449902
12	bielski	0,10642451	93,37142468	0,61672618	2,25011053	0,150799068	-0,04437456
13	bielski	0,09712773	68,21023235	0,55115362	8,60218546	0,183633707	-0,08650598
14	bielsko	0,10666292	139,5591745	0,64588516	0,58683139	0,113854908	-0,00719199
15	bierunsk	0,10678335	72,31736085	0,62652837	1,75211136	0,186819713	-0,08003636
16	bieszcz	0,24676314	97,12826486	0,6095605	8,17615698	0,248928781	-0,00216564
17	bilgoraj	0,16447494	59,60548885	0,5809129	4,910425	0,17867209	-0,01419715
18	bochensk	0,11816367	68,91857311	0,58917038	3,04723886	0,145681592	-0,02751792
19	boleslaw	0,19472202	69,72202035	0,6255337	3,34075218	0,219538689	-0,02481667

cd. tab. 1.7.1

Lp.	obs	bezrobot	regon	lud_prod	noclegi	bezrobot_t	reszta
20	braniewski	0,29777688	64,96623819	0,60874442	7,91663863	0,283802414	0,013974466
21	brodnicki	0,22093662	65,73995956	0,59297202	4,84110723	0,192308658	0,028627962
22	brzeski	0,15175769	62,16931217	0,58637118	2,86409291	0,145472663	0,006285027
23	brzeski	0,23634522	89,93619865	0,61858214	3,16635772	0,17679099	0,05955423
24	brzezinski	0,19058114	87,57657278	0,6204583	4,10332869	0,201840138	-0,011259
25	brzozowski	0,25033179	49,42565558	0,57306302	3,59861486	0,151085051	0,099246739
26	buski	0,12205082	65,61599231	0,58515266	7,88272659	0,237452906	-0,11540209
27	bydgoski	0,16613259	83,56451339	0,62552228	3,65509823	0,207827136	-0,04169455
28	Bydgoszcz	0,11421966	124,3337543	0,64257658	0,33686881	0,122308407	-0,00808875
29	Bytom	0,18391673	78,9212288	0,64397239	0,30109574	0,182899337	0,001017393
30	bytowski	0,31625195	74,80928226	0,60377308	7,79512987	0,259251651	0,057000299
31	Chelm	0,17973059	84,77036521	0,64336933	0,64013238	0,180870911	-0,00114032
32	chelmiński	0,22705399	63,06622246	0,6110624	4,34274359	0,220341623	0,006712367
33	chelmiński	0,19664897	37,77596513	0,56698715	7,73838974	0,236034024	-0,03938505
34	chojnicki	0,20626996	90,10649924	0,62630981	3,7285448	0,202321484	0,003948476
35	chojnicki	0,25012773	74,66068374	0,60446054	3,40982296	0,174464524	0,075663206
36	Chorzów	0,18373607	83,68835418	0,62929007	0,43198937	0,151410822	0,032325248
37	choszczowski	0,28483623	67,9078436	0,61360933	7,93877389	0,289686646	-0,00485042
38	chrzanowski	0,1449684	79,43983145	0,62816024	1,75517412	0,180784517	-0,03581612
39	ciechanowski	0,1929599	77,26678291	0,607051	3,84699215	0,184624974	0,008334926
40	cieszyński	0,12057551	104,863689	0,62279219	2,1682075	0,145877732	-0,02530222
41	czarnkowsko-trzcianowski	0,22369401	72,7488593	0,61686377	3,50591303	0,20240002	0,02129399
42	Częstochowa	0,14663569	107,3946464	0,64624578	0,50329429	0,154431067	-0,00779538
43	częstochowski	0,14817528	64,41460971	0,6023592	3,81314755	0,19163701	-0,04346173
44	człuchowski	0,31031876	76,28742726	0,61792992	7,62171001	0,280840003	0,029478757
45	Dąbrowa	0,18608169	95,18910741	0,6784947	0,69213313	0,23520656	-0,04912487
46	dąbrowski	0,1968168	43,99595676	0,58537923	4,60175778	0,201245068	-0,00442827
47	debicki	0,18682969	54,57289786	0,59657772	2,61297614	0,169741581	0,017088109
48	drawski	0,29538973	91,06803368	0,61712385	6,45539109	0,237272872	0,058116858
49	działowski	0,28138653	56,85883481	0,60242963	4,69332765	0,218846243	0,062540287
50	dzierżonowski	0,22868148	93,47332205	0,61790746	1,99138004	0,147822057	0,080859423

Źródło: Opracowanie własne na podstawie pliku [powiaty.gdt]; T. K u f e l, *Ekonomia. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2007.

Zastosowanie REGLINP dało następujące wyniki:

Tabela 1.7.2

Parametry oraz statystyka opisowa modelu stopy bezrobocia w powiatach

a_3	a_2	a_1	a_0	
0,0196761	1,9008	-0,00129131	-0,945176	parametry
0,00373272	0,425549	0,000474643	0,251316	błąd standardowy oszacowania parametrów
0,450304	0,047332			R^2 ; S_e
12,5609	46			F ; iss
0,08442122	0,103055			ESS ; RSS

Źródło: Opracowanie własne w Excelu.

gdzie:

R^2 – współczynnik determinacji,

S_e – odchylenie standardowe reszt,

F – statystyka Fishera,

iss – ilość stopni swobody ($n - m - 1$),

ESS – wyjaśniona przez model suma kwadratów,

RSS – resztowa suma kwadratów.

Tak więc model liniowy jest następujący:

$$\hat{Y} = -0,945176 - 0,00129131 \cdot (\text{regon}) + 1,9008 \cdot (\text{lud_prod}) + 0,0196761 \cdot (\text{noclegi}).$$

Opcja Excela (*Narzędzia/Analiza danych/Lista narzędzi statystycznych*) umożliwia określenie między innymi:

- korelacji, kowariancji,
- statystyki opisowej, histogramu, regresji.

Korelacja

Tabela 1.7.3

Macierz symetryczna współczynników korelacji

Zmienne	bezrobot	regon	lud_prod	noclegi
bezrobot	1			
regon	-0,273055	1		
lud_prod	-0,042493	0,701364	1	
noclegi	0,458887	-0,52811	-0,69885	1

Źródło: Opracowanie własne w Excelu.

Kowariancja

Tabela 1.7.4

Macierz symetryczna współczynników kowariancji

Zmienne	bezrobot	regon	lud_prod	noclegi
bezrobot	0,003826			
regon	-0,338496	401,658397		
lud_prod	-0,000070	0,373030	0,000704	
noclegi	0,072086	-26,879798	-0,047101	6,449715

Źródło: Opracowanie własne w Excelu.

Statystyka opisowa

Przykład dotyczy zmiennej objaśnianej „bezrobot”. Wprowadzamy do okna dialogowego następujące informacje:

- zakres wejściowy Y: \$C\$2:\$C\$52; tytuły w pierwszym wierszu,
- statystyka podsumowująca; zapis w nowym arkuszu,
- poziom ufności – 95%,
- największa oraz najmniejsza wartość zmiennej.

Tabela 1.7.5

Statystyka opisowa zmiennej Y „Stopa bezrobocia w powiatach”

bezrobot	
Średnia arytmetyczna	0,195567
Błąd standardowy średniej	0,008748
Mediana	0,188705
Odchylenie standardowe	0,061855
Wariancja próbki	0,003826
Kurtoza	-0,654921
Skośność	0,386007
Zakres	0,219124
Suma wartości obserwacji	9,778363
Licznik (ilość obserwacji)	50
Największa wartość	0,316252
Najmniejsza wartość	0,097128
Poziom	0,017579

Źródło: Opracowanie własne w Excelu.

W tabeli 1.7.5 odchylenie standardowe jest pierwiastkiem z wariancji. Mediana jest wartością środkową stopy bezrobocia z 50 obserwacji. Współczynnik kurtozy określony w statystyce opisowej jako kurtoza jest miarą płaskości rozkładu wartości zmiennej Y i określany według wzoru podanego w Excelu jako:

$$K = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{(x_j - \bar{x})}{S} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}. \quad (1.7.6)$$

Współczynnik skośności (skośność) wyraża odchylenie rozkładu względem pionu określonego przez średnią. Program Excel oblicza go według wzoru:

$$S_k = \frac{n(n+1)}{(n-1)(n-2)} \sum \left(\frac{(x_j - \bar{x})}{S} \right)^3. \quad (1.7.7)$$

Natomiast według K. Jajugi⁶⁰ klasyczny współczynnik asymetrii (skośności) określony jest według zależności:

$$S_k = \frac{1/n \cdot \sum (x_i - \bar{x})^3}{S^3}.$$

Z tabeli 1.7.5 wynika, że procent popełnienia błędu jest mniejszy od poziomu istotności wyrażonego procentowo jako $\alpha = 5\%$, gdyż wynosi 1,76 %.

Histogram

Skorzystanie z opcji (*Histogram*) umożliwia wstępne zorientowanie się co do rozkładu normalnego reszt 49 obserwacji. Obliczone zostają częstości występowania wartości $e = y - \hat{y}$ w określonych przedziałach oraz podany jest wykres Pareto.

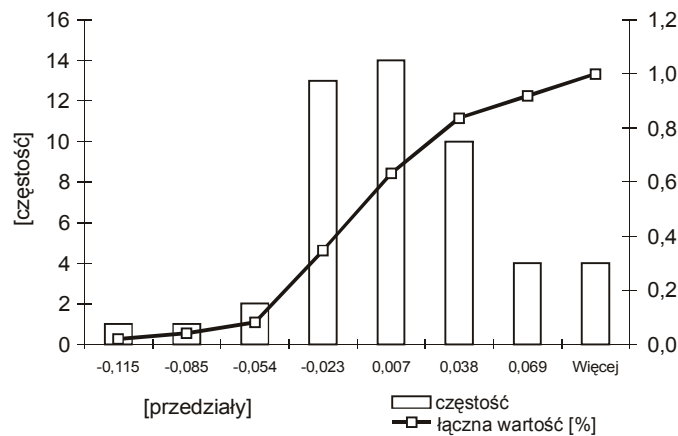
⁶⁰ K. Jajuga (red.), *Ekonometria. Metody...*, op.cit., rozdz. 1.4.

Tabela 1.7.6

Częstość występowania wartości reszt oraz wykres Pareto

Zbiór danych (koszyk)	Częstość	Łączna wartość [%]
-0,115	1	2,04
-0,085	1	4,08
-0,054	2	8,16
-0,023	13	34,69
0,007	14	63,27
0,038	10	83,67
0,069	4	91,84
Więcej	4	100,00

Źródło: Opracowanie własne w Excelu.



Źródło: Opracowanie własne w Excelu.

Wykres 1.7.1. Histogram reszt

Regresja

Do okna dialogowego w Excelu wprowadzamy następujące informacje:

- zakres wejściowy Y: \$C\$3:\$C\$52,
- zakres wejściowy X: \$D\$3:\$F\$52,
- poziom ufności – 95%,
- wyjście na nowy arkusz,
- składniki resztowe, a w tym standaryzowane składniki resztowe; rozkład reszt.

W wyniku otrzymujemy raport podsumowujący, którego dane zamieszczono w tab. 1.7.7–1.7.10.

Tabela 1.7.7

Statystyka opisowa regresji liniowej

Wielokrotność R	0,671047
R^2	0,450304
Dopasowany współczynnik determinacji R^2	0,414454
Odchylenie standardowe reszt (S_e)	0,0473321
Obserwacje n	50

Źródło: Opracowanie własne w Excelu.

W tabeli 1.7.7 wielokrotność R to omówiony wcześniej współczynnik korelacji wielorakiej R_w . Dopasowany współczynnik determinacji (R^2) jest skorygowanym o liczbę stopni swobody współczynnikiem determinacji i wyznaczany jest według wzoru:

$$\bar{R}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (1.7.8)$$

$$\text{lub } \bar{R}^2 = 1 - \frac{n-1}{n-2} (1 - R)^2.$$

Tabela 1.7.8

Analiza wariancji

Odniesienia	df	SS	MS	F	Istotność F
Regresja	3	0,0844	0,0281	12,5609	0,000004
Resztkowy	46	0,1031	0,0022		
Razem	49	0,1875			

gdzie:

df – ilość stopni swobody w odniesieniu do obliczania MS ($m = 3$ – ilość zmiennych objaśniających, $46 = 50 - m - 1$),

SS – suma kwadratów, przy czym:

regresja – ESS w REGLINP estymowana suma kwadratów, a resztowy – RSS :

$$ESS = \sum_{i=1}^n (\hat{y}_i - \hat{\bar{y}})^2. \quad (1.7.9)$$

$$RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2. \quad (1.7.10)$$

MS – średnie z sum kwadratów obliczane jako $MS = SS/df$,
 F – statyka Fishera określona dla łącznej oceny parametrów modelu,
Istotność F – prawdopodobieństwo błędu oszacowania F .

Źródło: Opracowanie własne w Excelu.

Tabela 1.7.9

Wnioskowanie statystyczne

Odniesienia	Współczynniki	Błąd standardowy	t -Stat	Wartość- p
Przecięcie	-0,945176	0,251316	-3,7609	0,000477
Zmienna 1	-0,00129131	0,000474643	-2,7206	0,009167
Zmienna 2	1,9008	0,425549	4,4667	0,00005
Zmienna 3	0,0196761	0,00373272	5,2712	0,000004

Źródło: Opracowanie własne w Excelu.

Ponieważ nie podano w programie nazw zmiennych tylko obszary ich wartości, system nadał nazwy umowne: Przecięcie, Zmienna1, Zmienna2, Zmienna3. W tabeli 1.7.9 współczynniki odpowiadają parametrom modelu, a ich wartości oraz błędy standardowe oszacowania a_0, a_1, a_2, a_3 są identyczne jak otrzymane funkcją REGLINP (tab. 1.7.2). Statystyki t -Stat odpowiadają wcześniej omówionym otrzymanym z testu t -Studenta, stosowanym przy badaniu istotności statystycznej określonej zmiennej X_j w modelu ekonometrycznym z prawdopodobieństwem popełnienia błędu określonym w kolumnie Wartość- p .

Kolejna tabela 1.7.10 zawiera reszty, które program oblicza automatycznie na podstawie oszacowanych współczynników (parametrów) modelu. W tabeli 1.7.10 wprowadzono dodatkowo kolumnę Standaryzacja reszt. Standaryzację reszt przeprowadzono według wzoru analogicznego do (1.4.4):

$$z_i = \frac{(e_i - \bar{e})}{S_e}, \quad (1.7.11)$$

gdzie:

e_i – kolejne wartości reszt obserwacji 1–50,

\bar{e} – średnia arytmetyczna reszt wynosząca 0,0000025 (w przybliżeniu zero),

S_e – odchylenie standardowe reszt (obliczone w tym przykładzie dla $n = 50$ obserwacji).

Odchylenie standardowe populacji $n = 50$ obserwacji wynosi 0,045399439.

Mając określony współczynnik determinacji (R^2) = 0,450304 oraz wartość średnią zmiennej objaśnianej Y (tab. 1.7.5), która wynosi 0,195567, możemy

obliczyć dalsze pomocnicze miary dopasowania modelu do danych rzeczywistych:

- współczynnik zbieżności: $\Phi^2 = 1 - R^2 = 1 - 0,450304 = 0,549696$,
- współczynnik wyrazistości $W = S_e / \bar{y} = 0,0473321/0,195567 = 0,242031$.

Tabela 1.7.10

Składniki resztowe 50 obserwacji

Obserwacje	\hat{Y}	Reszta	Standaryzacja reszt
1	0,177011406	0,060300054	1,328205901
2	0,179005575	0,014214825	0,313100216
3	0,27842227	0,03420381	0,753391652
4	0,226926608	-0,026246368	-0,578126502
5	0,198276844	-0,025140394	-0,553765538
6	0,193019235	-0,036819585	-0,811019609
7	0,177964923	-0,000958903	-0,021127004
8	0,166995835	0,018732875	0,412617956
9	0,225104011	0,085418619	1,88148508
10	0,210259973	-0,076616553	-1,687615642
11	0,15044986	-0,03450196	-0,759969986
12	0,150801731	-0,044377221	-0,977489418
13	0,183635756	-0,086508026	-1,905492184
14	0,113857839	-0,007194919	-0,158485886
15	0,186822408	-0,080039058	-1,763002146
16	0,248931136	-0,002167996	-0,047759329
17	0,178674426	-0,014199486	-0,312773397
18	0,145684066	-0,027520396	-0,606189127
19	0,2195413	-0,02481928	-0,546692459
20	0,28380473	0,01397215	0,30775489
21	0,192311057	0,028625563	0,630521267
22	0,145475123	0,006282567	0,138378705
23	0,176793611	0,059551609	1,311720114
24	0,201842719	-0,011261579	-0,248060982
25	0,151087402	0,099244388	2,1860212
26	0,237455126	-0,115404306	-2,541981987
27	0,207829753	-0,041697163	-0,918456588
28	0,122311315	-0,008091655	-0,178238008
29	0,182902184	0,001014546	0,02234158
30	0,259253967	0,056997983	1,255472163

cd. tab.1.7.10

Obserwacje	\hat{Y}	Reszta	Standaryzacja reszt
31	0,180873748	-0,001143158	-0,025185528
32	0,220341623	0,006712367	0,147845785
33	0,236034024	-0,039385054	-0,867528432
34	0,202321484	0,003948476	0,086966392
35	0,174464524	0,075663206	1,666605483
36	0,151410822	0,032325248	0,712013123
37	0,289688985	-0,004852755	-0,106895727
38	0,18078723	-0,03581883	-0,788976274
39	0,184627495	0,008332405	0,183529881
40	0,145880441	-0,025304931	-0,557389737
41	0,202402592	0,021291418	0,46897423
42	0,154433956	-0,007798266	-0,171775619
43	0,191639495	-0,043464215	-0,957378911
44	0,280842387	0,029476373	0,649261797
45	0,235209556	-0,049127866	-1,082130474
46	0,201247414	-0,004430614	-0,097597345
47	0,169744085	0,017085605	0,376334037
48	0,237275332	0,058114398	1,280063109
49	0,218848675	0,062537855	1,377497285
50	0,147824737	0,080856743	1,781001992

Źródło: Opracowanie własne w Excelu.

1.7.3. Dalsze badania modelu ekonometrycznego (w Excelu)

a) Autokorelacja składnika resztowego

Stosujemy statystykę Durbina-Watsona, podaną w publikacji J. Dziechciarza, która jest analogiczna do wzoru (1.1.19)⁶¹:

$$DW = \frac{\sum_{t=1}^{n-1} (e_{t+1} - e_t)}{\sum_{t=1}^n e_t^2},$$

gdzie:

e_t – reszta dla obserwacji zmiennej objaśnianej w okresie t .

⁶¹ J. Dziechciarz (red.), *Ekonometria. Metody...*, op.cit.

Procedura postępowania jest następująca:

- dla poszczególnych obserwacji wyznaczamy wartości DW ,
- dla przyjętego poziomu istotności, np. $\alpha = 0,05$, przy danym n i ilości zmiennych $k = m + 1$ odczytujemy z tablicy rozkładu statystyki Durбина-Watsona wartości krytyczne d_L oraz d_U ,
- sprawdzamy, czy:
 - $DW < d_L$, jeśli tak, to odrzucamy hipotezę zerową H_0 o braku autokorelacji na rzecz hipotezy alternatywnej H_1 o dodatniej autokorelacji,
 - $DW > d_U$ – brak jest podstaw do odrzucenia H_0 ,
 - $d_L \leq DW \leq d_U$ – nie podejmujemy żadnej decyzji.

b) Heteroscedastyczność

Stosowanie KMNK zakłada stałość wariancji składnika losowego (reszty) w czasie. Jednakże w modelach praktycznie stosowanych wariancje mają tendencję wzrostową lub malejącą. Do zbadania tego zjawiska służy test Goldfelda i Quandta, a postępowanie w tym zakresie jest przykładowo następujące⁶²:

- dzielimy n obserwacji intuicyjnie według układu wektora reszt otrzymanego KMNK na dwie grupy: n_1, n_2 ,
- szacujemy wektory parametrów KMNK oddzielnie dla każdej grupy określonej jako A i B ,
- określamy wariancje dla grup oraz statystykę, korzystając ze wzorów, które są analogiczne do 1.1.21, 1.1.22, przy czym $m = k$:

$$S_1^2 = \frac{\sum_{i \in A} e_i^2}{n_1 - m - 1}, \quad S_2^2 = \frac{\sum_{i \in B} e_i^2}{n_2 - m - 1}, \quad F_e = \frac{S_1^2}{S_2^2},$$

- odczytujemy F_α z tablicy rozkładu F dla ilości stopni $n_1 - m - 1$ w liczniku oraz $n_2 - m - 1$ w mianowniku na poziomie istotności α , np. 0,05,
- sprawdzamy, czy $F_e > F_\alpha$, jeśli tak, to odrzucamy hipotezę zerową H_0 równości wariancji każdej z grup na rzecz hipotezy alternatywnej H_1 :

$$H_0 : \sigma_1^2 = \sigma_2^2, \quad H_1 : \sigma_1^2 > \sigma_2^2.$$

Jeżeli H_1 jest postaci jak niżej, stosujemy statystykę odwrotną:

$$H_1 : \sigma_1^2 < \sigma_2^2, \quad F_e = \frac{S_2^2}{S_1^2}.$$

i odrzucamy H_0 , jeśli $F_e > F_{\alpha(n_2-m-1, n_1-m-1)}$.

⁶² Ibidem.

c) Normalność rozkładu składnika losowego (wyrażonego przez reszty)

Jeśli próba jest mała ($n \leq 30$), to przeprowadzamy test Hellwiga w sposób następujący:

– szacujemy odchylenie standardowe S dla ciągu e_t reszt kolejnych obserwacji i dokonujemy ich standaryzacji, według analogicznych wzorów (1.1.9), (1.1.10):

$$S = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}, \quad e_t' = \frac{e_t}{S}.$$

(We wzorze na e_t' przyjmuje się wartość średniej reszt równą zero oraz traktuje się ciąg reszt jako „populację” do obliczenia odchylenia standardowego. Jest to zbieżne z obliczeniem wartości standaryzowanych reszt zamieszczonych w tabeli 1.7.10).

– sortujemy narastająco zestandaryzowane reszty i określamy dla nich, na podstawie tablicy rozkładu normalnego, wartości dystrybuanty,

– dzielimy odcinek możliwych wartości dystrybuanty $[0; 1]$ na cele, czyli równe części i określamy liczbę K cel, w których nie ma żadnej wartości,

– odczytujemy z tablicy testu Hellwiga dla poziomu istotności $\alpha = 0,05$ wartości krytyczne K_1 oraz K_2 ,

– sprawdzamy hipotezę zerową H_0 : reszty mają rozkład normalny, gdy $K_1 \leq K \leq K_2$, to nie ma podstaw na odrzucenie hipotezy zerowej.

Dla dużych prób zmienna losowa K ma asymptotycznie rozkład Poissona z parametrem λ i według propozycji podanej w publikacji J. Dziechciarza⁶³ stosujemy test λ lub test Kołomogorowa o statystyce testowej dla maksymalnej różnicy:

$$D_n = \sup_x |F_n(x) - F_0(x)|, \quad (1.7.12)$$

gdzie:

$x = e_t$, $F_0(x)$ – dystrybuanta rozkładu normalnego $N(0, S)$,

$F_n(x)$ – dystrybuanta empiryczna z próby.

Procedura postępowania z przeprowadzeniem testu Kołomogorowa ze statystyką λ , w której możemy korzystać z wyrażeń Excela, jest następująca:

– obliczamy odchylenie standardowe S dla próby n obserwacji na podstawie podanego wzoru (1.7.12); dokonujemy standaryzacji reszt,

– sortujemy narastająco reszty, a następnie określamy na podstawie tablicy VI w *Aneksie* dystrybuanty rozkładu normalnego,

– obliczamy dystrybuanty empiryczne,

⁶³ Ibidem.

– obliczamy różnice:

$$|F_n(e_t) - F_0(e_t)|$$

i odczytujemy różnicę maksymalną, kierując się wzorem (1.7.12),

– określamy wartość obliczeniową statystyki⁶⁴:

$$\lambda = \sqrt{n}D_n, \quad (1.7.13)$$

– określamy wartość krytyczną $\lambda_0 = Q(\lambda)$ na podstawie rozdziału IV oraz tablicy 48 rozkładu $K(y)$ Kołomogorowa⁶⁵, przy czym:

$$Q(\lambda) = K(y) = P(\sqrt{n}D_n < \lambda = y),$$

– sprawdzamy, czy dla danego poziomu istotności, np. $\alpha = 0,05$,

$$\lambda = \sqrt{n}D_n \geq \lambda_0,$$

co oznacza niewystępowanie normalności rozkładu reszt i odrzucenie hipotezy zerowej:

$$H_0 : \varepsilon_t \sim N(0, S), (t = 1, \dots, n).$$

d) Losowość

W zależności od ilości obserwacji w próbie do badania losowości reszt modelu liniowego stosujemy⁶⁶:

– test serii ($n \leq 30$),

– twierdzenie von Misesa ($n > 30$).

Procedura testu serii:

– przyporządkowujemy symbole A resztom ($e_t > 0$), a wartościom ujemnym symbol B (w celu wyeliminowania reszt ujemnych możemy zwiększyć w Excelu dokładność obliczeń),

– obliczamy empiryczną liczbę serii K_e (ciągów o jednakowym symbolu),

– odczytujemy z tablicy liczby serii przy teście dwustronnym K_1 oraz K_2 ,

– sprawdzamy, czy: $K_1 \leq K_e \leq K_2$, co oznacza potwierdzenie hipotezy zerowej:

$$H_0 : \varepsilon_t \text{ losowy.}$$

W przypadku dużej próby $n > 30$ nieuporządkowanych narastająco określamy:

– maksymalną długość serii r z liczby serii N_n o długości r w n doświadczeniach,

⁶⁴ Ibidem, rozdz. 6.5.

⁶⁵ R. Z i e l i ń s k i, *Tablice statystyczne*, PWN, Warszawa 1972.

⁶⁶ J. Dziechciarz (red.), *Ekonometria. Metody...*, op.cit.

– obliczamy:

$$\mu = \frac{1 - 0,5^r}{0,5^{r+1}}, \quad \sigma = \sqrt{2^{2r+2} - (2r+1)2^{r+1}}$$

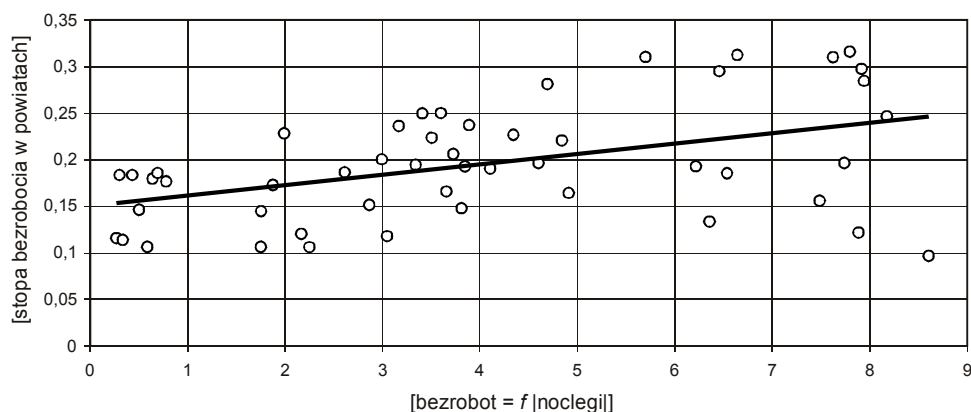
– szacujemy prawdopodobieństwo pojawienia się serii o długości r :

$$P(N_n \geq 1) = 1 - \Phi \left(\frac{\left(1 - \frac{n}{\mu}\right) \mu^{3/2}}{\sigma \sqrt{n}} \right),$$

– sprawdzamy, czy $P(N_n \geq 1) < 0,05$ (poziom istotności), jeśli tak, to brak podstaw do odrzucenia hipotezy o losowości reszt.

e) Obserwacje odróżniające się od pozostałych – nietypowe, wpływowe

Interesują nas skutki oddziaływania określonych obserwacji na model ekonometryczny. Nietypowe cechują duże reszty, a po ich wyeliminowaniu następuje poprawa współczynnika determinacji (R^2), natomiast w przypadku wpływowych następuje wyraźna zmiana parametrów modelu. Do wykrywania obserwacji wpływowych stosujemy analizę wykresu rozrzutu obserwacji dla modelu z jedną zmienną objaśniającą, np. z „noclegi”.



Źródło: Opracowanie własne w Excelu.

Wykres 1.7.2. Wykres punktowy rozrzutu par obserwacji zmiennych „bezrobot” i „noclegi”

Wpływowe są obserwacje najbardziej oddalone od linii regresji „bezrobot” = $f(\text{„noclegi”})$, przy czym zmienną objaśnianą jest „stopa bezrobocia w powiatach”.

Dla wielu zmiennych objaśniających m modelu ekonometrycznego mamy następującą liczbę wykresów⁶⁷:

$$\binom{m}{2} = \frac{m \cdot (m-1)}{2}.$$

W dalszej części opieramy się na przykładzie modelu liniowego stopy bezrobocia „bezrobot” w powiatach o trzech zmiennych objaśniających:

- regon – liczba jednostek regon na 1 tys. mieszkańców,
- lud_prod – udział ludności w wieku produkcyjnym,
- noclegi – liczba miejsc noclegowych na 1 tys. mieszkańców.

Analizę graficzną poszczególnych zmiennych stosuje się z uwagi na znaczne odbieganie niektórych obserwacji od pozostałych. Może być jednak obserwacja, której wpływowość możemy zaobserwować dopiero w przestrzeni wielowymiarowej. Obserwacje wpływowe wykrywane są przy zastosowaniu klasyfikacji wielowymiarowej, np. taksonomii wrocławskiej. Wówczas obserwacje dzielone są na klasy, przy czym obserwacje wpływowe przydzielane są do osobnych klas ze względu na znaczne oddalenie od pozostałych obserwacji (podrozdz. 1.4.2.4).

Jednak podstawowym narzędziem do wykrywania obserwacji wpływowych jest macierz rzutowania, która zależy tylko od macierzy zmiennych objaśniających:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T. \quad (1.7.14)$$

Ponieważ w modelu liniowym $\hat{\mathbf{y}} = \mathbf{X}\mathbf{a} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, tak więc:

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} = \sum_{j=1}^n h_{ij} y_j.$$

Wielkość i -tego elementu h_i głównej przekątnej macierzy H określa wpływ i -tej obserwacji na parametry modelu. Korzystamy z następującej formuły, aby zaobserwować, jak zmienia się wektor parametrów \mathbf{a} pod wpływem usunięcia z danych i -tej obserwacji o (y_i, x_i) ⁶⁸:

$$\mathbf{a}(i) = \mathbf{a} - (\mathbf{X}^T \mathbf{X})^{-1} x_i^T \frac{e_i}{1 - h_i},$$

gdzie:

- e_i – i -ta reszta obserwacji n określona dla obserwacji modelu oszacowanego KMNK,
- $e_i/(1-h_i)$ – współczynnik proporcjonalności.

⁶⁷ Ibidem, rozdz. 6.7.

⁶⁸ Ibidem, wzór 6.15.

Usunięcie obserwacji, której h_i jest duże, ma wpływ na estymację. W modelach z jedną zmienną objaśniającą elementy głównej macierzy H określają, jak dana obserwacja (y_i, x_i) różni się od średniej:

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}.$$

W metodzie macierzy rzutowania wprowadzono pojęcie „kwadrat odległości Mahalonobisa”⁶⁹:

$$MD_i^2 = (n-1)\left(h_i - \frac{1}{n}\right). \quad (1.7.15)$$

Najmniejsza wartość $h_i = 1/n$ występuje wtedy, gdy obserwacja równa się średniej, natomiast gdy wartości h_i są bliskie jedności, obserwacja jest bardzo oddalona od średniej i jednocześnie od wszystkich innych obserwacji. Zatem wartości diagonalne macierzy rzutowania pozwalają wyłonić obserwacje wpływowe⁷⁰. Skoncentrujemy się jednak na naszym przykładzie, którego podstawą jest plik 50 obserwacji stopy bezrobocia w powiatach. Do takiej bazy oraz modelu liniowego o trzech zmiennych objaśniających parametry określone w tabeli 1.7.2 funkcją REGLINP są następujące:

a_3	a_2	a_1	a_0
0,0196761	1,9008	-0,00129131	-0,945176

Przypomnijmy sobie model liniowy naszego przykładu:

$$\hat{Y} = -0,945176 - 0,00129131 \cdot (\text{regon}) + 1,9008 \cdot (\text{lud_prod}) + 0,0196761 \cdot (\text{noclegi}).$$

W analitycznym wykrywaniu obserwacji wpływowych metodą macierzy rzutowania postępowanie jest następujące:

- określenie h^* , czyli krytycznego poziomu h_i :
wartość przeciętna $h_i = m/n = 3/50 = 0,06$, przy czym $h^* = 2(m/n) = 0,12$ lub $h^* = 3(m/n) = 0,18$; przyjmijmy $h^* = 0,18$,
alternatywnym sposobem jest podzielenie przedziału $[0,1]$ dla h_i na trzy części informujące o „statusie odległościowym” obserwacji:
 $h_i \in [1/n; 0,2]$ bezpieczne,

⁶⁹ Ibidem, wzór 6.19.

⁷⁰ Ibidem. Przykład 6.15 realizacji graficznej oraz analitycznej wyłonienia obserwacji wpływowych dla modelu z jedną zmienną objaśniającą.

$h_i \in [0,2; 0,5]$ ryzykowne,

$h_i \in [0,5; 1]$ niedopuszczalne,

– wyznaczenie macierzy rzutowania \mathbf{H} o wymiarach (50x50), po wcześniejszym określeniu poniższych macierzy z uwzględnieniem zmiennej stałej tożsamościowo równej 1:

\mathbf{X} – o wymiarach (50x4),

\mathbf{X}^T – macierzy transponowanej o wymiarach (4x50).

Prace te wykonano w Excelu, korzystając z funkcji: TRANSPONUJ(...), MACIERZ.ILOCZYN(...;...), MACIERZ.ODW(...).

Obszary macierzy \mathbf{X} oraz \mathbf{X}^T są duże i dlatego zamieszczamy tylko iloczyn i odwrotność macierzy ($\mathbf{X}^T\mathbf{X}$), a do innych działań podano w nawiasach obszary macierzy:

$$\mathbf{X}^T\mathbf{X} \rightarrow (4 \times 50) \cdot (50 \times 4) = (4 \times 4)$$

Tabela 1.7.11

Macierz iloczynu $\mathbf{X}^T\mathbf{X}$

50	3909,075505	30,588462	200,360680
3 909,075505	325 298,687567	2 409,730620	14 347,390431
30,588462	2 409,730620	18,747590	120,266571
200,360680	14 347,390431	120,2665707	1 118,924083

Źródło: Opracowanie własne w Excelu.

$$(\mathbf{X}^T\mathbf{X})^{-1} \rightarrow (4 \times 4)$$

Tabela 1.7.12

Macierz odwrotna $(\mathbf{X}^T\mathbf{X})^{-1}$

28,192225	0,022077	-47,136638	-0,264898
0,022077	0,000101	-0,049324	0,000059
-47,136638	-0,049324	80,832990	0,384740
-0,264898	0,000059	0,384740	0,006219

Źródło: Opracowanie własne w Excelu.

$$\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} (50 \times 4) \times (4 \times 4) = (50 \times 4); \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T (50 \times 4) \times (4 \times 50) = (50 \times 50).$$

Otrzymaliśmy zatem macierz rzutowania o wymiarach (50x50), z zaokrągleniem do dwóch miejsc po przecinku. Obserwacje wyeliminowano według wstępnego kryterium o $h_i < 0,10$. W wyniku tego działania pozostało 14 obserwacji (1–14), w tym dwie o wartości $h_i > 0,18$ (tab. 1.7.13).

Tabela 1.7.13

Macierz rzutowania

0,11	0,01	-0,04	-0,03	-0,02	-0,09	0,04	0,10	-0,01	-0,07	0,07	0,11	0,09	0,08	3	1
0,01	0,12	0,08	-0,05	0,12	-0,04	0,01	0,02	0,09	-0,03	0,09	0,01	0,00	-0,10	6	2
-0,04	0,08	0,13	0,02	0,14	0,09	0,08	0,00	0,04	0,06	0,03	0,00	0,01	-0,10	8	3
-0,03	-0,05	0,02	0,10	-0,02	0,14	0,03	-0,02	-0,03	0,11	-0,07	-0,02	0,00	0,06	11	4
-0,02	0,12	0,14	-0,02	0,18	0,05	0,09	0,02	0,06	0,03	0,06	0,01	0,02	-0,13	13	5
-0,09	-0,04	0,09	0,14	0,05	0,24	0,09	-0,05	-0,04	0,19	-0,10	-0,04	-0,01	0,00	14	6
0,04	0,01	0,08	0,03	0,09	0,09	0,17	0,09	-0,07	0,04	0,00	0,10	0,11	0,00	16	7
0,10	0,02	0,00	-0,02	0,02	-0,05	0,09	0,11	-0,04	-0,05	0,06	0,12	0,11	0,06	20	8
-0,01	0,09	0,04	-0,03	0,06	-0,04	-0,07	-0,04	0,13	-0,01	0,07	-0,05	-0,06	-0,06	25	9
-0,07	-0,03	0,06	0,11	0,03	0,19	0,04	-0,05	-0,01	0,15	-0,07	-0,04	-0,02	0,02	28	10
0,07	0,09	0,03	-0,07	0,06	-0,10	0,00	0,06	0,07	-0,07	0,11	0,05	0,03	-0,03	33	11
0,11	0,01	0,00	-0,02	0,01	-0,04	0,10	0,12	-0,05	-0,04	0,05	0,13	0,12	0,07	37	12
0,09	0,00	0,01	0,00	0,02	-0,01	0,11	0,11	-0,06	-0,02	0,03	0,12	0,12	0,07	44	13
0,08	-0,10	-0,10	0,06	-0,13	0,00	0,00	0,06	-0,06	0,02	-0,03	0,07	0,07	0,19	45	14
3	6	8	11	13	14	16	20	25	28	33	37	44	45		

W ostatnim wierszu tabeli podano numery obserwacji, główną przekątną macierzy jest macierz od lewego górnego rogu (wartość 0,11 do wartości 0,19 nad obserwacją 45).

Źródło: Opracowanie własne w Excelu.

– w wyniku przeanalizowania elementów $h_i > 0,18$ głównej przekątnej macierzy **H** usuwamy i -tą obserwację o największym h_i ; jest nią obserwacja 14.,

– ponowna estymacja parametrów modelu funkcją REGLINP i znowu badanie jak poprzednio,

– jeśli $h_i \leq h^*$, wówczas kończymy proces iteracyjny eliminacji obserwacji wpływowych.

Po usunięciu 14. obserwacji z macierzy **X** i ponownym obliczeniu macierzy rzutowania o wymiarach (49x49), na głównej przekątnej wystąpiły dwie obserwacje (28 oraz 45) o $h_i = 0,19$. Ponieważ mają tę samą wartość, usuwamy je jednocześnie z macierzy **X**. Przystąpmy w kolejnej iteracji do ponownego obliczenia skorygowanej macierzy rzutowania o wymiarach (47x47).

Pojawiły się teraz dwie obserwacje o $h_i > 0,18$, tj. obserwacja 13. oraz 19.:

$$13 (h_i = 0,21), 16 (h_i = 0,19).$$

Aby przyspieszyć obliczenia, usuwamy zarówno obserwację 13., jak i 16. Uzyskujemy macierz rzutowania, w której dla jednej obserwacji, tj. ósmej, wartość $h_i = 0,21$. Ze względu na objętość niniejszego materiału pozostawia-

my Czytelnikowi dalsze badanie, aż do doprowadzenia do warunku $h_i \leq 0,18$. Uznajemy jednak, że w przybliżeniu spełniony jest warunek bezpieczny $h_i \in [1/n; 0,2]$ i dokonujemy ponownej estymacji modelu liniowego 45 obserwacji po trzech iteracjach funkcją REGLINP.

Tabela 1.7.14

Raport z realizacji funkcją REGLINP

a_3	a_2	a_1	a_0	
0,020603	1,724477	-0,000944	-0,864073	parametry
0,004015	0,505905	0,000623	0,291230	błąd standardowy
0,424877	0,047287			$R^2; S_e$
10,096372	41			$F; iss$
0,067729	0,091680			$ESS; RSS$

Źródło: Opracowanie własne w Excelu.

A teraz przypomnijmy sobie określony w tabeli 1.7.2 funkcją REGLINP model liniowy o następujących parametrach i statystyce opisowej.

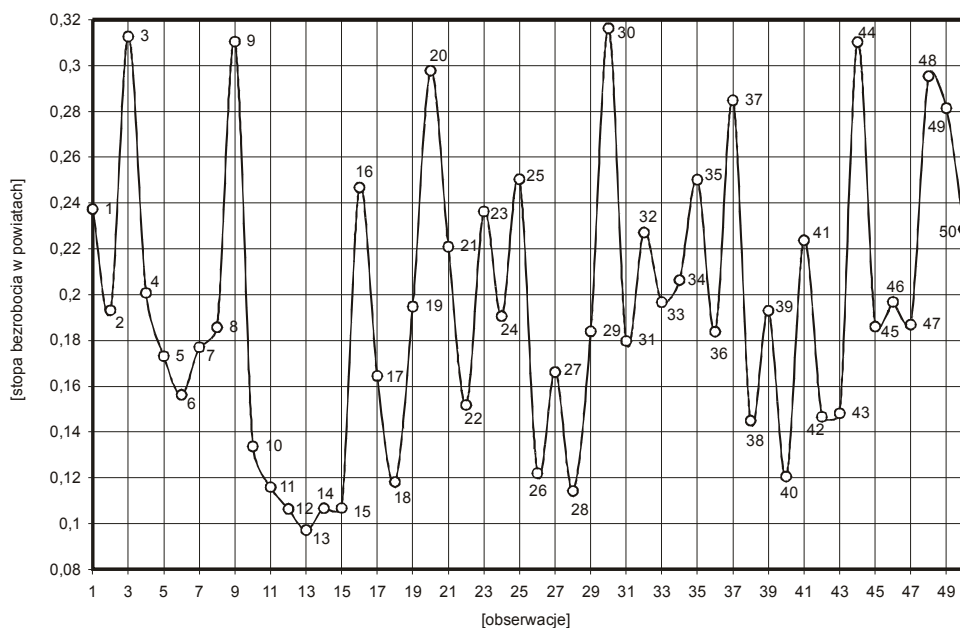
a_3	a_2	a_1	a_0	
0,0196761	1,9008	-0,00129131	-0,945176	parametry
0,00373272	0,425549	0,000474643	0,251316	błąd standardowy
0,450304	0,047332			$R^2; S_e$
12,5609	46			$F; iss$
0,08442122	0,103055			$ESS; RSS$

Źródło: Opracowanie własne w Excelu.

Według publikacji J. Dziechciarza⁷¹: „Obserwacje uważa się za wpływową, jeśli w wyniku [...] usunięcia jej z modelu znacznie zmieniają się oszacowane parametry modelu”.

Porównanie nieznacznej zmiany współczynnika determinacji R^2 oraz wyrażonej zmiany parametrów a_0, a_1, a_2 przekonuje nas, że usunięte obserwacje: 14., 28., 45., 13. oraz 16. były obserwacjami wpływowymi. Punktową wartość zmiennej objaśnianej w kolejnych obserwacjach (powiatach) ukazano na wykresie 1.7.3. Wykres jest tylko orientacją umiejscowienia wartości stopy bezrobocia względem średniej – 0,2. Wpływ zmiennych objaśniających na obserwacje uznane za wpływowe należy bowiem, jak już wspomniano, rozpatrywać w przestrzeni trójwymiarowej.

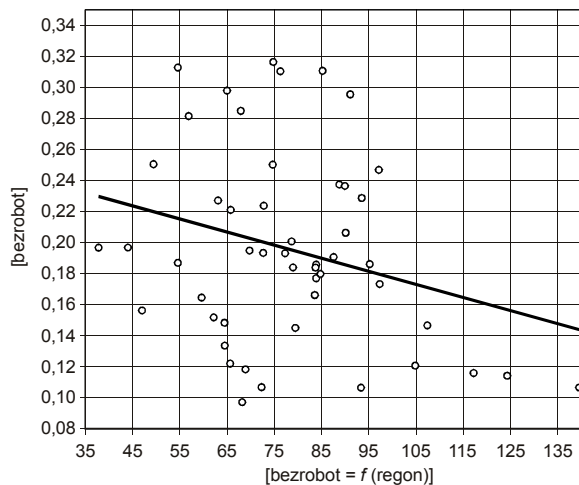
⁷¹ Ibidem, rozdz. 6.7.



Źródło: Opracowanie własne w Excelu.

Wykres 1.7.3. Wykres punktowy zmiennej „bezrobot”

Zobaczymy jeszcze umiejscowienie obserwacji uznanych za wpływowe na wykresie punktowym przykładowej zmiennej objaśniającej „regon” oraz zmiennej objaśnianej „bezrobot”.



Źródło: Opracowanie własne w Excelu.

Wykres 1.7.4. Wykres punktowy obserwacji pary zmiennych „bezrobot” i „regon”

Przykład wskazuje na niewystępowanie wyraźnej obserwacji nietypowej, która byłaby znacznie oddalona od innych obserwacji, lecz w przybliżeniu leżała wzdłuż w linii regresji. Natomiast z punktu widzenia jednej zmiennej objaśniającej występują obserwacje wpływowe (np. 13. o wartości $\text{regon} \approx 68,22$), znacznie oddalone równoległe od linii regresji. Jako ćwiczenie połączmy linią ciągłą na wykresie 1.7.4 otrzymane wcześniej według iteracji, macierzy rzutowania, obserwacje wyeliminowane, tj. 13., 14., 16., 28., 45. Następnie sprawdźmy wzrokowo ich wpływ na linię regresji. Jednak należy pamiętać, że w przestrzeni trójwymiarowej wpływ na model ekonometryczny jest inny.

1.7.4. Zastosowanie programu GRETL

Jak już wspomniano w podrozdz. 1.2.2.1, pakiet GRETL jest bezpłatnym, dostępnym w internecie oprogramowaniem do estymacji i weryfikacji modeli ekonometrii. Jego zastosowanie zaprezentowane zostanie na przykładzie niektórych miar oceny weryfikacji otrzymanego modelu ekonometrycznego.

a) Ocena istotności statystycznej parametrów modelu o procedurze:

– określenie zmiennych niezależnych X_j (objaśniających) oraz zmiennej objaśnianej Y ,

– zastosowanie testu t -Studenta istotności parametru α_j . Hipoteza zerowa zakłada brak istotności statystycznej danego parametru $H_0: [\alpha_j = 0]$, natomiast hipoteza alternatywna $H_1: [\alpha_j \neq 0]$. Przypomnijmy sobie statystykę, która określona jest wzorem:

$$t_j = \frac{|a_j|}{S(a_j)},$$

– oszacowanie parametrów liniowego modelu ekonometrycznego na bazie danych dotyczących stopy bezrobocia w pięćdziesięciu powiatach, lecz rozszerzonej np. o pięć dalszych zmiennych objaśniających⁷²:

zatr_przem – udział zatrudnionych w przemyśle (sekcja C+D+E+F),

sklep_zat – liczba zatrudnionych na 1 sklep,

mieszk_new – liczba mieszkań oddanych do użytku na 1 tys. mieszkańców,

ha_urol – udział użytków rolnych w powierzchni powiatu,

lud_250tys – zmienna (0–1), przy czym 1= miasto o liczbie ludności powyżej 250 tys., 0 – pozostałe.

⁷² [www.kufel.torun.pl], zbiór [powiaty2002.gdt].

Zmienne objaśniające regon, lud_prod, noclegi już były w naszej wcześniej omawianej bazie, stosowanej do obliczeń w Excelu (tab. 1.7.1).

Po zdefiniowaniu struktury bazy w programie GRETL, korzystamy z menu (*Model/Klasyczna metoda najmniejszych kwadratów*) i rezultatem jest model 1 estymacji wraz ze statystyką opisową dla zmiennej zależnej „bezrobot”, której fragment podano w tabeli.

Tabela 1.7.15

Model 1 wraz ze statystyką opisową

Zmienna	Współczynnik	Błąd stand.	Statystyka <i>t</i>	Wartość <i>p</i>	
const	-0,763574	0,285219	-2,6772	0,01063	**
regon	-0,00112881	0,000533592	-2,1155	0,04051	**
zatr_przem	-0,0900188	0,064573	-1,3941	0,17081	
sklep_zat	-0,0512426	0,0330824	-1,5489	0,12908	
mieszk_new	-0,00520729	0,00518114	-1,0050	0,32077	
lud_prod	1,91159	0,451284	4,2359	0,00013	***
ha_urol	-0,0299996	0,0495286	-0,6057	0,54805	
noclegi	0,0126064	0,00492128	2,5616	0,01419	**
lud_250tys	-0,0255601	0,0345797	-0,7392	0,46402	

Średnia arytmetyczna zmiennej zależnej = 0,195567

Odchylenie standardowe zmiennej zależnej = 0,061855

Suma kwadratów reszt = 0,0909375

Błąd standardowy reszt = 0,0470955

Współczynnik determinacji (R^2) = 0,514938

Skorygowany współczynnik determinacji (\bar{R}^2) = 0,420292

Statystyka $F(8, 41) = 5,44066$ (wartość $p = 0,000107$)

Wyłączając stałą, największa wartość p jest dla zmiennej **ha_urol**.

Źródło: Opracowanie własne w GRETL.

W tabeli 1.7.15 w celu szybkiej orientacji symbolami gwiazdek program oznacza parametry istotnie różniące się od zera, przy poziomach istotności: 1% – ***, 5% – **, 10% – *. Na zakończenie raportu program proponuje zmienną do usunięcia, która charakteryzuje się wysokim prawdopodobieństwem popełnienia błędu p w konstrukcji modelu z tą zmienną oraz niską statystyką t .

– usunięcie zmiennej **ha_urol** ze struktury modelu 1 i dokonanie ponownej estymacji, otrzymując kolejny model 2.

Tabela 1.7.16

Model 2 wraz ze statystyką opisową

Zmienna	Współczynnik	Błąd stand.	Statystyka t	Wartość p	
const	-0,837147	0,256108	-3,2687	0,00216	***
regon	-0,00106249	0,000518285	-2,0500	0,04664	**
zatr_przem	-0,0911435	0,064058	-1,4228	0,16217	
sklep_zat	-0,0462816	0,0318099	-1,4549	0,15312	
miesz_k_new	-0,00578826	0,00505306	-1,1455	0,25849	
lud_prod	1,97947	0,433837	4,5627	0,00004	***
noclegi	0,0130892	0,00481956	2,7158	0,00956	***
lud_250 tys	-0,0241506	0,0342403	-0,7053	0,48450	

Średnia arytmetyczna zmiennej zależnej = 0,195567

Odchylenie standardowe zmiennej zależnej = 0,061855

Suma kwadratów reszt = 0,0917513

Błąd standardowy reszt = 0,0467392

Współczynnik determinacji (R^2) = 0,510598

Skorygowany współczynnik determinacji (\bar{R}^2) = 0,429031

Statystyka $F(7, 42) = 6,25986$ (wartość $p = 4,69e-005$)

Wyłączając stałą, największa wartość p jest dla zmiennej **lud_250 tys**

Źródło: Opracowanie własne w GRETL.

– postępując analogicznie jak w modelu 1, eliminujemy teraz zmienną **lud_250 tys** i ponownie szacujemy parametry. Program podpowiada nam w kolejnych iteracjach modeli wyeliminowanie zmiennych objaśniających:

model3 – miesz_new,

model4 – zatr_przem,

model5 – sklep_zatr.

Model 6 zawiera wszystkie zmienne łącznie ze stałą oznaczone trzema gwiazdkami. Tak więc parametry tego modelu, a przez nie zmienne objaśniające, są statystycznie istotne nawet na poziomie istotności 1% i nadają się do praktycznego wykorzystania, np. do celów prognozowania.

Tabela 1.7.17

Model 6 wraz ze statystyką opisową

Zmienna	Współczynnik	Błąd stand.	Statystyka t	Wartość p	
const	-0,945176	0,251316	-3,7609	0,00048	***
regon	-0,00129131	0,000474643	-2,7206	0,00917	***
lud_prod	1,9008	0,425549	4,4667	0,00005	***
noclegi	0,0196761	0,00373272	5,2712	<0,00001	***

Średnia arytmetyczna zmiennej zależnej = 0,195567

Odchylenie standardowe zmiennej zależnej = 0,061855

Suma kwadratów reszt = 0,103055

Błąd standardowy reszt = 0,0473321

Współczynnik determinacji (R^2) = 0,450304

Skorygowany współczynnik determinacji (\bar{R}^2) = 0,414454

Statystyka $F(3, 46) = 12,5609$ (wartość $p < 0,00001$)

Źródło: Opracowanie własne w GRETL.

Warto zauważyć, że w efekcie naszej analizy parametrów modelu uzyskaliśmy w programie GRETL postać modelu identyczną jak wcześniej w Excelu.

$$\hat{Y} = -0,945176 - 0,00129131 \cdot (\text{regon}) + 1,9008 \cdot (\text{lud_prod}) + 0,0196761 \cdot (\text{noclegi}).$$

b) Całościowa ocena przydatności modelu

Stosujemy statystykę F , na którą wzór (1.1.6) możemy zapisać również w postaci:

$$F_{obl} = \frac{R^2 / m}{(1 - R^2) / (n - m - 1)},$$

gdzie:

R^2 – współczynnik determinacji,

m – ilość zmiennych objaśniających.

Hipoteza zerowa zakłada nieistotność statystyczną każdego z parametrów modelu:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0.$$

Do naszego modelu 6 przyjmijmy poziom istotności $\alpha = 0,05$, $m = 3$, $n - m - 1 = 50 - 3 - 1 = 46$. Statystyka Fishera, czyli F_{obl} podana w tabeli 1.7.17, przy bardzo małym prawdopodobieństwie błędu p wynosi:

$$F(3;46) = 12,5609.$$

Tak więc kierując się tylko p , uznajemy wszystkie zmienne modelu 6 jako istotne statystycznie. Innym sposobem jest porównanie F_{obl} z wartością krytyczną wyznaczoną z tablicy V wartości krytycznych rozkładu $F(\alpha, r_1, r_2)$, która wynosi $F^* = 2,81$. Ponieważ $F_{obl} > F^*$, to odrzucamy hipotezę zerową o nieistotności całego zestawu parametrów modelu.

c) Ocena stopnia dopasowania modelu

Na podstawie statystyki opisowej dołączonej do modelu 6 (tabela 1.7.17) możemy wstępnie określić dopasowanie modelu do danych rzeczywistych, stosując miary:

– błąd standardowy reszt $S_e = 0,0473321$, średnia arytmetyczna zmiennej zależnej $\bar{y} = 0,195567$,

– współczynnik zmienności resztowej, przy czym 10% jest przyjętą wartością graniczną:

$$V_e = \frac{S_e}{\bar{y}} \cdot 100 = \frac{0,0473321}{0,195567} \cdot 100 = 24,2\% > V_e^* = 10\%,$$

– współczynnik determinacji $R^2 = 0,450304$ oraz pomocniczo współczynnik zbieżności, czyli stopień niewyjaśnienia zmienności Y przez oszacowany model 6:

$$\Phi^2 = 1 - R^2 = 1 - 0,450304 = 0,549696,$$

– skorygowany (dopasowany) współczynnik determinacji $R^2 = \bar{R}^2 = 0,414454$ stosujemy w modelach o różnej liczbie obserwacji i oszacowanych parametrach; porównajmy zatem model 1 i model 6 odpowiednio o ośmiu oraz trzech zmiennych objaśniających:

$R_1^2 = 0,420292$, $R_6^2 = 0,414454$; nastąpiło nieznaczne zmniejszenie, choć model 6 ma właściwie dobrane zmienne objaśniające.

Model 6 o 50 obserwacjach ma \bar{R}^2 równy 0,450304. Natomiast model 1 z wyeliminowanymi pięcioma obserwacjami wpływowymi, dla którego $R^2 = 0,424877$ (tab. 1.7.14), obliczymy następująco:

$$\bar{R}^2 = 1 - \frac{n-1}{n-m-1} (1 - R^2) = 1 - \frac{45-1}{45-3-1} (1 - 0,424877) = 0,382794.$$

W miarę eliminacji zmiennych nieistotnych oraz obserwacji wpływowych nastąpił nieznaczny spadek skorygowanego współczynnika determinacji. W tej sytuacji warto się zastanowić nad pobraniem ze zbioru 380-elementowego „stopy bezrobocia w powiatach” liczniejszej próby, np. 100 obserwacji. Pozostawia się jednak to badanie Czytelnikowi.

d) Ocena normalności składnika losowego (reszt)

Powróćmy do modelu 6 i zastosujmy w celu użycia programu GRETL testu Jarque'a-Bery (*JBT*) do oceny normalności reszt. Hipoteza zerowa zakłada normalność składnika resztowego. Statystykę *JBT*, która jest analogiczna do wzoru 1.1.15, lecz pomija *k*, określana jest według wzoru:

$$JBT = n \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right],$$

gdzie:

S – skośność,

K – kurtoza.

Test oparty jest na rozkładzie χ^2 z dwoma stopniami swobody. Możemy też skorzystać z menu (*Testy/Test normalności rozkładu reszt*) w oknie oszacowanego modelu 6. Pojawi się wykres słupkowy testu normalności rozkładu. Kliknięcie na wykresie powoduje pojawienie się menu, z którego wybieramy (*Kopiuj do schowka*), aby potem wprowadzić do dokumentu Worda.

Tabela 1.7.18

Częstość występowania reszt dla obserwacji modelu 6

Przedział	Średnia	Liczba	Częstość [%]	Skumulowana [%]
$\leq 0,097517$	-0,1154	1	2	2
-0,097517 – -0,061742	-0,07963	3	6	8
-0,061742 – -0,025967	-0,043855	10	20	28
-0,025967 – -0,0098074	-0,00808	18	36	64
0,0098074 – 0,045582	0,027695	9	18	82
0,045582 – 0,081357	0,06347	7	14	96
$\geq 0,081357$	0,099244	2	4	100

Test Jarqu'a-Bery – H_0 : dystrybuanta empiryczna posiada rozkład normalny: $\chi^2_{(2)} = 0,45$ z wartością $p = 0,79856$.

Źródło: Opracowanie własne w GRETL.

W końcowej uwadze program orzeka o normalności rozkładu reszt. Potwierdzeniem jest duże prawdopodobieństwo $p = 0,798556$, co pozwala nam przyjąć hipotezę zerową⁷³. Ponadto wartość krytyczna statystyki $\chi^2_{(2)}$ dla $\alpha = 0,05$ i dwóch stopni swobody według tablicy II w *Aneksie* wynosi 5,991. Określona w GRETL-u równa się 0,45, tak więc:

⁷³ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria...*, op.cit, rozdz. 1.91.

0,450 < 5,991 i nie ma podstaw do odrzucenia hipotezy zerowej.

e) Ocena jednorodności wariancji składnika resztowego (heteroskedastyczności)

Program GRETL do oceny jednorodności wariancji składnika losowego proponuje test White'a⁷⁴. W modelu 6 o 50 obserwacjach i równaniu podstawowym:

$$\hat{Y} = -0,945176 - 0,00129131 \cdot (\text{region}) + 1,9008 \cdot (\text{lud_prod}) + 0,0196761 \cdot (\text{noclegi})$$

test White'a przyjmuje sprawdzenie istotności regresji wyznaczonej dla $m = 3$ zmiennych objaśniających modelu, ich kwadratami oraz iloczynami⁷⁵.

Powstanie równanie pomocnicze wariancji reszt w postaci ogólnej 9 zmiennych:

$$\begin{aligned} \sigma_i^2 = & \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i1}^2 + \\ & + \alpha_5 x_{i1} x_{i2} + \alpha_6 x_{i1} x_{i3} + \alpha_7 x_{i2}^2 + \alpha_8 x_{i2} x_{i3} + \alpha_9 x_{i3}^2 + v_i \end{aligned}$$

a hipoteza zerowa: parametry modelu pomocniczego są równe zero (wariancja składnika resztowego modelu podstawowego jest jednorodna).

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_9 = 0$$

ma statystykę $T \cdot R^2$ o rozkładzie χ^2 z $J = 9$ stopniami swobody, przy czym:

T – liczba obserwacji,

R^2 – współczynnik determinacji dla modelu pomocniczego,

J – ilość zmiennych w równaniu pomocniczym wariancji reszt.

Hipoteza alternatywna:

$$H_1 := \alpha_1 = \alpha_2 = \dots = \alpha_9 \neq 0,$$

co najmniej jeden parametr modelu pomocniczego jest różny od zera – wariancja składnika resztowego modelu podstawowego jest niejednorodna⁷⁶.

W przeprowadzeniu testu korzystamy z menu (*Testy/Test heteroskedastyczności*) z okna estymacji modelu 6 po estymacji KMNK. W wynikach zamiast przecinka występuje kropka dziesiętna. Obraz testu White'a na heteroskedastyczność reszt (zmienność wariancji resztowej) po wprowadzeniu w formacie *rtf* do Worda podano w tab. 1.7.19.

⁷⁴ M. Osińska (red.), *Ekonometria współczesna...*, op.cit., rozdz. 4.3.4: *Badanie jednorodności wariancji składnika losowego*.

⁷⁵ T. K u f e l, *Ekonometria. Rozwiązywanie problemów...*, op.cit., rozdz. 4.3.4.

⁷⁶ T. K u f e l, *Ekonometria. Rozwiązywanie problemów...*, op.cit.; M. Osińska (red.), *Ekonometria współczesna*, op.cit.

Tabela 1.7.19

Parametry modelu pomocniczego ze zmienną zależną kwadratów reszt

Zmienna	Współczynnik	Błąd standardowy	Statystyka t	Wartość p
const	-0,232181	0,533736	-0,435	0,66589
regon	-0,00110049	0,00159292	-0,691	0,49364
lud_prod	0,893611	1,78239	0,501	0,61887
noclegi	0,00681558	0,0134336	0,507	0,614
sq_regon	-1,94509E-06	1,84918E-06	-1,052	0,29917
regon_lud_pr	0,00230064	0,00285027	0,807	0,42435
regon_nocleg	5,57372E-06	2,06490E-05	0,270	0,78860
sq_lud_prod	-0,868303	1,50452	-0,577	0,56709
lud_pr_nocleg	-0,0114819	0,0221247	-0,519	0,60665
sq_noclegi	-1,55966E-05	0,000122828	-0,127	0,89959

Źródło: Opracowanie własne w GRETL.

gdzie:

 sq – kwadrat danej zmiennej, a np. **regon_lud_prod** iloczyn zmiennych.Współczynnik determinacji $R^2 = 0,139724$, statystyka testu: $T \cdot R^2 = 6,986206$, z wartością $p = \chi^2(9) = 0,638556$.Prawdopodobieństwo popełnienia błędu (wartość $p = 0,638556$) wskazuje, że nie ma podstaw do odrzucenia hipotezy H_0 , gdyż:

$$T \cdot R^2 = 50 \cdot 0,139724 = 6,9862 < \chi^2_{(0,05;9)} = 16,919.$$

Według tablicy II na poziomie istotności $\alpha = 0,05$ oraz przy 9 stopniach swobody wartość krytyczna $\chi^2_{(0,05;9)} = 16,919$. Występuje zatem jednorodność wariancji i wszystkie nietypowe obserwacje poprawnie opisuje model 6.GRETL ma wbudowane tablice statystyczne. W tym przypadku do reszt możemy stosować menu (*Narzędzia/Tablice statystyczne*), aby otrzymać wartość krytyczną (np. do określenia wartości krytycznej $\alpha = 0,05$ oraz $J = 9$). Program posiada również tablice rozkładu normalnego, t -Studenta, F -Fishera oraz Durбина-Watsona.

f) Ocena liniowości

Wykorzystanie liniowej analitycznej postaci przykładowego modelu 6 może stanowić o jego słabości. Test nieliniowości White'a sprawdza do postaci logarytmicznej zasadność zbudowanego modelu liniowego względem potęgowego o zmiennych zlogarytmowanych. Hipoteza zerowa H_0 : zależność liniowa:

$$H_0 = \gamma_1 = \gamma_2 = \dots = \gamma_j = 0,$$

natomiast hipoteza alternatywna H_1 : nieliniowa.

Równanie pomocnicze odchylenia standardowego wobec równania podstawowego modelu 6 w postaci ogólnej jest następujące⁷⁷:

$$\sigma_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \gamma_1 \ln x_{i1} + \gamma_2 \ln x_{i2} + \gamma_3 \ln x_{i3} + \nu_i.$$

W programie GRETL korzystamy z menu (*Testy/Test nieliniowości (logarytmy)*) na poziomie oszacowania KMNK modelu 6. Program szacuje według metody KMNK parametry pomocniczego równania regresji z wykorzystaniem 50 obserwacji zmiennej zależnej reszty. Zastosowany jest tu test White'a dla nieliniowości z użyciem logarytmów trzech zmiennych (l_regon, l_lud_pr, l_nocleg). Wyniki GRETL-a prezentowane przez ten program w formie raportu podaje tab. 1.7.20.

Tabela 1.7.20

Parametry modelu pomocniczego ze zmienną zależną logarytmów zmiennych reszt

Zmienna	Współczynnik	Błąd standardowy	Statystyka t	Wartość p
const	21,8423	8,81987	2,476	0,01728
regon	0,000668302	0,00188353	0,355	0,72446
lud_prod	-23,8385	9,60029	-2,483	0,01700
noclegi	-0,00183896	0,00664546	-0,277	0,78332
l_regon	-0,0442363	0,145723	-0,304	0,76292
l_lud_pr	14,4515	5,80043	2,491	0,01666
l_nocleg	0,00260413	0,0201328	0,129	0,89769

Współczynnik determinacji $R^2 = 0,161275$, statystyka testu $T \cdot R^2$ przy trzech stopniach swobody dla logarytmów zmiennych z wartością $p = 0,0447127$.

Źródło: Opracowanie własne w GRETL.

Na poziomie istotności np. $\alpha = 0,05$ i trzech stopniach swobody wartość krytyczna $\chi^2_{(0,05;3)}$ odczytana z tablicy statystycznej χ^2 w GRETL wynosi 7,815.

Mała wartość prawdopodobieństwa – mniejsza niż $\alpha = 0,05$ – wskazuje, że należy odrzucić hipotezę zerową (postać liniowa). Obliczmy jednak statystykę:

⁷⁷ T. K u f e l, *Ekonometria. Rozwiązywanie problemów...*, op.cit., rozdz. 4.3.5.

$$T \cdot R^2 = 50 \cdot 0,161275 = 8,06375 < \chi^2_{(0,05;3)} = 7,815.$$

Tak więc mamy podstawę do odrzucenia postaci liniowej, a zastosowania postaci potęgowej modelu ze zmiennymi zlogarytmowanymi⁷⁸.

Na poziomie studiujących ekonometrię zastosowanie mikrokomputera do weryfikacji modeli ekonometrycznych może być znacznie szersze niż zaprezentowane w niniejszym rozdziale. W tym materiale zwrócono jednak szczególną uwagę na zastosowanie arkusza kalkulacyjnego Excel do określenia macierzy rzutowania jako sprawnego narzędzia do wyłonienia obserwacji wpływowych. Godny uwagi jest również sposób usuwania z modelu liniowego nieistotnych zmiennych objaśniających z zastosowaniem programu GRETL.

⁷⁸ Ibidem, s. 60.

1.8. Zależność między jedną zmienną objaśnianą a wieloma zmiennymi objaśniającymi

Jednorównaniowe modele zależności z wieloma zmiennymi objaśniającymi prezentują przykłady funkcji, których parametry oszacowano KMNK w programie Excel po wcześniejszej transformacji liniowej:

a) Funkcja produkcji:

$$\hat{Q} = 0,06 \cdot X_1^{0,3} \cdot X_2^{0,6},$$

gdzie:

Q – produkcja w tys. zł,

X_1 – majątek trwały w tys. zł,

X_2 – zatrudnienie w osobach.

b) Funkcja popytu na wyrób w , z uwzględnieniem ceny innego produktu z :

$$\hat{P}_w = 6 \cdot C_w^{-1,5} \cdot C_z \cdot R^2 \cdot D^{2,5},$$

gdzie:

C_w – cena jednostkowa produktu w ,

C_z – cena jednostkowa produktu z ,

R – wydatki na reklamę,

D – przeciętny dochód na mieszkańca danego regionu.

Do określenia postaci analitycznej funkcji liniowej wielu zmiennych objaśnianych przyjmijmy oznaczenia⁷⁹:

y_i ($i = 1, 2, \dots, n$) – obserwacje zmiennej objaśnianej,

x_{ij} – i -ta obserwacja zmiennej objaśniającej X_j ($j = 1, 2, \dots, m$),

wektor zmiennej \mathbf{y} :

⁷⁹ K. Jajuga (red.), *Ekonometria. Metody...*, op.cit., rozdz. 4.4.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_i \\ \dots \\ y_n \end{bmatrix},$$

macierz wartości zmiennych objaśniających:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nm} \end{bmatrix},$$

model ekonometryczny zależności Y od zmiennych objaśniających:

$$Y = f(X_1, X_2, \dots, X_m) + \varepsilon,$$

gdzie:

ε – czynnik losowy,

$f(X_1, X_2, \dots, X_m)$ – postać analityczna funkcji, którą, gdy jest liniowa, zapisujemy jako:

$$Y = \alpha_0 + \alpha_1 \cdot X_1 + \alpha_2 \cdot X_2 + \dots + \alpha_m \cdot X_m.$$

Wymienione parametry α_i podlegają oszacowaniu w trakcie budowy modelu ekonometrycznego.

1.8.1. Dobór modelu regresji poprzez obserwację wykresów korelacyjnych

Jeśli w wyniku doboru zmiennych objaśniających najlepszy okaże się model z jedną zmienną lub gdy zjawisko gospodarcze opisywane jest tylko jedną zmienną, to do sprawdzenia rodzaju regresji posługujemy się wykresem punktowym, zwanym też korelacyjnym, który możemy wykonać w Excelu. Jak już wspomniano w podrozdz. 1.4.1, wykres korelacyjny stosuje się również w sytuacji funkcji liniowej wielu zmiennych, prezentując kolejne pary zmiennych na wykresach.

Przykład

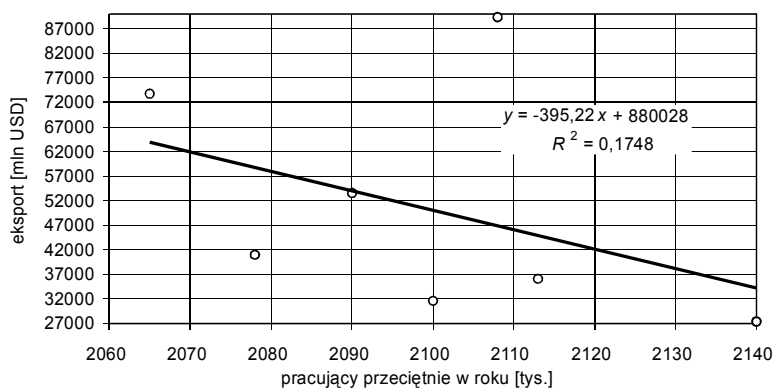
Szukamy zależności między eksportem Y [mln USD] a ilością pracujących w Polsce X [tys.]. Aby sformułować postać analityczną modelu regresji, rozpoczynamy od wykresu punktowego, dobierając do niego różne linie regresji.

Tabela 1.8.1

Dane statystyczne

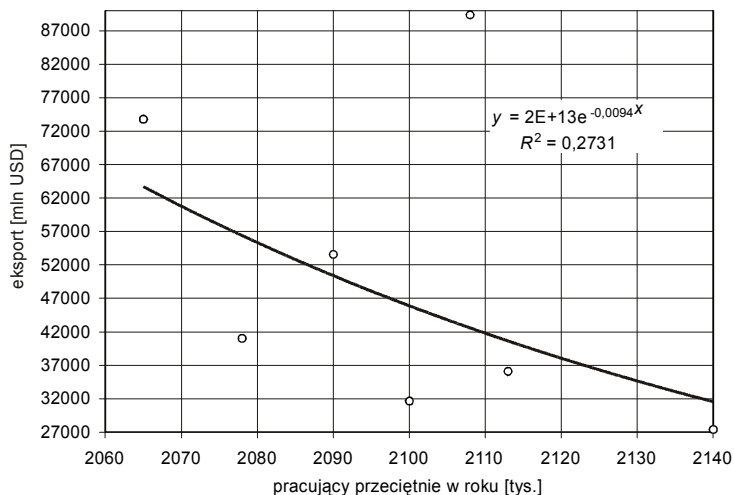
Y	27 407	31 651	36 092	41 010	53 577	73 781	89 378
X	2 140	2 100	2 113	2 078	2 090	2 065	2 108

Źródło: Opracowanie własne; „Rocznik Statystyczny Rzeczypospolitej Polskiej”, GUS, Warszawa 2006, s. 54.



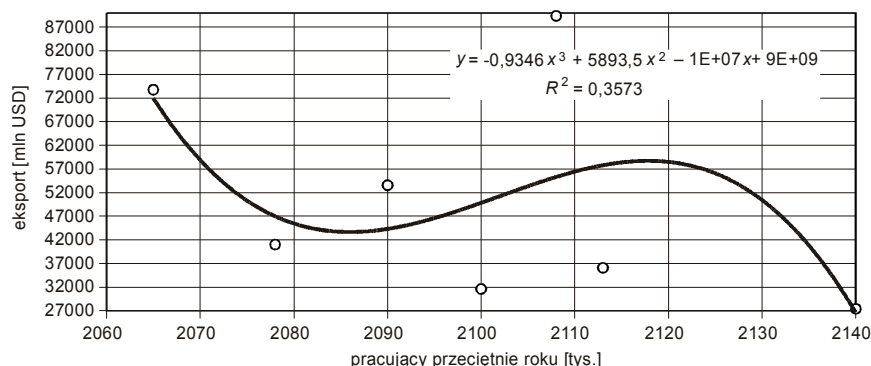
Źródło: Opracowanie własne w Excelu.

Wykres 1.8.1. Regresja liniowa. Eksport w Polsce w latach 1999–2005



Źródło: Opracowanie własne.

Wykres 1.8.2. Regresja wykładnicza. Eksport w Polsce w latach 1999–2005



Źródło: Opracowanie własne.

Wykres 1.8.3. Regresja wielomianowa. Eksport w Polsce w latach 1999–2005

Podano trzy przykłady doboru funkcji regresji. Najlepszy współczynnik dopasowania R^2 , chociaż wciąż niski, ma regresja wielomianowa. Z powodu braku danych odnośnie pracujących we wcześniejszych latach, musimy nasze badania nad modelem kontynuować w kolejnych latach, rozszerzając szeregi danych.

Przykład

Mamy wybrane dane statystyczne z tabeli *Ważniejsze dane o sytuacji społeczno-gospodarczej*⁸⁰. Przyjmijmy, że budujemy model zależności zmiennej objaśnianej Y od trzech zmiennych objaśniających, których potrzeba została wcześniej zweryfikowana:

gdzie:

Y – pozyskanie drewna (grubizny) w hm^3 ,

X_1 – powierzchnia lasów (stan w dniu 31 XII) w mln ha,

X_2 – lesistość (stan w dniu 31 XII) w % (stosunek powierzchni porośniętej lasami do ogólnej powierzchni terenu – stopień zalesienia),

X_3 – odnowienia i zalesienia w tys. ha.

Tabela 1.8.2

Dane statystyczne do modelu pozyskania drewna

Rok	1946	1950	1960	1970	1980	1990	1995	1999	2000	2001	2002	2003	2004	2005
X_1	6,5	6,9	7,7	8,4	8,6	8,7	8,8	8,9	8,9	8,9	8,9	8,9	9	9
X_2	20,8	22,2	24,6	27	27,6	27,8	28	28,3	28,4	28,4	28,5	28,6	28,7	28,8
X_3	42	193	172	115	95,3	66,8	77,8	63,3	68	65,4	56,8	66,1	61,7	62
Y	11,3	13,7	15,9	18,5	20,8	17,6	20,7	24,3	26	25	27,1	28,7	30,4	29,7

Źródło: Opracowanie własne według „Rocznika Statystycznego” 2006.

⁸⁰ Ibidem, s. 50.

Określamy model zależności liniowej $\hat{Y} = a_0 + a_1X_1 + a_2X_2 + a_3X_3$, korzystając z funkcji REGLINP.

a_3	a_2	a_1	a_0
-0,02	7,44	-17,7	-27,5

$$\hat{Y} = -27,5 - 17,7 \cdot X_1 + 7,44 \cdot X_2 - 0,02 \cdot X_3.$$

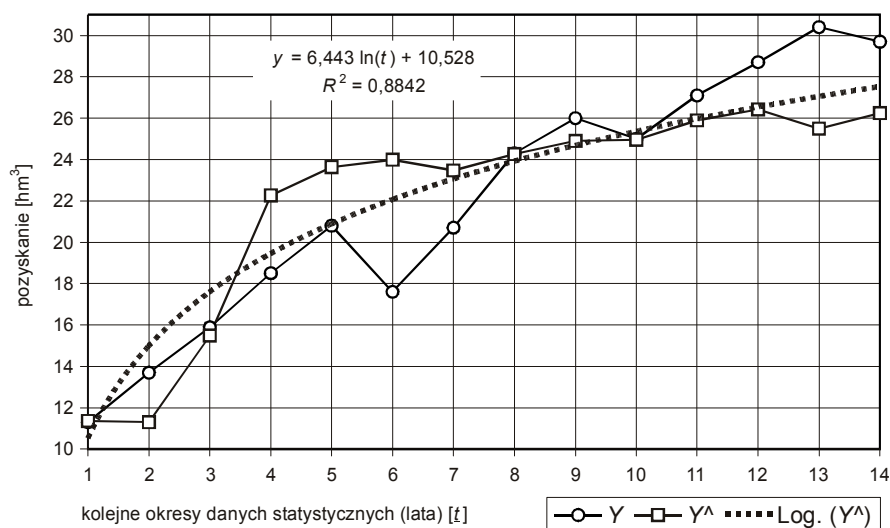
Obliczmy wartości modelowe z powyższej funkcji i przedstawmy to na wykresie.

Tabela 1.8.3

Obliczenie wartości modelowych

Y	11,3	13,7	15,9	18,5	20,8	17,6	20,7	24,3	26	25	27,1	28,7	30,4	29,7
\hat{Y}	11,4	11,3	15,5	22,3	23,6	24	23,5	24,3	24,9	25	25,9	26,4	25,5	26,2
n	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Źródło: Opracowanie własne w Excelu.



Źródło: Opracowanie własne w Excelu, gdzie: $Y^{\wedge} \rightarrow \hat{Y}$.

Wykres 1.8.4. Dobrany trend logarytmiczny do danych modelowych. Pozyskanie drewna (grubizny) w latach 1946–2005

Na podstawie rozrzutu punktów wartości modelowych w kolejnych latach obserwacji możemy korzystając z funkcji Excela i sprawdzając wzrokowo, dopasować adekwatny trend. W tym przykładzie jest nim trend logarytmiczny, który wykazuje dobre dopasowanie danych teoretycznych do empirycznych ($R^2 = 0,8842$).

1.8.2. Mierniki dopasowania modelu

W doborze trendu kierowaliśmy się współczynnikiem determinacji (R^2), który jest miarą dobroci dopasowania modelu do danych empirycznych. Temat mierników dopasowania znalazł już swoje odbicie we wzorach zamieszczonych w rozdziałach 1.1 oraz 1.7. Na potrzeby niniejszego przykładu przypomnijmy sobie te wzory:

– współczynnik determinacji (R^2); współczynnik ten w odniesieniu do modeli z wieloma zmiennymi objaśniającymi równy jest kwadratowi współczynnika korelacji wielorakiej (wzór 1.6.6),

Mając dane empiryczne y_i z obserwacji, określamy wartość średnią \bar{y} oraz obliczamy wartości teoretyczne z modelu \hat{y}_i . W tej sytuacji na określenie współczynnika determinacji posługujemy się wzorem (1.1.7):

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

– współczynnik zbieżności, określamy również jako ϕ^2 (wzór 1.7.3), który zapisać można jako:

$$\phi^2 = 1 - R^2,$$

– odchylenie standardowe reszt S_e , oznaczane również jako S_u (wzór 1.6.4), stanowiące pierwiastek z wariancji resztowej:

$$S_e = \sqrt{S_e^2},$$

$$S_e^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k},$$

gdzie:

$k = m + 1$,

m – ilość zmiennych objaśniających modelu.

Wariancję resztową w wersji macierzowej określamy według wzoru⁸¹:

$$S^2(e) = \frac{y^T y - a^T X^T y}{n - m - 1}, \quad (1.8.1)$$

– współczynnik wyrazistości W [%]; $W = (S_e / \bar{y}) \cdot 100$ (wzór 1.7.4).

⁸¹ J. Dziechciarz (red.), *Ekonometria. Metody...*, op.cit, s. 67, wzór (3.5).

Przykład

Obliczamy mierniki dopasowania: współczynnik determinacji (R^2), współczynnik zbieżności (Φ^2), odchylenie standardowe reszt (S_e), współczynnik wyrazistości (W) na podstawie danych w tab. 1.8.4.

Tabela 1.8.4

Dane modelowe oraz obliczenia pomocnicze

n	y_i	y_i^{\wedge}	$(y_i - \bar{y})^2$	$(y_i^{\wedge} - \bar{y})^2$	$u^2 = (y_i - y_i^{\wedge})^2$
1	11,3	11,4	117,1	115,7	0,004
2	13,7	11,3	70,9	117,0	5,733
3	15,9	15,5	38,7	43,9	0,165
4	18,5	22,3	13,1	0,0	14,140
5	20,8	23,6	1,7	2,3	8,026
6	17,6	24,0	20,4	3,5	40,906
7	20,7	23,5	2,0	1,8	7,663
8	24,3	24,3	4,7	4,6	0,002
9	26,0	24,9	15,0	7,7	1,216
10	25,0	25,0	8,3	8,0	0,002
11	27,1	25,9	24,8	14,2	1,455
12	28,7	26,4	43,3	18,5	5,162
13	30,4	25,5	68,5	11,4	23,974
14	29,7	26,2	57,4	17,0	11,966
Średnia	22,1		486,2	365,8	120,413

Źródło: Opracowanie własne w Excelu, gdzie: $y_i^{\wedge} \rightarrow \hat{y}_i$, $y^- \rightarrow \bar{y}$.

Współczynnik determinacji:

$$R^2 = \frac{365,8}{486,2} = 0,752.$$

Współczynnik determinacji w 75,2% wyjaśnia, w jakim stopniu oszacowany we wcześniejszym przykładzie model $\hat{Y} = -27,5 - 17,7 \cdot X_1 + 7,44 \cdot X_2 - 0,02 \cdot X_3$ jest dopasowany do danych z obserwacji.

Współczynnik zbieżności:

$$\Phi^2 = 1 - 0,752 = 0,248.$$

Współczynnik zbieżności w 24,8% informuje nas, w jakim stopniu zmienna Y nie jest objaśniana przez model.

Wariancja resztowa, odchylenie standardowe reszt oraz współczynnik wyrazistości:

$$S_u^2 = \frac{120,413}{14 - 4} = 12,0413, \quad S_u = \sqrt{12,0413} = 3,47,$$

$$W = \frac{3,47}{22,1} \cdot 100 = 15,7\%.$$

We wzorze na wariancję resztową: $n = 14$ obserwacji, $k = m + 1 = 3 + 1 = 4$, $m = 3$, zmienne objaśniające: X_1, X_2, X_3 .

$S_u = 3,47$ określa przeciętną różnicę między wartościami zmiennej objaśnianej Y a wartościami określonymi przez model Y .

Współczynnik wyrazistości (15,7%) informuje nas, jaki jest udział odchylenia standardowego zmiennej Y w stosunku do jej średniej.

1.8.3. Pytania i zadania

1. Podaj postać i funkcję współczynnika zmienności.
2. Napisz i wyjaśnij poszczególne elementy wzoru na obliczanie współczynnika korelacji między zmienną objaśnianą a zmienną objaśniającą.
3. Wymień kolejne etapy w analizie korelacyjnej.
4. Opisz procedurę doboru zmiennych objaśniających według metody pojemności nośników informacji.
5. Napisz analogicznie do modeli trendu modele regresji z jedną zmienną objaśniającą.
6. Przedstaw mierniki dopasowania modelu do danych empirycznych.
7. Do czego i w jaki sposób stosujemy test parametryczny Fishera?

1.8.4. Zadania

1. W Polsce spożycie masła [kg] w przeliczeniu na mieszkańca kształtowało się następująco:

Tabela 1.8.5

Dane statystyczne spożycia masła

Rok	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Spożycie	4,5	3,9	3,7	3,9	4,3	4,4	4,6	4,2	4,3	4,5	4,7	4,4	4,2

Źródło: Opracowanie własne według „Małych Roczników Statystycznych” z lat 1999–2006; tabele „Ważniejsze dane o sytuacji społeczno-gospodarczej kraju w przeliczeniu na mieszkańca”.

Określ współczynnik zmienności V_x , po wcześniejszym obliczeniu odchylenia standardowego S_x i wartości średniej \bar{x} zmiennej X (spożycie indywidualne masła).

2. Dany jest wektor i macierz współczynników korelacji między trzema zmiennymi X_1, X_2, X_3 :

$$\mathbf{R}_0 = \begin{bmatrix} 0,9 \\ 0,95 \\ 0,8 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0,8 & 0,7 \\ 0,8 & 1 & 0,9 \\ 0,7 & 0,9 & 1 \end{bmatrix}.$$

Określ: macierz rozszerzoną \mathbf{R}^* , wyznaczniki macierzy R^* oraz R , współczynnik korelacji wielorakiej R_k .

3. Określ współczynnik korelacji między zmiennymi Y – eksport w USD w przeliczeniu na mieszkańca oraz X – import w tych samych jednostkach również w przeliczeniu na USD, opierając się na danych tab. 1.8.6 (t – kolejne lata 1985–2005).

Tabela 1.8.6

Dane statystyczne dotyczące eksportu i importu

t	1	2	3	4	5	6	7	8	9	10	
X	291	299	288	323	271	250	406	415	490	559	
Y	309	322	324	369	355	376	390	344	368	447	
t	11	12	13	14	15	16	17	18	19	20	21
X	753	962	1 094	1 217	1 188	1 266	1 301	1 442	1 781	2 303	2 664
Y	593	633	666	730	709	819	934	1 073	1 403	1 933	2 342

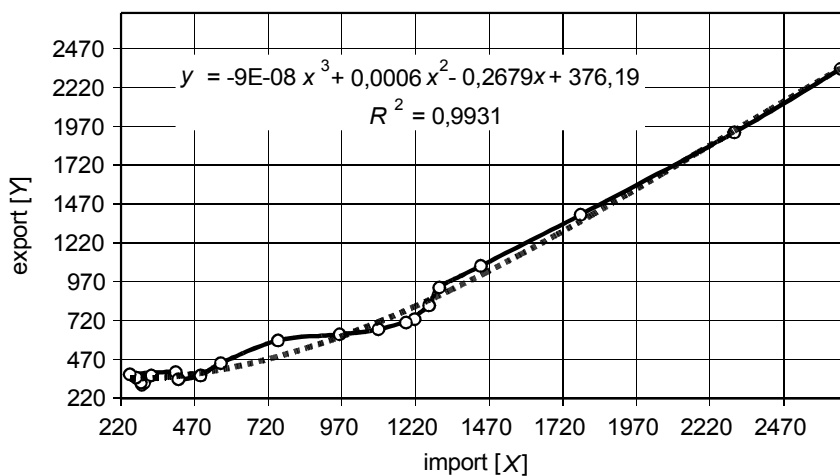
Źródło: „Małe Roczniki Statystyczne” z lat 1999–2005.

Oblicz współczynnik korelacji, korzystając z podanego wzoru (analogia do wzoru 1.6.12), a następnie alternatywnie zastosuj funkcję Excela [=WSP.KORELACJI(...)].

$$r_{yx} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

4. Spośród zestawów, jakie tworzą kombinacje zmiennych X_1 , X_2 oraz X_3 , wybierz zestaw, w którym integralna pojemność informacji $H_I = H_{\max}$, stosując metodę pojemności nośników informacji. Przyjmij jako dane wektor \mathbf{R}_0 oraz macierz współczynników korelacji \mathbf{R} z zadania drugiego.

5. Sprawdź w Excelu, dobierając rodzaje linii regresji, czy zamieszczony model funkcji wielomianowej trzeciego stopnia dla zależności zmiennej Y (eksport) od X (import) do danych w zadaniu trzecim jest najlepszy. Jako kryterium przyjmij współczynnik determinacji R^2 .



Źródło: Opracowanie własne według danych zadania trzeciego.

Wykres 1.8.5. Zależność eksportu od importu

6. Oblicz współczynnik wyrazistości modelu liniowego o trzech zmiennych objaśniających do danych stanowiących fragment tabeli 1.8.4 w zakresie 10 obserwacji y_1 oraz wartości modelowej \hat{y}_i .

Tabela 1.8.7

Dane umowne do zadania szóstego

n	y_i	\hat{y}_i
1	11,3	11,4
2	13,7	11,3
3	15,9	15,5
4	18,5	22,3
5	20,8	23,6
6	17,6	24
7	20,7	23,5
8	24,3	24,3
9	26	24,9
10	25	25

Źródło: Opracowanie własne.

7. Zastosuj funkcję REGLINP ze statystyką opisową Excela do określenia modelu liniowego trzech zmiennych $\hat{Y} = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_3$. Na-

stepnie zbadaj istotność całego zestawu zmiennych objaśniających jak też poszczególne parametry modelu.

T a b e l a 1.8.8

Dane umowne do zadania siódmego

i	Y	X_1	X_2	X_3
1	12	8	12	10
2	10	6	14	12
3	18	11	16	11
4	16	10	18	14
5	12	8	18	10
6	14	10	18	12
7	18	11	20	13
8	20	12	24	12
9	20	12	26	12
10	22	14	26	10

Ź r ó d ł o: Opracowanie własne.

1.9. Zagadnienie modeli wielorównaniowych z zastosowaniem programu GRETL

1.9.1. Model wielorównaniowy

Istnieje wiele modeli ekonometrycznych, a do szacowania ich parametrów literatura podaje różne sposoby. Istnieją też różnorodne programy komputerowe wspomagające rozwiązywanie modeli jedno- i wielorównaniowych. Oprócz tradycyjnej już funkcjonalności Excela w zakresie rachunku macierzowego do celów dydaktycznych proponuje się korzystanie z programu GRETL. Obejmuje on liczne modele ekonometryczne oraz można skorzystać z komputerowej estymacji ich parametrów różnymi sposobami. W niniejszym rozdziale wymieniono te sposoby, których teoretyczne podstawy spotykamy w pracy G.S. Maddali⁸². Ponadto zaprezentowano przykładowo podwójną metodę najmniejszych kwadratów do modeli wielorównaniowych, opierając się na rozważaniu teoretycznym zawartym w publikacji *Ekonometria. Metody, przykłady, zadania*⁸³ oraz pracy T. Kufla⁸⁴. Przetestowano tę metodę na programie z internetu, w wyniku czego otrzymano procedurę realizacji, jak się wydaje bardzo pomocną dla studiujących ekonometrię. Warto zaznaczyć, że program GRETL wersja 1.6.5 jest oprogramowaniem bezpłatnym należącym do grupy Open Source.

Model strukturalny wielorównaniowy (bez zmiennych opóźnionych w czasie) w postaci macierzowej jest następujący:

$$\mathbf{BY} + \mathbf{CX} = \boldsymbol{\varepsilon},$$

gdzie:

B – macierz nieosobliwa parametrów przy zmiennych endogenicznych,

C – macierz parametrów przy zmiennych egzogenicznych,

Y – wektor kolumnowy zmiennych objaśnianych,

X – wektor kolumnowy zmiennych objaśniających,

$\boldsymbol{\varepsilon}$ – wektor kolumnowy składników losowych.

⁸² G.S. Maddala, *Ekonometria...*, op.cit.

⁸³ J. Dziechciarz (red.), *Ekonometria. Metody...*, op.cit.

⁸⁴ T. Kufla, *Ekonometria. Rozwiązywanie...*, op.cit.

Macierzą nieosobliwą nazywamy macierz kwadratową, której wyznacznik jest różny od zera⁸⁵.

$$\mathbf{B} = \begin{bmatrix} 1 & -b_{12} & \dots & -b_{1m} \\ -b_{21} & 1 & \dots & -b_{2m} \\ \dots & \dots & \dots & \dots \\ -b_{m1} & -b_{m2} & \dots & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} -c_{11} & -c_{12} & \dots & -c_{1k} \\ -c_{21} & -c_{22} & \dots & -c_{2k} \\ \dots & \dots & \dots & \dots \\ -c_{m1} & -c_{m2} & \dots & -c_{mk} \end{bmatrix}, \quad \mathbf{Y} = \mathbf{P}\mathbf{X} + \boldsymbol{\eta}.$$

Po pomnożeniu równania $\mathbf{B}\mathbf{Y} + \mathbf{C}\mathbf{X} = \boldsymbol{\varepsilon}$ przez macierz odwrotną \mathbf{B}^{-1} otrzymamy postać zredukowaną modelu $\mathbf{Y} = \mathbf{P}\mathbf{X} + \boldsymbol{\eta}$, w którym Y zależy tylko od zmiennych egzogenicznych X :

gdzie:

$\mathbf{P} = -\mathbf{B}^{-1}\mathbf{C}$ – parametry postaci zredukowanej,

$\boldsymbol{\eta} = \mathbf{B}^{-1}\boldsymbol{\varepsilon}$ – składnik losowy postaci zredukowanej.

1.9.2. Model prosty

W tym modelu postać zredukowana jest identyczna ze strukturalną. Oznaczając przez \mathbf{I}_m macierz jednostkową zmiennych objaśniających, otrzymamy:

$$\mathbf{B} = \mathbf{I}_m.$$

W zbiorze modeli jednorównaniowych w żadnym z nich nie występuje zmienna endogeniczna. Parametry w poszczególnych równaniach możemy określić niezależnie od siebie klasyczną metodą najmniejszych kwadratów, zawartą w programie GRET. Alternatywnym rozwiązaniem jest zastosowanie funkcji Excela ([TRANSPONUJ]; [ILOCZYN.MACIERZ]; [MACIERZ.ODW]) do oszacowania parametrów równań modelu prostego z zależności określonej wzorem (1.7.1):

$$a = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{y}^T,$$

gdzie:

\mathbf{X} – macierz zmiennych objaśniających wraz ze zmienną tożsamościowo równą 1, odpowiadającą stałej w danym równaniu,

\mathbf{X}^T – ta sama macierz, lecz transponowana,

\mathbf{y} – wektor określonej zmiennej objaśniającej.

⁸⁵ *Leksykon naukowo-techniczny*, WNT, Warszawa 1984.

1.9.3. Model rekurencyjny

Niech macierz \mathbf{B} będzie macierzą trójkątną z zerami nad główną przekątną mającą jedynki lub dającą się do takiej postaci sprowadzić. Poszczególne kolumny odpowiadają równaniom zmiennych w modelu wielorównaniowym. Przyjmijmy, że w pierwszym równaniu nie występują zmienne endogeniczne jako zmienne objaśniające, tak więc parametry tego równania można oszacować KMNK. W kolejnych równaniach mogą się znajdować zmienne endogeniczne pełniące funkcje zmiennych objaśniających. Estymację parametrów modelu rekurencyjnego przeprowadzamy według procedury:

- obliczenie wartości teoretycznych \hat{y}_{i1} pierwszego równania poszczególnych obserwacji i , po wcześniejszym oszacowaniu parametrów KMNK, o czym już powiedziano,
- zastąpienie w macierzach obserwacji zmiennych wartości empirycznych równania drugiego wartościami obliczonymi – teoretycznymi zmiennej objaśnianej,
- oszacowanie parametrów równania drugiego również KMNK,
- obliczenie wartości teoretycznych \hat{y}_{i2} kolejnych obserwacji zmiennej Y_2 ,
- postępowanie analogiczne z kolejnymi równaniami modelu rekurencyjnego.

1.9.4. Przykład modelu rekurencyjnego

Jak już wspomniano program GRETL obejmuje m.in. metody ekonometryczne i sposoby estymacji parametrów równania (równań) określonego modelu. Polską wersję instalacyjną programu można pozyskać z internetu pod adresem [www.kufel.torun.pl]. Program pomyślany jest jako pomoc na zajęciach z ekonometrii⁸⁶. W konfrontacji obliczeń manualnych z szybką realizacją komputerową zabazowano na danych w publikacji J. Dziechciarza⁸⁷.

Tabela 1.9.1, stanowiąca przeredagowany w Excelu wydruk ze zbioru GRETL-a (z kropką dziesiętną zamiast przecinka), zawiera obserwacje zmiennych i obliczone wartości według równań modelu.

Naszym zadaniem jest oszacowanie parametrów modelu trójrównaniowego, przy czym zmienne endogeniczne to X_1 , X_2 , X_3 , z których X_1 oraz X_2 stanowią jednocześnie zmienne objaśniające. Po przeniesieniu zmiennych endogenicznych na lewą stronę otrzymamy:

⁸⁶ T. K u f e l, *Ekonometria. Rozwiązywanie...*, op.cit. (*Wstęp do wydania drugiego*).

⁸⁷ J. Dziechciarz (red.), *Ekonometria. Metody...*, op.cit., przykład 8.1.

$$\begin{aligned} \hat{x}_1 &= a_1x_2 + a_2x_4 + a_0, & 1\hat{x}_1 - a_1\hat{x}_2 + 0\hat{x}_3 &= a_2x_4 + a_0, \\ \hat{x}_2 &= b_1x_4 + b_2x_7 + b_0, & 0\hat{x}_1 + 1\hat{x}_2 + 0\hat{x}_3 &= b_1x_4 + b_2x_7 + b_0, \\ \hat{x}_3 &= c_1x_1 + c_2x_5 + c_0, & -c_1\hat{x}_1 + 0\hat{x}_2 + 1\hat{x}_3 &= c_2x_5 + c_0, \end{aligned} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ -a_1 & 1 & 0 \\ 0 & -c_1 & 1 \end{bmatrix}.$$

Tabela 1.9.1

Obserwacje zmiennych „ $X_1 - X_9$ ” oraz obliczone wartości „ X_1T ”, „ X_2T ”, „ X_3T ” według równań modelu

Obser	Rok	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_2T	X_1T	X_3T
1	1970	112.2	764	8621	331	537	5389	4215	137.6	3263	858.359673	112.288	9097.0382
2	1971	131.1	839	8666	348	578	5186	4709	140.4	3414	878.660583	133.444	9169.0946
3	1972	137.6	899	8725	364	609	4978	5200	136.1	3487	898.770352	149.152	9177.9566
4	1973	135.8	940	8852	398	639	4779	5687	140.7	3610	920.088181	148.191	9139.1558
5	1974	139.5	969	8950	427	649	4306	6100	142.7	3717	938.179336	145.001	9151.4030
6	1975	143.9	981	8974	468	622	4448	6472	140.8	3848	955.600376	128.978	9202.6034
7	1976	159.5	1046	9060	491	582	4327	6820	144.2	3929	970.735608	143.122	9331.6370
8	1977	149.7	1081	9128	491	538	4257	7170	131.6	4075	984.207458	157.017	9315.6962
9	1978	143.7	1070	9199	485	491	4217	7474	116.1	4179	995.454714	155.692	9324.8342
10	1979	151.2	1010	9252	469	450	4217	7708	107.9	4224	1003.250920	139.984	9407.3762
11	1980	147.1	1033	9315	454	415	4260	7954	97.5	4356	1011.584686	156.720	9416.3366
12	1981	133.9	909	9355	426	393	4342	8188	100.1	4136	1018.472876	121.688	9359.7614
13	1982	178.1	965	9442	397	381	4465	8347	89.4	4344	1022.398573	158.623	9629.9006
14	1983	194.9	1026	9572	370	373	4327	8542	108.1	4495	1027.861282	196.529	9735.8174
15	1984	229.8	1042	9728	350	375	4771	8765	127.6	4588	1034.931415	213.021	9938.3078
16	1985	246.3	1071	9899	341	383	4928	9468	107.1	4571	1061.309576	229.097	10027.3958
17	1986	249.4	1100	10000	334	400	5058	9692	94.3	4605	1069.401884	244.159	10029.5354
18	1987	267.6	1158	10129	343	422	5165	9868	95.3	4611	1076.857312	262.622	10115.3906
19	1988	245.3	1220	10248	356	444	5213	10031	96.4	4622	1084.115029	280.645	9964.0778
20	1989	215.6	1164	10313	378	463	5257	10055	69.6	4520	1086.703509	247.259	9772.2566
21	1990	175.6	924	10269	404	494	5301	9919	32.8	4597	1083.436101	138.797	9508.8146
22	1991	125.5	946	9936	428	549	5323	9809	20.9	4680	1081.018123	135.363	9163.6190
23	1992	125.8	1031	9770	496	607	5324	10043	13.3	4742	1095.170441	134.632	9110.7398
24	1993	102.5	1070	9605	584	660	5288	10111	14.9	4934	1104.446613	105.499	8924.3690

X_1 – liczba wydanych książek i broszur (mln. egz.), X_7 – liczba abonentów telewizyjnych (mln),
 X_2 – produkcja papieru (tys. t.), X_8 – liczba widzów w kinach (mln),
 X_3 – liczba bibliotek publicznych, X_9 – liczba chorych w szpitalach.
 X_4 – liczba studentów (tys.),
 X_5 – liczba uczniów w szkołach średnich (tys.),
 X_6 – liczba uczniów w szkołach podstawowych (tys.),

Źródło: Opracowanie własne w GRETL bazujące na wartościach obserwacji zmiennych w publikacji: J. Dziechciarz (red.) *Ekonometria. Metody, przykłady, zadania*, Akademia Ekonomiczna, Wrocław 2003, tab. 8.1.

Macierz współczynników **B** przy zmiennych endogenicznych po doprowadzeniu do postaci macierzy trójkątnej dolnej (zamiana kolumny drugiej na pierwszą oraz wiersza pierwszego na drugi) wskazuje na istnienie modelu rekurencyjnego. Musimy zatem wpierw oszacować parametry równania drugiego, bazując na szeregach 24 obserwacji zmiennych X_4 , X_7 oraz X_0 (równiej 1) oraz wektorze X_2 stanowiącym zmienną objaśnianą.

$$\hat{x}_2 = b_1x_4 + b_2x_7 + b_0.$$

W tym miejscu skorzystamy z funkcji estymacji parametrów równania KMNK występującej w programie GRETl, przy czym na potrzeby języka skryptowego tego programu zmienne empiryczne zapisano symbolami: X1, X2, X3, X4, X5, X6, X7, X8, X9, a zmienne endogeniczne obliczone (teoretyczne) jako X1T, X2T, X3T.

Początkowa procedura postępowania jest następująca:

- zainstalowanie pełnej wersji polskiej programu GRETl,
- w menu głównym korzystamy z opcji (*Plik/Nowy zbiór danych*),
- mając dane dla kolejnych zmiennych, w oknie dialogowym wpisujemy liczbę obserwacji 24 (domyślnie 50), *OK*,
- w oknie (*Struktura bazy danych*) zaznaczamy opcję (*Dane przekrojowe*) i naciskamy przycisk (*Naprzód*),
- wpisujemy nazwę zmiennej (*Rok*) i wprowadzamy dane (lata 1970–1993), akceptując strzałką w dół,
- po wprowadzeniu pełnego szeregu 24 lat naciskamy tylko klawisz (*Zastosuj*),
- w menu danego podokna wybieramy (*Zmienna*) oraz opcję (*Dodaj*),
- wpisujemy nową zmienną X_1 , wprowadzamy dane, naciskamy (*Zastosuj*),
- postępujemy podobnie z dalszymi zmiennymi X_2 – X_9 , a na zakończenie zbioru danych naciskamy klawisz (*Zamknij*),
- w menu (*Plik*) wybieramy opcję (*Zapisz dane jako*) i zapisujemy kolejno w dwóch formatach:
 - standardowy plik danych GRETl-a, np. (*E4.gdt*),
 - bazy danych (*E4.bin*) z wprowadzeniem nazwy bazy np. (*Obserwacje*),
 - całą pracę z GRETl-em jako sesję zapisujemy w formie pliku (*E4.gretl*) w podkatalogu (*GRETl*) dysku (*D*).

Zaprezentowana poniżej struktura zbioru zapisana została na dysku jako plik tekstowy (*E4.txt*). Widzimy tu dodatkowo wprowadzone zmienne wygenerowane w postaci szeregów odpowiadających określonym obserwacjom zmiennych wchodzących do trzech równań rozpatrywanego modelu rekurencyjnego. Zmienne wprowadzamy w nawiasach, parametry zmiennych zapisujemy również w nawiasach z kropką dziesiętną. Całość poprzedzamy poleceniem (*genr*), np. dla *X2T* – co widać w strukturze zbioru danych przepisanych z ekranu GRETL-a:

```

genrX2T=(0.075668)*(X4)+(0.038491)*(X7)+(671.074)
# Obserwacje 1 – 24 n = 24
Rok
X1
X2
X3
X4
X5
X6
X7
X8
X9
genrX2T(0.075668)*(X4)+(0.038491)*(X7)+(671.074)
genrX1T(0.397263)*(genrX2T)-(0.507356)*(X4)-(23.2019)
genrX3T(5.83453)*(genrX1T)-(0.948584)*(X5)+(8953.55)

```

W procedurze oszacowania parametrów równań oraz wygenerowania wartości teoretycznych, mając zbiór zmiennych, przystępujemy do oszacowania parametrów równań, lecz najpierw drugiego, gdyż zawiera tylko zmienne objaśniające X_4 oraz X_7 .

$$\hat{x}_2 = b_1x_4 + b_2x_7 + b_0.$$

Z menu głównego wybieramy (*Model*) a następnie opcję (*Klasyczna metoda najmniejszych kwadratów*). Pojawia się okno dialogowe (*Specyfikacja modelu*). Wybieramy jako zmienną zależną X_2 (według symbolizacji składni GRETL-a), a jako zmienne niezależne: const (stała), X_4 , X_7 . Program natychmiast wyświetla dane z estymacji, podając dodatkowo statystykę opisową.

Wyniki zapisujemy jako plik w formacie *rtf* (MS Word) na dysku (*D:\GRETL\E4.rtf*).

Dla równania drugiego uzyskaliśmy model liniowy:

$$\hat{x}_2 = 0,075668x_4 + 0,038491x_7 + 671,074.$$

Generujemy szereg wartości teoretyczne dla \hat{x}_2 . W tym celu korzystamy z menu (*Zmienna/Definiowanie nowej zmiennej*) i wprowadzamy wcześniej już podaną formułę na wyliczenie zmiennej określonej jako X2T. Następnie szacujemy parametry równania pierwszego również KMNK w GRETL, wstawiając zamiast X2 wygenerowane wartości X2T:

$$\hat{x}_1 = 0,397263x_2 - 0,507356x_4 - 23,2019.$$

Program numeruje kolejne wygenerowane modele i podaje dodatkowo statystykę opisową. Przykładowo dla powyższego równania zmiennej zależnej X1 ze zmienną teoretyczną X2T, zmienną objaśniającą X4 oraz stałą dla zbioru danych o 24 obserwacjach otrzymamy wielkości jak w tab. 1.9.2.

T a b e l a 1.9.2

Model oraz statystyka opisowa zmiennej „X1” (według notacji w GRETL)

Zmienna	Współczynnik	Błąd standardowy	Statystyka <i>t</i>
const	-23,2019	84,0681	-0,2760
X2T	0,397263	0,080979	4,9058
X4	-0,507356	0,0880348	-5,7631

Średnia arytmetyczna zmiennej zależnej = 168,4,
 Odchylenie standardowe zmiennej zależnej = 48,5454,
 Suma kwadratów reszt = 16207,7,
 Błąd standardowy reszt = 27,7812,
 Współczynnik determinacji $R^2 = 0,700981$,
 Skorygowany współczynnik determinacji $\bar{R}^2 = 0,672503$.

Ź r ó d ł o: Opracowanie własne według danych z tab. 1.9.1.

Generujemy, posługując się symbolami notacji GRETL, szereg wartości dla X1, wstawiając wartości z szeregów X2T oraz X4 i szacujemy parametry równania trzeciego, wprowadzając zamiast X1 wartości szeregu dla X1T oraz wartości obserwacji zmiennej objaśniającej X5.

$$\hat{x}_3 = 5,83453x_1 - 0,948584x_5 + 8953,55.$$

Możemy dodatkowo (tab. 1.9.1) wygenerować również wartości szeregu zmiennej X3T (oznaczonej w GRETL jako X3T).

1.9.5. Model o równaniach łącznie współzależnych

W tego typu modelach wśród zmiennych objaśniających występują takie, których nie można zastąpić wartościami teoretycznymi z równań wcześniej oszacowanych KMNK, np.:

$$Y_1 = a_1 Y_2 + a_2 X_1 + a_0, \quad Y_2 = b_1 Y_1 + b_2 X_1 + b_0.$$

Warunkiem estymacji jest identyfikowalność równań, czyli każde równanie musi być inne niż poprzednie. Stosujemy tu twierdzenie podane w publikacji J. Dziechciarza⁸⁸:

„Warunkiem identyfikowalności danego równania modelu m -równaniowego jest to, by macierz \mathbf{A}_i utworzona ze współczynników stojących przy zmiennych występujących w innych równaniach modelu, a niewystępujących w badanym (i -tym) równaniu, była rzędu $m-1$. Ponadto dane równanie jest identyfikowalne jednoznacznie, gdy liczba zmiennych występujących w modelu, a niewystępujących w badanym równaniu jest $m-1$ ”.

„Rzędem macierzy nazywamy największy stopień niezerowego minora tej macierzy, czyli wyznacznika obliczonego z wybranych wierszy i kolumn tej macierzy”⁸⁹.

Zastosowanie określonej metody estymacji zależy od badania identyfikowalności i tak, gdy równanie jest identyfikowalne jednoznacznie, stosujemy jedną z dwóch metod:

- podwójną metodę najmniejszych kwadratów,
- pośrednią metodę najmniejszych kwadratów.

Rozpatrzmy model o $m = 3$ równaniach łącznie współzależnych⁹⁰ i podanej macierzy (po przeniesieniu elementów na lewą stronę równań):

$$\begin{aligned} \hat{x}_1 &= a_1 x_2 + a_2 x_4 + a_0, \\ \hat{x}_2 &= b_1 x_3 + b_2 x_7 + b_0, \\ \hat{x}_7 &= c_1 x_1 + c_2 x_9 + c_0, \end{aligned} \quad \mathbf{B} = \begin{bmatrix} 1 & -a_1 & 0 \\ 0 & 1 & -b_2 \\ -c_1 & 0 & 1 \end{bmatrix}.$$

Zmienne endogeniczne X_1, X_2, X_7 są jednocześnie zmiennymi objaśniającymi. Sprawdźmy zatem identyfikowalność równań poprzez badanie macierzy \mathbf{A}_i , obejmującej np. w równaniu pierwszym zmienne egzogeniczne X_3, X_7, X_9 z równań drugiego i trzeciego, które nie występują w równaniu pierwszym

$$\mathbf{A}_1 = \begin{bmatrix} -b_1 & -b_2 & 0 \\ 0 & 1 & -c_2 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 1 & -a_2 & 0 \\ -c_1 & 0 & -c_2 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} -a_1 & 0 & -a_2 \\ 1 & -b_1 & 0 \end{bmatrix}.$$

Rząd macierzy $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ jest drugiego rzędu, czyli równa się $m - 1$ ⁹¹. Zatem nasze trzy równania są identyfikowalne. Aby dana macierz była drugiego rzędu, potrzeba, aby chociaż jeden z wyznaczników – w naszym przy-

⁸⁸ Ibidem, s. 196.

⁸⁹ T. Jurlewicz, Z. Skoczylas, *Algebra liniowa. Przykłady i zadania...*, op.cit.

⁹⁰ J. Dziechciarz (red), *Ekonometria. Metody...*, op.cit., s. 197.

⁹¹ E. Nowak, *Zarys metod ekonometrii. Zbiór zadań...*, op.cit.

kładzie „par kolumn” – był różny od zera. Macierze $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ warunek ten spełniają. Niech k oznacza liczbę zmiennych, które występują w modelu, a nie występują w określonym równaniu. Dla $m = 3$ równania, w każdym z równań $k = 3$. Tak więc $m - 1 < k$, co oznacza, że równania są niejednoznacznie identyfikowalne i umożliwia to zastosowanie podwójnej metody najmniejszych kwadratów (2MNK) do estymacji parametrów równań. Warto nadmienić, że metoda ta stosowana jest również w równaniach identyfikowalnych jednoznacznie, czyli gdy występuje warunek $m - 1 = k$.

Procedura postępowania w metodzie 2MNK jest następująca:

Następuje szacowanie KMNK postaci zredukowanej modelu o równaniu macierzowym $\mathbf{Y} = \mathbf{P}\mathbf{X} + \boldsymbol{\eta}$, co wyjaśniono we wstępie rozdziału. W wyniku otrzymujemy trzy równania, w których występuje ta sama macierz zmiennych objaśniających \mathbf{X} . Możemy szacować każde zredukowane równanie osobno, ponieważ doprowadziliśmy do modelu prostego równań, lub przeliczyć wszystkie równania jednocześnie⁹². Macierz transponowaną \mathbf{P}^T obliczamy z zależności:

$$\mathbf{P}^T = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y},$$

gdzie:

\mathbf{X} – macierz obserwacji zmiennych egzogenicznych łącznie ze zmienną X_0 równą 1 o wymiarze $n \times k$,

\mathbf{Y} – macierz o wymiarze $n \times m$ obserwacji zmiennych, endogenicznych.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & 1 \\ x_{21} & x_{22} & \dots & 1 \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & 1 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{bmatrix}, \quad \mathbf{P}^T = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{m1} \\ p_{12} & p_{22} & \dots & p_{m2} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & p_{mk} \end{bmatrix}.$$

Kolumny macierzy \mathbf{P}^T odpowiadają elementom i -tych równań postaci zredukowanej modelu.

Obliczenie wartości teoretycznych kolejnych zmiennych objaśniających rachunkiem macierzowym.

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{P}^T.$$

Zasadniczo interesują nas te wartości teoretyczne zmiennych, które w kolejnych równaniach występują jako zmienne objaśniające. Jeśli stosujemy

⁹² J. Dziechciarz (red.), *Ekonometria. Metody...*, op.cit., rozdz. 8.6.

funkcję [ILOZYN.MACIERZ(...;...)] Excela, to możemy szybko obliczyć wszystkie wartości teoretyczne.

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_{11} & \hat{y}_{12} & \dots & \hat{y}_{1m} \\ \hat{y}_{21} & \hat{y}_{22} & \dots & \hat{y}_{2m} \\ \dots & \dots & \dots & \dots \\ \hat{y}_{n1} & \hat{y}_{n2} & \dots & \hat{y}_{nm} \end{bmatrix}.$$

Mając wartości teoretyczne zmiennych, możemy estymować parametry równań postaci strukturalnej modelu. W tym względzie stosujemy ponownie KMNK, budując macierz \mathbf{X}_i oraz zastępując wartości obserwacji zmiennych endogenicznych ich wartościami teoretycznymi, pobranymi z odpowiedniej kolumny macierzy $\hat{\mathbf{Y}}$. Korzystamy tu ze znanego nam wzoru na wektor ocen parametrów strukturalnych i -tego równania modelu, bazując na wektorze obserwacji zmiennej objaśnianej y_i .

$$\mathbf{b}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y}_i.$$

Ocenę parametrów przeprowadzamy oddzielnie dla każdego równania postaci strukturalnej, chociaż istnieje możliwość zredukowania tego do jednego etapu. Wyjaśnienie postępowania w drugim przypadku znajdujemy w publikacji J. Dziechciarza⁹³.

1.9.6. Przykład zastosowania podwójnej metody najmniejszych kwadratów (2MNK) w Excelu

Jednym z powszechnie dostępnych sposobów estymacji parametrów według 2MNK jest zastosowanie funkcji Excela w zakresie rachunku macierzowego. Przypomnijmy sobie jednak nasz model o równaniach łącznie współzależnych:

$$\hat{x}_1 = a_1 x_2 + a_2 x_4 + a_0,$$

$$\hat{x}_2 = b_1 x_3 + b_2 x_7 + b_0,$$

$$\hat{x}_7 = c_1 x_1 + c_2 x_9 + c_0.$$

Widzimy, że zmiennymi endogenicznymi są X_1 , X_2 , X_7 a egzogenicznymi X_3 , X_4 , X_9 oraz $X_0 = 1$. Rozwiązanie oparte na postępowaniu według przykładu w publikacji J. Dziechciarza⁹⁴ jest następujące:

⁹³ Ibidem, rozdz. 8.7.

⁹⁴ Ibidem, przykład 8.3.

Określamy macierz \mathbf{X} obejmującą zmienne X_3, X_4, X_9, X_0 oraz \mathbf{Y} zawierającą zmienne X_1, X_2, X_7) na podstawie oznaczeń GRETL-a wartości zmiennych zbioru danych z tab. 1.9.1:

$$\mathbf{X} = \begin{bmatrix} 8621 & 331 & 3263 & 1 \\ 8666 & 348 & 3414 & 1 \\ \dots & \dots & \dots & \dots \\ 9605 & 584 & 4934 & 1 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 112,2 & 764 & 4215 \\ 131,1 & 839 & 4709 \\ \dots & \dots & \dots \\ 102,5 & 1070 & 10111 \end{bmatrix}.$$

Transponujemy macierz \mathbf{X} , korzystając z funkcji Excela [TRANSPONUJ(...)] poprzedzonej znakiem równości.

Szacujemy parametry postaci zredukowanej zmiennych X_3, X_4, X_9 , korzystając ze wzoru:

$$\mathbf{P}^T = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

Do obliczenia powyższego wektora używamy kolejno następujących funkcji Excela:

[MACIERZ.ILOCZYN(...;...)], wstawiając zakresy dla macierzy \mathbf{X}^T oraz \mathbf{X} ,

[MACIERZ.ILOCZYN(...;...)], wprowadzając zakres macierzy \mathbf{X}^T i wektora \mathbf{Y} ,

[MACIERZ.ODW(...)], wpisując obszar macierzy uzyskanej z iloczynu $(\mathbf{X}^T \mathbf{X})$,

[MACIERZ.ILOCZYN(...;...)], wprowadzając adresy arkusza kalkulacyjnego dla macierzy odwrotnej $(\mathbf{X}^T \mathbf{X})^{-1}$ oraz macierzy $(\mathbf{X}^T \mathbf{Y})$.

W wyniku działania uzyskujemy macierz \mathbf{P}^T , w której kolumnach występują trzy równania modelu, a w wierszach parametry odpowiadające zmiennym X_3, X_4, X_9 .

$$\mathbf{P}^T = \begin{bmatrix} 0,026963 & 0,1105 & 1,76378 \\ -0,44326 & 0,27825 & 2,3231 \\ 0,032775 & 0,033545 & 2,0681 \\ -41,860 & -291,485 & -18380,367 \end{bmatrix}.$$

Wartości z macierzy \mathbf{P}^T możemy zapisać jako równania postaci zredukowanej modelu:

$$\hat{x}_1 = 0,026963x_3 - 0,44326x_4 + 0,032775x_9 - 41,860,$$

$$\hat{x}_2 = 0,1105x_3 - 0,27828x_4 + 0,033545x_9 - 291,485,$$

$$\hat{x}_7 = 1,76378x_3 + 2,3231x_4 + 2,0681x_9 - 18380,367.$$

Obliczamy wartości teoretyczne zmiennych endogenicznych X_1 , X_2 , X_7 wszystkich obserwacji. W realizacji mikrokomputerowej do obliczenia wartości teoretycznych korzystamy z funkcji elementarnych Excela.

Tabela 1.9.3

Wartości teoretyczne zmiennych „ X_1 ”, „ X_2 ”, „ X_7 ”

\hat{x}_1	\hat{x}_2	\hat{x}_7
150,8182	862,768	4 342,299
149,4452	877,537	4 773,444
146,3365	890,9582	5 065,647
138,7214	918,5804	5 623,007
132,0162	941,0697	6 084,512
118,7833	958,5259	6 493,01
113,562	978,1472	6 865,641
120,1807	990,5593	7 287,52
128,1632	1 000,224	7 613,893
138,1593	1 003,138	7 763,268
150,8333	1 010,354	8 112,529
157,1125	999,6023	7 663,053
179,1301	1 008,124	8 179,297
199,5524	1 020,041	8 658,148
215,672	1 034,834	9 079,169
223,7149	1 050,656	9 324,71
230,6554	1 061,01	9 556,905
230,3410	19 077,971	9 817,749
228,1478	1 095,108	10 080,59
216,8056	1 104,992	10 035,39
206,6182	1 109,948	10 177,43
189,7215	1 082,612	9 817,499
157,1360	1 085,271	9 810,903
119,9730	1 097,967	10 121,38

Źródło: Na podstawie: J. Dziechciarz (red.) *Ekonometria. Metody, przykłady, zadania*, Akademia Ekonomiczna, Wrocław 2003, tab. 8.2.

Określamy macierz X_i kolejno dla trzech równań postaci strukturalnej modelu o równaniach łącznie współzależnych, zastępując w każdym rów-

naniu zmiennej endogenicznej X_1, X_2, X_7 wartości empiryczne obserwacji wartościami teoretycznymi. Następnie obliczamy parametry równań postaci strukturalnej według KMNK, bazując na wzorze:

$$\mathbf{b}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y}_i,$$

przy czym wektorem \mathbf{y}_i jest szereg obserwacji \mathbf{y}_i zmiennej objaśnianej.

W obliczeniach korzystamy z funkcji Excela w zakresie rachunku macierzowego, otrzymując w rezultacie pierwsze równanie trójrównaniowego modelu zmiennych łącznie współzależnych:

$$\mathbf{X}_1 = \begin{bmatrix} 862,768 & 331 & 1 \\ 877,537 & 348 & 1 \\ \dots & \dots & \dots \\ 10917,967 & 584 & 1 \end{bmatrix}, \quad \begin{aligned} \hat{x}_1 &= a_1 \hat{x}_2 + a_2 x_4 + a_0, \\ \hat{x}_1 &= 0,393 \hat{x}_2 - 0,507 x_4 - 18,69, \end{aligned}$$

– określenie równania drugiego $\hat{x}_2 = b_1 \hat{x}_3 + b_2 x_7 + b_0$ jest analogiczna jak pierwszego i uzyskujemy:

$$\mathbf{X}_2 = \begin{bmatrix} 8621 & 4342,30 & 1 \\ 8666 & 4773,444 & 1 \\ \dots & \dots & \dots \\ 9605 & 10121,38 & 1 \end{bmatrix}, \quad \begin{aligned} \hat{x}_2 &= b_1 \hat{x}_3 + b_2 x_7 + b_0, \\ \hat{x}_2 &= 0,00089 \hat{x}_3 + 0,039 x_7 + 688,692, \end{aligned}$$

– określenie macierzy \mathbf{X}_3 oraz oszacowanie parametrów równania trzeciego:

$$\mathbf{X}_3 = \begin{bmatrix} 150,8182 & 3263 & 1 \\ 149,4452 & 3414 & 1 \\ \dots & \dots & \dots \\ 119,9730 & 4934 & 1 \end{bmatrix}, \quad \begin{aligned} \hat{x}_7 &= c_1 \hat{x}_1 + c_2 x_9 + c_0, \\ \hat{x}_7 &= 5,770 \hat{x}_1 + 3,624 x_9 - 8288,367. \end{aligned}$$

Trzy równania dla $\hat{x}_1, \hat{x}_2, \hat{x}_7$ z oszacowanymi parametrami stanowią określenie modelu zmiennych łącznie współzależnych.

Innym sposobem estymacji parametrów równań zarówno postaci zredukowanej, jak i strukturalnej jest zastosowanie funkcji Excela REGLINP o składni:

[REGLINP(zakres wektora \mathbf{Y} ; zakres macierzy \mathbf{X} – bez \mathbf{X}_0 ;
PRAWDA; FAŁSZ]

Słowo PRAWDA oznacza podanie również stałej równania, a słowo FAŁSZ – niegenerowanie statystyki opisowej.

Oczywiście możemy również skorzystać z programu GRET i opcji (*Klasyczna metoda najmniejszych kwadratów*), którą szczegółowo przedstawiono w podrozdz. 1.9.4. Program GRET dysponuje także funkcją obliczeń bezpośrednich podwójnej metody najmniejszych kwadratów. Procedura postępowania w tym zakresie przedstawiona zostanie w podrozdz. 1.9.7 na przykładzie równań łącznie współzależnych. Warto jednak w tym miejscu wymienić modele, jakimi dysponuje program GRET, oczywiście poza klasyczną metodą najmniejszych kwadratów. Wymienione modele można wywołać z menu (*Model*):

- inne liniowe modele:
 - ważona metoda najmniejszych kwadratów (UMNK),
 - podwójna metoda najmniejszych kwadratów (2MNK),
 - MNK wysokiej precyzji,
- modele szeregów czasowych,
- modele panelowe,
- modele nieliniowe: logitowy, probitowy, tobitowy, Poissona, logistyczny, nieliniowa metoda najmniejszych kwadratów,
- odporne estymatory:
 - metoda najmniejszych wartości bezwzględnych,
 - korelacja rang,
- metoda największej wiarygodności,
- metoda momentów GMM,
- model równań, współzależnych, a w jego ramach, w zależności od identyfikowalności równań, następujące sposoby estymacji parametrów:
 - metoda pozornie niepowiązanych równań (sur),
 - potrójna metoda najmniejszych kwadratów (3sls),
 - metoda największej wiarygodności z pełną informacją (fiml),
 - metoda największej wiarygodności z ograniczoną informacją (liml),
 - klasyczna metoda najmniejszych kwadratów (ols),
 - podwójna metoda najmniejszych kwadratów (tsls),
 - ważona metoda najmniejszych kwadratów (wls).

Podstawy teoretyczne części z wymienionych sposobów estymacji spotykamy w pracy G. S. Maddali⁹⁵, a także w publikacji J. Dziechciarza⁹⁶ oraz w pracy E. Nowaka⁹⁷.

⁹⁵ G. S. M a d d a l a, *Ekonometria...*, op.cit.

⁹⁶ J. Dziechciarz (red), *Ekonometria. Metody...*, op.cit.

⁹⁷ E. N o w a k, *Zarys metod ekonometrii. Zbiór zadań...*, op.cit.

1.9.7. Przykład zastosowania 2MNK w programie GRETL

Podstawą jest plik danych GUS z lat 1962–1985 dla Polski, dostępny w internecie⁹⁸. Plik obejmuje 24 obserwacje dotyczące przemysłu chemicznego scharakteryzowanego dziewięcioma zmiennymi z określeniem w nawiasach ich pól w strukturze zbioru danych.

Tabela 1.9.4

Obserwacje zmiennych „ P_t ”, „ Z_t ”, „ I_t ”

Obserw.	produk	zatrud	inwest	time1	time2	time3	produk1 = produk(-1)	zatrud1 = zatrud(-1)	inwest1 = inwest(-1)
n	P_t	Z_t	I_t	t	t^2	t^3	P_{t-1}	Z_{t-1}	I_{t-1}
1	176	301	52	1	1	1	163	280	51
2	185	317	52	2	4	8	176	301	52
3	188	342	53	3	9	27	185	317	52
4	198	356	53	4	16	64	188	342	53
5	199	373	56	5	25	125	198	356	53
6	201	381	59	6	36	216	199	373	56
7	205	407	62	7	49	343	201	381	59
8	211	432	66	8	64	512	205	407	62
9	216	436	70	9	81	729	211	432	66
10	214	438	70	10	100	1 000	216	436	70
11	215	440	72	11	121	1 331	214	438	70
12	220	444	73	12	144	1 728	215	440	72
13	223	460	71	13	169	2 197	220	444	73
14	224	480	65	14	196	2 744	223	460	71
15	223	475	63	15	225	3 375	224	480	65
16	224	483	57	16	256	4 096	223	475	63
17	226	510	59	17	289	4 913	224	483	57
18	225	501	60	18	324	5 832	226	510	59
19	224	514	61	19	361	6 859	225	501	60
20	226	520	61	20	400	8 000	224	514	61
21	227	524	62	21	441	9 261	226	520	61
22	230	538	65	22	484	10 648	227	524	62
23	232	543	64	23	529	12 167	230	538	65
24	235	560	63	24	576	13 824	232	543	64

P_t – produkcja (produk),

Z_t – zatrudnienie (zatrud),

I_t – inwestycje (inwest),

t – zmienna czasowa (time1),

t^2 = t^2 – zmienna czasowa (time2),

stanowiąca kwadrat zmiennej czasowej t ,

t^3 = t^3 – zmienna czasowa (time3),

P_{t-1} – produk_1 = produk(-1); przesunięcie szeregu danych wielkości produkcji o 1 okres (rok),

Z_{t-1} – zatrud_1 = zatrud(-1),

I_{t-1} – inwest_1 = inwest(-1).

Źródło: Opracowanie własne na podstawie danych [www.kufel.torun.pl].

⁹⁸ [www.kufel.torun.pl].

Wartości danych z GRETL-a po edycji w Excelu podaje tabela 1.9.4. Ponieważ występują przesunięcia czasowe w produkcji, zatrudnieniu oraz inwestycjach, dlatego do tych informacji wprowadzono przykładowe umowne wartości obserwacji pierwszej. W tym względzie skorzystano ze ścieżki poprawy zbioru danych (*Edycja/Zastosuj/Zamknij*).

Wydruk zbioru danych można otrzymać bezpośrednio z GRETL-a w postaci tekstowej poprzez wywołanie menu (*Dane/Zaznacz wszystko*) (wszystkie pola); (*Dane/Pokaż wartości/Drukuj*).

Realizacja komputerowa w GRETL-u opiera się na modelu wielorównaniowym o równaniach łącznie współzależnych wykorzystujących⁹⁹ niektóre ze zmiennych tab. 1.9.4.

$$P_t = \beta_{12}Z_t + \alpha_{10} + \alpha_{11}P_{t-1} + \alpha_{13}I_{t-1} + \varepsilon_{1t},$$

$$Z_t = \beta_{21}P_t + \alpha_{20} + \alpha_{22}I_t + \alpha_{23}I_{t-1} + \varepsilon_{2t}.$$

P_t i Z_t to zmienne endogeniczne; P_{t-1} , I_t , I_{t-1} są zmiennymi egzogenicznymi; ε_{1t} oraz ε_{2t} stanowią reszty poszczególnych równań. Po przekształceniu powyższy model dwurównaniowy można zapisać jako $\mathbf{BY} + \mathbf{CX} = \boldsymbol{\varepsilon}$:

$$\begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \cdot \begin{bmatrix} P_t \\ Z_t \end{bmatrix} + \begin{bmatrix} -\alpha_{10} & -\alpha_{11} & 0 & -\alpha_{13} & 0 \\ -\alpha_{20} & 0 & -\alpha_{22} & -\alpha_{23} & -\alpha_{24} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ P_{t-1} \\ I_t \\ I_{t-1} \\ t \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

Mając równanie w modelu równań łącznie współzależnych (chcąc dobrać odpowiednią metodę estymacji parametrów α przy poszczególnych zmiennych egzogenicznych), musimy najpierw sprawdzić warunek identyfikowalności równań $G = 2$ według macierzy \mathbf{C} :

\mathbf{P}_t

– dwie zmienne niewystępujące: I_t , t (ich parametry w macierzy \mathbf{C} mają wartość zero),

– minimalna liczba zmiennych: $G - 1 = 2 - 1 = 1$.

Jest to więc równanie identyfikowalne niejednoznacznie¹⁰⁰.

\mathbf{Z}_t

– jedna zmienna P_{t-1} o zerowym parametrze.

⁹⁹ T. K u f e l, *Ekonometria. Rozwiązywanie problemów...*, op.cit., rozdz. 11.1.

¹⁰⁰ Ibidem, s. 151.

Występuje równość ilości zmiennych i oraz minimalnej ilości, tak więc równanie drugie jest jednoznacznie identyfikowalne. W tej sytuacji w obu równaniach można zastosować podwójną metodę najmniejszych kwadratów o procedurze realizacji:

- określenie zmiennych współzależnych występujących również jako zmienne objaśniające (P_t, Z_t),
- zastąpienie tych zmiennych wartościami teoretycznymi:

$$\hat{P}_t; \hat{Z}_t,$$

- szacowanie KMNK parametrów postaci zredukowanej, w której skład wchodzi cztery zmienne objaśniające (P_{t-1}, I_t, I_{t-1}, t) oraz stała (const), czyli zmienne z góry ustalone.

W odniesieniu do dwóch naszych równań postać zredukowana jest następująca¹⁰¹:

$$\begin{aligned}\hat{P}_t &= \pi_{10} + \pi_{11}P_{t-1} + \pi_{12}I_t + \pi_{13}I_{t-1} + \pi_{14}t, \\ \hat{Z}_t &= \pi_{20} + \pi_{21}P_{t-1} + \pi_{22}I_t + \pi_{23}I_{t-1} + \pi_{24}t.\end{aligned}$$

Wymienione zmienne oraz stałe są podstawą określenia wartości teoretycznych zmiennych endogenicznych P_t, Z_t , a więc posłużą nam w kolejnej estymacji do określenia postaci strukturalnej rozpatrywanego modelu. Tyle odniesień teoretycznych, a teraz zastosowanie metody 2MNK w programie GRETL do modelu dwóch równań współzależnych, co można przeprowadzić dwoma sposobami:

- a) osobno każde równanie,
- b) jednocześnie system równań.

Opierając się na opracowaniu T. Kufla¹⁰² na fragmentach wydruków komputerowych, przedstawione zostaną oba sposoby przetestowane przez autorów niniejszego opracowania.

Ad a)

Wybieramy ścieżkę z menu (*Model/Inne liniowe modele/Podwójna metoda najmniejszych kwadratów (2MNK...)*)

Specyfikacja pojedynczego równania dla „produkt”:

zmienna zależna – „produkt”,

zmienne niezależne (X), domyślnie: „const”, „zatrud”, „produkt1”, „inwest1”,

¹⁰¹ Ibidem, s. 152.

¹⁰² T. K u f e l, *Ekonometria. Rozwiązywanie problemów...*, op.cit.

instrumenty (zmiennie objaśniające postaci zredukowanej): „const”, „produk1”, „inwest”, „inwest1”, „time1”,
 OK (wyświetlane są wyniki estymacji zmiennej zależnej endogenicznej „produk” postaci zredukowanej pod określeniem *Model 1*).
 Zakres informacji dotyczący tego modelu zamieszczono w tabeli 1.9.5, a pod nią fragment statystyki opisowej.

Tabela 1.9.5

Model 1

Zmienne	Współczynnik (parametr)	Błąd standardowy	Statystyka <i>t</i>
const	64,9576	9,07634	7,157
zatrud	0,089998	0,0230224	3,909
produk1	0,452118	0,10279	4,398
Inwest1	0,219223	0,0848503	2,584

Średnia arytmetyczna zmiennej zależnej – 214,458,
 Odchylenie standardowe zmiennej zależnej – 15,8772,
 Suma kwadratów reszt – 62,2475,
 Błąd standardowy reszt – 1,76419,
 Współczynnik determinacji $R^2 = 0,989264$,
 Skorygowany współczynnik determinacji $\bar{R}^2 = 0,987654$.

Źródło: Opracowanie własne w GRETL.

Specyfikacja równania dla zmiennej endogenicznej „zatrud”:
 zmienna zależna – „zatrud”,
 zmienne niezależne: „const”, „produk”, „inwest”, „inwest1”, „time1”,
 instrumenty: „const”, „produk1”, „inwest”, „inwest1”, „time1”,
 OK (pojawia się raport z oszacowaniem parametrów oraz statystyką opisową pod nazwą *Modelu 2*, którego fragment zamieszczono w tabeli 1.9.6).

Tabela 1.9.6

Model 2

Zmienne	Współczynnik (parametr)	Błąd standardowy	Statystyka <i>t</i>
const	-147,971	67,5382	-2,191
produk	2,62417	0,451839	5,808
inwest	1,07858	0,596393	1,809
inwest1	-1,56254	0,610357	-2,560
time1	5,07182	0,8482234	5,979

Średnia arytmetyczna zmiennej zależnej – 448,958,
Odchylenie standardowe zmiennej zależnej – 73,8927,
Suma kwadratów reszt – 886,435,
Błąd standardowy reszt – 6,83041,
Współczynnik determinacji $R^2 = 0,992943$,
Skorygowany współczynnik determinacji $\bar{R}^2 = 0,991457$.

Źródło: Opracowanie własne w GRETL.

Po wywołaniu (*Plik/Drukuj/OK*) otrzymujemy wydruk obu modeli. Program GRETL umożliwia także sporządzenie wykresów rozrzutu danych z obserwacji, np. wskazanych kolumn bazy danych, tj.: „produk”, „zatrud”, „inwest” w zależności od okresów „time1”. Mając oszacowane parametry równań zmiennej zależnej „produk” czy też „zatrud”, możemy w GRETL sporządzić wykresy:

- empirycznej i wyrównywanej wartości danej zmiennej,
- reszty regresji dla zmiennej „produk” oraz „zatrud”.

Gdy skorzystamy ze ścieżki (*Narzędzia/Pokaż dziennik poleceń*), program GRETL udostępni nam wykaz zarejestrowanych komend języka skryptowego tworzonych podczas pracy z menu, np.:

```
open (C:\Documents and Settings\Dane aplikacji\GRETL\PMNK.gdt),  
gnuplot produk zatrud inwest time1,  
tsls produk const zatrud produk1 inwest1; const produk1  
inwest\inwest1 time1,  
tsls zatrud const produk inwest inwest1 time1; const produk1  
inwest\inwest1 time1.
```

Przy czym symbol *tsls* oznacza 2MNL, a składnia tego polecenia zawiera:

- zmienną endogeniczną,
- stałą oraz zmienne objaśniające danego równania,
- po średniku – ustalone z góry zmienne modelu.

Program GRETL umożliwia wybór podzbioru poleceń (komend) do określonego tematu, np. estymacji, a w ramach niego opis komendy *tsls*.

Ad b)

Do estymacji systemu równań wywołujemy menu (*Model/Model równań współzależnych*). W specyfikacji wprowadzamy trzy polecenia:

- equation – wskazanie zmiennych równania (najpierw zmiennych endogenicznych); powtarzamy to polecenie dla każdego równania,
- endog – wskazanie nazw zmiennych endogenicznych,
- instr – nazwy zmiennych z góry ustalonych (instrumentów) tworzących postać zredukowaną modelu.

W naszym przykładzie specyfikacja poleceń jest następująca:

```
equation produk const zatrud produk1 inwest1,  
equation zatrud const produk inwest inwest1 time1,  
endog produk zatrud,  
instr produk1 inwest inwest1 time1.
```

Okno specyfikacji modelu równań współzależnych pozwala wybrać metodę estymacji spośród siedmiu już wcześniej wymienionych, a wśród nich *tsls*. Wybór metody zależy od wcześniej przeprowadzonej identyfikacji równań. Do naszego przykładu, jak już wcześniej wykazano, odpowiednia jest metoda *tsls*. Przy realizacji estymacji systemu równań GRETl korzysta z pełnej bazy danych, tj. 24 obserwacji, i umożliwia wydrukowanie pełnego zestawu parametrów oraz statystyki opisowej do wszystkich równań jednocześnie.

Po opanowaniu składni poleceń – komend możemy dokonać estymacji modelu, pisząc instrukcję do programu i nadając tej instrukcji nazwę np. *Program.int*. Polecenia języka skryptowego tworzymy w oknie wywołanym menu (*Plik/Pliki poleceń skryptowych/Nowy plik skryptowy*). System sygnalizuje błędy w składni komend. Po naciśnięciu ikony (*Wykonaj*) następuje realizacja, przy czym program w naszym przykładzie pobiera obserwacje 2–24, pomijając pierwszą, gdyż kieruje się przesunięciem jednookresowym zmiennych „produkt1”, „inwest1”. Zaprezentowanie wyników estymacji parametrów jest analogiczne do otrzymywanych z systemu równań.

W niniejszym podrozdziale zaprezentowano tylko element oprogramowania GRETl-a w zastosowaniu do modeli wielorównaniowych. Szerszą panoramę jego możliwości spotykamy w pracy T. Kufla¹⁰³. Opanowanie tego narzędzia wymaga jednak bezpośrednich testowań na komputerze poszczególnych opcji rozbudowanego menu. Autorzy mają nadzieję, że wykonana przez nich próba prześledzenia procedur metody 2MNL w kontekście założeń teoretycznych wynikających z literatury przedmiotu będzie zachętą do szerszego poznania dalszych metod estymacji z zastosowaniem tego programu.

¹⁰³ Ibidem.

1.10. Modele stacjonarnych szeregów czasowych

Zmienność stacjonarnych szeregów czasowych można wyjaśnić ich zmiennością z przeszłości oraz składnikiem losowym. Zmiennymi objaśniającymi zatem jest opóźniona zmienna zależna oraz składnik resztowy.

Etapy budowy modeli zwanych modelami Boxa i Jenkinsa.

- 1) Badanie stacjonarności szeregu czasowego przy określonych warunkach.
- 2) Badanie rzędu p autoregresji.
- 3) Badanie rzędu średniej ruchomej.

1.10.1. Badanie stacjonarności przy określonych warunkach

Jeśli szereg czasowy ma stałą średnią i wariancję, a kowariancja zależy tylko od rzędu opóźnienia k , to taki szereg nazywamy stacjonarnym¹⁰⁴.

Te warunki można zapisać jako:

$$E(Y_t) = \text{const} = \bar{y}$$

$$\text{VAR}(Y_t) = \text{const} = S^2$$

$$\text{COV}(Y_t, Y_{t-k}) = \text{COV}_k$$

Do ilustracji rodzaju szeregu weźmy znany nam już (podrozdz. 1.3) zbiór danych przedstawiający odczytane z licznika Geigera ilości cząstek emitowanych przez substancję radioaktywną w 100 kolejnych 40-sekundowych okresach.

T a b e l a 1.10.1

Ilość emitowanych cząstek

1	2	3	4	5	6	7	8	9	10
23	20	16	18	30	22	26	15	13	18
11	12	13	14	15	16	17	18	19	20
14	17	11	37	21	16	10	20	22	25
21	22	23	24	25	26	27	28	29	30
19	19	19	20	12	23	24	17	18	16

¹⁰⁴ M. Lipiec-Zajchowska (red), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria*, rozdz. 3.3..., op.cit.

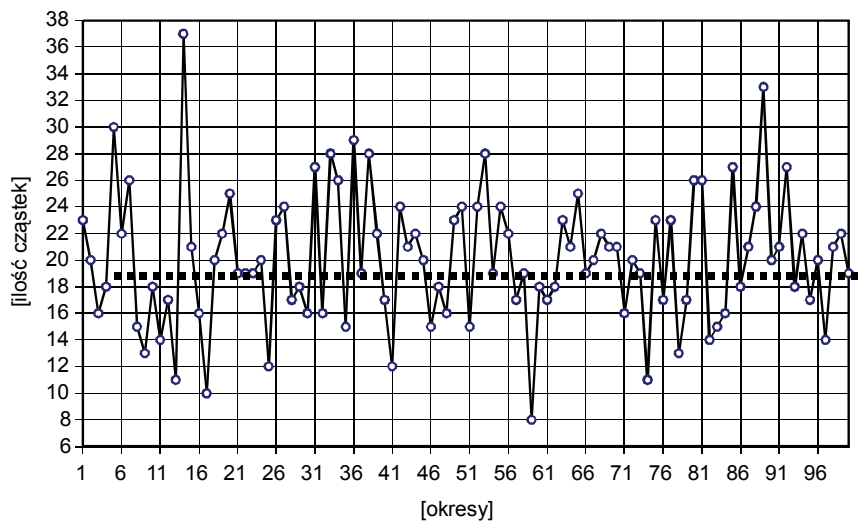
cd. tab. 1.10.1

31	32	33	34	35	36	37	38	39	40
27	16	28	26	15	29	19	28	22	17
41	42	43	44	45	46	47	48	49	50
12	24	21	22	20	15	18	16	23	24
51	52	53	54	55	56	57	58	59	60
15	24	28	19	24	22	17	19	8	18
61	62	63	64	65	66	67	68	69	70
17	18	23	21	25	19	20	22	21	21
71	72	73	74	75	76	77	78	79	80
16	20	19	11	23	17	23	13	17	26
81	82	83	84	85	86	87	88	89	90
26	14	15	16	27	18	21	24	33	20
91	92	93	94	95	96	97	98	99	100
21	27	18	22	17	20	14	21	22	19

pierwszy wiersz – okres; drugi wiersz – ilość cząstek

Źródło: Opracowanie własne w Excelu; dane według tab. 1.3.1.

Przedstawmy ten szereg czasowy na wykresie punktowym wokół wartości średniej $\bar{y} = 20,04$ dla wariancji próby ($n - 1 = 99$) wynoszącej $S^2 = 24,77$.



Źródło: Opracowanie własne w Excelu.

Wykres 1.10.1. Ilość wyemitowanych cząstek w kolejnych 100 obserwacjach

Badanie stacjonarności opiera się na funkcji autokorelacji lub teście Dickey-Fullera¹⁰⁵. Autokorelacją nazywamy korelację zmiennej samej ze sobą, lecz opóźnioną o rząd k okresów. W przykładzie skupimy się na zastosowaniu funkcji autokorelacji. W budowie modeli opartych o tę funkcję obliczamy współczynniki autokorelacji ρ_k :

$$\rho_k = \frac{\sum_{t=1+k}^n (y_t - \bar{y}) \cdot (y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}. \quad (1.10.1)$$

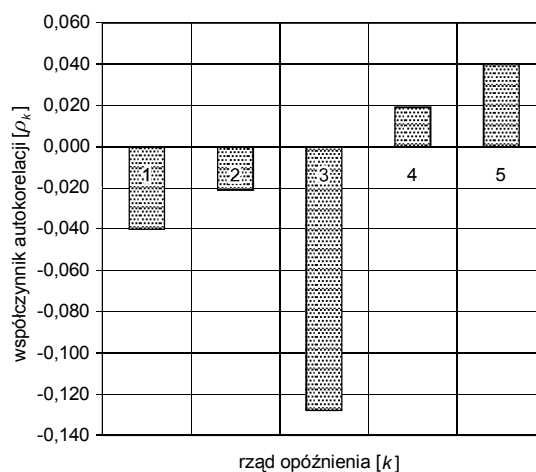
gdzie:

- y – wartość zmiennej w okresie t ,
- \bar{y} – średnia arytmetyczna zmiennej,
- k – rząd opóźnienia.

P r z y k ł a d

Obliczamy współczynniki autokorelacji do przedstawionej na wykresie (1.10.1) zmiennej Y (ilość cząstek emitowanych przez substancję radioaktywną) dla rzędów opóźnienia $k = -2, -1, 0, 1, 2, 3, 4, 5$ próby zawierającej 100 obserwacji. Fragment obliczeń pomocniczych dla 61 obserwacji podano w tabeli 1.10.2.

Zależność między ρ_k a k nazywamy funkcją autokorelacji $ACF(k)$. W naszym przykładzie, gdy k [1–5], obserwujemy nieznanikanie wykładnicze funkcji autokorelacji, co potwierdza, że na podstawie tego badania mamy do czynienia z szeregiem niestacjonarnym.



Ź r ó d ł o: Opracowanie własne w Excelu.

Wykres 1.10.2. Zależność między współczynnikiem autokorelacji a rzędem opóźnień

¹⁰⁵ Ibidem, s. 111.

Tabela 1.10.2

Obliczenie pomocnicze do współczynników autokorelacji

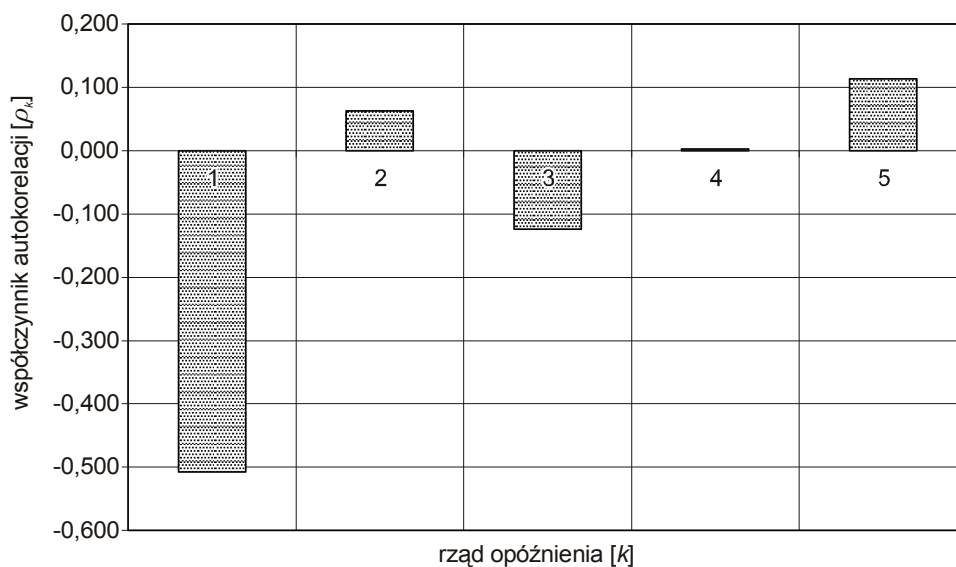
t	y_t	y_{t-2}	y_{t-1}	y_{t+1}	y_{t+2}	$y_t - y_t'$	$y_{t-2} - y_t'$	$y_{t-1} - y_t'$	$y_{t+1} - y_t'$	$y_{t+2} - y_t'$	$(y_t - y_t')^2$	$g \times h$	$g \times i$
a	b	c	d	e	f	g	h	i	j	k	l	m	n
1	23			20	16	2,96			-0,04	-4,04	8,7616		
2	20		23	16	18	-0,04		2,96	-4,04	-2,04	0,002		-0,118
3	16	23	20	18	30	-4,04	2,96	-0,04	-2,04	9,96	16,322	-11,958	0,162
4	18	20	16	30	22	-2,04	-0,04	-4,04	9,96	1,96	4,162	0,082	8,242
5	30	16	18	22	26	9,96	-4,04	-2,04	1,96	5,96	99,202	-40,238	-20,318
6	22	18	30	26	15	1,96	-2,04	9,96	5,96	-5,04	3,842	-3,998	19,522
7	26	30	22	15	13	5,96	9,96	1,96	-5,04	-7,04	35,522	59,362	11,682
8	15	22	26	13	18	-5,04	1,96	5,96	-7,04	-2,04	25,402	-9,878	-30,038
9	13	26	15	18	14	-7,04	5,96	-5,04	-2,04	-6,04	49,562	-41,958	35,482
10	18	15	13	14	17	-2,04	-5,04	-7,04	-6,04	-3,04	4,162	10,282	14,362
11	14	13	18	17	11	-6,04	-7,04	-2,04	-3,04	-9,04	36,482	42,522	12,322
12	17	18	14	11	37	-3,04	-2,04	-6,04	-9,04	16,96	9,242	6,202	18,362
13	11	14	17	37	21	-9,04	-6,04	-3,04	16,96	0,96	81,722	54,602	27,482
14	37	17	11	21	16	16,96	-3,04	-9,04	0,96	-4,04	287,642	-51,558	-153,318
15	21	11	37	16	10	0,96	-9,04	16,96	-4,04	-10,04	0,922	-8,678	16,282
16	16	37	21	10	20	-4,04	16,96	0,96	-10,04	-0,04	16,322	-68,518	-3,878
17	10	21	16	20	22	-10,04	0,96	-4,04	-0,04	1,96	100,802	-9,638	40,562
18	20	16	10	22	25	-0,04	-4,04	-10,04	1,96	4,96	0,002	0,162	0,402
19	22	10	20	25	19	1,96	-10,04	-0,04	4,96	-1,04	3,842	-19,678	-0,078
20	25	20	22	19	19	4,96	-0,04	1,96	-1,04	-1,04	24,602	-0,198	9,722
21	19	22	25	19	19	-1,04	1,96	4,96	-1,04	-1,04	1,082	-2,038	-5,158
22	19	25	19	19	20	-1,04	4,96	-1,04	-1,04	-0,04	1,082	-5,158	1,082
23	19	19	19	20	12	-1,04	-1,04	-1,04	-0,04	-8,04	1,082	1,082	1,082
24	20	19	19	12	23	-0,04	-1,04	-1,04	-8,04	2,96	0,002	0,042	0,042
25	12	19	20	23	24	-8,04	-1,04	-0,04	2,96	3,96	64,642	8,362	0,322
26	23	20	12	24	17	2,96	-0,04	-8,04	3,96	-3,04	8,762	-0,118	-23,798
27	14	12	23	17	18	3,96	-8,04	2,96	-3,04	-2,04	15,682	-31,838	11,722
28	17	23	24	18	16	-3,04	2,96	3,96	-2,04	-4,04	9,242	-8,998	-12,038
29	18	24	17	16	27	-2,04	3,96	-3,04	-4,04	6,96	4,162	-8,078	6,202
30	16	17	18	27	16	-4,04	-3,04	-2,04	6,96	-4,04	16,322	12,282	8,242
31	27	18	16	16	28	6,96	-2,04	-4,04	-4,04	7,96	48,442	-14,198	-28,118
32	16	16	27	28	26	-4,04	-4,04	6,96	7,96	5,96	16,322	16,322	-28,118
33	28	27	16	26	15	7,96	6,96	-4,04	5,96	-5,04	63,362	55,402	-32,158
34	26	16	28	15	29	5,96	-4,04	7,96	-5,04	8,96	35,522	-24,078	47,442
35	15	28	26	29	19	-5,04	7,96	5,96	8,96	-1,04	25,402	-40,118	-30,038
36	29	26	15	19	28	8,96	5,96	-5,04	-1,04	7,96	80,282	53,402	-45,158
37	19	15	29	28	22	-1,04	-5,04	8,96	7,96	1,96	1,082	5,242	-9,318
38	28	29	19	22	17	7,96	8,96	-1,04	1,96	-3,04	63,362	71,322	-8,278
39	22	19	28	17	12	1,96	-1,04	7,96	-3,04	-8,04	3,842	-2,038	15,602
40	17	28	22	12	24	-3,04	7,96	1,96	-8,04	3,96	9,242	-24,198	-5,958
41	12	22	17	24	21	-8,04	1,96	-3,04	3,96	0,96	64,642	-15,758	24,442
42	24	17	12	21	22	3,96	-3,04	-8,04	0,96	1,96	15,682	-12,038	-31,838
43	21	12	24	22	20	0,96	-8,04	3,96	1,96	-0,04	0,922	-7,718	3,802
44	22	24	21	20	15	1,96	3,96	0,96	-0,04	-5,04	3,842	7,762	1,882
45	20	21	22	15	18	-0,04	0,96	1,96	-5,04	-2,04	0,002	-0,038	-0,078
46	15	22	20	18	16	-5,04	1,96	-0,04	-2,04	-4,04	25,402	-9,878	0,202
47	18	20	15	16	23	-2,04	-0,04	-5,04	-4,04	2,96	4,162	0,082	10,282
48	16	15	18	23	24	-4,04	-5,04	-2,04	2,96	3,96	16,322	20,362	8,242
49	23	18	16	24	15	2,96	-2,04	-4,04	3,96	-5,04	8,762	-6,038	-11,958
50	24	16	23	15	24	3,96	-4,04	2,96	-5,04	3,96	15,682	-15,998	11,722
51	15	23	24	24	28	-5,04	2,96	3,96	3,96	7,96	25,402	-14,918	-19,958
52	24	24	15	28	19	3,96	3,96	-5,04	7,96	-1,04	15,682	15,682	-19,958
53	28	15	24	19	24	7,96	-5,04	3,96	-1,04	3,96	63,362	-40,118	31,522
54	19	24	28	24	22	-1,04	3,96	7,96	3,96	1,96	1,082	-4,118	-8,278
55	24	28	19	22	17	3,96	7,96	-1,04	1,96	-3,04	15,682	31,522	-4,118
56	22	19	24	17	19	1,96	-1,04	3,96	-3,04	-1,04	3,842	-2,038	7,762
57	17	24	22	19	8	-3,04	3,96	1,96	-1,04	-12,04	9,242	-12,038	-5,958
58	19	22	17	8	18	-1,04	1,96	-3,04	-12,04	-2,04	1,082	-2,038	3,162
59	8	17	19	18	17	-12,04	-3,04	-1,04	-2,04	-3,04	144,962	36,602	12,522
60	18	19	8	17	18	-2,04	-1,04	-12,04	-3,04	-2,04	4,162	2,122	24,562
61	17	8	18	18	23	-3,04	-12,04	-2,04	-2,04	2,96	9,242	36,602	6,202

cd. tab. 1.10.2

$g \times k$	$g \times j$	y_{t+3}	$y_{t+3} - \bar{y}$	$g \times s$	y_{t+4}	$y_{t+4} - \bar{y}$	$g \times w$	y_{t+5}	$y_{t+5} - \bar{y}$	$g \times s$
o	P	r	s	t	u	w	x	y	z	a_1
-0,1184	-11,9584	18	-2,04	-6,0384	30	9,96	29,4816	22	1,96	5,8016
0,1616	0,0816	30	9,96	-0,3984	22	1,96	-0,0784	26	5,96	-0,2384
8,2416	-40,2384	22	1,96	-7,9184	26	5,96	-24,0784	15	-5,04	20,3616
-20,3184	-3,9984	26	5,96	-12,1584	15	-5,04	10,2816	13	-7,04	14,3616
19,5216	59,3616	15	-5,04	-50,1984	13	-7,04	-70,1184	18	-2,04	-20,318
11,6816	-9,8784	13	-7,04	-13,7984	18	-2,04	-3,9984	14	-6,04	-11,838
-30,0384	-41,9584	18	-2,04	-12,1584	14	-6,04	-35,9984	17	-3,04	-18,118
35,4816	10,2816	14	-6,04	30,4416	17	-3,04	15,3216	11	-9,04	45,5616
14,3616	42,5216	17	-3,04	21,4016	11	-9,04	63,6416	37	16,96	-119,4
12,3216	6,2016	11	-9,04	18,4416	37	16,96	-34,5984	21	0,96	-1,9584
18,3616	54,6016	37	16,96	-102,438	21	0,96	-5,7984	16	-4,04	24,4016
27,4816	-51,5584	21	0,96	-2,9184	16	-4,04	12,2816	10	-10,04	30,5216
-153,3184	-8,6784	16	-4,04	36,5216	10	-10,04	90,7616	20	-0,04	0,3616
16,2816	-68,5184	10	-10,04	-170,278	20	-0,04	-0,6784	22	1,96	33,2416
-3,8784	-9,6384	20	-0,04	-0,0384	22	1,96	1,8816	25	4,96	4,7616
40,5616	0,1616	22	1,96	-7,9184	25	4,96	-20,0384	19	-1,04	4,2016
0,4016	-19,6784	25	4,96	-49,7984	19	-1,04	10,4416	19	-1,04	10,4416
-0,0784	-0,1984	19	-1,04	0,0416	19	-1,04	0,0416	19	-1,04	0,0416
9,7216	-2,0384	19	-1,04	-2,0384	19	-1,04	-2,0384	20	-0,04	-0,0784
-5,1584	-5,1584	19	-1,04	-5,1584	20	-0,04	-0,1984	12	-8,04	-39,878
1,0816	1,0816	20	-0,04	0,0416	12	-8,04	8,3616	23	2,96	-3,0784
1,0816	0,0416	12	-8,04	8,3616	23	2,96	-3,0784	24	3,96	-4,1184
0,0416	8,3616	23	2,96	-3,0784	24	3,96	-4,1184	17	-3,04	3,1616
0,3216	-0,1184	24	3,96	-0,1584	17	-3,04	0,1216	18	-2,04	0,0816
-23,7984	-31,8384	17	-3,04	24,4416	18	-2,04	16,4016	16	-4,04	32,4816
11,7216	-8,9984	18	-2,04	-6,0384	16	-4,04	-11,9584	27	6,96	20,6016
-12,0384	-8,0784	16	-4,04	-15,9984	27	6,96	27,5616	16	-4,04	-15,998
6,2016	12,2816	27	6,96	-21,1584	16	-4,04	12,2816	28	7,96	-24,198
8,2416	-14,1984	16	-4,04	8,2416	28	7,96	-16,2384	26	5,96	-12,158
-28,1184	16,3216	28	7,96	-32,1584	26	5,96	-24,0784	15	-5,04	20,3616
-28,1184	55,4016	26	5,96	41,4816	15	-5,04	-35,0784	29	8,96	62,3616
-32,1584	-24,0784	15	-5,04	20,3616	29	8,96	-36,1984	19	-1,04	4,2016
47,4416	-40,1184	29	8,96	71,3216	19	-1,04	-8,2784	28	7,96	63,3616
-30,0384	53,4016	19	-1,04	-6,1984	28	7,96	47,4416	22	1,96	11,6816
-45,1584	5,2416	28	7,96	-40,1184	22	1,96	-9,8784	17	-3,04	15,3216
-9,3184	71,3216	22	1,96	17,5616	17	-3,04	-27,2384	12	-8,04	-72,038
-8,2784	-2,0384	17	-3,04	3,1616	12	-8,04	8,3616	24	3,96	-4,1184
15,6016	-24,1984	12	-8,04	-63,9984	24	3,96	31,5216	21	0,96	7,6416
-5,9584	-15,7584	24	3,96	7,7616	21	0,96	1,8816	22	1,96	3,8416
24,4416	-12,0384	21	0,96	-2,9184	22	1,96	-5,9584	20	-0,04	0,1216
-31,8384	-7,7184	22	1,96	-15,7584	20	-0,04	0,3216	15	-5,04	40,5216
3,8016	7,7616	20	-0,04	-0,1584	15	-5,04	-19,9584	18	-2,04	-8,0784
1,8816	-0,0384	15	-5,04	-4,8384	18	-2,04	-1,9584	16	-4,04	-3,8784
-0,0784	-9,8784	18	-2,04	-3,9984	16	-4,04	-7,9184	23	2,96	5,8016
0,2016	0,0816	16	-4,04	0,1616	23	2,96	-0,1184	24	3,96	-0,1584
10,2816	20,3616	23	2,96	-14,9184	24	3,96	-19,9584	15	-5,04	25,4016
8,2416	-6,0384	24	3,96	-8,0784	15	-5,04	10,2816	24	3,96	-8,0784
-11,9584	-15,9984	15	-5,04	20,3616	24	3,96	-15,9984	28	7,96	-32,158
11,7216	-14,9184	24	3,96	11,7216	28	7,96	23,5616	19	-1,04	-3,0784
-19,9584	15,6816	28	7,96	31,5216	19	-1,04	-4,1184	24	3,96	15,6816
-19,9584	-40,1184	19	-1,04	5,2416	24	3,96	-19,9584	22	1,96	-9,8784
31,5216	-4,1184	24	3,96	15,6816	22	1,96	7,7616	17	-3,04	-12,038
-8,2784	31,5216	22	1,96	15,6016	17	-3,04	-24,1984	19	-1,04	-8,2784
-4,1184	-2,0384	17	-3,04	3,1616	19	-1,04	1,0816	8	-12,04	12,5216
7,7616	-12,0384	19	-1,04	-4,1184	8	-12,04	-47,6784	18	-2,04	-8,0784
-5,9584	-2,0384	8	-12,04	-23,5984	18	-2,04	-3,9984	17	-3,04	-5,9584
3,1616	36,6016	18	-2,04	6,2016	17	-3,04	9,2416	18	-2,04	6,2016
12,5216	2,1216	17	-3,04	3,1616	18	-2,04	2,1216	23	2,96	-3,0784
24,5616	36,6016	18	-2,04	24,5616	23	2,96	-35,6384	21	0,96	-11,558
6,2016	4,1616	23	2,96	-6,0384	21	0,96	-1,9584	25	4,96	-10,118
6,2016	-8,9984	21	0,96	-2,9184	25	4,96	-15,0784	19	-1,04	3,1616

Źródło: Opracowanie własne w Excelu, gdzie: $y^- \rightarrow \bar{y}$.

Wartości współczynników autokorelacji są symetryczne względem $k = 0$. Przyjmują wartości z przedziału $\langle -1; 1 \rangle$. Ponieważ badanie stacjonarności oparte o funkcję autokorelacji szeregu czasowego wykazało jego niestacjonarność, to następnie badamy stacjonarność różnic pierwszego rzędu¹⁰⁶: $\Delta y_t = y_t - y_{t-1}$. Wynik tego badania pokazano na wykresie 1.10.3 jako funkcję autokorelacji szeregu pierwszych różnic autokorelacji.



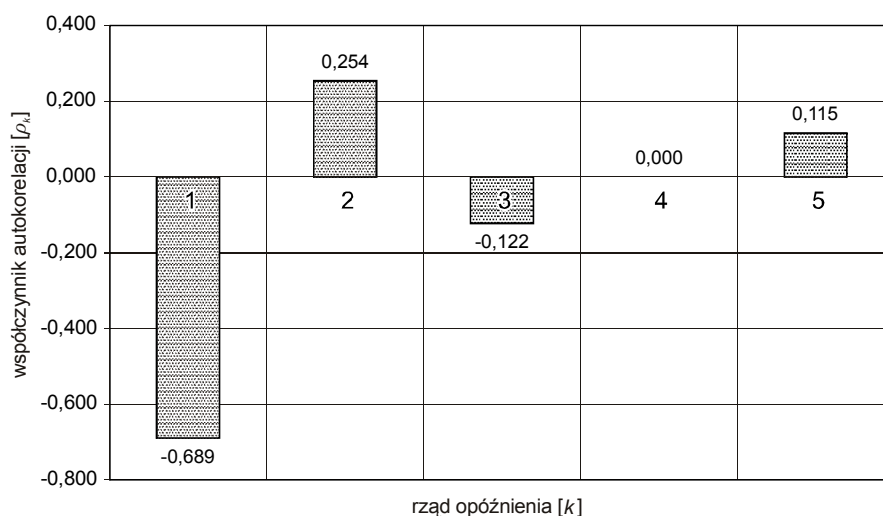
Źródło: Opracowanie własne w Excelu.

Wykres 1.10.3. Zależność współczynnika autokorelacji od rzędu opóźnienia w szeregu pierwszych różnic

Analizując wykres, stwierdzamy nadal niestacjonarność szeregu czasowego, gdyż wartość współczynnika autokorelacji wraz ze wzrostem rzędu opóźnień nie dąży wykładniczo do zera. Przeprowadzamy więc dalsze badania, bazując na szeregu różnic drugiego rzędu.

$$\Delta_2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}).$$

¹⁰⁶ M. Lipiec-Zajchowska (red.), *Wspomaganie...*, t. 2, op.cit., rozdz. 3.3.1.



Źródło: Opracowanie własne w Excelu.

Wykres 1.10.4. Zależność współczynnika autokorelacji od rzędu opóźnienia w szeregu drugich różnic

Obserwujemy nadal nieznikające wykładnicze współczynniki autokorelacji dla $k = 1-5$. Należy dalej badać stacjonarność, budując szeregi różnic trzeciego i wyższych rzędów, lecz dalsze badanie pozostawiamy Czytelnikowi. W celu przejścia do omówienia dalszych etapów wymienionych w podrozdziale 1.10.1, przyjmujemy umownie, że nasz szereg czasowy jest niestacjonarny. Sprawdźmy to jednak jeszcze raz na etapie badania rzędu p autoregresji.

1.10.2. Badanie rzędu p autoregresji

Model autoregresyjny rzędu p :

$$Y_t = c + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t,$$

przyjmując dla y_{t-k} symbol $AR(k)$, możemy przedstawić jako:

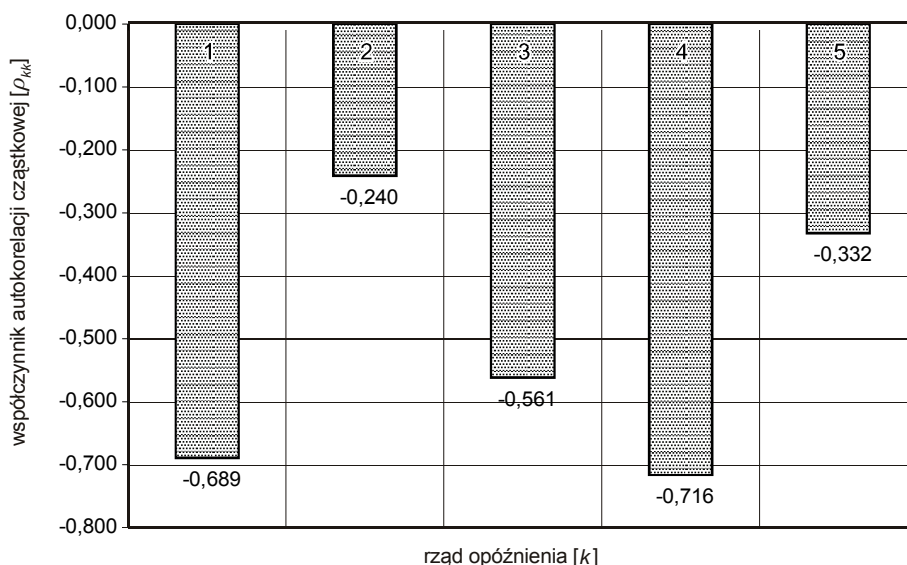
$$Y_t = c + \alpha_1 AR(1) + \alpha_2 AR(2) + \dots + \alpha_p AR(p) + \varepsilon_t.$$

Budowa takich modeli wymaga rzędu opóźnień, który ustalamy dla $k > 1$ na podstawie funkcji autokorelacji cząstkowej:

$$\rho_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \rho_{kk-1} \cdot \rho_{k-1}}{1 - \sum_{j=1}^{k-1} \rho_{kk-1} \rho_j}. \quad (1.10.2)$$

(Dla $k = 1$ – patrz wykres 1.10.4, $\rho_{kk} = \rho_1$, zatem $\rho_{11} = -0,689$).

Funkcja autokorelacji cząstkowej stanowi zależność pomiędzy współczynnikiem autokorelacji cząstkowej ρ_{kk} a rzędem opóźnienia k . Sporządźmy wykres słupkowy funkcji współczynników autokorelacji cząstkowej dla danych rzędów opóźnień k .



Źródło: Opracowanie własne w Excelu.

Wykres 1.10.5. Zależność między funkcją autokorelacji cząstkowej a rzędem opóźnienia

Współczynniki autokorelacji cząstkowej podane na wykresie 1.10.5 obliczono dla k od 1 do 5 na podstawie wzoru (1.10.2). Po którym opóźnieniu k urywa się funkcja autokorelacji cząstkowej, czyli jest bliska zero, możemy dopiero stwierdzić po k o znacznie wyższej wartości, np. >10 . To urwanie umożliwia nam określenie rzędu autoregresji i sformułowanie modelu. Manualne takie badanie nawet ze wspomaganie Excelem jest bardzo pracochłonne i dlatego stosuje się odpowiednie programy komputerowe z estymacją testu DF (Dickey-Fullera)¹⁰⁷. Na podstawie dotychczasowego badania podtrzymujemy stwierdzenie o niestacjonarności naszego szeregu czasowego.

Jeśli wystąpiłoby zanikanie zarówno funkcji autokorelacji, jak i autokorelacji cząstkowej, można zbudować model mieszany o nazwie $ARMA(p,q)$. Stanowi on połączenie wcześniej omawianego modelu autoregresyjnego rzędu p oraz modelu średniej ruchomej rzędu q , o którym powiemy w kolejnym, trzecim etapie.

¹⁰⁷ Ibidem, szerzej: rozdz. 3.3.

1.10.3. Badanie rzędu średniej ruchomej

Badanie rzędu średniej ruchomej oparte jest na średnich ruchomych składnika ε o stałej wariancji i średniej równej zero¹⁰⁸.

Model średnich ruchomych rzędu q jest liniową kombinacją składników resztowych:

$$Y_t = c + \beta_1 \cdot \varepsilon_{t-1} + \beta_2 \cdot \varepsilon_{t-2} + \dots + \beta_q \cdot \varepsilon_{t-q}.$$

Stosowana jest metoda polegająca na obliczaniu średnich arytmetycznych reszt:

$$\begin{aligned} \varepsilon_1 &= y_1 - \bar{y}, \\ \varepsilon_2 &= y_2 - \bar{y} - \beta_1 \cdot \varepsilon_1, \\ \varepsilon_3 &= y_3 - \bar{y} - \beta_1 \cdot \varepsilon_1 - \beta_2 \cdot \varepsilon_2, \\ &\dots \end{aligned}$$

P r z y k ł a d

Zbadamy wstępnie występowanie stacjonarności szeregu czasowego na podstawie spożycia ziemniaków w Polsce w latach 1985–2005 w przeliczeniu na jednego mieszkańca.

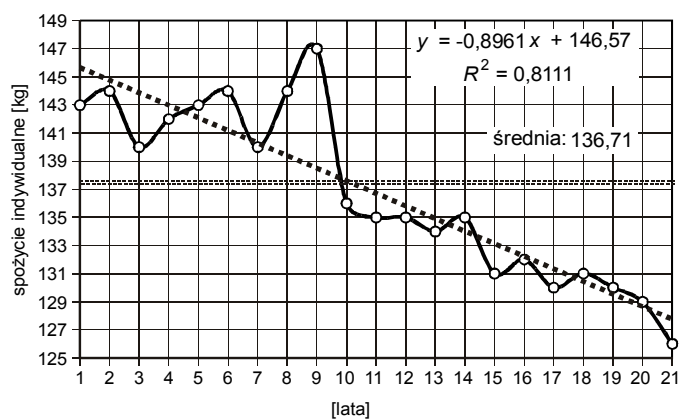
T a b e l a 1.10.3

Dane statystyczne

Okres	Rok	Spożycie indywidualne [kg]
1	1985	143
2	1986	144
3	1987	140
4	1988	142
5	1989	143
6	1990	144
7	1991	140
8	1992	144
9	1993	147
10	1994	136
11	1995	135
12	1996	135
13	1997	134
14	1998	135
15	1999	131
16	2000	132
17	2001	130
18	2002	131
19	2003	130
20	2004	129
21	2005	126

Ź r ó d ł o: „Małe Roczniki Statystyczne” (lata 1999–2006).

¹⁰⁸ Ibidem, rozdz. 3.3.3.



Źródło: Opracowanie własne w Excelu.

Wykres 1.10.6. Spożycie indywidualne ziemniaków w latach 1985–2005. Wykres punktowy i linia trendu

Poziomy zmiennej szeregu czasowego nie oscylują wokół średniej o wartości 136,71, lecz mają tendencję zmian wobec trendu malejącego, zatem szereg czasowy jest niestacjonarny.

Przykład

Firma produkująca okna plastikowe osiągnęła zyski ze swej działalności w latach 1985–1994. Przeprowadzimy badanie stacjonarności szeregu czasowego.

Tabela 1.10.4

Dane umowne

Okres	Rok	Zysk [tys. zł]
1	1985	-300
2	1986	280
3	1987	150
4	1988	820
5	1989	-10
6	1990	1 130
7	1991	880
8	1992	1 010
9	1993	670
10	1994	2 050

Źródło: Opracowanie własne w Excelu.

a) Obliczenie współczynników autokorelacji ρ_k

Tabela 1.10.5

Obliczenia pomocnicze oraz ustalenie współczynników autokorelacji (fragment)

t	y_t	y_{t-2}	y_{t-1}	y_{t+1}	y_{t+2}	$y_t - \bar{y}$	$y_{t-2} - \bar{y}$	$y_{t-1} - \bar{y}$	$y_{t+1} - \bar{y}$	$y_{t+2} - \bar{y}$	$(y_t - \bar{y})^2$				
t	Z	Z(-2)	Z(-1)	Z(+1)	Z(+2)	a	b	c	d	e		a · b	a · c	a · d	a · e
1	-300			280	150	-968			-388	-518	937 024			375 584	501 424
2	280		-300	150	820	-388		-968	-518	152	150 544		375 584	200 984	-58 976
3	150	-300	280	820	-10	-518	-968	-388	152	-678	268 324	501 424	200 984	-78 736	351 204
4	820	280	150	-10	1 130	152	-388	-518	-678	462	23 104	-58 976	-78 736	-103 056	70 224
5	-10	150	820	1 130	880	-678	-518	152	462	212	459 684	351 204	-103 056	-313 236	-143 736
6	1 130	820	-10	880	1 010	462	152	-678	212	342	213 444	70 224	-313 236	97 944	158 004
7	880	-10	1 130	1 010	670	212	-678	462	342	2	44 944	-143 736	97 944	72 504	424
8	1 010	1 130	880	670	2 050	342	462	212	2	1 382	116 964	158 004	72 504	684	472 644
9	670	880	1 010	2 050		2	212	342	1 382		4	424	684	2 764	
10	2 050	1 010	670			1 382	342	2			1 909 924	472 644	2 764		
\bar{y}	668										4 123 960	1 351 212	255 436	255 436	1 351 212
		ϕ_0	ϕ_2	ϕ_{-1}	ϕ_1	ϕ_2									
		1	0,328	0,062	0,062	0,328									

Źródło: Opracowanie własne w Excelu, gdzie: $y^- \rightarrow \bar{y}$.

Dla $k = 1, 2$ nie stwierdzamy zanikania funkcji autokorelacji. Sprawdźmy, czy to zanikanie wystąpi dla różnic pierwszego rzędu:

$$\Delta y_t = y_t - y_{t-1}$$

Tabela 1.10.6

Różnice pierwszego rzędu

t	y_t	y_{t-1}	$y_t - y_{t-1}$
t	Z	Z(-1)	
1	-300		
2	280	-300	580
3	150	280	-130
4	820	150	670
5	-10	820	-830
6	1130	-10	1140
7	880	1130	-250
8	1010	880	130
9	670	1010	-340
10	2050	670	1380

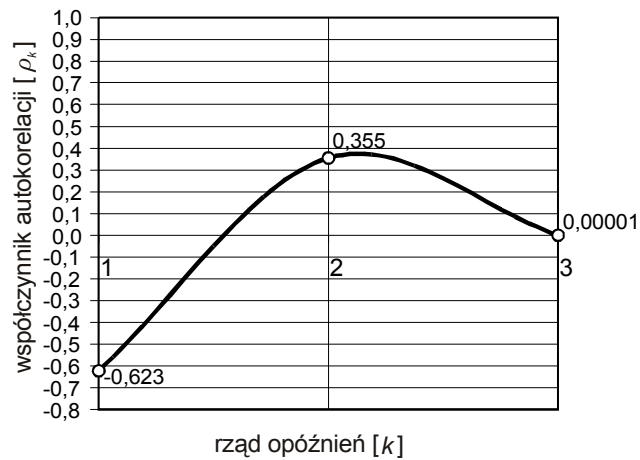
Źródło: Opracowanie własne.

Tabela 1.10.7

Obliczenia pomocnicze oraz ustalenie współczynników autokorelacji dla różnic pierwszego rzędu

t	$Z_t = (Y_t - Y_{t-1})$	Z_{t-2}	Z_{t-1}	Z_{t+1}	Z_{t+2}	$Z_t \cdot Z_{t-2}$	$Z_t \cdot Z_{t-1}$	$Z_t \cdot Z_{t+1}$	$Z_t \cdot Z_{t+2}$	$(Z_t - Z_t)^2$	$a \cdot b$	$a \cdot c$	$a \cdot d$	$a \cdot e$	Z_{t-3}	$Z_{t-3} \cdot Z_t$	$a \cdot f$
	Z	$Z(-2)$	$Z(-1)$	$Z(+1)$	$Z(+2)$	a	b	c	d	e					f		
1	580			-130	670	319				409							
2	-130		580	670	-830	-391	319	409	-1091	152 968		-124 721	-159 921	426 746			
3	670	580	-130	-830	1140	409	319	-391	-1 091	879	130 390	-159 921	-446 143	359 368			
4	-830	-130	670	1140	-250	-391	409	879	-511	1 190 523	426 746	-446 143	-958 965	557 679	580	319	-1 410
5	1140	670	-830	-250	130	879	409	-1 091	-511	-131	772 446	-958 965	-449 210	-115 232	-130	-391	1 270
6	-250	-830	1140	130	-340	-511	-1 091	879	-131	-601	261 235	-449 210	67 012	307 235	670	409	-920
7	130	1140	-250	-340	1 380	-131	879	-511	-601	1 119	17 190	-115 232	67 012	-146 699	-830	-1 091	960
8	-340	-250	130	1 380		-601	-511	-131	1 119	-261	361 335	78 812	-672 577		1 140	879	-1 480
9	1 380	130	-340			1 119	-131	-601	-261		1 251 912	-146 699	-672 577		-250	-511	1 630
Z	261										4 276 489	1 519 486	-2 665 712	1 519 486			50
											ρ_0	ρ_{-1}	ρ_1	ρ_2			ρ_{-3}
											1,000	-0,623	-0,623	0,355			0,00001

Źródło: Opracowanie własne w Excelu, gdzie: $\bar{z} \rightarrow \bar{z}$.



Źródło: Opracowanie własne w Excelu.

Wykres 1.10.7. Zależność współczynnika autokorelacji od rzędu opóźnienia $k = 1, 2, 3$ dla różnic pierwszego rzędu

Przy $k = 3$ nastąpiło prawie zaniknięcie funkcji autokorelacji, należałoby przeprowadzić dalsze badanie dla $k > 3$. W ćwiczeniach spróbujmy jednak szukać potwierdzenia stacjonarności, budując szereg różnic drugiego rzędu¹⁰⁹.

$$N_t = \Delta_2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

Tabela 1.10.8

Szereg różnic drugiego rzędu

	y_t	y_{t-1}	$y_t - y_{t-1}$	y_{t-2}	$y_{t-1} - y_{t-2}$	N_t
t	Z	Z(-1)	a	b	c	a - c
1	-300					
2	280	-300	580			
3	150	280	-130	-300	580	-710
4	820	150	670	280	-130	800
5	-10	820	-830	150	670	-1500
6	1130	-10	1140	820	-830	1970
7	880	1130	-250	-10	1140	-1390
8	1010	880	130	1130	-250	380
9	670	1010	-340	880	130	-470
10	2050	670	1380	1010	-340	1720
					N_t^-	100

Źródło: Opracowanie własne w Excelu, gdzie: $N_t^- \rightarrow \bar{N}_t$.

¹⁰⁹ Ibidem, s. 112.

Tabela 1.10.9

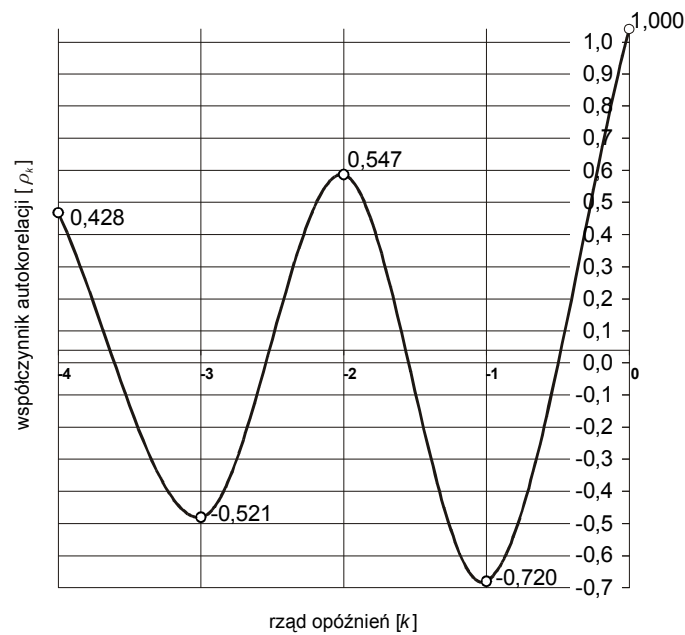
Obliczenia pomocnicze oraz ustalenie współczynników autokorelacji dla szeregu różnic drugiego rzędu

t	N_t	N_{t-3}	N_{t-2}	N_{t-1}	N_{t-4}	$N_t - N_t^*$	$N_{t-3} - N_t^*$	$N_{t-2} - N_t^*$	$N_{t-1} - N_t^*$	$N_{t-4} - N_t^*$	$(N_t - N_t^*)^2$				
	N	$N(-3)$	$N(-2)$	$N(-1)$	$N(-4)$	a	b	c	d	e		$a \cdot b$	$a \cdot c$	$a \cdot d$	$a \cdot e$
1	-710					-810					656 100				
2	800					700					490 000				
3	-1 500					-1 600					2 560 000				
4	1 970					1 870					3 496 900				
5	-1 390	800				-1 490	700	-1 600	1 870	-810	2 220 100	-1 043 000	2 384 000	-2 786 300	1 206 900
6	380	-1 500	1 970			280	-1 600	1 870	-1 490	700	78 400	-448 000	523 600	-417 200	196 000
7	-470	1 970	-1 390	380		-570	1 870	-1 490	280	-1 600	324 900	-1 065 900	849 300	-159 600	912 000
8	1 720	-1 390	380	-470	1 970	1 620	-1 490	280	-570	1 870	2 624 400	-2 413 800	453 600	-923 400	3 029 400
N^*	100										12 450 800	-6 485 400	6 815 500	-8 965 500	5 344 300
											ρ_0	ρ_{-3}	ρ_{-2}	ρ_{-1}	ρ_{-4}
											1,000	-0,521	0,547	-0,720	0,429

Źródło: Opracowanie własne w Excelu, gdzie $N^* \rightarrow \bar{N}$.

Ponieważ współczynniki k autokorelacji są symetryczne, obliczenia funkcji autokorelacji przeprowadzono dla rzędu opóźnień $k = 0-4$.

Z wykresu 1.10.8 wynika, że funkcja autokorelacji („fala”) zanika, co świadczy o stacjonarności szeregu czasowego drugich różnic.



Źródło: Opracowanie własne.

Wykres 1.10.8. Funkcja autokorelacji dla szeregu różnic drugiego rzędu

1.10.4. Pytania

1. Wymień i opisz etapy budowy modeli Boxa i Jenkinsa.
2. Co to jest funkcja autokorelacji i jak ją przedstawiamy graficznie?
3. Podaj i objaśnij wzór na funkcję autokorelacji.
4. Wyjaśnij pojęcia stacjonarność różnic pierwszego oraz drugiego rzędu.
5. Omów zastosowanie autokorelacji cząstkowej do wyznaczenia rzędu p autoregresji.
6. Na czym oparte jest badanie rzędu średniej ruchomej?

1.10.5. Zadania

1. Opady pyłu radioaktywnego [w microcurie na metr sześcienny powietrza] w określonym regionie USA w kolejnych 31 dniach maja roku 1964 wynosiły:

T a b e l a 1.10.10

Dane statystyczne

9,3	6,8	9,8	6,6	4,3	6,7	6,4	10,1	8,9	3,7	
5,3	6,5	7,4	8,3	4,6	7,9	6,5	5,1	7,2	8,7	
7,9	6,3	2,7	5,3	8,8	7,3	9,0	7,7	8,4	7,8	5,8

Ź r ó d ł o: Fragment danych z publikacji: J. F r e d, *Podstawy nowoczesnej statystyki*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1971, s. 33.

Przedstaw na wykresie punktowym poziomy (wartości) zmiennej dla $t = 31$ okresów (dni) i zaznacz wartość średnią. Oszacuj rodzaj występującego szeregu czasowego.

2. Dla powyższych danych oblicz współczynniki autokorelacji dla rzędów opóźnień $k = 1, 2, 3, 4$. Przedstaw wyniki graficznie jako funkcje autokorelacji. Oceń, czy występuje stacjonarność szeregu czasowego Y_t .

3. Jeśli wystąpi stacjonarność, oblicz autokorelacje cząstkowe. Sporządź wykres i oceń, czy występuje zanikanie funkcji autokorelacji cząstkowej.

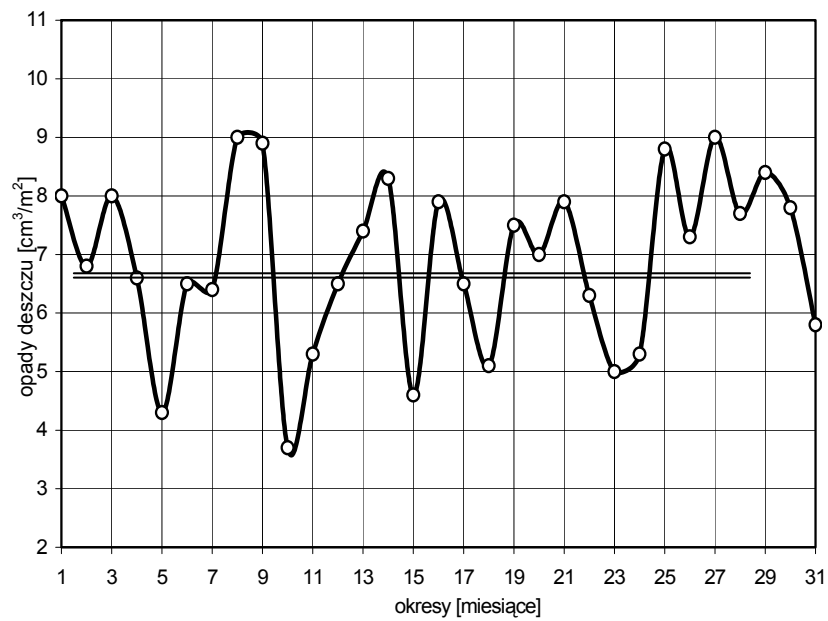
4. Scharakteryzuj podany wykres (1.10.9) punktowy poziomów zmiennej szeregu czasowego.

Tabela 1.10.11

Opady deszczu w cm^3/m^2

8	6,8	8	6,6	4,3	6,5	6,4	9	8,9	3,7		Σ	
5,3	6,5	7,4	8,3	4,6	7,9	6,5	5,1	7,5	7		68,2	
7,9	6,3	5	5,3	8,8	7,3	9	7,7	8,4	7,8	5,8	66,1	
											Σ	79,3
											Σ	213,6
											Średnia	6,8903

Źródło: Opracowanie własne; dane umowne.



Źródło: Opracowanie własne.

Wykres 1.10.9. Obserwacje opadów deszczu w miesiącach

1.11. Modele zmiennych dychotomicznych¹¹⁰

1.11.1. Idea uogólnionej metody najmniejszych kwadratów

Klasyczna metoda najmniejszych kwadratów (KMNK) oparta jest na założeniach jak:

- a) brak autokorelacji odchyłeń losowych,
- b) stałość wariancji odchyłeń losowych,
- c) nielosowy charakter zmiennych objaśniających.

Natomiast uogólniona metoda najmniejszych kwadratów (UMNK) stosowana jest do szacowania parametrów liniowego modelu ekonometrycznego, gdy niespełnione są założenia a) oraz b). Macierz wariancji i kowariancji dla tej metody ma postać:

$$\mathbf{D}^2(\varepsilon) = \sigma^2 \mathbf{V}.$$

gdzie:

\mathbf{V} – dowolna dodatnio określona macierz symetryczna stopnia n ,
 σ^2 – nieznaną stałą parametr.

Otrzymany według UMNK wektor ocen parametrów strukturalnych modelu ekonometrycznego jest następujący:

$$\mathbf{a} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}. \quad (1.11.1)$$

Jeśli występuje niestałość wariancji odchyłeń losowych, to macierz odwrotna \mathbf{V}^{-1} jest macierzą diagonalną:

$$\mathbf{V}^{-1} = \begin{bmatrix} \frac{1}{v_1} & 0 & \dots & 0 \\ 0 & \frac{1}{v_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{v_n} \end{bmatrix}.$$

¹¹⁰ Podrozdział opracowano na podstawie: E. Nowak, *Zarys metod ekonometrii. Zbiór zadań...*, op.cit., s. 109.

Jednym ze sposobów wyznaczenia elementów macierzy \mathbf{V} jest przyjęcie za v_t modułów reszt e_t modelu oszacowanego KMNK.

$$v_t = |e_t| \quad (t = 1, 2, \dots, n).$$

Gdy występuje autokorelacja odchyłeń losowych, najczęściej przyjmuje się, że ciąg składnika losowego $\{\varepsilon_t\}$ podlega procesowi autokorelacji pierwszego rzędu dla $t = 1, 2, \dots, n$

$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t,$$

gdzie:
 $|\rho| < 1.$

Wariancje odchyłeń losowych S_e^2 oraz macierz wariancji i kowariancji $\mathbf{D}^2(a)$ określa się według (1.11.2) oraz (1.11.3), przy czym k oznacza ilość zmiennych objaśniających.

$$S_e^2 = \frac{1}{n-k-1} \mathbf{e}^T \mathbf{V}^{-1} \mathbf{e}. \quad (1.11.2)$$

$$\mathbf{D}^2(a) = S_e^2 (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}. \quad (1.11.3)$$

We wzorze na wariancję odchyłeń losowych e oznacza wektor reszt modelu oszacowanego według UMNK. Pierwiastki z wartości na głównej przekątnej otrzymanej ze wzoru na $\mathbf{D}^2(a)$ macierzy wariancji i kowariancji stanowią standardowe błędy szacunku parametrów strukturalnych oszacowanego modelu.

1.11.2. Liniowy model prawdopodobieństwa (LPM)

Zmienną dychotomiczną nazywamy taką zmienną endogeniczną Y , która jest zmienną zero-jedynkową, określającą wystąpienie określonego zdarzenia lub wybór pomiędzy dwoma opcjami. Modele zmiennych dychotomicznych należą do metod statystycznych analizy jakościowej zjawiska ekonomicznego. Modele ekonometryczne z tego zakresu to:

- liniowy model prawdopodobieństwa,
- model logitowy,
- model probitowy.

W modelach tego typu jednostki tworzące próbę oprócz jednej cechy wspólnej, przykładowo zakup określonej marki samochodu w klasie standard lub z wyposażeniem dodatkowym, mają inne cechy jakościowe – miejsce zamieszkania, poziom wykształcenia itp., jak też cechy ilościowe. Głównym celem badania jest związek między zestawem cech ilościowych X , przykłado-

wo: ceną, pojemnością bagażnika, stawką roczną ubezpieczenia, a prawdopodobieństwem Y zakupu samochodu. Niech jeden oznacza, że dana osoba zakupiła wybrany produkt (TAK), a zero – przeciwnie (NIE).

Liniowymi modelami prawdopodobieństwa nazywamy te modele, które wyrażają dychotomiczną zmienną Y jako liniową funkcję zmiennych egzogenicznych (objaśniających) X .

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i. \quad (1.11.4)$$

Model taki stanowi szczególny przypadek regresji liniowej, do której można stosować klasyczną metodę najmniejszych kwadratów. Wartość oczekiwana $E(Y_i|X_i)$ stanowi prawdopodobieństwo warunkowe p_i – zajścia zdarzenia, np. zakup samochodu przy danej wartości x_i .

Wartość ta powinna zawierać się w przedziale (0–1). Przekształcając wzór 1.11.4, zmienna losowa i składnik losowy przyjmują¹¹¹:

$$Y_i = 1 \quad \varepsilon_i = 1 - \beta_1 - \beta_2 X_i \text{ z prawdopodobieństwem } P_i,$$

$$Y_i = 0 \quad \varepsilon_i = -\beta_1 - \beta_2 X_i \text{ z prawdopodobieństwem } (1-P_i).$$

Gdy z modelu wypadają wartości mniejsze od zera oraz $\leq 0,5$, to traktujemy je jako zero, a wszystkie wartości większe od 0,5, a także większe od jedynki, przyjmujemy jako jeden. Wartość graniczną stanowi 0,5. Tak więc w trakcie analizy wyników wszystkie prawdopodobieństwa mniejsze lub równe wartości granicznej traktuje się jako zero, wartości wyższe od 0,5 jako 1.

Podstawowym problemem LPM jest założenie, że prawdopodobieństwo wzrasta liniowo w relacji do X . W praktyce nie zawsze tak musi być, gdyż określony przyrost zmiennej X w jednej grupie osób wywoła inną reakcję niż u innej. Według M. Lipiec-Zajchowskiej¹¹² zależność między P_i a X_i podobna jest kształtem do dystrybuanty zmiennej losowej. Stosowane są dystrybuanty rozkładu normalnego oraz funkcji logistycznej.

1.11.3. Model logitowy (logit)

Logit jest liniową funkcją X , lecz prawdopodobieństwa już nie są liniowe tak jak w metodzie LPM:

$$L_i = \ln(P_i/(1-P_i)) = Z_i = \beta_1 + \beta_2 X_i. \quad (1.11.5)$$

L_i – współczynnik szansy (logit), czyli stosunek prawdopodobieństwa wystąpienia danego zdarzenia do prawdopodobieństwa jego niewystąpienia.

¹¹¹ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria...*, op.cit., rozdz. 4.

¹¹² Ibidem, rys. 4.2.

Do oszacowania parametrów modelu logitowego potrzebne są wartości X_i oraz logitów:

$L_i = \ln(1/0)$ – zajście określonego zdarzenia,

$L_i = \ln(0/1)$ – niezajście określonego zdarzenia.

Na potrzeby estymacji prawdopodobieństwa zastępuje się częstościami w próbie.

1.11.4. Model probitowy (probit)

Wyrażenie $F^{-1}P_i$ nazywa się probitem. Model probitowy wykorzystuje dystrybuantę rozkładu normalnego. Transformacja probitowa polega na przekształceniu danego prawdopodobieństwa na wartość dystrybuanty F standaryzowanego rozkładu normalnego¹¹³:

$$\text{Probit}_i = F^{-1}P_i = \beta_1 + \beta_2 X_i + \varepsilon_i. \quad (1.11.6)$$

Jeżeli p_i oznacza prawdopodobieństwo podjęcia inwestycji przez przedsiębiorcę przy rocznym zysku z działalności X_i , to, aby uniknąć wyników ujemnych, wartość otrzymaną z przekształcenia prawdopodobieństwa p_i na postać probitową powiększa się o 5. Model probitowy, zwany również normit, stosowany jest w zmiennych losowych o rozkładzie normalnym lub zbliżonym do normalnego. Parametry szacuje się KMNK.

Między wartościami logit i probit występuje zależność:

$$\text{Logit/Probit} = \frac{\pi}{\sqrt{3}} = 1,8. \quad (1.11.7)$$

1.11.5. Zastosowanie programu Excel do modelowania zmiennych dychotomicznych

P r z y k ł a d 1

Duży salon sprzedaży samochodów marki S przeprowadził badania na próbie 500 nabywców dużych samochodów osobowych. Rozpatrywano dwie opcje:

- wykonanie standardowe,
- wyposażenie dodatkowe.

W analizie brano pod uwagę:

- cenę,
- pojemność silnika,
- pojemność bagażnika,
- roczną stawkę ubezpieczenia pojazdu.

¹¹³ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonomia...*, op.cit., rozdz. 4.2.2.

Zastosujemy model LPM. Wprowadźmy oznaczenie N zmiennej objaśnianej:

$N = 1$ – nabywca, który zakupił samochód w wersji standardowej,

$N = 0$ – nabywca, który zakupił samochód z dodatkowym wyposażeniem.

Zmienne objaśniające:

C – cena samochodu w tys. zł. (standard + wyposażenie),

S – pojemność silnika w cm^3 ,

B – pojemność bagażnika w m^3 ,

U – roczna stawka ubezpieczenia w tys. zł.

Przyjmijmy założenie, że stawka ubezpieczenia jest proporcjonalna do pojemności silnika, przy czym minimalna wynosi 700 zł. Współczynnik wzrostu stawki ubezpieczenia w_i wynosi:

1 – do 1000 cm^3 ,

1,1 – do 1200 cm^3 ,

1,2 – do 1500 cm^3 ,

1,3 – powyżej 1500 cm^3 .

Przykładowy wzór na obliczenie stawki ubezpieczenia (dane umowne):

$$U_i = 700 + w_i \cdot 0,35 \cdot S,$$

gdzie:

0,35 – przelicznik pojemności silnika na stawkę ubezpieczenia.

Postać ogólna funkcji liniowej dla i -tego nabywcy jest następująca:

$$N_i = \beta_1 + \beta_2 \cdot C_i + \beta_3 \cdot S_i + \beta_4 \cdot B_i + \beta_5 \cdot U_i.$$

Symulowanie tego przykładu, znając N_i według zachowań nabywców w relacji z czterema zmiennymi objaśniającymi C , S , B , U , jest zbyt obszerne w ramach tej publikacji. Z tego względu przyjmijmy, że w wyniku zastosowania funkcji REGLINP Excela uzyskano parametry β_i i liniowy model prawdopodobieństwa LPM:

$$N_i = 0,7 + 0,004 \cdot C_i + 0,0005 \cdot S_i + 0,0002 \cdot B_i - 0,001 \cdot U_i.$$

Jeśli dane orientacyjne, np. dla samochodu Peugeot 206, będą:

$C = 35$ tys. zł,

$S = 1200$ cm^3 ,

$B = 0,375$ m^3 ,

$U = 1200$ zł,

to:

$$N_1 = 0,7 + 0,004 \cdot 35 + 0,0005 \cdot 1200 + 0,0002 \cdot 0,375 - 0,001 \cdot 1200 = 0,2475.$$

Ponieważ $N_1 < 0,5$, przyjmujemy 0, czyli nabywca zakupił samochód Peugeot 206 z wyposażeniem dodatkowym. W analizie poszczególnych jednostek badawczych (nabywców) posługujemy się probabilistycznymi krzywymi odpowiedzi sporządzonymi odrębnie dla każdej zmiennej objaśniającej modelu¹¹⁴. W naszym modelu wzrost ceny samochodu oraz pojemności silnika wpływa na zwiększenie prawdopodobieństwa zaku-

¹¹⁴ Ibidem, rozdz. 4.2.

pu samochodu z dodatkowym wyposażeniem. Wzrost pojemności bagażnika ma minimalny wpływ na prawdopodobieństwo, natomiast wzrost stawki ubezpieczenia w istotny sposób wpływa na zmniejszenie zakupu samochodu.

Spośród nabywców, którzy zakupili samochód z dodatkowym wyposażeniem, a więc zakwalifikowanych jako $N_i = 0$, możemy sporządzić ranking według malejącego prawdopodobieństwa zakupu samochodu o danych cechach ilościowych (C, S, B, U). Może to być wykorzystane w prognozowaniu ilości opcji sprzedawanych samochodów osobowych.

Przykład 2

Dane dotyczą sytuacji finansowej 112 branżowych hurtowni funkcjonujących w roku 2003, z podaniem liczby hurtowni, które upadły w roku 2005. Zastosujemy model probitowy do określenia prawdopodobieństwa upadku hurtowni, której współczynnik zadłużenia wobec dostawców wynosi 0,5, a wobec pracowników 0,45.

Tabela 1.11.1

Sytuacja finansowa hurtowni

Grupa	Współczynnik zadłużenia wobec dostawców X_1	Współczynnik zadłużenia wobec pracowników X_2	Liczba analizowanych hurtowni m	Liczba upadłych hurtowni y
1	0,1–0,2	0,1–0,3	12	8
2	0,2–0,4	0,4–0,5	20	12
3	0,4–0,6	0,3–0,4	35	10
4	0,6–0,8	0,5–0,6	20	15
5	0,8–1,0	0,7–0,8	25	14

Współczynnik zadłużenia wobec dostawców (pojęcie umowne) to iloraz sumy kwot niezapłaconych faktur oraz sumy kwot faktur ogółem na 31.12.2003 r. w horyzoncie *ex post* pięciu lat; współczynnik zadłużenia wobec pracowników (pojęcie umowne) to iloraz sumy niezrealizowanych wypłat oraz sumy należności pracownikom ogółem w roku 2003 również w tym horyzoncie czasowym.

Źródło: Opracowanie własne; dane umowne.

Wyznamy prawdopodobieństwo upadku finansowego hurtowni, której współczynnik zadłużenia wobec dostawców wynosi 0,5, a współczynnik zadłużenia wobec pracowników 0,45.

W obliczeniach jako podstawę teoretyczną przyjęto wzory¹¹⁵:

a) transformacja danego prawdopodobieństwa na wartości dystrybuanty F standaryzowanego rozkładu normalnego:

$$\text{Probit} = Pr = F(p_j) + 5,$$

(Jak już wspomniano stałą 5 dodajemy, aby uniknąć wartości ujemnych.)

¹¹⁵ R. C z y ż e w s k i, M. H u n d e r t, R. K l ó s k a, *Wybrane zagadnienia z prognozowania*, Economicus, Szczecin 2006.

b) częstość p_i występowania upadłości hurtowni w danej grupie jako iloraz liczby wystąpienia zdarzenia y_j oraz liczby wszystkich zdarzeń w danej grupie m_j :

$$p_j = y_j/m_j,$$

c) oszacowanie parametrów strukturalnych a_i modelu poprzez zastosowanie uogólnionej metody najmniejszych kwadratów:

$$\mathbf{a} = (\mathbf{X}^T \cdot \mathbf{V}^{-1} \mathbf{X})^{-1} \cdot \mathbf{X}^T \mathbf{V}^{-1} \cdot \mathbf{Pr}. \quad (1.11.8)$$

gdzie:

\mathbf{X} – macierz parametrów obserwacji zmiennych objaśniających (łącznie z parametrem równym 1 dla X_0 ,

\mathbf{Pr} – wektor zawierający wartości probitów,

\mathbf{V} – macierz, w której elementy głównej przekątnej obliczono z poniższego wzoru, a pozostałe elementy są równe zero.

$$\mathbf{V} = \frac{p_j \cdot (1 - p_j)}{m_j \cdot F(p_j)} = \frac{p_j \cdot (1 - p_j)}{m_j \cdot (Pr)^2}.$$

Wyznaczenie probitów Pr (dla zmiennych podano wartości środkowe)

T a b e l a 1.11.2

Obliczenie probitów

x_{1j}	x_{2j}	m_j	y_j	p_j	Pr
0,15	0,2	12	8	0,67	5,44
0,3	0,45	20	12	0,60	5,26
0,5	0,35	35	10	0,29	4,45
0,7	0,55	20	15	0,75	5,67
0,9	0,75	25	14	0,56	5,15

Ź r ó d ł o: Opracowanie własne w Excelu.

Wartości probitów wyznaczono na podstawie wzoru $Pr = F(p_j) + 5$. Dla funkcji $F(p_j)$ według dystrybuanty rozkładu normalnego (tab. VI, *Aneks* niniejszego opracowania) odnajdujemy przy określonym p_j wartość u . Na przykład dla $p_j = 0,67$ $u = 0,44$.

Jeżeli dodamy stałą 5, otrzymamy probit 5,44. Natomiast dla prawdopodobieństwa $< 0,5$, np. 0,29, obliczamy wpierw różnicę $(0,5 - 0,29) = 0,21$. Znajdujemy w tablicy wartość $u = 0,55$ według $(0,5 + 0,21) = 0,71$. Ponieważ wartość dystrybuanty 0,29 występuje przed zerem, tj. po stronie wartości ujemnych, to $u = -0,55$. Zatem probit $Pr = -0,55 + 5 = 4,45$.

Wygodniej jest skorzystać z tab. 1.11.3, stanowiącej przekształcenie prawdopodobieństw na probity.

Tabela 1.11.3

Zamiana prawdopodobieństw na probity

p_j	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0		2,67	2,95	3,12	3,25	3,36	3,45	3,52	3,59	3,66
0,1	3,72	3,77	3,82	3,87	3,92	3,96	4,01	4,05	4,08	4,12
0,2	4,16	4,19	4,23	4,26	4,29	4,33	4,36	4,39	4,44	4,45
0,3	4,48	4,50	4,53	4,56	4,59	4,61	4,64	4,67	4,69	4,72
0,4	4,75	4,77	4,80	4,82	4,85	4,87	4,90	4,92	4,95	4,97
0,5	5,00	5,03	5,05	5,08	5,10	5,13	5,15	5,18	5,20	5,23
0,6	5,25	5,28	5,31	5,33	5,36	5,39	5,41	5,44	5,47	5,50
0,7	5,52	5,55	5,58	5,61	5,64	5,67	5,71	5,74	5,77	5,81
0,8	5,84	5,88	5,92	5,95	5,99	6,04	6,08	6,13	6,18	6,23
0,9	6,28	6,34	6,41	6,48	6,55	6,64	6,75	6,88	7,05	7,33
	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009
0,99	7,33	7,37	7,41	7,46	7,51	7,58	7,65	7,75	7,88	8,09

Źródło: Opracowanie własne a podstawie: M. Cieślak (red.), *Prognozowanie gospodarcze*, Wydawnictwo Naukowe PWN, Warszawa 2005, tab. 5.1.

Obliczenia pomocnicze v_j oraz $1/v_j$ elementów macierzy V wykonano w Excelu.

Tabela 1.11.4

Elementy macierzy diagonalnej V

m_j	p_j	Pr	v_j	$1/v_j$
12	0,67	5,44	0,00062	1606,1655
20	0,6	5,26	0,00043	2305,6333
35	0,29	4,45	0,00030	3366,1365
20	0,75	5,67	0,00029	3429,2160
25	0,56	5,15	0,00037	2691,0004

Źródło: Opracowanie własne.

Macierz diagonalna V

$$V = \begin{bmatrix} 0,00062 & 0 & 0 & 0 & 0 \\ 0 & 0,00043 & 0 & 0 & 0 \\ 0 & 0 & 0,00030 & 0 & 0 \\ 0 & 0 & 0 & 0,00029 & 0 \\ 0 & 0 & 0 & 0 & 0,00037 \end{bmatrix}$$

Macierz odwrotna V^{-1} (według własności macierzy diagonalnej)

$$V^{-1} = \begin{bmatrix} 1606,1655 & 0 & 0 & 0 & 0 \\ 0 & 2305,6333 & 0 & 0 & 0 \\ 0 & 0 & 3366,1365 & 0 & 0 \\ 0 & 0 & 0 & 3429,2160 & 0 \\ 0 & 0 & 0 & 0 & 2691,0004 \end{bmatrix}.$$

Macierz X oraz X^T zmiennych objaśniających, a w tym: X_0 – parametr stały, X_1 – współczynnik zadłużenia wobec dostawców, X_2 – współczynnik zadłużenia wobec pracowników:

$$X = \begin{bmatrix} 1 & 0,15 & 0,2 \\ 1 & 0,3 & 0,45 \\ 1 & 0,5 & 0,35 \\ 1 & 0,7 & 0,55 \\ 1 & 0,9 & 0,75 \end{bmatrix}, \quad X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0,15 & 0,3 & 0,5 & 0,7 & 0,9 \\ 0,2 & 0,45 & 0,35 & 0,55 & 0,75 \end{bmatrix}.$$

Wektor Pr i wektor parametrów a :

$$Pr = \begin{bmatrix} 5,44 \\ 5,26 \\ 4,45 \\ 5,67 \\ 5,15 \end{bmatrix}, \quad a = (X^T \cdot V^{-1} X)^{-1} \cdot X^T V^{-1} \cdot Pr.$$

Obliczenia parametrów a_0, a_1, a_2 z zastosowaniem rachunku macierzowego i funkcji Excela są następujące:

$$a = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0,15 & 0,3 & 0,5 & 0,7 & 0,9 \\ 0,2 & 0,45 & 0,35 & 0,55 & 0,75 \end{bmatrix} \begin{bmatrix} 1606,1655 & 0 & 0 & 0 & 0 \\ 0 & 2305,6333 & 0 & 0 & 0 \\ 0 & 0 & 3366,1365 & 0 & 0 \\ 0 & 0 & 0 & 3429,2160 & 0 \\ 0 & 0 & 0 & 0 & 2691,0004 \end{bmatrix} \begin{bmatrix} 1 & 0,15 & 0,2 \\ 1 & 0,3 & 0,45 \\ 1 & 0,5 & 0,35 \\ 1 & 0,7 & 0,55 \\ 1 & 0,9 & 0,75 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 1606,1655 & 0 & 0 & 0 & 0 \\ 0 & 2305,6333 & 0 & 0 & 0 \\ 0 & 0 & 3366,1365 & 0 & 0 \\ 0 & 0 & 0 & 3429,2160 & 0 \\ 0 & 0 & 0 & 0 & 2691,0004 \end{bmatrix} \begin{bmatrix} 5,44 \\ 5,26 \\ 4,45 \\ 5,67 \\ 5,15 \end{bmatrix}.$$

Macierz iloczynu ($\mathbf{X}^T \mathbf{V}^{-1}$) określono przy zastosowaniu funkcji Excela MACIERZ.ILOCZYN(...):

$$(\mathbf{X}^T \mathbf{V}^{-1}) = \begin{bmatrix} 1606,166 & 2305,633 & 3366,137 & 3429,216 & 2691,000 \\ 240,925 & 691,690 & 1683,068 & 2400,451 & 2421,900 \\ 321,233 & 1037,535 & 1178,148 & 1886,068 & 2018,250 \end{bmatrix}.$$

Macierz iloczynu ($\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}$):

$$(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}) = \begin{bmatrix} 13398,152 & 7438,035 & 6441,235 \\ 7438,035 & 4945,206 & 4085,193 \\ 6441,235 & 4085,193 & 3494,515 \end{bmatrix}.$$

Macierz odwrotną ($(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$) określono przy zastosowaniu funkcji MACIERZ.ODW(...):

$$(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} = \begin{bmatrix} 0,00068 & 0,00037 & -0,00168 \\ 0,00037 & 0,00610 & -0,00781 \\ -0,00168 & -0,00781 & 0,01251 \end{bmatrix}.$$

Iloczyn macierzy ($\mathbf{X}^T \mathbf{V}^{-1} \mathbf{Pr}$):

$$(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{Pr}) = \begin{bmatrix} 69146,7857 \\ 38521,9193 \\ 33535,6988 \end{bmatrix}.$$

Iloczyn macierzy do określenia parametrów \mathbf{a} :

$$(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \cdot \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Pr} = \begin{bmatrix} 4,719 \\ -1,466 \\ 2,613 \end{bmatrix}.$$

Otrzymaliśmy parametry:

$\mathbf{a}_0 = 4,719$; $\mathbf{a}_1 = -1,466$; $\mathbf{a}_2 = 2,613$,

a model probitowy jest następujący:

$$\hat{P}_r = 4,719 - 1,466 \cdot X_{1j} + 2,613 \cdot X_{2j}.$$

Prawdopodobieństwo upadku firmy (przy $x_1 = 0,5$ $x_2 = 0,45$):

$$\hat{P}_r = 4,719 - 1,466 \cdot 0,5 + 2,613 \cdot 0,45 = 5,1619.$$

Z tabeli 1.11.3 dla probitu 5,1619 odczytujemy prawdopodobieństwo 0,57 upadku hurtowni o współczynniku zadłużenia wobec dostawców $x_1 = 0,5$ oraz

współczynnika zadłużenia wobec pracowników $x_2 = 0,45$. Ponieważ $p_i > 0,5$, możemy założyć, że hurtownia o danych cechach zadłużenia upadnie.

Przykład 3

Rozwiążmy to samo zadanie, wykorzystując model logitowy. Interesuje nas również prawdopodobieństwo upadku hurtowni o cechach $x_1 = 0,5$ oraz $x_2 = 0,45$.

Dane jak w tab. 1.11.2 opisujące przykład z modelem probitowym.

Tabela 1.11.5

Dane do modelu logitowego

Grupa	Współczynnik zadłużenia wobec dostawców x_1	Współczynnik zadłużenia wobec pracowników x_2	Liczba analizowanych hurtowni m	Liczba upadłych hurtowni y
1	0,15	0,2	12	8
2	0,3	0,45	20	12
3	0,5	0,35	35	10
4	0,7	0,55	20	15
5	0,9	0,75	25	14

Źródło: Opracowanie własne.

Obliczenia pomocnicze (a–c) wykonano według wzorów¹¹⁶:

a) wyrażenie prawdopodobieństwa w formie logitu:

$$L_i = \ln \frac{p_j}{1 - p_j},$$

b) częstość p_j występowania badanego zjawiska stanowi iloraz liczby zdarzeń, w którym dane zjawisko wystąpiło, i liczby wszystkich zdarzeń w danej kategorii:

$$p_j = y_j/m_j,$$

c) oszacowanie wyrażenia na \mathbf{a} dla określenia parametrów strukturalnych modelu logitowego, stosując uogólnioną metodę najmniejszych kwadratów:

$$\mathbf{a} = (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \cdot \mathbf{X}^T \mathbf{W}^{-1} \mathbf{L}, \quad (1.11.9)$$

\mathbf{X} – macierz parametrów obserwacji zmiennych objaśniających (łącznie z parametrem = 1 dla X_0),

\mathbf{L} – wektor zawierający wartości logitów,

\mathbf{W} – macierz o elementach na głównej przekątnej określonych według wzoru:

¹¹⁶ R. C z y ż e w s k i, M. H u n d e r t, R. K l ó s k a, *Wybrane zagadnienia z prognozowania...*, op.cit., rozdz. III.2.

$$w_j = \frac{1}{m_j p_j \cdot (1 - p_j)}$$

– wyznaczenie logitów (przykładowe wyrażenie w Excelu na obliczenie logitu [=LN(C3/(1-C3))]):

Tabela 1.11.6

Obliczenie logitów

x_{1j}	x_{2j}	m_j	y_j	p_j	L_j
0,15	0,2	12	8	0,67	0,7082
0,3	0,45	20	12	0,60	0,4055
0,5	0,35	35	10	0,29	-0,8954
0,7	0,55	20	15	0,75	1,0986
0,9	0,75	25	14	0,56	0,2412

Źródło: Opracowanie własne w Excelu.

– określenie elementów macierzy **W** (wyrażenie w Excelu na element w_j , np. [=1/(B16*C16*(1-C16))])

Tabela 1.11.7

Wyznaczenie elementów „ w_i ” oraz („ $1/w_j$ ”)

A	B	C	D	E
15	m_j	p_j	w_j	$1/w_j$
16	12	0,67	0,3769	2,6532
17	20	0,60	0,2083	4,8000
18	35	0,29	0,1388	7,2065
19	20	0,75	0,2667	3,7500
20	25	0,56	0,1623	6,1600

Źródło: Opracowanie własne w Excelu.

Macierz diagonalna **W**:

$$W = \begin{bmatrix} 0,3769 & 0 & 0 & 0 & 0 \\ 0 & 0,2083 & 0 & 0 & 0 \\ 0 & 0 & 0,1388 & 0 & 0 \\ 0 & 0 & 0 & 0,2667 & 0 \\ 0 & 0 & 0 & 0 & 0,1623 \end{bmatrix}$$

Macierz odwrotna \mathbf{W}^{-1} :

$$\mathbf{W}^{-1} = \begin{bmatrix} 2,6532 & 0 & 0 & 0 & 0 \\ 0 & 4,8000 & 0 & 0 & 0 \\ 0 & 0 & 7,2065 & 0 & 0 \\ 0 & 0 & 0 & 3,7500 & 0 \\ 0 & 0 & 0 & 0 & 6,1600 \end{bmatrix}.$$

Macierz \mathbf{X} :

$$\mathbf{X} = \begin{bmatrix} 1 & 0,15 & 0,2 \\ 1 & 0,3 & 0,45 \\ 1 & 0,5 & 0,35 \\ 1 & 0,7 & 0,55 \\ 1 & 0,9 & 0,75 \end{bmatrix}.$$

Wektor \mathbf{L} :

$$\mathbf{L} = \begin{bmatrix} 0,7082 \\ 0,4055 \\ -0,8954 \\ 1,0986 \\ 0,2412 \end{bmatrix}.$$

Wymienione macierze zapisano w arkuszu kalkulacyjnym Excela.

Wektor parametrów (zastosowanie rachunku macierzowego funkcjami Excela):

$$\mathbf{a} = (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \cdot \mathbf{X}^T \mathbf{W}^{-1} \mathbf{L}.$$

Obliczenie parametrów \mathbf{a}_0 , \mathbf{a}_1 , \mathbf{a}_2 .

Przed macierzami podano przykłady wyrażeń Excela, przy czym w nawiasach wpisywane są odpowiednie obszary adresów arkusza kalkulacyjnego. Funkcja poprzedzona jest znakiem „=”.

=TRANSPONUJ(...):

$$\mathbf{X}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0,15 & 0,30 & 0,50 & 0,70 & 0,90 \\ 0,20 & 0,45 & 0,35 & 0,55 & 0,75 \end{bmatrix}.$$

=MACIERZ.ILOCZYN(...):

$$(\mathbf{X}^T \mathbf{W}^{-1}) = \begin{bmatrix} 2,6532 & 4,8000 & 7,2065 & 3,7500 & 6,1600 \\ 0,3980 & 1,4400 & 3,6033 & 2,6250 & 5,5440 \\ 0,5303 & 2,1600 & 5,5223 & 2,0625 & 4,6200 \end{bmatrix}.$$

=MACIERZ.ILOCZYN(...):

$$(\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X}) = \begin{bmatrix} 24,5697 & 13,6102 & 11,8954 \\ 13,6102 & 9,1204 & 7,5905 \\ 11,8954 & 7,5905 & 6,5603 \end{bmatrix}.$$

=MACIERZ.ODW(...):

$$(\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} = \begin{bmatrix} 0,3412 & 0,1546 & -0,7975 \\ 0,1546 & 3,0288 & -3,7848 \\ -0,7975 & -3,7848 & 5,9776 \end{bmatrix}.$$

=MACIERZ.ILOCZYN(...):

$$(\mathbf{X}^T \mathbf{W}^{-1} \mathbf{L}) = \begin{bmatrix} 2,9782 \\ 1,8605 \\ 2,3734 \end{bmatrix}.$$

=MACIERZ.ILOCZYN(...):

$$(\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W}^{-1}) \mathbf{L} = \begin{bmatrix} -0,5891 \\ -2,8875 \\ 4,7710 \end{bmatrix}.$$

Model logitowy:

$$Y_{L_j} = -0,5891 - 2,8875 x_{1j} + 4,7710 x_{2j}.$$

Wartość logitu dla: $x_1 = 0,5$; $x_2 = 0,45$:

$$L = -0,5891 - 2,8875 \cdot 0,5 + 4,7710 \cdot 0,45 = 0,1141.$$

Niech wynik będzie zapisany przykładowo w komórce B79.

Prawdopodobieństwo upadku firmy:

$$L = \ln [p_j / (1 - p_j)] \quad e^L = p_j / (1 - p_j) \quad e^{-L} = (1 - p_j) / p_j = 1/p_j - 1 \quad 1 + e^{-L} = 1/p_j \\ p_j = 1 / [1 + e^{-L}] \quad p_j = 1 / [1 + \exp(-L)]$$

Wyrażenie Excela: [=1/(1+EXP(B79))] → $p_j = 0,4715$.

Funkcja [EXP(...)] oblicza wartość $e = 2,718282$ podniesioną do potęgi określonej przez podaną liczbę.

Prawdopodobieństwo upadku hurtowni o współczynniku zadłużenia wobec dostawców $x_1 = 0,5$ oraz współczynniku zadłużenia wobec pracowników $x_2 = 0,45$ wynosi 0,4715, a więc wynik naszego przykładu jest o 10% mniejszy niż w modelu probitowym.

1.11.6. Pytania

1. Wymień oraz podaj główne cechy modelu ekonometrycznego przy zmiennych dychotomicznych.
2. Jak traktujemy wartości y_i w modelu LPM?
3. Co nazywamy logitem?
4. Wyjaśnij zależność $Pr = F(p_j) + 5$.
5. Opisz elementy estymatora:

$$\mathbf{a} = [\mathbf{X}^T \cdot \mathbf{V}^{-1} \mathbf{X}]^{-1} \cdot \mathbf{X}^T \mathbf{V}^{-1} \cdot \mathbf{Pr}.$$

6. Do jakiego estymatora stosujemy elementy macierzy obliczane według wzoru:

$$w_j = \frac{1}{m_j p_j \cdot (1 - p_j)}.$$

1.11.7. Zadania

1. Dany jest liniowy model prawdopodobieństwa LPM, o zmiennych $\bar{x}_1, \bar{x}_2, \bar{x}_3$:

$$\hat{Y} = 7 - 0,08 \cdot X_1 - 0,005 \cdot X_2 + 0,003 \cdot X_3.$$

Określ wartość funkcji tego modelu dla: $x_1 = 30, x_2 = 1100, x_3 = 0,3$.

Jak zinterpretujesz wynik przy założeniu:

$Y = 1$ – klient nabył skuter,

$Y = 0$ – klient nie zakupił skutera.

2. Wyznacz prawdopodobieństwa oraz probity Pr do podanych wartości zmiennych X_1, X_2 , liczby obserwacji obiektu m oraz wystąpienia określonego zjawiska Y w danej kategorii.

Tabela 1.11.8

Rozplanowanie wstępne do obliczenia probitów

x_{1j}	x_{2j}	m_j	y_j	p_j	Pr
0,1	0,2	15	7		
0,3	0,4	20	15		
0,6	0,3	30	10		
0,7	0,5	25	15		
0,8	0,7	25	18		

Źródło: Opracowanie własne; dane umowne.

3. Wyznacz prawdopodobieństwa i logity do podanych zmiennych X_1, X_2 , liczbę przebadanych pacjentów m oraz liczbę stwierdzenia choroby y w badaniach okresowych pracowników poszczególnych wydziałów huty bieli cynkowej.

Tabela 1.11.9

Rozplanowanie wstępne do obliczenia logitów

x_{1j}	x_{2j}	m_j	y_j	p_j	L_j
0,18	0,25	120	10		
0,32	0,40	200	9		
0,45	0,38	350	15		
0,73	0,57	180	5		
0,89	0,70	250	3		

Źródło: Opracowanie własne; dane umowne.

1.12. Zastosowanie modelu regresji oraz modeli opartych na szeregach czasowych w prognozowaniu (użycie Excela)

Wypracowane przez ekonometrię modele regresji oraz trendu wykorzystywane są między innymi w prognozowaniu wartości zmiennych. W niniejszym rozdziale podejmiemy próbę omówienia na przykładach tej tematyki, ograniczając się do:

- zastosowania modelu regresji,
- wykorzystania modeli trendów,
- użycia modeli wyrównywania wykładniczego,
- wykorzystania klasycznego modelu autoregresji,
- modeli Holta-Wintersa.

Występuje tu także wstępne wprowadzenie do modelowania symulacyjnego działalności gospodarczej. Rozdział kończy się pytaniami oraz przykładami zadań do samodzielnego rozwiązywania ze wspomaganie mikrokomputerowym.

W początkowej części rozdziału jest wprowadzenie w problematykę prognozowania gospodarczego, klasyfikacja rodzaju prognoz, wskazanie metod prognozowania oraz podano miary błędów prognoz. Omówiono procedurę projektowania i weryfikacji liniowego modelu regresji na potrzeby prognozowania. Zamieszczono przykład modelowania oraz określenia prognozy z zakresu produkcji przedsiębiorstwa.

Następnie poruszono zagadnienie modeli nieliniowych oraz pokazano testy porównywania modeli, w tym kryterium Schwarza. Zamieszczono przykład prezentujący metodę najmniejszych kwadratów na określenie parametrów modelu z wieloma zmiennymi objaśniającymi na wyznaczenie prognoz z użyciem analizy współczynników korelacji w doborze zmiennych objaśniających.

Prognozowanie to przede wszystkim obserwacje w czasie zjawisk i wielkości społeczno-gospodarczych. Nie mogło więc zabraknąć przykładów modeli opartych na szeregach czasowych, a więc trendu pełzającego, modeli średnich ruchomych, aż po modele wyrównywania wykładniczego włącznie. W ramach przykładów występuje obliczenia błędów *ex post* i *ex ante*, a także określenie prognozy kolejnego miesiąca z równania modelu adaptacyjnego.

Obserwowane wielkości w procesach gospodarki rynkowej wykazują sezonowość, np. w ciągu roku. Z tego względu podano użycie modeli Holt-Wintersa z sezonowością multiplikatywną i addytywną. Czasem w prognozowaniu wygodne jest zastosowanie szeregu czasowego tej samej zmiennej, lecz z przesunięciem o okres lub kilka okresów. Celowo więc pokazano przykład z klasycznym modelem autoregresyjnym.

Potrzeby między innymi prognozowania, jak też wspierały rozwój nauki, jaką jest ekonometria, sprawiły, że zbudowano konstrukcje matematyczne coraz to nowszych doskonalszych modeli ekonometrycznych ze wspomaganie techniką komputerową. Liczne modele dla współczesnej gospodarki znajdzie Czytelnik w klasycznym już podręczniku G.S. Maddali¹¹⁷.

Modelowanie symulacyjne, do którego nawiązano w niniejszym rozdziale, to wejście w nowoczesną obecnie problematykę symulacji stosującą złożone modele ekonometryczne z zastosowaniem komputera. Treningi na podstawie tych modeli przeprowadzane są już przez niektóre ośrodki w procesie dydaktycznym doskonalenia przyszłych i obecnych menedżerów firm. Dla zainteresowanych tą problematyką wyszczególniono oprogramowane modele symulacyjne do zarządzania produkcją.

W zamieszczonych przykładach zastosowano program WinQSB oraz funkcje Excela. Prognozowanie utworowało sobie drogę jako samodzielna dyscyplina nauki. Prognozowanie, zwłaszcza gospodarcze, określa bowiem kierunki rozwoju oraz umożliwia wskazanie przyszłych wielkości ekonomicznych. Opiera się na danych statystycznych wykorzystywanych między innymi przez modele ekonometryczne. Obszerne informacje na temat metod i zastosowania prognozowania gospodarczego znajdziemy w publikacji M. Cieślak¹¹⁸.

Prognozy możemy podzielić ze względu na¹¹⁹:

– horyzont czasowy prognoz: krótkoterminowe, średnioterminowe, długoterminowe – strategiczne wielowariantowe (dla podejmowania długofalowych decyzji gospodarczych), operatywne (wykonanie planu, kształtowanie się zmiennych ekonomicznych),

– charakter i strukturę prognoz: proste i złożone, ilościowe – opisywane zmienną ilościową i jakościowe, punktowe i przedziałowe, jednorazowe i powtarzalne,

– stopień szczegółowości prognoz (ogólne i szczegółowe),

¹¹⁷ G.S. Maddala, *Ekonometria. Metody...*, op.cit.

¹¹⁸ M. Cieślak (red.), *Prognozowanie gospodarcze...*, op.cit.

¹¹⁹ Ibidem, rozdz. 1.3; M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria...*, rozdz. 1.11.2.

- zasięg terenowy prognoz (światowe, międzynarodowe, krajowe, regionalne),
- metody opracowania (modelowe, indukcyjne, dedukcyjne, średnie),
- cel lub funkcję: ostrzegawcze, badawcze, aktywne, pasywne, samorealizujące się (np. ogłoszenie wzrostu cen wywołuje wykup towarów, a w konsekwencji dalszy wzrost cen), unicestwiający się (np. zapowiedź rekordowego napływu gości wywołuje zniechęcenie ich do przyjazdu).

Zmiany ilościowe w prognozowaniu określamy zgodnie z trendem lub funkcją regresji dla kolejnych okresów. Natomiast zmiany jakościowe dotyczą zmiany postaci trendu, charakteru związków między zmiennymi czy też zaniku niektórych zależności.

W prognozowaniu występuje pojęcie „predykcja”, które obejmuje diagnozowanie przeszłości, a więc estymację parametrów i sformułowanie modelu, określenie przyszłości – prognozowanie, stosując określoną metodę. Jeśli mówimy o kompleksowości predykcji, to dotyczy ona wnioskowania o wektory zmiennych reprezentujących współzależne zjawiska. Natomiast sekwencyjność predykcji oznacza budowę prognoz dla danej zmiennej w następujących po sobie okresach.

Wymieńmy teraz metody prognozowania¹²⁰:

- analizy i prognozowania krótkookresowego szeregów czasowych (czas, przeszłe wartości zmiennej prognozowanej); stosujemy tu modele: średnie ruchome, wygładzanie wykładnicze, autoregresyjne, inne,
- prognozowania przyczynowo-skutkowego z zastosowaniem modeli: ekonometryczne jedno-, wielorównaniowe, behawioralne (według praw zachowania ekonomicznego firm), symptomatyczne, gdy brak teorii do budowy modelu lub możliwości pozyskania wartości szeregów zmiennych; metody prognozowania przyczynowo-skutkowego stosujemy, gdy znane są przyszłe wartości zmiennych objaśniających lub istnieje możliwość prognozowanie tych zmiennych, a ponadto do eksperymentowania – symulacji na przypuszczalnych wartościach zmiennych objaśniających,
- analogowe (bazowanie na zmiennych podobnych, np. dalszy rozwój telefonii komórkowej w Polsce na podstawie wyższego poziomu w krajach zachodnich),
- heurystyczne (jedną z metod jest „burza mózgów” na temat nowych odkryć naukowych w określonej dziedzinie, z pozyskiwaniem opinii wielu ekspertów z różnych środowisk naukowych, następnie przeprowadzenie analizy poglądów i wybór według największego prawdopodobieństwa),

¹²⁰ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria...*, rozdz. 1.11.4.

– kombinowane (średnia prognoza z różnych modeli z zastosowaniem wag zaufania do tych metod określonych przez ekspertów).

O wyborze określonej metody decyduje dostępność danych i właściwości metody.

Określone prognozy w okresach minionych możemy ocenić *ex post*, stosując następujące miary¹²¹:

– pierwiastek średniego kwadratu błędu:

$$RMSE = \sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} (\hat{y}_n - y_n)^2}, \quad (1.12.1)$$

gdzie:

$S+h$ – zakres próby P , dla której oceniamy wygasłe już prognozy,

$n=S$ – miejsce startu – obserwacja n -ta,

h – horyzont próby poza obserwacją n -tą,

y_n – wartości rzeczywiste,

\hat{y}_n – wartości prognozowane.

– średni absolutny błąd:

$$MAE = \frac{1}{1+h} \sum_{n=S}^{S+h} |\hat{y}_n - y_n|, \quad (1.12.2)$$

– średni procentowy absolutny błąd:

$$MAPE = \frac{1}{1+h} \sum_{n=S}^{S+h} \left| \frac{\hat{y}_n - y_n}{y_n} \right| \cdot 100, \quad (1.12.3)$$

– współczynnik Theila:

$$THEIL = \frac{\sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} (\hat{y}_n - y_n)^2}}{\sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} \hat{y}_n^2} + \sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} y_n^2}}. \quad (1.12.4)$$

Dekompozycja współczynnika Theila ($U^M + U^S + U^C = 1$):

– współczynnik średniego błędu U_M oraz współczynnik błędu wariancji U_S :

¹²¹ Ibidem, rozdz. 11.11.5.

$$U^M = \frac{(\hat{y} - \bar{y})^2}{\sum (\hat{y}_n - y_n)^2 / h}, \quad U^S = \frac{(S_{\hat{y}} - S_y)^2}{\sum (\hat{y}_n - y_n)^2 / h},$$

– współczynnik błędu kowariancji:

$$U^C = \frac{2 \cdot (1-r) \cdot S_{\hat{y}} \cdot S_y}{\sum (\hat{y}_n - y_n)^2 / h}.$$

W wymienionych wzorach:

S – odchylenia standardowe w odniesieniu do prognozy \hat{y} i wartości średniej \bar{y} ,

r – współczynnik korelacji pomiędzy wartościami modelowymi a empirycznymi z obserwacji,

U^C – miara błędu niesystematycznego.

Gdy $U^M > 0,2$, to występuje błąd systematyczny prognozy. Jeśli U^S ma wysoką wartość, wówczas zmienność zmiennej prognozowanej jest niedostateczna.

1.12.1. Błąd prognozy *ex ante* oraz badanie aktualności modelu

Błąd prognozy *ex ante* jest miarą dokładności statystycznej. Wyróżniamy: względny błąd prognozy *ex ante* i bezwzględny błąd prognozy *ex ante*.

Względny błąd prognozy *ex ante*, w %, obliczamy według wzoru z publikacji M. Cieślak, przy czym zamiast symbolu (*) zastosowano (^) w odniesieniu do zmiennej prognozowanej¹²²:

$$\eta_t = \frac{v_t}{\hat{y}_t} \cdot 100,$$

$t > n$.

Bezwzględny błąd prognozy *ex ante* dla modelu trendu liniowego¹²³:

$$v_t = \sqrt{\frac{(T - \bar{t})^2}{n} + \frac{1}{n} + 1} \cdot S_u,$$

gdzie:

T – numer okresu, dla którego wyznacza się prognozę,

\bar{t} – średnia wartość zmiennej czasowej,

S_u – odchylenie standardowe reszt, stanowiące pierwiastek z poniższej wariancji resztowej:

¹²² M. Cieślak (red.), *Prognozowanie gospodarcze. Metody...*, op.cit., rozdz. 2.4.3.

¹²³ Ibidem, s. 56.

$$S^2_u = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n - k}, \quad S_u = \sqrt{S^2_u},$$

gdzie:

$k = m + 1$,

m – ilość zmiennych objaśniających,

y_t – wartość zmiennej Y w okresie t ,

\hat{y}_t – modelowa (teoretyczna) wartość zmiennej Y ,

\bar{y} – średnia wartość zmiennej Y dla szeregu czasowego o n obserwacjach.

Wymienione wzory są analogiczne do wcześniej podanych, odpowiednio: (1.6.5), (1.6.3), (1.6.4).

Badanie aktualności modelu przeprowadzamy współczynnikiem Janusowym J^{124} :

$$J^2 = \frac{\frac{1}{T-n} \sum_{t=n+1}^T (y_t - y_{tP})^2}{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad (1.12.5)$$

gdzie:

y_t – wartości zmiennej prognozowanej (mianownik) w okresach próby i weryfikacji prognoz (licznik),

y_{tP} – prognozy w okresie weryfikacji prognoz,

\hat{y}_t – wartości obliczone na podstawie modelu dla n obserwacji.

Jeśli spełniony jest warunek $J^2 \leq 1$, to model prognostyczny można stosować dalej.

P r z y k ł a d

Liczba komputerów y_t w firmie w latach 1995–2001 wyniosła (stan na dzień 31.12 danego roku):

T a b e l a 1.12.1

Liczba komputerów w firmie

Lata	1995	1996	1997	1998	1999	2000	2001
Y_t	114	139	183	205	231	233	234

Ź r ó d ł o: Opracowanie własne; dane umowne.

¹²⁴ R. C z y ż e w s k i, M. H u n d e r t, R K l ó s k a, *Wybrane zagadnienia z prognozowania...*, op.cit., rozdz. I.2.

Etapy prac:

1. Do obliczenia parametrów modelu trendu liniowego określamy model ekonometryczny klasyczną metodą najmniejszych kwadratów, a ponadto wyznaczymy:

- współczynnik korelacji (R),
- współczynnik determinacji (R^2),
- skorygowany współczynnik determinacji (\bar{R}^2).

2. Na podstawie modelu określamy prognozy liczby komputerów w firmie w latach 2002–2004.

3. Zakładając, że mamy rzeczywiste liczby komputerów w latach 2002–2004, które wyniosły odpowiednio 265, 290, 315, badamy trafność prognoz przy pomocy średniego bezwzględnego błędu procentowego prognoz *ex post*.

4. Sprawdzamy współczynnikiem Janusowym, czy model jest nadal aktualny.

Ad 1) Model trendu liniowego:

$$\hat{Y} = a_0 + a_1 \cdot t, \quad a_1 = \frac{\sum_{t=1}^n (y_t - \bar{y}) \cdot (t - \bar{t})}{\sum_{t=1}^n (t - \bar{t})^2}, \quad a_0 = \bar{y} - a_1 \cdot \bar{t}.$$

Tabela 1.12.2

Obliczenia pomocnicze oraz określenie parametrów

Lata	1995	1996	1997	1998	1999	2000	2001	średnia	
t	1	2	3	4	5	6	7	4	
y_t	114	139	183	205	231	233	234	191,29	
$y_t - \bar{y}$	-77,29	-52,29	-8,29	13,71	39,71	41,71	42,71		
$t - \bar{t}$	-3	-2	-1	0	1	2	3		
$(y_t - \bar{y}) (t - \bar{t})$	231,86	104,57	8,29	0,00	39,71	83,43	128,14	SUMA	596,00
$(t - \bar{t})^2$	9	4	1	0	1	4	9	SUMA	28

$$a_1 = 596/28 = 21,29, \quad a_0 = \bar{y} - a_1 = 106,14 \text{ (obliczenia w Excelu),}$$

$$\hat{y} = 106,14 + 21,29 t.$$

Źródło: Opracowanie własne w Excelu, gdzie: $\bar{y} \rightarrow \bar{y}, \bar{t} \rightarrow \bar{t}, \hat{y} \rightarrow \hat{y}$.

Ad 1a) Współczynnik korelacji R (analogicznie do wzoru 1.4.2):

$$R_{yt} = \frac{\text{cov}_{yt}}{\sqrt{\text{var}_y \cdot \text{var}_t}} = \frac{\text{cov}_{yt}}{S(y_t)S(t)},$$

$$R_{yt} = [\Sigma(y_t - \bar{y})(t - \bar{t})] / [n \cdot S(y_t) \cdot S(t)],$$

$$n = 7.$$

Kowariancja według tabeli 1.12.2:

$$\text{cov}_{yt} = [\Sigma (y_t - \bar{y}) \cdot (t - \bar{t})] / n = 596 / 7 = 85,14$$

$S(y_t)$ oraz $S(t)$ – odchylenie standardowe zmiennej Y i okresów t jako pierwiastka z wariancji var_y oraz var_t .

Tabela 1.12.3

Tabela pomocnicza do obliczenia wariancji zmiennej Y oraz t

$y_t - \bar{y}$	-77,29	-52,29	-8,29	13,71	39,71	41,71	42,71		Wariancja	S
$(y_t - \bar{y})^2$	5 973,08	2 733,80	68,65	188,08	1 577,22	1 740,08	1 824,51	Suma/7	2 015,0612	44,89
$(t - \bar{t})^2$	9	4	1	0	1	4	9	Suma/7	4	2,00

$$R = 85,14 / (44,89 \cdot 2) = 0,94836,$$

$$R = 94,84\%.$$

Źródło: Opracowanie własne w Excelu, gdzie: $y^- \rightarrow \bar{y}$, $t^- \rightarrow \bar{t}$.

Ad 1b) Współczynnik determinacji $(R^2) = 0,94836^2 = 0,8994$.

Ad 1c) Skorygowany współczynnik determinacji (\bar{R}^2) dla $k = 2$:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k} \cdot (1 - R^2), \quad (1.12.6)$$

$$\bar{R}^2 = 1 - \frac{7-1}{7-2} \cdot (1 - 0,8994) = 0,9793.$$

Ad 2) Określenie prognozy liczby komputerów w firmie w latach 2002–2004:

$$\hat{y}_t = 106,14 + 21,29 \cdot t, \quad t = 8, 9, 10, \quad n = 7, T = 10.$$

Tabela 1.12.4

Prognozy na lata 2002–2004

t	8	9	10
\hat{y}	276,46	297,75	319,04

Źródło: Opracowanie własne w Excelu.

Ad 3) Przyjmując, że wartości zmiennej y_t w latach 2002–2004 były odpowiednio: 265, 290, 315, zbadamy trafność prognoz przy pomocy średniego bezwzględnego (absolutnego) błędu procentowego prognoz *ex post*. Kolejny wzór jest analogiczny do (1.12.3).

$$MAPE = \frac{1}{T-n} \sum_{t=n+1}^T \left| \frac{\hat{y}_t - y_{tp}}{y_t} \right| \cdot 100.$$

Tabela 1.12.5

Obliczenia pomocnicze do błędu MAPE

<i>t</i>	8	9	10
<i>y</i> [^]	276,46	297,75	319,04
<i>t</i>	<i>y_t</i>	<i>y_{tp}</i>	<i> (y_t - y_{tp})/y_t </i>
8	265	276,46	0,0432
9	290	297,75	0,0267
10	315	319,04	0,0128
		Suma	0,0828
		MAPE [%]	2,7598

Źródło: Opracowanie własne w Excelu, gdzie: $y^{\wedge} \rightarrow \hat{y}$.

$$MAPE = [1/(10 - 7)] \cdot 0,0828 \cdot 100 \cong 2,76\%.$$

Dokonując prognozy liczby komputerów w firmie, pomyliliśmy się średnio o 2,76%. Zatem otrzymane wyniki prognozowania można uznać za trafne, są bowiem w granicach błędu 5% (tab. 1.12.5).

Ad 4) Sprawdzenie współczynnikiem Janusowym na podstawie wzoru (1.5), czy model jest nadal aktualny.

Tabela 1.12.6

Obliczenia pomocnicze do współczynnika Janusowego

1995	1	114	127,43	180,3649
1996	2	139	148,72	94,4784
1997	3	183	170,01	168,7401
1998	4	205	191,3	187,69
1999	5	231	212,59	338,9281
2000	6	233	233,88	0,7744
2001	7	234	255,17	448,1689
Suma				1419,1448
Lata	<i>t</i>	<i>y_t</i>	<i>y_{tp}</i>	<i>(y_t - y_{tp})²</i>
2002	8	265	276,46	131,3316
2003	9	290	297,75	60,0625
2004	<i>T</i> = 10	315	319,04	16,3216
Suma				207,7157

$$J^2 = \frac{1/(10 - 7)207,7157}{(1/7)1419,1448},$$

$$J^2 = 0,15.$$

Ponieważ współczynnik Janusowy nie przekracza 1, model można wykorzystywać do dalszego prognozowania.

Źródło: Opracowanie własne w Excelu, gdzie: $y_{tp}^{\wedge} \rightarrow \hat{y}_{tp}$.

1.12.2. Pytania

1. Objasnij elementy wzoru na względny procentowy błąd prognozy *ex ante*.
2. Do czego służy współczynnik Janusowy i jak obliczyć jego poszczególne składniki?
3. Od kiedy model szeregu czasowego nie nadaje się do dalszego prognozowania?

1.12.3. Zadania

1. W przedsiębiorstwie liczba stanowisk pracy bezpośrednio produkcyjnych w latach 2000–2006 według stanu na 31.12 danego roku wynosiła:

Tabela 1.12.7

Liczba stanowisk pracy

t	1	2	3	4	5	6	7
Lata	2000	2001	2002	2003	2004	2005	2006
Y_t	11	14	18	20	23	25	28

Źródło: Opracowanie własne w Excelu; dane umowne.

Określony funkcją REGLINP model trendu liniowego jest następujący:

$$\hat{Y} = 8,714 + 2,786 t.$$

Tabela 1.12.8

Parametry modelu liniowego oraz fragment statystyki opisowej

a_1	a_0	
2,786	8,714	parametry modelu
0,106	0,474	standardowe błędy oceny parametrów
0,993		współczynnik determinacji (R^2)

Źródło: Opracowanie własne w Excelu.

W ramach zadania:

- a) zbadaj dopasowanie modelu do danych z obserwacji przy pomocy skorygowanego współczynnika determinacji \bar{R}^2 ; $n = 7$; $k = 2$,
- b) oblicz wygasłe już prognozy na lata 2000–2006,
- c) określ prognozę liczby stanowisk pracy bezpośrednio produkcyjnych na rok 2007,
- d) oblicz względny błąd prognozy *ex ante* [%] określonej na rok 2007.

2. Dane są obserwacje rozwoju osiedla domków jednorodzinnych na lata 2000–2006 według stanu na 31.12 danego roku:

T a b e l a 1.12.9

Liczby domów jednorodzinnych

t	1	2	3	4	5	6	7
Lata	2000	2001	2002	2003	2004	2005	2006
Y_t	10	12	15	16	20	22	24

Ź r ó d ł o: Opracowanie własne; dane umowne.

- a) zdefiniuj model trendu liniowego $\hat{Y} = a_0 + a_1 \cdot t$, korzystając z KMNK.
- b) określ prognozy na lata 2007–2009.
- c) sprawdź współczynnikiem Janusowym według warunku $J^2 \leq 1$, czy model jest nadal aktualny.

1.12.4. Budowa i weryfikacja modelu regresji

P r z y k ł a d 1

Rozpocznijmy od przykładu modelu regresji z jedną zmienną objaśniającą. Do celów dydaktycznych niech to będzie założenia model liniowy obrazujący zależność liczby zawartych małżeństw M , w tys., od liczby przeciętnie zatrudnionych w roku Z , również w tys.:

$$\hat{M} = f(Z) + \varepsilon.$$

gdzie:

ε – składnik losowy.

W tabeli 1.12.10 wprowadzono jeszcze jedną zmienną, która będzie później zastosowana, tj. R – liczba rozwodów, w tys., która, jak się wydaje, zniechęca potencjalne pary do zawierania małżeństw.

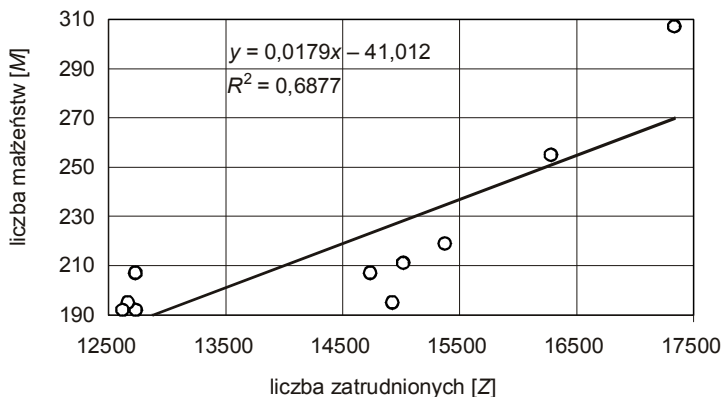
T a b e l a 1.12.10

Wartości zmiennych „ M ”, „ Z ”, „ R ”

		y	x_1	x_2
n	Rok	M	Z	R
1	1980	307	17 334	39,8
2	1990	255	16 280	42,4
3	1995	207	14 735	38,1
4	1999	219	15 373	42
5	2000	211	15 018	42,8
6	2001	195	14 924	45,3
7	2002	192	12 729	45,4
8	2003	195	12 663	48,6
9	2004	192	12 615	56,3
10	2005	207	12 728	67,6

Ź r ó d ł o: Opracowanie własne na podstawie: „Rocznik Statystyczny Rzeczypospolitej Polskiej” 2006, tab. „Ważniejsze dane o sytuacji społeczno-gospodarczej kraju”.

Przedstawmy naszą zależność $M = f(Z)$ w formie diagramy korelacyjnego.



Źródło: Opracowanie własne w Excelu, gdzie $y \rightarrow M$, $x \rightarrow Z$.

Wykres 1.12.1. Zależność zmiennych M i Z

Między badanymi zmiennymi istnieje dodatnia zależność korelacyjna. Współczynnik determinacji (R^2) podany na diagramie wynosi 0,6877, co świadczy o średnim dopasowaniu modelu regresji liniowej do danych empirycznych. Obliczmy współczynnik korelacji między zmienną objaśnianą M a zmienną objaśniającą Z :

$$R_{MZ} = \sqrt{R^2} = \sqrt{0,6877} = 0,829.$$

Oceniamy istotność współczynnika korelacji, korzystając z testu w publikacji M. Lipiec-Zajchowskiej¹²⁵:

Przystępujemy do procedury weryfikacji hipotezy:

$H_0: r_{xy} = 0$ (nie istnieje zależność pomiędzy badanymi zmiennymi),

$H_1: r_{xy}$ (istnieje zależność pomiędzy badanymi zmiennymi).

przy czym: $x = Z$, $y = M$.

Weryfikujemy poprzez porównanie wartości statystyki t_{obl} obliczonej z próby z wartością teoretyczną t_α , odczytaną z rozkładu t -Studenta. Na podstawie próby obliczamy statystyki t_{obl} dla $N = n$ obserwacji według wzoru (1.1.5)¹²⁶:

$$t_{obl} = R_{xy} \cdot \sqrt{\frac{n-2}{1-R_{xy}^2}}.$$

Jeśli $|t_{obl}| \geq |t_\alpha|$, to hipotezę zerową odrzucamy, oznacza to bowiem istnienie zależności pomiędzy zmienną objaśnianą M a zmienną objaśniającą Z w populacji generalnej.

Przyjmujemy poziom istotności α , np. 0,05, oraz ilość stopni swobody:

$$n' = (n - 1) = 10 - 1 = 9.$$

¹²⁵ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2. *Ekonometria...*, op.cit., rozdz. 1.3.

¹²⁶ Ibidem.

Na podstawie tablicy rozkładu t -Studenta $t_{(0,05; 9)} = 2,2622$:

$$t_{obl} = 0,829 \cdot \sqrt{\frac{10-2}{1-0,829^2}} = 4,1927.$$

Zatem $|t_{obl}| > |t_{\alpha}|$.

Istnieje więc zależność pomiędzy badanymi zmiennymi określonego modelu:

$$\hat{M} = -41,012 + 0,0179 \cdot Z.$$

Obliczamy skorygowany współczynnik determinacji jako miarę dopasowania modelu przy $k = 2$ (ilość szacowanych parametrów):

$$\bar{R}^2 = 1 - \frac{n-1}{n-k} \cdot (1 - R^2) = 1 - \frac{10-1}{10-2} \cdot (1 - 0,6877) = 0,6487,$$

czyli w około 65% zmienna objaśniana M jest wyjaśniana zmiennością zmiennej objaśniającej Z .

P r z y k ł a d 2

Alternatywnym rozwiązaniem do określenia regresji liniowej jest zastosowanie funkcji REGLINP ze statystyka opisową.

T a b e l a 1.12.11

Raport z funkcji REGLINP

a_1	a_0	
0,0179	-41,012	parametry a_i
0,0043	62,082	standardowy błąd szacunku parametrów a_i
0,6877	21,635	współczynnik determinacji (R^2); odchylenie standardowe reszt (S_e)
17,6203	8	statystyka F ; ilość stopni swobody (iss)
8247,4640	3744,536	regresyjna suma kwadratów (RSS); resztowa suma kwadratów (RSS)

Ź r ó d ł o: Opracowanie własne w Excelu.

Wariancja resztowa jest ilorazem resztowej sumy kwadratów (RSS) i ilości stopni swobody ($iss = n - m - 1 = n - k$); przy czym m – ilość zmiennych objaśniających (wzór 1.8.1).

$$S_u^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n - k} = \frac{3744,536}{10 - 2} = 468,07.$$

Kolejną miarą oceny dobroci dopasowania modelu do danych empirycznych jest odchylenie standardowe reszt stanowiące pierwiastek z wariancji resztowej. Można go pobrać również z raportu funkcji REGLINP:

$$S_u = \sqrt{S_u^2} = \sqrt{468,07} = 21,635.$$

Przykład 3

Oceniamy zmienność reszt w stosunku do wartości średniej \bar{m} zmiennej objaśnianej M .

Tabela 1.12.12

Obliczenie resztowej sumy kwadratów

n	$Y = M$	$X_1 = Z$	\hat{M}	$e = u = M - \hat{M}$	u^2
1	307	17 334	269,2666	37,7334	1 423,8095
2	255	16 280	250,4000	4,6000	21,1600
3	207	14 735	222,7445	-15,7445	247,8893
4	219	15 373	234,1647	-15,1647	229,9681
5	211	15 018	227,8102	-16,8102	282,5828
6	195	14 924	226,1276	-31,1276	968,9275
7	192	12 729	186,8371	5,1629	26,6555
8	195	12 663	185,6557	9,3443	87,3159
9	192	12 615	184,7965	7,2035	51,8904
10	207	12 728	186,8192	20,1808	407,2647
Średnia: $\bar{m} = 218$				Σ	3 747,4638

Źródło: Opracowanie własne w Excelu.

Wystąpiła różnica w obliczeniu resztowej sumy kwadratów (RSS) w stosunku do raportu funkcji REGLINP w przykładzie 2. Wynika to z zaokrąglenia parametrów modelu $\hat{M} = -41,012 + 0,0179 \cdot Z$ do trzeciego miejsca po przecinku.

Obliczenie współczynnika zmienności według wzoru (1.3.15) i $S_e = 21,635$ według tab. 1.12.11.

$$V_u = \frac{S_e}{\bar{m}} = \frac{21,635}{218} \approx 0,1 = 10\%.$$

Przykład 4

Badanie statystycznej istotności parametrów a_0 oraz a_1 testem t -Studenta według wzoru (1.1.8):

$$t_{obl} = \frac{|a_i|}{S(a_i)}, \quad |t_{obl}| > t_\alpha,$$

hipoteza zerowa: $H_0: |a_i| = 0$,

hipoteza alternatywna: $H_1: |a_i| \neq 0$.

Gdy ten warunek jest spełniony, to hipotezę zerową odrzucamy na rzecz hipotezy alternatywnej; parametr a_i różni się w sposób istotny od zera, a zmienna objaśniająca Z wpływa w sposób istotny na zmienną objaśnianą M .

Dla $\alpha = 0,05$ i $iss = n - 2 = 10 - 2 = 8$, $t_\alpha = 2,306$ (według tablicy wartości krytycznych rozkładu t -Studenta)

$$t_{a_0} = \frac{|-41,012|}{62,082} = 0,664, \quad t_{a_1} = \frac{0,0179}{0,0043} = 4,163.$$

Jeżeli $t_{a_0} < t_{\alpha}$, to nie mamy podstaw do odrzucenia hipotezy H_0 . Gdy $t_{a_1} > t_{\alpha}$, odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej, parametr a_1 różni się w sposób istotny od zera i zmienna objaśniająca Z w sposób istotny oddziałuje na zmienną objaśnianą M .

Ponieważ zakładamy, że parametr a_0 pozostaje w modelu, badanie dla t_{a_0} było niekonieczne.

Przykład 5

Zastosujmy test Jargu'ea-Bery badania normalności składnika resztowego¹²⁷.

Test ten służy do weryfikacji hipotezy, że rozkład składnika resztowego jest rozkładem normalnym i określony jest wzorem (1.1.15), którego podobna formuła występuje również w publikacji M. Lipiec-Zajchowskiej¹²⁸:

$$JBT = \frac{n-k}{6} \cdot [S^2 + \frac{1}{4}(K-3)^2],$$

gdzie:

$k = 2$ – liczba szacowanych parametrów (a_0, a_1),

$S = A_s$ – wartość współczynnika skośności, gdy nie dysponujemy Excelem, obliczana według wzoru analogicznego do (1.3.17):

$$A_s = S = \frac{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^3}{S^3}.$$

K – kurtოza opisana formułą analogiczną do (1.1.14):

$$K = \frac{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^4}{S^4},$$

we wzorach:

S – odchylenie standardowe,

$T = n$ – liczba obserwacji,

y_t – wartość y w okresie t .

\bar{y} – średnia wartość y .

Hipoteza zerowa H_0 zakłada, że rozkład składnika resztowego jest rozkładem normalnym. Hipoteza alternatywna H_1 zakłada, że rozkład ten nie jest rozkładem normalnym. Test oparty jest na rozkładzie χ^2 z dwoma stopniami swobody. Hipoteza zerowa jest spełniona gdy¹²⁹:

¹²⁷ Ibidem.

¹²⁸ Lipiec-Zajchowska M.(red.), *Wspomaganie procesów decyzyjnych*, t. 1: *Statystyka*, op.cit. Aneks.

¹²⁹ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria...*, op.cit., rozdz. 1.9.

$$JBT < \chi^2_{(\alpha, 2)}$$

Dla poziomu istotności $\alpha = 0,05$ i $r = 2$ stopnie swobody wartość krytyczna $\chi^2_{(0,05, 2)}$ wynosi 5,991. Skorzystajmy z funkcji Excela do obliczenia skośności oraz kurtozy szeregu reszt modelu (reszty z tab. 1.12.12 zapisane są w obszarze M30:M39).

-0,1802	[=KURTOZA(M30:M39)]
0,2459	[=SKOŚNOŚĆ(M30:M39)]

Zatem:

$$JBT = \frac{10-2}{6} \cdot [0,2459^2 + \frac{1}{4}(-0,1802-3)^2] = \frac{4}{3} \cdot (0,0605 + \frac{10,1137}{4}) = \frac{4}{3} \cdot 2,5889 = 3,4519.$$

Ponieważ $JBT = 3,4519 < \chi^2_{(0,05, 2)} = 5,991$, nie mamy podstaw do odrzucenia hipotezy H_0 i stwierdzamy, że rozkład reszt jest rozkładem normalnym.

Przykład 6

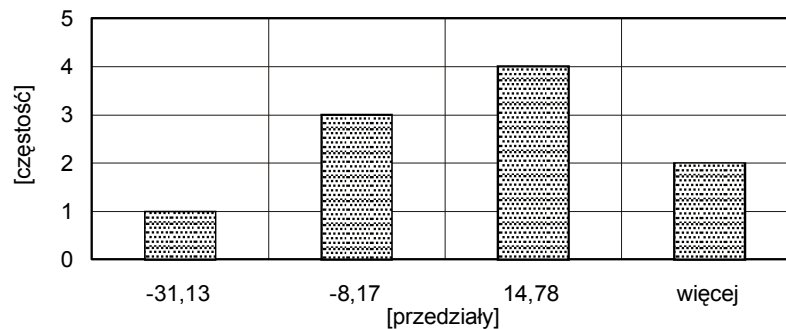
Zobaczmy rozkład składnika resztowego e z tab. 1.12.12 na histogramie

Tabela 1.12.13

Częstość występowania reszt w przedziałach

Przedział	Częstość
-31,13	1
-8,17	3
14,78	4
Więcej	2

Źródło: Opracowanie własne w Excelu.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.2. Histogram reszt według danych z tab. 1.12.12

Przykład 7

Początkowa część tego przykładu nawiązuje do procedury zaprezentowanej w rozdziale 1.6.11.

Opracowanie modelu regresji z czterema zmiennymi objaśniającymi X_1, X_2, X_3, X_4 . Wyznaczamy prognozy na miesiące 11. i 12, bazując na danych:

Tabela 1.12.14

Dane umowne zmiennych

t	y	x_1	x_2	x_3	x_4
1	18,5	4,4	10,5	2,80	8
2	19,2	4,3	10,0	2,83	10
3	20,0	4,5	11,7	3,01	15
4	20,8	4,5	12,1	3,17	12
5	21,7	4,4	13,0	3,42	7
6	20,8	4,6	13,9	3,56	10
7	23,7	4,6	15,1	3,73	11
8	25,0	4,6	13,9	3,81	13
9	26,7	4,5	18,1	4,07	8
10	28,3	4,5	20,0	4,32	17

gdzie zmienne oznaczają:

Y – produkcja przedsiębiorstwa w tys. szt.,

X_1 – zatrudnienie (w tys. osób),

X_2 – moc zainstalowanych maszyn w kW,

X_3 – dostawa surowców w tys. ton,

X_4 – przestoje z powodu awarii w dniach.

Źródło: Opracowanie własne (kopia tabeli 1.6.11).

W ramach zadania:

1. Dla ogólnej postaci modelu:

$$\ln \hat{Y} = f(\ln X_1, \dots, \ln X_j, c), \text{ czyli } \ln \hat{Y} = c + a_1 \ln X_1 + \dots + a_n \ln X_n$$

określamy podstawowe parametry zmiennej objaśnianej Y , zmiennych objaśniających X_i w postaci zlogarytmowanej.

Tabela 1.12.15

Ramka pomocnicza do obliczania miar zmienności logarytmów zmiennych

ln ...	średnia	odchylenie standardowe	maksimum	minimum
ln Y				
ln X_1				
ln X_2				
ln X_3				
ln X_4				

Źródło: Opracowanie własne (kopia tabeli 1.6.12).

2. Wyznaczamy współczynniki kowariancji i korelacji pomiędzy logarytmem zmiennej objaśnianej a logarytmami zmiennych objaśniających.

T a b e l a 1.12.16

Ramka pomocnicza do obliczenia kowariancji i korelacji między logarytmami zmiennych

Nr	Relacje zmiennych	Kowariancja	Korelacja
1	ln Y, ln Y		
2	ln Y, ln X ₁		
3	ln Y, ln X ₂		
4	ln Y, ln X ₃		
5	ln Y, ln X ₄		
6	ln X ₁ , ln X ₁		
7	ln X ₁ , ln X ₂		
8	ln X ₁ , ln X ₃		
9	ln X ₁ , ln X ₄		
10	ln X ₂ , ln X ₂		
11	ln X ₂ , ln X ₃		
12	ln X ₂ , ln X ₄		
13	ln X ₃ , ln X ₃		
14	ln X ₃ , ln X ₄		
15	ln X ₄ , ln X ₄		

Ź r ó ł o: Opracowanie własne (kopia tabeli 1.6.13).

Do obliczenia korelacji między zmienną objaśnianą Y a zmiennymi objaśniającymi X_j lub między zmiennymi objaśniającymi dla n = 10 stosujemy wzory analogiczne do (1.4.1), (1.4.2), przy czym wartości zmiennych zgodnie z wymogiem modelu podamy w postaci logarytmów.

$$\text{cov}_{yx} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}),$$

$$r_{yx} = \frac{\text{cov}_{yx}}{\sqrt{\text{var}_y \cdot \text{var}_x}} = \frac{\text{cov}_{yx}}{S_y \cdot S_x}.$$

3. Zastosujemy analizę korelacji do doboru zmiennych objaśniających według procedury:

a) wyznaczenie wartości krytycznej współczynnika korelacji, która określa poziom istotności współczynnika korelacji w populacji generalnej. Wartość tę wyznaczamy ze wzoru (1.6.9):

$$r^* = \sqrt{t_\alpha^2 / (t_\alpha^2 + n - 2)},$$

gdzie:

t_α – wartość statystyki odczytanej z tablicy t-Studenta dla danego poziomu istotności α oraz dla n – 2 stopni swobody.

Dla $n = 10$, $iss = n - 2 = 8$ oraz, przykładowo, $P = \alpha = 0,01$ według tablicy wartości krytycznych rozkładu t -Studenta (tabl. I w *Aneksie*) $t_\alpha = 3,3554$.

b) wylimowanie zmiennej objaśniającej, dla której korelacja między $\ln Y$ a $\ln X_i$ jest statystycznie nieistotna według zasady: $|r_j| \leq r^*$,

c) wybór z spośród pozostałych korelacji $|r_j|$ zmiennej o korelacji maksymalnej między $(\ln Y, \ln X_i)$; $|r_{hi}| = \max |r_j|$,

d) eliminacja zmiennych o wzajemnej korelacji większej od krytycznej (dla zmiennej o korelacji max) $|r_{hi}| > r^*$,

e) wybór następnej zmiennej objaśniającej przez powrót do pkt. d),

f) gdy takiej nie ma, to sformułowanie modelu np. $\ln \hat{Y} = a_1 \ln X_2 + a_0$.

4. Oszacowanie parametrów KMNK dla wielu zmiennych, przy czym m wartości zmiennych występują w postaci logarytmów.

5. Określenie funkcji produkcji na podstawie zadanej na wejściu:

$$\ln \hat{Y} = a_0 + a_1 \ln X_1 + \dots + a_4 \ln X_4$$

po zamianie jej w postać funkcji wykładniczej $\hat{Y} = e^{(a_0 + a_1 \ln X_1 + \dots + a_4 \ln X_4)}$.

6. Określenie prognozy na miesiące 11. i 12. poprzez:

a) określenie przez eksperta zmiany wielkości zmiennych objaśniających lub przez wyznaczenie ich trendu,

b) obliczenie prognoz analitycznie na podstawie sformułowanego modelu.

Rozwiązanie tego przykładu przeprowadzono z zastosowaniem wyrażeń Excela.

Ad 1) Podstawowe parametry zmiennej objaśnianej Y oraz zmiennych objaśniających X_i w postaci zlogarytmowanej.

Tabela 1.12.17a

Tabela pomocnicza do obliczania miar zmienności

t	Y	$Y^* = \ln Y$	$(\ln Y - \ln Y^*)^2$	X_1	$X_1^* = \ln X_1$	$(\ln X_1 - \ln X_1^*)^2$	X_2	$X_2^* = \ln X_2$	$(\ln X_2 - \ln X_2^*)^2$
1	18,5	2,9177707	0,0342176	4,4	1,4816045	0,0004010	10,5	2,3513753	0,0639113
2	19,2	2,9549103	0,0218568	4,3	1,4586150	0,0018503	10	2,3025851	0,0909607
3	20	2,9957323	0,0114530	4,5	1,5040774	0,0000060	11,7	2,4595888	0,0209072
4	20,8	3,0349530	0,0045965	4,5	1,5040774	0,0000060	12,1	2,4932055	0,0123158
5	21,7	3,0773123	0,0006471	4,4	1,4816045	0,0004010	13	2,5649494	0,0015392
6	20,8	3,0349530	0,0045965	4,6	1,5260563	0,0005966	13,9	2,6318888	0,0007677
7	23,7	3,1654750	0,0039343	4,6	1,5260563	0,0005966	15,1	2,7146947	0,0122131
8	25	3,2188758	0,0134850	4,6	1,5260563	0,0005966	13,9	2,6318888	0,0007677
9	26,7	3,2846636	0,0330923	4,5	1,5040774	0,0000060	18,1	2,8959119	0,0851063
10	28,3	3,3428618	0,0576533	4,5	1,5040774	0,0000060	20	2,9957323	0,1533116
♦		31,0275078	0,1855325		15,0163026	0,0044662		26,0418206	0,4418005
Średnia/wariancja		3,1027508	0,0185533		1,5016303	0,0004466		2,6041821	0,0441800
Odchylenie standardowe			0,1362103			0,0211334			0,2101905
Maksimum		3,3428618			1,5260563			2,9957323	
Minimum		2,9177707			1,4586150			2,3025851	

Źródło: Opracowanie własne w Excelu, gdzie: $\ln Y^- \rightarrow \ln \bar{Y}$, $\ln X_1^- \rightarrow \ln \bar{X}_1$, $\ln X_2^- \rightarrow \ln \bar{X}_2$.

Tabela 1.12.17b

Tabela pomocnicza do obliczania miar zmienności (cd.)

X_3	$X_3' = \ln X_3$	$(\ln X_3 - \ln X_3)^2$	X_4	$X_4' = \ln(X_4)$	$(\ln X_4 - \ln X_4)^2$
2,8	1,0296194	0,0420131	8	2,0794415	0,0843649
2,83	1,0402767	0,0377578	10	2,3025851	0,0045310
3,01	1,1019401	0,0175961	15	2,7080502	0,1143470
3,17	1,1537316	0,0065382	12	2,4849066	0,0132270
3,42	1,2296406	0,0000245	7	1,9459101	0,1797655
3,56	1,2697605	0,0012369	10	2,3025851	0,0045310
3,73	1,3164082	0,0066941	11	2,3978953	0,0007839
3,81	1,3376292	0,0106170	13	2,5649494	0,0380451
4,07	1,4036430	0,0285788	8	2,0794415	0,0843649
4,32	1,4632554	0,0522877	17	2,8332133	0,2146613
	12,3459047	0,2033442		23,6989782	0,7386215
	1,2345905	0,0203344		2,3698978	0,0738622
		0,1425988			0,2717759
	1,4632554			2,8332133	
	1,0296194			1,9459101	

Źródło: Opracowanie własne w Excelu, gdzie: $\ln X_3^- \rightarrow \ln \bar{X}_3$, $\ln X_4^- \rightarrow \ln \bar{X}_4$.

Tabela 1.12.18

Obliczone miary pomocnicze zmiennych w postaci zlogarytmowanej

ln ...	Średnia	Odchylenie standardowe	Maksimum	Minimum
ln Y	3,102750776	0,136210322	3,342861805	2,917770732
ln X_1	1,50163026	0,021133429	1,526056303	1,458615023
ln X_2	2,604182064	0,210190508	2,995732274	2,302585093
ln X_3	1,234590472	0,142598816	1,463255402	1,029619417
ln X_4	2,369897824	0,271775928	2,833213344	1,945910149

Źródło: Opracowanie własne w Excelu.

Ad 2) Współczynniki kowariancji i korelacji między zmiennymi w postaci zlogarytmowanej.

Tabela 1.12.19a

Obliczenia pomocnicze do określenia kowariancji i korelacji między logarytmami zmiennych

	$(\ln Y - \ln Y)^2$	$(\ln Y - \ln Y)$	$(\ln X_1 - \ln X_1)^2$	$(\ln X_1 - \ln X_1)$	b · d	$(\ln X_2 - \ln X_2)^2$	$(\ln X_2 - \ln X_2)$	b · g
	a	b	c	d	e	f	g	h
	0,0342176	0,1849800	0,0004010	0,0200257	0,0037044	0,0639113	0,2528068	0,0467642
	0,0218568	0,1478405	0,0018503	0,0430152	0,0063594	0,0909607	0,3015970	0,0445882
	0,0114530	0,1070185	0,0000060	0,0024471	0,0002619	0,0209072	0,1445932	0,0154742
	0,0045965	0,0677978	0,0000060	0,0024471	0,0001659	0,0123158	0,1109766	0,0075240
	0,0006471	0,0254385	0,0004010	0,0200257	0,0005094	0,0015392	0,0392327	0,0009980
	0,0045965	0,0677978	0,0005966	0,0244260	0,0016560	0,0007677	0,0277068	0,0018785
	0,0039343	0,0627243	0,0005966	0,0244260	0,0015321	0,0122131	0,1105127	0,0069318
	0,0134850	0,1161250	0,0005966	0,0244260	0,0028365	0,0007677	0,0277068	0,0032175
	0,0330923	0,1819128	0,0000060	0,0024471	0,0004452	0,0851063	0,2917299	0,0530694
	0,0576533	0,2401110	0,0000060	0,0024471	0,0005876	0,1533116	0,3915502	0,0940155
Σ	0,1855325				0,0180583			0,2744613
kowariancja	0,0185533				0,0018058			0,0274461
korelacja	1				0,6273325			0,9586453

Źródło: Opracowanie własne w Excelu.

Tabela 1.12.19b

Obliczenia pomocnicze do określenia kowariancji
i korelacji między logarytmami zmiennych (cd.)

$(\ln X_3 - \ln X_3)^2$	$(\ln X_3 - \ln X_3)$	b · j	$(\ln X_4 - \ln X_4)^2$	$(\ln X_4 - \ln X_4)$	b · m
<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>	<i>m</i>	<i>n</i>
0,0420131	0,2049711	0,0379156	0,0843649	0,2904563	0,0537286
0,0377578	0,1943138	0,0287274	0,0045310	0,0673127	0,0099515
0,0175961	0,1326504	0,0141960	0,1143470	0,3381524	0,0361886
0,0065382	0,0808589	0,0054821	0,0132270	0,1150088	0,0077973
0,0000245	0,0049499	0,0001259	0,1797655	0,4239877	0,0107856
0,0012369	0,0351701	0,0023845	0,0045310	0,0673127	0,0045637
0,0066941	0,0818178	0,0051320	0,0007839	0,0279974	0,0017561
0,0106170	0,1030387	0,0119654	0,0380451	0,1950515	0,0226504
0,0285788	0,1690525	0,0307528	0,0843649	0,2904563	0,0528377
0,0522877	0,2286649	0,0549050	0,2146613	0,4633155	0,1112472
		0,1915866			0,3115067
		0,0191587			0,0311507
		0,9863685			0,8414850

Źródło: Opracowanie własne w Excelu.

Obliczenia dalszych zależności, tj. 8–15, występujących w tabeli 1.12.16 wykonano analogicznie. W rezultacie otrzymano tabelę kowariancji i korelacji (1.12.20).

Tabela 1.12.20

Obliczone kowariancje i korelacje

Nr	Relacje	Kowariancja	Korelacja
1	$\ln Y, \ln Y$	0,0185533	1
2	$\ln Y, \ln X_1$	0,0018058	0,6273325
3	$\ln Y, \ln X_2$	0,0274461	0,9586453
4	$\ln Y, \ln X_3$	0,0191587	0,9863685
5	$\ln Y, \ln X_4$	0,0311507	0,8414850
6	$\ln X_1, \ln X_1$	0,0004466	1
7	$\ln X_1, \ln X_2$	0,0025172	0,5666754
8	$\ln X_1, \ln X_3$	0,0019432	0,6448232
9	$\ln X_1, \ln X_4$	0,0027249	0,4744211
10	$\ln X_2, \ln X_2$	0,0441800	1
11	$\ln X_2, \ln X_3$	0,0290493	0,9691856
12	$\ln X_2, \ln X_4$	0,0448532	0,7851803
13	$\ln X_3, \ln X_3$	0,0203344	1
14	$\ln X_3, \ln X_4$	0,0308671	0,7851803
15	$\ln X_4, \ln X_4$	0,0738622	1

Źródło: Opracowanie własne w Excelu (kopia tabeli 1.6.14).

Ad 3) Analiza korelacji doboru zmiennych objaśniających.

a) wartość krytyczna współczynnika korelacji:

$$r^* = \sqrt{3,3554^2 / (3,3554^2 + 8 - 2)} = 0,76,$$

b) wyeliminowanie zmiennej objaśniającej, której korelacja między $\ln Y$ a $\ln X_j$ jest statystycznie nieistotna, według zasady:

$|r_j| \leq r^*$, $|r_1| = 0,627$, czyli jest $< r^*$, wyeliminowaniu podlega zmienna X_1 ,

c) wybór spośród pozostałych zmiennej o korelacji maksymalnej między $(\ln Y, \ln X_j)$:

$|r_h| = \max |r_3| = 0,986$, jest to zmienna X_3 .

d) eliminacja zmiennych o wzajemnej korelacji większej od krytycznej:

$|r_{hj}| > r^*$, są nimi X_4 o korelacji 0,785 oraz X_2 o korelacji 0,969.

e) pozostała tylko zmienna X_3 i ona wchodzi do modelu regresji liniowej.

Ponieważ w dalszym toku chcemy pokazać zastosowanie KMNK do wielu zmiennych, założymy, że do naszego modelu wchodzi oprócz X_3 również zmienna X_2 :

$$\ln Y = f(\ln X_2, \ln X_3), \quad Y' = a_1 \cdot X_2' + a_2 \cdot X_3' + a_0.$$

Ad 4) Oszacowanie parametrów modelu $Y' = a_1 \cdot X_2' + a_2 \cdot X_3' + a_0$ z zastosowaniem wzorów matematycznych Cramera¹³⁰.

W analogii do poniższych równań według symbolizacji Cramera przyjmijmy założenia: $b_i = a_i$, $Y = \ln Y$, $X_1 = \ln X_2$, $X_2 = \ln X_3$, $n = 10$.

Wynikające z naszego przykładu:

$$\Sigma Y = b_0 \cdot n + b_1 \cdot (\Sigma X_1) + b_2 \cdot (\Sigma X_2),$$

$$\Sigma X_1 Y = b_0 \cdot (\Sigma X_1) + b_1 \cdot (\Sigma X_1^2) + b_2 \cdot (\Sigma X_1 X_2),$$

$$\Sigma X_2 Y = b_0 \cdot (\Sigma X_2) + b_1 \cdot (\Sigma X_1 X_2) + b_2 \cdot (\Sigma X_2^2).$$

Obliczmy więc poszczególne sumy, biorąc wartości zlogarytmowane zmiennych Y , X_1 , X_2 , X_3 .

Tabela 1.12.21

Obliczenia pomocnicze sum, iloczynów oraz kwadratów logarytmów zmiennych

Y	$X_1 = X_2'$	$X_2 = X_3'$	$X_1 Y$	$X_2 Y'$	X_1^2	X_2^2	$X_1 X_2$
$\ln Y$	$\ln X_2$	$\ln X_3$	$\ln X_2 \ln Y$	$\ln X_3 \ln Y$	$(\ln X_2)^2$	$(\ln X_3)^2$	$\ln X_2 \ln X_3$
2,9177707	2,3513753	1,0296194	6,8607739	3,0041934	5,5289656	1,0601161	2,4210216
2,9549103	2,3025851	1,0402767	6,8039324	3,0739243	5,3018981	1,0821756	2,3953256
2,9957323	2,4595888	1,1019401	7,3682697	3,3011175	6,0495773	1,2142719	2,7103195
3,0349530	2,4932055	1,1537316	7,5667613	3,5015211	6,2160734	1,3310966	2,8764899
3,0773123	2,5649494	1,2296406	7,8931501	3,7839879	6,5789652	1,5120159	3,1539657
3,0349530	2,6318888	1,2697605	7,9876589	3,8536636	6,9268389	1,6122918	3,3418686
3,1654750	2,7146947	1,3164082	8,5932985	4,1670574	7,3695676	1,7329306	3,5736465
3,2188758	2,6318888	1,3376292	8,4717234	4,3056623	6,9268389	1,7892518	3,5204913
3,2846636	2,8959119	1,4036430	9,5120964	4,6104950	8,3863060	1,9702137	4,0648265
3,3428618	2,9957323	1,4632554	10,0143190	4,8914606	8,9744119	2,1411164	4,3835214
31,028	26,042	12,346	81,072	38,493	68,259	15,445	32,441

Źródło: Opracowanie własne w Excelu; w ostatnim wierszu obliczono sumy odpowiadające podstawieniu.

¹³⁰ J. Jurlewicz, Z. Skoczylas, *Algebra liniowa I...*, op.cit., rozdz. 4: *Układy równań liniowych*.

$$\begin{aligned}
31,028 &= 10 \cdot b_0 + 26,042 \cdot b_1 + 12,346 \cdot b_2, \\
81,072 &= 26,042 \cdot b_0 + 68,256 \cdot b_1 + 32,441 \cdot b_2, \\
38,493 &= 12,346 \cdot b_0 + 32,441 \cdot b_1 + 15,445 \cdot b_2.
\end{aligned}$$

Wzór Cramera na macierzach¹³¹: $\mathbf{A} \cdot \mathbf{X} = \mathbf{B}$.

$$\mathbf{X}_1 = \det \mathbf{A}_1 / \det \mathbf{A}, \dots, \mathbf{X}_k = \det \mathbf{A}_k / \det \mathbf{A}, \dots, \mathbf{X}_n = \det \mathbf{A}_n / \det \mathbf{A},$$

$$\mathbf{X}_1 = b_0; \mathbf{X}_2 = b_1; \mathbf{X}_3 = b_2$$

$$\mathbf{A} = \begin{bmatrix} 10 & 26,042 & 12,346 \\ 26,042 & 68,256 & 32,441 \\ 12,346 & 32,441 & 15,445 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 31,028 \\ 81,072 \\ 38,493 \end{bmatrix}, \quad \mathbf{X} = [b_0 \quad b_1 \quad b_2].$$

Wyznacznik macierzy \mathbf{A} określamy według reguły Sarrusa lub funkcją Excela:

$$\det \mathbf{A} = 0,0480.$$

Macierze \mathbf{A}_1 , \mathbf{A}_2 , \mathbf{A}_3 tworzymy przez wstawienie kolumny macierzy \mathbf{B} w miejsce niewiadomej b_i .

$$\mathbf{A}_1 = \begin{bmatrix} 31,028 & 26,042 & 12,346 \\ 81,072 & 68,256 & 32,441 \\ 38,493 & 32,441 & 15,445 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 10 & 31,028 & 12,346 \\ 26,042 & 81,072 & 32,441 \\ 12,346 & 38,493 & 15,445 \end{bmatrix},$$

$$\mathbf{A}_3 = \begin{bmatrix} 10 & 26,042 & 31,028 \\ 26,042 & 68,256 & 81,072 \\ 12,346 & 32,441 & 38,493 \end{bmatrix}.$$

$$\det \mathbf{A}_1 = 0,0888; \det \mathbf{A}_2 = 0,0068; \det \mathbf{A}_3 = 0,0343,$$

$$a_0 = b_0 = \det \mathbf{A}_1 / \det \mathbf{A} = 0,0888 / 0,0480 = 1,85,$$

$$a_1 = b_1 = \det \mathbf{A}_2 / \det \mathbf{A} = 0,0068 / 0,0480 = 0,14,$$

$$a_2 = b_2 = \det \mathbf{A}_3 / \det \mathbf{A} = 0,0343 / 0,0480 = 0,72.$$

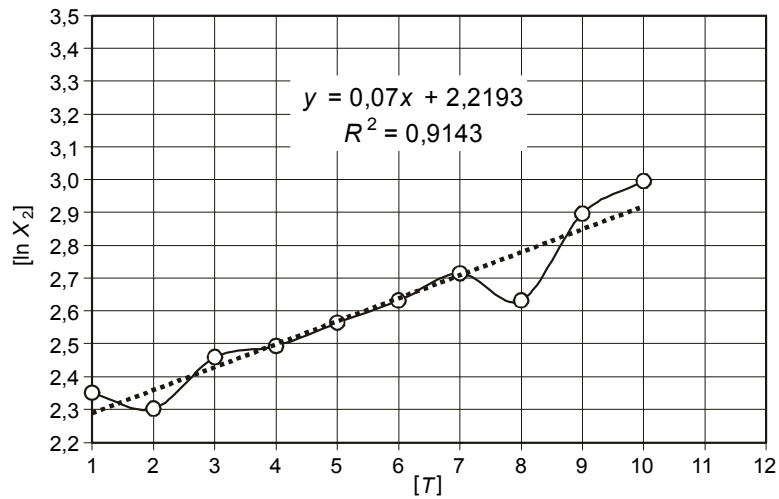
Ad 5) Określenie funkcji produkcji:

$$\ln \hat{Y} = a_0 + a_1 \cdot \ln X_2 + a_2 \cdot \ln X_3,$$

$$\ln \hat{Y} = 1,85 + 0,14 \cdot \ln X_2 + 0,72 \cdot \ln X_3.$$

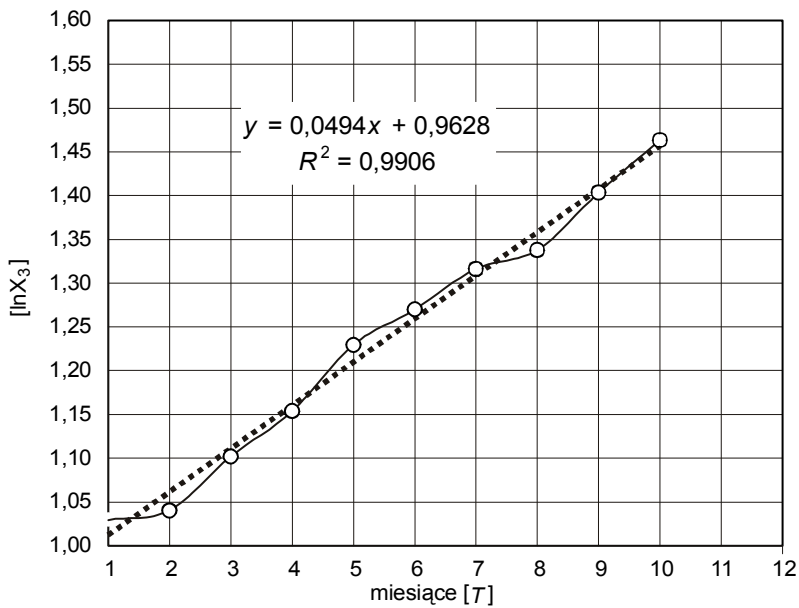
Ad 6a) Wyznaczenie wartości zmiennych $\ln X_2$, $\ln X_3$ w miesiącach 11. i 12. na podstawie trendów liniowych.

¹³¹ Ibidem.



Źródło: Opracowanie własne w Excelu, gdzie: $y \rightarrow \ln X_2$, $x \rightarrow T$.

Wykres 1.12.3. Wykres punktowy zmiennej $\ln X_2$



Źródło: Opracowanie własne w Excelu, gdzie $y \rightarrow \ln X_3$, $x \rightarrow T$.

Wykres 1.12.4. Wykres punktowy zmiennej $\ln X_3$

Obliczenia wartości zmiennych X_2 i X_3 na podstawie wykresów 1.12.3 i 1.12.4 dla miesiąca 11, 12.

$Y' = \ln X_2 = 0,07 \cdot 11 + 2,2193 = 2,989$; w miesiącu 10. było 2,996; $\ln X_2 = 0,07 \cdot 12 + 2,2193 = 3,059$.

$Y'' = \ln X_3 = 0,0494 \cdot 11 + 0,9628 = 1,506$; w miesiącu 10. – 1,463; $\ln X_3 = 0,0494 \cdot 12 + 0,9628 = 1,556$.

Prognoza według zlogarytmowanej funkcji produkcji dla:
miesiąca 11.:

$$\ln \hat{Y} = 1,85 + 0,14 \cdot \ln X_2 + 0,72 \cdot \ln X_3$$

$$\ln \hat{Y} = 1,85 + 0,14 \cdot 2,989 + 0,72 \cdot 1,506 = 3,35278,$$

$$\hat{Y} = e^{3,35278} = 28,58; \text{ w miesiącu 10. było } 28,3;$$

miesiąca 12.:

$$\ln \hat{Y} = 1,85 + 0,14 \cdot 3,059 + 0,72 \cdot 1,556 = 3,39859$$

$$\hat{Y} = e^{3,39859} = 29,92.$$

Modele:

Ad 6b) Model oryginalny: $\ln \hat{Y} = 1,85 + 0,14 \cdot \ln X_2 + 0,72 \cdot \ln X_3$,

$$\hat{Y} = e^{1,85 + 0,14 \cdot \ln X_2 + 0,72 \cdot \ln X_3} = e^{1,85} \cdot (e^{0,14})^{\ln X_2} \cdot (e^{0,72})^{\ln X_3}.$$

Do obliczenia e do określonej potęgi korzystamy z funkcji Excela [=EXP(...)] lub obliczamy jako potęgi wartości e , przy czym: $e = 2,718282$ ¹³².

$$\hat{Y} = 6,36 \cdot 1,15^{\ln X_2} \cdot 2,05^{\ln X_3}.$$

Przyjmując zaokrąglone wartości parametrów modelu, w tabeli 1.12.22 obliczono wygasłe prognozy oraz prognozy produkcji na miesiące 11. i 12. dla oszacowanych według linii trendów zmiennych objaśniających X_2 i X_3 .

T a b e l a 1.12.22

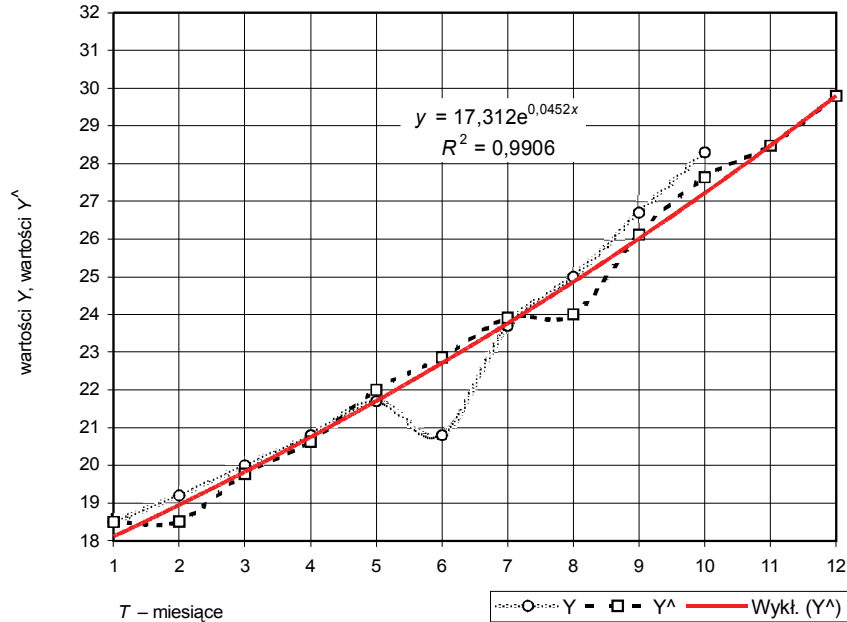
Wygasłe prognozy oraz prognozy na miesiące 11. i 12

$t = T$	Y	Y^{\wedge}	$\ln X_2$	$\ln X_3$
1	18,5	18,50	2,35	1,03
2	19,2	18,52	2,30	1,04
3	20,0	19,78	2,46	1,10
4	20,8	20,63	2,49	1,15
5	21,7	22,00	2,56	1,23
6	20,8	22,86	2,63	1,27
7	23,7	23,91	2,71	1,32
8	25,0	24,00	2,63	1,34
9	26,7	26,11	2,90	1,40
10	28,3	27,64	3,00	1,46
11		28,48	2,99	1,51
12		29,79	3,06	1,56

Ź r ó d ł o: Opracowanie własne w Excelu, gdzie:
 T – okres prognozowany, $Y^{\wedge} \rightarrow \hat{Y}$.

¹³² I. Bronsztejn, K. Siemieniediw, *Poradnik encyklopedyczny. Matematyka*, PWN, Warszawa 1959.

Model trendu bazujący na danych teoretycznych



Źródło: Opracowanie własne w Excelu, gdzie: $y^{\wedge} \rightarrow \hat{y}$.

Wykres 1.12.5. Dane empiryczne i modelowe z trendem wykładniczym

Na wykresie przedstawiono kształtowanie się w czasie dwunastu miesięcy $T = 1-12$ wartości empirycznych Y , modelowych \hat{Y} wraz z dopasowaniem do nich trendu o funkcji:

$$\hat{Y} = 17,312 \cdot e^{0,0452 \cdot t}$$

Jest to funkcja wykładnicza postaci:

$$\hat{Y} = a_0 \cdot a_1^t$$

Pozostawia się Czytelnikowi przeprowadzenie alternatywnych obliczeń dla modelu:

$$\ln Y = a_0 + a_1 \cdot \ln X_3,$$

gdyż według analizy korelacji zmienna X_2 wypada ze względu na silne powiązanie ze zmienną X_3 .

W prognozowaniu oprócz modeli liniowych regresji stosowane są również **modele nieliniowe**, w tym z jedną zmienną objaśniającą. Mogą one tworzyć złożone modele i dla uproszczenia pomijany jest wtedy parametr stały a_0 . Do nieliniowych należą modele wymienione wcześniej w podrozdz. 1.5.6.3, 1.5.6.4, 1.5.6.5, dla których przeprowadzamy transformację liniową:

- potęgowy: $Y = X^a$, $\ln Y = a \ln X$,
gdy X zmieni się o 1%, to Y o 0,25%,
- wykładniczy: $Y = a^X$, $\ln Y = X \ln a$,
gdy X zmieni się o 1, to Y przeciętnie o $a\%$,
- liniowo-logarytmiczny: $Y = a \ln X$,
 a to przeciętna zmiana Y wywołana jednoprocenową zmianą X .

W celu zastosowania metody najmniejszych kwadratów w modelach nieliniowych przeprowadzamy, jak już wspomniano, transformację liniową według podanych przykładów. Jak już wiemy z podrozdziału 1.7.1, ocenę dobroci dopasowania modeli nieliniowych oraz ich weryfikację poprzez zastosowanie odpowiednich miar i testów statystycznych przeprowadzamy analogicznie do modeli liniowych.

1.12.5. Przykłady różne z zakresu kilku zmiennych objaśniających

Przykład

Powróćmy teraz do naszego wcześniejszego przykładu (podrozdz. 1.12.4) dotyczącego zależności liczby zawieranych małżeństw M od liczby pracujących Z , których wzrost jest wyrazem możliwości stabilizacji i zakładania ogniska domowego. Do zobrazowania zastosowania modelu regresji liniowej z więcej niż jedną zmienną objaśniającą wprowadzono wcześniej w tab. 1.12.10 zmienną R „liczba rozwodów”:

$$M = f(Z, R) + \varepsilon.$$

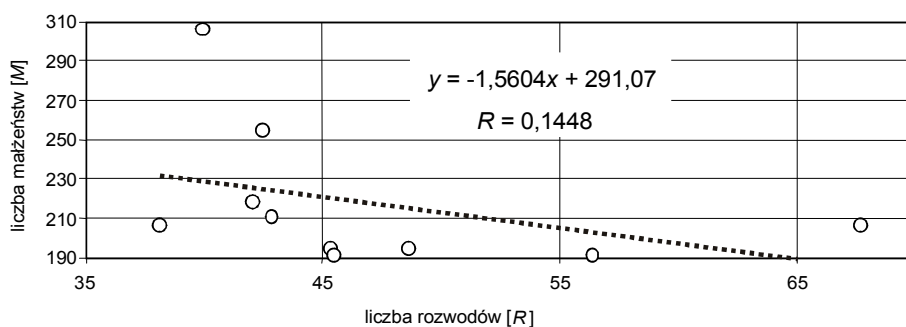
Tabela 1.12.23

Wartości zmiennych „ M ”, „ Z ”, „ R ”

		Y	X1	X2
n	Rok	M	Z	R
1	1980	307	17 334	39,8
2	1990	255	16 280	42,4
3	1995	207	14 735	38,1
4	1999	219	15 373	42
5	2000	211	15 018	42,8
6	2001	195	14 924	45,3
7	2002	192	12 729	45,4
8	2003	195	12 663	48,6
9	2004	192	12 615	56,3
10	2005	207	12 728	67,6

Źródło: Dane tab. 1.12.10, gdzie: Y , X_1 , X_2 – odpowiadające oznaczeniu zmiennych M , Z , R .

Zobaczymy na diagramie korelacyjnym, jak układa się zależność M od R . Z wykresu wynika korelacja ujemna między zmiennymi M i R . Wzrost liczby rozwodów wywołuje spadek liczby małżeństw. Ta bezpośrednia zależność liniowa jest słaba, o czym mówi współczynnik determinacji (R^2).



Źródło: Opracowanie własne, gdzie: $y \rightarrow M$, $x \rightarrow R$.

Wykres 1.12.6. Zależność zmiennych M oraz R

Oszacujmy jednak model $M = f(Z, R)$, korzystając z funkcji Excela REGLINP.

Tabela 1.12.24

Raport z realizacji funkcją REGLINP

a_2	a_1	a_0	
1,415	0,023	-180,764	parametry a_i
1,056	0,006	119,908	standardowy błąd szacunku parametrów a_i
0,752	20,632		współczynnik determinacji (R^2); odchylenie standardowe reszt (S_e)
10,585	7		statystyka F ; ilość stopni swobody (iss)
9012,139	2979,861		regresyjna suma kwadratów (ESS); resztowa suma kwadratów (RSS)

Źródło: Opracowanie własne w Excelu.

Należy zauważyć wzrost współczynnika determinacji w stosunku do sytuacji, gdy istniała tylko zmienna Z (zob. tab. 1.12.11, gdzie $R^2 = 0,6877$). Model regresji liniowej jest następujący:

$$\hat{M} = -180,764 + 0,023 \cdot Z + 1,415 \cdot R.$$

Przyjmując zgodnie z trendem w kolejnym roku wartość zmiennych Z oraz R , możemy, mając powyższy model, szacować M .

P r z y k ł a d

Wyznamy procedurę określenia spadku wagi u osoby zażywającej określony lek odchudzający w dawce 2 cm^3 przez 16 miesięcy, bazując na danych tab. 1.12.26 przedstawiających badanie skuteczności leku.

a) przyjmujemy model liniowy dwóch zmiennych:

$$\hat{Y} = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2,$$

gdzie:

X_1 – dawka zażywanego leku w cm^3 ,

X_2 – okres trwania kuracji w miesiącach,

\hat{Y} – spadek wagi w funtach (1 funt = 0,4536 kg); wartość modelowa,

b) skorzystajmy z równań normalnych wynikających z KMNK:

$$\sum Y = a_0 \cdot n + a_1 \cdot (\sum X_1) + a_2 \cdot (\sum X_2),$$

$$\sum X_1 Y = a_0 \cdot (\sum X_1) + a_1 \cdot (\sum X_1^2) + a_2 \cdot (\sum X_1 X_2),$$

$$\sum X_2 Y = a_0 \cdot (\sum X_2) + a_1 \cdot (\sum X_1 X_2) + a_2 \cdot (\sum X_2^2).$$

c) utwórzmy w Excelu tabelę pomocniczą do obliczania wymienianych składowych oraz ich sum.

T a b e l a 1.12.25

Ramka tabeli pomocniczej do obliczania składowych równań normalnych

n	Y	X_1	$(X_1)^2$	$X_1 Y$	X_2	$(X_2)^2$	$X_2 Y$	$X_1 X_2$
1								
...								
22								
	$\sum Y$	$\sum X_1$	$\sum (X_1)^2$	$\sum X_1 Y$	$\sum X_2$	$\sum (X_2)^2$	$\sum X_2 Y$	$\sum X_1 X_2$

Ź r ó d ł o: Opracowanie własne.

d) wprowadźmy sumy do równań normalnych oraz określmy macierze **A**, **B** oraz **A**₁, **A**₂, **A**₃ przez analogię do wcześniejszego przykładu,

e) zastosujmy wzory Cramera na określenie parametrów a_i naszego modelu:

$$a_0 = \det \mathbf{A}_1 / \det \mathbf{A}, \quad a_1 = \det \mathbf{A}_2 / \det \mathbf{A}, \quad a_2 = \det \mathbf{A}_3 / \det \mathbf{A},$$

gdzie:

A₁, **A**₂, **A**₃ – macierze powstałe po zastąpieniu kolumny odpowiadającej w równaniach normalnych parametrowi przez kolumnę wyrazów wolnych,

f) wstawmy obliczone parametry do modelu,

g) obliczmy spadek wagi, wprowadzając do modelu $X_1 = 2$ oraz $X_2 = 16$.

T a b e l a 1.12.26

Wartości zmiennych „Y”, „X₁”, „X₂”

Spadek wagi (funty) Y	Dawka (cm^3) X_1	Okres (miesiące) X_2
9	1,5	3
24	2,5	3
13	1,5	30
94	2	11
45	2	5
30	2	14
30	2	3

Spadek wagi (funty) Y	Dawka (cm ³) X_1	Okres (miesiące) X_2
20	2	1
15	2	4
16	2	1
10	2	1
39	2	3
15	1,5	2
19	2	6
56	3	6
45	2,5	12
37	3	36
59	5	12
26	2	24
7	1,5	2
37	3	7
34	2	18

Ź r ó d ł o: Opracowanie własne na podstawie: J. Freund, *Podstawy nowoczesnej statystyki*, PWE, Warszawa 1971, s. 336.

1.12.6. Przykłady na zastosowanie testów porównywania modeli

Już z przytoczonych wcześniej przykładów wynika, że do prognozowania możemy użyć różnych modeli, w tym modeli trendu czy regresji. Aby porównać, który z kilku modeli opisujących tę samą zależność zmiennej objaśnianej od zmiennej/zmiennych objaśniających jest najlepszy, stosujemy określone kryteria, których wzory podano w podrozdz. 1.6.7:

– kryterium Akeike'a:

$$AK = n \cdot \ln(\sum u^2) + 2k,$$

gdzie:

n – ilość obserwacji,

$u = e = z_n$ – reszta z modelu ($y - \hat{y}$),

k – liczba parametrów a_i w modelu,

– kryterium Schwarza:

$$SCHW = n \cdot \ln(\sum u^2) + k \cdot \ln(n).$$

P r z y k ł a d

Do określonej zależności zmiennej objaśnianej od zmiennych objaśniających, w tym również czasu t , dla 32 obserwacji określono trzy modele:

A – regresji wielowymiarowej z trzema zmiennymi objaśniającymi X_1, X_2, X_3 ,
B – trendu ze zmienną t ,
C – Holta-Wintersa ze zmienną t (model ten będzie omówiony w dalszej części niniejszej pracy).

Zastosujemy kryterium Schwarz, opierając się na danych umownych:

T a b e l a 1.12.27

Wartości do obliczenia kryterium Schwarz

Model	Suma kwadratów reszt ($\sum u^2$)	Liczba parametrów (k)
A	12	4
B	18	2
C	24	2

Ź r ó d ł o: Opracowanie własne, dane umowne.

$$SCHW_A = 32 \cdot \ln 12 + 4 \cdot \ln 32 = 93,38, \quad SCHW_B = 32 \cdot \ln 18 + 2 \cdot \ln 32 = 99,42,$$

$$SCHW_C = 32 \cdot \ln 24 + 2 \cdot \ln 32 = 108,63.$$

Min: 93,38; 99,42; 108,63 – model *A*.

P r z y k ł a d

Zastosujemy kryterium Akaike'a dla 20 obserwacji wyboru modeli spośród podanych w tab. 1.12.28:

T a b e l a 1.12.28

Sumy kwadratów reszt modeli

Model	Suma kwadratów reszt ($\sum u^2$)
A	15
B	20
C	25
D	10

gdzie:

A – model regresji wielowymiarowej z trzema zmiennymi objaśniającymi X_1, X_2, X_3 ,

B – model trendu liniowego ze zmienną t ,

C – model prostego wyrównywania wykładniczego ze zmienną t ,

D – model Holta-Wintersa ze zmienną t .

Ź r ó d ł o: Opracowanie własne; dane umowne.

Modele *B–D* oparte są na szeregach czasowych.

$$AK_A = 20 \cdot \ln 15 + 2 \cdot 4 = 62,16, \quad AK_B = 20 \cdot \ln 20 + 2 \cdot 2 = 63,91,$$

$$AK_C = 20 \cdot \ln 25 + 2 \cdot 2 = 68,38, \quad AK_D = 20 \cdot \ln 10 + 2 \cdot 2 = 50,05.$$

Najniższą wartość kryterium informacyjnego Akeike'a wykazuje model *D*.

1.12.7. Pytania (do 1.12.4–1.12.6)

1. Wymień i opisz fazy predykcji (przewidywania).
2. Scharakteryzuj metody prognozowania.
3. Napisz i objaśnij wzór na pierwiastek średniego kwadratu błędu.
4. Podaj i opisz elementy funkcji reprezentujących modele:
 - liniowy moduł regresji z jedną zmienną objaśniającą,
 - nieliniowe modele regresji z jedną zmienną objaśniającą stosowane w prognozowaniu.
5. Objaśnij przykłady liniowych i nieliniowych modeli z wieloma zmiennymi objaśniającymi:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k,$$

$$\hat{Y}_t = a_0 M_t^\alpha L_t^{1-\alpha}$$

$$(\ln \hat{Y}_t) = a_0 + a_1 \ln Y_{t-1} + a_2 \ln X_1 + \dots + a_{n-1} \ln X_n.$$

6. Objaśnij zastosowanie równania: $a = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$.

7. Równaniami normalnymi określono parametry modelu regresji liniowej popytu na mięso wołowe i cielęce Y , bazując na cenie wołowiny X_1 , cenie wieprzowiny X_2 oraz dochodzie na członka rodziny w określonym regionie kraju X_3 :

$$\hat{Y} = 3500 - 0,1 X_1 + 0,06 X_2 + 0,02 X_3$$

Podaj postaci ogólne równań normalnych powyższych zmiennych.

8. Opisz postępowanie przy obliczaniu prognozy Y_p dla $x_1 = 4,7$, $x_2 = 360$, $x_3 = 11$, $x_4 = 1,2$, bazując na modelu regresji liniowej:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4$$

oraz funkcji REGLINP Excela, przyjmując dane tab. 1.12.29.

Tabela 1.12.29

Wartości zmiennych

x_1	x_2	x_3	x_4	y
4,4	424	9	1,1	0,9
4,5	414	9,2	1,1	1,0
4,9	432	9,5	1,1	1,05
5,1	414	9,7	1,2	1,05
5,0	410	10,1	1,3	1,0
4,9	394	10,8	1,3	1,1
4,8	375	10,5	1,2	1,2
4,6	366	10,7	1,2	1,25
4,2	348	11,1	1,1	1,2
4,2	333	10,3	1,1	1,3
4,3	338	9,8	1,1	1,35
4,5	380	10,1	1,1	1,3

Źródło: Opracowanie własne; dane umowne.

1.12.8. Zadania (do 1.12.4–1.12.6)

- Zastosuj wzory Cramera do określenia X , Y , Z na podstawie równań:
 $X + 4Y - 2Z = 10$,
 $3X + 2Y + Z = 20$,
 $X - 10Z = 10$.
- Dane są obserwacje zmiennej objaśnianej Y oraz zmiennej objaśniającej X :

Tabela 1.12.30

Obserwacje zmiennych

x	y
10	40
20	60
30	70
40	75
50	80
60	89
70	92
80	96

Źródło: Opracowanie własne; dane umowne.

Określ parametry a_0, a_1 modelu regresji $\hat{Y} = a_0 + a_1X$ podanymi sposobami wynikającymi z KMNK $\sum(Y - \hat{Y})^2 \rightarrow \min$, a następnie oblicz prognozę y_p , jeśli $x = 55$:

– zastosowanie wzorów (1.5.1) i (1.5.2):

$$a_1 = \frac{\sum_{t=1}^n (y_t - \bar{y}) \cdot (x - \bar{x})}{\sum_{t=1}^n (x - \bar{x})^2}, \quad a_0 = \bar{y} - a_1 \cdot \bar{x},$$

– użycie równań normalnych do dwóch parametrów a_0 oraz a_1 :

$$\begin{aligned} \sum Y &= n \cdot a_0 + a_1 \cdot \sum X, \\ \sum XY &= a_0 \cdot \sum X + a_1 \cdot \sum X^2, \end{aligned}$$

– zastosowanie wzorów (w miejsce \bar{Y}, \bar{X} wstawić wartości zmiennych):

$$\begin{aligned} a_0 &= [(\sum Y)(\sum X^2) - (\sum X)(\sum XY)] / [n(\sum X^2) - (\sum X)^2], \\ a_1 &= [n(\sum XY) - (\sum X)(\sum Y)] / [n(\sum X^2) - (\sum X)^2], \end{aligned}$$

– określenie regresji przez sporządzenie wykresu w Excelu.

3. Określono wektor oraz macierz korelacji między pięcioma zmiennymi objaśniającymi X_1, X_2, X_3, X_4, X_5 .

Przeprowadź analizę korelacyjną doboru zmiennych objaśniających.

$$\mathbf{R}_0 = \begin{bmatrix} 0,58 \\ 0,86 \\ 0,48 \\ 0,87 \\ 0,83 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0,79 & 0,26 & 0,64 & 0,10 \\ 0,79 & 1 & 0,33 & 0,86 & 0,59 \\ 0,26 & 0,33 & 1 & 0,17 & 0,51 \\ 0,64 & 0,86 & 0,17 & 1 & 0,62 \\ 0,10 & 0,59 & 0,51 & 0,62 & 1 \end{bmatrix}.$$

4. Określ prognozę dla $x_1 = 3, x_2 = 8$ po wcześniejszym określeniu parametrów funkcji regresji liniowej $\hat{Y} = a_0 + a_1X_1 + a_2X_2$, korzystając z rachunku macierzowego na określenie wektora $a = (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T\mathbf{y}$.

Tabela 1.12.31

Obserwacje zmiennych „Y”, „X₁”, „X₂”

y	x_1	x_2
9	1,5	3
24	2,5	5
13	1,5	9
34	2	11
45	2	5
30	1,5	3
30	2	6
20	1,5	2
15	1,5	4
16	2	2

Źródło: Opracowanie własne; dane umowne.

1.12.9. Modele oparte na szeregach czasowych

Opracowanie i stosowanie odpowiednich modeli regresji, zwłaszcza o wielu zmiennych objaśniających X_j , jest pracochłonne i kłopotliwe w bieżącym pozyskiwaniu wartości tych zmiennych. Wymaga szacowania przyszłych wartości zmiennych objaśniających. Z tego względu zwrócono uwagę na budowanie modeli na podstawie szeregów czasowych, które jednak mogą charakteryzować się dużą zmiennością oraz sezonowością. W prognozowaniu opartym na obserwacji zmiennych objaśnianych Y w kolejnych okresach czasu t stosowane są modele:

- oparte na trendzie pelzającym (omówiony został w podrozdz. 1.5.6.2),
- prostych średnich ruchomych,
- ważonych średnich ruchomych,
- prostego wyrównywania wykładniczego,
- adaptacyjne wyrównywania wykładniczego,
- podwójnych średnich ruchomych.

Modele oparte na szeregach czasowych zaprezentowane zostaną na podstawie danych statystycznych opisujących spożycia masła w kg w przeliczeniu na jednego mieszkańca Polski w latach 1994–2005.

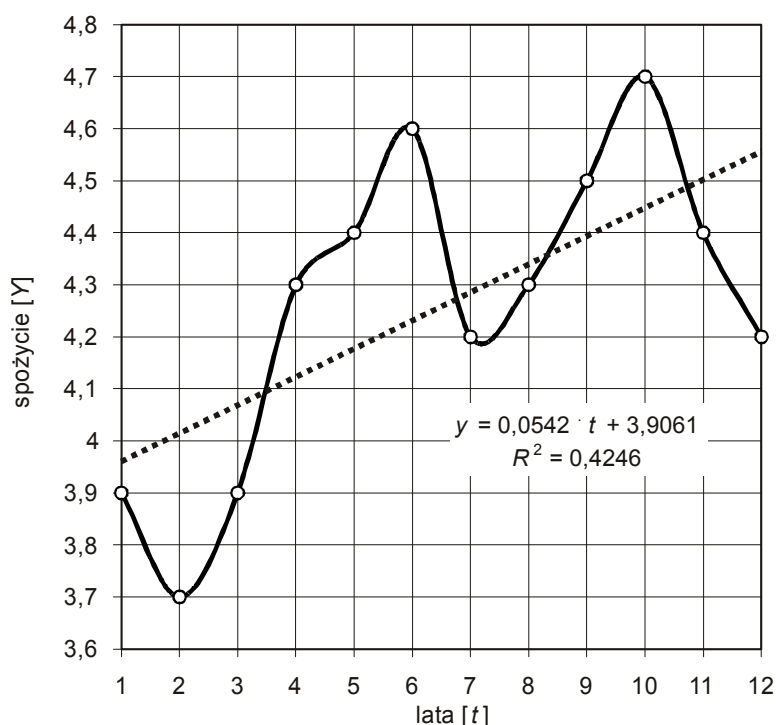
Tabela 1.12.32

Indywidualne spożycie masła [kg]

Rok t	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Okres	1	2	3	4	5	6	7	8	9	10	11	12
Spożycie y	3,9	3,7	3,9	4,3	4,4	4,6	4,2	4,3	4,5	4,7	4,4	4,2

Źródło: Opracowanie własne, fragment tab. 1.8.5.

Przedstawmy zależność Y od X na wykresie punktowym.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.7. Indywidualne spożycie masła [kg]

W spożyciu masła w latach od 1994 do 2005 wystąpił trend dodatni i cykliczność ze zmienną amplitudą. Spożycie tego tłuszczu zależy od mody i intensywności reklam na inne tłuszcze substytucyjne (margaryna, olej jadalny, smalec); w okresie 12 lat wykazuje wyraźne tendencje wzrostu. Analizując ponadto dodatkowe dane kwartalne spożycia masła w ciągu roku, spodziewać się można wystąpienia sezonowości, gdyż zapotrzebowanie na masło w po-

rach chłodnych jest większe niż w gorących. W przykładach zrealizowanych w Excelu zastosowano oznaczenie wartości zmiennych \bar{Y} oraz prognoz F dużymi literami.

1.12.9.1. Modele prostych średnich ruchomych

W modelach tych, oznaczonych skrótem SMA, przyjmuje się określone liczby okresów do obliczenia średniej arytmetycznej, np.: 2, 3, 4. W tabeli 1.12.33 obliczono wygasłe prognozy, opierając się na średnich arytmetycznych dla dwóch okresów (SMA2) oraz czterech okresów (SMA4), a ponadto prognozie na okres przyszły – trzynasty. W wierszu końcowym obliczono pierwiastek średniego kwadratu błędu dla wygasłych prognoz *ex post* ze wspólnego przedziału 5–12 okresów dla SMA2 i SMA4, korzystając z analogicznego wzoru jak (1.12.1):

$$RMSE = \sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} (F_n - Y_n)^2} = \sqrt{\frac{1}{1+7} \sum_5^{5+7} (F_n - Y_n)^2}.$$

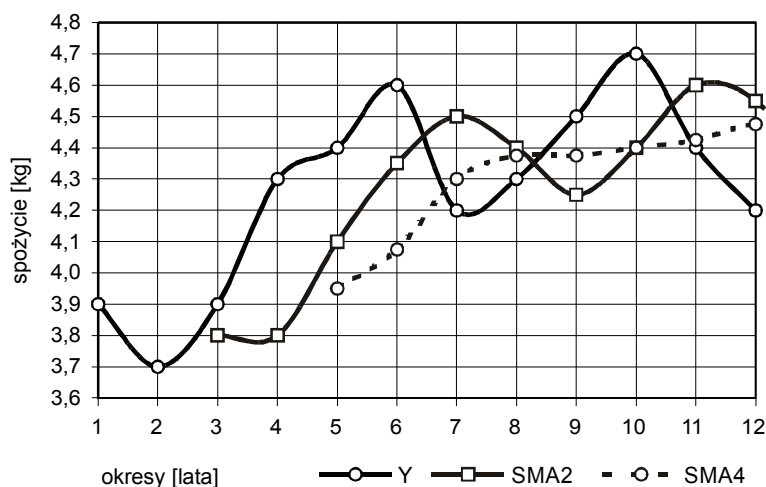
Tabela 1.12.33

Obliczenie wygasłych prognoz i przyszłej według modeli SMA2 i SMA4

Okres	Spożycie	SMA2			SMA4		
		F_n	$Y_n - F_n$	$(Y_n - F_n)^2$	F_n	$Y_n - F_n$	$(Y_n - F_n)^2$
1	3,9						
2	3,7						
3	3,9	3,80	0,10	0,01			
4	4,3	3,80	0,50	0,25			
5	4,4	4,10	0,30	0,09	3,95	0,45	0,20
6	4,6	4,35	0,25	0,06	4,08	0,53	0,28
7	4,2	4,50	-0,30	0,09	4,30	-0,10	0,01
8	4,3	4,40	-0,10	0,01	4,38	-0,08	0,01
9	4,5	4,25	0,25	0,06	4,38	0,13	0,02
10	4,7	4,40	0,30	0,09	4,40	0,30	0,09
11	4,4	4,60	-0,20	0,04	4,43	-0,02	0,00
12	4,2	4,55	-0,35	0,12	4,48	-0,27	0,08
13	-	4,30	$\sqrt{8}$	0,07	4,45		0,08
			RMSE	0,266			0,291

Źródło: Opracowanie własne w Excelu.

Minimum to $RMSE = 0,266$ dla metody SMA2. Przedstawmy graficznie obliczone prognozy.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.8. Empiryczne i modelowe wartości indywidualnego spożycia masła

Modele prostych średnich ruchomych mają zastosowanie w prognozowaniu krótkookresowym. Większa liczba okresów w obliczaniu średnich daje bardziej wygładzony szereg prognozowany.

Sprawdzenie występowania trendu w szeregu czasowym zmiennej Y uzyskamy przez obliczenie pierwszych różnic (różnica $d = 1$) $\Delta Y = Y_n - Y_{n-1}$, a następnie statystyki t -Studenta.

Tabela 1.12.34

Szereg pierwszych różnic

Y_n	Y_{n-1}	$Y_n - Y_{n-1}$
3,9		
3,7	3,9	-0,2
3,9	3,7	0,2
4,3	3,9	0,4
4,4	4,3	0,1
4,6	4,4	0,2
4,2	4,6	-0,4
4,3	4,2	0,1
4,5	4,3	0,2
4,7	4,5	0,2
4,4	4,7	-0,3
4,2	4,4	-0,2
Srednia		0,027
Odchylenie		0,257

Źródło: Opracowanie własne w Excelu.

Do obliczenia średniej oraz odchylenia standardowego (próby) szeregu pierwszych różnic zastosowano wzory analogiczne do (1.3.1), (1.3.13)¹³³:

$$\Delta \bar{Y} = \frac{\sum_{i=2}^n \Delta y_i}{n}, \quad S_x = \sqrt{\sum S_x^2}, \quad S_x^2 = \frac{\sum_{i=1}^n (\Delta y_i - \Delta \bar{Y})^2}{n-1},$$

oraz

$$t_{obl} = \frac{\Delta \bar{Y}}{S_x \cdot \sqrt{n-d}} > t_{0,05}.$$

Przeprowadzamy weryfikację hipotez:

H_0 : średnia arytmetyczna pierwszych różnic równa się zero – w szeregu nie występuje trend,

H_1 : średnia arytmetyczna pierwszych różnic jest różna od zera – w szeregu występuje trend.

Gdy $t_{obl} < t_{\alpha}$, to nie ma podstaw do odrzucenia hipotezy zerowej. Przyjęto poziom istotności $\alpha = 0,05$. Według tablicy wartości krytycznych rozkładu t -Studenta dla $\alpha = 0,05$ i $iss = n - 1 = 12 - 1 = 11$, $t_{(0,05; 11)} = 2,2010$.

$$t_{obl} = \frac{0,027}{0,257 \cdot \sqrt{12-1}} = 0,0317.$$

Ponieważ $t_{obl} < t_{0,05; 11}$, wobec tego w szeregu nie występuje trend liniowy.

1.12.9.2. Modele ważonych średnich ruchomych

Weźmy pod uwagę model SMA4, lecz nadajmy wagi w_i kolejnym obserwacjom brany po cztery (0,1; 0,2; 0,3; 0,4). Wtedy prognoza na okres piąty, liczona jako średnia ważona, wyniesie:

$$F_{05} = 0,1 \cdot y_1 + 0,2 \cdot y_2 + 0,3 \cdot y_3 + 0,4 \cdot y_4,$$

$$F_{05} = 0,1 \cdot 3,9 + 0,2 \cdot 3,7 + 0,3 \cdot 3,9 + 0,4 \cdot 4,3 = 4,02.$$

¹³³ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 1: *Ekonometria...*, op.cit., s. 69.

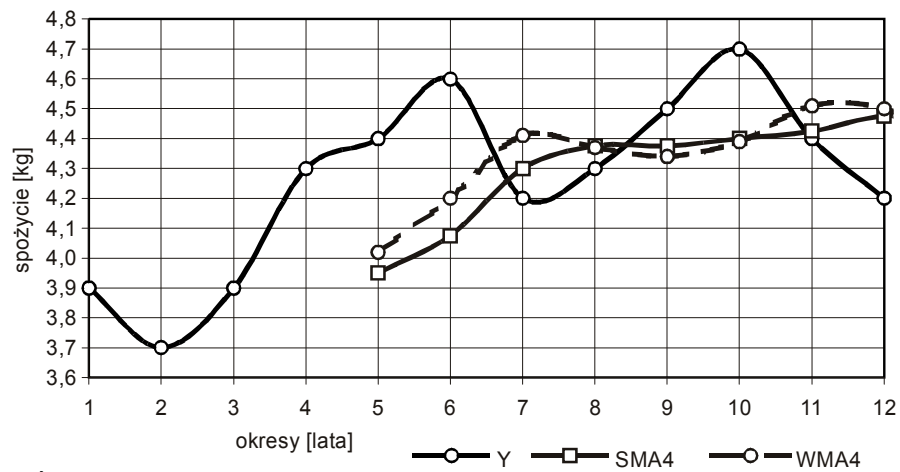
Tabela 1.12.35

Określenie prognozy na okres 13 metodą WMA4

Okres t	Spożycie Y_n	SMA4			WMA4		
		F_n	$Y_n - F_n$	$(Y_n - F_n)^2$	F_n	$Y_n - F_n$	$(Y_n - F_n)^2$
1	3,9						
2	3,7						
3	3,9						
4	4,3						
5	4,4	3,95	0,45	0,20	4,02	0,38	0,14
6	4,6	4,08	0,53	0,28	4,20	0,40	0,16
7	4,2	4,30	-0,10	0,01	4,41	-0,21	0,04
8	4,3	4,38	-0,08	0,01	4,37	-0,07	0,00
9	4,5	4,38	0,13	0,02	4,34	0,16	0,03
10	4,7	4,40	0,30	0,09	4,39	0,31	0,10
11	4,4	4,43	-0,02	0,00	4,51	-0,11	0,01
12	4,2	4,48	-0,27	0,08	4,50	-0,30	0,09
13	-	4,45	$\sqrt{8}$	0,084	4,39		0,072
			RMSE	0,291			0,269

Źródło: Opracowanie własne w Excelu.

RMSE dla średniej ważonej czterookresowej jest mniejszy.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.9. Porównanie wartości empirycznych i modelowych według metod SMA4 oraz WMA4

1.12.9.3. Model prostego wyrównywania wykładniczego

Model ten o skrócie SES określony jest funkcją¹³⁴:

$$F_n = \alpha \cdot Y_{n-1} + (1-\alpha) \cdot F_{n-1}, \quad (1.12.7)$$

gdzie:

α – parametr wyrównania (0;1); gdy celem jest silne wygładzanie szeregu, to α powinno mieć małą wartość, np. 0,1 lub 0,2.

Model wymaga znajomości wartości początkowej, którą przyjmuje się $F_2 = Y_1$. Prognoza w modelu SES na kolejny okres jest ważoną średnią ruchomą wartości rzeczywistej z okresu bezpośrednio poprzedzającego oraz wartości prognozowanej. Między modelem prostej średniej ruchomej, o liczbie okresów w , a modelem prostego wyrównywania wykładniczego, o parametrze α , istnieje zależność:

$$w = 2/\alpha - 1. \quad (1.12.8)$$

Na przykład jeżeli $w = 4$ okresy, to $\alpha = 2/(w + 1) = 2/5 = 0,4$. Możemy budować różne modele SES, dobierając α i kierując się minimalną wartością miary *RMSE*.

Obliczmy kolejne prognozy dla naszego przykładu spożycia masła, przyjmując $\alpha = 0,7$; $F_2 = Y_1$, a następnie, w celu porównania z wcześniejszymi metodami dla ośmiu ostatnich okresów, obliczmy *RMSE*.

Tabela 1.12.36

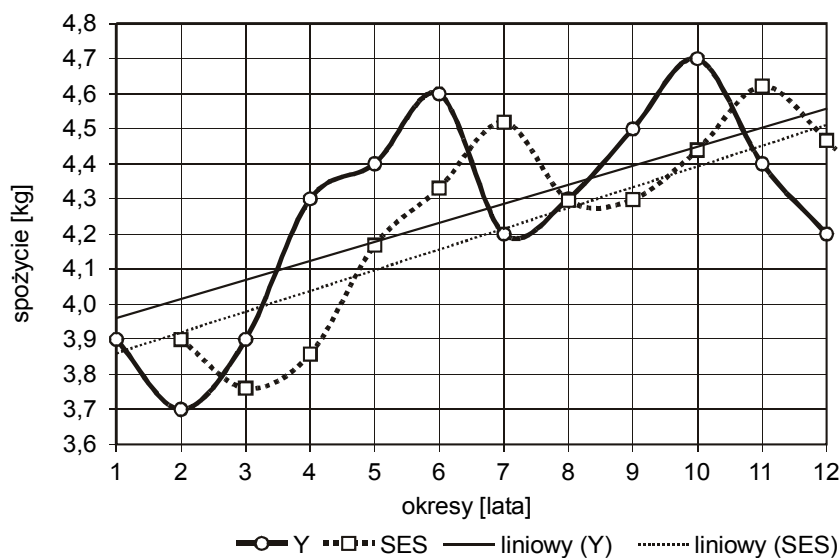
Obliczenie prognoz oraz określenie miary błędu prognoz *RMSE*

Okres t	Spożycie Y_n	SES		
		F_n	$Y_n - F_n$	$(Y_n - F_n)^2$
1	3,9			
2	3,7	3,9	-0,200	0,040
3	3,9	3,76	0,140	0,020
4	4,3	3,86	0,442	0,195
5	4,4	4,17	0,233	0,054
6	4,6	4,33	0,270	0,073
7	4,2	4,52	-0,319	0,102
8	4,3	4,30	0,004	0,000
9	4,5	4,30	0,201	0,041
10	4,7	4,44	0,260	0,068
11	4,4	4,62	-0,222	0,049
12	4,2	4,47	-0,267	0,071
13	-	4,28	$\Sigma/8$	0,057
			<i>RMSE</i>	0,239

Źródło: Opracowanie własne w Excelu.

¹³⁴ Ibidem, s. 71.

Z porównania modeli SMA2, SMA4, WMA4, SES, według miary błędu *ex post*, najniższy *RMSE* ma model SES.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.10. Porównanie wartości empirycznych i modelu SES

1.12.9.4. Model adaptacyjny wyrównywania wykładniczego

W modelu adaptacyjnym wyrównywania wykładniczego ARRES parametr α znany nam z modelu SES [$F_n = F_{n-1} + \alpha (Y_{n-1} - F_{n-1})$] jest dostosowywany do każdego następnego okresu. Zastąpiony został sygnałem adaptacyjnym TST_n ¹³⁵:

$$TST_n = \frac{SAD_n}{MAD_n}, \quad (1.12.9)$$

gdzie:

TST_n – sygnał adaptacyjny w okresie n , który staje się parametrem α w następnym okresie; przyjmuje wartości z przedziału (0;1).

$$SAD_n = \beta \cdot (Y_n - F_n) + (1 - \beta) \cdot SAD_{n-1},$$

$$MAD_n = \beta \cdot |Y_n - F_n| + (1 - \beta) \cdot MAD_{n-1}.$$

¹³⁵ Ibidem, s. 75.

Dodatkowy parametr β wprowadzony do średnich ważonych odchylenia SAD_n i MAD_n przyjmowany jest wstępnie jako 0,2.

Prognozę na kolejny okres obliczamy ze wzoru¹³⁶:

$$F_n = F_{n-1} + TST_{n-1} \cdot (Y_{n-1} - F_{n-1}). \quad (1.12.10)$$

Do obliczenia prognoz według modelu ARRES przyjmujemy wartości początkowe:

$$TST_1 = 0,2; F_2 = Y_1; SAD_1 = 0; MAD_1 = Y_2 - F_2 \quad TST_1 = \alpha.$$

Przykład

Zastosujemy wyrażenia Excela do określenia prognoz modelem ARRES oraz obliczenia błędu $RMSE$.

Tabela 1.12.37

Określenie wygasłych prognoz oraz prognozy przyszłej na okres 13.

1	B	C	D	E	F	G	H	I	J
2	Okres	Spożycie	SAD_n	MAD_n	(SAD_n/MAD_n)	TST_n	ARRES		
3		Y_n					F_n	$Y_n - F_n$	$(Y_n - F_n)^2$
4	1	3,9	0	-0,200		0,200			
5	2	3,7	-0,040	-0,120	0,333	0,333	3,900	-0,200	0,040
6	3	3,9	0,008	-0,056	-0,143	0,143	3,833	0,200	0,040
7	4	4,3	0,086	0,035	2,455	2,455	3,900	0,400	0,160
8	5	4,4	0,089	0,048	1,850	1,850	3,957	0,100	0,010
9	6	4,6	0,111	0,079	1,417	1,417	4,203	0,200	0,040
10	7	4,2	0,009	0,143	0,063	0,063	4,573	-0,400	0,160
11	8	4,3	0,027	0,134	0,203	0,203	4,006	0,100	0,010
12	9	4,5	0,062	0,147	0,419	0,419	4,012	0,200	0,040
13	10	4,7	0,089	0,158	0,566	0,566	4,053	0,200	0,040
14	11	4,4	0,012	0,186	0,062	0,062	4,137	-0,300	0,090
15	12	4,2	-0,031	0,189	-0,163	0,163	3,967	-0,200	0,040
16	13	-					3,954	$\Sigma/8$	0,054
17								$RMSE$	0,232

Źródło: Opracowanie własne w Excelu, gdzie pierwsza kolumna to numery wierszy, a pierwszy wiersz to kolejne kolumny arkusza kalkulacyjnego.

Przykładowo zastosowane wyrażenia Excela dla okresu trzeciego:

$$Y_3 - F_3 = C6 - W5,$$

$$(Y_3 - F_3)^2 = I6^2$$

$$SAD_3 = 0,2 * I6 + (1 - 0,2) * D5$$

$$MAD_3 = 0,2 * \text{MODUŁ.LICZBY}(I6) + (1 - 0,2) * E5$$

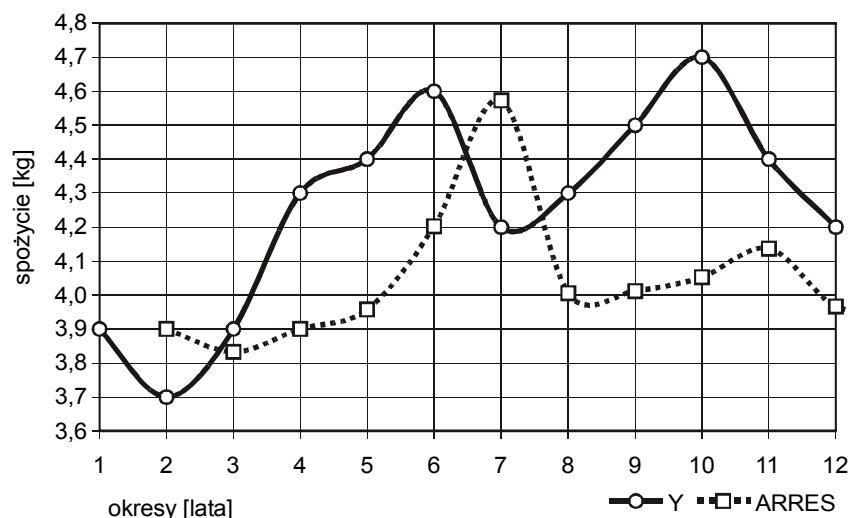
$$F_3 = H5 + (G5 * I5)$$

$$TST_3 = \text{MODUŁ.LICZBY}(F6)$$

¹³⁶ Ibidem.

Gdy porównamy wyniki $RMSE$ według pięciu dotychczasowych modeli, to z punktu miary $RMSE$ najlepszy jest model adaptacyjny wyrównywania wykładniczego ARRES.

Modelu adaptacyjnego ARRES nie stosuje się w sytuacji występowania trendu i wahań okresowych w szeregu zmiennej objaśnianej. O braku statystycznej istotności trendu wiemy z wcześniejszego testu.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.11. Wartości empiryczne i teoretyczne modelu ARRES

1.12.9.5. Modele podwójnych średnich ruchomych

Rozpatrzmy przykład z trendem. Pod uwagę bierzemy dziesięciolecie (lata 80.), w którym spożycie masła wzrosło z 6 do 8,9 kg na mieszkańca. Zakładamy, że w poszczególnych dziesięciu latach przyrost był jednakowy tj. $2,9/10 = 0,29$ kg.

Zbudujmy na tym założeniu model podwójnych średnich ruchomych bazujący na średniej arytmetycznej zmiennej Y w trzech kolejnych okresach. Średnia ruchoma trzyokresowa $MA3$ jest prognozą F_n na okres czwarty.

Tabela 1.12.38

Prognozy w modelu MA3

Okres	Y	MA 3	Y - MA 3 (trend)	F_n (MA 3)	$Y - F_n$
1	6,290				
2	6,580				
3	6,870	6,580	0,290		
4	7,160	6,870	0,290	6,580	0,580
5	7,450	7,160	0,290	6,870	0,580
6	7,740	7,450	0,290	7,160	0,580
7	8,030	7,740	0,290	7,450	0,580
8	8,320	8,030	0,290	7,740	0,580
9	8,610	8,320	0,290	8,030	0,580
10	8,900	8,610	0,290	8,320	0,580

Źródło: Opracowanie własne w Excelu.

Zauważamy występowanie błędu systematycznego $(Y - F_n) = 0,580$, który równa się $(2 \cdot \text{trend}) = 2 \cdot 0,290 = 0,580$.

W eliminacji tego błędu stosuje się podwójne średnie ruchome $MA(M \times W)$.

gdzie:

W – okres pierwszej średniej ruchomej, w naszym przykładzie – 3,

M – okres drugiej średniej ruchomej niech będzie również – 3.

Tabela 1.12.39

Obliczenie przykładowe dla modelu $MA(M \times W)$

a	b	c	d	e	f	g	h
Okres	Y	MA 3	b - c trend	MA (3 x 3)	c - e	$F_n = c + f +$ trend	$Y - F_n$
1	6,290						
2	6,580						
3	6,870	6,580	0,290				
4	7,160	6,870	0,290				
5	7,450	7,160	0,290	6,870	0,290		
6	7,740	7,450	0,290	7,160	0,290	7,740	0
7	8,030	7,740	0,290	7,450	0,290	8,030	0
8	8,320	8,030	0,290	7,740	0,290	8,320	0
9	8,610	8,320	0,290	8,030	0,290	8,610	0
10	8,900	8,610	0,290	8,320	0,290	8,900	0

Źródło: Opracowanie własne w Excelu.

Przykładowa prognoza dla czerwca przy umownej stałej wartości trendu 0,29:

$$F_6 = MA(3)_5 + [MA(3)_5 - MA(3 \times 3)_5] + \text{trend},$$

$$F_6 = 7,160 + [7,160 - 6,870] + 0,290 = 7,740.$$

Trend szacuje się według wzoru (1.12.11), przy czym W to okres pierwszej średniej ruchomej¹³⁷. Do naszego przykładu o stałej wartości trendu:

$$\text{trend} = \frac{2}{W-1} (MA(W)_n - MA(M \times W)_n), \quad (1.12.11)$$

$$\text{trend} = \frac{2}{3-1} (7,160)_5 - 6,870_5 = 0,290.$$

Trend jest zatem różnicą między pojedynczą a podwójną średnią ruchomą. Gdy budujemy prognozę na m okresów, wówczas wartość prognozowaną określamy ze wzoru:

$$F_{n+m} = a_n + b_n \cdot m, \quad (1.12.12)$$

gdzie:

$$a_n = S'_n + (S'_n - S''_n) = 2 S'_n - S''_n - \text{uśredniona wartość zmiennej},$$

$$S'_n = (Y_n + Y_{n-1} + \dots + Y_{n-W+1})/W,$$

$$S''_n = (S'_n + S'_{n-1} + \dots + S'_{n-M+1})/M,$$

$$b_n = [2/(W-1)] (S'_n - S''_n) - \text{uśredniona wartość trendu}.$$

Model podwójnych średnich ruchomych, który do celów informatycznych nazwiemy SMAD, jest stosowany, gdy występuje trend i wahania przypadkowe, lecz bez wahań sezonowych.

P r z y k ł a d

Indywidualne spożycie masła (w kg) w ciągu dwudziestolecia (lata 70. i 80.) według danych statystycznych wzrosło od 4,91 do 8,9 kg, czyli średnio $(8,9 - 4,91)/20 = 0,2 =$ = kg/rok. Przyjmijmy, że przyrost ten był nierównomierny w poszczególnych latach.

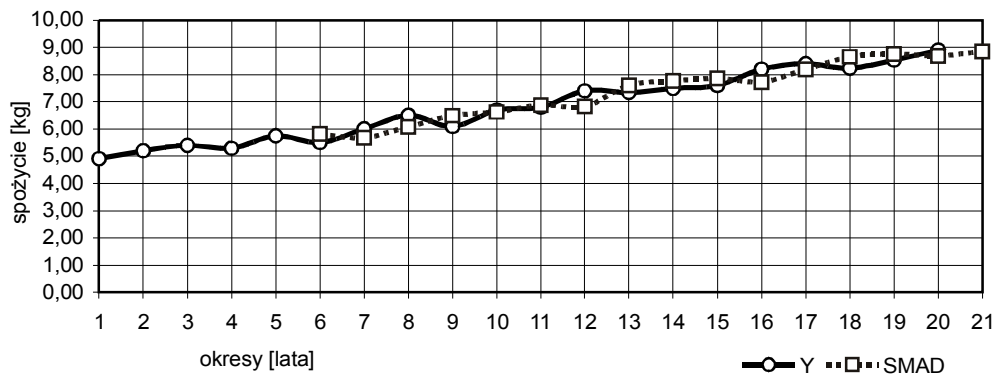
¹³⁷ Ibidem, rozdz. 2.12.

Tabela 1.12.40

Obliczenia pomocnicze oraz określenie prognoz w SMAD

1	2	3	4	5	6	7	8	9
N	Y	$MA3$ S'_n	$MA(3 \times 3)$ S''_n	$2S'_n - S''_n$ a_n	$S'_n - S''_n$ b_n	$F_{n+1} = a_n + b_n$ $m = 1$	$Y - F_n$ błąd	$(Y - F_n)^2$
1	4,91							
2	5,20							
3	5,40	5,17						
4	5,30	5,300						
5	5,75	5,483	5,318	5,649	0,166			
6	5,50	5,517	5,433	5,600	0,083	5,814	-0,314	0,099
7	6,00	5,750	5,583	5,917	0,167	5,683	0,317	0,100
8	6,50	6,000	5,756	6,244	0,244	6,083	0,417	0,174
9	6,10	6,200	5,983	6,417	0,217	6,489	-0,389	0,151
10	6,70	6,433	6,211	6,656	0,222	6,633	0,067	0,004
11	6,80	6,533	6,389	6,678	0,144	6,878	-0,078	0,006
12	7,40	6,967	6,644	7,289	0,322	6,822	0,578	0,334
13	7,35	7,183	6,894	7,472	0,289	7,611	-0,261	0,068
14	7,50	7,417	7,189	7,644	0,228	7,761	-0,261	0,068
15	7,60	7,483	7,361	7,606	0,122	7,872	-0,272	0,074
16	8,20	7,767	7,556	7,978	0,211	7,728	0,472	0,223
17	8,40	8,067	7,772	8,361	0,294	8,189	0,211	0,045
18	8,24	8,280	8,038	8,522	0,242	8,656	-0,416	0,173
19	8,54	8,393	8,247	8,540	0,147	8,764	-0,224	0,050
20	8,90	8,560	8,411	8,709	0,149	8,687	0,213	0,046
21					Prognoza	8,858	Σ	1,615
							$\Sigma/15$	0,108
							RMSE	0,328

Źródło: Opracowanie własne w Excelu; dane umowne bazujące na danych statystycznych dla dwudziestolecia (lata 70. i 80.), gdzie: N – kolejne lata 1970–1989.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.12. Graficzny obraz dopasowania danych modelowych do empirycznych (zastosowanie SMAD) oraz określenie prognozy na okres 21.

1.12.9.6. Model Browna

Metoda podwójnego wyrównywanie wykładniczego, zwana modelem Browna, stanowi analogię do metody podwójnych średnich ruchomych SMAD. Zastosowano w nim parametr wyrównujący α dla dwóch procedur wyrównywania. Oblicza się różnice między pojedynczo S'_n i podwójnie S''_n wygładzonymi wartościami zmiennej w celu estymacji trendu. Model Browna zapisany jest w postaci równań¹³⁸:

$$S'_n = \alpha Y_n + (1 - \alpha) \cdot S'_{n-1}, \quad S''_n = \alpha S'_n + (1 - \alpha) \cdot S''_{n-1},$$

$$a_n = S'_n + (S'_n - S''_n) = 2 S'_n - S''_n, \quad b_n = [\alpha/(1 - \alpha)] \cdot (S'_n - S''_n),$$

$$F_{n+m} = a_n + b_n \cdot m.$$

Przed rozpoczęciem obliczeń przyjmujemy następujące wartości początkowe:

$$S'_1 = S''_1 = Y_1, \quad a_1 = Y_1, \quad b_1 = [(Y_2 - Y_1) + (Y_4 - Y_3)]/2.$$

Metodą prób i błędów dokonujemy doboru parametru wyrównującego α , tak aby osiągnąć minimalną wartość miary błędu *RMSE* prognoz *ex post*, rozpoczynając od $\alpha = 0,1$.

Przykład

Przedstawmy obliczenia prognoz i błędów według modelu Browna na podstawie danych z tab. 1.12.41.

Tabela 1.12.41

Obliczenia pomocnicze oraz określenie prognozy w modelu Browna nazwanego SESD

A	B	C	D	E	F	G	H	I	J	
1	$\alpha =$	0,1	\$C\$154		$m =$	1				
2	Okres	Y_n	S'_n	S''_n	a_n	b_n	$F_{n+m} = a_n + b_n \cdot m$	$Y - F_n$ (błąd)	$(Y - F_n)^2$	
3	1	4,91	4,910	4,910	4,910	0,095				
4	2	5,20	4,939	4,913	4,965	0,003	5,005	0,195	0,038	
5	3	5,40	4,985	4,920	5,050	0,007	4,968	0,432	0,187	
6	4	5,30	5,017	4,930	5,103	0,010	5,057	0,243	0,059	
7	5	5,75	5,090	4,946	5,234	0,016	5,113	0,637	0,406	
8	6	5,50	5,131	4,964	5,298	0,019	5,250	0,250	0,062	
9	7	6,00	5,218	4,990	5,446	0,025	5,316	0,684	0,468	
10	8	6,50	5,346	5,025	5,667	0,036	5,471	1,029	1,058	
11	9	6,10	5,421	5,065	5,778	0,040	5,702	0,398	0,158	
12	10	6,70	5,549	5,113	5,985	0,048	5,818	0,882	0,779	
13	11	6,80	5,674	5,169	6,179	0,056	6,034	0,766	0,587	
14	12	7,40	5,847	5,237	6,457	0,068	6,235	1,165	1,356	
15	13	7,35	5,997	5,313	6,681	0,076	6,524	0,826	0,682	
16	14	7,50	6,148	5,397	6,898	0,083	6,757	0,743	0,552	
17	15	7,60	6,293	5,486	7,099	0,090	6,982	0,618	0,382	
18	16	8,20	6,483	5,586	7,381	0,100	7,189	1,011	1,022	
19	17	8,40	6,675	5,695	7,655	0,109	7,481	0,919	0,845	
20	18	8,24	6,832	5,809	7,855	0,114	7,764	0,476	0,226	
21	19	8,54	7,002	5,928	8,077	0,119	7,968	0,572	0,327	
22	20	8,90	7,192	6,054	8,330	0,126	8,196	0,704	0,495	
23	21	Prognoza					8,456	Σ	9,688	
24	22	Prognoza					8,583	$\Sigma/19$	0,510	
25								<i>RMSE</i>	0,714	

Źródło: Opracowanie własne w Excelu.

¹³⁸ Ibidem, rozdz. 2.13.

Zastosowane formuły do otrzymania przykładowych wyników zaznaczono czcionką półgrubą w tabeli 1.12.41:

$$\begin{aligned}
 4,939 &= \text{C\$154} * \text{C157} + (1 - \text{C\$154}) * \text{D156} \\
 4,913 &= \text{C\$154} * \text{D157} + (1 - \text{C\$154}) * \text{E156} \\
 4,965 &= 2 * \text{D157} - \text{E157} \\
 0,003 &= (\text{C\$154} / (1 - \text{C\$154})) * (\text{D157} - \text{E157}) \\
 5,005 &= \text{F156} + \text{G156} * \text{G\$154} \\
 0,195 &= \text{C157} - \text{H157} \\
 0,038 &= \text{B157}^2 \\
 0,714 &= \text{PIERWIASTEK}(\text{J177})
 \end{aligned}$$

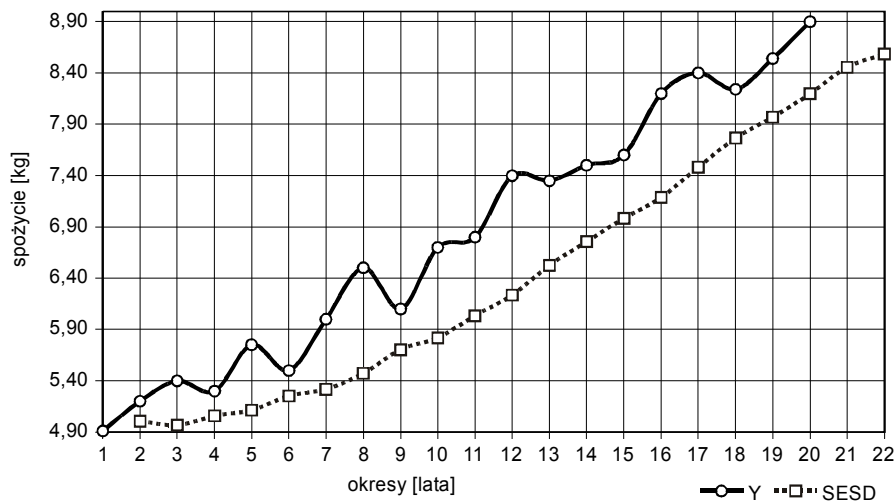
Przykładowe obliczenia:

$$\begin{aligned}
 S'_1 &= S''_1 = Y_1 = 4,91 \quad a_1 = Y_1 = 4,910, \\
 b_1 &= [(5,20 - 4,91) + (5,30 - 5,40)] / 2 = 0,095, \\
 S'_2 &= \alpha Y_2 + (1 - \alpha) \cdot S'_1 = 0,1 \cdot 5,20 + (1 - 0,1) \cdot 4,910 = 4,939, \\
 S''_2 &= \alpha S'_2 + (1 - \alpha) S''_1 = 0,1 \cdot 4,935 + (1 - 0,1) \cdot 0,4,910 = 4,913, \\
 F_2 &= a_1 + b_1 \cdot 1 = 4,910 + 0,095 = 5,005.
 \end{aligned}$$

Mając dane dla okresu 20., można określić prognozy dla okresów, np. 21 oraz 22:

$$F_{20+1} = a_{20} + b_{20} \cdot 1 = 8,330 + 0,126 = 8,456,$$

$$F_{20+2} = a_{20} + b_{20} \cdot 2 = 8,330 + 0,126 \cdot 2 = 8,582.$$



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.13. Graficzny obraz danych empirycznych i określonych metodą SESD dla $\alpha = 0,1$.

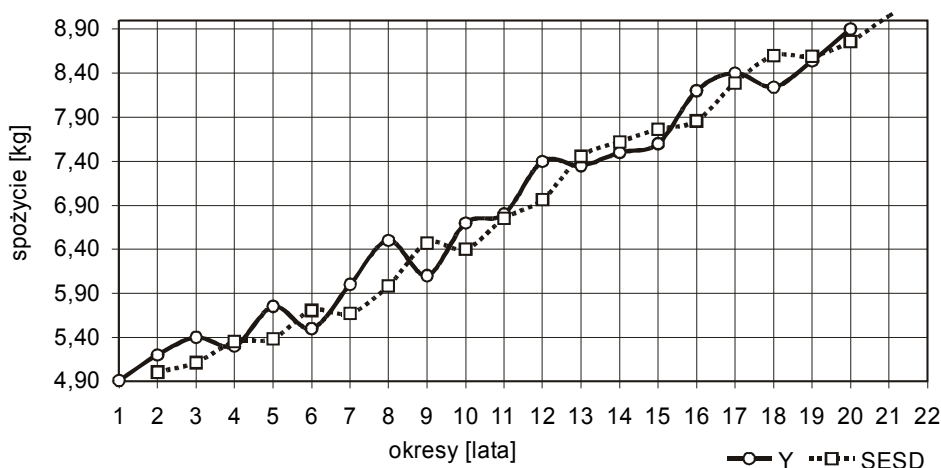
Z wykresu wynika zbyt niskie $\alpha = 0,1$, należy zatem dokonać symulacji zmieniając parametr α . Przykładowo, przyjmijmy kolejno: $\alpha = 0,2; 0,3; 0,4; 0,5$, a z pośród nich wybierzmy najlepsze, biorąc pod uwagę minimum $RMSE$. W tym celu zmienimy dokładność obliczeń do dwóch miejsc po przecinku. W Excelu wystarczy w tab. 1.12.41 zmienić tylko wartość w adresie bezwzględnym $\$C\1 arkusza kalkulacyjnego. Rezultat kilku kroków symulacji podaje tab. 1.12.42 oraz wykres 1.12.14.

Tabela 1.12.42.

Symulacja parametru α (minimum $RMSE$ dla $\alpha = 0,35$)

α	0,1	0,2	0,3	0,4	0,5			
$RMSE$	0,7141	3491	0,278	0,2775	0,2975			
α	0,41	0,39	0,38	0,37	0,36	0,35	0,34	0,33
$RMSE$	0,2789	0,2763	0,2752	0,2744	0,2739	0,2736	0,2737	0,2741

Źródło: Opracowanie własne w Excelu.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.14. Obraz dopasowania danych empirycznych i modelowych (Y - $SESD$) zmiennej Y z zastosowaniem modelu $SESD$ przy $\alpha = 0,35$

1.12.9.7. Pytania

1. Podaj procedurę wyznaczania trendu pelzającego dla dwunastu obserwacji zmiennej objaśnianej w kolejnych latach.
2. Wymień i objaśnij wzory modeli prognoz:
 - modeli prostych średnich ruchomych (SMA_2 , SMA_4) dla okresów $t = 1-10$ i Y_t , przy czym: SMA_2 – średnie dwuokresowe, SMA_4 – średnie cztero-kresowe,

– model ważonych średnich ruchomych czterookresowych (WMA4) o wagach:

$$w_1 = 0,15; w_2 = 0,2; w_3 = 0,25; w_4 = 0,4.$$

– model prostego wyrównywania wykładniczego (SES) dla $\alpha = 0,3$.

3. Wyjaśnij pojęcie sygnał adaptacyjny stosowany w modelach ARRES.

4. Podaj procedurę budowy prognoz według modelu podwójnych średnich ruchomych SMAD.

5. Określ kolejne elementy $S'_n, S''_n, a_n, b_n, F_{n+m} = a_n + b_n \cdot m$, gdzie $m = 1$, w celu wyznaczenia prognozy F_3 według modelu Browna (podwójnego wyrównywania wykładniczego), mając fragment danych i obliczenia początkowe (tab. 1.12.43):

T a b e l a 1.12.43

Zainicjowanie tabeli obliczeń prognoz w modelu Browna

Okres	Y_n	S'_n	S''_n	a_n	b_n	$F_{n+m} = a_n + b_n \cdot m$
1	26	26	26	26	1	27
2	24					
3	32					
3	36					

Ź r ó d ł o: Opracowanie własne; dane umowne.

1.12.9.8. Zadania

1. Oblicz wartości prognoz oraz sporządź wykres trendu pełzającego na okres 13., bazując na szeregu czasowym inflacji:

T a b e l a 1.12.44

Inflacja w latach [%]

t (lata)	1	2	3	4	5	6	7	8	9	10	11	12
Y_t	4	5	7	8	10	12	16	15	17	19	21	20

Ź r ó d ł o: Opracowanie własne; dane umowne.

W ramach tego zadania:

– określ ilość podszeregów według wzoru: $n - k + 1$,

– wyznacz i określ podszeregi o długości segmentu $k = 3$,

– oblicz trendy liniowe dla każdego podszeregu, przy czym parametry a_0 i a_1 określ z równań KMNK,

– na podstawie równań trendów liniowych dla każdego podszeregu (segmentu) oblicz wartości teoretyczne na poszczególne okresy t , odpowiadające danemu podszeregowi,

– oblicz wartości średnich arytmetycznych z określonych wartości teoretycznych dla kolejnych 1–12 okresów (lat),

- narysuj schemat trendu pełzającego obejmującego wartości empiryczne Y oraz teoretyczne \hat{Y} na poszczególne lata,
- oblicz prognozę inflacji na okres 13. na podstawie modelu trendu z okresów 10–12 lub znając trend z trzech średnich arytmetycznych obliczonych dla okresów 10–12.

2. Oblicz prognozy *ex post* na kolejne miesiące oraz prognozy *ex ante* na miesiąc 11. oraz 12., stosując model prostego wyrównywania wykładniczego i wykorzystując dane z tab. 1.12.45.

$$\hat{Y} = \alpha Y_{n-1} + (1 - \alpha) F_{n-1},$$

gdzie:

$\alpha = 0,6$; prognoza początkowa $F_2 = Y_1$.

T a b e l a 1.12.45

Obserwacje zmiennej

t (miesiące)	1	2	3	4	5	6	7	8	9	10
Y_n	110	114	112	113	115	118	119	117	119	118

Ź r ó d ł o: Opracowanie własne; dane umowne.

3. Określ prognozę miesiąca czwartego F_4 z równania modelu adaptacyjnego ARRES, przy zastosowaniu równań tego modelu:

$$F_n = F_{n-1} + TST_{n-1} \cdot (Y_{n-1} - F_{n-1}),$$

$$TST_n = |SAD_n / MAD_n|,$$

$$SAD_n = \beta (Y_n - F_n) + (1 - \beta) SAD_{n-1},$$

$$MAD_n = \beta (Y_n - F_n) + (1 - \beta) MAD_{n-1}.$$

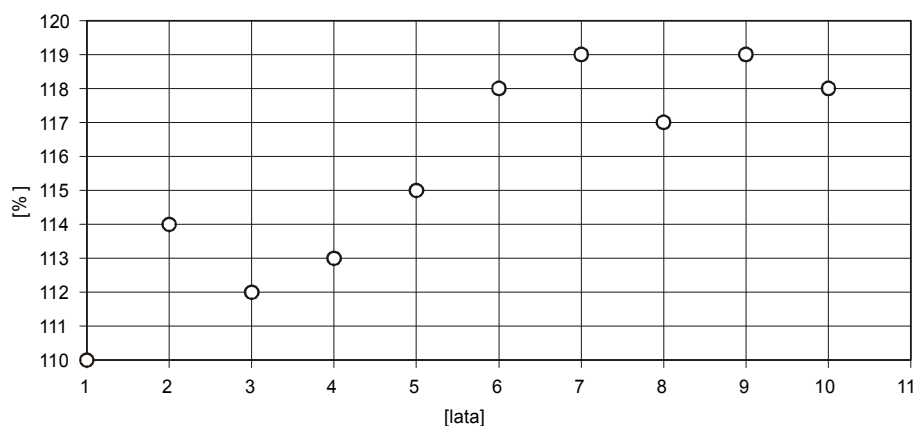
Przyjmij dane początkowe:

$$Y_1 = 110, Y_2 = 114; Y_3 = 112; F_2 = Y_1,$$

$$SAD_1 = 0; MAD_1 = 4; TST_1 = 0,2.$$

4. Stosując test t -Studenta sprawdź, czy w szeregu czasowym z zadania drugiego, który poglądowo przedstawiono w formie wykresu punktowego, występuje trend liniowy dla różnicy $d = 1$. Zastosuj wzór:

$$t_{obl} = \frac{\Delta \bar{Y}}{s_x \cdot \sqrt{n - d}}.$$



Źródło: Opracowanie własne w Excelu; dane umowne.

Wykres 1.12.15. Wykres punktowy zmiennej Y (inflacja) [%] w latach

5. Bazując na szeregu czasowym z zadania drugiego, oblicz prognozę na okres 11., stosując model podwójnej średniej ruchomej.

6. Zastosuj model Browna (podwójnego wyrównywania wykładniczego) do określenia prognozy okresu 12., mając $\alpha = 0,3$ oraz szereg czasowy według tab. 1.12.46.

Tabela 1.12.46

Obserwacje zmiennej „ Y ”

Okres	1	2	3	4	5	6	7	8	9	10
Y_n	26	24	32	36	34	41	43	47	51	54

Źródło: Opracowanie własne; dane umowne.

W ramach zadania:

a) określ dla każdego okresu wymienione elementy, stosując podane wzory:

$S'_n = \alpha Y_n + (1 - \alpha)S'_{n-1}$ (pojedynczo wyrównana wartość szeregu zmiennej),

$S''_n = \alpha S'_n + (1 - \alpha)S''_{n-1}$ (podwójnie wyrównana wartość szeregu zmiennej),

$a_n = S'_n + (S'_n - S''_n) = 2 S'_n - S''_n$ (wyrównana wykładniczo wartość zmiennej dla prognozy),

$b_n = [\alpha/(1 - \alpha)] (S'_n - S''_n)$ (wyrównana wykładniczo wartość trendu dla prognozy).

Przyjmij wartości początkowe:

$$S'_1 = S''_1 = Y_1, a_1 = Y_1, b_1 = [(Y_2 - Y_1) + (Y_4 - Y_3)]/2,$$

$$S'_1 = S''_1 = Y_1 = 26, a_1 = Y_1 = 26, b_1 = [(24 - 26) + (36 - 32)]/2 = 1;$$

b) określ miarę *RMSE* błędu prognozy:

$$RMSE = \sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} (Y_n - F_n)^2}.$$

c) mając dane a_{10} , b_{10} , m , określ prognozę dla okresu 12;

d) przedstaw na wykresie dane rzeczywiste i teoretyczne (modelowe) otrzymane z zastosowania modelu Browna;

e) oceń trafność przyjętego parametru wyrównującego $\alpha = 0,3$ jednolitego do wygładzania wartości szeregu oraz wyrównywania trendu;

f) jeśli masz takie warunki, wprowadź model Browna do arkusza kalkulacyjnego Excel i dokonaj symulacji doboru parametru α spośród przedziału wartości 0,1–0,9;

g) dla najlepszego α z punktu widzenia minimum *RMSE* określ kolejne prognozy F_n dla okresów obserwacji oraz okresu prognozowanego 12.

1.12.10. Klasyczny model autoregresyjny (zastosowanie programu WinQSB)

Model autoregresyjny służy do budowy prognoz krótkoterminowych. Ma zastosowanie w sytuacji, kiedy trudno jest zebrać dane dotyczące zmiennych objaśniających, którymi w tym modelu są opóźnione zmienne objaśniane.

Postać modelu:

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-n}, \varepsilon_t).$$

W tym modelu zmienną objaśnianą – endogeniczną wyraża funkcja poprzednich wartości tej zmiennej oraz składnika losowego.

P r z y k ł a d

Zbudujmy model autoregresyjny, bazując na eksporcie w handlu zagranicznym w przeliczeniu na mieszkańca według cen bieżących w USD. Mamy dane od roku 1993 do 2005.

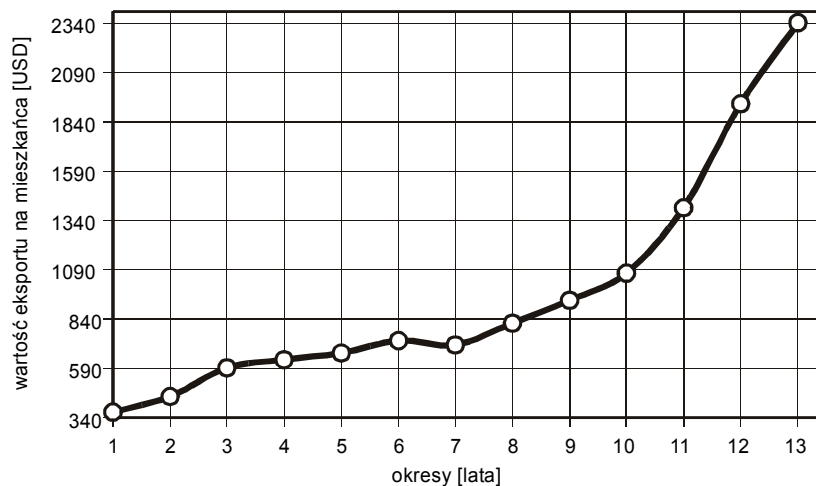
Tabela 1.12.47

Eksport na mieszkańca [USD]

Rok	Okres [t]	Eksport [Y]
1993	1	368
1994	2	447
1995	3	593
1996	4	633
1997	5	666
1998	6	730
1999	7	709
2000	8	819
2001	9	934
2002	10	1 073
2003	11	1 403
2004	12	1 932
2005	13	2 342

Ź r ó d ł o: Opracowanie własne na podstawie: „Mały Rocznik Statystyczny Polski”, GUS, Warszawa, za lata 1993–2005.

Przedstawmy szereg czasowy na wykresie punktowym.



Ź r ó d ł o: Opracowanie własne w Excelu.

Wykres 1.12.16. Wartość eksportu na mieszkańca w latach 1993–2005

Tym razem zastosujemy program WinQSB, w którym zamiast przecinka wpisywana jest kropka dziesiętna. Utwórzmy szeregi czasowe opóźnione o 1, 2, 3 okresy wartości zmiennej objaśnianej Y .

T a b e l a 1.12.48

Opóźnione zmienne objaśniane $Y(-1)$, $Y(-2)$, $Y(-3)$

t	Y	$Y(-1)$	$Y(-2)$	$Y(-3)$	
1	368				
2	447	368			
3	593	447	368		
4	633	593	447	368	1
5	666	633	593	447	2
6	730	666	633	593	3
7	709	730	666	633	4
8	819	709	730	666	5
9	934	819	709	730	6
10	1 073	934	819	709	7
11	1 403	1 073	934	819	8
12	1 932	1 403	1 073	934	9
13	2 342	1 932	1 403	1 073	10

Ź r ó ł o: Opracowanie własne w Excelu.

Wartości zmiennych od 4. do 13. obserwacji, dotyczące zmiennych Y , $Y(-1)$, $Y(-2)$, $Y(-3)$ wprowadzono do programu WinQSB jako 1–10. Procedura określenia modelu autoregresyjnego z zastosowaniem tego programu jest następująca:

– wybór modułu (*Forecasting – FC*) prognozowanie; korzystanie z opcji (*File*) zbiór; (*New Problem*) nowy problem; (*Problem Type*) typ problemu, a w ramach niego wybór:

(*Time Series Forecasting*) szeregu czasowego,

(*Regression Forecasting*) prognozowanie według regresji zmiennych.

W naszym przykładzie wybieramy regresję oraz wypełniamy informacje: (*Problem Title*) tytuł problemu, np. *Eksport*; *Number of Variables* (liczba zmiennych) – 4, tj. Y , $Y(-1)$, $Y(-2)$, $Y(-3)$; *Number of Observations* – 10; *OK*,

– po pojawieniu się tabeli zmieniamy nazwy standardowe na własne: Y , $Y(-1)$, $Y(-2)$, $Y(-3)$; wprowadzamy obserwacje według tab. 1.12.48 od 4 do 13. Jeśli mamy dane w innym pliku, to możemy je skopiować i wkleić (*Paste*); zachowujemy plik danych (*Save Problem AS*), np. (*E.ECC*) w podkatalogu (*D:\SKRYPTI*) oraz jako (*E.DAT*) do wklejenia w edytorze Word. Rezultat naszego dotychczasowego działania przedstawia tab. 1.12.49:

Tabela 1.12.49

Baza danych do obliczeń w WinQSB

FC Obserwacje	Eksport Y	Y(-1)	Y(-2)	Y(-3)
1	633	593	447	368
2	666	633	593	447
3	730	666	633	593
4	709	730	666	633
5	819	709	730	666
6	934	819	709	730
7	1 073	934	819	709
8	1 403	1 073	934	819
9	1 932	1 403	1 073	934
10	2 342	1 932	1 403	1 073

Źródło: Opracowanie własne w Excelu na podstawie danych z ekranu programu WinQSB.

– (Solve and Analyze) rozwiąż i przeanalizuj; (Perform Linear Regression) wykonaj regresję liniową; (Select a Dependent Variable) wybierz zmienną objaśnianą, np. Y; (Select as Many Independent Variables) wybierz zmienne objaśniające, np. Y(-1), Y(-2), Y(-3); OK.

– program daje rozwiązanie; możemy dokonać edycji czcionek oraz dokładności obliczeń, np. ARIAL 12, 0.00. Fragment tabeli modułu FC po dodatkowej edycji obramowań w Excelu podano w tabeli 1.12.50.

Tabela 1.12.50

Zestawienie sumaryczne dla „Eksport”

08-15-2007 12:44:04	Variable Name	Mean	Standard Deviation	Regression Coefficient	Standard Error	t value
Dependent	Y	1124,1	589,186			
Y-intercept	Constant			-242,8467	156,7105	-1,5497
1	Y(-1)	949,2	423,8904	1,3595	0,4707	2,8879
2	Y(-2)	800,7	275,466	-0,3228	1,0311	-0,3131
3	Y(-3)	697,2	210,8911	0,4805	0,6669	0,7206
			R-square =	0,9774	R-adjusted =	0,9661

Variable Name – nazwa zmiennej,

Dependent – zależna,

Constant – stała (a_0),

Mean – średnia zmiennej Y – (\bar{Y}),

Standard Deviation – odchylenie standardowe zmiennych,

Regression Coefficient – parametry linii regresji (a_i),

R-square – R^2 , czyli współczynnik determinacji; zwróćmy uwagę na jego wysoką wartość,

R-adjusted – rozszerzony współczynnik determinacji (\bar{R}^2),

Standard Error – błąd standardowy określenia parametru,

Value – wartość obliczeniowa (t_{obl}) testu t-Studenta dla sprawdzenia istotności statystycznej przy małej próbie danego parametru a_i przy $m = 3$ zmienne objaśniające.

Źródło: Opracowanie własne, moduł FC program WinQSB.

Sprawdźmy warunek:

$|t_{obl}| > t^*$ dla $iss = n - m - 1 = 10 - 3 - 1 = 6$ oraz na poziomie istotności $\alpha = 0,05$, wartość krytyczna t^* według rozkładu t -Studenta wynosi 2,4469. Zatem spoglądając na kolumnę t value w tab. 1.12.50 możemy w modelu uwzględnić tylko parametr a_1 ze zmienną objaśnianą $Y(-1)$.

Ponadto moduł FC w opcji (*Results*) wyniki umożliwia wyświetlenie i zapisanie w pliku tabel:

a) Współczynniki korelacji (*Correlations Analysis for Eksport*).

T a b e l a 1.12.51

Współczynniki korelacji między zmiennymi

08-15-2007	Variable	Variable	Correlation
1	Y	Y(-1)	0,9873
2	Y	Y(-2)	0,9719
3	Y	Y(-3)	0,9207
4	Y(-1)	Y(-2)	0,9795
5	Y(-1)	Y(-3)	0,9134
6	Y(-2)	Y(-3)	0,9582

Ź r ó d ł o: Opracowanie własne w WinQSB.

Największy współczynnik korelacji jest między zmienną Y a tą samą zmienną, lecz opóźnioną o jeden okres. Jest to zbieżne z wcześniejszym określeniem modelu według testu istotności statystycznej parametrów a_i , a poprzez niego oszacowanie wpływu zmiennych opóźnionych na zmienną badaną Y .

Na podstawie zależności korelacyjnej można wstępnie sformułować model jako:

$$\hat{Y} = a_0 + a_1 Y(-1).$$

b) Analiza reszt (*Residual Analysis for Eksport*). Na początek przeprowadźmy ją dla modelu uwzględniającego trzy zmienne objaśniane $Y(-1)$, $Y(-2)$, $Y(-3)$.

T a b e l a 1.12.52

Standaryzacja reszt dla „Eksport”

08-15-2007 13:26:26	Actual Y	Prediction Y^	Std. Dev. of Prediction	Residual ui	[%] Residual	Standardized Residual
1	633	595,86	95,15	37,14	6,23	0,42
2	666	641,06	79,16	24,94	3,89	0,28
3	730	743,17	43,24	-13,17	-1,77	-0,15
4	709	838,75	43,92	-129,75	-15,47	-1,46
5	819	805,40	68,23	13,60	1,69	0,15
6	934	992,47	77,53	-58,47	-5,89	-0,66
7	1 073	1 103,21	40,13	-30,21	-2,74	-0,34
8	1 403	1 307,91	50,51	95,09	7,27	1,07
9	1 932	1 766,93	62,65	165,07	9,34	1,86

gdzie: $Y^{\wedge} \rightarrow \hat{Y}$.

Actual Y – wartości rzeczywiste zmiennej objaśnianej Y ,

Prediction \hat{Y} – wartości modelowe obliczone z modelu autoregresyjnego trzeciego rzędu dla zmiennej Y , otrzymanego opcją (*Regression Equation for Eksport*) równania regresji.

T a b e l a 1.12.53

Równanie regresji dla zmiennej objaśnianej Y

08-15-2007 13:38:13	Dependent Variable	Independent Variable
Equation:	$Y^{\wedge} =$	$- 242,85 + 1,36 Y(-1) - 0,32 Y(-2) + 0,48 Y(-3)$

Dependent Variable – zmienna objaśniana,

Independent Variable – zmienna objaśniająca $Y(1)$, $Y(-2)$, $Y(-3)$.

W modelu wartości parametrów zaokrąglono do drugiego miejsca po przecinku.

Standard Deviation of Prediction – standardowe odchylenie od wartości prognozowanej,

Residual – reszta ($u_i = Y - \hat{Y}$),

% Residual – procent reszt określony według:

$$\%u_i = \frac{u_i}{\hat{Y}_i} \cdot 100.$$

Standardized Residual – standaryzacja reszt, która np. dla obserwacji pierwszej wynosi:

$$S_i = \frac{u_i - \bar{u}}{S_u}, \quad S_i = \frac{37,14 - (-0,011)}{88,63} = 0,42.$$

Ź r ó d ł o: Opracowanie własne w WinQSB, gdzie: $Y^{\wedge} \rightarrow \hat{Y}$.

Standaryzację określamy stosunkiem różnicy (reszty z obserwacji i -tej minus średnia arytmetyczna reszt) do odchylenia standardowego reszt.

Do obliczenia średniej oraz odchylenia standardowego użyto wzory:

$$\bar{u} = \frac{\sum_{i=1}^n u_i}{n}, \quad S_i = \sqrt{S_i^2} = \sqrt{\frac{\sum_{i=1}^n (u_i - \bar{u})^2}{n}}.$$

P r z y k ł a d

Zbudujmy jeszcze jeden model autoregresyjny bazujący na spożyciu masła w gramach w przeliczeniu na jednego mieszkańca (zob. tabela 1.12.40). Przeprowadźmy symulacje doboru zmiennych opóźnionych w czasie, biorąc pod uwagę test istotności statystycznej parametrów a_i w celu określenia prognozy na okres 21. (tab. 1.12.41), czyli 18. według tab. 1.12.54 ze względu na przesunięcie czasowe ($t-3$). Posłużymy się w tym celu modulem *FC*.

Tabela 1.12.54

Szeregi czasowe zmiennych „Y”, „Y(-1)”, „Y(-2)”, „Y(-3)”

Spożycie masła				
FC	Y	Y(-1)	Y(-2)	Y(-3)
Obserwacja				
1	5 300	5 400	5 200	4 910
2	5 750	5 300	5 400	5 200
3	5 500	5 750	5 300	5 400
4	6 000	5 500	5 750	5 300
5	6 500	6 000	5 500	5 750
6	6 100	6 500	6 000	5 500
7	6 800	6 700	6 100	6 500
8	7 400	6 800	6 700	6 100
9	6 700	6 100	6 500	6 000
10	7 350	7 400	6 800	6 700
11	7 500	7 350	7 400	6 800
12	7 600	7 500	7 350	7 400
13	8 200	7 600	7 500	7 350
14	8 400	8 200	7 600	7 500
15	8 240	8 400	8 200	7 600
16	8 540	8 240	8 400	8 200
17	8 900	8 540	8 240	8 400

Źródło: Opracowanie własne, fragment tab. 1.12.40 z przykładową zmianą wartości obserwacji od 7. do 9.

Tabela 1.12.55

Zestawienie sumaryczne dla „Spożycie masła”

08-12-2007 16:21:16	Variable Name	Mean	Standard Deviation	Regression Coefficient	Standard Error	t value
Dependent	Y	7 104,71	1 132,41			
Y – intercept	Constant			288,66	427,92	0,67
1	Y(-1)	6 898,82	1 103,41	0,22	0,26	0,85
2	Y(-2)	6 702,35	1 090,20	0,37	0,24	1,52
3	Y(-3)	6 506,47	1 095,80	0,44	0,26	1,7
			R – square =	0,95	R – adjusted =	0,94

Źródło: Opracowanie własne w WinQSB.

Według tablicy I z *Aneksu* wartości krytycznych rozkładu *t*-Studenta przy poziomie istotności $\alpha = 0,05$ i przy ilości stopni swobody $iss = n - k = 17 - 4 = 13$, $t_\alpha = t^* = 2,1604$ ($k - 4$ parametry, $n = 17$ – ilość obserwacji danych dla modelu autoregresyjnego).

Z tabeli regresji (1.12.55) wynika, że żaden parametr dla modelu autoregresyjnego trzeciego rzędu nie spełnia warunku $|t_{obl}|$ (*t value*) $> t^*$. W tej sytuacji przeanalizujemy tabelę współczynników korelacji między zmiennymi, korzystając z opcji (*Correlations Analysis...*).

T a b e l a 1.12.56
Współczynniki korelacji między zmiennymi
„*Y*”, „*Y(-1)*”, „*Y(-2)*”, „*Y(-3)*”

08-13-2007	Variable	Variable	Correlation
1	Y	Y(-1)	0,9597
2	Y	Y(-2)	0,9634
3	Y	Y(-3)	0,9672
4	Y(-1)	Y(-2)	0,9585
5	Y(-1)	Y(-3)	0,9635

Ź r ó d ł o: Opracowanie własne w WinQSB.

Najsłabsza zależność występuje między spożyciem masła w okresie bieżącym a spożyciem w okresie poprzednim i wynosi 0,9597. Przeprowadźmy zatem obliczenia dla modelu:

$$\hat{Y} = a_0 + a_1Y(-2) + a_2Y(-3).$$

T a b e l a 1.12.57

Zestawienie sumaryczne dla modelu bez zmiennej „Y(-1)” „Spożycie masła”

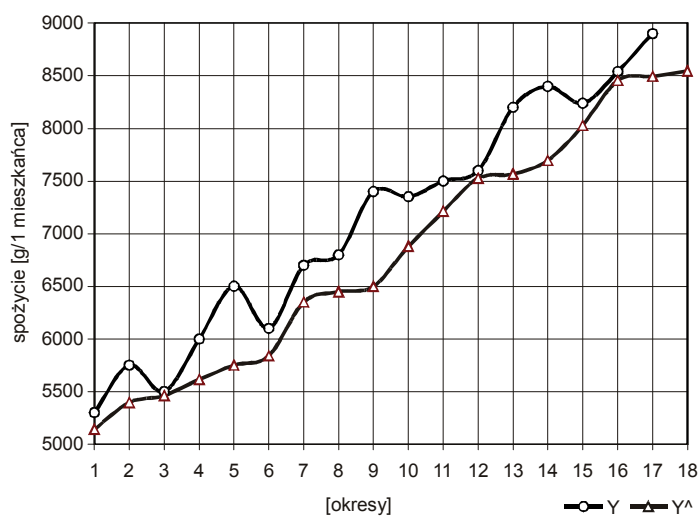
08-13-2007 10:09:26	Variable Name	Mean	Standard Deviation	Regression Coefficient	Standard Error	t value
Dependent	Y	7 104,71	1 132,41			
Y-intercept	Constant			369,19	413,41	0,89
1	Y(-2)	6 702,35	1 090,20	0,46	0,21	2,18
2	Y(-3)	6 506,47	1 095,80	0,56	0,21	2,63
			R - square =	0,95	R - adjusted =	0,94

Ź r ó d ł o: Opracowanie własne w WinQSB.

Według tablicy I *Aneksu* przy zmienionej ilości parametrów ($iss = 17 - 3 = 14$, $\alpha = 0,05$), $t^* = 2,1448$. Z powodu braku statystycznej istotności $|t \text{ value}| > t^*$ w modelu autoregresyjnym można pominąć stałą $a_0 = 369,19$. Tak więc bez tej stałej model jest następujący:

$$\hat{Y} = 0,46 \cdot Y(-2) + 0,56 \cdot Y(-3).$$

Określmy zatem wygasłe prognozy oraz prognozę na okres 21. (czyli 18. w tab. 1.12.58). Obliczmy również pierwiastek średniego błędu wygasłych prognoz *RMSE*. Wyniki tych prac przedstawiono na wykresie 1.12.17 oraz zapisano w tab. 1.12.58.



Źródło: Opracowanie własne w Excelu.

Wykres 1.12.17. Dane empiryczne i teoretyczne – modelowe spożycia masła

Tabela 1.12.58

Tabela pomocnicza do określenia prognozy

Okres	Y	Y(-2)	Y(-3)	Y [^]	u _i = Y - Y [^]	(Y - Y [^]) ²
	4 910					
	5 200					
	5 400	4 910				
1	5 300	5 200	4 910	5 141,60	158,4	25 090,56
2	5 750	5 400	5 200	5 396	354,0	125 316
3	5 500	5 300	5 400	5 462	38,0	1 444
4	6 000	5 750	5 300	5 613	387,0	149 769
5	6 500	5 500	5 750	5 750	750,0	562 500
6	6 100	6 000	5 500	5 840	260,0	67 600
7	6 700	6 500	6 000	6 350	350,0	122 500
8	6 800	6 100	6 500	6 446	354,0	125 316
9	7 400	6 700	6 100	6 498	902,0	813 604
10	7 350	6 800	6 700	6 880	470,0	220 900
11	7 500	7 400	6 800	7 212	288,0	82 944
12	7 600	7 350	7 400	7 525	75,0	5 625
13	8 200	7 500	7 350	7 566	634,0	401 956
14	8 400	7 600	7 500	7 696	704,0	495 616
15	8 240	8 200	7 600	8 028	212,0	44 944
16	8 540	8 400	8 200	8 456	84,0	7 056
17	8 900	8 240	8 400	8 494,40	405,6	164 511,36
18	Prognoza			8 542,80	Σ	3 416 691,92
					Σ/17	200 981,88
					RMSE	448,31

Źródło: Opracowanie własne w Excelu, $i = 1 - 20$; $Y^{\wedge} \rightarrow \hat{Y}$.

Przykład

W omawianym modelu $\hat{Y} = 0,46 \cdot Y(-2) + 0,56 \cdot Y(-3)$ sprawdzimy normalność składnika resztowego, posługując się testem Jarque'a-Bery (*JBT* czasem określanym również jako *JB*), który służy do weryfikacji hipotezy zerowej (rozkład składnika resztowego jest rozkładem normalnym). Obliczymy *JBT* według wzoru (1.1.15), a następnie sprawdzamy, czy $JBT < \chi^2$ przy dwóch stopniach swobody i przy określonym poziomie istotności α :

$$JBT = \frac{n-k}{6} \cdot \left[S^2 + \frac{1}{4}(K-3)^2 \right].$$

W celu szybkiego określenia skośności S oraz kurtozy K skorzystajmy z opcji Excela (*Analiza danych/Statystyka opisowa*). Uwzględniamy 17 reszt od u_4 do u_{20} na bazie obliczenia w kolumnie $u_i = Y - \hat{Y}$ (tab. 1.12.58).

Tabela 1.12.59

Statystyka opisowa dotycząca reszt

Statystyka opisowa reszt ($Y - \hat{Y}$)	
Średnia	378
Błąd standardowy	60,26
Mediana	354
Odchylenie standardowe	248,45
Wariancja próbki	61 729
Kurtoza	-0,24
Skośność	0,63
Zakres	864
Minimum	38
Maksimum	902
Suma	6 426
Licznik	17
Największy(1)	902
Najmniejszy(1)	38

Źródło: Opracowanie własne w Excelu.

Obliczamy *JBT* dla $n = 17$, $k = 2$ (dwa parametry modelu: $a_1 = 0,46$, $a_2 = 0,56$), S – skośność = 0,63, K – kurtoza = - 0,24.

$$JBT = \frac{17-2}{6} \cdot \left[(0,63)^2 + \frac{1}{4}(-0,24-3)^2 \right] = 7,55.$$

Jak już wspomniano, test oparty jest na rozkładzie χ^2 z dwoma stopniami swobody. Hipoteza zerowa H_0 zakłada, że rozkład składnika jest rozkładem normalnym. Hipoteza alternatywna H_1 zakłada, że rozkład ten nie jest rozkładem normalnym. Hipoteza zerowa jest spełniona, gdy $JBT < \chi^2$.

Przy poziomie istotności $\alpha = 0,05$ i $r = 2$ $\chi^2 = 5,991$ (tab. II w *Aneksie*). Ponieważ $JBT > \chi^2 = 5,991$, odrzucamy hipotezę H_0 o normalności składnika resztowego na rzecz hipotezy alternatywnej H_1 , mówiącej o niezgodności rozkładu składnika losowego z rozkładem normalnym.

Przykład

W hurtowni regionalnej kwiatów sprzedaż wiązanek gerberów po pięć sztuk w poszczególnych miesiącach czterech lat, począwszy od stycznia do grudnia w przeliczeniu na jeden punkt sprzedaży, kształtowała się jak w tab. 1.12.60. Obliczamy prognozę na miesiąc styczeń następnego roku przy zastosowaniu modelu autoregresyjnego.

W celu sformułowania modelu autoregresyjnego rozpoczniemy jak w poprzednich przykładach od modelu trzeciego rzędu zmiennych opóźnionych $Y(-1)$, $Y(-2)$, $Y(-3)$. Obserwacje (4–16) jako (1–13) wprowadzamy do modułu *FC* programu WinQSB.

Model wejściowy ma postać: $\hat{Y} = a_0 + a_1 \cdot Y(-1) + a_2 \cdot Y(-2) + a_3 \cdot Y(-3)$

Tabela 1.12.60

Sprzedaż wiązanek gerberów

Okres	Y	Y(-1)	Y(-2)	Y(-3)
1	12			
2	9	12		
3	8	9	12	
4	11	8	9	12
5	9	11	8	9
6	8	9	11	8
7	6	8	9	11
8	8	6	8	9
9	7	8	6	8
10	7	7	8	6
11	6	7	7	8
12	8	6	7	7
13	6	8	6	7
14	5	6	8	6
15	3	5	6	8
16	5	3	5	6

Źródło: Opracowanie własne w Excelu.

Obliczamy współczynniki korelacji.

Tabela 1.12.61

Współczynniki korelacji między zmiennymi „Y”, „Y(-1)”, „Y(-2)”, „Y(-3)”

08-13-2007	Variable	Variable	Correlation
1	Y	Y(-1)	0,5830
2	Y	Y(-2)	0,5350
3	Y	Y(-3)	0,5356
4	Y(-1)	Y(-2)	0,5606
5	Y(-1)	Y(-3)	0,4547
6	Y(-2)	Y(-3)	0,4605

Źródło: Opracowanie własne w WinQSB.

Biorąc pod uwagę tylko współczynniki korelacji między zmienną objaśnianą a zmiennymi objaśniającymi, pomijamy $Y(-2)$ i model z tego punktu widzenia może być następujący:

$$\hat{Y} = a_0 + a_1 \cdot Y(-1) + a_2 \cdot Y(-3).$$

Sprawdzamy założony model analizą korelacyjną dla $n' = 13$ obserwacji (4–16).

Dla $n = 13$, $iss = n - k = 13 - 3 = 10$, $\alpha = 0,05$, wartość krytyczna $t_\alpha = 2,2281$ rozkładu t -Studenta (tabl. I w *Aneksie*).

Wartość krytyczna współczynnika korelacji (wzór 1.6.9):

$$r^* = \sqrt{t_\alpha^2 / (t_\alpha^2 + n - 2)} = \sqrt{2,2281^2 / (2,2281^2 + 13 - 2)} = 0,5576.$$

Sprawdzamy, czy $|r_j| \leq r^*$. Gdy porównamy z tab. 1.12.61, takimi zmiennymi są $Y(-2)$ oraz $Y(-3)$. Pozostało tylko $Y(-1)$, dla którego $|r_{hi}| = \max|r_j| = 0,583$. Następnie sprawdzamy warunek: $|r_{hi}| > r^*$. Według tego kryterium eliminujemy $Y(-2)$. Pozostaje jednak zmienna $Y(-3)$, ale ta zmienna już wcześniej wypadła.

Sformułowanie modelu:

$$\hat{Y} = a_0 + a_1 \cdot Y(-1).$$

Dokonajmy estymacji parametrów a_i regresji liniowej funkcją REGLINP dla obserwacji 4–16 zmiennej objaśniającej $Y(-1)$.

T a b e l a 1.12.62

Raport z funkcji REGLINP

a_1	a_0	
0,6	2,6	parametry
0,252	1,847	standardowy błąd szacunku a_i
0,340	1,727	współczynnik determinacji; standardowy błąd oceny Y
5,665	11	statystyka F ; ilość stopni swobody
16,892	32,800	regresyjna suma kwadratów; resztowa suma kwadratów

Ź r ó d ł o: Opracowanie własne w Excelu.

Postać analityczna modelu: $\hat{Y} = 2,6 + 0,6 \cdot Y(-1)$.

Badanie istotności parametrów strukturalnych a_0, a_1 powyższego liniowego modelu ekonometrycznego polega na sprawdzeniu, czy zmienna objaśniająca $Y(-1)$ istotnie wpływa na zmienną objaśnianą Y . Weryfikujemy hipotezę zerową:

$$H_0 : [a_i = 0], \quad i = 1, 2$$

wobec hipotezy alternatywnej:

$$H_1 : [a_i \neq 0],$$

a sprawdzaniem tej hipotezy jest statystyka:

$$t_0 = \frac{|2,6|}{1,847} = 1,408, \quad t_1 = \frac{|0,6|}{0,252} = 2,380.$$

Określamy wartość krytyczną $t_\alpha = t^* = 2,201$ dla przyjętego poziomu istotności $\alpha = 0,05$ i ilości stopni swobody $r = 13 - 2 = 11$ (tabl. I w *Aneksie*).

Przyjmujemy, że w modelu pozostaje parametr a_0 . Ponieważ $|t_1| > t^*$, to hipotezę zerową H_0 odrzucamy na rzecz hipotezy alternatywnej H_1 . Parametr a_1 różni się w sposób istotny od zera i zmienna objaśniająca $Y(-1)$ wpływa w sposób istotny statystycznie na zmienną objaśnianą Y .

1.12.10.1. Pytania

1. Podaj i objaśnij następującą postać modelu autoregresyjnego:

$$\hat{Y} = 0,694 + 0,44 \cdot Y(-1) + 0,376 \cdot Y(-3).$$

2. Wymień kolejność postępowania w określeniu parametrów linii regresji powyższego modelu dwoma sposobami:

- funkcjami Excela,
- modułem prognozowanie (*Forecasting*) programu WinQSB.

3. W jaki sposób i czym musimy dysponować, aby sprawdzić statystyczną istotność parametrów modelu w pytaniu pierwszym?

1.12.10.2. Zadania

1. Określ model autoregresyjny trzeciego rzędu na podstawie danych o imporcie w handlu zagranicznym według cen bieżących w USD w przeliczeniu na mieszkańca Polski w latach 1985–2005 (tab. 1.12.63). Zastosuj arkusz kalkulacyjny Excel.

T a b e l a 1.12.63

Import w latach 1985–2005

t	1	2	3	4	5	6	7	8	9	10	
X	291	299	288	323	271	250	406	415	490	559	
t	11	12	13	14	15	16	17	18	19	20	21
X	753	962	1 094	1 217	1 188	1 266	1 301	1 442	1 781	2 303	2 664

Ź r ó d ł o: Opracowanie własne; fragment tabeli 1.8.6.

2. Sprawdź istotność statystyczną modelu, który otrzymasz w zadaniu pierwszym, zweryfikuj zmienne, a następnie określ błąd *RMSE* wygaśłych już prognoz dla przedziału okresów 1–21.

3. Zastosuj test *JBT* do sprawdzenia normalności szeregu czasowego składnika resztowego ($u = Y - \hat{Y}$) z wykorzystaniem opcji (*Statystyka opisowa*) programu Excel.

Tabela 1.12.64

Składnik resztowy

t	1	2	3	4	5	6	7	8	9	10	11
u_t	0,71	-0,11	0,38	-0,29	1,21	0,76	-0,41	0,95	1,23	1,73	0,01
t	12	13	14	15	16	17	18	19	20	21	22
u_t	0,06	2,18	0,91	0,42	0,73	2,58	1,02	1,43	-0,01	2,61	1,02

Źródło: Opracowanie własne; dane umowne.

1.13. Wyrównywanie trendu z zastosowaniem modelu Holta-Wintersa

Do modelu Holta-Wintersa (bez sezonowości) wprowadza się drugi parametr β w celu wygładzania trendu¹³⁹. Wygładzona jest wartość trendu z poprzedniego okresu, by usunąć błędy przypadkowe. Następujące równania służą do zapisu tego modelu:

$$S_n = \alpha Y_n + (1 - \alpha) \cdot (S_{n-1} + b_{n-1}),$$

$$b_n = \beta \cdot (S_n - S_{n-1}) + (1 - \beta) \cdot b_{n-1},$$

$$F_{n+m} = S_n + b_n \cdot m.$$

Wiele z zastosowanych symboli znamy z wcześniejszych modeli, wyjaśnienia wymaga jednak b_n , które oznacza wyrównywany trend w okresie n .

W obliczeniach z użyciem modelu Holta-Wintersa przyjmujemy następujące wartości początkowe:

$$S_1 = Y_1 \quad b_1 = [(Y_2 - Y_1) + (Y_4 - Y_3)]/2.$$

Staramy się dobrać parametry α , β tak, aby doprowadzić do minimalnej wartości pierwiastka kwadratu średniego błędu *RMSE* prognozy *ex post*. Model Holta-Wintersa stosuje się wtedy, gdy występuje trend i wahania okresowe.

P r z y k ł a d

Skorzystajmy z danych szeregu czasowego o zmiennym trendzie (tab. 1.12.41) użytych do modelu Browna. Przyjmijmy początkowo, że $\alpha = 0,4$, a $\beta = 0,2$. W obliczaniu wygasłych prognoz i prognozy na okres 21. przyjmujemy $m = 1$, natomiast do prognozy na okres 22. $m = 2$.

¹³⁹ Ibidem, rozdz. 2.14.

Tabela 1.13.1

Obliczenia pomocnicze oraz określenie prognoz na okres 21. i 22.

$\alpha =$		$\beta =$				
Okres	Y_n	S_n	b_n	$F_{n+m} = a_n + b_n \cdot m, m = 1$	$Y - F_n$ (błąd)	$(Y - F_n)^2$
1	4,91	4,910	0,095			
2	5,20	5,083	0,111	5,005	0,195	0,038
3	5,40	5,276	0,127	5,194	0,206	0,043
4	5,30	5,362	0,119	5,403	-0,103	0,011
5	5,75	5,588	0,140	5,481	0,269	0,072
6	5,50	5,637	0,122	5,729	-0,229	0,052
7	6,00	5,856	0,141	5,759	0,241	0,058
8	6,50	6,198	0,182	5,997	0,503	0,253
9	6,10	6,268	0,159	6,380	-0,280	0,078
10	6,70	6,536	0,181	6,427	0,273	0,075
11	6,80	6,750	0,188	6,717	0,083	0,007
12	7,40	7,123	0,225	6,938	0,462	0,213
13	7,35	7,348	0,225	7,347	0,003	0,000
14	7,50	7,544	0,219	7,573	-0,073	0,005
15	7,60	7,698	0,206	7,763	-0,163	0,027
16	8,20	8,022	0,230	7,904	0,296	0,088
17	8,40	8,311	0,241	8,252	0,148	0,022
18	8,24	8,428	0,216	8,553	-0,313	0,098
19	8,54	8,602	0,208	8,644	-0,104	0,011
20	8,90	8,846	0,215	8,811	0,089	0,008
21	Prognoza			9,062	Σ	1,158
22	Prognoza			9,277	$\Sigma/19$	0,061
					<i>RMSE</i>	0,247

Źródło: Opracowanie własne w Excelu.

Przykładowo na okres 2.:

$$S_2 = \alpha \cdot Y_2 + (1 - \alpha) \cdot (S_1 + b_1) = 0,4 \cdot 5,20 + (1 - 0,4) \cdot (4,91 + 0,095) = 5,083,$$

$$b_2 = \beta \cdot (S_2 - S_1) + (1 - \beta) \cdot b_1 = 0,2 \cdot (5,083 - 4,910) + (1 - 0,2) \cdot 0,095 = 0,111,$$

$$F_2 = S_1 + b_1 \cdot 1 = 4,910 + 0,095 \cdot 1 = 5,005.$$

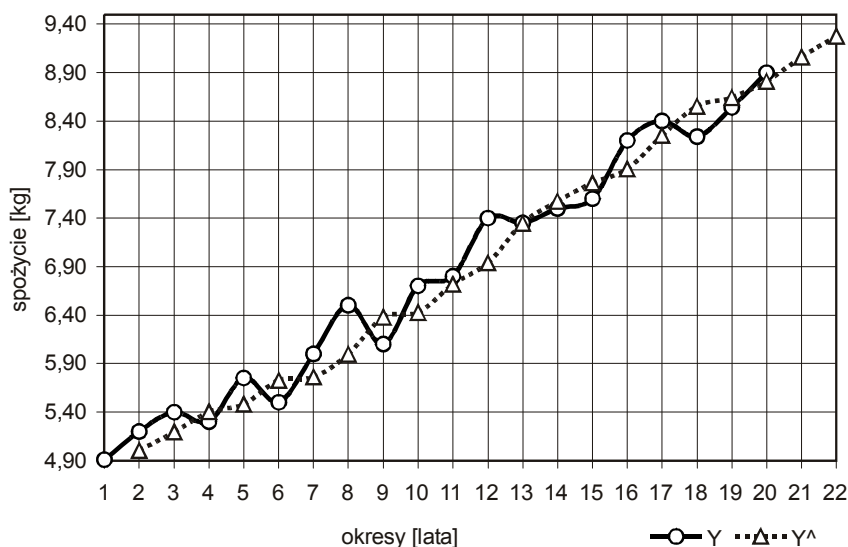
Mając dane z okresu 20., obliczamy prognozy np. na okres 21. oraz 22. (różnica wyników w tab. 1.13.1 oraz poniższych spowodowana jest większą precyzją obliczeń w Excelu):

$$F_{20+1} = S_{20} + b_{20} \cdot 1 = 8,846 + 0,215 = 9,061,$$

$$F_{20+2} = S_{20} + b_{20} \cdot 2 = 8,846 + 0,215 \cdot 2 = 9,276.$$

Błąd wygaśniętych prognoz *RMSE* (według tab. 1.13.1) równa się 0,247. W celu zminimalizowania tego błędu proponuje się Czytelnikowi przeprowadzenie symulacji na kombinacji dwóch parametrów: α i β . Efekt końcowy tej wielokrotnej symulacji przedstawia tab. 1.13.2.

A teraz dla zobrazowania dopasowania modelu omawianego przykładu z danymi empirycznymi przedstawimy oba szeregi danych Y oraz \hat{Y} na wykresie.



Źródło: Opracowanie własne w Excelu.

Wykres 1.13.1. Dane empiryczne i teoretyczne zmiennej „Spożycie”

Przykład

Wykonajmy symulację parametrów α i β , bazując na zapisanej w arkuszu kalkulacyjnym tab. 1.13.1 oraz wartościach wyjściowych $\alpha = 0,4$ i $\beta = 0,2$. Zmieniając najpierw parametr α powyżej a później poniżej 0,4, przy stałej wartości $\beta = 0,2$, wybieramy najlepsze α . Następnie, postępując podobnie z β względem najlepszego α , równego jednak 0,4, otrzymamy zestaw tych parametrów dający najmniejsze $RMSE$.

Tabela 1.13.2

Wyniki symulacji parametrów α i β

α	β	$RMSE$
0,5	0,2	0,252
0,4	0,2	0,247
0,3	0,2	0,252
0,4	0,1	0,265
<u>0,4</u>	<u>0,3</u>	<u>0,243</u>

Źródło: Opracowanie własne w Excelu.

W tabeli wyróżniono podkreśleniem najlepszy zestaw parametrów $\alpha = 0,4$ i $\beta = 0,3$.

1.13.1. Wprowadzenie składowej sezonowej do modelu Holta-Wintersa (z użyciem Excela)

Zmiany w czasie wywołane są takimi przyczynami, jak trend, wahania sezonowe/okresowe, wahania cykliczne oraz wahania przypadkowe¹⁴⁰. Trendem nazywamy długookresowe zmiany powodujące wzrost lub spadek zmiennej objaśniającej. Wahania sezonowe to zmiany poziomu badanego zjawiska w tym samym czasie każdego roku. Jeśli występują nieregularne zmiany określonego zjawiska (badanej zmiennej) i mają one charakter falowy wokół trendu, to mówimy o wahaniami cyklicznych. Wahania przypadkowe (losowe) to zakłócenia w regularnych zmianach obserwowanej zmiennej objaśniającej.

W rzeczywistych szeregach czasowych zmiany nie występują tylko w wymienionych formach, lecz są z nich złożone. Jeśli np. dla kolejnych okresów amplituda zmian jest stała przy istniejącym trendzie i wahaniami cyklicznych, to taki model nazywamy addytywnym, a jeśli nie – to modelem multiplikatywnym.

W modelu Holta-Wintersa bez sezonowości występowały dwa parametry wyrównywania, przy czym α odnosiło się do wartości szeregu zmiennej objaśnianej, a parametr β dotyczył trendu.

Wyrównywanie sezonowości w modelu Holta-Wintersa przeprowadza się według określonej procedury poprzez zastosowanie parametru γ (gamma), stosowanego do obliczania indeksów sezonowości. Sezonowość multiplikatywna modelu Holta-Wintersa ma postać:

$$Y_{n+1} = (S_n + b_n) \cdot I_{n-L+1} + z_{n+1},$$

gdzie:

I_{n-L+1} – wyrównana wartość indeksu sezonowości na okres $n + 1$,

z_{n+1} – błąd w okresie $n + 1$,

L – długość cyklu sezonowości (12 dla danych miesięcznych, 4 dla kwartalnych).

Wyrównana wykładniczo wartość Y_n po usunięciu sezonowości :

$$S_n = \alpha \cdot (Y_n / I_{n-L}) + (1 - \alpha) \cdot (S_{n-1} + b_{n-1}).$$

Wyrównana wykładniczo wartość trendu w okresie n :

$$b_n = \beta (S_n - S_{n-1}) + (1 - \beta) \cdot b_{n-1}.$$

Wyrównany wykładniczo indeks sezonowości w okresie n :

¹⁴⁰ Ibidem, rozdz. 2.2.

$$I_n = \gamma \cdot (Y_n/S_n) + (1 - \gamma) I_{n-L}.$$

Prognoza na okres $n + m$, przy czym m to horyzont prognozy:

$$F_{n+m} = (S_n + b_n \cdot m) \cdot I_{n-L+1}. \quad (1.13.1)$$

Przykład

Spożycie masła jako jednego z podstawowych tłuszczów jest większe w porach chłodnych, a mniejsze w ciepłych. Obserwacja zużycia masła w gospodarstwie domowym pozwala na wstępne oszacowanie średnimi współczynnikami zużycia całorocznego w kwartałach na około:

- 0,25 (styczeń, luty, marzec),
- 0,30 (kwiecień, maj, czerwiec),
- 0,10 (lipiec, sierpień, wrzesień),
- 0,35 (październik, listopad, grudzień).

Suma współczynników = 1.

Rozpatrzmy cztery kolejne lata – od 2000 do 2003 – pod względem spożycia masła wyrażonego w gramach, stanowiące dane statystyczne przeliczone na mieszkańca kraju. Roczne wartości spożycia (tab. 1.12.32) zostały rozpisane na kwartały z uwzględnieniem podanych współczynników sezonowości.

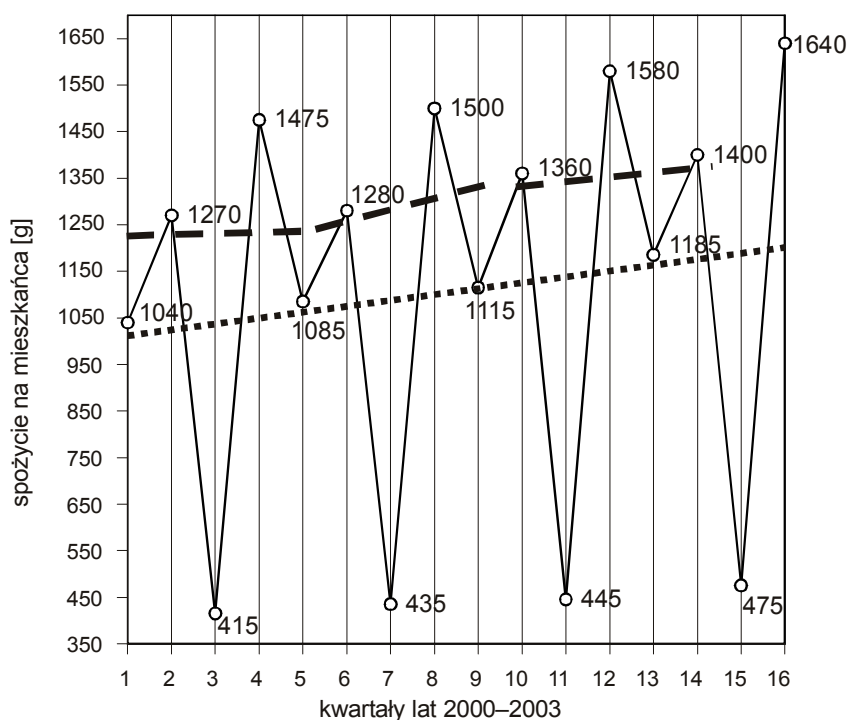
Tabela 1.13.3

Indywidualne spożycie masła w kwartałach

Rok	Kwartał/współczynnik	1/0,25	2/0,30	3/0,10	4/0,35	Spożycie/rok
2000		1 050	1 260	420	1 470	4 200
2001		1 075	1 290	430	1 505	4 300
2002		1 125	1 350	450	1 575	4 500
2003		1 175	1 410	470	1 645	4 700

Źródło: Opracowanie własne w Excelu.

Przedstawmy przytoczone dane na wykresie punktowym z zaznaczeniem wartości kwartalnych nieco zróżnicowanych, aby przypominały dane rzeczywiste. Linia prostą zaznaczono trend liniowy, a linią przerywaną połączono wartości obserwacji w kwartale drugim.



Źródło: Opracowanie własne w Excelu.

Wykres 1.13.2. Dane modelowe i trend liniowy indywidualnego spożycia masła w kwartałach

Postępowanie z uwzględnieniem sezonowości było następujące:

a) Określenie wartości początkowych na podstawie siedmiu początkowych obserwacji¹⁴¹.

Dla ustalenia średniej wartości trendu między kolejnymi latami przyjmijmy trzy kwartały (1, 2, 3).

Tabela 1.13.4

Obliczenia pomocnicze do ustalenia trendu

Rok/kwartał	$Y_5 - Y_1$	$Y_6 - Y_2$	$Y_7 - Y_3$	Suma
2000.1	1 040	1 270	415	1 475
2001.1	1 085	1 280	435	1 500
Σ	45	10	20	75

Źródło: Opracowanie własne w Excelu, kwartały zapisano jako rok.kw.

¹⁴¹ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria...*, op.cit., rozdz. 2.15.

Dzielimy sumę 75 przez liczbę wszystkich okresów (16), traktując wynik jako średnią wartość trendu: $75/16 = 4,6875$.

b) Obliczamy:

S_n – średnią arytmetyczną pierwszych czterech wartości Y :

$S_1 = S_2 = S_3 = S_4 = (1040 + 1270 + 415 + 1475)/4 = 1050$; wartość ta jest scentrowana na kwartale 1,5.

c) Dokonujemy korekty wartości trendu b_n odpowiednim mnożnikiem, np. dla pierwszych czterech kwartałów (w zależności od wzrostu czy spadku spożycia względem wartości średniej wynoszącej 1050).

1 – $1040/1050$ (-0,1),

2 – $1270/1050$ (1,2),

3 – $415/1050$ (-2,5),

4 – $1475/1050$ (+1,4).

Na przykład kwartał trzeci: $1050/415 = 2,53 \approx 2,5$, spadek (-2,5)

T a b e l a 1.13.5

Wstępnie skorygowana wartość trendu w kwartałach

Kwartał	Trend b_n
1	$-0,1 \cdot 4,6875 = -0,469$
2	$1,2 \cdot 4,6875 = 5,625$
3	$-2,5 \cdot 4,6875 = -11,719$
4	$1,4 \cdot 4,6875 = 6,563$

Ź r ó d ł o: Opracowanie własne w Excelu.

d) Obliczamy wyrównaną wykładniczo wartość zmiennej i trendu na koniec pierwszych czterech kwartałów.

T a b e l a 1.13.6

Wyrównana wartość „ $S_n + b_n$ ”

Kwartał	Formuła	Obliczenie
1	$S_1 + b_1$	$1\ 050 - 0,469 = 1\ 049,531$
2	$S_2 + b_2$	$1\ 050 + 5,625 = 1\ 055,625$
3	$S_3 + b_3$	$1\ 050 - 11,719 = 1\ 038,281$
4	$S_4 + b_4$	$1\ 050 + 6,563 = 1\ 056,563$

Ź r ó d ł o: Opracowanie własne w Excelu.

e) Wyznaczamy indeksy sezonowości:

$$I_n = Y_n / (S_n + b_n)$$

Tabela 1.13.7

Wstępnie ustalone indeksy sezonowości

Kwartał	Y_n	$S_n + b_n$	I_n
1	1 040	1 049,531	0,991
2	1 270	1 055,625	1,203
3	415	1 038,281	0,400
4	1 475	1 056,563	1,396

Źródło: Opracowanie własne w Excelu.

f) Ponowne ustalenie wartości początkowych:

– skorygowane S_4 o indeks sezonowości dla czwartego kwartału:

$$S_4 = Y_4/I_4 = 1475/1,396 = 1056,590,$$

– ustalenie $b_4 = 4,6875$ (wartość wstępnego trendu),

– wstępne przyjęcie wartości parametrów jednakowo: $\alpha = 0,2$; $\beta = 0,2$; $\gamma = 0,2$.

g) Przeprowadzenie obliczeń prognoz przy długości cyklu $L = 4$ oraz przyjętych parametrach α, β, γ .

Zastosowane wzory na przykładzie obserwacji piątej i umiejscowienia danych w arkuszu kalkulacyjnym według tabeli 1.13.8:

$$S_5 = \alpha (Y_5/I_{5-L}) + (1 - \alpha)(S_{5-1} + b_{5-1}),$$

$$b_5 = \beta (S_5 - S_{5-1}) + (1 - \beta) \cdot b_{5-1},$$

$$I_5 = \gamma \cdot (Y_5/S_5) + (1 - \gamma) I_{5-L}, \quad F_{5+1} = (S_5 + b_5) \cdot I_{5-L+1}.$$

Tabela 1.13.8

Zastosowane wyrażenia Excela

Przykład wyrażen Excela i obliczenia manualnego dla obserwacji 5	
S_5	$=(\$B\$59*C65/G61)+((1-\$B\$59)*F64)$ $(0,2 * 1085/0,991) + ((1 - 0,2) * 1061,278 = 1067,993$
b_5	$=(\$D\$59*(D65-D64))+((1-\$D\$59)*E64)$ $(0,2*(1067,993 - 1056,590)) + ((1 - 0,2) * 4,6875)$
$S_5 + b_5$	$=D65+E65$ $1067,993+6,0305 = 1074,023$
I_5	$=\$F\$59*C65/D65)+((1-\$F\$59)*G61)$ $(0,2 * 1085/1067,993) + ((1 - 0,2) * 0,991)$
F_5	$=D64-E64)*G61$ $(1056,590 + 4,6875) * 0,991$
$Y_5 - F_5$	$=C65-H65$ $1085 - 1276,717 = -191,717$
$(Y_5 - F_5)^2$	$=I65^2$ $1276,717^2 = 36755,344$
F_{20}	$=(\$D\$76+\$E\$76)*G76$ $(1179,97 + 11,150) * 1,395 = 1661,219$

Źródło: Opracowanie własne.

Tabela 1.13.9

Obliczenia pomocnicze wartości danych do określenia prognoz

$\alpha=0,2$		$\beta=0,2$		$\gamma=0,2$		$L=4$			
Okres	kw.	Y_k	S_n	b_n	$S_n + b_n$	I_n	F_n	$Y_n - F_n$	$(Y_n - F_n)^2$
1	2000.1	1 040			1049,531	0,991			
2	2000.2	1 270			1055,625	1,203			
3	2000.3	415			1038,281	0,400			
4	2000.4	1 475	1056,590	4,6875	1061,278	1,396			
5	2001.1	1 085	1067,993	6,0305	1074,023	0,996	1051,726	33,274	1107,159
6	2001.2	1 280	1072,02	5,6299	1077,650	1,201	1292,050	-12,050	145,203
7	2001.3	435	1079,62	6,0239	1085,644	0,401	431,060	3,940	15,524
8	2001.4	1 500	1083,415	5,5781	1088,993	1,394	1515,559	-15,559	242,073
9	2002.1	1 115	1095,093	6,7982	1101,891	1,000	1084,620	30,380	922,921
10	2002.2	1 360	1107,953	8,0105	1115,964	1,206	1323,594	36,406	1325,425
11	2002.3	445	1114,947	7,8071	1122,754	0,400	447,037	-2,037	4,150
12	2002.4	1 580	1124,937	8,2438	1133,181	1,396	1564,784	15,216	231,520
13	2003.1	1 185	1143,444	10,296	1153,741	1,008	1133,661	51,339	2635,709
14	2003.2	1 400	1155,077	10,564	1165,641	1,208	1391,941	8,059	64,947
15	2003.3	475	1169,839	11,403	1181,243	0,401	466,596	8,404	70,626
16	2003.4	1 640	1179,974	11,150	1191,123	1,395	1648,857	-8,857	78,451
17							1200,185	Σ	6843,706
18							1438,371	$\Sigma/12$	570,309
19							478,166	RMSE	23,881
20							1661,219		

Źródło: Opracowanie własne w Excelu.

Prognozy na kwartały kolejnego roku bazują na danych okresu 16., mianowicie:

 $S_{16}, b_{16}, I_{13}, I_{14}, I_{15}, I_{16}$.

Wyrażenie do ich obliczenia podaje tabela 1.13.8.

h) Obliczamy błąd wygasłych 16 prognoz, czyli $RMSE = 23,881$.

Realizując dalej, możemy dokonać symulacji na parametrach α, β, γ tab. 1.13.9 (będącej w arkuszu kalkulacyjnym programu Excel) w celu zmniejszenia $RMSE$. Przykładowo dla $\alpha = 0,2; \beta = 0,6$, wprowadzenie zamiast $\gamma = 0,2$ parametru $\gamma = 0,6$ co daje poprawę $RMSE$ na 20,797.

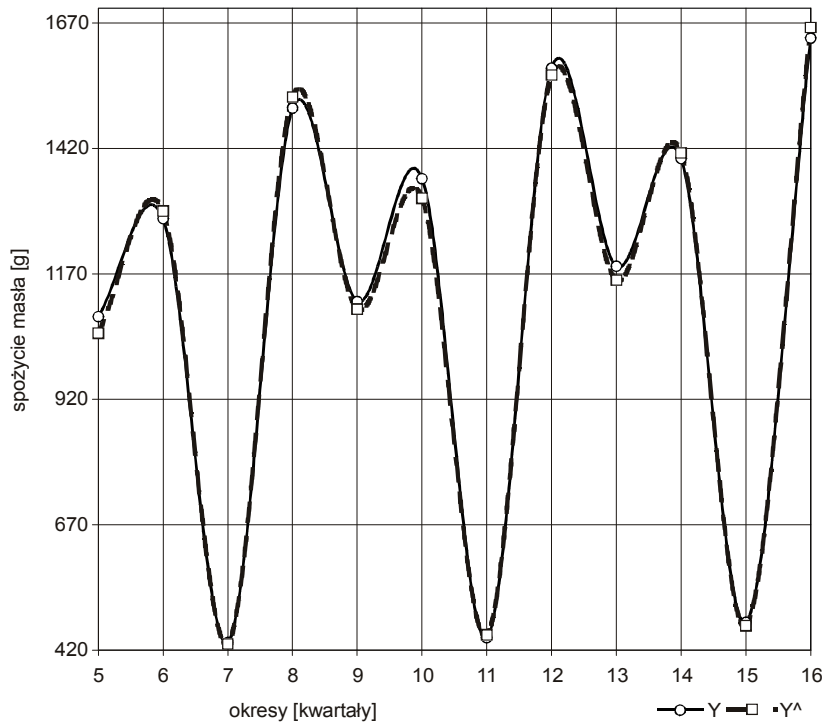
Tabela 1.13.10

Obliczone prognozy dla
 $\alpha = 0,2, \beta = 0,6, \gamma = 0,6$

Okres	Y	$Y^{\wedge} = F$
5	1 085	1 052
6	1 280	1 295
7	435	433
8	1 500	1 522
9	1 115	1 100
10	1 360	1 321
11	445	450
12	1 580	1 566
13	1 185	1 158
14	1 400	1 411
15	475	469
16	1 640	1 661

Źródło: Opracowanie własne w Excelu.

Przedstawmy ten wariant rozwiązania na wykresie punktowym.



Źródło: Opracowanie własne w Excelu.

Wykres 1.13.3. Dane empiryczne i teoretyczne według modelu multiplikatywnego Holta-Wintersa

Widzimy bardzo dobre dopasowanie modelu do danych empirycznych spożycia masła w kwartałach lat 2000–2001.

Warto jeszcze podać równania stanowiące model Holta-Wintersa z **sezonowością addytywną**¹⁴²:

$$\begin{aligned}
 S_n &= \alpha (Y_n - I_{n-L}) + (1 - \alpha)(S_{n-1} + b_{n-1}), & b_n &= \beta (S_n - S_{n-1}) + (1 - \beta) b_{n-1}, \\
 I_n &= \gamma (Y_n - S_n) + (1 - \gamma)I_{n-L}, & F_{n+m} &= (S_n + b_n m) + I_{n-L+1}.
 \end{aligned}
 \tag{1.13.2}$$

Manualne obliczenia oraz wykres wspomagane Excelem sporządzamy analogicznie do przedstawionego modelu multiplikatywnego.

¹⁴² Ibidem, s. 94.

1.13.2. Przykłady doboru najlepszych parametrów modelu (zastosowanie modułu FC programu WinQSB)

Manualne budowanie modeli adaptacyjnych jest pracochłonne. Z tego względu stosuje się programy komputerowe do automatyzacji obliczeń, zwłaszcza w celu doboru najlepszych z punktu widzenia miary prognozy, np. *RMSE* parametrów wyrównywania wykładniczego. Podano procedury korzystania z programu WinQSB dotyczące modelu podwójnego wyrównywania wykładniczego z trendem (*DEST*) oraz modelu Holta-Wincersa z sezonowością multipikatywną.

1.13.2.1. Podwójne wyrównywanie wykładnicze z trendem

Procedura realizacji bazująca na danych statystycznych (lata 70. i 80.) podanych w tab. 1.12.41 jest następująca:

– WinQSB; (*Forecasting*); (*File*); (*New Problem*); (*Time Series Forecasting*).

– (*Problem Title*) (*DEST* – podwójne wyrównywanie wykładnicze z trendem); (*Time Unit*) (rok); (*Number of Time Units*) (*Periods*) (12); *OK*,

– (*Historical Date*)...; (*File*); (*Save Problem AS*) (*DEST.fcc*),

– (*Solve an Analyze*); (*Perform Forecasting*); (*Double exponential smoothing with trend – DEST*); (*Metod Parameters*); (*Assign Values*) lub (*Search the best*); (*Search Criterion*) (*MSE*); (*Number of periods to forecast*) (2); (*Smoothing constat alpha*) (**0.1**); (*Initial value F() if known*); (*MSE =*; *Enter Search Domain Alpha Start*) (0) (*End*) (1) (*Step*) (0.01).

Obok opcji podano wprowadzone informacje:

– tytuł problemu: *DEST* – podwójne wyrównywanie wykładnicze,

– jednostka szeregu czasowego – rok, liczba okresów – 12,

– nazwa pliku: *DEST.fcc*, horyzont prognozowania – 2,

– krok iteracji – 0.01 (z kropką dziesiętną).

Rozwiązanie przebiega w zależności od zadanej opcji:

– przypisanie wartości początkowej parametru α ,

– komputerowy – automatyczny dobór parametru α w kolejnych krokach dla osiągnięcia minimum *RMSE*.

Rozwiązanie modelu przy pierwszej opcji (*Assign Values*) (przypisanie wartości) modułem *FC* jest analogiczne do wykonanego wcześniej funkcjami Excela (tab. 1.12.41).

Tabela 1.13.11

Raport z prognozowania według podwójnego wyrównywania wykładniczego z trendem opcja „Assign Values”

08-15-2007 Rok	Actual Data	Forecast by DEST	Forecast Error	CFE	MAD	MSE	MAPE [%]
1	4,91						
2	5,2	4,91	0,29	0,29	0,29	0,084	5,577
3	5,4	4,968	0,432	0,722	0,361	0,135	6,788
4	5,3	5,057	0,243	0,965	0,322	0,11	6,052
5	5,75	5,113	0,637	1,602	0,4	0,184	7,308
6	5,5	5,25	0,25	1,852	0,37	0,16	6,755
7	6	5,316	0,684	2,535	0,423	0,211	7,529
8	6,5	5,471	1,029	3,564	0,509	0,332	8,714
9	6,1	5,702	0,398	3,962	0,495	0,31	8,44
10	6,7	5,818	0,882	4,844	0,538	0,362	8,965
11	6,8	6,034	0,766	5,61	0,561	0,385	9,196
12	7,4	6,235	1,165	6,775	0,616	0,473	9,79
13	7,35	6,524	0,826	7,6	0,633	0,49	9,911
14	7,5	6,757	0,743	8,343	0,642	0,495	9,91
15	7,6	6,982	0,618	8,961	0,64	0,487	9,783
16	8,2	7,189	1,011	9,972	0,665	0,523	9,953
17	8,4	7,481	0,919	10,892	0,681	0,543	10,015
18	8,24	7,764	0,476	11,367	0,669	0,524	9,765
19	8,54	7,968	0,572	11,939	0,663	0,513	9,595
20	8,9	8,196	0,704	12,643	0,665	0,512	9,506
21		8,456					
22		8,583					
CFE		12,643					
MAD		0,665					
MSE		0,512					
MAPE		9,506					
Trk.Signal		19					
R – sqaure		1					
		Alpha = 0,1					
		F(0) = 4,91					
		F'(0) = 4,91					

Źródło: Opracowanie własne w WinQSB, moduł FC, opcja DEST.

Błąd wygasłych prognoz według MSE w tabeli 1.13.11:

$$RMSE = \sqrt{MSE} = \sqrt{0,512} = 0,715.$$

Prognozy dla $\alpha = 0,1$ na okres 21. oraz 22. są identyczne jak w obliczeniach manualnych (tab. 1.12.41, rozdz. 1.12.9.6).

Wykonajmy teraz automatyczny dobór parametru α drugą opcją (*Search the best*) z punktu widzenia minimum RMSE.

Tabela 1.13.12

Raport z prognozowania z wyborem DEST i z użyciem opcji wyboru najlepszego rozwiązania

08-15-2007 Rok	Actual Data	Forecast by DEST	Forecast Error	CFE	MAD	MSE	MAPE [%]
1	4,91						
2	5,2	4,910	0,290	0,290	0,290	0,084	5,577
3	5,4	5,113	0,287	0,577	0,289	0,083	5,446
4	5,3	5,349	-0,049	0,528	0,209	0,056	3,941
5	5,75	5,386	0,364	0,892	0,248	0,075	4,541
6	5,5	5,705	-0,205	0,687	0,239	0,069	4,379
7	6	5,671	0,329	1,016	0,254	0,075	4,564
8	6,5	5,985	0,515	1,531	0,291	0,102	5,043
9	6,1	6,470	-0,370	1,160	0,301	0,107	5,171
10	6,7	6,399	0,301	1,462	0,301	0,105	5,096
11	6,8	6,752	0,048	1,510	0,276	0,095	4,658
12	7,4	6,965	0,435	1,946	0,290	0,103	4,769
13	7,35	7,454	-0,104	1,841	0,275	0,096	4,490
14	7,5	7,620	-0,120	1,722	0,263	0,089	4,267
15	7,6	7,761	-0,161	1,560	0,256	0,085	4,114
16	8,2	7,859	0,341	1,901	0,261	0,087	4,117
17	8,4	8,289	0,111	2,012	0,252	0,082	3,942
18	8,24	8,600	-0,360	1,652	0,258	0,085	3,967
19	8,54	8,594	-0,054	1,598	0,247	0,080	3,782
20	8,9	8,759	0,141	1,739	0,241	0,077	3,666
21		9,053					
22		9,266					
CFE		1,739					
MAD		0,241					
MSE		0,077					
MAPE		3,666					
Trk.Signal		7,204					
R-qaure		1					
		Alpha = 0,35					
		F(0) = 4,91					
		F(0) = 4,91					

Źródło: Opracowanie własne.

Program ustalił parametr wyrównujący $\alpha = 0,35$. Błąd prognozy uległ znacznemu zmniejszeniu:

$$RMSE = \sqrt{MSE} = \sqrt{0,077} = 0,277.$$

1.13.2.2 Model Holta-Wintersa z sezonowością multiplikatywną

We wcześniejszej tab. 1.13.9, opracowanej w Excelu, model Holta-Wintersa z sezonowością multiplikatywną (do określenia prognoz na okresy 17–20) zbudowano dla:

- jednakowych wartości parametrów $\alpha, \beta, \gamma = 0,2$,
- długości cyklu sezonowości $L = 4$,
- wyznaczonych wartości początkowych S_n, b_n na okres 4.,
- określonych wstępnie indeksów sezonowości na okresy 1–4.

Według tabeli 1.13.9 błąd $RMSE$ wynosi 23,881. Następnie podjęto manualną próbę dobierania parametrów i zmniejszono $RMSE$ do 20,797.

Skorzystajmy jednak z automatycznego doboru parametrów opcją (*Holt-Winters Multiplicative Algorithm* – HWM) modułu *FC* programu komputerowego WinQSB, aby $RMSE$ osiągnęło minimum. Dobór przeprowadzono ze skokiem co 0.01 (z kropką dziesiętną) bez podawania uzyskanych wartości początkowych. W celu lepszego porównania wykonajmy dwie wersje:

- a) wersję (*Assign Values*) – podanie wartości parametrów

Tabela 1.13.13

Raport z prognozowania z sezonowością multiplikatywną z podaniem wartości początkowych $\alpha = 0,2$ i $\beta = 0,2$; $\gamma = 0$,

08-17-2007 Kwartał	Actual Data	Forecast by HWM	Forecast Error	CFE	MAD	MSE	MAPE [%]
1	1040						
2	1270						
3	415						
4	1475						
5	1085	1040	45	45	45	2025	4,1475
6	1280	1283,189	-3,1885	41,8115	24,0942	1017,583	2,1983
7	435	419,7778	15,2222	57,0337	21,1369	755,627	2,632
8	1500	1507,371	-7,3708	49,6628	17,6954	580,3026	2,0968
9	1115	1072,107	42,8932	92,556	22,7349	832,2072	2,4468
10	1360	1315,924	44,0757	136,6317	26,2917	1017,284	2,5792
11	445	438,0271	6,9729	143,6046	23,5319	878,9033	2,4346
12	1580	1561,356	18,6438	162,2484	22,9209	812,4893	2,2778
13	1185	1126,505	58,4949	220,7433	26,8735	1102,396	2,5731
14	1400	1390,384	9,6162	230,3595	25,1478	1001,404	2,3845
15	475	460,5273	14,4727	244,8322	24,1774	929,4088	2,4447
16	1640	1651,602	-11,6021	233,2302	23,1294	863,1753	2,3
17		1198,017					
CFE		233,2302					

cd. tab. 1.13.13

08-17-2007 Kwartał	Actual Data	Forecast by <i>HWM</i>	Forecast Error	<i>CFE</i>	<i>MAD</i>	<i>MSE</i>	<i>MAPE</i> [%]
<i>MAPE</i>		2,3					
<i>Trk.Signal</i>		10,0837					
<i>R-sqaure</i>		1					
		$c = 4$					
		Alpha = 0,2					
		Beta = 0,2					
		Gamma = 0,2					
		$F(0) = 1050$					
		$T(0) = 0$					
		$S(1) = 0,9905$					
		$S(2) = 1,2095$					
		$S(3) = 0,3952$					
		$S(4) = 1,4048$					

Źródło: Opracowanie własne programem WinQSB, moduł *FC*, prognozowanie z opcją (*HWM*).

Program przyjął jako prognozę początkową $F(0) = 1050$, czyli średnią arytmetyczną z obserwacji 1–4 i określił indeksy sezonowości $S(1)$, $S(2)$, $S(3)$, $S(4)$. Po uwzględnieniu indeksu sezonowości dla kwartału pierwszego prognoza $F_1 \approx 1040$. Gdy porównamy wyniki z tab. 1.13.9, zauważymy duże przybliżenie z wynikami realizacji manualnej modelu. Błąd wygasłych prognoz jako pierwiastek z wartości *MSE* (tab. 1.13.13) wynosi 29,380; jest więc rzędu obliczonego z zastosowaniem funkcji elementarnych Excela, który wynosi 23,881;

b) wersja z wyszukaniem najlepszej wartości parametrów (*Search the best*).

Tabela 1.13.14

Raport z realizacji prognozowania z opcją wyboru najlepszego rozwiązania według kryterium *MSE*

08-17-2007 Kwartał	Actual Data	Forecast by <i>HWM</i>	Forecast Error	<i>CFE</i>	<i>MAD</i>	<i>MSE</i>	<i>MAPE</i> (%)
	10401						
2	1270						
3	415						
4	1475						
5	1085	1040	45	45	45	2025	4,1475
6	1280	1285,387	-5,3866	39,6134	25,1933	1027,008	2,2841
7	435	422,049	12,951	52,5644	21,1125	740,5815	2,5152
8	1500	1521,002	-21,0015	31,563	21,0848	665,7015	2,2364
9	1115	1104,046	10,9545	42,5174	19,0587	556,5612	1,9856

08-17-2007 Kwartał	Actual Data	Forecast by <i>HWM</i>	Forecast Error	<i>CFE</i>	<i>MAD</i>	<i>MSE</i>	<i>MAPE</i> (%)
10	1360	1327,577	32,4229	74,94	21,286	639,01	20,052
11	445	449,07	-4,07	70,87	18,827	550,09	1,8895
12	1580	1572,617	7,3834	78,2537	17,3962	488,1409	1,7118
13	1185	1163,858	21,1422	99,3959	17,8124	483,5689	1,7198
14	1400	1418,912	-18,9124	80,4836	17,9224	470,9797	1,6829
15	475	468,5322	6,4678	86,9513	16,8811	431,9663	1,6537
16	1640	1662,922	-22,922	64,0293	17,3845	439,7539	1,6324
17		1236,599					
<i>CFE</i>		64,0293					
<i>MAD</i>		17,3845					
<i>MSE</i>		439,7539					
<i>MAPE</i>		1,6324					
<i>Trk.Signal</i>		3,6831					
<i>R - square</i>		1					
		c = 4					
		Alpha = 0,14					
		Beta = 1					
		Gamma = 0,64					
		F(0) = 1050					
		T(0) = 0					
		S(1) = 0,9905					
		S(2) = 1,2095					
		S(3) = 0,3952					
		S(4) = 1,4048					

Źródło: Opracowanie własne.

Zastosowanie automatyzacji pozwoliło zmniejszyć błąd wygasłych prognoz *RMSE*.

$$RMSE = \sqrt{MSE} = \sqrt{439,7539} = 20,97.$$

Wynik jest bliski wartości $RMSE = 20,797$, uzyskanej wcześniej metodą prób i błędów przy parametrach $\alpha = 0,2$, $\beta = 0,6$, $\gamma = 0,6$.

1.13.3. Pytania

1. Podaj oraz wyjaśnij równania opisu modelu Holta-Wintersa (bez sezonowości).
2. Podaj przykład doboru wartości początkowych w budowie tego modelu.
3. Napisz i wyjaśnij równania modelu Holta-Wintersa z sezonowością multiplikatywną.

4. Wyjaśnij różnicę między modelem z sezonowością multiplikatywną a addytywną.

5. Napisz procedurę korzystania z modułu prognozowania (*Forecasting*) pakietu WinQSB.

1.13.4. Zadania

1. Określ brakujące informacje w tab. 1.13.15 dla modelu Holta-Wintersa (bez sezonowości), a więc: wartości początkowe, prognozy *ex post*, prognozę *ex ante* dla okresu piątego, przyjmując $\alpha = 0,2$, $\beta = 0,1$.

Tabela 1.13.15

Ramka inicjująca do obliczeń prognoz według modelu Holta-Wintersa

Okres	Y_n	S_n	b_n	$F_{n+m} = a_n + b_n \cdot m$	$Y_n - F_n$	$(Y - F_n)^2$
1	25					
2	28					
3	32					
4	36					
5						

Źródło: Opracowanie własne.

2. W supermarkiecie kwartalny popyt na wodę mineralną, w tys. litrów, w latach 2000–2003 był następujący:

Tabela 1.13.16.

Popyt na wodę mineralną

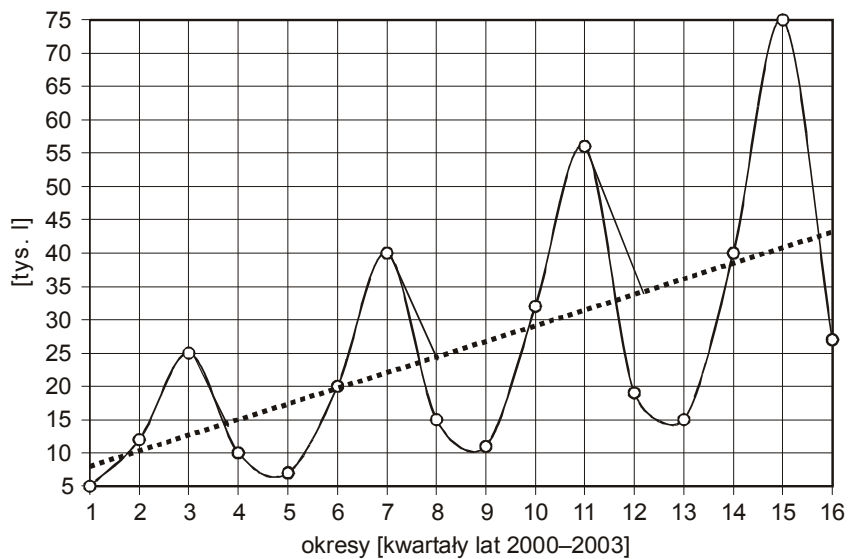
Rok \ kwartały	1	2	3	4
2000	5	12	25	10
2001	7	20	40	15
2002	11	32	56	19
2003	15	40	75	27

Źródło: Opracowanie własne.

Zbuduj model Holta-Wintersa z sezonowością multiplikatywną dla $\alpha = 0,2$, $\beta = 0,2$ oraz $\gamma = 0,1$, wcześniej określając: średnią wartość trendu, średnią wartość zmiennej dla kwartału/sezonu $L = 4$, indeksy sezonowości. Określ prognozę *ex post* dla 16 obserwacji.

3. Oblicz błąd *RMSE* prognozy określonej w zadaniu drugim.

4. Wykonaj analizę podanego wykresu punktowego popytu na wodę mineralną, określonego na podstawie danych zadania drugiego.



Źródło: Opracowanie własne w Excelu.

Wykres 1.13.4. Dane empiryczne oraz trend liniowy popytu na wodę mineralną

1.14. Model trendu, sezonowości i autoregresji (przykład w programie GRETL)

1.14.1. Wielomianowe modele trendu

Procesy ekonomiczne charakteryzują się niestacjonarnością średniej. Do celów prognozowania stosowane są złożone modele trendu, sezonowości i autoregresji danych rzeczywistych zmiennej objaśnianej. Istnieje potrzeba określenia stopnia wielomianu trendu, amplitud sezonowości, np. w kwartałach roku, oraz rzędu autokorelacji. W tym celu oraz do ostatecznej estymacji parametrów modelu złożonego interesujące jest zastosowanie funkcji ogólnie dostępnego (bezpłatnie w internecie) programu GRETL.

Założmy, że mamy do czynienia z procesem niestacjonarnym typu trendowo-sezonowo-autoregresyjnym o podejściu określonym modelem:

$$y_t = P_t + S_t + AR_p + e'_t$$
$$Y_t = \sum_{j=0}^r a_j t^j + \sum_{i=1}^{m-1} d_i Q_{it} + \sum \rho_p Y_{t-p} + e'_t,$$

gdzie:

P_t – wielomian j -ty trendu,

S_t – sezonowość w m kwartałach roku,

AR_p – autoregresja stopnia p ,

e'_t – reszta.

t – zmienna czasowa, $t = 1, 2, \dots, n$.

a_j – parametr wielomianu j -tego stopnia.

Pozostałe symbole wyjaśnione w podrozdziałach 1.14.2 i 1.14.3.

Według T. Kufli¹⁴³ przy założeniu stacjonarności wariancji składnikowy model procesu ekonomicznego możemy zapisać również jako:

$$Y_t = P_t + S_t + C_t + \eta_t,$$

gdzie:

P_t – wielomian j -ty trendu,

S_t – sezonowość w m kwartałach roku,

C_t – składnik koniunkturalny zależny od rozwoju gospodarczego,

$\eta_t = e'_t$ – składnik przypadkowy, czyli reszta, obliczana jako różnica między wartością obserwacji a wynikiem z modelu dla danego t .

¹⁴³ T. K u f e l, *Ekonometria. Rozwiązywanie problemów...*, op.cit., rozdz. 6.

Trend jako wielomian zmiennej czasowej r -tego stopnia przed oszacowaniem parametrów określamy wzorem:

$$Y_t = \sum_{j=0}^r \alpha_j t^j + \eta_t.$$

Aby określić stopień wielomianu r , oprzemy się na przykładzie liczby turystów (w tys.) korzystających z noclegów w pewnym mieście Polski, porównując modele $r = 1, 2, 3$ dla zmiennej czasowej kwartałów z lat 2001–2005.

Tabela 1.14.1

Turyści korzystający z noclegów

Kwartał \ Rok	2001	2002	2003	2004	2005
1	133	137	134	150	188
2	238	257	258	302	336
3	262	267	273	333	393
4	159	170	177	224	265

Źródło: Opracowanie własne na podstawie: R. C z y ż e w s k i, M. H u n d e r t, R. K ł ó s k a, *Wybrane zagadnienia z prognozowania*, Szczecin 2006, przykład 1.7.

Przedstawione dane charakteryzują się trendem rosnącym oraz sezonowością w poszczególnych kwartałach roku. Obliczenia ułatwi funkcja (*Model/Klasyczna metoda najmniejszych kwadratów...*) programu GRETL, który, jak już wspomniano, można zainstalować na własnym komputerze ze strony [www.kufel.torun.pl]. Program ten w wersji 1.6.5 wchodzi w skład bezpłatnego oprogramowania (*Open Source*).

Przed dokonaniem estymacji danego modelu trendu stopnia r musimy dodać zmienne czasowe: „time”, „time2”, „time3” do bazy danych, która w naszym przykładzie obejmuje 20 obserwacji (5 lat x 4 kwartały) zmiennej objaśnianej Y w okresach t . Zmienne niezależne definiujemy jako:

$$\text{time2} = \text{time}^2,$$

$$\text{time3} = \text{time}^3.$$

Wpierw inicjujemy bazę danych składającą się z 20 obserwacji (2001.1–2005.4), stosując funkcję (*Plik/Nowy zbiór danych*), wprowadzając zmienną niezależną Y nazwaną „Nocleg”. Następnie dodajemy zmienną czasową (*Dodawanie zmiennych/time-zmienna czasowa t*) (1–20). W kolejności definiujemy zmienne „time2” oraz „time3”, wprowadzając powyższe formuły. Rezultatem jest baza danych dla poszczególnych kwartałów $Q1$ – $Q4$; wydruk uzyskany z programu GRETL prezentuje tabela 1.14.2:

Tabela 1.14.2

Wydruk bazy danych z programu GRETL

obs	Nocleg	time	time2	time3
2001Q1	133	1	1	1
2001Q2	238	2	4	8
2001Q3	262	3	9	27
2001Q4	159	4	16	64
2002Q1	137	5	25	125
2002Q2	257	6	36	216
2002Q3	267	7	49	343
2002Q4	170	8	64	512
2003Q1	134	9	81	729
2003Q2	258	10	100	1 000
2003Q3	273	11	121	1 331
2003Q4	177	12	144	1 728
2004Q1	150	13	169	2 197
2004Q2	302	14	196	2 744
2004Q3	333	15	225	3 375
2004Q4	224	16	256	4 096
2005Q1	188	17	289	4 913
2005Q2	336	18	324	5 832
2005Q3	393	19	361	6 859
2005Q4	265	20	400	8 000

obs – identyfikator kolejnej obserwacji w kwartale roku,
 Nocleg – liczba turystów korzystających z noclegów,
 time – zmienna czasowa kwartałów lat 2001–2005,
 time2 – kwadrat zmiennej czasowej *time*,
 time3 – sześćcian zmiennej czasowej *time*.

Źródło: Opracowanie własne na podstawie opisu kolumn bazy danych w GRETL.

Teraz dokonujemy estymacji kolejnych trzech modeli trendu:

– liniowego:

$$\hat{Y} = a_0 + a_1t,$$

– kwadratowego:

$$\hat{Y} = a_0 + a_1t + a_2t^2,$$

– wielomianu trzeciego stopnia:

$$\hat{Y} = a_0 + a_1t + a_2t^2 + a_3t^3.$$

Zastosujemy funkcję GRETL-a (*Model/Klasyczna metoda najmniejszych kwadratów...*), traktując kolejno jako zmienne objaśniające:

$$t = \text{time},$$

$$t^2 = \text{time2},$$

$$t^3 = \text{time3}.$$

I tak dla zmiennej objaśnianej Y oraz zmiennej czasowej t uzyskamy statystykę, której fragment opisują modele.

Model 1

Estymacja KMNK z wykorzystaniem 20 obserwacji 2001.1–2005.4.

Zmienna zależna: Nocleg

Zmienna	Współczynnik	Błąd standardowy	Statystyka t
const	165,221	30,9422	5,3397
time	6,43609	2,583	2,4917

Średnia arytmetyczna zmiennej zależnej = 232,8,

Odchylenie standardowe zmiennej zależnej = 75,1872,

Suma kwadratów reszt = 79862,7,

Błąd standardowy reszt $S_e = 66,6094$,

Współczynnik determinacji (R^2) = 0,256463,

Skorygowany współczynnik determinacji (\bar{R}^2) = 0,215155,

Ilość stopni swobody (iss) = 18.

$$\hat{Y} = 165,221 + 6,43609t.$$

Model ten możemy zapisać jako ikonę sesji GRETL-a. Wykazuje on bardzo niskie dopasowanie do danych empirycznych $R^2 = 0,256463$.

A teraz kolejne dwa modele (trendu kwadratowego oraz wielomianu trzeciego stopnia) również dla 20 obserwacji zmiennej zależnej Nocleg.

Model 2

Zmienna	Współczynnik	Błąd standardowy	Statystyka t
const	198,844	49,9128	3,9838
time	-2,73377	10,9467	-0,2497
time2	0,43666	0,506334	0,8624

Średnia arytmetyczna zmiennej zależnej = 232,8,

Odchylenie standardowe zmiennej zależnej = 75,1872,

Suma kwadratów reszt = 76515,3,
 Błąd standardowy reszt = 67,0887,
 Współczynnik determinacji (R^2) = 0,287628,
 Skorygowany współczynnik determinacji (\bar{R}^2) = 0,20382,
 Ilość stopni swobody (iss) = 17.

$$\hat{Y} = 198,844 - 2,73377t + 0,43666t^2.$$

Model 3

Zmienna	Współczynnik	Błąd standardowy	Statystyka t
const	180,809	75,2872	2,4016
time	6,46852	30,2928	0,2135
time2	-0,632584	3,30941	-0,1911
time3	0,0339442	0,103755	0,3272

Średnia arytmetyczna zmiennej zależnej = 232,8,
 Odchylenie standardowe zmiennej zależnej = 75,1872,
 Suma kwadratów reszt = 76006,8,
 Błąd standardowy reszt = 68,9233,
 Współczynnik determinacji (R^2) = 0,292362,
 Skorygowany współczynnik determinacji (\bar{R}^2) = 0,15968,
 Ilość stopni swobody (iss) = 16.

$$\hat{Y} = 180,809 + 6,46852t - 0,632584t^2 + 0,0339442t^3.$$

Do porównywania modeli wykorzystamy wariancję resztową S_e^2 dla par modeli na podstawie próby oraz zastosujemy test F^{144} , który jest analogiczny do wzoru (1.1.21):

a) model trendu liniowego (l) i kwadratowego (k):

$$F_{obl} = \frac{(S_{el}^2)}{(S_{ek}^2)} = \frac{66,6094^2}{67,0887^2} = 0,9858.$$

Przyjmijmy hipotezy dotyczące wariancji resztowej populacji:

$$H_0 = \delta^2_l = \delta^2_k, \quad H_1 = \delta^2_l > \delta^2_k.$$

Gdy $\alpha = 0,05$ oraz $r_1 = 18$, $r_2 = 17$, to wartość krytyczna odczytana z tablicy rozkładu F wynosi $F_\alpha = 2,26$. Jeżeli $F_{obl} < F_\alpha$, czyli brak istotnego spadku wariancji resztowej, nie możemy odrzucić hipotezy zerowej.

¹⁴⁴ Ibidem, rozdz. 6.1.

Określenie wartości krytycznej F_α umożliwia również funkcja GRETŁ-a (*Narzędzia/Tablice statystyczne*):

$F(18, 17)$:

prawostronne prawdopodobieństwo = 0,05,

prawdopodobieństwo dopełnienia = 0,95,

krytyczna wartość = 2,25667 \approx 2,26;

b) model trendu kwadratowego (k) i wielomianu trzeciego stopnia (s):

$$F_{obl} = \frac{(S_{ek}^2)}{(S_{es}^2)} = \frac{67,0887^2}{68,9233^2} = 0,9475.$$

Gdy $\alpha = 0,05$, $r_1 = 17$, $r_2 = 16$, wartość krytyczna $F_\alpha = 2,30$, również występuje ta sama sytuacja.

Spróbujmy ocenić istotność zestawu parametrów strukturalnych poszczególnych modeli, bazując na $n = 20$ obserwacjach, liczbie zmiennych objaśniających m oraz współczynnika determinacji (R^2). Skorzystajmy ze statystyki analogicznej do wzoru (1.1.6), określonej w publikacji J. Dziechciarza¹⁴⁵:

$$F_{obl} = \frac{n - m - 1}{m} \cdot \frac{R^2}{1 - R^2}.$$

Wartości F_{obl} oraz wartości krytyczne F_α według tablicy rozkładu F ($\alpha = 0,05$, $r_1 = m$, $r_2 = n - m - 1$) dla poszczególnych modeli trendu wynoszą:

– liniowego: $m = 1$, $R^2 = 0,256463$, $F_{obl} = 6,21$, $F_\alpha = 4,41$, zatem $F_{obl} > F_\alpha$,

– kwadratowego: $m = 2$, $R^2 = 0,287628$, $F_{obl} = 3,43$, $F_\alpha = 3,59$, zatem $F_{obl} < F_\alpha$,

– wielomianu trzeciego stopnia: $m = 3$, $R^2 = 0,292362$, to $F_{obl} = 2,20$, $F_\alpha = 3,24$, $F_{obl} < F_\alpha$.

Ponieważ ocena parametrów wielomianu trzeciego stopnia jest nieistotna statystycznie, przyjmujemy trend kwadratowy, czyli stopień wielomianu zmiennej czasowej $r = 2$. Informację o potrzebie takiego wyboru podaje również program GRETŁ w wyświetlonej pełnej statystyce opisowej modelu trzeciego (pozostawia to się do sprawdzenia Czytelnikowi).

Warto nadmienić, że poszczególne modele możemy zapisać w formie ikony, „przeciągnąć myszką” do ikony zbiorczej sesji (*Tabela model*), a następnie wyświetlić prezentacje wszystkich modeli, których może być maksymalnie dwanaście.

¹⁴⁵ J. Dziechciarz (red.), *Ekonometria. Metody...*, op.cit.

1.14.2. Model Kleina ze zmiennymi sezonowymi

Nasz wieloskładnikowy model procesu ekonomicznego ma już określony stopień wielomianu zmiennej czasowej. W modelu występuje jednak też człon S_t obejmujący zmienne Q_i czterech kwartałów roku. Do estymacji modeli z trendem kwadratowym i zmiennymi sezonowymi umieszczamy dodatkowo w bazie danych zmienne o wartości zero lub jeden w zależności od odniesienia do określonego kwartału. Przyjmijmy, że nasz wstępny (z oszacowanymi parametrami) model jest bez ostatniej zmiennej sezonowej dla kwartału czwartego, a parametr (tzw. amplitudę) ostatniej zmiennej określamy z podanego warunku, że suma parametrów (amplitud) równa się zero:

$$Y_t = a_0 + a_1 t + a_2 t^2 + \sum_{k=1}^{m-1} d_{ok} Q_{kt} + e'_t, \quad \sum_{i=1}^m \hat{d}_{ok} = 0, \quad \hat{d}_{04} = -(\hat{d}_{01} + \hat{d}_{02} + \hat{d}_{03}).$$

Do zbudowania modelu trendu kwadratowego z periodycznymi składnikami sezonowymi w kwartałach roku zastosujemy model Kleina¹⁴⁶, korzystając z rachunku macierzowego i wyrażenia na wektor parametrów.

$$\hat{\mathbf{a}} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{y}.$$

$$\hat{\mathbf{a}} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ d_{01} \\ d_{02} \\ d_{03} \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 9 & 0 & 0 & 1 \\ 1 & 4 & 16 & -1 & -1 & -1 \\ 1 & 5 & 25 & 1 & 0 & 0 \\ 1 & 6 & 36 & 0 & 1 & 0 \\ 1 & 7 & 49 & 0 & 0 & 1 \\ 1 & 8 & 64 & -1 & -1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 20 & 400 & -1 & -1 & -1 \end{bmatrix}, \quad \hat{\mathbf{Y}} = \begin{bmatrix} 133 \\ 238 \\ 262 \\ 159 \\ 137 \\ 257 \\ 267 \\ 170 \\ \vdots \\ 265 \end{bmatrix}.$$

Kolumny macierzy \mathbf{T} dotyczą w kolejności zmiennych:
 t_0 – przy parametrze α_0 tożsamościowo równej 1,

¹⁴⁶ R. C z y ż e w s k i, M. H u n d e r t, R. K l ó s k a, *Wybrane zagadnienia z prognozowania...*, op.cit., rozdz. 1.21.

- t (time) – w bazie danych programu GRETL,
- t^2 (time2),
- Q_1 – I kwartału, przyjmującej, w zależności od obserwacji w kwartałach roku, wartości 1 lub 0,
- Q_2 – II kwartału,
- Q_3 – III kwartału,
- \mathbf{T}^T – macierz transponowana (zamiana kolumn wierszami),
- \mathbf{y} – wektor zmiennej objaśnianej.

Zastosowanie programu GRETL

Do oszacowania modelu trendu i sezonowości musimy wprowadzić do bazy GRETL-a dodatkowo zmienne Q_1 , Q_2 , Q_3 . Skorzystamy wstępnie z funkcji (*Dodawanie zmiennych/Periodyczne zmienne 0-1*). Program definiuje zmienne sezonowe i generuje je automatycznie jako dg1, dg2, dg3, dg4 dla danych kwartalnych, przy czym np. wartość dg1 dla kwartału pierwszego wynosi jeden, a dla pozostałych zero.

Aby wprowadzić do macierzy \mathbf{T} metodę Kleina, wykonujemy korekty co czwartej zmiennej z (1) na (-1), stosując funkcję (*Dane/Edycja wartości*). Przeprowadzamy estymację parametrów modelu trendowo-sezonowego, stosując klasyczną metodę najmniejszych kwadratów dla zmiennych:

- objaśnianej Y (Nocleg),
- zmiennych objaśniających z macierzy \mathbf{T} (const, time, time2, dg1, dg2, dg3).

Fragment statystyki opisowej modelu z procesu estymacji KMNK dla 20 obserwacji 2001.1–2005.4 podaje model 1 według metody Kleina, uzyskany z programu GRETL:

Model 1

Zmienna	Współczynnik	Błąd standardowy	Statystyka t
const	208,783	8,644	24,1535
time	-4,6083	1,88721	-2,4419
time2	0,504562	0,0871991	5,7863
dq1	-75,9233	4,52516	-16,7780
dq2	48,8983	4,47889	10,9175
dq3	70,3108	4,47889	15,6983

Średnia arytmetyczna zmiennej zależnej = 232,8,
 Odchylenie standardowe zmiennej zależnej = 75,1872,
 Suma kwadratów reszt = 1866,74,

Błąd standardowy reszt = 11,5472,
Współczynnik determinacji (R^2) = 0,98262,
Skorygowany współczynnik determinacji (\bar{R}^2) = 0,976413,
Statystyka $F(5, 14) = 158,308$,
Statystyka testu Durбина-Watsona = 1,79609,
Autokorelacja reszt rzędu pierwszego = 0,0735706,
Kryterium informacyjne Akaike'a = 159,482.

Na uwagę zasługuje bardzo dobre dopasowanie następującego modelu do danych rzeczywistych ($R^2 = 0,98262$);

$$\hat{Y}_t = 208,783 - 4,6083t + 0,504562t^2 - 75,9233Q_1 + 48,8983Q_2 + 70,3108Q_3.$$

Obliczmy jeszcze brakujący parametr dla okresu czwartego:

$$\hat{d}_{04} = -(-75,9233 + 48,8983 + 70,3108) = -43,2858.$$

Tak więc do prognozowania wszystkich kwartałów możemy stosować model:

$$\hat{Y}_t = 208,783 - 4,6083t + 0,504562t^2 - 75,9233Q_1 + 48,8983Q_2 + 70,3108Q_3 - 43,2858Q_4.$$

Realizacja w Excelu

Międzyprogramowa konwersja bazy danych z GRETL-a do Excela jest następująca:

Gretl: (*Dane/Zaznacz wszystko/Pokaz wartości*); (*Prezentacja danych/Kopiuj*),
Excel: (*Edycja/Wklej*); usuwamy time3, zamieniamy dg1, dg2, dg3 na zmienne Q_1 , Q_2 , Q_3 i dopisujemy Q_4 dla kwartału 4, obliczenie wartości teoretycznych Noclegt dla zmiennej zależnej (objaśnianej) według podanego modelu \hat{Y}_t , w którym nie występuje time3.

Tabela 1.14.3

Baza danych w Excelu po imporcie z bazy GRETL-a
oraz estymacja funkcją REGLINP

Rok kw.	Nocleg	time	time2	Q ₁	Q ₂	Q ₃	Q ₄	Noclegt	Noclegtr
20011	133	1	1	1	0	0	0	128,8	128,8
20012	238	2	4	0	1	0	0	250,5	250,5
20013	262	3	9	0	0	1	0	269,8	269,8
20014	159	4	16	0	0	0	1	155,2	155,1
20021	137	5	25	1	0	0	0	122,4	122,4
20022	257	6	36	0	1	0	0	248,2	248,2
20023	267	7	49	0	0	1	0	271,6	271,6
20024	170	8	64	0	0	0	1	161,0	160,9
20031	134	9	81	1	0	0	0	132,3	132,3
20032	258	10	100	0	1	0	0	262,1	262,1
20033	273	11	121	0	0	1	0	289,5	289,5
20034	177	12	144	0	0	0	1	182,9	182,9
20041	150	13	169	1	0	0	0	158,2	158,2
20042	302	14	196	0	1	0	0	292,1	292,1
20043	333	15	225	0	0	1	0	323,5	323,5
20044	224	16	256	0	0	0	1	221,0	220,9
20051	188	17	289	1	0	0	0	200,3	200,3
20052	336	18	324	0	1	0	0	338,2	338,2
20053	393	19	361	0	0	1	0	373,7	373,7
20054	265	20	400	0	0	0	1	275,2	275,2

-32	81,5966	60,1841	-64,6375	0,5046	-4,6083	197,497
a ₆	a ₅	a ₄	a ₃	a ₂	a ₁	a ₀

Źródło: Opracowanie własne w Excelu.

Do estymacji parametrów modelu trendowo-sezonowego zastosowano alternatywnie funkcję arkusza kalkulacyjnego Excel o przykładowej składni:

=REGLINP(C3:C22;D3:I22;PRAWDA;PRAWDA)

gdzie:

C3:C22 – obszar zmiennej zależnej Y , czyli Nocleg,

D3:I22 – obszar kolumn zmiennych objaśniających (time, time2, Q₁, Q₂, Q₃, Q₄).

Efekt oszacowania parametrów modelu podano w małej tabelce na zakończenie tabeli 1.14.3 oraz obliczono wartości teoretyczne zmiennej Y w kolumnie Noclegtr.

Warto zauważyć, że według REGLINP parametr stałej a_0 oraz parametry przy zmiennych sezonowych różnią się od obliczenia według modelu Kleina, lecz wyniki można przyjąć za identyczne.

Excel umożliwia również sporządzenie wykresu zmiennej objaśnianej Y (Nocleg) oraz Noclegt. Na wykresie 1.14.1 zaznaczono również trend kwadratowy, którego dopasowanie względem zmiennej Nocleg jest niskie. Współczynnik determinacji wynosi bowiem: $R^2 = 0,2876$. Przedstawiony graficznie model jest zbliżony do wielorównaniowego modelu Holta-Wintersa z sezo-

nowością addytywną, który, przypomnijmy sobie z podrozdziału 1.13.1, stosowany jest do określenia prognozy F na przyszły okres $n+m$ ¹⁴⁷. W modelu tym amplituda wahań zjawiska w czasie jest stała.

$$F_{n+m} = (S_n + b_n m) + I_{n-L+m},$$

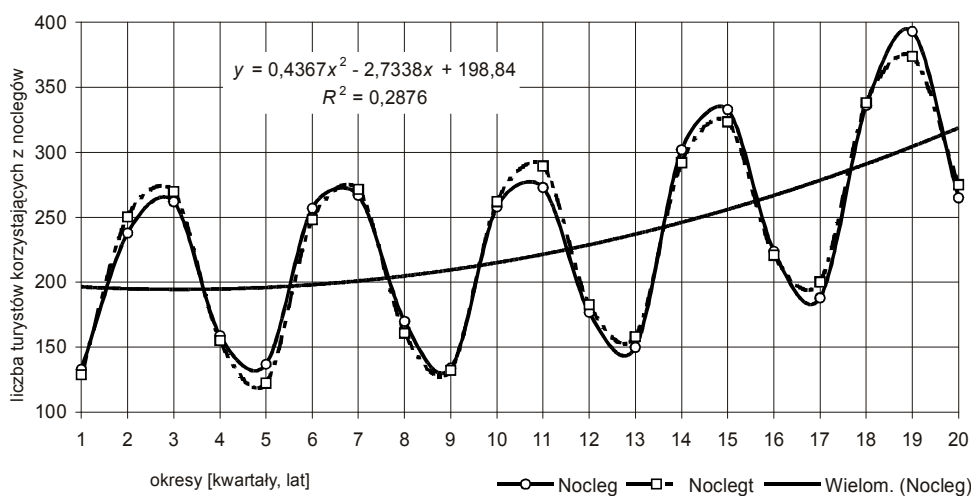
$$S_n = \alpha(Y_n - I_{n-L}) + (1 - \alpha)(S_{n-1} + b_{n-1}),$$

$$b_n = \beta(S_n - S_{n-1}) + (1 - \beta)b_{n-1},$$

$$I_n = \gamma(Y_n - S_n) + (1 - \gamma)I_{n-L},$$

gdzie:

- S_n – wyrównana wykładniczo wartość zmiennej Y_n po usunięciu sezonowości,
- b_n – wyrównana wykładniczo wartość trendu na koniec okresu n ,
- m – horyzont prognozy,
- n – liczba obserwacji,
- I_{n-L+1} – wyrównany wykładniczo indeks sezonowości w okresie $n-L+1$,
- L – długość rocznego cyklu sezonowości,
- Y_n – wartość rzeczywista zmiennej w okresie n ,
- α – parametr wyrównywania; β – parametr wyrównywania do oszacowania trendu b_n ,
- γ – parametr wyrównywania stosowany do obliczania indeksów sezonowości.



Źródło: Opracowanie własne w Excelu.

Wykres 1.14.1. Wartości empiryczne i teoretyczne zmiennej Y

¹⁴⁷ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonomia...*, op.cit., rozdz. 2.15.

1.14.3. Przykład modelu autoregresyjnego

W procesach ekonomicznych trendowo-sezonowo-autoregresyjnych rząd modelu autoregresyjnego p wyznaczany jest z reszt obliczonych według modelu trendu i sezonowości¹⁴⁸, który określiliśmy wcześniej jako:

$$\hat{Y}_t = 208,783 - 4,6083t + 0,504562t^2 - 75,9233Q_1 + 48,8983Q_2 + 70,3108Q_3 - 43,2858Q_4.$$

Model autoregresyjny procesu stacjonarnego o rzędzie opóźnienia p , w którym wartość bieżąca określana jest na podstawie wartości z poprzednich okresów, ma postać:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t.$$

Nawiązując do rozdziału 1.10 niniejszej pracy, warto przypomnieć, że procesem stacjonarnym nazywamy taki proces, dla którego szereg czasowy ma stałą średnią i wariancję, i kowariancja zależy od rzędu opóźnienia a nie od czasu¹⁴⁹.

Rząd opóźnienia można określić funkcją autokorelacji cząstkowej, która wyraża zależność pomiędzy współczynnikiem autokorelacji cząstkowej a rzędem opóźnienia (rozd. 1.10.1). Wartość współczynnika autokorelacji cząstkowej zawiera się w przedziale $<-1:1>$.

Przykładowo, współczynnik autokorelacji cząstkowej $p = 3$ bada zależność między Y_t i Y_{t-3} , gdy pominiemy wpływ zmiennych Y_{t-1} oraz Y_{t-2} . Gdy funkcja autokorelacji zanika, tj. wartości współczynników autokorelacji maleją wykładniczo i funkcja ta urywa się, to rząd autoregresji jest równy p .

W GRETL-u, oceniając składnik autoregresyjny procesu, skorzystać można z opcji (*Wykresy/Wykres reszt modelu/Korelogram*), przyjmując rząd opóźnienia 4 w menu okna wyników statystyki opisowej dla opisanego modelu trendowo-sezonowego. Maksymalna liczba opóźnienia p_{\max} nie powinna przekraczać 15–20% długości szeregu.

Dla ilości obserwacji $n = 20$:

$$p_{\max} = [(0,15 \cdot 20); (0,20 \cdot 20)] = 3; 4.$$

Podana jest tu funkcja autokorelacji (*ACF*) oraz autokorelacji cząstkowej (*PACF*) procesu resztowego. Fragment uzyskanych wyników z GRETL-a,

¹⁴⁸ T. K u f e l, *Ekonometria. Rozwiązywanie...*, op.cit., rozdz. 6.3.

¹⁴⁹ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2; *Ekonometria...*, op.cit., rozdz. 3.3.

zawiera tab. 1.14.4, przy czym nie występuje tu wyraźna tendencja wykładnicza zanikania oraz, w końcu, urywania się (przyjmowanie wartości bliskich zero) autokorelacji cząstkowej *PACF*. Zaobserwowanie tego wymaga znacznie większej liczby obserwacji zmiennej zależnej.

T a b e l a 1.14.4

Wartości funkcji autokorelacji i funkcji autokorelacji cząstkowej dla rzędu opóźnień „p”

<i>p</i>	<i>ACF</i>	<i>PACF</i>
1	0,0695	0,0695
2	-0,2720	-0,2782
3	-0,2226	-0,1950
4	0,0199	-0,0338

Ź r ó d ł o: Opracowanie własne w GRETL.

Współczynnik korelacji cząstkowej określonego rzędu opóźnienia *p* przy poziomie istotności $\alpha = 0,05$ porównujemy z wartością krytyczną, dla $n = 20$ obserwacji, która obliczana jest według wzoru¹⁵⁰:

$$PACF_{\alpha} = u_{\alpha} / \sqrt{n} = 1,96 / \sqrt{20} = 0,438.$$

Żadna z wartości opóźnień 1, 2, 3, 4 nie spełnia warunku $|PACF| > PACF_{\alpha}$ i niska jest również wartość współczynnika autokorelacji. Wszystkie wartości oszacowanych współczynników autokorelacji cząstkowej co do wartości bezwzględnej są mniejsze od wartości granicznej $PACF_{\alpha}$.

W celu skrócenia rozważań i dalszej prezentacji przykładu wieloskładnikowego modelu procesu ekonomicznego niestacjonarnego typu trendowo-sezonowo-autoregresyjnego przyjmijmy, że występuje autokorelacja rzędu pierwszego:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t.$$

W tej sytuacji model będzie miał postać:

$$\hat{Y}_t = a_0 + a_1 t + a_2 t^2 + a_3 Q_1 + a_4 Q_2 + a_5 Q_3 + a_6 Q_4 + a_7 Y_{t-1}.$$

W uzupełnieniu bazy danych o zmienną Y_{t-1} skorzystamy z funkcji (*Dodawanie zmiennych/Definiowanie nowej zmiennej*):

$$\text{Noclego} = \text{Nocleg} (-1).$$

¹⁵⁰ T. K u f e l, *Ekonometria. Rozwiązywanie...*, op.cit., rozdz. 6.3.

Uzyskaliśmy wektor zmiennej objaśnianej przesunięty o jeden kwartał. Pozwoliło to na dokonanie estymacji programem GRETl jednocześnie trzech członów klasyczną metodą najmniejszych kwadratów dla obserwacji 2–20 (2001.2–2005.4) dla zmiennej zależnej Nocleg i uzyskanie modelu 1.

M o d e l 1

Zmienna	Współczynnik	Błąd standardowy	Statystyka t
const	185,171	66,7923	2,7723
time	-3,33253	3,17657	-1,0491
time2	0,426265	0,201907	2,1112
dq1	-73,2031	13,382	-5,4703
dq2	56,3966	23,5836	2,3913
dq3	66,2029	15,2764	4,3337
Noclego	0,0922543	0,301296	0,3062

Średnia arytmetyczna zmiennej zależnej = 238,053,
 Odchylenie standardowe zmiennej zależnej = 73,3806,
 Suma kwadratów reszt = 1817,96,
 Błąd standardowy reszt = 12,3084,
 Współczynnik determinacji (R^2) = 0,981244,
 Skorygowany współczynnik determinacji (\bar{R}^2) = 0,971865,
 Statystyka $F(6, 12) = 104,63$,
 Ilość stopni swobody $iss = n - k = 20 - (m + 1) = 20 - 7 = 13$.

Ze względu na nieznaczne pogorszenie się współczynnika determinacji po wprowadzeniu członu z autoregresją pierwszego rzędu sprawdzimy istotność tego parametru przy pomocy testu t -Studenta dla $\alpha = 0,05$ i $iss = 13$:

$$|t_{obl}| > t_{\alpha} = 2,1604.$$

Warunku tego nie spełniają parametry zmiennych time oraz wyraźnie Noclego, więc zrezygnujemy ze zmiennej Noclego i wstępny model pozostaje w postaci wcześniej już zdefiniowanej, a więc trendowo-sezonowej.

$$\hat{Y}_t = a_0 + a_1 t + a_2 t^2 + a_3 Q_1 + a_4 Q_2 + a_5 Q_3 + a_6 Q_4.$$

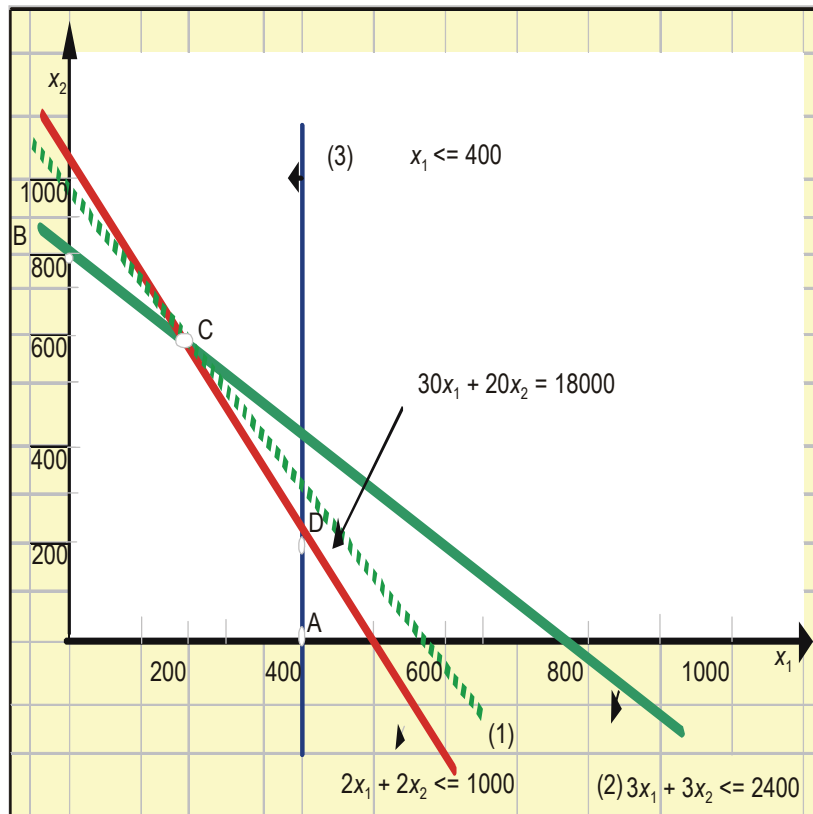
W niniejszym rozdziale zaprezentowano jedno z podejść do formułowania modelu trendowo-sezonowo-autoregresyjnego niestacjonarnego procesu ekonomicznego. Zarówno bowiem struktura modeli, jak i etapy ich uzyskania mogą być różnorodne. Występuje jednak znaczna pracochłonność ich budowy i weryfikacji. Z tego względu już w trakcie studiowania ekonometrii, jak

i w praktyce gospodarczej powinniśmy intensywnie korzystać ze wspomagania komputerowego, w tym z obszernych możliwości, jakie daje bezpłatnie dostępny w internecie program GRET. Próbkę tych możliwości w formie instrukcji korzystania z oprogramowania podano w niniejszej pracy.

Autorzy mają nadzieję, że będzie to kluczem do dalszego prześledzenia przydatności tego programu w celu poznania kolejnych metod i testów.

Cz ę ś ć 2

Elementy badań operacyjnych



2.1. Zastosowanie programu WinQSB oraz rozszerzenia Excel-Solver

2.1.1. O pakiecie WinQSB w zakresie badań operacyjnych

Pakiet WinQSB składa się z następujących modułów:

- zagregowane planowanie,
- alokacja środków,
- zarządzanie zapasami,
- planowanie pracy,
- procesy Markowa,
- gospodarka materiałowa,
- karty kontrolne,
- teoria kolejek,
- symulacja systemu kolejek,
- programowanie celowe,
- programowanie kwadratowe,

a także modułów będących obiektami naszego szczególnego zainteresowania w niniejszej części opracowania, mianowicie:

- analizy decyzji,
- programowania dynamicznego,
- prognozowania i regresji liniowej,
- modeli sieciowych,
- programowania nieliniowego,
- PERT_CPM,
- programowania liniowego i całkowitoliczbowego.

Korzystanie z menu tego pakietu zostanie zademonstrowane na przykładzie modułu (*Linear and Integer Programming*) programowanie liniowe i całkowitoliczbowe. Podstawowe menu główne tego modułu obejmuje rozwijalne opcje:

- plik (*File*):
 - nowy problem,
 - otwarcie istniejącego problemu,
 - zapisanie problemu,

- wydrukowanie,
- wyjście;
- edycja (*Edit*):
 - wycięcie,
 - kopiowanie,
 - wklejanie;
- format:
 - format zapisu liczbowego,
 - czcionka,
 - wyrównanie do lewej,
 - wyśrodkowanie,
 - wyrównanie do prawej,
 - wysokość wiersza,
 - szerokość kolumny;
- rozwiązanie i analiza (*Solve and Analyze*):
 - rozwiązanie,
 - rozwiązanie według kolejności tablic simpleksowych,
 - metoda graficzna;
- wyniki (*Results*):
 - rozwiązanie sumaryczne (*Solution Summary*),
 - podsumowanie ograniczeń (*Constraint Summary*),
 - raport kombinowany (*Combined Report*) stanowiący zestawienie dwóch powyższych zestawień.

P r z y k ł a d

Zakład stolarski produkuje drzwi X_1 i okna X_2 drewniane, w sztukach. Do produkcji zużywa się drewno w mb określonej szerokości desek oraz lakier w litrach. Normy jednostkowe tych materiałów oraz ograniczenia zasobów podano w tab. 2.1.1:

T a b e l a 2.1.1

Normy jednostkowe materiałów oraz ograniczenia zasobów

X_1	X_2	Znak (kierunek)	Zasób (limit)
5	3	\leq	200
2	1	\leq	50

Ź r ó d ł o: Opracowanie własne; dane umowne.

Zakład dąży do maksymalizacji zysku z produkcji. Zysk jednostkowy ze sprzedaży drzwi wynosi 150 zł, a ze sprzedaży okna 100 zł. Funkcja celu dąży do maksimum:

$$FC = 150 X_1 + 100 X_2 \rightarrow \text{maksimum}$$

X_1 – ilość produkowanych drzwi,

X_2 – ilość produkowanych okien.

Ze względów magazynowych $X_1 \leq 30, X_2 \leq 40$.

Zastosujemy moduł programowania liniowego i całkowitoliczbowego (LP–ILP) – zwany dalej LP – na określenie produkcji zakładu stolarskiego przy warunkach brzegowych: $X_1 \geq 0, X_2 \geq 0$.

Po wywołaniu modułu pojawia się okno dialogowe, do którego wpisujemy lub definiujemy opcje:

Tytuł problemu (*Problem Title*) – maksymalizacja zysków,

Ilość zmiennych (*Number of Variables*) – 2,

Ilość ograniczeń (*Number of Constraints*) – 4,

Kryterium (*Objective Criterion*) – maksymalizacja (*Maximization*),

Format danych wejściowych (*Data Entry Format*) – arkusz (*Spreadsheet Matrix Form*),

Typ zmiennej (*Default Variable Type*) – (całkowitoliczbowe) *Integer*.

Naciskamy *OK* i otrzymujemy tab. 2.1.2, do której następnie wpisujemy dane.

T a b e l a 2.1.2

Definiowanie zadania decyzyjnego

LP	MatrixFormat	Maksymalizacja zysku	2	4
Variable →	X1	X2	Direction	RHS
Maximize	150	100		
C1	5	3	<=	200
C2	2	1	<=	50
C3	1	0	<=	30
C4	0	1	<=	40
LowerBound	0	0		
UpperBound	M	M		
VariableType	Integer	Integer		

Ź r ó d ł o: Opracowanie własne.

gdzie:

Direction – znak (kierunek); według zapisu w module *LP* programu WinQSB,
RHS – zasób (limit).

Po wywołaniu z menu funkcji (*Solve and Analyze*) rozwiąż i analizuj otrzymujemy rozwiązanie optymalne funkcji celu przy danych ograniczeniach i warunkach brzegowych.

Tabela 2.1.3

Rozwiązanie sumaryczne maksymalizacji zysku

08-31-2007 17:10:00	Decision Variable	Solution Value	Unit Cost or Profit $C(j)$	Total Contribution
1	X1	5	150	750
2	X2	40	100	4000
	Objective	Function	Max =	4750

Źródło: Opracowanie własne.

Rozwiązanie sumaryczne podaje $X_1 = 5$ oraz $X_2 = 40$, a funkcja celu:

$$FC = 150 \cdot 5 + 100 \cdot 40 = 4750 \text{ zł.}$$

2.1.2. Rozszerzenie arkusza kalkulacyjnego Excel (Solver)

Według instrukcji podanej w menu (*Pomoc Excela*) instalowanie dodatku Solver przebiega następująco:

a) W menu (*Narzędzia*) klikamy na polecenie (*Dodatki*). Jeżeli dodatek Solver nie został umieszczony w oknie dialogowym (*Dodatki*), klikamy przycisk (*Przełóżaj*) i szukamy stacji dysków i folderu, gdzie znajduje się plik dodatku Solver o nazwie (*solver.xla*). Przeważnie jest to folder (*Library/Solver*). Po odnalezieniu pliku naciskamy klawisz *Enter* i następuje zainstalowanie dodatku Solver. Możemy go później wywołać z otwartego skoroszytu arkusza kalkulacyjnego. Gdy nie odnajdujemy pliku, uruchamiamy program instalacyjny Excel.

b) W oknie dialogowym (*Dodatki*) zaznaczamy pole wyboru (*Solver*). Dodatki, których pola wyboru zaznaczono w oknie dialogowym, są aktywne do czasu ich usunięcia.

Solver wykorzystuje zależność między komórkami arkusza kalkulacyjnego. Można poszukiwać maksymalnej lub minimalnej wartości jednej komórki przez zmianę wartości innej lub innych. Przykład podany w tab. 2.1.4 bazuje na opisie funkcjonalności Solvera zawartym w (*Pomocy Excel*) oraz modelu marketingowym. Model ten pokazuje wzrost sprzedaży w zależności od nakładów na reklamę. Umożliwia sprawdzenie, czy np. budżet na reklamę nie jest za mały, jak go podzielić, zwłaszcza jeśli występuje czynnik sezonowości na określony produkt w kwartałach roku.

W tabeli 2.1.4 danymi wejściowymi są: wydatki służbowe, nakłady na reklamę, sezonowość, cena produktu, koszt produktu, model prognozy sprzeda-

ży, według którego szacuje się poziom sprzedanych jednostek w przyszłych kwartałach roku.

Manewrując danymi wejściowymi (ograniczeniami) poszukujemy dogodnej dla biznesu wartości funkcji celu, w przykładzie zysku ze sprzedaży produktu.

Tabela 2.1.4

Dane wejściowe, obliczenia pośrednie oraz wyników

A	B	C	D	E	F	G	1
Kwartał roku:	1	2	3	4	Razem:		2
Sezonowość	0,8	1,2	0,9	1,1			3
						Przykłady dla kolumny E:	4
Sprzedane jednostki	2 156	4 852	2 729	4 077	13 815	=35*E3*(E11+2000)^0,5	5
Przychód ze sprzedaży [zł]	53 912	121 303	68 233	101 928	345 376	=E5*\$B\$18	6
Koszt zakupu [zł]	32 347	72 782	40 940	61 157	207 225	=E5*\$B\$19	7
Marża brutto [zł]	21 565	48 521	27 293	40 771	138 150	=E6-E7	8
							9
Wydatki służbowe [zł]	3 000	4 000	3 000	3 600	13 600		10
Reklama [zł]	3 932	11 346	5 507	9 215	30 000		11
Koszt ogólnozakładowy [zł]	5 391	12 130	6 823	10 193	34 538	=0,1*E6	12
Koszt całkowity [zł]	12 323	27 477	15 331	23 007	78 138	=SUMA(E10:E12)	13
							14
Zysk z produktów [zł]	9 242	21 044	11 963	17 764	60 013	=E8-E13	15
Rentowność sprzedaży	17%	17%	18%	17%	17%	=E15/E6	16

Cena produktu	25 zł	← B18
Koszt produktu	15 zł	B19

Źródło: Opracowanie własne na przykładzie modelu marketingowego w Excel-Solver.

c) Wyszukiwanie wartości maksymalizującej inną wartość

Poszukujemy maksymalnego zysku, zmieniając wydatki na reklamę. W menu (*Narzędzia*) klikamy polecenie Solver. W polu (*Komórka celu*) wpisujemy B15 lub zaznaczamy w arkuszu komórkę B15 (zysk w pierwszym kwartale). Wybieramy opcję (*Maks*). W polu (*Komórki zmieniane*) wpisujemy B11. Klikamy przycisk (*Rozwiąż*). Na pasku stanu pojawi się komunikat o rozpoczęciu rozwiązywania, później, że Solver znalazł rozwiązanie wielkości nakładów na reklamę w pierwszym kwartale. Następnie przywracamy wartość wyjściową w komórce B11 (Reklama – kwartał 1).

Jeśli chcemy przywrócić wszystkie początkowe ustawienia, klikamy przycisk (*Przywróć wszystko*).

d) Wyszukiwanie wartości przez zmianę kilku wartości

Można znaleźć budżet na reklamę w każdym kwartale, tak aby uzyskać największe zyski w całym roku. Tak więc Solver może służyć do określenia najlepszego sposobu podziału nakładów na kwartały.

W tym celu w menu (*Narzędzia*) klikamy polecenie Solver. W polu (*Komórka celu*) wpisujemy F15 (Zysk z produktów w zł). Wybieramy opcję maksymalizacji, a jako komórki zmieniane wpisujemy B11:E11. Klikamy przycisk (*Rozwiąż*). W ten sposób można rozwiązać problem nieliniowy, gdyż w wierszu piątym występuje potęgowanie w wyrażeniu na „Sprzedane jednostki”.

W rzeczywistości w problemach modelowania występują czynniki ograniczające do pewnych wartości. Dotyczy to zarówno komórki docelowej, jak i komórek zmienianych, dowolnych wartości powiązanych formułami (wyrażeniami).

e) Dodawanie ograniczeń

Jeśli nie ma pewności, że odpowiedź rynku na reklamę będzie odpowiednia przy danym modelu marketingowym, wprowadza się ograniczenie (więzy) poziomu wydatków na reklamę, zamiast, jak w przykładowym modelu, swobodne inwestowanie. W menu (*Narzędzia*) klikamy polecenie (*Solver*), a następnie naciskamy przycisk (*Dodaj*). W oknie (*Dodaj warunek ograniczający*) wpisujemy $F11 \leq 30\ 000$, przy czym F11 – całkowite nakłady na reklamę (tab. 2.1.4). Naciskamy przycisk (*Rozwiąż*).

f) Zmiana ograniczeń

W menu (*Narzędzia*) klikamy polecenie (*Solver*), następnie polecenie (*Zmień ograniczenie*) na $F11 \leq 35\ 000$, a następnie przyciski (*Rozwiąż*). Solver znajduje optymalne rozwiązanie w komórce F13 – całkowity zysk (tab. 2.1.5). Dzięki Solverowi możemy eksperymentować i poprawiać poprzedni wynik finansowy firmy.

Tabela 2.1.5

Znalezienie optymalnego rozwiązania po zmianie ograniczenia

A	B	C	D	E	F	G	1
Kwartał roku:	1	2	3	4	Razem:		2
Sezonowość	0,8	1,2	0,9	1,1			3
						Przykłady dla kolumny E:	4
Sprzedane jednostki	2 294	5 161	2 903	4 337	14 696	=35*E3*(E11+2000)^0.5	5
Przychód ze sprzedaży [zł]	57 350	129 037	72 583	108 427	367 396	=E5*\$B\$18	6
Koszt zakupu [zł]	34 410	77 422	43 550	65 056	220 437	=E5*\$B\$19	7
Marża brutto [zł]	22 940	51 615	29 033	43 371	146 958	=E6-E7	8
							9
Wydatki służbowe [zł]	3 000	4 000	3 000	3 600	13 600		10
Reklama [zł]	4 712	13 102	6 495	10 690	35 000		11
Koszt ogólnozakładowy [zł]	5 735	12 904	7 258	10 843	36 740	=0,1*E6	12
Koszt całkowity [zł]	13 447	30 006	16 753	25 133	85 340	=SUMA(E10:E12)	13
							14
Zysk z produktów [zł]	9 493	21 609	12 280	18 238	61 619	=E8-E13	15
Rentowność sprzedaży	17%	17%	17%	17%	17%	=E15/E6	16

Cena produktu	25 zł	← B18
Koszt produktu	15 zł	B19

Źródło: Kolejna wersja rozwiązania na podstawie tab. 2.1.4.

g) Zapisywanie modelu problemu

Jeśli klikniemy polecenie (*Zapisz*) w menu plik, ostatnie dokonania w oknie dialogowym (*Solver – Parametry*) zostaną przyłączone do arkusza i zachowane przy jego zapisywaniu. Można zdefiniować więcej niż jeden model (problem) w arkuszu, zapisując problemy oddzielnie za pomocą przycisku (*Zapisz scenariusz*). Wywołanie scenariusza następuje poprzez (*Narzędzia Scenariusze...*). Pojawia się okno dialogowe (*Menedżer scenariuszy*). Wskazujemy interesujący nas scenariusz i pojawiają się komórki i zmienne wprowadzone do danego modelu.

2.2. Analiza procesów decyzyjnych

2.2.1. Drzewo celów

Korzystając z teorii grafów, przedstawiamy strukturę problemu metodą drzewa celów. W drzewie celów wyróżniamy:

- X – węzeł pierwszego poziomu (cel główny),
- U_k – łuki (wiążą cel główny z celami cząstkowymi),
- Y_j – węzeł/element drugiego poziomu (cele cząstkowe),
- Y'_j – węzeł/element kolejnego poziomu.

Drzewa celów dzielimy na:

- a) decyzyjne drzewa celów,
- b) robocze drzewa celów.

Decyzyjne drzewa celów składają się z celów cząstkowych alternatywnych, z wyjątkiem elementów będących końcami grafu. Robocze drzewo celów zawiera tylko elementy, które wymagają rozwiązania pozycji bezpośrednio je poprzedzających. Niniejszy rozdział opracowano na podstawie procedur i wzorów zawartych w publikacji M. Lipiec-Zajchowskiej¹⁵¹.

Procedura postępowania:

1. Określenie struktury drzewa celów.
2. Opracowanie systemu oceny elementów drzewa celów.
3. Podanie wskaźników oceny.
4. Analiza wyników obliczeń.

Przykład procedury:

1) Opracowanie struktury drzewa celów do problemu „Uruchomienie w Europie Wschodniej produkcji skrzyń biegów do samochodów firmy J ”.

Etap ten powinien zawierać: opis, zatwierdzone warunki zgodne z przyjętymi zasadami w firmie macierzystej J , zdefiniowanie celów cząstkowych i elementarnych zarówno ze strony ilościowej, jak i jakościowej, określenie terminów realizacji poszczególnych celów. Elementy drzewa celów tego samego poziomu to grupa elementów, natomiast elementy drzewa celów z różnych poziomów tworzą gałąź drzewa celów.

Niech celem głównym, czyli problemem kompleksowym X_1 , będzie uruchomienie w Europie Wschodniej produkcji skrzyń biegów do samochodów firmy J .

Cele pośrednie, obejmujące problemy cząstkowe:

¹⁵¹ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne...*, op.cit., rozdz. 1.

Y_1 – wybór i zatwierdzenie lokalizacji fabryki,
 Y_2 – wybór technologii oraz wybudowanie, wyposażenie i uruchomienie wstępnej produkcji.

Zadania elementarne (Z_1 i Z_2 związane z Y_1 , natomiast Z_3, Z_4, Z_5 dotyczą Y_2):

Z_1 – analizy i prace koncepcyjne,
 Z_2 – uzgodnienie z władzami lokalnymi,
 Z_3 – wykonanie dokumentacji budowlanej,
 Z_4 – realizacja budowy i wyposażenia,
 Z_5 – testowanie technologii produkcji części i montażu.

W naszym drzewie celów wyróżniamy gałęzie, np. według wykresu 2.2.1 Z_1 - Y_1 - X_1 oraz grupy na tym samym poziomie, takie jak X, Y, Z . Opracowanie wielowariantowego drzewa celów powstaje w wyniku modelowania conceptualnego, np. w wyniku „burzy mózgów” w gronie specjalistów. Następnie po analizie dokonany jest wybór właściwego wariantu. Etapem kolejnym jest rozpisanie problemu kompleksowego na problemy cząstkowe i elementarne. Pozostaje jeszcze solidne przedyskutowanie i zatwierdzenie struktury drzewa celów.

2) Ocena elementów drzewa celów

Do oceny znaczenia kryteriów oraz oceny zadań elementarnych Z_1 - Z_5 (zob. wykres 2.2.1) przyjmijmy trzy kryteria: k_1, k_2, k_3 , którym przypiszmy wagi w skali dziesięciopunktowej. Ocena tę w punktach przeprowadzi grupa specjalistów.

Tabela 2.2.1

Ocena punktowa kryteriów „ k ” oraz zadań „ Z_i ”

Kryteria	Znaczenie punktowe kryterium Q_k	Z_1	Z_2	Σ	Z_3	Z_4	Z_5	Σ
k_1	8	6	4	10	6	3	8	17
k_2	6	4	7	11	7	2	6	15
k_3	6	5	4	9	9	4	8	21

Źródło: Opracowanie własne; dane umowne.

Sprowadźmy oceny punktowe kryteriów do wartości względnych:

$$q_k = \frac{Q_k}{\sum_{i=1}^m Q_i},$$

$$\sum q_i = 1$$

gdzie:

Q_i – ocena punktowa znaczenia kryterium dla danego zadania elementarnego i ,

$i = 1, 2, \dots, m$.

Ocena normatywna elementu zadań drzewa celów w tym przykładzie: $i = 1, 2, 3 \dots l$ przy k -tym kryterium:

$$Z_k^l = \frac{S_k^l}{\sum_{j=1}^m S_k^j},$$

gdzie:

$$k = 1, \dots, m$$

$$\sum_{k=1}^m Z_k^l = 1,$$

$$j = 1, 2, \dots, n.$$

Ocena normatywna elementów Z_1 i Z_2 związanych z elementem Y_1 .

T a b e l a 2.2.2

Ocena normatywna elementów „ Z_1 ” i „ Z_2 ”

Kryteria	q_k	Z_1	Z_2	Σ
k_1	0,4	0,6	0,4	1,0
k_2	0,3	0,36	0,64	1,0
k_3	0,3	0,56	0,44	1,0
r_l		0,516	0,484	1,0

r_l – wskaźnik względnej ważności l -tego elementu w realizacji elementu nadrzędnego, czyli Y_1 określony ze wzoru:

$$r_l = \sum_{i=1}^m q_i \cdot Z_i^l.$$

Dla elementów jednego poziomu : $\sum r_l = 1, \quad l = 1, 2, \dots, n.$

Ź r ó d ł o: Opracowanie własne; obliczenia pomocnicze.

Ocena normatywna elementów Z_3, Z_4 i Z_5 związanych z elementem Y_2 .

T a b e l a 2.2.3

Ocena normatywna elementów „ Z_3 ”, „ Z_4 ”, „ Z_5 ”

Kryteria	q_k	Z_3	Z_4	Z_5	Σ
k_1	0,4	0,35	0,18	0,47	1,0
k_2	0,3	0,47	0,13	0,40	1,0
k_3	0,3	0,43	0,19	0,38	1,0
r_l		0,41	0,17	0,42	1,0

Ź r ó d ł o: Opracowanie własne; obliczenia pomocnicze.

Ocena normatywna elementów Y_1, Y_2 związanych z celem głównym X_1 (wcześniejsza ocena punktowa ekspertów).

T a b e l a 2.2.4

Ocena punktowa elementów „ Y_1 ”, „ Y_2 ” według kryteriów

Kryteria	Znaczenie punktowe kryterium Q_k	Y_1	Y_2	Σ
k_1	8	7	5	12
k_2	6	6	8	14
k_3	6	5	7	12

Ź r ó d ł o: Opracowanie własne; dane umowne.

Tabela 2.2.5

Ocena normatywna elementów „Y₁”, „Y₂” według kryteriów

Kryteria	q_k	Y_1	Y_2	Σ
k_1	0,4	0,583	0,417	1,0
k_2	0,3	0,429	0,571	1,0
k_3	0,3	0,417	0,583	1,0
r_l		0,487	0,513	1,0

Źródło: Opracowanie własne; obliczenia pomocnicze.

3) Ogólny wskaźnik względnej ważności l -tego elementu z uwagi na realizację celu głównego.

$$R_l = \prod_{\alpha} r_{\alpha},$$

gdzie:

$$\sum R_l = 1,$$

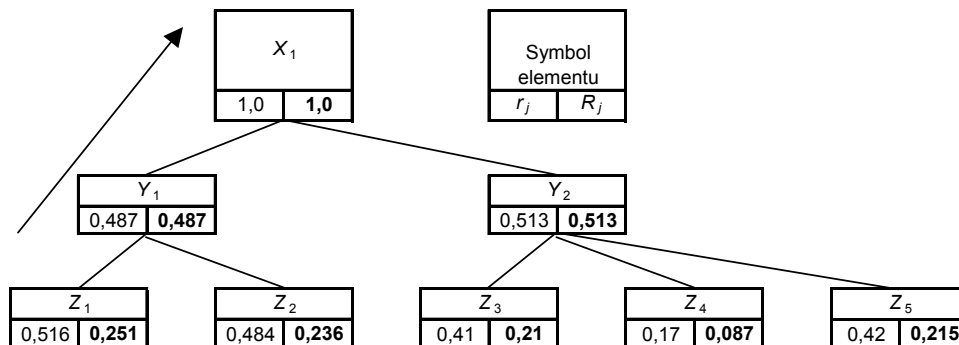
 α – indeksy elementów tworzących wspólną gałąź, np. (Y_1 – X_1):

Przykładowo:

$$R_{y_1} = \prod_{\alpha=y_1}^{x_1} r_{\alpha} = r_{y_1} \cdot r_{x_1} = 0,487 \cdot 1 = 0,487.$$

$$R_{z_1} = \prod_{\alpha=z_1}^{x_1} r_{\alpha} = r_{z_1} \cdot r_{y_1} \cdot r_{x_1} = 0,516 \cdot 0,487 \cdot 1 = 0,251.$$

Dla pozostałych analogicznie.



Źródło: Opracowanie własne w Excelu.

Wykres 2.2.1. Wyniki obliczeń drzewa celów

4) Analiza wyników

Jeżeli ogólny wskaźnik względnej ważności jest wysoki ($R_l > 1,2 R_0$), to elementy mają istotny wpływ na osiągnięcie celu głównego. Natomiast jeśli wartość jest przeciętna ($1,2 \cdot R_0 > R_l > 0,6 \cdot R_0$), to w planie realizacji elementy powinny być uwzględnione

na ogólnych zasadach. Jeśli wartość R_i jest niska ($R_i < 0,6 \cdot R_0$), to nie muszą one występować w planie realizacji danego przedsięwzięcia¹⁵².

W najniższym poziomie (w naszym trzecim przykładzie) R_0 jako średnia wartość ogólnego wskaźnika względnej ważności wynosi:

$$R_0 = (0,251 + 0,236 + 0,21 + 0,0087 + 0,215)/5 = 0,1998.$$

Zatem dla elementu Z_4 : $R_{z_4} = 0,6 \cdot 0,1998 = 0,1188$. Ponieważ $R_{z_4} = 0,087 < 0,1188$, to może być wyeliminowane z planu realizacji przedsięwzięcia.

2.2.2. Drzewo decyzyjne jako narzędzie do podejmowania decyzji w warunkach niepewności

Drzewo decyzyjne jest obrazem etapów oceny i wyboru w procesie rozwiązywania problemu decyzyjnego¹⁵³. Komputerowy pakiet WinQSB jest interakcyjnym systemem wspomagającym podejmowanie decyzji i zawiera moduł (*Decision Analysis – DA*) narzędzia analizy decyzji.

P r z y k ł a d

Na początku bieżącego roku wystąpiła korzystna koniunktura na rynku budowlanym. Szef firmy dekarskiej wykonującej pokrycia dachowe wobec licznych zamówień rozważa podjęcie decyzji:

- czy powiększyć znacznie swoją ekipę budowlaną i zrealizować zamówienia jeszcze w tym roku?
 - czy poczekać z realizacją do roku następnego?
- Księgowy oszacował warianty.

T a b e l a 2.2.6

Warianty

Zatrudnienie	Koniunktura	Zysk/strata [zł]
Powiększyć (<i>p</i>)	dobra <i>d</i>	120 000
	„załamanie” rynku <i>z</i>	-40 000
Utrzymać (<i>u</i>)	dobra <i>d</i>	85 000
	„załamanie” rynku <i>z</i>	60 000

Ź r ó d ł o: Opracowanie własne; dane umowne.

Prawdopodobieństwo recesji *z* wynosi 0,25, koniunktury *r* 0,75. Decyzja początkowa o powiększeniu lub utrzymaniu stanu ekipy dekarskiej należy do decydenta, pozostałe zależne są od losu zwanego tu „stanem natury”.

¹⁵² Ibidem, s. 11.

¹⁵³ Ibidem, rozdz. 1.2.

Wykorzystajmy do rozwiązania tego problemu moduł pakietu WinQSB (*Decision Analysis*). Proponujemy wybrać z tego modułu opcję (*Decision Tree Analysis*) analiza drzewa decyzyjnego. Na początku nadamy nazwę problemowi: „Zatrudnienie w firmie” oraz określimy liczbę wszystkich węzłów (*Number of Nodes/Events*) – 7. Liczba węzłów problemu decyzyjnego wynika z tabeli 2.2.6, a graficznie z wykresu 2.2.2. Definicję problemu w formie tekstowej z pliku modułu (*DA*) przedstawiono w tab. 2.2.7.

Tabela 2.2.7

Definicja sieci zależności między siedmioma węzłami

DA	Zatrudnienie w firmie	DT	Immediate Following Node	Node Payoff (+ profit, cost)	Probability (if available)
Node/Event Number	Node Name or Description	Node Type (enter <i>D</i> or <i>C</i>)			
1	Event1	<i>D</i>	2,3		
2	Event2	<i>C</i>	4,5		
3	Event3	<i>C</i>	6,7		
4	Event4	<i>C</i>		120 000	0.75
5	Event5	<i>C</i>		-40 000	0.25
6	Event6	<i>C</i>		85 000	0.75
7	Event7	<i>C</i>		60 000	0.25

Źródło: Opracowanie własne; moduł (*DA*) programu WinQSB; prawdopodobieństwo zapisane z kropką dziesiętną.

gdzie:

np. event1 – nazwa węzła (wydarzenia) nadana przez moduł (*DA*),

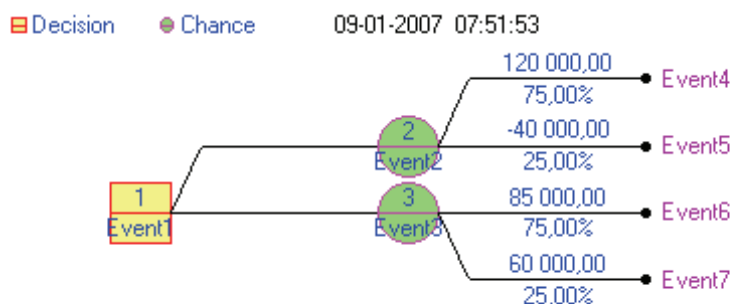
D lub *C* – typ węzła (decyzyjny, zmieniany przez program),

Immediate Following Node – węzły następne,

profit, cost – zysk, strata,

probability – prawdopodobieństwo.

Wyniki analizy w formie graficznej przedstawiające sieć zależności pobrane z modułu (*DA*) niniejszego problemu przedstawia wykres 2.2.2:



Źródło: Opracowanie własne moduł (*DA*); kółko górne węzeł 2., dolne węzeł 3.

Wykres 2.2.2. Sieć zależności problemu decyzyjnego

Na wykresie moduł podaje prawdopodobieństwo w procentach (tab. 2.2.7) oraz zyski/straty. Z tabeli 2.2.8 wynika, że najlepsza jest decyzja powiększenia obsady ekipy dekararskiej jeszcze w tym roku i wówczas wartość oczekiwanego zysku wyniesie 80 000 zł.

Tabela 2.2.8

*Wartości określone na podstawie drzewa decyzyjnego (wykres 2.2.2).
Analiza drzewa decyzyjnego „Zatrudnienie w firmie”*

9.01.07	Node/Event	Type: D, C	Expected value [zł]	Decision
1	Event1	D	80 000	Event2
2	Event2	C	80 000	
3	Event3	C	78 750	
4	Event4	C	0	
5	Event5	C	0	
6	Event6	C	0	
7	Event7	C	0	
Overall	Expected	Value =	80 000	

Źródło: Opracowanie własne z zastosowaniem modułu (DA).

Sprawdzenie wartości węzłów 1,2,3:

$120\,000 \cdot 0,75 = 90\,000$ zł, węzły 2–4,

$-40\,000 \cdot 0,25 = -10\,000$ zł, węzły 2–5.

Suma: $90\,000 - 10\,000 = 80\,000$ zł.

$85\,000 \cdot 0,75 = 63\,750$ zł,

$60\,000 \cdot 0,25 = 15\,000$ zł.

Suma: 78 750 zł.

Tak więc ścieżka od węzła 1. do 2. i następnie sumarycznie dla węzła 4. i 5. daje większy zysk. Potwierdza to moduł DA, wskazując na węzeł 2. (Event2).

Warto zauważyć, że obliczenie efektu węzła pierwszego jest wyborem spośród średnich ważonych efektów oraz prawdopodobieństw dla węzłów drugiego i trzeciego.

2.2.3. Przedstawienie sytuacji decyzyjnej za pomocą macierzy

Macierz decyzyjna obrazuje sytuację decyzyjną m.in. kształtowania się popytu na rynku na określony produkt. Tym produktem mogą być iglaki w donicach w supermarkecie ogrodniczym.

Przykład

Obserwacje przez kolejne 100 dni wykazały, że dni, w których nie sprzedano żadnego iglaka, było pięć. Natomiast liczbę dni ze sprzedażą w przedziale „od do” oraz oszacowaniem prawdopodobieństwa jako częstości ich występowania podano w tab. 2.2.9.

Tabela 2.2.9

Częstości występowania okresów

Popyt (od – do)	0	1–5	6–10	11–15	16–20	21–25
Liczba dni sprzedaży	5	10	15	30	25	15
Prawdopodobieństwo	0,05	0,1	0,15	0,3	0,25	0,15

Źródło: Opracowanie własne; dane umowne.

Przyjmijmy, że koszt produkcji iglaka wynosi 3 zł, a cena sprzedaży 5 zł. Zatem zysk wynosi 2 zł. Rotacja dostawy jest tygodniowa (stare niesprzedane iglaki są zabierane). Iglaki na sali supermarketu wymagają podlewania. Dzienny koszt wody i obsługi wynosi 0,1 zł /sztukę, co zmniejsza zysk w zależności od dni magazynowania. Iglaki oferowane do sprzedaży układane są na dostawianych stolikach w przejściu głównym. Możliwości magazynowania iglaków według decydenta wynoszą: 0, 5, 10, 15, 20 sztuk.

Kierownika supermarketu interesuje, jaki powinien być zapas iglaków, aby przy określonym prawdopodobieństwie straty były najmniejsze. Sformułujmy macierz decyzyjną¹⁵⁴.

Tabela 2.2.10

Macierz potencjalnych zysków ze sprzedaży iglaków

Popyt (średnio)		0	3	8	13	18	23	Wartość oczekiwania T_e
Prawdopodobieństwo		0,05	0,1	0,15	0,3	0,25	0,15	
Zapas [szt.]	0	0	0	0	0	0	0	0
	5	-2	4,4	4,4	4,4	4,4	4,4	4,08
	10	-4	3,2	15,2	15,2	15,2	15,2	13,04
	15	-6	1,2	13,2	25,2	25,2	25,2	19,44
	20	-8	-0,8	11,2	23,2	35,2	35,2	22,24

Źródło: Opracowanie własne; dane umowne.

Zapas stanowi ograniczenie możliwości magazynowania w cyklach tygodniowych. Elementami macierzy są w naszym przykładzie zyski obliczone jako średnia wartość popytu, przy założeniu średnio czterech dni magazynowania przy rotacji tygodniowej.

Na przykład przy zapasie pięciu sztuk i zerowym popycie strata wynosi: $5 \cdot 4 \cdot 0,1 = 2$ zł. Przy zapasie tym samym, lecz średniej sprzedaży trzech sztuk z dwuzłotowym zyskiem każda i koszcie magazynowania (podlewania) 0,1 zł/dzień przy czterech dniach magazynowania zysk wynosi:

$$3 \cdot 2 - 2 \cdot 4 \cdot 0,1 = 4,4 \text{ zł.}$$

Wartość oczekiwana T_e stanowi sumę iloczynów elementów macierzy (zysków) i prawdopodobieństw. Wartość oczekiwana zysku, przy zapasie np. pięciu sztuk, jest następująca:

$$T_e = -2 \cdot 0,05 + 4,4 \cdot 0,1 + 4,4 \cdot 0,15 + 4,4 \cdot 0,3 + 4,4 \cdot 0,25 + 4,4 \cdot 0,15 = 4,08 \text{ zł.}$$

Z tabeli 2.2.10 wynika, że składowanie 20 sztuk iglaków tygodniowo jest najlepszą decyzją przy obecnej możliwości powierzchni magazynowania.

¹⁵⁴ Ibidem, analogia do tablicy 1.13.

2.2.4. Reguły decyzyjne, gdy nieznane jest prawdopodobieństwo

Jeśli założymy, że w rozwiązywaniu problemu wystąpią tylko nieprzewidywalne sytuacje, możemy to określić jako „grę z naturą”, opartą na regułach decyzyjnych według kryteriów:

- Walda (reguła maksimum),
- Hurwicza (reguła współczynnika ostrożności),
- Bayesa (reguła największej przeciętnej wygranej),
- Savage (reguła minimalizacji oczekiwanych strat).

Przykład

Developer zamierza w okresie jesieni i zimy danego roku kontynuować budowę osiedla domów jednorodzinnych. Zdaje sobie sprawę, że pomyślność realizacji jego przedsięwzięcia oraz osiągnięcie zysków zależą od niepewności związanych z pogodą. Rozważył warianty kontynuacji inwestycji oraz „stany natury” – sytuacje pogodowe.

Warianty kontynuacji inwestycji:

B_1 – zamykanie etapów związanych z elewacją, uporządkowaniem terenu, nasadzeniami, ogrodzeniem, robotami instalacyjnymi na dachu,

B_2 – (B_1 + podjęcie budowy drogi osiedlowej),

B_3 – (B_1 + wykonanie fundamentów pod pięć kolejnych domów jednorodzinnych),

B_4 – (B_1 + wybudowanie stanu surowego pięciu domów jednorodzinnych),

B_5 – (B_1 + wykonanie trafostacji),

B_6 – (B_1 + odwodnienie terenu).

Sytuacje pogodowe:

P_1 – w miarę sucha i ciepła pogoda w okresie jesienno-zimowym,

P_2 – deszczowa jesień i łagodna bezśnieżna zima,

P_3 – sucha i ciepła jesień, lecz śnieżna zima,

P_4 – mokra jesień i mroźna zima.

Specjalista określił przewidywane zyski z realizacji przedsięwzięcia w mln zł w zależności od „stanów natury” (sytuacji pogodowej).

Tabela 2.2.11

Przewidywane zyski

Kontynuacja/Stany pogody	P_1	P_2	P_3	P_4	Min	Max	Średnia
B_1	1	0,5	0,6	0,25	0,25	1	0,588
B_2	1,5	0,8	0,7	0,3	0,3	1,5	0,825
B_3	1,2	0,6	0,5	0,4	0,4	1,2	0,675
B_4	3	1,5	1	0,8	0,8	3	1,575
B_5	1,3	0,6	0,4	0,3	0,3	1,3	0,625
B_6	1,1	0,5	0,4	0,2	0,2	1,1	0,550

Źródło: Opracowanie własne; dane umowne.

Określmy cztery kryteria na podstawie publikacji M. Lipiec-Zajchowskiej¹⁵⁵:

– kryterium Walda – wybieramy z tab. 2.2.11 minimalną wartość zysku (dla każdego wariantu inwestycyjnego); według tego kryterium należy wybrać wariant B_4 kontynuacji inwestycji, gdyż ma największy zysk spośród minimalnych,

– kryterium Hurwicza – wybór współczynnika ostrożności ryzyka $0 \leq r \leq 1$ dla wyliczenia zysku przeciętnego:

$$V(i) = r \cdot \min \alpha_{ij} + (1 - r) \max \alpha_{ij},$$

gdzie:

α_{ij} – wartość zysku z kontynuacji przedsięwzięcia według wariantu i przy sytuacji pogodowej (stanie natury) j .

Przyjmijmy $r = 0,8$ (nieduże ryzyko, bo wartość r jest wysoka) i obliczmy przeciętny zysk każdego z wariantów:

$$V(B_1) = 0,8 \cdot 0,25 + (1 - 0,8) \cdot 1 = 0,4 \text{ mln zł},$$

$$V(B_2) = 0,8 \cdot 0,3 + (1 - 0,8) \cdot 1,5 = 0,54 \text{ mln zł},$$

$$V(B_3) = 0,8 \cdot 0,4 + (1 - 0,8) \cdot 1,2 = 0,56 \text{ mln zł},$$

$$V(B_4) = 0,8 \cdot 0,8 + (1 - 0,8) \cdot 3 = 1,24 \text{ mln zł},$$

$$V(B_5) = 0,8 \cdot 0,3 + (1 - 0,8) \cdot 1,3 = 0,5 \text{ mln zł},$$

$$V(B_6) = 0,8 \cdot 0,2 + (1 - 0,8) \cdot 1,1 = 0,38 \text{ mln zł}.$$

Zatem powinniśmy wybrać wariant B_4 , dla którego $V(B_4) = 1,24$ mln zł.

– kryterium Bayesa – według średniej arytmetycznej wartości zysku, przy założeniu jednakowego prawdopodobieństwa zaistnienia dowolnego spośród „stanów natury” P_1, P_2, P_3, P_4 . Wyliczone średnie arytmetyczne wpisano w tab. 2.2.11. Najwyższą wartość według tego kryterium ma również wariant B_4 .

– kryterium Savage – minimum maksymalnych strat określonego wariantu.

Rozpatrujemy tab. 2.2.11 i bierzemy pod uwagę maksymalne straty, jakie możemy ponieść w wyniku przyjęcia wariantu gorszego niż najlepszy z możliwych danego „stanu natury”, np. dla wariantu ($B_1 - P_1$) to $1 - 3 = 2$.

T a b e l a 2.2.12

Straty w stosunku do najlepszego wariantu

Kontynuacja/Stany pogody	P_1	P_2	P_3	P_4	Max
B_1	-2	-1	-0,4	-0,55	-2
B_2	-1,5	-0,7	-0,3	-0,5	-1,5
B_3	-1,8	-0,9	-0,5	-0,4	-1,8
B_4	0	0	0	0	0
B_5	-1,7	-0,9	-0,6	-0,5	-1,7
B_6	-1,9	-1	-0,6	-0,6	-1,9

Ź r ó d ł o: Opracowanie własne; obliczenia pomocnicze.

¹⁵⁵ Ibidem, rozdz. 1.4.

Po przeanalizowaniu kolumny Max w tab. 2.2.12 stwierdzamy, że największa strata występuje dla wariantu B_1 , natomiast najmniejsza dla wariantu B_4 . W podsumowaniu wyboru spośród sześciu wariantów według czterech kryteriów możliwości zaistnienia przyjętych stanów natury (P_1 - P_4) otrzymaliśmy propozycję wariantu B_4 jako najlepszą do realizacji.

2.2.5. Bilans przepływów *input – output*

Do analizy powiązań techniczno-ekonomicznych, np. pomiędzy gałęziami gospodarki, stosowane są metody *input – output* (wejścia – wyjścia). Wykonywany bilans przepływów międzygałęziowych służy do tworzenia i podziału produktu globalnego. Zamieszczone tu wzory zaczerpnięto z publikacji M. Lipiec-Zajchowskiej¹⁵⁶. Bilans jest układem równań liniowych-bilansowych zapisanych jako:

$$X_i = x_{i1} + x_{i2} + \dots + x_{ij} + \dots + x_{in} + y_i,$$

gdzie:

X_i – wartość produkcji globalnej gałęzi (sektora) i ,

x_{ij} – wartość produkcji i -tej gałęzi, która zużywana jest na cele produkcyjne gałęzi j -tej (zużycia produkcyjne),

y_i – produkt końcowy gałęzi i .

Po wprowadzeniu współczynników technologicznych $\alpha = x_{ij}/X_j$, gdzie X_j jest sumą produkcji zużywaną przez gałęzie, układ równań można zapisać:

$$X_1 = \alpha_{11} \cdot X_1 + \alpha_{12} \cdot X_2 + \dots + \alpha_{1n} \cdot X_n + y_1,$$

$$\dots$$

$$X_n = \alpha_{n1} \cdot X_1 + \alpha_{n2} \cdot X_2 + \dots + \alpha_{nn} \cdot X_n + y_n.$$

Założenie: znane y_i oraz α_{ij} , nieznane X_i .

Po przekształceniu powstaje układ równań zwany układem Leontiewa:

$$(1 - \alpha_{11}) \cdot X_1 + \alpha_{12} \cdot X_2 + \dots - \alpha_{1n} \cdot X_n = y_1,$$

$$- \alpha_{21} \cdot X_1 + (1 - \alpha_{22}) \cdot X_2 + \dots - \alpha_{2n} \cdot X_n = y_2,$$

$$\dots$$

$$- \alpha_{n1} \cdot X_1 - \alpha_{n2} \cdot X_2 - \dots + (1 - \alpha_{nn}) \cdot X_n = y_n.$$

W zapisie macierzowo-wektorowym:

¹⁵⁶ Ibidem, rozdz. 1.5.

$$\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{y}, \quad (2.2.1)$$

gdzie:

\mathbf{I} – macierz jednostkowa,

$(\mathbf{I} - \mathbf{A})^{-1}$ – macierz odwrotna do macierzy $(\mathbf{I} - \mathbf{A})$, określanej jako macierz strukturalna,

\mathbf{A} – macierz współczynników technologicznych α_{ij} zwana macierzą Leontiewa, gdzie:
 $i, j = 1, 2, \dots, n$,

\mathbf{y} – wektor produkcji końcowej o elementach y_i ($i = 1, 2, \dots, n$),

\mathbf{X} – wektor produkcji globalnej o elementach X_i ($i = 1, 2, \dots, n$).

Analizę przepływów międzygałęziowych prowadzi się w jednostkach fizycznych lub pieniężnych.

Odpowiednio przekształcony układ Leontiewa zastosowano do analizy nakładów inwestycyjnych przy danych przyrostach produkcji:

$$\Delta \mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1} \Delta \mathbf{y}.$$

Założenia:

$\mathbf{A} = [\alpha_{ij}]$ ($i = 1, 2, \dots, n$) macierz współczynników technologicznych podana przykładowo w zł,

$\Delta \mathbf{X}$ – wektor kolumnowy przyrostów produkcji,

$\Delta \mathbf{y}$ – wektor kolumnowy przyrostów produktu końcowego.

Wprowadźmy współczynnik kapitałochłonności przyrostu produkcji w gałęzi i -tej:

$$m_i = J / \Delta x_i,$$

gdzie:

J – nakłady inwestycyjne niezbędne do uzyskania przyrostów produkcji.

Przepływy można badać również między przedsiębiorstwami w ramach sektora i wówczas wektor produkcji końcowej dotyczy produkcji, która wychodzi poza sektor. Ponadto tablica przepływów międzyproduktowych może służyć do rozdysponowania produkcji przedsiębiorstwa do odbiorców.

P r z y k ł a d

Ciepłownia wytwarza dwa produkty dwoma sposobami.

Produkty:

P_1 – ciepła woda na potrzeby mieszkańców osiedla mieszkaniowego oraz jako medium w instalacji CO w tys. m³,

P_2 – ciepło grzewcze do centralnego ogrzewania domów na osiedlu mieszkaniowym i do ogrzewania wody w tys. dżuli, przy czym 1 kaloria = 4,1868 J (dżula).

W ciepłowni pracują dwa zestawy pieców grzewczych o technologii:

S_1 – spalanie koksu,

S_2 – spalanie węgla brunatnego.

Przyjmijmy, że do wytworzenia określonego produktu potrzebny jest drugi produkt, ale nie on sam. Macierz współczynników technologicznych (normy) w tys. zł podano w tab. 2.2.13.

Tabela 2.2.13

Macierz współczynników technologicznych

Produkt	S ₁		S ₂	
	P ₁	P ₂	P ₁	P ₂
P ₁	0	0,1	0	0,2
P ₂	0,2	0	0,4	0

Źródło: Opracowanie własne; dane umowne.

Równania bilansowe produkcji ma dwa sposoby wytwarzania¹⁵⁷:

$$\begin{aligned} 0 \cdot P_{11}^1 + 0,1 \cdot P_{12}^1 + 0 \cdot P_{11}^2 + 0,2 \cdot P_{12}^2 + y_1 &= P_1, \\ 0,2 \cdot P_{21}^1 + 0 \cdot P_{22}^1 + 0,4 \cdot P_{21}^2 + 0 \cdot P_{22}^2 + y_2 &= P_2. \end{aligned}$$

Wiemy, że:

$$\begin{aligned} P_1^1 + P_1^2 &= P_1, & P_2^1 + P_2^2 &= P_2, \\ P_{12}^1 &= P_1^1, & P_{12}^2 &= P_1^2, & P_{21}^1 &= P_2^1, & P_{21}^2 &= P_2^2, \\ P_1^1 + P_1^2 - 0,1 \cdot P_1^1 - 0,2 \cdot P_1^2 &= y_1, \\ P_2^1 + P_2^2 - 0,2 \cdot P_2^1 - 0,4 \cdot P_2^2 &= y_2, \end{aligned}$$

$$\begin{aligned} 0,9 \cdot P_1^1 + 0,8 \cdot P_1^2 &= y_1, \\ 0,8 \cdot P_2^1 + 0,6 \cdot P_2^2 &= y_2. \end{aligned}$$

Ustalone wielkości planowane:

- produkty końcowe ($y_1 = 5, y_2 = 2$),
- wskaźniki produkcji danego produktu określonym sposobem:

$$P_1^1 = 1/4 \cdot P_1, \quad P_1^2 = 3/4 \cdot P_1, \quad P_2^1 = 2/5 \cdot P_2, \quad P_2^2 = 3/5 \cdot P_2.$$

Biorąc to pod uwagę, układ równań jest następujący:

$$\begin{aligned} 0,9 \cdot 0,25 \cdot P_1 + 0,8 \cdot 0,75 \cdot P_1 &= 5, & 0,825 \cdot P_1 &= 5, & P_1 &= 6,06 \text{ tys. m}^3, \\ 0,8 \cdot 0,40 \cdot P_2 + 0,6 \cdot 0,60 \cdot P_2 &= 2, & 0,68 \cdot P_2 &= 2, & P_2 &= 2,94 \text{ tys. kJ.} \end{aligned}$$

Przykład

Dana jest tablica przepływów międzygałęziowych (między gałęzią i a j). Oblicz przyrost produktów końcowych w gałęzi 1. i 2., jeżeli przyrost produktu globalnego wyniesie: $\Delta X_1 = 20; \Delta X_2 = 30$. Określ ponadto macierze $\mathbf{A}, \mathbf{I}, (\mathbf{I}-\mathbf{A}), (\mathbf{I}-\mathbf{A})^{-1}$ oraz układ równań modelu Leontiewa.

¹⁵⁷ Ibidem, rozdz. 1.5.2.

Tabela 2.2.14

Tablica przepływów międzygałęziowych

ji	x_{ij}		y_j	X_j
	1	2		
1	30	60	10	100
2	80	100	20	200

Źródło: Opracowanie własne; dane umowne.

Procedura:

– macierz współczynników technologicznych określona według wzoru: $a_{ij} = x_{ij}/X_j$:

$$\mathbf{A} = \begin{bmatrix} 0,3 & 0,6 \\ 0,4 & 0,5 \end{bmatrix},$$

– macierz jednostkowa:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

– macierz strukturalna (odejmowanie odpowiadających sobie elementów):

$$\mathbf{I} - \mathbf{A} = \begin{bmatrix} 0,7 & -0,6 \\ -0,6 & 0,5 \end{bmatrix},$$

– macierz odwrotna do macierzy strukturalnej: $(\mathbf{I} - \mathbf{A})^{-1} = 1/\det(\mathbf{I} - \mathbf{A}) \cdot (\mathbf{D}_{ij})^T$:

$$\det(\mathbf{I} - \mathbf{A}) = (0,7 \cdot 0,5) - [(-0,6) \cdot (-0,6)] = 0,35 - 0,36 = -0,01.$$

Elementy macierzy dopełnień algebraicznych \mathbf{D}_{ij} obliczamy według:

$$(-1)^{i+j} \cdot d_{ij},$$

przy czym d_{ij} otrzymujemy przez pominięcie kolumny j i wiersza i elementu obliczanego i wstawienie w to miejsce elementu, który pozostaje, np. 0,5.

Przed transpozycją:

$$(\mathbf{I} - \mathbf{A})^{-1} = (1/(-0,01)) \begin{bmatrix} 0,5 & 0,6 \\ 0,6 & 0,7 \end{bmatrix}^T = \begin{bmatrix} -50 & -60 \\ -60 & -70 \end{bmatrix}.$$

Po transpozycji macierz nie ulega zmianie:

$$(\mathbf{I} - \mathbf{A})^{-1} = \begin{bmatrix} -50 & -60 \\ -60 & -70 \end{bmatrix}.$$

Można też skorzystać z funkcji Excela do obliczenia macierzy odwrotnej [MACIERZ.ODW(...)]. W miejsce kropek wpisujemy obszar zajęty w arkuszu kalkulacyjnym

przez macierz $(\mathbf{I} - \mathbf{A})$, wstawiamy podaną funkcję do pierwszej komórki z lewego górnego rogu, po czym naciskamy jednocześnie trzy klawisze: *Shift+Ctrl+Enter* (prawy).

Wynik jest zgodny z uzyskanym po transpozycji.

Układ równań – model Leontiewa:

$$(1 - \alpha_{11}) \cdot X_1 + \alpha_{12} \cdot X_2 = y_1,$$

$$-\alpha_{21} \cdot X_1 + (1 - \alpha_{12}) \cdot X_2 = y_2,$$

$$(\mathbf{I} - \mathbf{A}) \mathbf{X} = \mathbf{y},$$

– zastosowane współczynniki według macierzy strukturalnej:

$$(\mathbf{I} - \mathbf{A}) = \begin{bmatrix} 0,7 & -0,6 \\ -0,6 & 0,5 \end{bmatrix},$$

$$0,7 \cdot X_1 - 0,6 \cdot X_2 = 10,$$

$$-0,6 \cdot X_1 + 0,5 \cdot X_2 = 20.$$

Przyrost produktu końcowego obliczamy według wzoru:

$$\Delta \mathbf{y} = (\mathbf{I} - \mathbf{A}) \cdot \Delta \mathbf{X},$$

$$\Delta \mathbf{y} = \begin{bmatrix} 0,7 & -0,6 \\ -0,6 & 0,5 \end{bmatrix} \cdot \begin{bmatrix} 20 \\ 30 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

– wielkość produktu końcowego: $y_j' = y_j + \Delta y$

$$y_1' = 10 + 2 = 12,$$

$$y_2' = 20 + 3 = 23.$$

2.2.6. Pytania

1. Wymień elementy drzewa celów.
2. Objasnij na przykładzie sposób obliczania ogólnego wskaźnika ważności l -tego elementu.
3. Podaj procedurę korzystania z modułu (*DA*) programu WinQSB do rozwiązywania problemu za pomocą drzewa decyzyjnego.
4. Opisz kryteria Walda, Hurwicza, Bayesa, Savage, które są stosowane, gdy nieznanne jest prawdopodobieństwo wystąpienia określonej sytuacji.
5. Napisz i objasnij układ równań Leontiewa.
6. Wyjaśni elementy zapisu macierzowego do określenia przyrostów nakładów inwestycyjnych.

2.2.7. Zadania

1. Narysuj strukturę drzewa problemu kompleksowego – zagospodarowanie działki budowlanej po wybudowaniu domu. Wyodrębnij problemy cząstkowe i elementarne.

2. Podane są oceny punktowe kryteriów zdarzeń elementarnych. Sprwadź te oceny do wartości względnych oraz opracuj oceny normatywne.

Tabela 2.2.15

Oceny punktowe zdarzeń

Kryteria	Znaczenie kryterium Q_k	Z_1	Z_2	Σ	Z_3	Z_4	Σ
k_1	7	5	4	9	6	5	11
k_2	8	6	7	13	7	3	10
k_3	6	5	3	8	8	4	12

Źródło: Opracowanie własne; dane umowne.

3. Oblicz czasy oczekiwania, znając popyt tygodniowy („od do”) w sklepie odzieżowym na koszule, oszacowane prawdopodobieństwo sprzedaży z danego przedziału, liczbę oraz możliwości magazynowania (zapasu). Oblicz także zyski ze sprzedaży koszul, przyjmując cenę jednostkową około 30 zł.

Tabela 2.2.16

Ramka do obliczenia zysków oraz czasów oczekiwania

Popyt (średnio)		0	5	10	15	20	T_e
Prawdopodobieństwo		0,05	0,40	0,30	0,15	0,10	
Zapas [szt.]	0						
	5						
	10						
	15						
	20						

Źródło: Opracowanie własne; dane umowne.

4. Dane są warianty inwestycyjne i „stany natury” oraz przewidywane zyski z realizacji wariantów oszacowane przez specjalistę. Wybierz wariant, rozpatrując według kryteriów Walda, Hurwicza, Bayesa, Savage. Przyjmij brakujące parametry.

Tabela 2.2.17

Warianty inwestycyjne i „stany natury”

Wariant/Stan natury	N_1	N_2	N_3
W_1	2	1	1,5
W_2	3	2	1,8
W_3	1,5	1	0,5
W_4	3	1,5	2
W_5	2,5	1	0,5
W_6	1	0,5	0,4

Źródło: Opracowanie własne; dane umowne.

5. Dana jest tablica przepływów międzygałęziowych. Określ:
- macierz współczynników technologicznych,
 - macierz strukturalną,
 - macierz odwrotną do macierzy strukturalnej,
 - układ równań modelu Leontiewa.

Tabela 2.2.18

Przepływy międzygałęziowe

j/i	x_{ij}			y_j	X_j
	1	2	3		
1	15	25	10	10	60
2	30	20	15	15	80
3	20	10	5	5	40

Źródło: Opracowanie własne; dane umowne.

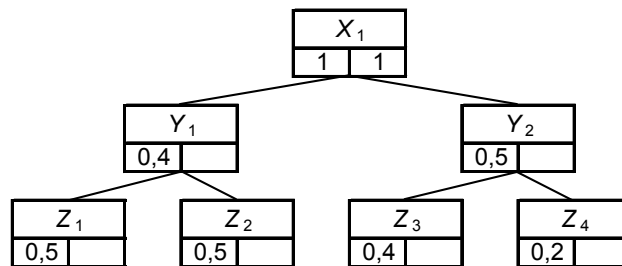
gdzie:

 y_j – produkt końcowy, X_j – produkt globalny.

6. Dana jest macierz Leontiewa oraz wektor przyrostów produkcji końcowej. Oblicz przyrosty produkcji globalnej, stosując rachunek macierzowy.

$$(\mathbf{I} - \mathbf{A}) = \begin{bmatrix} 0,6 & 0,3 & 0,4 \\ 0,2 & 0,7 & 0,5 \\ 0,5 & 0,1 & 0,3 \end{bmatrix}, \quad (\Delta \mathbf{y}) = \begin{bmatrix} 20 \\ 15 \\ 30 \end{bmatrix}.$$

7. Na wykresie podano wskaźniki względnej ważności i -tego elementu. Oblicz:
- wskaźniki realizacji elementu nadrzędnego (o poziom wyżej),
 - średnią wartość ogólnego wskaźnika względnej ważności poziomu z elementami Z_j .



Ź r ó d ł o: Opracowanie własne; dane umowne.

Wykres 2.2.3. Wskaźniki względnej ważności elementów

2.3. Rozwiązywanie programów liniowych

2.3.1. Idea algorytmu simpleks

Jedną z metod rozwiązywania zadań programowania liniowego jest algorytm simpleks. Badamy kolejne rozwiązania bazowe programu liniowego w postaci kanonicznej, czyli takiej, w której wszystkie warunki ograniczające są równościami. Rozpoczynamy od dowolnego rozwiązania podstawowego, sprawdzamy, czy jest optymalne, potem znajdujemy następne, lepsze rozwiązanie z punktu widzenia funkcji celu dążącej do minimum lub maksimum. Jeśli uznamy, że aktualnego rozwiązania bazowego nie można już poprawić, kończymy proces iteracyjny. Przejść możemy różnymi sposobami, np. tworząc kolejne tablice simpleksowe na podstawie przekształcenia – macierzowe lub algebraiczne. W niniejszym rozdziale zastosowano wzory zamieszczone w publikacji K. Kukuły¹⁵⁸.

Postać programu liniowego jest następująca:

– funkcja celu:

$$c_1 \cdot x_1 + c_2 \cdot x_2 + \dots + c_n \cdot x_n \rightarrow \max,$$

– ograniczenia:

$$a_{11} \cdot x_1 + a_{12} \cdot x_2 + \dots + a_{1n} \cdot x_n \leq b_1,$$

.....

$$a_{m1} \cdot x_1 + a_{m2} \cdot x_2 + \dots + a_{mn} \cdot x_n \leq b_m,$$

– warunki brzegowe:

$$x_1, x_2, \dots, x_n \geq 0.$$

Postać macierzowa tego modelu zadania decyzyjnego:

$$\mathbf{c}\mathbf{x} \rightarrow \max,$$

$$\mathbf{A}\mathbf{x} \leq \mathbf{b},$$

$$\mathbf{x} \geq 0,$$

przy czym macierz współczynników oraz wektory kolumnowe przedstawiamy jako:

¹⁵⁸ K. Kukuła (red.), *Badania operacyjne w przykładach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa 2004, rozdz. 1.4.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m1} & \dots & a_{mn} \end{bmatrix}, \quad \mathbf{c} = [c_1 \quad \dots \quad c_n], \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix}.$$

Pierwszym krokiem jest zamiana nierówności na równości w ograniczeniach przez wprowadzenie dodatkowych zmiennych; i tak jeśli:

\leq , to do lewych stron dodajemy zmienne swobodne, wchodzące do początkowego rozwiązania bazowego,

\geq , to od lewych stron odejmujemy zmienne swobodne i dodajemy zmienne sztuczne, w zakresie niezbędnym dla macierzy jednostkowej, co przedstawione będzie na przykładach.

Do funkcji celu zmienne swobodne wchodzą ze współczynnikami zero, natomiast zmienne sztuczne ze współczynnikami M , przy czym M jest liczbą bardzo dużą ($M \rightarrow \infty$). Zmienne swobodne mogą znaleźć się w końcowym rozwiązaniu zadania programowania liniowego (PL), lecz zmienne sztuczne nie. Jeśli funkcja celu jest maksymalizowana, to zmienne sztuczne występują ze współczynnikami $-M$, a gdy minimalizowana, to $+M$.

Pierwszą tablicę w postaci macierzowej można zapisać jak tab. 2.3.1:

Tabela 2.3.1

1 tablica w postaci macierzowej – tablica bazowa

c_b	c_j	25	10	0	0	0	Rozwiązanie b_i
	Zmienne bazowe x_b	x_1	x_2	x_3	x_4	x_5	
0	x_3	1	3	1	0	0	500
0	x_4	1,5	1	0	1	0	600
0	x_5	2	0	0	0	1	700
	z_j	0	0	0	0	0	0
	$k_j = c_j - z_j$	25	10	0	0	0	

Źródło: Opracowanie własne.

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 1,5 & 1 \\ 2 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 500 \\ 600 \\ 700 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{c} = [25 \quad 10 \quad 0 \quad 0 \quad 0], \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

gdzie:

\mathbf{A} – macierz współczynników występujących po lewej stronie warunków ograniczających,

\mathbf{b} – wektor wyrazów wolnych warunków ograniczających,

\mathbf{c} – wektor wierszowy współczynników funkcji celu,

x_b – wektor zmiennych bazowych,

- \mathbf{c}_b – wektor kolumnowy bazowy współczynników funkcji celu,
- \mathbf{I} – macierz jednostkowa o wymiarach $m \times m$,
- $\mathbf{0}$ – wektor zer (w wierszu z_j).

Liczba kolumn modelu w postaci kanonicznej odpowiada łącznej liczbie zmiennych:

- decyzyjnych,
- swobodnych,
- sztucznych.

Współczynniki przy zmiennych bazowych tworzą macierz jednostkową. Wartości z_j poszczególnych zmiennych oblicza się jako sumę iloczynów współczynników a_{ij} , odpowiadających poszczególnym zmiennym, i współczynników funkcji celu zmiennych bazowych c_{bi} :

$$z_j = \sum_{i=1}^m a_{ij} \cdot c_{bi}.$$

Wiersz ($k_j = c_j - z_j$) – kryterium simpleks – informuje, o ile zmieni się aktualna w danej iteracji wartość funkcji celu FC , jeżeli jednostkę tej zmiennej wprowadzimy do kolejnej bazy. Kryterium to służy do sprawdzenia, czy aktualne rozwiązanie jest już optymalne, i tak:

- $FC \rightarrow \max, k_j \leq 0,$
- $FC \rightarrow \min, k_j \geq 0.$

W niektórych publikacjach stosowane jest kryterium:

$$g_j = z_j - c_j.$$

Każdą kolejną tablicę simpleksową można przedstawić w postaci zapisu macierzowego, która wygląda w l -tej iteracji następująco:

T a b e l a 2.3.2

Kolejna tablica simpleksowa

	Zmienne bazowe	\mathbf{c}		Rozwiązanie b_i
\mathbf{c}_{bl}	\mathbf{x}_{bl}	$\mathbf{B}_l^{-1} \mathbf{A}$	\mathbf{B}_l^{-1}	$\mathbf{B}_l^{-1} \mathbf{b}$
	z_j	$\mathbf{c}_b^T \mathbf{B}_l^{-1} \mathbf{A}$	$\mathbf{c}_b^T \mathbf{B}_l^{-1}$	$\mathbf{c}_b^T \mathbf{B}_l^{-1} \mathbf{b}$
	$k_j = c_j - z_j$	$\mathbf{c}_j - \mathbf{c}_b^T \mathbf{B}_l^{-1} \mathbf{A}$	$-\mathbf{c}_b^T \mathbf{B}_l^{-1}$	

gdzie:

\mathbf{B}_l^{-1} – macierz odwrotna do wartości zmiennych bazowych danej l -tej iteracji,

$\mathbf{c}_b^T \mathbf{B}_l^{-1} \mathbf{b}$ – wartości funkcji celu.

Ź r ó ł o: Opracowanie na podstawie: K. Kukuła, (red.), *Badania operacyjne w przykładach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa 2004, tabl. 40.

Zatem elementem podstawowym w przekształceniach macierzowych kolejnych tablic iteracji jest macierz odwrotna \mathbf{B}_j^{-1} . Na przykładzie tab. 2.3.3 widzimy, że macierz odwrotna zastępuje macierz jednostkową w tablicy bazowej.

Tabela 2.3.3

Przekształcona tablica bazowa

c_b	c_j	25	10	0	0	0	Rozwiązanie b_i
	Zmienne bazowe	x_1	x_2	x_3	x_4	x_5	
0	x_3	0	3	1	0	-0,5	150
0	x_4	0	1	0	1	-0,75	125
25	x_1	1	0	0	0	0,5	350
	z_j	25	0	0	0	12,5	8 750
	$k_j = c_j - z_j$	0	10	0	0	-12,5	

Źródło: Opracowanie własne w Excelu.

Sposób obliczenia elementów tej tablicy przedstawiony zostanie w dalszej części opracowania. Z tablicy simpleksowej wynika, że rozwiązanie nie jest nadal optymalne, występuje bowiem wartość dodatnia (10) w wierszu kryterium przy założeniu, że:

$$FC \rightarrow \max.$$

Wartość 8750 to obliczona funkcja celu według:

$$FC_1 = \sum_i c_{bi} \cdot b_i = \sum (0 \cdot 150) + (0 \cdot 125) + (25 \cdot 350) = 8750.$$

2.3.2. Formułowanie zadań decyzyjnych

Problemy, dla których szukamy najlepszego rozwiązania przy pomocy programowania liniowego, wymagają najpierw sformułowania zadania decyzyjnego, a w ramach niego określenia funkcji celu oraz ograniczeń. Przedstawmy to na przykładach.

Przykład 1

Firma betoniarska produkuje duże bloczki betonowe oraz kręgi do studzienek ściekowych. Do produkcji wykorzystuje dwa materiały: żwir i cement (nie licząc wody, którą ma bezpłatnie z własnego stawu). Do wykonania bloczka betonowego potrzebne jest 10 kg żwiru i 2 kg cementu. Wykonanie kręgu wymaga 100 kg żwiru oraz 25 kg cementu. Aktualnie w magazynie znajduje się 30 t żwiru oraz 5 t cementu. Firma dąży do maksymalizacji przychodu. Sformułujemy zadanie decyzyjne, przyjmując ceny:

– bloczek – 3 zł/szt.,

– krąg – 50 zł.

Zadaniem naszym jest ułożenie liniowego zadania decyzyjnego (LZD).

Tabela 2.3.4

Normy jednostkowe i zasoby

Wyrób	Bloczek	Krąg	Zasoby
Żwir [kg]	10	100	30 000
Cement [kg]	2	25	5 000
Cena [zł]	3	50	

Źródło: Opracowanie własne; dane umowne.

x_1 – liczba bloczków,

x_2 – liczba kręgów.

Funkcja celu oraz ograniczenia:

$$FC: 3 \cdot x_1 + 50 \cdot x_2 \rightarrow \max,$$

$$C1: 10 \cdot x_1 + 100 \cdot x_2 \leq 30\,000,$$

$$C2: 2 \cdot x_1 + 25 \cdot x_2 \leq 5000,$$

$$x_1, x_2 \geq 0,$$

$$x_1, x_2 \in C \text{ (liczby całkowite).}$$

Przykład 2

Ogrodnik uprawiający astry w doniczkach przyjmuje, że wymagają one składników nawozowych, których codzienne dawki wynoszą przynajmniej: azot – 25 g, potas – 15 g, wapń – 20 g, magnez – 1 g.

Może zakupić trzy rodzaje nawozów, których ceny za 1000 g wynoszą odpowiednio: N_1 – 1 zł, N_2 – 2 zł, N_3 – 1,5 zł. Aby nie przedawkować azotu, należy go stosować nie więcej niż 50 g dziennie na doniczkę. Zawartość składników nawozowych w 1000 g poszczególnych nawozów podaje tab. 2.3.5. Ułożymy liniowe zadanie decyzyjne minimalizujące koszt uprawy astrów.

Tabela 2.3.5

Zawartość składników nawozowych

Nawóz	N_1	N_2	N_3	Dawki
S_1 – azot [g]	250	150	100	25
S_2 – potas [g]	150	200	100	15
S_3 – wapń [g]	50	70	150	20
S_4 – magnez [g]	1	0	0	50
Cena [zł]	1	2	1,5	

Źródło: Opracowanie własne; dane umowne.

x_1 – liczba kg nawozu N_1 ,

x_2 – liczba kg nawozu N_2 ,

x_3 – liczba kg nawozu N_3 ,

x_4 – liczba kg nawozu N_3 .

Funkcja celu oraz ograniczenia:

$$FC: x_1 + 2 \cdot x_2 + 1,5 \cdot x_3 \rightarrow \min,$$

$$C_1: 250 \cdot x_1 + 150 \cdot x_2 + 100 \cdot x_3 \geq 25,$$

$$C_2: 150 \cdot x_1 + 200 \cdot x_2 + 100 \cdot x_3 \geq 15,$$

$$C_3: 50 \cdot x_1 + 70 \cdot x_2 + 150 \cdot x_3 \geq 20,$$

$$C_4: x_1 \leq 50,$$

$$x_1, x_2, x_3 \geq 0.$$

Przykład 3

Zakład stolarski planuje produkować dwa nowe rodzaje okien drewnianych O_1 i O_2 . Możliwości uruchomienia są limitowane zdolnościami produkcyjnymi trzech warsztatów; toczenia – T , montażu – M , lakierni – L . Pracochłonności jednostkowe w godz./szt., zdolności produkcyjne warsztatów w miesiącu oraz zyski ze sprzedaży danego rodzaju okna podaje tab. 2.3.6. Zakład dąży do maksymalizacji zysku. Sformułujmy ten problem w postaci zadania decyzyjnego.

Tabela 2.3.6

Pracochłonności jednostkowe [godz./szt.]

Okno \ Warsztat	O_1 [szt.]	O_2 [szt.]	Zdolności [godz.]
T	5	4	190
M	8	6	250
L	3	2	150
Zysk [zł]	200	250	

Źródło: Opracowanie własne; dane umowne.

x_1 – liczba okien O_1 ,

x_2 – liczba okien O_2 .

Funkcja celu oraz ograniczenia:

$$FC: 200 \cdot x_1 + 250 \cdot x_2 \rightarrow \max,$$

$$C_1: 5 \cdot x_1 + 4 \cdot x_2 \leq 190,$$

$$C_2: 8 \cdot x_1 + 6 \cdot x_2 \leq 250,$$

$$C_3: 3 \cdot x_1 + 2 \cdot x_2 \leq 150,$$

$$x_1, x_2 \geq 0,$$

$$x_1, x_2 \in C.$$

Przykład 4

Hodowca drobiu podaje kurczętom niezbędne witaminy w tygodniowych dawkach:

– witamina D, przynajmniej 100 j. (jednostek),

– witamina A, nie więcej niż 150 j.,

– antybiotyk, nie więcej niż 100 j.

Stosuje mieszanki paszowe, których cena za kg wynosi: $M_1 - 1$ zł, $M_2 - 1,5$ zł, $M_3 - 2$ zł. Zawartość witamin oraz antybiotyku w poszczególnych paszach podaje tab. 2.3.7. Ułożmy zadanie decyzyjne, minimalizujące koszt tygodniowy hodowli drobiu.

T a b e l a 2.3.7

Zawartość witamin i antybiotyku w paszy

Mieszanka Składnik	M1	M2	M3	Porcje
D [j.]	25	20	30	100
A [j.]	10	15	25	150
Antybiotyk [j.]	30	35	20	100
Cena [kg/zł]	1	1,5	2	

Ź r ó d ł o: Opracowanie własne; dane umowne.

x_1 – liczba kg mieszanki M_1 ,

x_2 – liczba kg mieszanki M_2 ,

x_3 – liczba kg mieszanki M_3 .

$FC: x_1 + 1,5 \cdot x_2 + 2 \cdot x_3 \rightarrow \min,$

$C_1: 25 \cdot x_1 + 20 \cdot x_2 + 30 \cdot x_3 \geq 100,$

$C_2: 10 \cdot x_1 + 15 \cdot x_2 + 25 \cdot x_3 \leq 150,$

$C_3: 30 \cdot x_1 + 35 \cdot x_2 + 20 \cdot x_3 \leq 100,$

$x_1, x_2, x_3 \geq 0.$

P r z y k ł a d 5

Zakład blacharski produkuje ośmiolitrowe konewki do podlewania wody. Z arkuszy blachy ocynkowanej wycina elementy poszycia (dno, płaszcz, lejek, zwężka, sitko) trzema sposobami na różnych wycinarkach. Proces technologiczny warunkuje, że trzeci sposób cięcia nie może przekraczać 60% pierwszego. Liczbę elementów uzyskanych z arkusza podaje tab. 2.3.8. Na konewkę potrzebny jest komplet elementów: dwie zwężki i dwa lejki, a sitko, denko i płaszcz po sztuce. Zakład dysponuje 100 arkuszami blachy i zamierza wykonać co najmniej 150 konewek po cenach zależnych od sposobu cięcia w zł/szt., dążąc do maksymalizacji przychodu. Sformułujemy problem jako zadanie PL.

T a b e l a 2.3.8

Liczba elementów z arkusza

Sposoby Elementy	S_1 [szt.]	S_2 [szt.]	S_3 [szt.]
D (denko)	10	15	8
P (płaszcz)	5	7	4
L (lejek)	20	25	25
Z (zwężka)	25	20	20
S (sitko)	30	25	40
Cena [zł/szt.]	30	25	40

Ź r ó d ł o: Opracowanie własne; dane umowne.

x_1 – liczba sposobu cięcia S_1 ,
 x_2 – liczba sposobu cięcia S_2 ,
 x_3 – liczba sposobu cięcia S_3 .

$$FC: 30 \cdot x_1 + 25 \cdot x_2 + 40 \cdot x_3 \rightarrow \max,$$

$$C_1: x_1 + x_2 + x_3 \leq 100,$$

$$C_2: 10 \cdot x_1 + 15 \cdot x_2 + 8 \cdot x_3 \geq 150,$$

$$C_3: 5 \cdot x_1 + 7 \cdot x_2 + 4 \cdot x_3 \geq 150,$$

$$C_4: 20 \cdot x_1 + 25 \cdot x_2 + 25 \cdot x_3 \geq 300,$$

$$C_5: 25 \cdot x_1 + 20 \cdot x_2 + 20 \cdot x_3 \geq 300,$$

$$C_6: 30 \cdot x_1 + 25 \cdot x_2 + 25 \cdot x_3 \geq 150,$$

$$C_7: x_3 \leq 0,6 \cdot x_1,$$

$$x_1, x_2, x_3 \geq 0,$$

$$x_1, x_2, x_3 \in C.$$

Przykład 6

Hodowca bydła oprócz produkcji podstawowej posiada również trzy tereny uprawne: 20, 10, 25 ha. Uprawia na nich koniczynę, żyto oraz zbiera siano na paszę. Plony liczone są w kwintalach z ha:

Tabela 2.3.9

		Plony [q/ha]			Obszar [ha]
		Koniczyna	Żyto	Siano	
Teren	Uprawa				
	T_1	20	30	50	10
	T_2	35	40	60	20
T_3		25	35	40	25

Źródło: Opracowanie własne; dane umowne.

Gdyby sprzedał produkty, to uzyskałby ceny: koniczyna – 10 zł/q, żyto – 50 zł/q, siano – 10 zł/q. Farmer może hodować co najmniej 110 krów dojnych. Każda krowa zjada w ciągu roku co najmniej: 7,5 q koniczyny, 2,5 q żyta w formie otrębów oraz 10 q siana, które musi skarmić w pełni (gdyż nie ma gdzie zbyć) oraz własną słomę z wymłóconego żyta. Uzyskuje 1500 litrów mleka od krowy w ciągu roku, które przekazuje mleczarni po 1,5 zł/litr. Farmer chce uzyskać maksymalny zysk w danym roku. Wyznaczmy więc zadanie decyzyjne, które umożliwi rozwiązanie tego problemu.

x_{ij} – obszar uprawy na i -tym terenie j -tego produktu, $i = 1, 2, 3$, $j = 1, 2, 3$,

x_1 – liczba krów dojnych [szt.].

Zysk stanowi różnicę między przychodem ze sprzedaży mleka a kosztami wynikającymi z karmienia krów.

$$x_1 \cdot 1500 \cdot 1,5 - x_1 \cdot ((7,5 \cdot 10) + (2,5 \cdot 50) + (10 \cdot 10)) = 1955 \cdot x_1,$$

$$FC: 1955 \cdot x_1 \rightarrow \max.$$

Ograniczenie areału:

$$C_1: x_{11} + x_{12} + x_{13} \leq 20,$$

$$C_2: x_{21} + x_{22} + x_{23} \leq 10,$$

$$C_3: x_{31} + x_{32} + x_{33} \leq 25.$$

Ograniczenie wykarmienia:

$$C_4: 20x_{11} + 35x_{21} + 25x_{31} \geq 7,5x_1,$$

$$C_5: 30x_{12} + 40x_{22} + 35x_{32} \geq 2,5x_1,$$

$$C_6: 50x_{13} + 60x_{23} + 40x_{33} = 10x_1.$$

Ograniczenie obory:

$$C_7: x_1 \leq 110.$$

Ograniczenia brzegowe:

$$x_{ij} \geq 0 \quad x_1 \geq 1 \quad x_1 \in C.$$

Przykład 7

Zakład ogólnobudowlany zatrudnia trzech murarzy, którym w danym dniu trzeba wyznaczyć roboty do wykonania. Mistrz oszacował czas pracy w godzinach na wykonanie zadań w zależności od doświadczenia robotników.

Tabela 2.3.10

Pracochłonność zadań

Zadanie \ Murarz	Murarz		
	M_1 [godz.]	M_2 [godz.]	M_3 [godz.]
Z_1	1	2	1,5
Z_2	3	4,5	2
Z_3	2	2,5	1,5
Z_4	4	6	5
Z_5	2,5	3	4

Źródło: Opracowanie własne; dane umowne.

Przyjmijmy założenie: każdy murarz może wykonywać nie więcej niż dwa zadania, a każde zadanie musi być wykonywane tylko przez jednego pracownika. Ułóżmy liniowe zadanie decyzyjne, którego rozwiązanie minimalizuje łączny czas pracy murarzy.

Zmienne w tym przykładzie bazują na liczbach binarnych $\{0, 1\}$.

x_{ij} – i -ty murarz wykonuje j -te zadanie,

0 – gdy M_j nie wykonuje zadania Z_j ;

1 – gdy M_j wykonuje zadanie Z_j .

Z opisu wynikają możliwości wykonywania zadań, co przedstawiono w tab. 2.3.11:

Tabela 2.3.11

Możliwości wykonywania zadań przez murarzy

Zadanie \ Murarz	Murarz			Suma
	M_1 [godz.]	M_2 [godz.]	M_3 [godz.]	
Z_1	1	0	0	1
Z_2	0	1	0	1
Z_3	1	0	1	1
Z_4	0	1	0	1
Z_5	0	0	1	1
\leq	2	2	2	

Źródło: Opracowanie własne; dane umowne.

$$FC: x_{11} + 2x_{12} + 1,5x_{13} + 3x_{21} + 4,5x_{22} + 2x_{23} + 2x_{31} + 2,5x_{32} + 1,5x_{33} + 4x_{41} + 6x_{42} + 5x_{43} + 2,5x_{51} + 3x_{52} + 4x_{53} \rightarrow \min.$$

Ograniczenia poziome:

$$C_1: x_{11} + x_{12} + x_{13} = 1,$$

$$C_2: x_{21} + x_{22} + x_{23} = 1,$$

$$C_3: x_{31} + x_{32} + x_{33} = 1,$$

$$C_4: x_{41} + x_{42} + x_{43} = 1,$$

$$C_5: x_{51} + x_{52} + x_{53} = 1.$$

Ograniczenia pionowe:

$$C_6: x_{11} + x_{21} + x_{31} + x_{41} + x_{51} \leq 2,$$

$$C_7: x_{12} + x_{22} + x_{32} + x_{42} + x_{52} \leq 2,$$

$$C_8: x_{13} + x_{23} + x_{33} + x_{43} + x_{53} \leq 2.$$

Ograniczenie na liczbach binarnych:

$$x_{11}, \dots, x_{53} \in \{0, 1\}.$$

2.3.3. Zastosowanie modułu (LP) pakietu WinQSB

Sformułowaliśmy siedem różnych przykładów na liczbach całkowitych, rzeczywistych i binarnych. Rozwiązania możemy szybko uzyskać, posługując się modułem programowania liniowego i całkowitoliczbowego (LP) pakietu WinQSB. Poszczególne prezentacje danego przykładu obejmują:

- sformułowanie zadania decyzyjnego w układzie arkusza kalkulacyjnego,
- rozwiązanie sumaryczne.

W wersji komputerowej zmienne oznaczono jako X_1, X_2, \dots , natomiast ograniczenia jako C_1, C_2, \dots .

Ad 1) Dane według tab. 2.3.4. Maksymalizacja przychodu z produkcji prefabrykatów betonowych (błoczki, kręgi).

Tabela 2.3.12

Sformułowanie zadania decyzyjnego produkcji prefabrykatów

LP	MatrixFormat	Maksymalizacja przychodu	2	2
Variable -->	X1	X2	Direction	RHS
Maximize	3	50		
C1	10	100	<=	30 000
C2	2	25	<=	5 000
LowerBound	0	0		
UpperBound	M	M		
VariableType	Integer	Integer		

gdzie:

Ilość zmiennych (Variable) – 2, ilość ograniczeń – 2,

RHS – wartość ograniczenia; Direction (kierunek) – znak nierówności,

Integer – określenie całkowitoliczbowości zmiennych,

LowerBound, UpperBound – przedział zmienności (warunki brzegowe).

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Tabela 2.3.13

Rozwiązanie sumaryczne problemu decyzyjnego maksymalizacji przychodu

09-06-2007	Decision	Solution	Unit Cost or	Total
16:36:19	Variable	Value	Profit $C(j)$	Contribution
1	X1	0	3	0
2	X2	200	50	10 000
	Objective	Function	Max =	10 000

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Rozwiązaniem optymalnym jest produkcja tylko kręgów betonowych w ilości 200 szt. Przychód wynosi 10 000 zł.

Ad 2) Dane według tabeli 2.3.5. Minimalizacja kosztu uprawy astrów.

Tabela 2.3.14

Sformułowanie zadania decyzyjnego uprawy astrów

LP	MatrixFormat	Astry	3	4	
Variable -->	X1	X2	X3	Direction	RHS
Minimize	1	2	1.5		
C1	250	150	100	>=	25
C2	150	200	100	>=	15
C3	50	70	150	>=	20
C4	1	0	0	<=	50
LowerBound	0	0	0		
UpperBound	M	M	M		
VariableType	Continuous	Continuous	Continuous		

Continuous – dane rzeczywiste, przy czym w tabeli wejściowej zamiast przecinka wstawiamy kropkę dziesiętną.

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Tabela 2.3.15

Rozwiązanie sumaryczne problemu decyzyjnego minimalizacji kosztu

09-06-2007	Decision	Solution	Unit Cost or	Total
16:56:48	Variable	Value	Profit $C(j)$	Contribution
1	X1	0,0538	1	0,0538
2	X2	0	2	0
3	X3	0,1154	1,5	0,1731
	Objective	Function	(Min.) =	0,2269

Źródło: Opracowanie własne, moduł (LP), program WinQSB, wyniki w kg.

Optymalne rozwiązanie nawożenia doniczki astrów: N_1 – 53,8 g, N_3 – 115,4 g. Koszt nawożenia – 0,23 zł/szt.

Ad 3) Dane według tabeli 2.3.6. Maksymalizacja zysku ze sprzedaży okien.

Tabela 2.3.16

Sformułowanie zadania decyzyjnego sprzedaży okien

LP	MatrixFormat	Okna	2	3
Variable -->	X1	X2	Direction	RHS
Maximize	200	250		
C1	5	4	<=	190
C2	8	6	<=	250
C3	3	2	<=	150
LowerBound	0	0		
UpperBound	M	M		
VariableType	Integer	Integer		

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Tabela 2.3.17

Rozwiązanie sumaryczne problemu decyzyjnego maksymalizacji zysku ze sprzedaży okien

09-07-2007	Decision	Solution	Unit Cost or	Total
08:38:11	Variable	Value	Profit C (j)	Contribution
1	X1	0	200	0
2	X2	41	250	10 250
	Objective	Function	(Max.) =	10 250

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Program proponuje produkcję tylko okien O_2 i uzyskanie dzięki temu zysku 10 250 zł.

Ad 4) Dane według tabeli 2.3.7. Minimalizacja kosztu tygodniowego chowu kurcząt.

Tabela 2.3.18

Sformułowanie zadania decyzyjnego chowu kurcząt

LP	MatrixFormat	Kurki	3	3
Variable -->	X1	X2	X3	Direction
Minimize	1	1.5	2	
C1	25	20	30	>=
C2	10	15	25	<=
C3	30	35	20	<=
LowerBound	0	0	0	
UpperBound	M	M	M	
VariableType	Continuous	Continuous	Continuous	

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Tabela 2.3.19

Rozwiązanie sumaryczne problemu decyzyjnego minimalizacji
kosztu chowu kurcząt

09-07-2007	Decision	Solution	Unit Cost or	Total
08:50:42	Variable	Value	Profit $C(j)$	Contribution
1	X1	2,5	1	2,5
2	X2	0	1,5	0
3	X3	1,25	2	2,5
	Objective	Function	(Min.) =	5

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Koszt tygodniowy wynosi 5 zł.

Ad 5) Dane według tabeli 2.3.8. Maksymalizacja przychodu z produkcji konewek.

Tabela 2.3.20

Sformułowanie zadania decyzyjnego produkcji konewek

LP	MatrixFormat	Konewki	3	7	
Variable -->	X1	X2	X3	Direction	RHS
Maximize	30	25	40		
C1	1	1	1	<=	100
C2	10	15	8	>=	150
C3	5	7	4	>=	150
C4	20	25	25	>=	300
C5	25	20	20	>=	300
C6	30	25	25	>=	150
C7	-0.6	0	1	<=	0
LowerBound	0	0	0		
UpperBound	M	M	M		
VariableType	Integer	Integer	Integer		

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Tabela 2.3.21

Rozwiązanie sumaryczne problemu decyzyjnego maksymalizacji przychodu
z produkcji konewek

09-07-2007	Decision	Solution	Unit Cost or	Total
08:59:49	Variable	Value	Profit $C(j)$	Contribution
1	X1	63	30	1 890
2	X2	0	25	0
3	X3	37	40	1 480
	Objective	Function	(Max.) =	3 370

W tabelach rozwiązania sumarycznego:

Unit Cost or Profit $C(j)$ – jednostka kosztu lub zysku,

Total Contribution – podsumowanie (ilość rozwiązania · cena),

Reduced Cost, Basic Status – redukcja kosztu, status bazowy.

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Rozwiązanie: 63 rozkroje blachy sposobem S_1 oraz 37 rozkroi blachy sposobem S_3 .

Ad 6) Dane według tabeli 2.3.9. Maksymalizacja zysku ze sprzedaży mleka.

Tabela 2.3.22

Sformułowanie zadania decyzyjnego sprzedaży mleka

LP	MatrixFormat	Mleko	10	7								
Variable -->	X11	X12	X13	X21	X22	X23	X31	X32	X33	X1	Direction	RHS
Maximize	0	0	0	0	0	0	0	0	0	1955		
C1	1	1	1	0	0	0	0	0	0	0	<=	20
C2	0	0	0	1	1	1	0	0	0	0	<=	10
C3	0	0	0	0	0	0	1	1	1	0	<=	25
C4	20	0	0	35	0	0	25	0	0	-7.5	>=	0
C5	0	30	0	0	40	0	0	35	0	-2.5	>=	0
C6	0	0	50	0	0	60	0	0	40	-10	=	0
C7	0	0	0	0	0	0	0	0	0	1	<=	110
LowerBound	0	0	0	0	0	0	0	0	0	0		
UpperBound	M	M	M	M	M	M	M	M	M	M		
VariableType	Continuous	C	C	C	C	C	C	C	C	C	Integer	

C – typ zmiennej (continuous); *Integer* – całkowitoliczbowe

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Tabela 2.3.23

Rozwiązanie sumaryczne problemu decyzyjnego maksymalizacji zysku ze sprzedaży mleka

09-10-2007	Decision	Solution	Unit Cost or	Total
10:36:22	Variable	Value	Profit C (j)	Contribution
1	X11	0	0	0
2	X12	0	0	0
3	X13	20	0	0
4	X21	9,5	0	0
5	X22	0	0	0
6	X23	0,5	0	0
7	X31	17,6	0	0
8	X32	7,4	0	0
9	X33	0	0	0
10	X1	103	1955	201 365
	Objective	Function	(Max.) =	201 365

Źródło: Opracowanie własne, moduł (LP), programu WinQSB.

Farmer powinien zebrać:

- siano z 20 ha terenu T_1 , z 0,5 ha terenu T_2 ,
- koniczynę jako paszę zieloną z 9,5 ha terenu T_2 , z 17,6 ha terenu T_3 ,
- żyto z 7,4 ha terenu T_3 .

Pozwoli to utrzymać 103 krowy dojne, a ze sprzedaży mleka uzyskać 201 365 zł.

Ad 7) Dane według tabeli 2.3.10. Minimalizacja łącznego czasu pracy murarzy

Tabela 2.3.24

Sformułowanie zadania decyzyjnego czasu pracy murarzy

Variable -->	X11	X12	X13	X21	X22	X23	X31	X32	X33	X41	X42	X43	X51	X52	X53	Direction	RHS
Minimize	1	2	1.5	3	4.5	2	2	2.5	1.5	4	6	5	2.5	3	4		
C1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	=	1
C2	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	=	1
C3	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	=	1
C4	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	=	1
C5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	=	1
C6	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	<=	2
C7	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	<=	2
C8	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	<=	2
LowerBound	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
UpperBound	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
VariableType	Binary	Binary	B	B	B	B	B	B	B	B	B	B	B	B	B		

B (Binary) – format danych binarny.

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Tabela 2.3.25

Rozwiązanie sumaryczne problemu decyzyjnego minimalizacji łącznego czasu murarzy

09-10-2007	Decision	Solution	Unit Cost or	Total
10:56:57	Variable	Value	Profit C(j)	Contribution
1	X11	1	1	1
2	X12	0	2	0
3	X13	0	1,5	0
4	X21	0	3	0
5	X22	0	4,5	0
6	X23	1	2	2
7	X31	0	2	0
8	X32	0	2,5	0
9	X33	1	1,5	1,5
10	X41	1	4	4
11	X42	0	6	0
12	X43	0	5	0
13	X51	0	2,5	0
14	X52	1	3	3
15	X53	0	4	0
	Objective	Function	(Min.) =	11,5

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Rozdysponowanie pięciu zadań dla murarzy o dobrej wydajności pracy:

– M_1 wykonuje zadania Z_1 i Z_4 (1 godz., 4 godz.)

– M_2 – zadanie Z_5 (3 godz.),

– M_3 – zadania Z_2, Z_3 (2 godz., 1,5 godz.)

Łączny czas pracy wynosi 11,5 godz.

2.3.4. Metoda geometryczna rozwiązywania zadań programowania liniowego

Jeśli zadanie decyzyjne zawiera dwie zmienne X_1, X_2 , to rozwiązać je można graficznie w układzie współrzędnych X_1-X_2 . Ograniczenia w postaci nierówności przyjmujemy jako równości. Powstają w ten sposób izolinie, czyli funkcje stałej wartości. Rysujemy izolinie w pierwszej ćwiartce układu współrzędnych, gdyż przyjmujemy warunek, że: $X_1, X_2 \geq 0$. Do każdej narysowanej linii zaznaczamy obszar rozwiązania dopuszczalnego. Funkcję celu też traktujemy jako izolinię, przyjmując dla niej dowolną wartość. Jeśli ta funkcja dąży do minimum, to wygodnie jest przyjąć dla niej taką wartość, aby była nad izoliniami ograniczeń. Natomiast gdy do maksimum, to poniżej izolinii ograniczeń. Gdy do tak przygotowanych funkcji przyjmiemy kolejno $X_1 = 0$, a następnie $X_2 = 0$, to wyznaczymy punkty przecięcia się z odciętą i rzędną układu współrzędnego. Izolinie wyznaczają nam wielobok, który jest obszarem rozwiązań dopuszczalnych. Przesuwając równoległe izolinię funkcji celu w zależności od kierunku do jakiego dąży (maksimum, minimum), znajdujemy punkt wieloboku. Jest to rozwiązanie optymalne. Powróćmy teraz do danych według tab. 2.3.1 i określmy rozwiązanie optymalne sposobem graficznym:

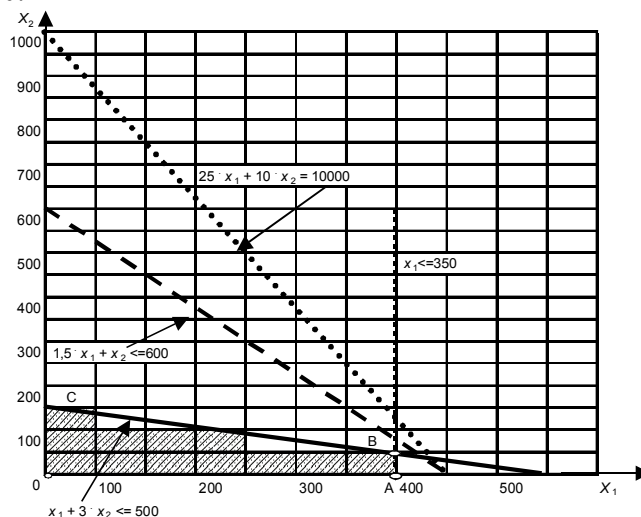
$$FC: 25 \cdot x_1 + 10 \cdot x_2 \rightarrow \max,$$

$$C_1: x_1 + 3 \cdot x_2 \leq 500,$$

$$C_2: 1,5 \cdot x_1 + x_2 \leq 600,$$

$$C_3: 2 \cdot x_1 \leq 700,$$

$$x_1, x_2 \geq 0.$$



Źródło: Opracowanie własne w Excelu.

Wykres 2.3.1. Wyznaczenie rozwiązania optymalnego

Etapy postępowania:

– wyznaczenie izolinii dla C_1 :

$$\begin{aligned} x_1 + 3 \cdot x_2 &= 500, \\ x_1 = 0 \quad x_2 &= 500/3 = 166,7, \\ x_2 = 0 \quad x_1 &= 500, \end{aligned}$$

– wyznaczenie izolinii dla C_2 :

$$1,5 \cdot x_1 + x_2 = 600, \quad x_1 = 0, \quad x_2 = 600, \quad x_2 = 0, \quad x_1 = 400,$$

– wyznaczenie izolinii dla C_3 :

$$x_1 = 700/2 = 350,$$

– wyznaczenie izolinii dla FC :

$$25 \cdot x_1 + 10 \cdot x_2 = 10\,000 \text{ – przyjęto punkt } (400, 0),$$

$$x_1 = 0 \quad x_2 = 1000, \quad x_2 = 0, \quad x_1 = 10\,000/25 = 400.$$

Czworobok A0CB to obszar (zbiór) rozwiązań dopuszczalnych

Rozwiązanie optymalne – punkt B (350, 50) przecięcia izolinii $x_1 + 3 \cdot x_2 = 500$ oraz $x_1 = 350$.

Posłużmy się teraz modulem (LP) pakietu WinQSB w celu sporządzenia rozwiązania graficznego. Wpierw jednak musimy sformułować zadanie decyzyjne. Zobaczmy też rozwiązanie sumaryczne.

T a b e l a 2.3.26

Sformułowanie zadania decyzyjnego do przykładu według tab. 2.3.1 w module (LP)

LP	MatrixFormat	Wyroby	2	3
Variable -->	X1	X2	Direction	RHS
Maximize	25	10		
C1	1	3	<=	500
C2	1.5	1	<=	600
C3	2	0	<=	700
LowerBound	0	0		
UpperBound	M	M		
VariableType	Integer	Integer		

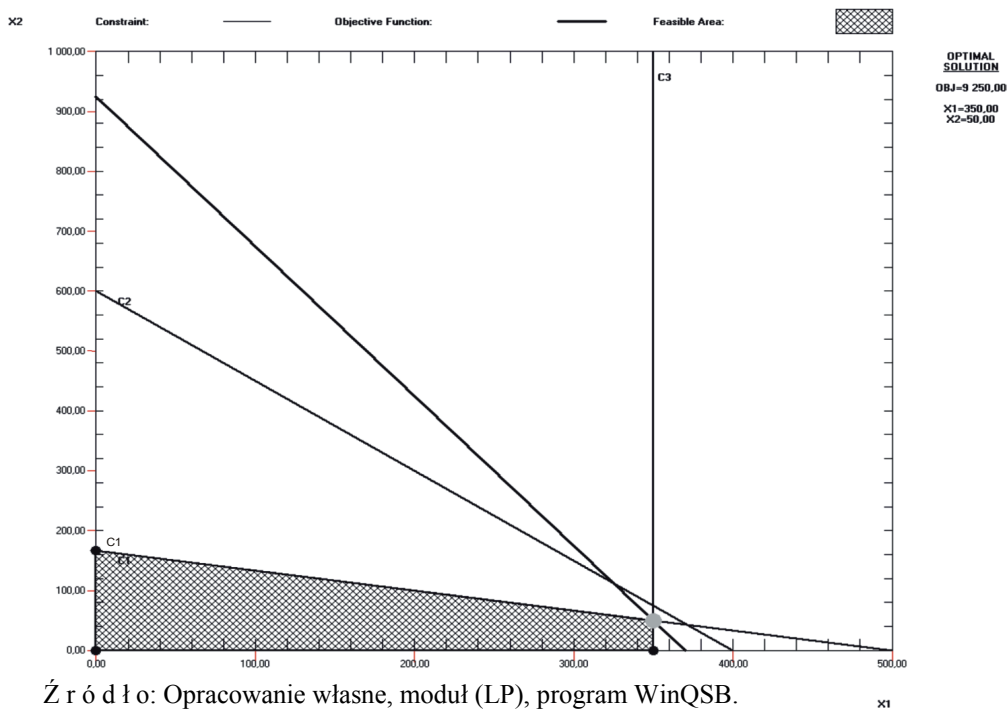
Ź r ó d ł o: Opracowanie własne, moduł (LP), program WinQSB.

Tabela 2.3.27

Rozwiązanie sumaryczne przykładowego problemu decyzyjnego (według danych tab. 2.3.26)

09-07-2007 14:52:23	Decision Variable	Solution Value	Unit Cost or Profit C(j)	Total Contribution
1	X1	350	25	8 750
2	X2	50	10	500
	Objective Function		(Max.) =	9 250

Źródło: Opracowanie własne, moduł (LP), program WinQSB.



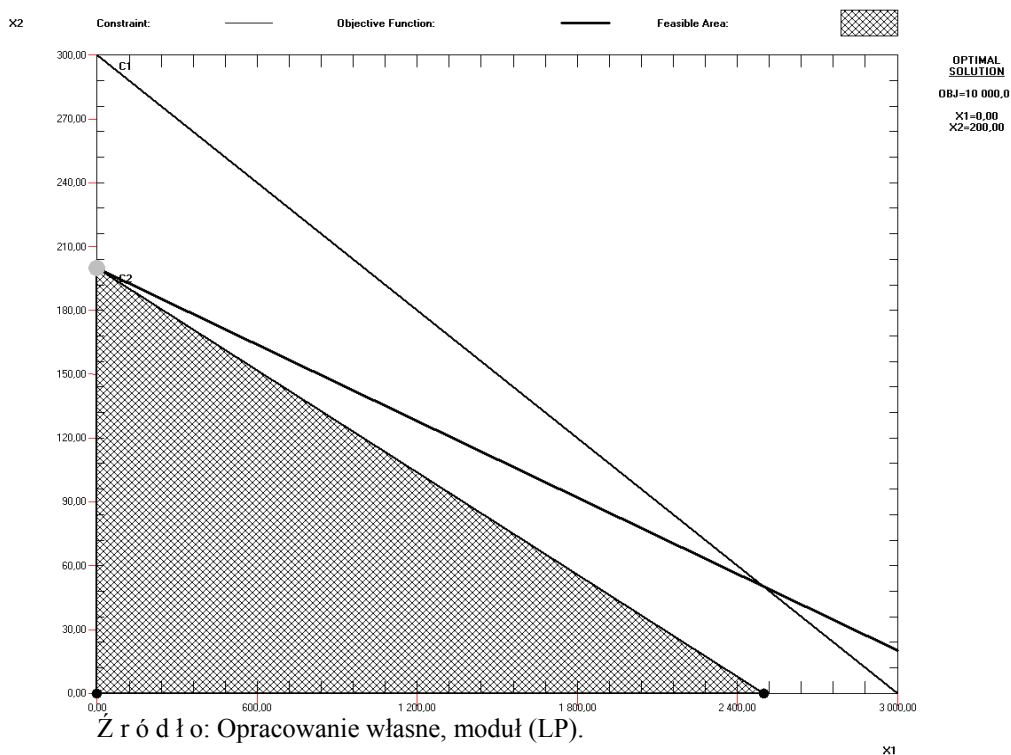
Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Wykres 2.3.2. Rozwiązanie graficzne przykładowego zadania decyzyjnego według danych tabeli 2.3.26

Program zaznacza izolinie skrótami: C_1 , C_2 , C_3 , funkcja celu $FC = OBJ$ wynosi 9250. Rozwiązanie optymalne zaznaczono szarą kropką i jest identyczne (350, 50) jak wykonane w Excelu. Obszar rozwiązań dopuszczalnych jest też identyczny, program zaznacza go w kratkę. Upewniliśmy się co do zgodności rozwiązania graficznego dwoma sposobami i mamy już rozwiązania analityczne zaprezentowane wcześniej na przykładach.

Zastosujmy teraz moduł (LP) pakietu WinQSB do dwóch wcześniejszych przykładów (1 i 3) o dwóch zmiennych.

Ad 1. Maksymalizacja przychodu z produkcji prefabrykatów betonowych (błoczeki, kręgi)



Wykres 2.3.3. Rozwiązanie graficzne zadania decyzyjnego produkcji prefabrykatów

Rozwiązanie optymalne (0, 200) jest zgodne z wcześniejszym obliczeniem.

Ad 3. Maksymalizacja zysku ze sprzedaży okien

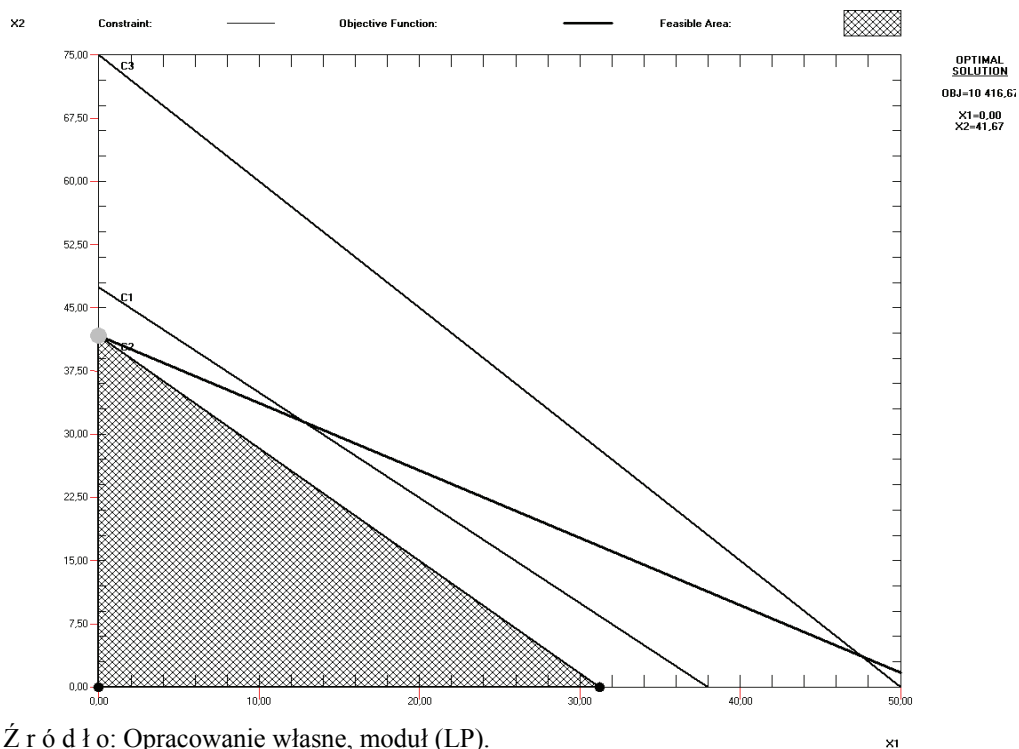
Rozwiązanie optymalne bezpośrednio otrzymane metodą geometryczną nie uwzględnia definicji formatu danych jako całkowitoliczbowe (*Integer*). Stąd różnica w wynikach.

Tabela 2.3.28

Wyniki rozwiązania analitycznego i graficznego problemu decyzyjnego sprzedaży okien

Zmienne i funkcja celu	Analitycznie (zob. tab. 2.3.17)	Graficznie (zob. tab. 2.3.4)
X_1	0	0
X_2	41	41,67
FC	10 250	10 416,67

Źródło: Opracowanie własne, tabela pomocnicza.



Źródło: Opracowanie własne, moduł (LP).

Wykres 2.3.4. Rozwiązanie graficzne zadania decyzyjnego sprzedaży okien

2.3.5. Zastosowanie dodatku Excel-Solver do rozwiązania zadania decyzyjnego

Przykład

Maksymalizacja przychodu ze sprzedaży dwóch wyrobów wykonanych z trzech surowców.

Tabela 2.3.29

Sformułowanie zadania decyzyjnego w arkuszu kalkulacyjnym w zakresie przychodu ze sprzedaży wyrobów

A	B	C	D	E	F
2				$W_1 = x_1$	$W_2 = x_2$
3	Nazwa surowca	Ilość	Wykorzystanie	350	50
4	S_1	500	500	1	3
5	S_2	600	575	1,5	1
6	S_3	700	700	2	0
7			Ceny:	25	10
8	Maksymalizacja przychodu ze sprzedaży		9 250		

Źródło: Opracowanie własne w dodatku programowym Excel-Solver.

Procedura postępowania w definiowaniu komórek arkusza kalkulacyjnego (tab. 2.3.29):

Zadanie decyzyjne	Wyrażenia Excela	
FC: $25 * x_1 + 10 * x_2$ → max	=E7*E3+F7*F3	komórka D8
C ₁ : $x_1 + 3 * x_2 \leq 500$ C ₂ : $1,5 * x_1 + x_2 \leq 600$ C ₃ : $2 * x_1 \leq 700$	=D\$4:\$D\$6<= \$C\$4:\$C\$6	warunki ograniczające
$x_1, x_2 \geq 0$	\$E\$3:\$F\$3>=0 \$E\$3:\$F\$3int	całkowita

Początkowo: $x_1, x_2 = 0$

Wykorzystanie:

=E3*E4+F3*F4

=E3*E5+F3*F5

=E3*E6+F3*F6

Przed rozwiązaniem ustawienie na komórce D8.

Narzędzia: (Solver/okno dialogowe)

Wypełniamy:

komórka celu – \$D\$8,

równa – Maks,

komórki zmieniane – \$E\$3:\$F\$3,

warunki ograniczające; dodaj; \$D\$4:\$D\$6<= \$C\$4:\$C\$6.

Przycisk: Rozwiąż.

Okno dialogowe – Solver znalazł rozwiązanie. Wszystkie ograniczenia i warunki optymalizacji są spełnione.

Dwie możliwości zakończenia:

Przechowaj rozwiązanie.

Przywróć wartości początkowe.

Przechowujemy rozwiązanie.

Wyniki: $x_1 = 350, x_2 = 50$ oraz $FC = 9250$ są zgodne z otrzymanymi w tabeli 2.3.27

metodą analityczną i wykreślą z zastosowaniem modułu (LP) pakietu WinQSB.

Kopiując arkusz tego przykładu, lecz zmieniając parametry wejściowe, rozwiążemy kolejne przykłady z użyciem dodatku Solver.

Przykład

Maksymalizacja przychodu ze sprzedaży prefabrykatów betonowych (dane wejściowe według tab. 2.3.4)

Tabela 2.3.30

Sformułowanie zadania decyzyjnego w arkuszu w zakresie przychodu ze sprzedaży prefabrykatów betonowych

A	B	C	D	E	F
2				Bloczek - x_1	Krąg - x_2
3	Materiał	Zapasy	Wykorzystanie	0	200
4	M_1	30 000	20 000	10	100
5	M_2	5 000	5 000	2	25
7			Ceny:	3	50
8	Maksymalizacja przychodu		10 000		

Źródło: Opracowanie własne w dodatku programowym Excel-Solver.

Procedura postępowania w zakresie definiowania komórek arkusza (tab. 2.3.31)

Zadanie decyzyjne	Wyrażenia Excela	
FC: $x_1 + 2 * x_2 + 1,5 * x_3 \rightarrow \min$	=E8*E3+F8*F3+G8*G3	komórka D9
C ₁ : $250 * x_1 + 150 * x_2 + 100 * x_3 \geq 25$ C ₂ : $150 * x_1 + 200 * x_2 + 100 * x_3 \geq 15$ C ₃ : $50 * x_1 + 70 * x_2 + 150 * x_3 \geq 20$	=D\$4:D\$7<= \$C\$4:C\$7	warunki ograniczające
$x_1, x_2, x_3 \geq 0$	\$E\$3:\$G\$3>=0	

Początkowo: $x_1, x_2, x_3 = 0$

Wykorzystanie:

=E4*E3+F4*F3+G4*G3

=E6*E3+F6*F3+G6*G3

=E5*E3+F5*F3+G5*G3

=E7*E3+F7*F3+G7*G3

Przed rozwiązaniem ustawienie na komórce D9.

Narzędzia: (Solver/okno dialogowe)

Wypełniamy:

komórka celu – \$D\$9,

równa – min,

komórki zmieniane – \$E\$3:\$G\$3,

warunki ograniczające; dodaj; \$D\$4:\$D\$7<= \$C\$4:C\$7.

Przechowujemy rozwiązanie.

Wyniki: $x_1 = 53,8$ g, $x_3 = 115,4$ g oraz $FC = 0,23$ zł/szt. są zgodne z otrzymanymi w tab. 2.3.15.

Przykład

Maksymalizacja zysku przy określonych zdolnościach produkcyjnych warsztatów (dane wejściowe według tab. 2.3.6).

Tabela 2.3.32

Sformułowanie zadania decyzyjnego w arkuszu w zakresie produkcji warsztatów

A	B	C	D	E	F
2				Okno $O_1 - x_1$	Okno $O_1 - x_2$
3	Warsztaty	Zdolności produkcyjne	Wykorzystanie	0	41
4	T	190	164	5	4
5	M	250	246	8	6
6	L	150	82	3	2
7			Zysk:	200	250
8	Maksymalizacja zysku		10 250		

Źródło: Opracowanie własne w dodatku programowym Excel-Solver.

Procedura postępowania w zakresie definiowania komórek w arkuszu (zob. tab. 2.3.32):

Zadanie decyzyjne:	Wyrażenia Excela:	
FC: $200 * x_1 + 250 * x_2$ → max	=E7*E3+F7*F3	komórka D8
C ₁ : $5 * x_1 + 4 * x_2 \leq 190$ C ₂ : $8 * x_1 + 6 * x_2 \leq 250$ C ₂ : $3 * x_1 + 2 * x_2 \leq 150$	=D\$4:\$D\$6<=\$C\$4:\$C\$6	warunki ograniczające
$x_1, x_2 \geq 0$	\$E\$3:\$F\$3int	całkowita

Początkowo: $x_1, x_2 = 0$.

Wykorzystanie:

$$=E4*E3+F4*F3 \qquad =E6*E3+F6*F3$$

$$=E5*E3+F5*F3$$

Przed rozwiązaniem ustawienie na komórce D8.

Narzędzia: (Solver/okno dialogowe)

Wypełniamy:

komórka celu – \$D\$8,

równa – Maks,

komórki zmieniane – \$E\$3:\$F\$3,

warunki ograniczające; dodaj; \$D\$4:\$D\$6<=\$C\$4:\$C\$6.

Przechowujemy rozwiązanie.

Wyniki: $x_1 = 0, x_2 = 41$ oraz $FC = 10\,250$ są zgodne z otrzymanymi w tab. 2.3.17.

Wykonajmy jeszcze dwa przykłady charakterystyczne ze względu na typ zastosowanych zmiennych:

- rzeczywiste i całkowitoliczbowe w jednym problemie,
- binarne.

Przykład

Maksymalizacja zysku ze sprzedaży mleka przy korzystaniu z własnych plonów (według danych wejściowych w tab. 2.3.9 oraz sformułowania zadania decyzyjnego w przykładzie 6).

Tabela 2.3.33

Arkusz wejściowy i wyrażenia Excela w zakresie maksymalizacji zysku ze sprzedaży mleka

A	B	C	D	E	F	G	H	I	J	K	L	M	N
2				x_{11}	x_{12}	x_{13}	x_{21}	x_{22}	x_{23}	x_{31}	x_{32}	x_{33}	x_1
3		Ograniczenie	Wykorzystanie	0,0	0,0	20,0	9,5	0,0	0,5	17,6	7,4	0,0	103
4		20	20	1	1	1	0	0	0	0	0	0	0
5		10	10	0	0	0	1	1	1	0	0	0	0
6		25	25	0	0	0	0	0	0	1	1	1	0
7		0	1	20	0	0	35	0	0	25	0	0	-7,5
8		0	0	0	30	0	0	40	0	0	35	0	-2,5
9		0	0	0	0	50	0	0	60	0	0	40	-10
10		110	103,00	0	0	0	0	0	0	0	0	0	1
11			Cena:	10			50			10			1,5
12		Maksymalizacja zysku	201 365	0	0	0	0	0	0	0	0	0	1 955

Cena dotyczy odpowiednio: koniczyny, żyta, siana, mleka

Wyrażenia w kolumnie „Wykorzystanie” oraz funkcji celu stanowią sumę iloczynów wartości w danej

komórce i odpowiadającej wartości zmiennej np.:

$$=E4*E$3+F4*F$3+G4*G$3+H4*H$3+I4*I$3+J4*J$3+K4*K$3+L4*L$3+M4*M$3+N4*N$3$$

Warunki ograniczające:

$$D$4:D$6<=C$4:C$6$$

$$D$10<=C$10$$

$$D$7:D$8>=C$7:C$8$$

$$E$3:N$3>=0$$

$$D$9=C$9$$

$$N$3int \text{ (zmienna całkowitoliczbowa)}$$

Źródło: Opracowanie własne w dodatku programowym Excel-Solver.

Otrzymane wyniki obliczeń są zgodne z wcześniejszymi otrzymanymi modelem (LP) (zob. tab. 2.3.23).

P r z y k ł a d

Minimalizacja łącznego czasu pracy murarzy (według danych wejściowych w tab. 2.3.10 i 2.3.11 i sformułowania zadania decyzyjnego w przykładzie 7., podrozdział 2.3.2).

T a b e l a 2.3.34

Arkusz wejściowy i wyrażenia Excela w zakresie minimalizacji czasu murarzy

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
2				X ₁₁	X ₁₂	X ₁₃	X ₂₁	X ₂₂	X ₂₃	X ₃₁	X ₃₂	X ₃₃	X ₄₁	X ₄₂	X ₄₃	X ₅₁	X ₅₂	X ₅₃
3		Ogr.	Wyk.	1	0	0	0	0	1	0	0	1	1	0	0	0	1	0
4	C1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
5	C2	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
6	C3	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
7	C4	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
8	C5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
9	C6	2	2	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
10	C7	2	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
11	C8	2	2	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1
12			Norma	1,0	2,0	1,5	3,0	4,5	2,0	2,0	2,5	1,5	4,0	6,0	5,0	2,5	3,0	4,0
13		Min	11,5															

Wyrażenia na funkcję celu jak i „Wykorzystanie” stanowią sumę iloczynów odpowiadających sobie komórek w linii ograniczenia (celu) oraz w wierszu zmiennych.

Warunki ograniczające:

$\$D\$4:\$D\$8=\$C\$4:\$C\$8,$

$\$D\$9:\$D\$11\leq\$C\$9:\$C\$11,$

$\$E\$3:\$S\$3=\text{binarny}.$

Wynik jest zatem identyczny jak w tab. 2.3.25.

Ź r ó d ł o: Opracowanie własne w dodatku programowym Excel-Solver.

Następne trzy przykłady zrealizowane z użyciem dodatku Solver opierają się na propozycjach obliczeń zawartych w programie Excel. Po zainstalowaniu programu, np. Excel 97, dostęp do przykładów wzorcowych jest następujący:

(C:/Program Files/Microsoft office/Office/Examples/Solver/Solverpd.xls).

P r z y k ł a d

Minimalizacja ryzyka zainwestowania w akcje spółek. Danymi początkowymi są zainwestowane udziały w akcjach spółkach.

T a b e l a 2.3.35

Zainwestowanie udziałów w akcje spółek

a) Definiowanie początkowych udziałów w w akcje:

A	B	C
Akcje spółek	Stopa zwrotu	Udziały
A	4,0%	45,0%
B	7,0%	22,0%
C	5,0%	6,0%
D	2,0%	9,0%
E	6,0%	16,0%
F	-1,0%	2,0%
	Suma	100,0%

b) Sformułowanie arkusza, wyrażenia oraz obliczenia według przykładu wzorcowego:

A	B	C	D	E	F	G
Akcje spółek	Stopa zwrotu	Udziały	Stopa zwrotu x udział	Udział ²	Odchylenie standardowe (dane)	Wynik wg wyrażen
A	4,0%	41,0%	0,01639	0,16796	12,0%	0,00571
B	7,0%	21,7%	0,01517	0,04693	21,0%	0,00229
C	5,0%	18,4%	0,00918	0,03368	13,0%	0,00127
D	2,0%	5,3%	0,00107	0,00285	12,5%	0,00068
E	6,0%	13,7%	0,00820	0,01867	18,0%	0,00011
F	-1,0%	0,0%	0,00000	0,00000	12,8%	0,00000
		Suma	Suma		Suma:	0,01006
		100%	5,00%		Odchylenie standardowe portfela:	10,03%

Ryzyko

c) Macierz współczynników korelacji (dane)

(PIERWIASTEK(G9))

A	B	C	D	E	F	G
A	1					
B	0,03	1				
C	0,50	0,32	1			
D	0,35	0,13	0,40	1		
E	0,31	0,22	0,48	0,34	1	
F	0,30	0,20	0,10	-0,16	0,04	1

Inwestor chce zminimalizować ryzyko określane jako odchylenie standardowe stopy zwrotu portfela.

Minimalna stopa zwrotu portfela – co najmniej 5%.

d) Zastosowane wyrażenia (formuły) Excela:

Komórka	Opis
B3:B8	stopy zwrotu (dane)
C3:C8	Udziały początkowe
C10	=SUMA(C3:C8)
D3	=B3*C3
D10	= SUMA(D3:D8), stopa zwrotu portfela [%]
E2	=C3^2
F3:F8	odchylenie standardowe stóp zwrotu (dane)
G3	=E3*F3^2+E4*F4^2+E5*F5^2+E6*F6^2+E7*F7^2+E8*F8^2
G4	=2*(C3*C4*F3*F4*B14+C3*C5*F3*F5*B15+C3*C6*F3*F6*B16+C3*C7*F3*F7*B17+C3*C8*F3*F8*B18)
G5	=2*(C4*C5*F4*F5*C15+C4*C6*F4*F6*C16+C4*C7*F4*F7*C17+C4*C8*F4*F8*C18)
G6	=2*(C5*C6*F5*F6*D16+C5*C7*F5*F7*D17+C5*C8*F5*F8*D18)
G7	=2*(C6*C7*F6*F7*E17+C6*C8*F6*F8*E18)
G8	=2*C7*C8*F7*F8*F18
G9	=SUMA(G3:G8)
G10	=PIERWIASTEK(SUMA(G3:G8)), odchylenie standardowe portfela [%]
B11:G18	Macierz współczynników korelacji – symetryczna (dana)

e) Elementy zadania decyzyjnego:

komórka celu – \$G\$10,

równa – Min,

Warunki ograniczające: \$C\$3:\$C\$8>=0; \$D\$10=5%.

komórki zmieniane – \$C\$3:\$C\$8,

Tabela 2.3.36

Raport wyników z dodatku Solver programu Excel

Komórka celu (min)			
Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$G\$10	Odchylenie standardowe portfela: według wyrażeń	9,81%	10,03%
Komórki decyzyjne			
Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$C\$3	A Udziały	45,0%	41,0%
\$C\$4	B Udziały	22,0%	21,7%
\$C\$5	C Udziały	6,0%	18,4%
\$C\$6	D Udziały	9,0%	5,3%
\$C\$7	E Udziały	16,0%	13,7%
\$C\$8	F Udziały	2,0%	0,0%

Microsoft Excel 8.0 Raport wyników
 Arkusz: [S265.xls]Roboczy
 Raport utworzony: 2007-09-10 18:32:16

Źródło: Opracowanie własne według danych z tab. 2.3.35.

Solver zaproponował procenty udziałów w akcjach spółek i jednocześnie utrzymał minimalną stopę zwrotu portfela zainwestowania, tj. 5%. Początkowo była to wartość 4,76%.

Przykład

Zakłady samochodowe montujące samochody ciężarowe, pożarnicze, kabiny, podwozia stosują typowe podzespoły: silniki, skrzynie biegów, układ kierowniczy, układ hamowania, ogrzewanie, koła. Wielkość produkcji jest ograniczona zapasami. Firmę interesuje produkcja przynosząca największy zysk, który zależny jest od promocji i zmniejsza się wraz z liczbą wyprodukowanych pojazdów.

Tabela 2.3.37

Maksymalizacja zysku z produkcji pojazdów

A	B	C	D	E	F	G	H	I	1
		Plan	C	P	K	PC	PP		2
Nazwa podzespołu	Stan mag.	Wykorzystanie	247	53	300	98	2		3
silnik	600	400	1	1	0	1	1		4
skrzynia biegów	400	400	1	1	0	1	1		5
układ kierowniczy	700	700	1	1	1	1	1		6
układ hamowania	450	400	1	1	0	1	1		7
ogrzewanie	600	600	1	1	1	0	0	Wskaźnik	8
koła	3 000	2 690	7	5	0	7	5	0,9	9
		Marża [tys. zł]	35	30	25	15	10		10
		Zysk [tys. zł]	4 983	1 069	4 240	929	19		11
		Zysk całkowity [tys. zł]	11 240						12

C – samochody ciężarowe,
 P – samochody pożarnicze,
 K – kabiny,
 PC – podwozia do samochodów ciężarowych,
 PP – podwozia do sam. pożarniczych.
 Wskaźnik: malejący wskaźnik zwrotu.
 Stan magazynowy – zapas.
 Zysk ze sprzedaży produktów [zł]: np. =D10*D3^I\$9
 Przez przyjęcie w zysku potęgi model staje się nieliniowy i z tego względu zdejmujemy domyślną opcję (*Przyjmij model liniowy*).
 Cel: maksymalizacja zysku przy ustalonym stanie magazynu i obniżającym się wskaźniku zwrotu.
 Zadanie decyzyjne:
 komórka docelowa – D12,
 ograniczenia – C4:C9<=B4:B9,
 komórki zmienne – D3:H3, D3:H3=int,
 int – całkowitoliczbowe – D3:H3>=0

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

Tabela 2.3.38

Raport wyników z dodatku Solver programu Excel

Komórka celu (maks)

Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$D\$12	Zysk całkowity [tys. zł]	0	11 240
Komórki decyzyjne			
Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$D\$3	Wykorzystanie C	0	247
\$E\$3	Wykorzystanie P	0	53
\$F\$3	Wykorzystanie K	0	300
\$G\$3	Wykorzystanie PC	0	98
\$H\$3	Wykorzystanie PP	0	2

Microsoft Excel 8.0 Raport wyników
 Arkusz: [S252.xls]Roboczy
 Raport utworzony: 2007-09-11 14:34:25

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

W raporcie podana jest propozycja wielkości produkcji wyrobów oraz określony całkowity zysk w wysokości 11 240 tys. zł.

P r z y k ł a d

Poszukujemy takiej obsady tygodniowej stanowisk, aby zminimalizować koszty płac przy zapewnieniu wymaganego zatrudnienia na stanowiskach. Pracownik jest zatrudniony przez pięć kolejnych dni, a dwa dni odpoczywa. Naszym celem jest zapewnienie odpowiedniej liczby pracowników najtańszym kosztem. Stawka pracownika wynosi np. 50 zł/dzień.

Tabela 2.3.39

Minimalizacja tygodniowego kosztu płac

A	B	C	D	E	F	G	H	I	J	K	
	Dni wolne	Obsada		N	Pn	Wt	Sr	Cz	Pt	So	
A	niedziela, poniedziałek	2		0	0	1	1	1	1	1	1
B	poniedziałek, wtorek	5		1	0	0	1	1	1	1	4
C	wtorek, środa	3		1	1	0	0	1	1	1	5
D	środa, czwartek	5		1	1	1	0	0	1	1	6
E	czwartek, piątek	2		1	1	1	1	0	0	1	7
F	piątek, sobota	3		1	1	1	1	1	0	0	8
G	sobota, niedziela	3		0	1	1	1	1	1	0	9
				18	16	15	15	16	18	17	10
	Wymagane zatrudnienie na stanowiskach			18	16	15	15	15	17	17	11
	Stawka pracownika/dzień	50 zł									12
	Koszt tygodniowy	5 750 zł									13

Zadanie decyzyjne:

komórka docelowa – C13,

ograniczenia – C3:C9>=0,

E10:K10>=E11:K11

komórki zmieniane – C3:C9,

C3:C9=int, całkowitoliczbowe.

Liczba pracowników każdego dnia musi być większa lub równa wymaganiom.

Koszt tygodniowy stanowi iloczyn stawki i sumy pracowników w tygodniu:

$C13=C12*SUMA(E10:K10)$

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

Tabelę inicjujemy wartością zero w obsadzie oraz wpisaniem stawki i wymaganego zatrudnienia na stanowiskach.

Tabela 2.3.40

Raport wyników z dodatku Solver programu Excel

Komórka celu (min)

Komórka	Nazwa	Wartość początkowa [zł]	Wartość końcowa [zł]
\$C\$13	Koszt tygodniowy	0	5 750
Komórki decyzyjne			
Komórka	Dni/Obsada	Wartość początkowa	Wartość końcowa
\$C\$3	niedziela, poniedziałek	0	2
\$C\$4	poniedziałek, wtorek	0	5
\$C\$5	wtorek, środa	0	3
\$C\$6	środa, czwartek	0	5
\$C\$7	czwartek, piątek	0	2
\$C\$8	piątek, sobota	0	3
\$C\$9	sobota, niedziela	0	3

Microsoft Excel 8.0 Raport wyników

Arkusz: [S252.xls]Roboczy1

Raport utworzony: 2007-09-11 16:49:04

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

Z raportu wynika, że minimalny koszt płac tygodniowy wynosi 5750 zł.

2.3.6. Pytania

1. Podaj postać programu liniowego (funkcja celu, ograniczenia, warunki brzegowe).
2. Napisz macierz współczynników i wektory kolumnowe postaci macierzowej zadania decyzyjnego.
3. Dany jest przykład macierzy współczynników i wektory. Sporządź pierwszą tablicę bazową metody simpleks.

$$\mathbf{A} = \begin{bmatrix} 2 & 6 \\ 3 & 2 \\ 4 & 9 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 250 \\ 300 \\ 700 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{c} = [5 \ 20 \ 0 \ 0 \ 0], \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

4. Podaj różnicę między kryterium k_j a g_j .
5. Przedstaw postać każdej kolejnej iteracji tablic simpleksowych.
6. Na podstawie podanego przykładu przedstaw metodę geometryczną rozwiązywania zadań programowania liniowego, mając dwie zmienne.

Biuro architektoniczne wykonuje projekty indywidualne domów jednorodzinnych w dwóch odmianach: parterowe P_1 , z poddaszem mieszkalnym P_2 . Zatrudnia na pełnym etacie architektów, konstruktorów, instalatorów oraz weryfikatora. Liczba zatrudnionych osób przy pracy jednozmianowej (176 godz./miesiąc), zatrudnienie osób, pracochłonności jednostkowe w godz. oraz ceny projektów podaje tab. 2.3.41. Biuro dąży do maksymalizacji przychodu.

Tabela 2.3.41

Pracochłonności jednostkowe, obsady oraz ceny projektów

Projekt \ Wykonawcy	P_1	P_2	Obsada
Architekt	50	70	2
Konstruktor	90	80	3
Instalator	120	110	3
Weryfikator	40	30	1
Cena projektu [zł]	20 000	15 000	

Źródło: Opracowanie własne; dane umowne

7. Napisz procedurę postępowania do wyznaczenia rozwiązania optymalnego przy zastosowaniu dodatku Solver programu Excel.

2.3.7. Zadania

1. Warsztat ślusarski wykonuje trzy typy bram ogrodzeniowych B_1, B_2, B_3 z kątownika oraz siatki stalowej. Zapotrzebowanie jednostkowe materiałów, zapasy magazynowe oraz ceny podaje tabela 2.3.42. Właściciel warsztatu dąży do maksymalnego przychodu. Ułóż liniowe zadanie decyzyjne (LZD).

Tabela 2.3.42

Materialożłonność jednostkowa, zapasy oraz ceny bram

Brama Materiał	B_1	B_2	B_3	Zapas
Kątownik [m]	15	18	14	200
Siatka [m ²]	5	7	4,5	80
Cena [zł]	1 500	2 000	1 200	

Źródło: Opracowanie własne; dane umowne.

2. Developer zamierza zakupić obiekty do remontu i przeznaczyć je na cele mieszkalne po zbankrutowanych jednostkach:

- firmach O_1 ,
- gospodarstwach rolnych O_2 ,
- obiektach kultury O_3 .

Chce wykonać mieszkania standardu M3, M4, M5. W tabeli 2.3.43 podano typy mieszkań, zapotrzebowanie potencjalnych nabywców, koszt zakupu i wyremontowania oraz ceny mieszkań. Developer chce uzyskać maksymalny przychód ze sprzedaży mieszkań. Ułóż liniowe zadanie decyzyjne.

Tabela 2.3.43

Informacje wejściowe do maksymalizacji przychodu ze sprzedaży mieszkań

Obiekt Mieszkanie	O_1	O_2	O_3	Zapotrzebowanie [szt.]	Cena mieszkania [tys. zł]
M3	x_{11}	x_{12}	x_{13}	35	250
M4	x_{21}	x_{22}	x_{23}	30	300
M5	x_{31}	x_{32}	x_{33}	20	350
Koszt zakupu i remontu [tys. zł]	5 000	4 000	8 000		

Źródło: Opracowanie własne; dane umowne.

3. Zakład ogrodniczy do własnej produkcji kwiatów wykonuje cztery mieszanki torfowo-nawozowe składające się z torfu, mielonego siana, ziemi łąkowej. Zakupuje składniki w workach po 50 kg w cenach:

- torf – 15 zł,
- siano mielone – 5 zł,
- ziemia łąkowa – 3 zł.

Posiada siano mielone, które ze względu na warunki przechowywania musi w pełni wykorzystać, torfu i ziemi może zakupić więcej niż to wynika z potrzeb.

Proporcje jednostkowe wymienionych składników na 1 kg mieszanki podaje tab. 2.3.44. Celem zakładu jest minimalizacja kosztu wytworzenia mieszanek do doniczek kwiatowych z pelargoniami (P), begoniami (B), stokrotkami (S), kaktusami (K). Ceny sprzedaży doniczek z kwiatami podano w tabeli.

Ułóż liniowe zadanie decyzyjne.

Tabela 2.3.44

Informacje wejściowe do minimalizacji kosztu wytworzenia mieszanek

Mieszanka [kg] Składniki	P	B	S	K	Liczba worków
Torf	0,5	0,6	0,4	0,1	50
Mielone siano	0,2	0,1	0,3	0,5	25
Ziemia łąkowa	0,3	0,3	0,3	0,4	50
Cena doniczki [zł/kg]	5	7	3	4	

Źródło: Opracowanie własne; dane umowne.

4. Do podanych opisów sformułuj zadania decyzyjne programowania liniowego.

a) Hobbyista chomików oprócz suchej karmy, wody w pojemniku podaje również preparaty witaminowe oraz wzmacniające zawierające niezbędne komponenty: K_1 , K_2 , K_3 . Komponent K_1 należy podawać co najmniej po 20 g, a komponenty K_2 , K_3 w dawkach codziennych nie większych niż 25 g na chomika. Cenę preparatów za 100 g podaje tab. 2.3.45. Zamiarem hobbyisty jest minimalizacja kosztów pielęgnacji pojedynczego chomika.

Tabela 2.3.45

Materiałochłonność jednostkowa oraz ceny preparatów

Preparaty [g] Komponenty	P_1	P_2	P_3
K_1	5	8	5
K_2	8	6	15
K_3	12	10	10
Cena preparatu [100 g/zł]	5	4	6

Źródło: Dane umowne.

b) Dekarz wykonuje zamocowania rynien. Wycina z płaskownika elementy E_1 , E_2 . Pracę może wykonać trzema sposobami, uzyskując liczbę elementów podanych w tab. 2.3.46. Na zamocowanie rynny składają się dwa elementy E_1 oraz jeden element E_2 . Zapas płaskowników wynosi 100. Zamierzaniem dekarza jest uzyskanie maksymalnej liczby kompletów.

Tabela 2.3.46

Liczba sztuk uzyskanych elementów przy danej kombinacji

Sposoby Elementy	S_1	S_2	S_3
E_1	4	5	3
E_2	6	4	7

Źródło: Opracowanie własne; dane umowne.

c) Spawacz wykonuje pięć typów bramek ogrodzeniowych, wykorzystując w tym celu tlen i acetylen (ilość tlenu podano w tab. 2.3.47). Acetylen doprowadzany jest do palnika w ilości 0,5 tlenu. Spawacz dysponuje trzema butlami tlenu oraz dwoma butlami acetyleny o pojemnościach po 100 litrów gazu. Spawacz dąży do maksymalizacji przychodu ze sprzedaży bramek, których ceny podano w tabeli.

Tabela 2.3.47

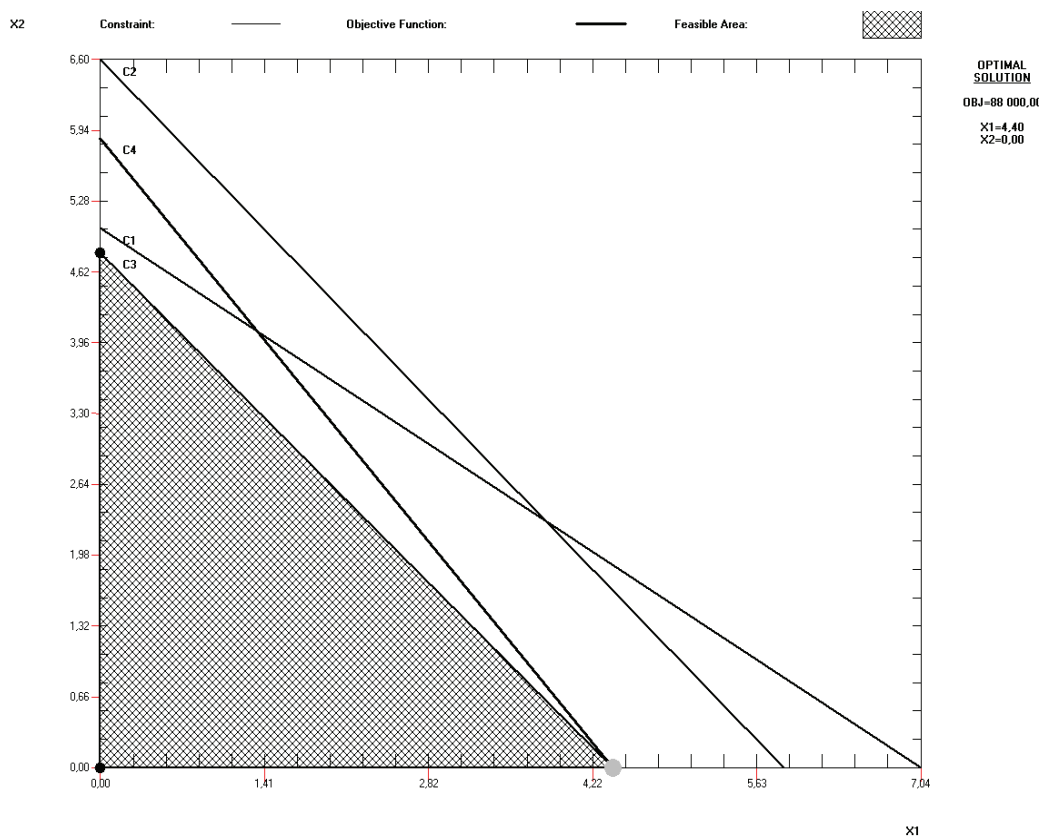
Informacje wejściowe do maksymalizacji przychodu ze sprzedaży bramek

Bramki Gazy	B_1	B_2	B_3
Tlen [l]	10	15	12
Acetylen [l] (oblicz)			
Ceny bramek [zł]	250	300	200

Źródło: Opracowanie własne; dane umowne.

5. Rozwiąż zadanie programowania liniowego metodą geometryczną, manualnie bazując na danych z pytania 6. i tab. 2.3.41.

6. Zastosuj moduł (LP) pakietu WinQSB w celu znalezienia rozwiązania optymalnego według danych z pytania 6. i tab. 2.3.41 oraz porównaj wyniki z rozwiązaniem manualnym. Ponadto porównaj efekt swej pracy z propozycją wykresu 2.3.5 rozwiązania, uzyskaną przy zastosowaniu modułu (LP) przez autorów.



Źródło: Opracowanie własne; moduł (LP), program WinQSB.

Wykres 2.3.5. Rozwiązanie graficzne problemu decyzyjnego z pytania 6.

7. Dane jest następujące zadanie programowania liniowego:

$$FC: 18x_1 + 9x_2 \rightarrow \min,$$

$$C_1: x_1 + 4x_2 \leq 100, \quad b_1 = 100,$$

$$C_2: 2x_1 + 4x_2 \leq 80, \quad b_2 = 80,$$

$$C_3: x_1 + x_2 = 40, \quad b_3 = 40,$$

$$x_1, x_2 \geq 0.$$

Rozwiąż to zadanie metodą geometryczną oraz zaznacz na wykresie, jak zmienia się optymalne rozwiązanie, jeżeli parametr b_2 należy do $\langle 80, 110 \rangle$.

8. Zastosuj dodatek Solver programu Excel do rozwiązania zadania siódmego, lecz bez zmienności wyrazu wolnego b_2 . Sprawdź, czy efekt pracy jest zgodny z rozwiązaniem podanym w tab. 2.3.48.

Tabela 2.3.48

Zdefiniowanie arkusza kalkulacyjnego do minimalizacji funkcji celu zadania szóstego

A	B	C	D	E	F
2				x_1	x_2
3	Nazwa	Zasób	Wykorzystanie	40,000002	0
4	b_1	100	40	1	4
5	b_2	80	80	2	4
6	b_3	40	40	1	1
7			Współczynniki	18	9
8	Minimalizacja		720		

Źródło: Opracowanie własne; dodatek Solver programu Excel.

9. Bazując na wcześniejszym przykładzie podanym w podrozdz. 2.3.5 i tab. 2.3.49, wykonaj następujące prace w celu zastosowania dodatku Solver programu Excel:

- wpisz wyrażenia w komórki arkusza kalkulacyjnego:
 - suma obsadodni = suma iloczynów obsad i zatrudnienia w tygodniu,
 - koszt tygodniowy: iloczyn stawki pracownika i sumy wymaganego zatrudnienia,
 - suma obsadodni \geq wymagane zatrudnienie;
- zainicjuj obsady jako zerowe oraz rozwiąż zadanie w Solverze,
- porównaj z wynikami w tab. 2.3.49.

Tabela 2.3.49

Informacje wejściowe oraz rozwiązanie problemu decyzyjnego

A	B	C	D	E	F	G	H	I	J	K	1
1	Dni wolne	Obsada		N	Pn	Wt	Śr	Cz	Pt	So	2
2	niedziela, poniedziałek	4		0	0	1	1	1	1	1	3
3	poniedziałek, wtorek	3		1	0	0	1	1	1	1	4
4	wtorek, środa	6		1	1	0	0	1	1	1	5
5	środa, czwartek	3		1	1	1	0	0	1	1	6
6	czwartek, piątek	4		1	1	1	1	0	0	1	7
7	piątek, sobota	4		1	1	1	1	1	0	0	8
8	sobota, niedziela	1		0	1	1	1	1	1	0	9
9	Suma obsadodni			20	18	16	16	18	17	20	10
10	Wymagane zatrudnienie			20	18	16	16	18	16	20	11
11	Stawka pracownika/dzień [zł/d]	40									12
12	Koszt tygodniowy [zł]	5 000									13

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

2.4. Algorytm simpleks – sposób według elementu rozwiązującego

Postać standardowa

Zamiana nierówności na równości, gdy:

\leq , to np. $a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + x_{n+1} = b_1$,

\geq , to np. $a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n - x_{n+1} = b_1$,

$x_{n+1} \geq 0$.

W pierwszym przypadku układ równań zadania decyzyjnego jest następujący:

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + x_{n+1} = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + x_{n+2} = b_2,$$

.....

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + x_{n+m} = b_m,$$

$$x_1, x_2, \dots, x_n, x_{n+1}, x_{n+2}, \dots, x_{n+m} \geq 0,$$

$$z = c_1x_1 + c_2x_2 + \dots + c_nx_n + 0x_{n+1} + \dots + 0x_{n+2} + \dots + 0x_{n+m} \rightarrow \max.$$

O postaci pierwszego rozwiązania oraz kolejnych algorytmu simpleks wspomnieliśmy w podrozdz. 2.3.1 dotyczącym rozwiązywania programów liniowych. W tym miejscu skupimy się na dwóch sposobach manualnego rozwiązywania zadań programowania liniowego, mianowicie:

– określeniu rozwiązania optymalnego metodą algebraiczną nazwane tu sposobem według elementu rozwiązującego¹⁵⁹,

– uzyskaniu rozwiązania z zastosowaniem rachunku macierzowego¹⁶⁰.

Pomocna w tym wypadku jest idea algorytmu simpleks, zaprezentowana w podrozdz. 2.3.1.

¹⁵⁹ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne...*, op.cit., rozdz. 2.4; W. Sikora (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii...*, op.cit., rozdz. 1.4.

¹⁶⁰ K. Kukuła (red.), *Badania operacyjne w przykładach i zadaniach...*, op.cit., rozdz. 1.4.

Rozwiązanie bazowe

$$x_1, x_2, \dots, x_n = 0,$$

$$x_{n+1} = b_1, x_{n+2} = b_2, \dots, x_{n+m} = b_m.$$

Tabela 2.4.1

I tablica simpleksowa – rozwiązanie bazowe

			x_1	x_2	...	x_n	x_{n+1}	x_{n+2}	...	x_{n+m}
			c_1	c_2	...	c_n	c_{n+1}	c_{n+2}	...	c_{n+m}
c	x_b	b	a₁	a₂	...	a_n	a_{n+1}	a_{n+2}	...	a_{n+m}
c_{n+1}	x_{n+1}	b_1	a_{11}	a_{12}	...	a_{1n}	1	0	...	0
c_{n+2}	x_{n+2}	b_2	a_{21}	a_{22}	...	a_{2n}	0	1	...	0
...	0	0	...	0
c_{n+m}	x_{n+m}	b_m	a_{m1}	a_{m2}	...	a_{mn}	0	0	...	1
			g_1	g_2	...	g_n	g_{n+1}	g_{n+2}	...	g_{n+m}

Źródło: Opracowanie własne na podstawie: M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003, s. 47.

Etapy postępowania

– określenie elementów wiersza kryterialnego:

$$g_1 = (c_{n+1} a_{11} + c_{n+2} a_{21} + \dots + c_{n+m} a_{m1}) - c_1,$$

.....

$$g_{n+m} = (c_{n+1} 0 + c_{n+2} 0 + \dots + c_{n+m} 1) - c_m,$$

– sprawdzenie rozwiązania bazowego, gdy $z \rightarrow \max$:

$$g_1 \leq 0, g_2 \leq 0, \dots, g_{n+m} \leq 0,$$

– wymiana wektora bazowego $g_n \leq 0$:

kolumna rozwiązująca: $\min g_i$

wiersz rozwiązujący: $\min b_j/a_{ij}$

element rozwiązujący (na przecięciu kolumny i wiersza rozwiązującego),

tworzenie kolejnej tablicy.

Przykład

$$\begin{aligned} 6x_1 + 2x_2 &\leq 24, \\ x_1 + 8x_2 &\leq 32, \\ x_1, x_2 &\geq 0, \\ z = 40x_1 + 20x_2 &\rightarrow \max, \end{aligned}$$

– postać standardowa:

$$\begin{aligned} 6x_1 + 2x_2 + x_3 &= 24, \\ x_1 + 8x_2 + x_4 &= 32, \\ x_1, x_2, x_3, x_4 &\geq 0, \\ z = 40x_1 + 20x_2 + 0x_3 + 0x_4 &\rightarrow \max. \end{aligned}$$

Tabela 2.4.2

I tablica simpleksowa

			x_1	x_2	x_3	x_4
			40	20	0	0
c	x_b	b	a_1	a_2	a_3	a_4
0	x_3	24	6	2	1	0
0	x_4	32	1	8	0	1
		g_j	-40	-20	0	0

Źródło: Opracowanie własne w Excelu, gdzie x_b – zmienne bazowe.

rozwiązanie bazowe:

$$x_1, x_2 = 0, x_3 = 24, x_4 = 32,$$

– wiersz kryterialny np.:

$$g_1 = (0 \cdot 6 + 0 \cdot 1) - 40 = -40,$$

– sprawdzenie rozwiązania bazowego $g_n \leq 0$:

$$g_1 \leq 0, g_2 \leq 0, g_3 = 0, g_4 = 0,$$

– wymiana wektora bazowego b :

$$\text{kolumna rozwiązująca} - x_1 \quad \min [(-40), (-20)] = -40,$$

$$\text{wiersz rozwiązujący} - x_4 \quad \min (24/6, 32/1) = 4,$$

$$\text{element rozwiązujący } a_{11} = 6,$$

– tworzenie II tablicy:

Tabela 2.4.3

II tablica simpleksowa

			x_1	x_2	x_3	x_4
			40	20	0	0
c	x_b	b	a_1	a_2	a_3	a_4
40	x_1	4	1	0,33	0,17	0
0	x_4	28	0	7,67	-0,17	1
		g_j	0	-6,8	6,8	0

\downarrow $2/6$ \downarrow $(6 \cdot 0 - 1 \cdot 1)/6$

Źródło: Opracowanie własne w Excelu.

- wymiana wektora bazowego: element rozwiązujący – $a_{22} = 7,67$,
- tworzenie III tablicy:

Tabela 2.4.4

III tablica simpleksowa

			x_1	x_2	x_3	x_4
			40	20	0	0
c	x_b	b	a_1	a_2	a_3	a_4
40	x_1	2,80	1	0	0,17	-0,04
20	x_2	3,65	0	1	-0,02	0,13
		g_j	0	0	6,4	1

Źródło: Opracowanie własne w Excelu.

W przedstawionych obliczeniach przyjęto zaokrąglenia do drugiego miejsca po przecinku i otrzymano przybliżone wyniki:

rozwiązanie primalne: $x_1 = 2,80$, $x_2 = 3,65$,

rozwiązanie dualne: $y_1 = 6,4$, $y_2 = 1$ ¹⁶¹,

funkcja celu: $40 \cdot 2,80 + 20 \cdot 3,65 = 185$.

Alternatywnym rozwiązaniem z większą dokładnością jest zastosowanie modułu (LP) pakietu WinQSB.

¹⁶¹ Rozwiązywanie zagadnień dualnych przedstawiono w podrozdziale 2.8 niniejszej publikacji.

Tabela 2.4.5

Rozwiązanie sumaryczne

09-18-2007	Decision	Solution	Unit Cost or	Total
16:01:58	Variable	Value	Profit C (j)	Contribution
1	X1	2,78	40	111,3
2	X2	3,65	20	73,04
	Objective	Function	(Max.) =	184,35

Źródło: Opracowanie własne, moduł (LP).

Moduł (LP) umożliwia także uzyskanie raportu kombinowanego z podaniem rozwiązania dualnego w „Shadow Price” (Cena dualna)

$$y_1 = 6,52; y_2 = 0,87.$$

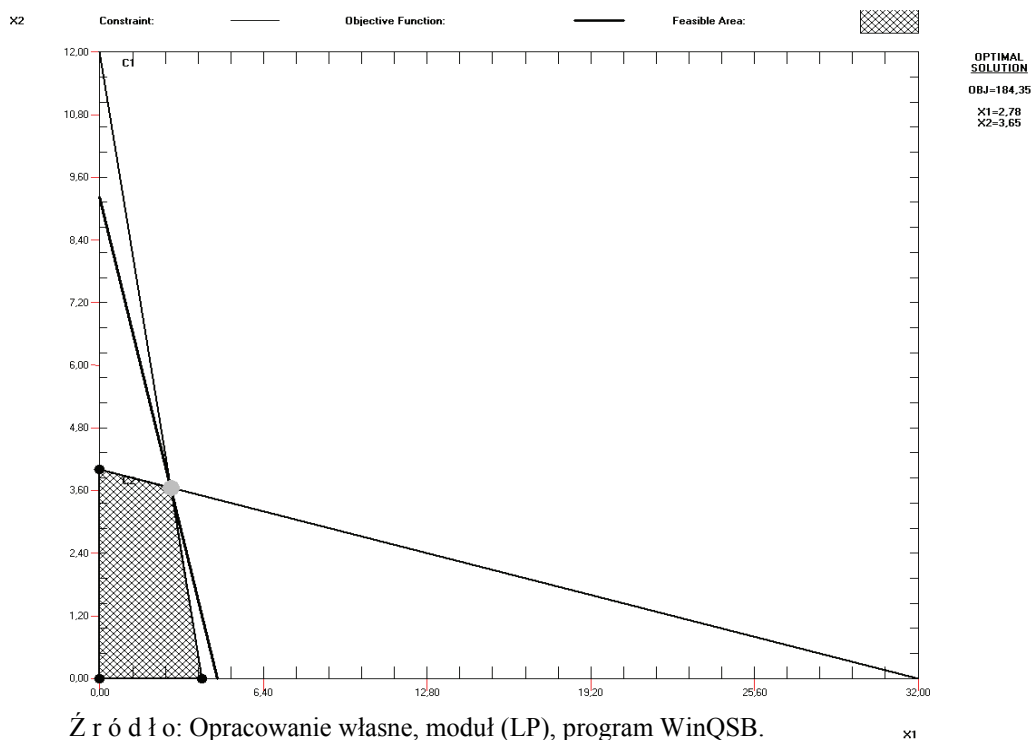
Tabela 2.4.6

Raport kombinowany

	16:08:42		Tuesday	September	18	2007		
	Decision	Solution	Unit Cost or	Total	Reduced	Basis	Allowable	Allowable
	Variable	Value	Profit C (j)	Contribution	Cost	Status	Min. C (j)	Max. C (j)
1	X1	2,78	40	111,3	0	Basic	2,5	60
2	X2	3,65	20	73,04	0	Basic	13,33	320
	Objective	Function	(Max.) =	184,35				
		Left Hand		Right Hand	Slack	Shadow	Allowable	Allowable
	Constraint	Side	Direction	Side	or Surplus	Price	Min. RHS	Max. RHS
1	C1	24	<=	24	0	6,52	8	192
2	C2	32	<=	32	0	0,87	4	96

Źródło: Opracowanie własne, moduł [LP], program WinQSB.

Przyjrzyjmy się rozwiązaniu geometrycznemu problemu w postaci graficznej.



Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Wykres 2.4.1 Rozwiązanie graficzne zadania prymalnego

Otrzymaliśmy wyniki: $x_1^* = 2,78$; $x_2^* = 3,65$, funkcja celu: $FC = 184,35$
 Rozwiązanie problemu graficznie jest identyczne jak w tab. 2.4.5 i 2.4.6.

Przykład

Do podanego opisu sformułujemy zadanie decyzyjne minimalizujące koszt chowu gołębi pocztowych, a następnie określimy rozwiązanie optymalne metodami:

- geometryczną,
- algebraiczną według elementu rozwiązującego.

Prawidłowo karmione ptaki muszą dostawać codziennie dawki preparatów wzmacniających w wysokości przynajmniej: preparat A – 15 j., B – 10 j., preparat C – 20 j. Zawartość poszczególnych preparatów w paszach P_1 oraz P_2 oraz ceny za 1 kg podaje tab. 2.4.7:

Tabela 2.4.7

Dane wejściowe

Pasze	P_1	P_2
Preparat		
A	1	2
B	2	1
C	0	3
Cena [kg/zł]	2	4

Źródło: Opracowanie własne; dane umowne.

Sformułowanie zadania decyzyjnego:

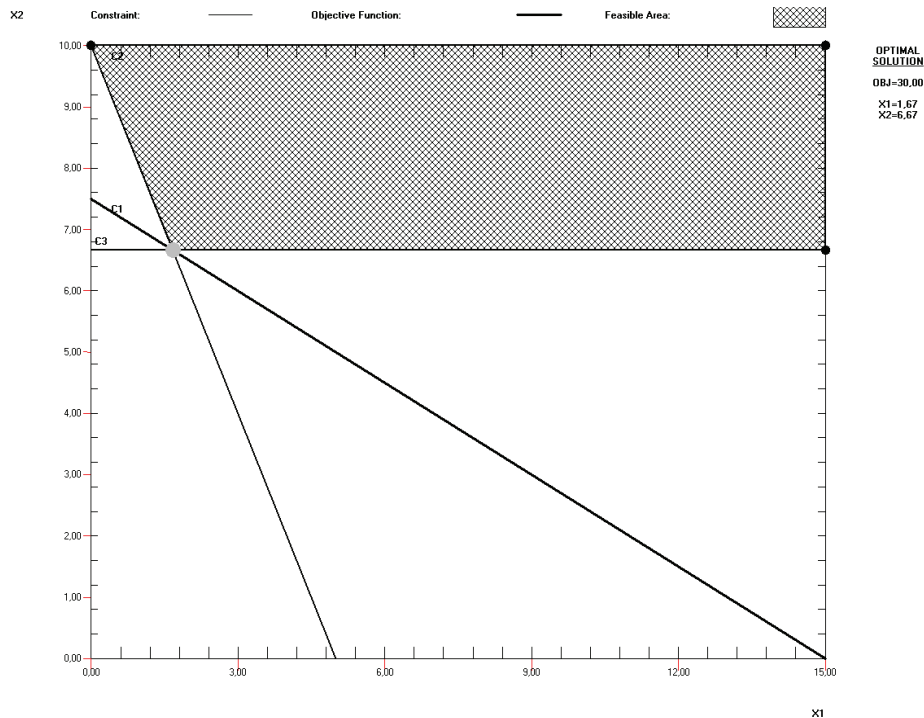
$$FC: 2x_1 + 4x_2 \rightarrow \min,$$

$$C_1: x_1 + 2x_2 \geq 15,$$

$$C_2: 2x_1 + x_2 \geq 10,$$

$$C_3: 3x_2 \geq 20,$$

$$x_1, x_2 \geq 0.$$



Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Wykres 2.4.2. Rozwiązanie graficzne zadania decyzyjnego minimalizującego koszt chowu gołębi pocztowych

A teraz zobaczymy kolejne tablice iteracyjne emitowane przez moduł (LP) programu WinQSB, przy czym:

SC_1, SC_2, SC_3 – zmienne swobodne konieczne dla postaci kanonicznej ograniczeń C_1, C_2, C_3 .

A_{C1}, A_{C2}, A_{C3} – zmienne sztuczne dla wprowadzenia macierzy jednostkowej.

Wprowadzono kryterium według oznaczeń modułu (LP)¹⁶²:

$$k_j = C(j) - \left(\sum C(j)_b a_{ij} \right). \quad (2.4.1)$$

Ratio – udział ilościowy (zasób określany przez program jako RHS dzielony przez odpowiadający współczynnik w kolumnie rozwiązującej – np. $15/2 = 7,5$),

M – zapis dużej wartości (∞) w algorytmie programu.

¹⁶² W. Sikora (red.), *Przykłady i zadania z badań operacyjnych...*, op.cit., s. 51.

Tabela 2.4.8

I tablica simpleksowa (Simplex Tableau – Iteration 1)

		X_1	X_2	S_{C1}	S_{C2}	S_{C3}	A_{C1}	A_{C2}	A_{C3}	RHS	Ratio
Basis	$C(j)$	2,00	4,00	0	0	0	M	M	M		
A_{C1}	M	1,00	2,00	-1,00	0	0	1,00	0	0	15,00	7,50
A_{C2}	M	2,00	1,00	0	-1,00	0	0	1,00	0	10,00	10,00
A_{C3}	M	0	3,00	0	0	-1,00	0	0	1,00	20,00	6,67
	$C(j)-Z(j)$	2,00	4,00	0	0	0	0	0	0	0	
	* Big M	-3,00	-6,00	1,00	1,00	1,00	0	0	0	0	

Źródło: Opracowanie własne, moduł (LP), program Win QSB; gdzie: BASIS – x_b – wektor zmiennych bazowych.

Elementem rozwiązującym jest zatem komórka na przecięciu kolumny X_2 oraz wiersza A_{C3} . Kryterium dla kolumny X_2 wynosi: $4-6M$. Zastępujemy zatem zmienną sztuczną A_{C3} zmienną decyzyjną X_2 i wyznaczamy w podobny sposób kolejne tablice simpleksowe.

Tabela 2.4.9

II tablica simpleksowa (Simplex Tableau – Iteration 2)

		X_1	X_2	S_{C1}	S_{C2}	S_{C3}	A_{C1}	A_{C2}	A_{C3}	RHS	Ratio
Basis	$c_j/C(j)$	2,00	4,00	0	0	0	M	M	M		
A_{C1}	M	1,00	0	-1,00	0	0,67	1,00	0	-0,67	1,67	1,67
A_{C2}	M	2,00	0	0	-1,00	0,33	0	1,00	-0,33	3,33	1,67
X_2	4,00	0	1,00	0	0	-0,33	0	0	0,33	6,67	M
	$C(j)-Z(j)$	2,00	0	0	0	1,33	0	0	-1,33	26,67	
	* Big M	-3,00	0	1,00	1,00	-1,00	0	0	2,00	0	

Źródło: Opracowanie własne, moduł (LP), prognoza WinQSB.

W tej tabeli mamy alternatywny wybór komórek rozwiązujących dla A_{C1} lub A_{C2} , gdyż Ratio jest jednakowe. Program przyjął przykładowo A_{C2} i zastąpił zmienną X_1 .

Tabela 2.4.10

III tablica simpleksowa (Simplex Tableau – Iteration 3)

		X_1	X_2	S_{C1}	S_{C2}	S_{C3}	A_{C1}	A_{C2}	A_{C3}	RHS	Ratio
Basis	$C(j)$	2,00	4,00	0	0	0	M	M	M		
A_{C1}	M	0	0	-1,00	0,50	0,50	1,00	-0,50	-0,50	0	0
X_1	2,00	1,00	0	0	-0,50	0,17	0	0,50	-0,17	1,67	M
X_2	4,00	0	1,00	0	0	-0,33	0	0	0,33	6,67	M
	$C(j)-Z(j)$	0	0	0	1,00	1,00	0	-1,00	-1,00	30,00	
	* Big M	0	0	1,00	-0,50	-0,50	0	1,50	1,50	0	

Źródło: Opracowanie własne, moduł (LP).

Wystąpiły dwie alternatywne kolumny rozwiązujące dla zmiennych swobodnych: S_C2, S_C3 .

Ratio o wartości zero jest w wierszu ze zmienną sztuczną A_C1 , którą program zamienia na zmienną swobodną S_C2 . Przy obliczaniu Ratio wartości ujemne traktowane są jako zero.

Tabela 2.4.11

IV tablica simpleksowa (Simplex Tableau -Iteration 4)

	x_1	x_2	S_C1	S_C2	S_C3	A_C1	A_C2	A_C3		RHS
Basis	C(j)	2,00	4,00	0	0	0	0	0	0	
S_C2	0	0	0	-2,00	1,00	1,00	2,00	-1,00	-1,00	0
x_1	2,00	1,00	0	-1,00	0	0,67	1,00	0	-0,67	1,67
x_2	4,00	0	1,00	0	0	-0,33	0	0	0,33	6,67
	$C(j)-Z(j)$	0	0	2,00	0	0	-2,00	0	0	30,00
	* Big M	0	0	0	0	0	1,00	1,00	1,00	0

Źródło: Opracowanie własne, moduł (LP).

W tej tablicy wszystkie kryteria funkcji celu dążącej do minimum i kryterium k_j są dodatnie, a więc mamy rozwiązanie optymalne: $x_1^* = 1,67$, $x_2^* = 6,67$, co jest zgodne z wcześniejszym rozwiązaniem graficznym (wykres 2.4.2).

2.4.1. Pytania

1. Podaj procedurę znalezienia rozwiązania optymalnego według metody elementu rozwiązującego.
2. Opisz na podanym przykładzie I tablicy simpleksowej sposób wyznaczania elementu rozwiązującego.

Tabela 2.4.12

I tablica simpleksowa

			x_1	x_2	x_3	x_4
			20	30	0	0
c	x_b	b	a_1	a_2	a_3	a_4
0	x_3	24	5	4	1	0
0	x_4	40	2,5	2	0	1

Źródło: Opracowanie własne; dane umowne.

2.4.2. Zadania

1. Dane są następujące zadania decyzyjne programowania liniowego:

Zadanie 1

$$FC: 9x_1 + 8x_2 \rightarrow \max,$$

$$C_1: 3x_1 + x_2 \leq 9,$$

$$C_2: x_1 + 2x_2 \leq 12,$$

$$C_3: x_1, x_2 \geq 0,$$

Zadanie 2

$$FC: 9x_1 + 8x_2 \rightarrow \min,$$

$$3x_1 + x_2 \geq 9,$$

$$x_1 + 2x_2 \geq 12,$$

$$x_1, x_2 \geq 0,$$

– określ postać kanoniczną zadań,
 – rozwiąż zadania metodą geometryczną,
 – określ przykładowo I i II tablicę simpleksową sposobem algebraicznym elementu rozwiązującego z zastosowaniem ułamków zwykłych.

2. Tabela 2.4.13 stanowi sformatowanie zadania decyzyjnego w programie WinQSB. Znajdź rozwiązanie optymalne wykonane manualnie metodą elementu rozwiązującego. Jeśli masz dostęp do modułu (LP) tego programu, sprawdź, czy twoje iteracje są zgodne z rozwiązaniami komputerowymi.

Tabela 2.4.13

Sformatowanie zadania w module (LP) programu WinQSB

Variable -->	X1	X2	Direction	RHS
Minimize	3	9		
C1	2.5	5	>=	100
C2	4.5	9	>=	180
C3	1	3	>=	30
LowerBound	0	0		
UpperBound	M	M		
VariableType	Continuous	Continuous		

Źródło: Opracowanie własne.

W tabeli występuje kropka dziesiętna jako przecinek. Typ danych Continuous – dane rzeczywiste.

2.5. Algorytm simpleks – zastosowanie rachunku macierzowego

Prezentację przykładu rozwiązywania zadań programowania liniowego z zastosowaniem rachunku macierzowego oparto na publikacji K. Kukuły¹⁶³.

Model matematyczny

$$FC: 40x_1 + 30x_2 \rightarrow \max,$$

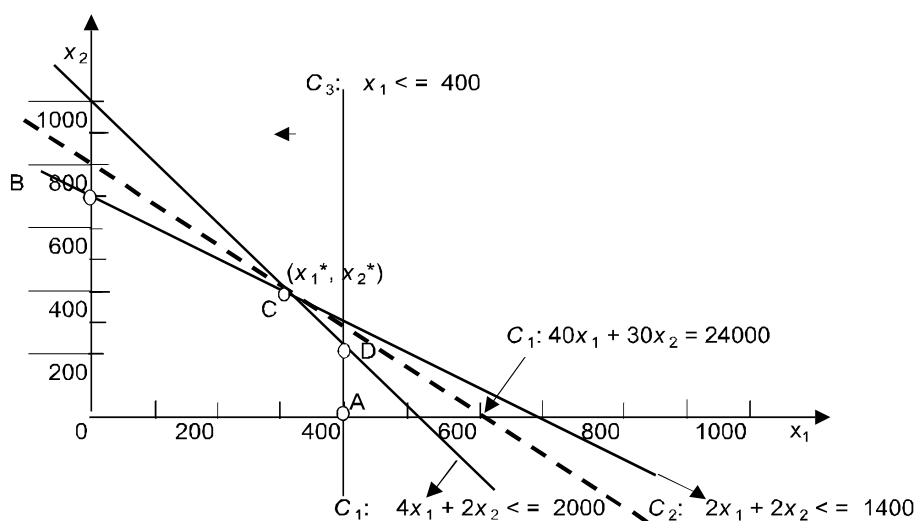
$$C_1: 4x_1 + 2x_2 \leq 2000,$$

$$C_2: 2x_1 + 2x_2 \leq 1400,$$

$$C_3: 3x_1 \leq 1200,$$

$$x_1, x_2 \geq 0.$$

Rozwiązanie metodą geometryczną



Źródło: Opracowanie własne w Excelu, gdzie: $x_1^* \rightarrow x_1^*$, $x_2^* \rightarrow x_2^*$.

Wykres 2.5.1. Graficzne rozwiązanie zadania decyzyjnego

Rozwiązaniem dopuszczalnym jest pięciobok A0BCD. Rozwiązanie dopuszczalne wyznacza punkt C o współrzędnych $x_1^* = 300$, $x_2^* = 400$.

¹⁶³ K. Kukuła (red.), *Badania operacyjne w przykładach i zadaniach...*, op.cit., rozdz. 1.4.

Postać kanoniczna modelu matematycznego

$$FC: 40x_1 + 30x_2 + 0x_3 + 0x_4 + 0x_5 \rightarrow \max,$$

$$C_1: 4x_1 + 2x_2 + 1x_3 + 0x_4 + 0x_5 = 2000,$$

$$C_2: 2x_1 + 2x_2 + 0x_3 + 1x_4 + 0x_5 = 1400,$$

$$C_3: 3x_1 + 0x_2 + 0x_3 + 0x_4 + 1x_5 = 1200,$$

$$x_1, x_2 \geq 0.$$

Iteracje rozwiązania:

Tabela 2.5.1

Pierwsza iteracja – postać macierzowa

		c		
c_b	x_b	A	I	b
	z_j	0		
	c_j - z_j	c		

Źródło: K. Kukuła (red.), *Badania operacyjne w przykładach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa 2004, tab. 39.

Kolejna iteracja: według tab. 2.5.2.

$$z_j = \sum_{i=0}^m a_{ij} c_{bi}, \quad (2.5.1.)$$

Rozwiązanie rachunkiem macierzowym

Tabela 2.5.2

I tablica simpleksowa

c_b	c_j	40	30	0	0	0	Rozwiązanie b_j
	Zmienne bazowe - x_b	x₁	x₂	x₃	x₄	x₅	
0	x₃	4	2	1	0	0	2 000
0	x₄	2	2	0	1	0	1 400
0	x₅	3	0	0	0	1	1 200
	z_j	0	0	0	0	0	0
	k_j = c_j - z_j	40	30	0	0	0	

Kolumna rozwiązująca – zmienna x_1 .

Wiersz rozwiązujący – zmienna x_5 .

Element rozwiązujący – 3 (cyfra półgruba).

Źródło: Opracowanie własne w Excelu.

Tabela 2.5.3

II tablica simpleksowa

c_b	c_j	40	30	0	0	0	Rozwiązanie b_i
	Zmienne bazowe x_b	x_1	x_2	x_3	x_4	x_5	
0	x_3	0	2	1	0	-4/3	400
0	x_4	0	2	0	1	-2/3	600
40	x_1	1	0	0	0	1/3	400
	z_j	40	0	0	0	40/3	16 000
	$k_j = c_j - z_j$	0	30	0	0	-40/3	

Źródło: Opracowanie własne w Excelu.

Elementy macierzowe II tablicy simpleksowej (tab. 2.5.2):

– określenie macierzy odwrotnej dla zmiennych bazowych X_3, X_4, X_1

$$\mathbf{B}_2 = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{bmatrix}, \quad \det(\mathbf{B}_2) = 1 \cdot \begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = 3, \quad \mathbf{B}_2^{-1} = \frac{1}{3} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ -4 & -2 & 1 \end{bmatrix}^T,$$

$$\mathbf{B}_2^{-1} = \frac{1}{3} \begin{bmatrix} 3 & 0 & -4 \\ 0 & 3 & -2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -4/3 \\ 0 & 1 & -2/3 \\ 0 & 0 & 1/3 \end{bmatrix}.$$

Otrzymaliśmy więc elementy odpowiadające kolumnom X_3, X_4, X_5 w II tablicy simpleksowej. W następnych działaniach wykorzystano funkcję excela [ILOZYN.MACIERZ(...;...)]. Program zamienia ułamki zwykle współczynników macierzy na dziesiętne.

– iloczyny z macierzą odwrotną:

$$\mathbf{B}_2^{-1}\mathbf{A} = \begin{bmatrix} 1 & 0 & -1,33 \\ 0 & 1 & -0,67 \\ 0 & 0 & 0,33 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 2 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B}_2^{-1}\mathbf{b} = \begin{bmatrix} 1 & 0 & -1,33 \\ 0 & 1 & -0,67 \\ 0 & 0 & 0,33 \end{bmatrix} \cdot \begin{bmatrix} 2000 \\ 1400 \\ 1200 \end{bmatrix} = \begin{bmatrix} 400 \\ 600 \\ 400 \end{bmatrix},$$

$$\mathbf{c}_b^T \mathbf{B}_2^{-1} \mathbf{A} = [0 \ 0 \ 40] \cdot \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 1 & 0 \end{bmatrix} = [40 \ 0], \quad \mathbf{c}_b^T \mathbf{B}_2^{-1} = [0 \ 0 \ 40] \cdot \begin{bmatrix} 1 & 0 & -1,33 \\ 0 & 1 & -0,67 \\ 0 & 0 & 0,33 \end{bmatrix} = [0 \ 0 \ 13,33],$$

– funkcja celu:

$$\mathbf{c}_b^T \mathbf{B}_2^{-1} \mathbf{b} = [0 \ 0 \ 13,33] \cdot \begin{bmatrix} 2000 \\ 1400 \\ 1200 \end{bmatrix} \approx 16000.$$

Kolumna rozwiązująca – zmienna x_2 .

Wiersz rozwiązujący – zmienna x_3 .

Element rozwiązujący – 2 (cyfra półgruba).

Tabela 2.5.4

III tablica simpleksowa

\mathbf{c}_b	\mathbf{c}_j	40	30	0	0	0	Rozwiązanie b_i
	Zmienne bazowe - \mathbf{x}_b	x_1	x_2	x_3	x_4	x_5	
30	x_2	0	1	0,5	0	-0,67	200
0	x_4	0	0	-1	1	0,67	200
40	x_1	1	0	0	0	0,33	400
	\mathbf{z}_j	40	30	15	0	-6,67	22 000
$k_j =$	$\mathbf{c}_j - \mathbf{z}_j$	0	0	-15	0	6,67	

Źródło: Opracowanie własne w Excelu.

Elementy macierzowe III tablicy simpleksowej:

– określenie macierzy odwrotnej dla zmiennych bazowych x_2, x_4, x_1 :

$$\mathbf{B}_3 = \begin{bmatrix} 2 & 0 & 4 \\ 2 & 1 & 2 \\ 0 & 0 & 3 \end{bmatrix}, \quad \det(\mathbf{B}_3) = 2 \cdot \begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = 6, \quad \mathbf{B}_3^{-1} = \frac{1}{6} \begin{bmatrix} 3 & -6 & 0 \\ 0 & 6 & 0 \\ -4 & 4 & 2 \end{bmatrix}^T,$$

$$\mathbf{B}_3^{-1} = \frac{1}{6} \begin{bmatrix} 3 & 0 & -4 \\ -6 & 6 & 4 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0,5 & 0 & -0,67 \\ -1 & 1 & 0,67 \\ 0 & 0 & 0,33 \end{bmatrix},$$

– iloczyny z macierzą odwrotną:

$$\mathbf{B}_3^{-1}\mathbf{A} = \begin{bmatrix} 0,5 & 0 & -0,67 \\ -1 & 1 & 0,67 \\ 0 & 0 & 0,33 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 2 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B}_3^{-1}\mathbf{b} = \begin{bmatrix} 0,5 & 0 & -0,67 \\ -1 & 1 & 0,67 \\ 0 & 0 & 0,33 \end{bmatrix} \cdot \begin{bmatrix} 2000 \\ 1400 \\ 1200 \end{bmatrix} = \begin{bmatrix} 200 \\ 200 \\ 400 \end{bmatrix},$$

$$\mathbf{c}_b^T \mathbf{B}_3^{-1} \mathbf{A} = [30 \ 0 \ 40] \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} = [40 \ 30], \quad \mathbf{c}_b^T \mathbf{B}_3^{-1} \mathbf{b} = [30 \ 0 \ 40] \cdot \begin{bmatrix} 0,5 & 0 & -0,67 \\ -1 & 1 & 0,67 \\ 0 & 0 & 0,33 \end{bmatrix} = [15 \ 0 \ 6,67],$$

– funkcja celu:

$$\mathbf{c}_b^T \mathbf{B}_3^{-1} \mathbf{b} = [15 \ 0 \ -6,67] \cdot \begin{bmatrix} 2000 \\ 1400 \\ 1200 \end{bmatrix} = 22000.$$

Kolumna rozwiązująca – zmienna x_5 .

Wiersz rozwiązujący – zmienna x_4 .

Element rozwiązujący – 6,67 (cyfra pogrubiona).

Tabela 2.5.5

IV tablica simpleksowa

\mathbf{c}_b	\mathbf{c}_j	40	30	0	0	0	Rozwiązanie b_{IV}
	Zmienne bazowe - \mathbf{x}_b	x_1	x_2	x_3	x_4	x_5	
30	x_2	0	1	-0,5	1	0	400
0	x_5	0	0	-1,5	1,5	1	300
40	x_1	1	0	0,5	-0,5	0	300
	z_j	40	30	5	10	0	24 000
	$\mathbf{c}_j - z_j$	0	0	-5	-10	0	

Źródło: Opracowanie własne w Excelu.

Tabela 2.5.5 stanowi końcową iterację. Rozwiązanie optymalne: $x_1^* = 300$, $x_2^* = 400$.

Elementy macierzowe IV tablicy simpleksowej:

– określenie macierzy odwrotnej dla zmiennych bazowych x_2, x_5, x_1 :

$$\mathbf{B}_4 = \begin{bmatrix} 2 & 0 & 4 \\ 2 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix}, \quad \det(\mathbf{B}_4) = \begin{vmatrix} 2 & 0 & 4 \\ 2 & 0 & 2 \\ 0 & 1 & 3 \end{vmatrix} = 4, \quad \mathbf{B}_4^{-1} = \frac{1}{4} \begin{bmatrix} -2 & -6 & 2 \\ 4 & 6 & -2 \\ 0 & 4 & 0 \end{bmatrix}^T,$$

do obliczenia wyznacznika macierzy \mathbf{B}_4 zastosowano regułę Sarrusa:

$$\mathbf{B}_4^{-1} = \frac{1}{4} \begin{bmatrix} -2 & 4 & 0 \\ -6 & 6 & 4 \\ 2 & -2 & 0 \end{bmatrix} = \begin{bmatrix} 0,5 & 1 & 0 \\ -1,5 & 1,5 & 1 \\ 0,5 & -0,5 & 0 \end{bmatrix},$$

– iloczyny z macierzą odwrotną:

$$\mathbf{B}_4^{-1} \mathbf{A} = \begin{bmatrix} -0,5 & 1 & 0 \\ -1,5 & 1,5 & 1 \\ 0,5 & -0,5 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 2 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B}_3^{-1} \mathbf{b} = \begin{bmatrix} -0,5 & 1 & 0 \\ -1,5 & 1,5 & 1 \\ 0,5 & -0,5 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2000 \\ 1400 \\ 1200 \end{bmatrix} = \begin{bmatrix} 400 \\ 300 \\ 300 \end{bmatrix},$$

$$\mathbf{c}_b^T \mathbf{B}_4^{-1} \mathbf{A} = [30 \ 0 \ 40] \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} = [40 \ 30], \quad \mathbf{c}_b^T \mathbf{B}_4^{-1} \mathbf{b} = [30 \ 0 \ 40] \cdot \begin{bmatrix} -0,5 & 1 & 0 \\ -1,5 & 1,5 & 1 \\ 0,5 & -0,5 & 0 \end{bmatrix} = [5 \ 10 \ 0],$$

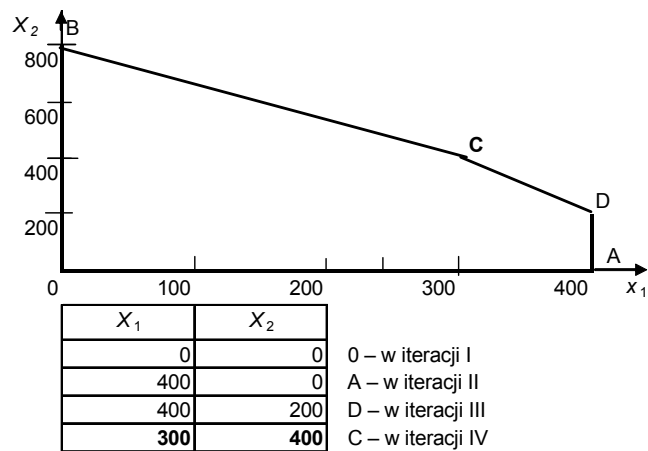
– funkcja celu:

$$\mathbf{c}_b^T \mathbf{B}_4^{-1} \mathbf{b} = [30 \ 0 \ 40] \cdot \begin{bmatrix} 400 \\ 300 \\ 300 \end{bmatrix} = 2400.$$

Wyrażenia Excela (na przykładowych obszarach arkusza kalkulacyjnego):

$$\begin{aligned} \mathbf{B}_i^{-1} &= \text{MACIERZ.ODW}(F194:H196) \\ \mathbf{B}_i^{-1} \mathbf{A} &= \text{MACIERZ.ILOCZYN}(F187:H189;B194:C196) \\ \mathbf{B}_i^{-1} \mathbf{b} &= \text{MACIERZ.ILOCZYN}(F187:H189;J194:J196) \\ \mathbf{c}_b^T \mathbf{B}_i^{-1} &= \text{MACIERZ.ILOCZYN}(B192:D192;F187:H189) \\ \mathbf{c}_b^T \mathbf{B}_i^{-1} \mathbf{A} &= \text{MACIERZ.ILOCZYN}(B192:D192;D187:E189) \\ \mathbf{c}_b^T \mathbf{B}_i^{-1} \mathbf{b} &= \text{MACIERZ.ILOCZYN}(B192:D192;CI187:I189) \end{aligned}$$

Rozwiązania w kolejnych iteracjach:



Funkcja celu: $F(x_1^*, x_2^*) = 300 * 40 + 400 * 30 = 24\ 000$,

Źródło: Opracowanie własne w Excelu, gdzie: $x_1^* \rightarrow x_1^*, x_2^* \rightarrow x_2^*$.

Wykres 2.5.2. Punkty wieloboku rozwiązań dopuszczalnych

Tabela 2.5.6

Rozwiązanie zadania decyzyjnego dodatkiem Solver programu Excel z przykładowymi wyrażeniami

A	B	C	D	E	Wiersz
			W_1	W_2	
Nazwa zasobu	Zasób	Wykorzystane	300	400	241
					=\$D\$241*D243+\$E\$241*E243
s_1	2 000	2 000	4	2	243
s_2	1 400	1 400	2	2	244
s_3	1 200	900	3	0	245
		Ceny	40	30	246
Max sprzedaży:		24 000	=(D246*D241)+(E246*E241)		

Źródło: Opracowanie własne, Solver, program Excel.

Zdefiniowanie zadania decyzyjnego w Solverze:

- maksymalizacja sprzedaży – funkcja celu (wyrażenie jak w tab. 2.5.6),
- warunki ograniczające: C243:C245 ≤ B243:B245,
- warunki brzegowe: D241:E241 ≥ 0; D241:E241 int (int – całkowitoliczbowe),
- zmienne decyzyjne X_1, X_2 (wielkość produkcji odpowiednio wyrobów W_1 oraz W_2),
- początkowe rozwiązanie: $x_1, x_2 = 0$.

W naszym przykładzie przed rozwiązaniem ustawiamy się na komórce celu: C247.

2.5.1. Pytania

1. Podaj kolejne kroki procedury algorytmu simpleks według elementu rozwiązującego w zadaniach programowania liniowego.

2. Podaj pierwszą tablicę rozwiązania bazowego algorytmu simpleks według rachunku macierzowego dla zadania decyzyjnego:

$$FC: 2x_1 + 3x_2 \rightarrow \max,$$

$$C_1: x_1 + 2x_2 \leq 100,$$

$$C_2: 3x_1 + 3x_2 \leq 100,$$

$$C_3: x_1 \leq 60,$$

$$x_1, x_2 \geq 0.$$

3. W poszukiwaniu rozwiązania optymalnego wymień elementy rachunku macierzowego, które zawierają kolejne tablice simpleksowe.

4. Podaj składnie (wyrażenia) funkcji Excela stosowane w rozwiązywaniu zadań programowania liniowego według rachunku macierzowego.

5. Wypełnij zaciemnione pola tablicy (tab. 2.5.7) rozwiązaniem bazowym podanego zadania decyzyjnego wykonywanego według rachunku macierzowego:

$$FC: x_1 + 4x_2 \rightarrow \min,$$

$$C_1: x_1 + 2x_2 \geq 10,$$

$$C_2: 8x_1 - 4x_2 \geq 20,$$

$$C_3: 4x_1 + 2x_2 \leq 40,$$

$$x_1, x_2 \geq 0.$$

Tabela 2.5.7

Zainicjowanie tablicy rozwiązania bazowego

			x_1	x_2	x_3	x_4	x_5	x_6	x_7
C_b	xb	b	a₁	a₂	a₃	a₄	a₅	a₆	a₇
			1	2					
			8	4					
			4	2					
	Kryterium	k_j							

Źródło: Opracowanie własne w Excelu.

2.5.2. Zadania

1. Rozwiąż manualnie następujące zadania programowania liniowego algorytmem simpleks według metody elementu wyróżniającego i sprawdź wyniki w swojej ostatniej tablicy z zamieszczonymi tutaj rozwiązaniami uzyskanymi w module (LP) programu WinQSB.

a) Rozwiązanie sumaryczne zadania decyzyjnego:

$$FC: 9x_1 + 6x_2 \rightarrow \max,$$

$$C_1: 4x_1 + x_2 \leq 200,$$

$$C_2: 2x_1 + 4x_2 \leq 100,$$

$$x_1, x_2 \geq 0.$$

Tabela 2.5.8

Rozwiązanie sumaryczne zadania decyzyjnego

09-19-2007	Decision	Solution	Unit Cost or	Total
10:55:40	Variable	Value	Profit $C(j)$	Contribution
1	X1	50	9	450
2	X2	0	6	0
	Objective	Function	(Max) =	450

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

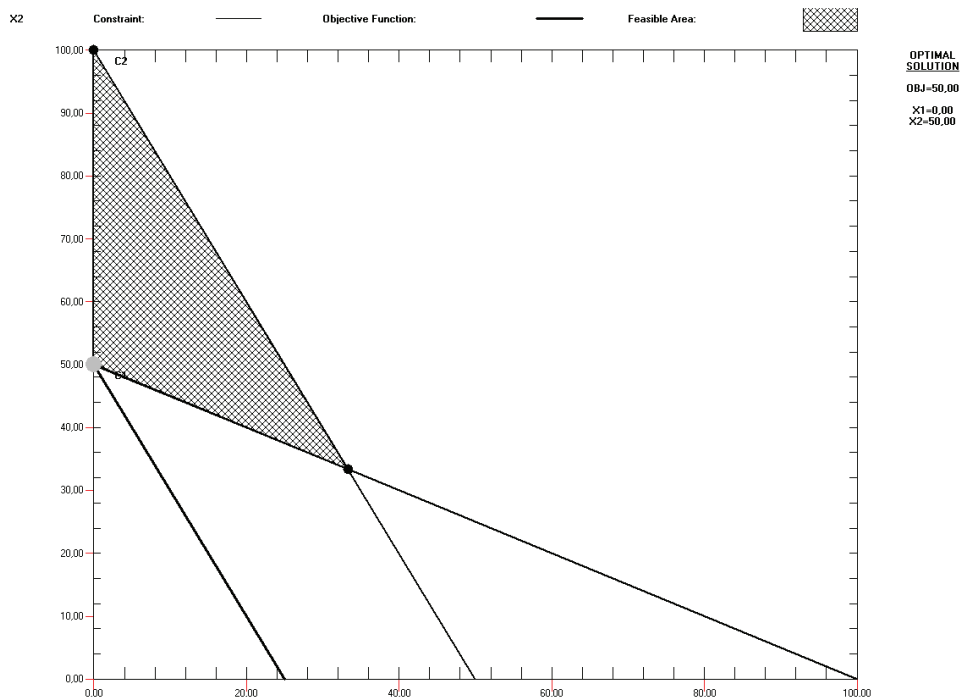
b) Rozwiązanie geometryczne (graficzne) zadania decyzyjnego:

$$FC: 2x_1 + x_2 \rightarrow \min,$$

$$C_1: 4x_1 + x_2 \geq 200,$$

$$C_2: 2x_1 + 4x_2 \leq 100,$$

$$x_1, x_2 \geq 0.$$



Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Wykres 2.5.3. Rozwiązanie graficzne zadania decyzyjnego

2. Po wykonaniu tablicy bazowej w ramach pytania piątego dokończ rozwiązywanie podanego tam zadania decyzyjnego, lecz zastosuj sposób według elementu rozwiązującego. Porównaj wyniki z zamieszczonym rozwiązaniem sumarycznym.

T a b e l a 2.5.9

Rozwiązanie sumaryczne zadania decyzyjnego

09-19-2007	Decision	Solution	Unit Cost or	Total
11:34:13	Variable	Value	Profit C (j)	Contribution
1	X1	10	1	10
2	X2	0	4	0
	Objective	Function	(Min) =	10

Ź r ó d ł o: Opracowanie własne, moduł (LP).

3. Rozwiąż zadanie decyzyjne programowania liniowego z pytania piątego, stosując rachunek macierzowy.

4. Rozwiąż zadanie decyzyjne rachunkiem macierzowym z zastosowaniem algorytmu simpleks:

$$FC: 3x_1 + 2x_2 \rightarrow \max,$$

$$C_1: x_1 + 2x_2 \leq 10,$$

$$C_2: 4x_1 + 4x_2 \leq 30,$$

$$C_3: 2x_2 \leq 40,$$

$$x_1, x_2 \geq 0.$$

Sprawdź swoje rozwiązanie z wynikami otrzymanymi według modułu (LP) programu WinQSB, mianowicie:

$$x_1^* = 7,5,$$

$$x_2^* = 0,$$

$$FC = 22,5.$$

2.6. Sztuczna baza danych

Jeżeli postać kanoniczna zagadnienia programowania liniowego nie zawiera macierzy jednostkowej, to tworzymy sztuczną bazę danych ze zmiennymi x_{n+m} :

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n + x_{n+1} = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n + x_{n+2} = b_2,$$

.....

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n + x_{n+m} = b_m,$$

$$x_1, x_2, \dots, x_n, x_{n+1}, x_{n+2}, \dots, x_{n+m} \geq 0,$$

$$z = c_1x_1 + c_2x_2 + \dots + c_nx_n + Mx_{n+1} + Mx_{n+2} + Mx_{n+m} \rightarrow \min.$$

Przykład

Rozwiązanie manualne

Tabela 2.6.1

Minimalizacja kosztów diety (spożycie produktów A,B,C zawierających składniki „S₁”, „S₂”, „S₃”)

	Model matematyczny zadania decyzyjnego	x_1	x_2	x_3 (2)	x_4	x_5 (1)	x_6 (3)		
FC:	$5x_1 + 2x_2 + x_3 \rightarrow \min$	5	2	1	0	M	M	→	min
C ₁ :	$x_1 + 2x_2 \geq 70$	1	2	0	-1	1	0	=	70
C ₂ :	$2x_1 + 3x_2 + x_3 = 140$	2	3	1	0	0	0	=	140
C ₃ :	$3x_1 + x_2 = 105$	3	1	0	0	0	1	=	105
	$x_1, x_2, x_3 \geq 0$								

gdzie:

M – bardzo duża liczba dodatnia; $x_1 \rightarrow A, x_2 \rightarrow B, x_3 \rightarrow C$; ograniczenia dotyczą odpowiednio składników: $C_1 \rightarrow S_1, C_2 \rightarrow S_2, C_3 \rightarrow S_3$.

Źródło: Opracowanie własne na podstawie: W. Sikora (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii*, Akademia Ekonomiczna, Poznań 2003, przykład. 2.

Macierz jednostkową tworzą zmienne: x_5, x_3, x_6 .

Postać kanoniczną a następnie uzupełnioną uzyskujemy po wprowadzeniu do modelu matematycznego zmiennych:

C_1 : zmiennej swobodnej x_4 ze współczynnikiem (-1) oraz zmiennej sztucznej x_5 ze współczynnikiem 1.

C_3 : zmiennej sztucznej x_6 ze współczynnikiem 1.

Ograniczenie C_2 nie wymaga wprowadzenia dodatkowej zmiennej, gdyż zmienna x_3 ze współczynnikiem 1 jest elementem macierzy jednostkowej.

Do funkcji celu $FC \rightarrow \min$ zmienne swobodne wchodzi z współczynnikiem zero, a zmienna sztuczna ze współczynnikiem (+ M).

Tabela 2.6.2

I tablica simpleksowa

	x_j	x_1	x_2	x_3	x_4	x_5	x_6	b_i	$\Theta = b_i/a_i$ (Ratio)
x_b	c_j c_b	5	2	1	0	M	M		
x_5	M	1	2	0	-1	1	0	70	70
x_3	1	2	3	1	0	0	0	140	70
$x_6(x_1)$	$M(5)$	3	1	0	0	0	1	105	35
	z_j	$4M+2$	$3M+3$	1	$-M$	M	M	$FC: 175M+140$	
	$k_j = c_j - z_j$	$3-4M$	$-1-3M$	0	M	0	0		

Kolumna wyróżniająca ze zmienną x_1 .

Wiersz wyróżniający ze zmienną x_6 .

Element wyróżniający – 3.

Źródło: Opracowanie na podstawie: W. Sikora (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii*, Akademia Ekonomiczna, Poznań 2003, tab. 11.

Tabela 2.6.3

II tablica simpleksowa

	x_j	x_1	x_2	x_3	x_4	x_5	x_6	b_i	$\Theta = b_i/a_i$ (Ratio)
x_b	c_j c_b	5	2	1	0	M	M		
$x_5(x_2)$	$M(2)$	0	$5/3$	0	-1	1	-1/3	35	21
x_3	1	0	7/3	1	0	0	-2/3	70	30
x_1	5	1	1/3	0	0	0	1/3	35	105
	z_j	5	$5/3M+4$	1	$-M$	M	$1-1/3M$	$FC: 35M+245$	
	$k_j = c_j - z_j$	0	$-2-5/3M$	0	M	0	$4/3M-1$		

Kolumna wyróżniająca ze zmiennej x_2 .

Wiersz wyróżniający ze zmiennej x_5 .

Element wyróżniający – 5/3.

Źródło: Opracowanie własne na podstawie: W. Sikora (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii*, Akademia Ekonomiczna, Poznań 2003, tab. 12.

Tabela 2.6.4

III tablica simpleksowa

x_j	x_1	x_2	x_3	x_4	x_5	x_6	b_i	$\Theta = b_i/a_i$ (Ratio)	
x_b	c_j	5	2	1	0	M	M		
c_b									
x_2	2	0	1	0	$-3/5$	$3/5$	$-1/5$	21	X
$x_3(x_4)$	1(0)	0	0	1	$7/5$	$-7/5$	$-1/5$	21	15
x_1	5	1	0	0	$1/5$	$-1/5$	$2/5$	28	140
z_j		5	2	1	$6/5$	$-6/5$	$7/5$	FC: 203	
$k_j = c_j - z_j$		0	0	0	$-6/5$	$M+6/5$	$M-7/5$		

Kolumna wyróżniająca ze zmienną x_4 .

Wiersz wyróżniający ze zmienną x_3 .

Element wyróżniający – $7/5$.

Źródło: Opracowanie na podstawie: W. Sikora, (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii*, Akademia Ekonomiczna, Poznań 2003, tab. 13.

Wiersz ze zmienną bazową x_2 o wartości ujemnej ($-3/5$) nie uwzględniony. Wynika to z warunków brzegowych.

Tabela 2.6.5

IV tablica simpleksowa (końcowa)

x_j	x_1 (A)	x_2 (B)	x_3	x_4	x_5	x_6	b_i	$\Theta = b_i/a_i$ (Ratio)	
x_b	c_j	5	2	1	0	M	M		
c_b									
x_2	2	0	1	$3/7$	0	0	$-2/7$	30	(B)
x_4	0	0	0	$5/7$	1	-1	$-1/7$	15	więcej S_1
x_1	5	1	0	$-1/7$	0	0	$3/7$	25	(A)
z_j		5	2	$y_2 = 1/7$	0	$y_1 = 0$	$y_3 = 11/7$	FC: 185	
$k_j = c_j - z_j$		0	0	$6/7$	0	0	$M-11/7$		

Źródło: Opracowanie na podstawie: W. Sikora, (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii*, Akademia Ekonomiczna, Poznań 2003, tab. 14.

Sprawdzenie rozwiązania według kryterium: $k_j \geq 0$. Wszystkie wartości k spełniają to kryterium, zatem mamy rozwiązanie optymalne.

$FC = 185$, $x_1^* = 25$, $x_2^* = 30$, $x_4 = 15$ nadwyżka (zapas) składnika S_1 .

Przykład

Rozwiązanie wcześniejszego zadania decyzyjnego z zastosowaniem modułu (LP) programu WinQSB.

a) Zdefiniowanie problemu decyzyjnego (raport)

Variable -->	X1	X2	X3	Direction	RHS
Minimize	5	2	1		
C1	1	2		>=	70
C2	2	3	1	=	140
C3	1	0		=	105
LowerBound	0	0	0		
UpperBound	M	M	M		
VariableType	Continuous	Continuous	Continuous		

b) Raporty z kolejnych iteracji

Tabela 2.6.6

I tablica simpleksowa (Simplex Tableau – Iteration 1)

		X1	X2	X3	Surplus_C1	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	C(j)	5,0	2,0	1,0	0	0	0	0	RHS	Ratio
Artificial_C1	M	1,0	2,0	0	-1,0	1,0	0	0	70,0	35,0
Artificial_C2	M	2,0	3,0	1,0	0	0	1,0	0	140,0	46,7
Artificial_C3	M	3,0	1,0	0	0	0	0	1,0	105,0	105,0
	C(j)-Z(j)	5,0	2,0	1,0	0	0	0	0	0	
	* Big M	-6,0	-6,0	-1,0	1,0	0	0	0	0	

gdzie:

Surplus_C1 – zmienna swobodna ograniczenia C₁.

Artificial_C1, Artificial_C2, Artificial_C3 – zmienna sztuczna odpowiednia do ograniczeń C₁, C₂, C₃.

Basis – wektor zmiennych bazowych.

c_j – ceny zmiennych bazowych.

Tabela 2.6.7

II tablica simpleksowa (Simplex Tableau – Iteration 2)

		X1	X2	X3	Surplus_C1	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	C(j)	5,0	2,0	1,0	0	0	0	0	RHS	Ratio
X2	2,0	0,5	1,0	0	-0,5	0,5	0	0	35,0	70,0
Artificial_C2	M	0,5	0	1,0	1,5	-1,5	1,0	0	35,0	70,0
Artificial_C3	M	2,5	0	0	0,5	-0,5	0	1,0	70,0	28,0
	C(j)-Z(j)	4,0	0	1,0	1,0	-1,0	0	0	70,0	
	* Big M	-3,0	0	-1,0	-2,0	3,0	0	0	0	

Tabela 2.6.8

III tablica simpleksowa (Simplex Tableau – Iteration 3)

		X1	X2	X3	Surplus_C1	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	C(j)	5,0	2,0	1,0	0	0	0	0	RHS	Ratio
X2	2,0	0,0	1,0	0	-0,6	0,6	0	-0,2	21,0	M
Artificial_C2	M	0,0	0,0	1,0	1,4	-1,4	1,0	-0,2	21,0	15,0
X1	5,0	1,0	0,0	0	0,2	-0,2	0	0,4	28,0	140,0
	C(j)-Z(j)	0	0	1,0	0,2	-0,2	0	-1,6	182,0	
	* Big M	0	0	-1,0	-1,4	2,4	0	1,2	0	

Tabela 2.6.9

IV tablica simpleksowa (Simplex Tableau – Iteration 4)

		5	X2	X3	Surplus_C1	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	C(j)	5,0	2,0	1,0	0	0	0	0	RHS	Ratio
X2	2,0	0,0	1,0	0,4	0	0	0,4	-0,3	30,0	
Surplus_C1	0	0,0	0,0	0,7	1,0	-1,0	0,7	-0,1	15,0	
X1	5,0	1,0	0,0	-0,1	0	0	-0,1	0,4	25,0	
	C(j)-Z(j)	0	0	0,9	0	0	-0,1	-1,6	185,0	
	* Big M	0	0	0	0	1,0	1,0	1,0	0	

Rozwiązanie optymalne: FC: 185, $x_1^* = 25$, $x_2^* = 30$, Surplus_C1 – zapas składnika dla ograniczenia C₁.

c) Zestawienie w postaci raportu kombinowanego

Tabela 2.6.10

Raport kombinowany z rozwiązania zadania decyzyjnego

	Decision Variable	Solution Value	Unit Cost or Profit C(j)	Total Contribution	Reduced Cost	Basis Status
1	X1	25,0	5,0	125,0	0	Basis
2	X2	30,0	2,0	60,0	0	Basis
3	X3	0	1,0	0	0,9	At bound
Objective		Function	(Min) =	185,0		
	Constraint	Left Hand Side	Direction	Right Hand Side	Slack or Surplus	Shadow Price
1	C1	85,0	>=	70,0	15,0	0
2	C2	140,0	=	140,0	0	0,1
3	C3	105,0	=	105,0	0	1,6

Źródło: Tabele 2.6.6–2.6.10 opracowanie własne, moduł (LP), program WinQSB.

Przykład

Określ I tablicę simpleksową bazową (z uwzględnieniem zmiennych bazy sztucznej) do podanego modelu matematycznego. Następnie rozwiąż zadanie w WinQSB.

$$FC: x_1 + x_2 + x_3 \rightarrow \max,$$

$$C_1: x_1 + x_2 \geq 8,$$

$$C_2: 2x_1 + x_2 + x_3 = 14,$$

$$C_3: x_1 + 2x_2 \leq 10,$$

$$x_1, x_2, x_3 \geq 0.$$

a) Sformatowanie I tablicy simpleksowej

Tabela 2.6.11

Manualne zdefiniowanie I tablicy simpleksowej (bazowej)

x_j	x_1	x_2	x_3	x_4	x_5	s_1	b_i	$\Theta = b_i/a_i$ (Ratio)
x_b	c_j	1	1	1	0	0	$-M$	
	c_b							
s_1	$-M$	1	1	0	-1	0	1	8
x_3	1	2	1	1	0	0	0	14
$x_5(x_2)$	0(1)	1	2	0	0	1	0	10
	z_j	$-M+2$	$-M+1$	1	M	0	$-M$	$-8M+14$
	$k_j = c_j - z_j$	$M-1$	M	0	$-M$	0	0	

gdzie:

s_1 – zmienna sztuczna odpowiadająca ograniczeniu C_1 ,

$(-M)$ – współczynnik, gdy $FC \rightarrow \max$.

Kolumna rozwiązująca ze zmienną x_2 .

Wiersz rozwiązujący ze zmienną x_5 .

Element rozwiązujący – 2.

Źródło: Opracowanie własne w Excelu.

b) Zdefiniowanie problemu decyzyjnego w module (LP) programu WinQSB, format Matrix (tablica)

Variable -->	X1	X2	X3	Direction	RHS
Maximize 1	1	1	1		
C1	1	1	0	\geq	8
C2	2	1	1	=	14
C3	1	2	0	\leq	10
LowerBound	0	0	0		
UpperBound	M	M	M		
VariableType	Continuous	Continuous	Continuous		

Program WinQSB nadaje własne nazwy zmiennym, i tak do ograniczenia C1 wprowadza zmienną swobodną Surplus_C1 oraz zmienną sztuczną Artificial_C1, natomiast dla

ograniczenia C2 zmienną sztuczną Artificial_C2. Do ograniczenia C3 wprowadza zmienną swobodną Slack_C3 (niedobór na ograniczeniu C3).

c) Rozwiązanie problemu decyzyjnego w module (LP):

T a b e l a 2.6.12

Raport kombinowany

	Decision Variable	Solution Value	Unit Cost or Profit C(j)	Total Contribution	Reduced Cost	Basis Status
1	X1	6	1	6	0	Basis
2	X2	2	1	2	0	Basis
3	X3	0	1	0	0	Basis
Objective Function (Max.) =				8		
	Constraint	Left Hand Side	Direction	Right Hand Side	Slack or Surplus	Shadow Price
1	C1	8	>=	8	0	-2
2	C2	14	=	14	0	1
3	C3	10	<=	10	0	1

Ź r ó d ł o: Opracowanie własne, moduł (LP), program WinQSB.

Rozwiązanie (Solution Value): $FC = 8$, $x_1^* = 6$, $x_2^* = 2$, $x_3 = 0$

Przykład początkowych dwóch iteracji w rozwiązaniu komputerowym

T a b e l a 2.6.13

I tablica simpleksowa (Simplex Tableau – Iteration 1)

		X1	X2	X3	Surplus_C1	Slack_C3	Artificial_C1	Artificial_C2		
Basis	C(j)	1	1	1	0	0	-M	-M	RHS	Ratio
Artificial_C1	-M	1	1	0	-1	0	1	0	8	8
Artificial_C2	-M	2	1	1	0	0	0	1	14	7
Slack_C3	0	1	2	0	0	1	0	0	10	10
	C(j)-Z(j)	1	1	1	0	0	0	0	0	
	* Big M	3	2	1	-1	0	0	0	0	

Ź r ó d ł o: Opracowanie własne, moduł (LP), program WinQSB.

T a b e l a 2.6.14

II tablica simpleksowa (Simplex Tableau – Iteration 2)

		X1	X2	X3	Surplus_C1	Slack_C3	Artificial_C1	Artificial_C2		
Basis	C(j)	1	1	1	0	0	-M	-M	RHS	Ratio
Artificial_C1	-M	0	0,5	-0,5	-1	0	1	-0,5	1	2
X1	1	1	0,5	0,5	0	0	0	0,5	7	14
Slack_C3	0	0	1,5	-0,5	0	1	0	-0,5	3	2
	C(j)-Z(j)	0	0,5	0,5	0	0	0	-0,5	7	
	* Big M	0	0,5	-0,5	-1	0	0	-1,5	0	

Ź r ó d ł o: Opracowanie własne, moduł (LP), program WinQSB.

2.6.1. Pytania

1. Kiedy tworzymy bazę sztuczną?

2. Sprowadź zadania programowania liniowego: a), b), c), d) do postaci kanonicznej, a jeśli nie będą zawierać macierzy jednostkowej, utwórz bazę sztuczną w formie I tablicy simpleksowej (bazowej).

a)

$$FC: 2x_1 + 5x_2 + x_3 \rightarrow \min,$$

$$C_1: 2x_1 + x_2 \geq 70,$$

$$C_2: 3x_1 + 2x_2 + x_3 = 140,$$

$$C_3: x_1 + 3x_2 = 105,$$

$$x_1, x_2, x_3 \geq 0,$$

b)

$$FC: 2x_1 + 5x_2 + x_3 \rightarrow \max,$$

$$C_1: 2x_1 + x_2 \leq 70,$$

$$C_2: 3x_1 + 2x_2 + x_3 = 140,$$

$$C_3: x_1 + 3x_2 = 105,$$

$$x_1, x_2, x_3 \geq 0,$$

c)

$$FC: 2x_1 + 5x_2 \rightarrow \min,$$

$$C_1: 2x_1 + x_2 = 20,$$

$$C_2: x_1 + 10x_2 \geq 20,$$

$$C_3: 5x_1 + x_2 = 20,$$

$$x_1, x_2 \geq 0,$$

d)

$$FC: 2x_1 + 5x_2 + x_3 \rightarrow \max,$$

$$C_1: 2x_1 + x_2 + 2x_3 = 70,$$

$$C_2: 3x_1 + 2x_2 + x_3 = 140,$$

$$C_3: x_1 = 35,$$

$$x_1, x_2, x_3 \geq 0.$$

2.6.2. Zadania

Mając określone przykładowe tablice iteracji wykonane w module (LP) do zadań programowania liniowego (a–d), określ kolejne (jeśli potrzeba) algorytmem simpleks według sposobu elementu wyróżniającego. W tablicach zmienne

swobodne oznaczono jako Surplus lub Slack, a zmienne sztuczne jako Artificial z dopisanym symbolem ograniczenia. Zaokrąglone wartości na wydruku wynikają z precyzji obliczeń.

Ad a)

Tabela 2.6.15

IV tablica simpleksowa (Simplex Tableau – Iteration 4)

	Zmienne:	X1	X2	X3	Surplus_C1	Artificial_C1	Artificial_C2	Artificial_C3		$\Theta = b_i/a_i$
Basis	$C(j)$	2	5	1	0	0	M	M	RHS	(Ratio)
X1		2	1	0	0,43	0	0	0,43	-0,29	30
Surplus_C1		0	0	0	0,71	1	-1	0,71	-0,14	15
X2		5	0	1	-0,14	0	0	-0,14	0,43	25
$k_j = C(j)-Z(j)$		0	0	0	0,86	0	0	-0,14	-1,57	185
* Big M		0	0	0	0	1	1	1	1	0

Kolumna rozwiązująca ze zmienną Artificial_C3

Wiersz rozwiązujący ze zmienną x_2

Element rozwiązujący – 0,43 wartości ujemnych nie uwzględniamy

Przykładowy odczyt kryterium z tablicy iteracji IV: $k_7 = 1,57 + 1M$.

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Ad b)

Tabela 2.6.16

V tablica simpleksowa (Simplex Tableau – Iteration 5)

	Zmienne:	X1	X2	X3	Slask_C1	Artificial_C2	Artificial_C3		$\Theta = b_i/a_i$	
Basis	$C(j)$		2	5	1	0	-M	-M	RHS	(Ratio)
Slask_C1		0	1,67	0	0	1	0	-0,33	35	
X3		1	2,33	0	1	0	1	-0,67	70	
X2		5	0,33	1	0	0	0	0,33	35	
$k_j = C(j)-Z(j)$			-2	0	0	0	-1	-1	245	
* Big M			0	0	0	0	-1	-1	0	

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Ad c)

Tabela 2.6.17

IV tablica simpleksowa (Simplex Tableau – Iteration 4)

	Zmienne:	X1	X2	Surplus_C2	Artificial_C1	Artificial_C2	Artificial_C3		$Q = b_i/a_i$
Basis	$C(j)$	2	5	0	M	M	M	RHS	(Ratio)
Surplus_C2		0	0	1	16,33	-1	-6,33	180	
X2		5	0	1	1,67	0	0,67	20	
X1		2	1	0	0,33	0	0,33	0	
$k_j = C(j)-Z(j)$			0	0	-7,67	0	2,67	100	
* Big M			0	0	0	1	1	0	

Kolumna rozwiązująca ze zmienną Artificial_C3.

Wiersz rozwiązujący ze zmienną x_1 .

Element rozwiązujący – 0,33.

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Ad d) Z obliczeniem udziałów w kolumnie Ratio

Tabela 2.6.18

II tablica simpleksowa (Simplex Tableau – Iteration 2)

	Zmienne:	X1	X2	X3	Artificial_C1	Artificial_C2	Artificial_C3		$\theta = b_i/a_i$
Basis	C(j)	2	5	1	-M	-M	-M	RHS	(Ratio)
Artificial_C1	-M	0	1	2	1	0	-2	0	0
Artificial_C2	-M	0	2	1	0	1	-3	35	17,5
X1	2	1	0	0	0	0	1	35	M
$k_j =$	$C(j)-Z(j)$	0	5	1	0	0	-2	70	
	* Big M	0	3	3	0	0	-6	0	

Kolumna rozwiązująca ze zmienną x_2 .

Wiersz rozwiązujący ze zmienną Artificial_C1.

Element rozwiązujący – 1.

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

W poprawionych już tabelach 2.6.15–2.6.18 w Excelu jako parametr funkcji celu zmiennych sztucznych moduł (LP) programu WinQSB powinien wpisać M ze znakiem (+) lub (–) odpowiednio do funkcji celu, a wpisuje w raportach komputerowych zero, lecz zadanie rozwiązywane jest jednak poprawnie z wartością M .

2.7. Analiza wrażliwości

Opracowanie w części manualnej nawiązuje do procedury postępowania zawartej w publikacji po redakcja W. Sikory¹⁶⁴.

Przykład

Firma budowlana produkuje trzy prefabrykaty betonowe z dwóch materiałów. Dąży do maksymalizacji przychodu. Normy jednostkowe, ceny oraz zasoby materiałów podano w tabeli 2.7.1.

Tabela 2.7.1

Normy jednostkowe, ceny i zasoby

Wyrób \ Materiał	X_1	X_2	X_3	Zasób
M_1	4	2	1	32
M_2	2	4	4	36
Cena [zł]	8	3	10	

Źródło: Opracowanie własne; dane umowne.

Określ manualnie ze wspomaganie komputerowym:

- wartość maksymalną przychodu,
- zmienne dualne,
- przedziały cen wyrobów przy zachowaniu rozwiązania optymalnego,
- granice zasobów przy zachowaniu struktury rozwiązania,
- przedziały nakładów jednostkowych na wyrób X_2 przy zachowaniu rozwiązania optymalnego.

Rozwiązanie z przykładami raportów modułu (LP) programu WinQSB ze względu na granice wrażliwości

Ad a) Model matematyczny i postać kanoniczna zadania decyzyjnego:

$$FC: 8x_1 + 3x_2 + 10x_3 \rightarrow \max,$$

$$C_1: 4x_1 + 2x_2 + x_3 \leq 32,$$

$$C_2: 2x_1 + 4x_2 + 4x_3 \leq 36,$$

$$x_1, x_2, x_3 \geq 0.$$

$$FC: 8x_1 + 3x_2 + 10x_3 + 0x_4 + 0x_5 \rightarrow \max,$$

$$C_1: 4x_1 + 2x_2 + x_3 + x_4 = 32,$$

$$C_2: 2x_1 + 4x_2 + 4x_3 + x_5 = 36,$$

$$x_1, x_2, x_3, x_4, x_5 \geq 0.$$

¹⁶⁴ W. Sikora (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii*, Akademia Ekonomiczna, Poznań 2003, rozdz. 1.4.

Kolejne iteracje rozwiązania zadania decyzyjnego metodą simpleks:

Tabela 2.7.2

I tablica simpleksowa

	x_j	x_1	x_2	x_3	x_4	x_5	b_i	θ_i
	x_b	c_b/c_j	8	3	10	0	0	
w_1	x_4	0	4	2	1	1	0	32
w_2	x_5	0	2	4	4	0	1	36
	z_j		0	0	0	0	0	$FC = 0$
	$k_j = c_j - z_j$		8	3	10	0	0	

Kolumna rozwiązująca ze zmienną x_3 .

Wiersz rozwiązujący ze zmienną x_5 .

Element rozwiązujący – 4.

Tabela 2.7.3.

	x_j	x_1	x_2	x_3	x_4	x_5	b_i	θ_i
	x_b	c_b/c_j	8	3	10	0	0	
w_1	x_4	0	3,5	1	0 **	1	-0,25	23
w_2	x_3	10	0,5	1	1 *	0	0,25	9
	z_j		5	10	10	0	2,5	$FC = 90$
	$k_j = c_j - z_j$		3	-7	0	0	-2,5	

II tablica simpleksowa

Kolumna rozwiązująca ze zmienną x_1 .

Wiersz rozwiązujący ze zmienną x_4 .

Element rozwiązujący – 3,5.

Tabela 2.7.4

III tablica simpleksowa (końcowa)

	x_j	x_1	x_2	x_3	x_4	x_5	Rozwiązanie b_i	θ_i
	x_b	c_b/c_j	8	3	10	0	0	
w_1	x_1	8	1	2/7	0	2/7	-1/14	46/7
w_2	x_3	10	0	6/7	1	-1/7	2/7	40/7
	z_j		8	76/7	10	6/7	16/7	$FC = 109,71$
	$k_j = c_j - z_j$		0	-7,86	0	-6/7	-16/7	

gdzie:

$$\theta_i = b_i / a_i; a_i \text{ dla kolumny rozwiązującej}; FC = \sum c_b \cdot b_i.$$

Źródło: Tabele 2.7.2–2.7.4 – opracowanie własne w Excelu.

Rozwiązaniem alternatywnym do sposobu według elementu rozwiązującego jest takie przekształcenie danej tablicy simpleksowej za pomocą operacji elementarnych na wierszach w_1 i w_2 , aby element rozwiązujący zwany również centralnym stał się jedynką, natomiast pozostałe elementy jego kolumny – zerami. W odniesieniu do tablicy II, która powstała poprzez przekształcenie pierwszej tablicy simpleksowej dla komórek oznaczonych ** oraz *.

$$** \quad w_1' = w_1 - 1/4 w_2,$$

$$* \quad w_2' = 1/4 w_2, \text{ przy czym } w - \text{ numer wiersza macierzy.}$$

Przykładowo dla elementu rozwiązującego 4 w kolumnie zmiennej x_3 (tab. 2.7.2):

$$w_1' x_3 = 1 - 1/4 \cdot 4 = 0,$$

$$w_2' = 1/4 \cdot 4 = 1.$$

Dla III tablicy sympleksowej i elementu rozwiązującego $3,5 = 7/2$ dokonaliśmy przekształcenia:

$$w_1' = \frac{2}{7} \cdot w_1,$$

$$w_2' = w_2 - \frac{2}{7} \cdot w_1.$$

W stosunku do zmiennej x_1 :

$$w_{1x_1}' = \frac{2}{7} \cdot 3,5 = \frac{2}{7} \cdot \frac{7}{2} = 1,$$

$$w_{2x_1}' = 0,5 - \frac{1}{7} \cdot 3,5 = 0,5 - \frac{1}{7} \cdot \frac{7}{2} = 0.$$

Przekształcenie dotyczy wszystkich zmiennych $x_1 - x_6$ macierzy uzupełnionej oraz wyrazu b_i .

Rozwiązanie uzyskaliśmy w trzech iteracjach i jest następujące:

$$x_1^* = 46/7 = 6,5714 \approx 6,57, \quad x_2^* = 0 \quad x_3^* = 40/7 = 5,7143 \approx 5,71.$$

Ad b) Zmienne dualne:

Rozwiązanie dualne zadania decyzyjnego występuje pod zmiennymi swobodnymi x_4 i x_5 w wierszu z_j :

$$y_1^* = 6/7 = 0,8571 \approx 0,86, \quad y_2^* = 16/7 = 2,2857 \approx 2,29.$$

Funkcja celu rozwiązania dualnego jest równa funkcji celu rozwiązania primalnego $FC_D = FC_P$.

Temat formułowania i rozwiązywania zadań dualnych występuje w rozdziale 2.8.

Ad c) Przedziały cen wyrobów:

W tabeli 2.7.4 zmienna X_1 jest bazowa, więc c_1 wpływa na kryteria zmiennych nie-bazowych. Dla realizacji w Excelu zamieniamy ułamki zwykle na liczby dziesiętne.

T a b e l a 2.7.5

Dane do analizy wrażliwości cen wyrobów

	x_j	x_1	x_2	x_3	x_4	x_5	Rozwiązanie b_i	θ_i
x_b	c_b/c_j	8	3	10	0	0		
x_1	c_1	1	0,29	0	0,29	-0,07	6,57	
x_3	10	0	0,86	1	-0,14	0,29	5,71	
	z_j	8	76/7	10	6/7	16/7	FC: 109,71	
	k_j	0	-7,86	0	-0,86	-2,29		

Ź r ó ł o: Opracowanie własne w Excelu; dane według tabeli 2.7.4.

Określamy równania:

$$k_2 = 3 - (0,29c_1 + 0,86 \cdot 10) \leq 0 \quad c_1 \geq -19,31,$$

$$k_4 = 0 - (0,29 \cdot c_1 - 0,14 \cdot 10) \leq 0 \quad c_1 \geq 4,83,$$

$$k_5 = 0 - (-0,07c_1 + 0,29 \cdot 10) \leq 0 \quad c_1 \geq 41,43.$$

Zatem cena c_1 przy zachowaniu rozwiązania optymalnego może być w przedziale:

$$c_1 = \langle 4,83; 41,43 \rangle,$$

$$k_2 = c_2 - 76/7 \leq 0.$$

Cena c_2 może być w przedziale $c_2 \in \langle -\infty; 10,86 \rangle$.

Zmienne X_2, X_4, X_5 są niebazowe, zatem ich ceny wpływają tylko na ich kryteria.

$$k_2 = 3 - (0,29 \cdot 8 + 0,86 \cdot c_3) \leq 0 \quad c_3 \geq 0,79,$$

$$k_4 = 0 - (0,29 \cdot 8 - 0,14 \cdot c_3) \leq 0 \quad c_3 \geq 16,57,$$

$$k_5 = 0 - (-0,07 \cdot 8 + 0,29 \cdot c_3) \leq 0 \quad c_3 \geq 1,93.$$

Cena c_3 wyrobu x_3 może być w przedziale $\langle 1,93; 16,57 \rangle$.

Przyjrzyjmy się teraz realizacji komputerowej modułem (LP) programu WinQSB dla zbadania wrażliwości na zmianę cen c_1, c_2, c_3 .

Tabela 2.7.6

Raport kombinowany uwzględniający wrażliwość na ceny

10:53:35		Thursday	December	6	2007			
Decision Variable	Solution Value	Unit Cost or Profit c(j)	Total Contribution	Reduced Cost	Basis Status	Allowable Min. C(j)	Allowable Max. C(j)	
1 X1	6,57	6	39,43	0	Basic	3	24	
2 X2	0	6	0	-4,71	At bound	-M	10,71	
3 X3	5,71	6	34,29	0	Basic	1,5	12	
Objective	Function (Max.) =		73,71					
	Left Hand		Right Hand	Slack	Shadow	Allowable	Allowable	
	Constraint	Side	Direction	Side	or Surplus	Price	Min. RHS	Max. RHS
1 C1	32	<=	32	0	0,86	9	72	
2 C2	36	<=	36	0	2,29	16	128	

Źródło: Opracowanie własne z zastosowaniem modułu (LP) programu WinQSB.

W danych wejściowych (tabela 2.7.1) ceny były następujące: $c_1 = 8, c_2 = 3, c_3 = 10$ zł.

W tabeli 2.7.6 występuje przykład analizy wrażliwości wykonanej w module (LP) dla zmiennych cen: $c_1 = c_2 = c_3 = 6$ spełniającej wcześniej określone możliwe przedziały cen wyrobów z zachowaniem rozwiązania optymalnego. Z raportu wynika, że rozwiąza-

nie optymalne nie ulega zmianie, a korekcie podlega jedynie funkcja celu, która wynosi $FC = 73,71$.

Ad d) Granice zasobów:

Zabazujemy na iteracji końcowej (tab. 2.7.4) lecz na danych przybliżonych z tab. 2.7.5 oraz na wzorach rachunku macierzowego określania składników kolejnej iteracji podanych w tab. 2.3.2.

Przeliczenia według wzorów rachunku macierzowego wymienionych w tab. 2.7.6 wykonano w Excelu.

Tabela 2.7.7

Dane tab. 2.3.2

	Zmienne bazowe	c		Rozwiązanie
c_{bl}	x_{bl}	B_j⁻¹ A	B_j⁻¹	B_j⁻¹ b
	z_j	c_b^T B_j⁻¹ A	c_b^T B_j⁻¹	c_b^T B_j⁻¹ b
	k_j = c_j - z_j	c_j - c_b^T B_j⁻¹ A	-c_b^T B_j⁻¹	

$$\mathbf{B}^{-1} = \begin{bmatrix} 0,29 & -0,07 \\ -0,14 & 0,29 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ 36 \end{bmatrix},$$

$$\begin{aligned} 0,29b_1 - 2,52 &\geq 0, \\ -0,14b_1 + 10,44 &\geq 0, \end{aligned}$$

$$\begin{aligned} 0,29 \cdot 32 - 0,07b_2 &\geq 0, \\ -0,14 \cdot 32 + 0,29b_2 &\geq 0. \end{aligned}$$

$$b_1 < 8,69; 74,57 > ,$$

$$b_2 < 15,45; 132,57 > .$$

Ad e) Przedziały nakładów jednostkowych na wyrób x_2 (według danych tab. 2.7.4):

$$\begin{bmatrix} 2/7 & -1/14 \\ -1,7 & 2/7 \end{bmatrix} \cdot \begin{bmatrix} a_{12} \\ 4 \end{bmatrix} = \begin{bmatrix} 2/7 a_{12} & -2/7 \\ -1/7 a_{12} & 4 \end{bmatrix}.$$

$$z_2 = 8 \cdot \left(\frac{2}{7} a_{12} - \frac{2}{7} \right) + 10 \cdot \left(-\frac{1}{7} a_{12} + \frac{8}{7} \right) = -\frac{4}{7} a_{12} + \frac{64}{7},$$

$$k_2 = 3 - \left(-\frac{4}{7} a_{12} + \frac{64}{7} \right) = -\frac{43}{7} + \frac{4}{7} a_{12} \leq 0,$$

$$a_{12} = \left\langle -\infty, \frac{43}{7} \right\rangle,$$

$$z_2 = 8 \cdot \left(\frac{4}{7} - \frac{1}{14} a_{22} \right) + 10 \cdot \left(-\frac{2}{7} + \frac{2}{7} a_{22} \right) = \frac{16}{7} a_{22} + \frac{12}{7},$$

$$k_2 = 3 - \left(\frac{16}{7} a_{22} + \frac{12}{7} \right) = -\frac{9}{7} - \frac{16}{7} a_{22} \leq 0,$$

$$a_{22} = \left\langle \frac{9}{16}; \infty \right\rangle.$$

Powracając do granic zasobów, dokonajmy zmiany zasobu b_1 z 32 oraz $b_2 = 36$ na 20 i zrealizujmy zadanie decyzyjne w module (LP).

Tabela 2.7.8

Raport kombinowany (zmiana zasobów materiałów)

09:59:01		Thursday	December	6	2007		
Decision	Solution	Unit Cost or	Total	Reduced	Basis	Allowable	Allowable
Variable	Value	Profit C(j)	Contribution	Cost	Status	Min. c(j)	Max. C(j)
1 X1	4,29	8	34,29	0	Basic	5	40
2 X2	0	3	0	-7,86	At bound	-M	10,86
3 X3	2,86	10	28,57	0	Basic	2	16
Objective	Function	(Max.) =	62,86				
	Left Hand		Right Hand	Slack	Shadow	Allowable	Allowable
Constraint	Side	Direction	Side	or Surplus	Price	Min. RHS	Max. RHS
1 C1	20	<=	20	0	0,86	5	40
2 C2	20	<=	20	0	2,29	10	80

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

A teraz przetestujmy działanie modułu (LP) związane ze zmienną parametrów przykładowej zmiennej X_2 .

Tabela 2.7.9

Raport kombinowany (zmiana parametrów zmiennej „X₂”)

10:16:14		Thursday	December	6	2007		
Decision	Solution	Unit Cost or	Total	Reduced	Basis	Allowable	Allowable
Variable	Value	Profit C(j)	Contribution	Cost	Status	Min. C(j)	Max. C(j)
1 X1	6,57	8	52,57	0	Basic	5	40
2 X2	0	3	0	-12,71	At bound	-M	15,71
3 X3	5,71	10	57,14	0	Basic	2	16
Objective	Function	(Max.) =	109,71				
	Left Hand		Right Hand	Slack	Shadow	Allowable	Allowable
Constraint	Side	Direction	Side	or Surplus	Price	Min. RHS	Max. RHS
1 C1	32	<=	32	0	0,86	9	72
2 C2	36	<=	36	0	2,29	16	128

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Jak już wspomniano, w tabelach 2.7.8 i 2.7.9 pokazano przykłady raportów uzyskanych przy wykorzystaniu modułu (LP) programu WinQSB. Raporty dotyczą zmiany zasobów oraz parametrów w granicach wrażliwości:

– tabela 2.7.8: zmiany zasobów surowców b_1, b_2 na 20; obserwujemy zachowanie struktury rozwiązania optymalnego, tj. x_1 oraz x_3 , a $x_2 = 0$,

– tabela 2.7.9: zmiany parametrów a_{12}, a_{22} dla zmiennej X_2 na 5; nastąpiło zachowanie rozwiązania optymalnego (wartości optymalne rozwiązania prymalnego zmiennych X_1, X_2 i odpowiadające im wartości dualne y_1, y_2 oraz wartość funkcji celu nie uległy zmianie).

A teraz zobaczymy rozwiązanie komputerowe zadania decyzyjnego prymalnego oraz odpowiadającego mu zadania dualnego, wychodząc z danych zawartych w tab. 2.7.1. Sposób formułowania zadania dualnego podano w rozdziale 2.8.

Tabela 2.7.10

Zdefiniowanie zadania decyzyjnego prymalnego

LP	MatrixFormat	X1	X2	X3	Direction	RHS
Variable -->						
Maximize		8	3	10		
C1		4	2	1	<=	32
C2		2	4	4	<=	36
LowerBound		0	0	0		
UpperBound	M	M	M	M		
VariableType	Continuous	Continuous	Continuous	Continuous		

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Tabela 2.7.11

Raport kombinowany rozwiązania prymalnego

Decision	Solution	Unit Cost or Profit C(j)	Total	Reduced Cost	Basis	Allowable Min. C(j)	Allowable Max. C(j)
1	X1	6,57	8	52,57	0	Basic	5, 40
2	X2	0	3	0	-7,86	At bound	-M, 10,86
3	X3	5,71	10	57,14	0	Basic	2, 16
Objective Function (Max.) =		109,71					
Constraint	Side	Direction	Side	Slack or Surplus	Shadow Price	Allowable Min. RHS	Allowable Max. RHS
1	C1	32	<=	32	0	0,86	9, 72
2	C2	36	<=	36	0	2,29	16, 128

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

W porównaniu z tab. 2.7.5 w raporcie (tab. 2.7.11) występuje zgodność wyników w zakresie rozwiązania optymalnego wyrobów X_1, X_2, X_3 , funkcji celu i zmiennych dualnych co do wartości bezwzględnej. Natomiast przedziały cen wyrobów oraz granice zasobów są zbliżone i zostały zaokrąglone do liczb całkowitych.

W tabelach 2.7.12–2.7.14 podano rozwiązanie zadania decyzyjnego prymalnego w trzech iteracjach wykonane w czasie 0,01 sekundy.

Tabela 2.7.12

I tablica simpleksowa (Simplex Tableau – Iteration 1)

		X1	X2	X3	Slack_C1	Slack_C2		
Basis	C(j)	8	3	10	0	0	RHS	Ratio
Slack_C1	0	4	2	1	1	0	32	32
Slack_C2	0	2	4	4	0	1	36	9
	C(j)-Z(j)	8	3	10	0	0	0	

Zmienne swobodne oznaczono przez moduł (LP), odpowiednio:
 X_4 – Slack_C1, X_5 – Slack_C2.

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Tabela 2.7.13

II tablica simpleksowa (Simplex Tableau – Iteration 2)

		X1	X2	X3	Slack_C1	Slack_C2		
Basis	C(j)	8	3	10	0	0	RHS	Ratio
Slack_C1	0	3,5	1	0	1	-0,25	23	6,57
X3	10	0,5	1	1	0	0,25	9	18
	C(j)-Z(j)	3	-7	0	0	-2,5	90	

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Tabela 2.7.14

III tablica simpleksowa (Simplex Tableau – Iteration 3)

		X1	X2	X3	Slack_C1	Slack_C2		
Basis	C(j)	8	3	10	0	0	RHS	Ratio
X1	8	1	0,29	0	0,29	-0,07	6,57	
X3	10	0	0,86	1	-0,14	0,29	5,71	
	C(j)-Z(j)	0	-7,86	0	-0,86	-2,29	109,71	

Źródło: Tabele 2.7.12–2.7.14 – opracowanie własne z zastosowaniem modułu (LP).

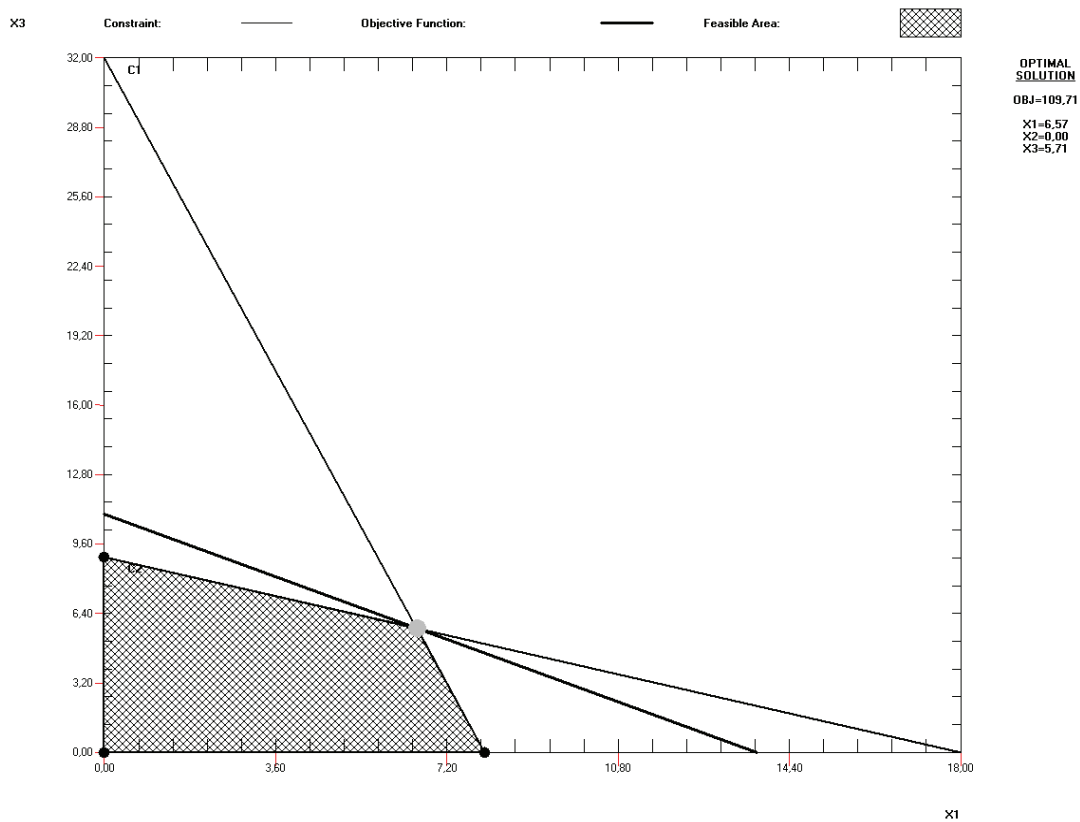
Rozwiązanie komputerowe dualne

Tabela 2.7.15

Zdefiniowanie zadania decyzyjnego dualnego

LP	MatrixFormat		2	3
Variable -->	Y1	Y2	Direction	RHS
Minimize	32	36		
C1	4	2	>=	8
C2	2	4	>=	3
C3	1	4	>=	10
LowerBound	0	0		
UpperBound	M	M		

Źródło: Opracowanie własne z zastosowaniem modułu (LP).



Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Wykres 2.7.1. Rozwiązanie graficzne (przy założeniu według rozwiązania optymalnego – $x_2^* = 0$)

Tabela 2.7.16

Raport kombinowany rozwiązania dualnego

	12:22:10	Thursday	December	6	2007		
Decision	Solution	Unit Cost or	Total	Reduced	Basis	Allowable	Allowable
Variable	Value	Profit C(j)	Contribution	Cost	Status	Min. C(j)	Max. C(j)
1 Y1	0,86	32	27,43	0	Basic	9	72
2 Y2	2,29	36	82,29	0	Basic	16	128
Objective	Function (Min.) =		109,71				
	Left Hand		Right Hand	Slack	Shadow	Allowable	Allowable
Constraint	Side	Direction	Side	or Surplus	Price	Min. RHS	Max. RHS
1 C1	8	>=	8	0	6,57	5	40
2 C2	10,86	>=	3	7,86	0	-M	10,86
3 C3	10	>=	10	0	5,71	2	16

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Rozwiązania zadania dualnego w pięciu iteracjach, gdzie ograniczenia zasobów stanowią ceny c_1, c_2, c_3 .

Tabela 2.7.17

I tablica simpleksowa (Simplex Tableau – Iteration 1)

		Y1	Y2	Surplus_C1	Surplus_C2	Surplus_C3	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	$C(j)$	32	36	0	0	0	0	0	0	RHS	Ratio
Artificial_C1	M	4	2	-1	0	0	1	0	0	8	4
Artificial_C2	M	2	4	0	-1	0	0	1	0	3	1
Artificial_C3	M	1	4	0	0	-1	0	0	1	10	3
	$C(j)-Z(j)$	32	36	0	0	0	0	0	0	0	
	* Big M	-7	-10	1	1	1	0	0	0	0	

gdzie:

Surplus_C1 – zmienna swobodna odpowiadająca ograniczeniu ceny C_1 .Artificial_C1 – zmienna sztuczna odpowiadająca ograniczeniu ceny C_1 .

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Tabela 2.7.18

II tablica simpleksowa (Simplex Tableau – Iteration 2)

		Y1	Y2	Surplus_C1	Surplus_C2	Surplus_C3	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	$C(j)$	32	36	0	0	0	0	0	0	RHS	Ratio
Artificial_C1	M	3	0	-1	0,5	0	1	-0,5	0	6,5	2,17
Y2	36	0,5	1	0	-0,25	0	0	0,25	0	0,75	1,5
Artificial_C3	M	-1	0	0	1	-1	0	-1	1	7	M
	$C(j)-Z(j)$	14	0	0	9	0	0	-9	0	27	
	* Big M	-2	0	1	-1,5	1	0	2,5	0	0	

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Tabela 2.7.19

III tablica simpleksowa (Simplex Tableau – Iteration 3)

		Y1	Y2	Surplus_C1	Surplus_C2	Surplus_C3	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	$C(j)$	32	36	0	0	0	0	0	0	RHS	Ratio
Artificial_C1	M	0	-6	-1	2	0	1	-2	0	2	1
Y1	32	1	2	0	-0,5	0	0	0,5	0	1,5	M
Artificial_C3	M	0	2	0	0,5	-1	0	-0,5	1	8,5	17
	$C(j)-Z(j)$	0	-28	0	16	0	0	-16	0	48	
	* Big M	0	4	1	-2,5	1	0	3,5	0	0	

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Tabela 2.7.20

IV tablica simpleksowa (Simplex Tableau – Iteration 4)

		Y1	Y2	Surplus_C1	Surplus_C2	Surplus_C3	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	$C(j)$	32	36	0	0	0	0	0	0	RHS	Ratio
Surplus_C2	0	0	-3	-0,5	1	0	0,5	-1	0	1	M
Y1	32	1	0,5	-0,3	0	0	0,3	0	0	2	4
Artificial_C3	M	0	3,5	0,3	0	-1	-0,3	0	1	8	2,3
	$C(j)-Z(j)$	0	20	8	0	0	-8	0	0	64	
	* Big M	0	-3,5	-0,3	0	1	1,3	1	0	0	

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Tabela 2.7.21

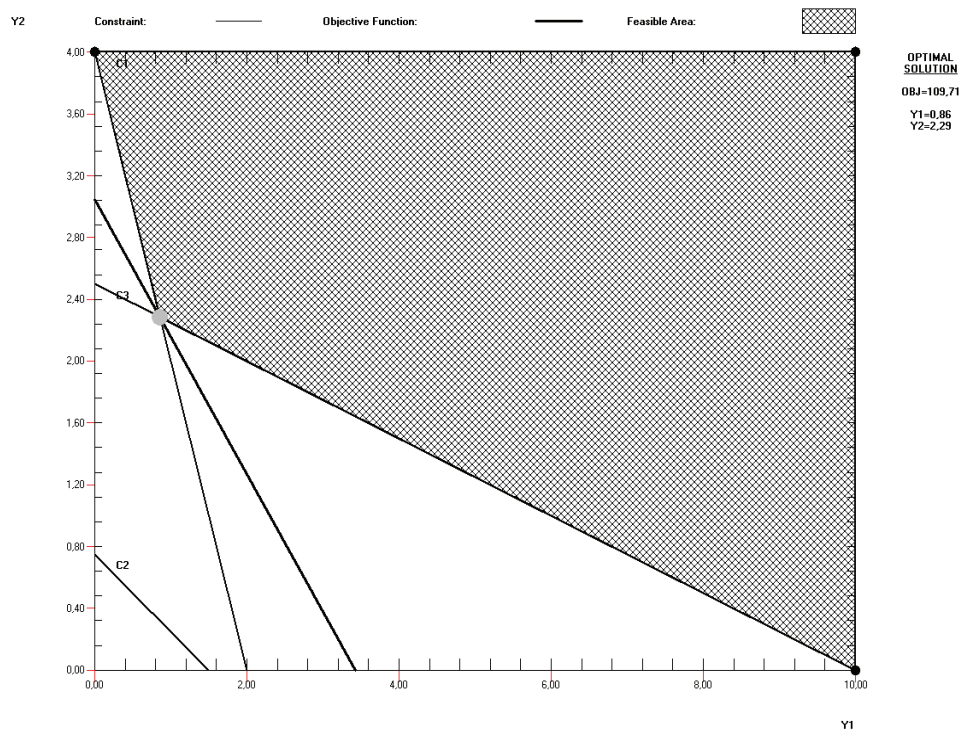
V tablica simpleksowa (Simplex Tableau – Iteration 5)

		Y1	Y2	Surplus_C1	Surplus_C2	Surplus_C3	Artificial_C1	Artificial_C2	Artificial_C3		
Basis	C(j)	32	36	0	0	0	0	0	0	RHS	Ratio
Surplus_C2	0	0	0	-0,29	1	-0,86	0,29	-1	0,86	7,86	
Y1	32	1	0	-0,29	0	0,14	0,29	0	-0,14	0,86	
Y2	36	0	1	0,07	0	-0,29	-0,07	0	0,29	2,29	
	C(j)-Z(j)	0	0	6,57	0	5,71	-6,57	0	-5,71	109,71	
	* Big M	0	0	0	0	0	1	1	1	0	

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Gdy porównamy rozwiązanie dualne zaznaczone czcionką pogrubioną z wynikami tabeli 2.7.11, to widzimy identyczność wyników w zakresie wartości zmiennych primalnych, dualnych oraz funkcji celu.

Interesuje nas z pewnością rozwiązanie geometryczne zadania dualnego. Zaprezentowano to graficznie, korzystając z modułu (LP) na wykresie 2.7.2.



Źródło: Opracowanie własne z zastosowaniem modułu (LP).

Wykres 2.7.2. Rozwiązanie graficzne zadania dualnego

Rozwiązanie optymalne – dualne: $y_1^* = 0,86$; $y_2^* = 2,29$

Przykład

Dana jest ostatnia tablica simpleksowa o rozwiązaniu optymalnym. Określ przedział zmienności ceny c_1 wyrobu X_1 .

Tabela 2.7.22

Końcowa tablica simpleksowa

	X_j	X_1	X_2	X_3	X_4	X_5	Rozwiązanie	Θ_i
x_b	c_b/c_j	5	2	7	0	0	b_i	
X_1	c_1	1	0	0	-1/3	2/3	8	
X_3	7	0	1	1	2/3	-1/3	4	
	z_j	5	7	0	3	1		
	k_j	0	-5	0	-3	-1	68	

Źródło: Opracowanie własne z zastosowaniem modułu (LP).

$$k_2 = 2 - (0 \cdot c_1 + 1 \cdot 7) \leq 0,$$

$$k_4 = 0 - (-1/3 \cdot c_1 + 2/3 \cdot 7) \leq 0,$$

$$k_5 = 0 - (2/3 \cdot c_1 - 1/3 \cdot 7) \leq 0,$$

zatem:

$$c_1 \leq 14,$$

$$c_1 \geq 3,5,$$

$$c_1 \in \langle 3,5, 14 \rangle.$$

Przykład

Określ granicę zmienności zasobu b_1 , mając daną macierz odwrotną rozwiązania optymalnego podanego w powyższym przykładzie oraz wektor wyrazów wolnych.

$$\mathbf{B}^{-1} = \begin{bmatrix} -1/3 & 2/3 \\ 2/3 & -1/3 \end{bmatrix}, \quad \mathbf{b}_i = \begin{bmatrix} b_1 \\ 20 \end{bmatrix},$$

$$-\frac{1}{3}b_1 + \frac{2}{3} \cdot 20 \geq 0, \quad b_1 \leq 40,$$

$$\frac{2}{3}b_1 + \left(-\frac{1}{3} \cdot 20\right) \geq 0, \quad b_1 \geq 10, \quad \text{a więc } b_1 \in \langle 10; 40 \rangle.$$

Przykład

Oblicz, w jakim przedziale mogą się zmieniać współczynniki a_{11} oraz a_{21} nakładów jednostkowych na wykonanie wyrobu X_1 . Dana jest macierz odwrotna z rozwiązania optymalnego (tab. 2.7.22) oraz wektor współczynników $a_{i1} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ z tablicy bazowej.

Obliczenie przedziału dla współczynnika a_{11} :

$$\mathbf{B}^{-1} = \begin{bmatrix} -1/3 & 2/3 \\ 2/3 & -1/3 \end{bmatrix}, \quad a_{i1} = \begin{bmatrix} a_{11} \\ 2 \end{bmatrix},$$

Mnożymy macierz odwrotną i wektor współczynników a_{i1} i otrzymujemy wektor:

$$\begin{aligned} \left(-\frac{1}{3}a_{11} + \frac{2}{3} \cdot 2\right) &= -\frac{a_{11}}{3} + \frac{4}{3}, & \begin{bmatrix} -\frac{a_{11}}{3} + \frac{4}{3} \\ \frac{2}{3}a_{11} - \frac{2}{3} \end{bmatrix} \\ \left(\frac{2}{3}a_{11} - \frac{1}{3} \cdot 2\right) &= \frac{2}{3}a_{11} - \frac{2}{3}, \end{aligned}$$

Według tabeli 2.7.7 parametr pomocniczy z_1 dla zmiennej x_1 obliczamy ze wzoru:
 $z_j = \mathbf{c}_b^T \mathbf{B}_l^{-1} \mathbf{A}$,

$$\mathbf{c}_b^T \text{ dla zmiennej } x_i, \quad \mathbf{c}_b = \begin{bmatrix} 5 \\ 7 \end{bmatrix}.$$

$$z_1 = 5 \cdot \left(-\frac{a_{11}}{3} + \frac{4}{3}\right) + 7 \cdot \left(\frac{2}{3}a_{11} - \frac{2}{3}\right) = 3a_{11} + 2,$$

$$k_1 = 5 - (3a_{11} + 2) = 3 - 3a_{11} \leq 0, \quad a_{11} \geq 1,$$

$$\text{zatem } a_{11} = \langle 1; \infty \rangle.$$

Obliczenie przedziału dla współczynnika a_{21} :

$$\mathbf{B}^{-1} = \begin{bmatrix} -1/3 & 2/3 \\ 2/3 & -1/3 \end{bmatrix}, \quad a_{i1} = \begin{bmatrix} 1 \\ a_{21} \end{bmatrix}.$$

Mnożymy macierz odwrotną i wektor współczynników a_{i1} i otrzymujemy wektor:

$$\begin{aligned} \left(-\frac{1}{3} \cdot 1 + \frac{2}{3} \cdot a_{21}\right) &= -\frac{1}{3} + \frac{2}{3}a_{21}, & \begin{bmatrix} -\frac{1}{3} + \frac{2}{3}a_{21} \\ \frac{2}{3} - \frac{1}{3}a_{21} \end{bmatrix} \\ \left(\frac{2}{3} \cdot 1 - \frac{1}{3} \cdot a_{21}\right) &= \frac{2}{3} - \frac{1}{3}a_{21}, \end{aligned}$$

$$z_1 = 5 \cdot \left(-\frac{1}{3} + \frac{2}{3}a_{21}\right) + 7 \cdot \left(\frac{2}{3} - \frac{1}{3}a_{21}\right) = 3 + a_{21},$$

$$k_1 = 5 - (3 + a_{21}) = 2 - a_{21} \leq 0, \quad a_{21} \geq 2,$$

$$\text{czyli } a_{21} = \langle 2; \infty \rangle.$$

2.7.1. Pytania

1. Wyjaśnij pojęcia „zmiennie bazowe” oraz „zmiennie niebazowe” w rozwiązaniu optymalnym według metody simpleks.
2. Opisz procedurę określania przedziałów parametrów w funkcji celu, aby rozwiązanie optymalne nie uległo zmianie dla zmiennych bazowych X_1, X_3 oraz niebazowych X_2, X_4, X_5 , przy czym X_4 i X_5 to zmienne swobodne.
3. Przyjmij, że dana jest macierz odwrotna z ostatniej tablicy simpleksowej (o rozwiązaniu optymalnym) oraz wektor wyrazów wolnych b_i . Podaj sposób określania przedziałów wartości wyrazów wolnych, aby struktura rozwiązania optymalnego nie uległa zmianie.
4. Dana jest macierz odwrotna z rozwiązania optymalnego oraz wektor parametrów współczynników nakładów jednostkowych na produkcję wyrobu X_1 . Opisz procedurę postępowania do wyznaczania przedziałów dla współczynników a_{11} oraz a_{21} zmiennej X_1 , aby rozwiązanie optymalne pozostawało niezmienione.

2.7.2. Zadania

Dana jest ostatnia tablica simpleksowa o rozwiązaniu optymalnym. Określ przedział zmienności ceny c_1 wyrobu X_1 .

Tabela 2.7.23

Końcowa tablica simpleksowa

	x_j	x_1	x_2	x_3	x_4	x_5	Rozwiązanie b_i	θ_i
x_b	c_b/c_j	3	2	5	0	0		
x_1	3	1	-1/3	0	2/3	-1/3	8/3	
x_3	5	0	5/3	1	-1/3	2/3	2/3	
	z_j	3	22/3	5	1/3	7/3	<i>FC: 94/3</i>	
	k_j	0	-16/3	0	-1/3	-7/3		

Źródło: Opracowanie własne w Excelu.

2. Określ granicę zmienności b_i (b_1, b_2) zasobów, mając daną macierz odwrotną rozwiązania optymalnego oraz wektor wyrazów wolnych.

$$\mathbf{B}^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}, \quad \mathbf{b}_i = \begin{bmatrix} 10 \\ 20 \end{bmatrix}.$$

3. Oblicz, w jakim przedziale mogą się zmieniać współczynniki a_{11} oraz a_{21} nakładów jednostkowych na wykonanie wyrobu X_1 . Dana jest macierz odwrotna z rozwiązania optymalnego oraz wektor współczynników a_{i1} z tablicy bazowej.

$$\mathbf{B}^{-1} = \begin{bmatrix} 2/3 & -1/3 \\ -1/3 & 2/3 \end{bmatrix}, \quad a_{i1} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

2.8. Zadanie dualne

Dotychczasowa nasza uwaga koncentrowała się na zadaniu decyzyjnym prymalnym. Bazując na zadaniu prymalnym (ZP), zwanym również pierwotnym, formułujemy zadanie dualne (ZD), zwane również programem dualnym¹⁶⁵. Zamieszczone w niniejszym rozdziale wzory oraz twierdzenia opierają się na publikacji pod redakcją W. Sikory¹⁶⁶.

Zadanie prymalne

$$FC_P = Z_P = \mathbf{c}\mathbf{x} \rightarrow \max,$$

$$C_i: \mathbf{A}\mathbf{x} \leq \mathbf{b},$$

$$x \geq \mathbf{0},$$

$$x \in R^n.$$

Zadanie dualne

$$FC_D = Z_D = \mathbf{b}^T \mathbf{y} \rightarrow \min,$$

$$C_i: \mathbf{A}^T \mathbf{y} \geq \mathbf{c}^T,$$

$$y \geq \mathbf{0},$$

$$y \in R^m.$$

P r z y k ł a d

Mając zadanie prymalne, określ zadanie dualne.

Zmienne: x_1, x_2, x_3

$$FC_P = f(x_1, x_2, x_3) = 3x_1 + 2x_2 + 5x_3 \rightarrow \max,$$

$$C_1: x_1 + x_2 + 2x_3 \leq 15,$$

$$C_2: x_1 + 2x_2 + x_3 \leq 18,$$

$$x_1, x_2, x_3 \geq 0,$$

y_1, y_2

$$FC_D = g(y_1, y_2) = 15y_1 + 18y_2 \rightarrow \min,$$

$$C_1: y_1 + y_2 \geq 3,$$

$$C_2: y_1 + 2y_2 \geq 2,$$

$$C_3: 2y_1 + y_2 \geq 5,$$

$$y_1, y_2 \geq 0.$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 15 \\ 18 \end{bmatrix}, \quad \mathbf{c} = [3 \ 2 \ 5], \quad \mathbf{A}^T = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{b}^T = [15 \ 18], \quad \mathbf{c}^T = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}.$$

Twierdzenia

1. Zadanie pierwotne jest zadaniem dualnym do swojego zadania dualnego.

¹⁶⁵ K. Kukuła (red.), *Badania operacyjne w przykładach i zadaniach...*, op.cit., s.59.

¹⁶⁶ W. Sikora (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii...*, op.cit., rozdz. 1.3.

2. Jeżeli zadanie prymalne jest zadaniem na maksimum, to zadanie dualne jest na minimum.
3. Ilość zmiennych w zadaniu prymalnym jest równa ilości ograniczeń w zadaniu dualnym.
4. Ilość ograniczeń w zadaniu prymalnym jest równa ilości zmiennych w zadaniu dualnym.
5. Współczynniki stojące przy zmiennych w funkcji celu w zadaniu prymalnym są wyrazami wolnymi w ograniczeniach zadania dualnego i na odwrót.
6. Macierz współczynników ograniczeń w zadaniu dualnym jest równa transponowanej macierzy współczynników ograniczeń zadania prymalnego.
7. Znak nierówności w ograniczeniu zadania prymalnego jest związany ze znakiem odpowiedniej zmiennej w zadaniu dualnym. Znak zmiennej zadania prymalnego jest związany z odpowiednim znakiem nierówności w ograniczeniu zadania dualnego.
8. Naturalne ograniczenia z zadania prymalnego przechodzą na naturalne ograniczenia zadania dualnego. Nienaturalne na nienaturalne.

T a b e l a 2.8.1

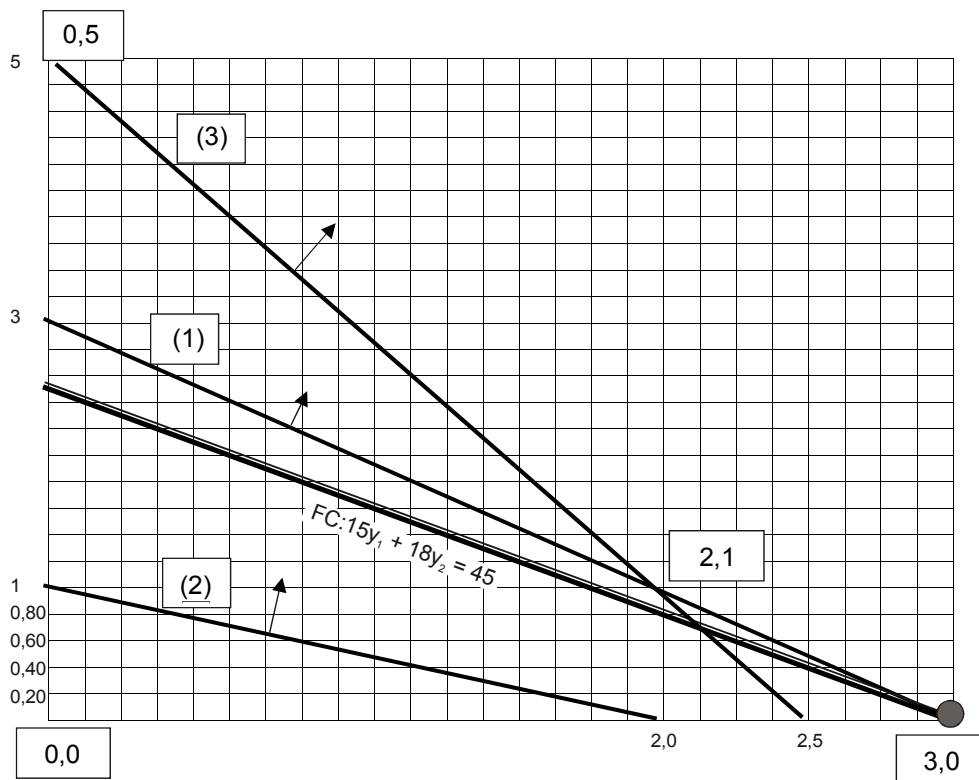
Ograniczenia naturalne

Zadanie na:	Znak ograniczenia	Znak zmiennej
Maksimum	\leq	\geq
Minimum	\geq	\geq

Ź r ó d ł o: Opracowanie własne.

9. Jeżeli istnieje rozwiązanie optymalne zadania prymalnego, to istnieje również rozwiązanie zadania dualnego i oba rozwiązania dają tę samą wartość funkcji celu.
10. Jeżeli w zadaniu prymalnego i -te ograniczenie ma postać równania, to i -ta zmienna w zadaniu dualnym jest dowolna co do znaku. Jeżeli j -ta zmienna zadania prymalnego jest dowolna co do znaku, to j -te ograniczenie zadania dualnego ma postać równania.
11. Jeżeli dla optymalnych rozwiązań dopuszczalnych zadania prymalnego i dualnego k -te ograniczenie dowolnego zadania jest spełnione z ostrą nierównością, to odpowiadająca mu k -ta zmienna w zadaniu dualnym jest spełniona z równością.

Rozwiązanie graficzne manualne zadania dualnego.



Źródło: Opracowanie własne w Excelu; siatka komputerowa 25x25, przy czym siatka rzeczywista kolumn 0,12, a wierszy 0,20; oś odcięta y_1 a oś rzędna y_2 .

Wykres 2.8.1. Graficzne rozwiązanie zadania dualnego

$$(y_1^*, y_2^*) = (3, 0), Z_D = g(y_1^*, y_2^*) = 45.$$

Warunki rozwiązania (y_1^*, y_2^*) :

$$C_1: 3 + 0 = 3, \quad (x_1^* \neq 0), \quad C_2: 3 + 0 > 2, \quad x_2^* = 0 \quad C_3: 6 + 0 > 5, \quad x_3^* = 0.$$

Z obliczeń zadania dualnego wynika, że:

$$C_2 \text{ i } C_3 \text{ zadania dualnego są spełnione z ostrą nierównością, zatem } x_2^* = x_3^* = 0 \text{ w zadaniu prymalnym.}$$

$$\text{Ponieważ według twierdzenia 9. } Z_D^* = Z_P^* = f(x_1, x_2, x_3) = 3x_1 + 2x_2 + 5x_3 = 3x_1 + 0 + 0 = 45, \text{ zatem } x_1^* = 15 \text{ w zadaniu prymalnym.}$$

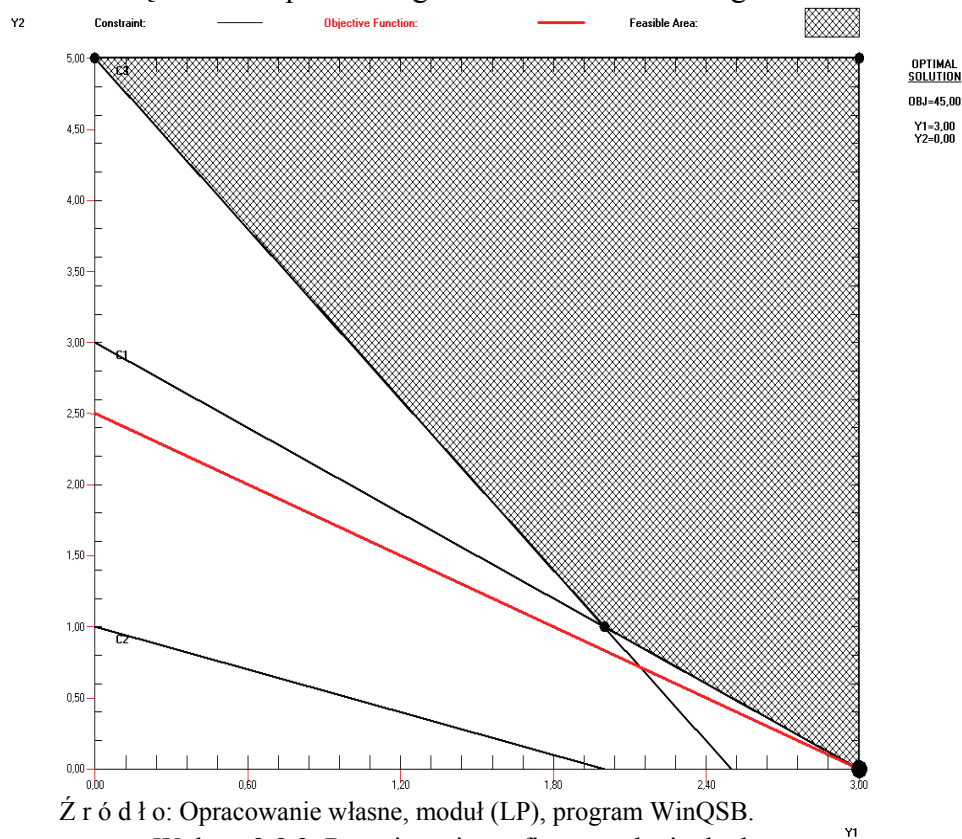
Tabela 2.8.2

Przykładowa interpretacja zadania primalnego i dualnego

ZP	ZD
W jakich ilościach produkować wyroby, by maksymalizować zysk?	Jakie powinny być minimalne ceny dualne zasobów, by zysk był większy lub równy cenie wyrobów.
Ograniczenia – zużycie zasobu i musi być mniejsze lub równe dostępnej ilości tego zasobu.	Ograniczenia – ceny dualne ilości zasobów potrzebne na wyprodukowanie jednostki wyrobu j muszą być większe lub równe cenie jednostki tego wyrobu.
Przykładowa funkcja celu: $Z_P = \mathbf{c} \cdot \mathbf{x} \rightarrow \max$ $A_x \leq b$ (ograniczenia) $x \geq 0$ Z_P – dochód ze sprzedaży wszystkich wyrobów (max) \mathbf{x} – produkcja wyrobów j \mathbf{c} – cena jednostkowa wyrobów j	Przykładowa funkcja celu: $Z_D = \mathbf{b}^T \cdot \mathbf{y} \rightarrow \min$ $A^T \mathbf{y} \geq \mathbf{c}^T$ $y \geq 0$ Z_D – zużycie wszystkich surowców (min) \mathbf{y} – udział zasobów i w cenie wyrobów j \mathbf{b} – dostępna ilość jednostek zasobów i

Źródło: Opracowanie własne.

Rozwiązanie komputerowe graficzne zadania dualnego.



Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Wykres 2.8.2. Rozwiązanie graficzne zadania dualnego

P r z y k ł a d

Rozwiąż zadanie decyzyjne z rozdziału 2.5, formułując zadanie dualne.

Procedura manualna rozwiązania zadania dualnego (ZD) bazuje na danych z przykładu zadania primalnego (ZP) wykonywanego z zastosowaniem rachunku macierzeowego i funkcji Excela.

Zadanie primalne (zamiana na postać kanoniczną):

$$\begin{array}{ll}
 FC_p : 40x_1 + 30x_2 \rightarrow \max, & FC_p : 40x_1 + 30x_2 + 0x_3 + 0x_4 + 0x_5 \rightarrow \max, \\
 C_1 : 4x_1 + 2x_2 \leq 2000, & C_1 : 4x_1 + 2x_2 + 1x_3 = 2000, \\
 C_2 : 2x_1 + 2x_2 \leq 1400, & C_2 : 2x_1 + 2x_2 + 1x_4 = 1400, \\
 C_3 : 3x_1 \leq 1200, & C_3 : 3x_1 + 1x_5 = 1200. \\
 x_1, x_2 \geq 0. & x_1, x_2 \geq 0.
 \end{array}$$

Zadanie dualne (zamiana na postać uzupełnioną w celu uzyskania macierzy jednostkowej):

$$\begin{array}{ll}
 FC_D : 2000y_1 + 1400y_2 + 1200y_3 \rightarrow \min, & FC_D : 2000y_1 + 1400y_2 + 1200y_3 + 0y_4 + 0y_5 + Ms_1 + Ms_2 \rightarrow \min, \\
 C_1 : 4y_1 + 2y_2 + 3y_3 \geq 40, & C_1 : 4y_1 + 2y_2 + y_3 - 1y_4 + 1s_1 = 40, \\
 C_2 : 2y_1 + 2y_2 \geq 30, & C_2 : 2y_1 + 2y_2 - 1y_5 + 1s_2 = 30. \\
 y_1, y_2 \geq 0. &
 \end{array}$$

gdzie: s_1, s_2 – zmienne sztuczne,
 y_4, y_5 – zmienne swobodne.

Rozwiązanie manualne zadania dualnego wspomagane komputerowo programem Excel.

T a b e l a 2.8.3.

I tablica simpleksowa (bazowa) do zadania dualnego

c_b	c_j	2 000	1 400	1 200	0	0	M	M	Rozwiązanie b_i
	Zmienne bazowe	y_1	y_2	y_3	y_4	y_5	s_1	s_2	
M	s_1	4	2	3	-1	0	1	0	40
M	s_2	2	2	0	0	-1	0	1	30
Suma	z_j	$6M$	$4M$	$3M$	$-M$	$-M$	M	M	$FC : 70M$
	$k_j = c_j - z_j$	$2\ 000 - 6M$	$1\ 400 - 4M$	$1\ 200 - 3M$	M	M	0	0	

Kolumna rozwiązująca ze zmienną y_1 .

Wiersz rozwiązujący ze zmienną s_1 ; określony poprzez obliczenia Θ_i (patrz podrozdz. 2.6)

Element (komórka) rozwiązujący (centralny) – 4.

Ź r ó ł o: Opracowanie własne w Excelu.

Tabela 2.8.4

II tablica simpleksowa (do zadania dualnego)

c_b	c_j	2000	1400	1200	0	0	M	M	Rozwiązanie b_j
	Zmienne bazowe - y_b	y_1	y_2	y_3	y_4	y_5	s_1	s_2	
2000	y_1	1	1/2	3/4	-1/4	0	1/4	0	10
M	s_2	0	1	-3/2	1/2	-1	-1/2	1	10
Suma	z_j	2000	1 000+ M	1 500-3/2 M	-500+1/2 M	- M	500-1/2 M	M	FC: 2000+10 M
	$k_j = c_j - z_j$	0	400-M	-300+3/2 M	500-1/2 M	M	-500+3/2 M	0	

Kolumna rozwiązująca ze zmienną y_2 .

Wiersz rozwiązujący ze zmienną s_2 .

Element rozwiązujący – 1.

Ź r ó d ł o: Opracowanie własne, przykładowe zastosowanie rachunku macierzowego wykonanego z zastosowaniem funkcji Excela.

W budowie II tablicy simpleksowej zastosowano przekształcenie tablicy bazowej według macierzowej postaci tablicy dla j -tej iteracji (tab. 2.8.5).

Tabela 2.8.5

Tablica simpleksowa „ l ” iteracji

Zmienne bazowe	c		Rozwiązanie
x_{bl}	$B_l^{-1} A$	B_l^{-1}	$B_l^{-1} b$
z_j	$c_b^T B_l^{-1} A$	$c_b^T B_l^{-1}$	FC: $c_b^T B_l^{-1} b$
$k_j = c_j - z_j$	$c_j - c_b^T B_l^{-1} A$	$-c_b^T B_l^{-1}$	

Ź r ó d ł o: Opracowanie własne, nawiązanie do tab. 2.3.2.

W odniesieniu do II tablicy simpleksowej (tab. 2.8.4) oraz danych z tabeli bazowej przykład poniżej przedstawia wyniki zastosowania funkcji Excela: MACIERZ.ODW(...); MACIERZ.ILOCZYN(...), bazujący na macierzy A , macierzy bazowej II iteracji B_2 , wektorach: b , y_b , c_b .

$$A = \begin{bmatrix} 4 & 2 & 3 \\ 2 & 2 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 40 \\ 30 \end{bmatrix}, \quad y_b = \begin{bmatrix} y_1 \\ s_2 \end{bmatrix}, \quad c_b = \begin{bmatrix} 2000 \\ M \end{bmatrix},$$

$$c = [2000 \quad 1400 \quad 1200 \quad 0 \quad 0 \quad M \quad M].$$

$$B_2^{-1} = \begin{bmatrix} 0,25 & 0 \\ -0,5 & 1 \end{bmatrix}, \quad B_2^{-1} \cdot A = \begin{bmatrix} 1 & 0,5 & 3/4 \\ 0 & 1 & -3/2 \end{bmatrix}, \quad B_2^{-1} b = \begin{bmatrix} 10 \\ 10 \end{bmatrix}.$$

Obliczenie wartości w wierszu z_j (jako sumy iloczynów $a_{ij} \cdot c_{bi}$).

Excel nie realizuje mnożenia przez M i dlatego liczymy manualnie wartości w wierszu z_j oraz k_j np. dla zmiennej y_2 :

$$z_j = (1/2 \cdot 2000) + (M \cdot 1) = 1000 + M, \quad k_j = 1400 - (1000 + M) = 400 - M.$$

Dla odmiany trzecią, a zarazem ostatnią tablicę simpleksową obliczono według sposobu elementu rozwiązującego algorytmu simpleks. Procedurę postępowania w tym zakresie podano w podrozdz. 2.4.

Tabela 2.8.6

III tablica simpleksowa (do zadania dualnego)

c_b	c_j	2 000	1 400	1 200	0	0	M	M	Rozwiązanie b_i
	Zmienne bazowe y_b	y_1	y_2	y_3	y_4	y_5	s_1	s_2	
2 000	y_1	1	0	3/2	-1/2	1/2	1/2	-1/2	5
1 400	y_2	0	1	-3/2	1/2	-1	-1/2	1	10
Suma	z_j	2 000	1 400	900	-300	-400	300	400	FC: 24 000
$k_j =$	$c_j - z_j$	0	0	300	300	400	$M - 300$	$M - 400$	

Źródło: Opracowanie własne w Excelu.

Sprawdzenie, czy występuje już rozwiązanie optymalne: $k_j \geq 0$ (dla wszystkich zmiennych).

Zatem mamy już rozwiązanie optymalne: $y_1^* = 5$; $y_2^* = 10$; $y_3^* = 0$. Sprawdźmy rozwiązanie manualne zadania dualnego, stosując moduł (LP) programu WinQSB (tab. 2.8.7).

Tabela 2.8.7

Raport kombinowany zadania dualnego

	10:54:51	Friday	December	7	2007			
	Decision	Solution	Unit Cost or	Total	Reduced	Basis	Allowable	Allowable
	Variable	Value	Profit $C(j)$	Contribution	Cost	Status	Min. $C(j)$	Max. $C(j)$
1	Y1	5	2 000	10 000	0	Basis	1 400	2 200
2	Y2	10	1 400	14 000	0	Basis	1 200	2 000
3	Y3	0	1 200	0	300	At bound	900	M
	Objective	Function	(Min.) =	24 000				
		Left Hand		Right Hand	Slack	Shadow	Allowable	Allowable
	Constraint	Side	Direction	Side	or Surplus	Price	Min. RHS	Max. RHS
1	C1	40	>=	40	0	300	30	60
2	C2	30	>=	30	0	400	20	40

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Nastąpiło potwierdzenie rozwiązania z tą samą wartością funkcji celu, jaka wystąpiła w tab. 2.8.6. Rozwiązaniem dualnym do naszego zadania dualnego jest wcześniej obliczone według rachunku macierzowego rozwiązanie prymalne (tab. 2.5.5), w którym $x_1^* = 300$, $x_2^* = 400$. Pokazuje to też tab. 2.8.7 w kolumnie Shadow Price.

2.8.1. Pytania

1. Podaj zasady wprowadzania zmiennych swobodnych oraz sztucznych do ograniczeń i funkcji celu programu liniowego.

2. Określ postać kanoniczną wymienionych modeli matematycznych programowania liniowego:

a)

$$FC: 9x_1 + 6x_2 \rightarrow \max,$$

$$C_1: 4x_1 + 2x_2 \leq 100,$$

$$C_2: 6x_1 + 6x_2 \leq 240,$$

$$C_3: 3x_1 \leq 300,$$

$$x_1, x_2 \geq 0,$$

b)

$$FC: 9x_1 + 6x_2 \rightarrow \min,$$

$$C_1: 4x_1 + 2x_2 \geq 100,$$

$$C_2: 6x_1 + 6x_2 \geq 240,$$

$$C_3: 3x_1 \geq 300,$$

$$x_1, x_2 \geq 0,$$

c)

$$FC: 10x_1 + 4x_2 + 2x_3 \rightarrow \min,$$

$$C_1: 2x_1 + 4x_2 \geq 140,$$

$$C_2: 4x_1 + 6x_2 + x_3 = 280,$$

$$C_3: 6x_2 + 2x_3 = 210,$$

$$x_1, x_2, x_3 \geq 0.$$

3. Zamień zadanie prymalne na dualne:

$$FC: 6x_1 + 4x_2 + 10x_3 \rightarrow \max,$$

$$C_1: 2x_1 + 2x_2 + 4x_3 \leq 30,$$

$$C_2: 2x_1 + 4x_2 + 2x_3 \leq 36,$$

$$C_3: x_1, x_2, x_3 \geq 0.$$

4. Objasnij twierdzenia o dualności na przykładzie podanych zadań prymalnego i dualnego:

ZP

$$FC_P: 3x_1 + 2x_2 + 5x_3 \rightarrow \max,$$

$$C_1: x_1 + x_2 + 2x_3 \leq 15,$$

$$C_2: x_1 + 2x_2 + x_3 \leq 18,$$

$$x_1, x_2, x_3 \geq 0.$$

ZD

$$FC_D: 15y_1 + 18y_2 \rightarrow \min,$$

$$C_1: y_1 + y_2 \geq 3,$$

$$C_2: y_1 + 2y_2 \geq 2,$$

$$C_3: 2y_1 + y_2 \geq 5.$$

5. Zinterpretuj na dowolnym przykładzie ceny dualne rozwiązania prymalnego o znanej wartości funkcji celu.

2.8.2. Zadania

1. Określ pierwszą tablicę bazową do podanego modelu matematycznego:

$$FC: x_1 + x_2 + x_3 \rightarrow \max,$$

$$C_1: x_1 + x_2 \geq 8,$$

$$C_2: 2x_1 + x_2 + x_3 = 14,$$

$$C_3: x_1 + 2x_2 \leq 10,$$

$$x_1, x_2, x_3 \geq 0.$$

2. Wyznacz graficznie wartości zmiennych y_1, y_2 do podanego modelu zadania dualnego:

$$FC_D: 30y_1 + 36y_2 \rightarrow \min,$$

$$C_1: 2y_1 + 2y_2 \geq 6,$$

$$C_2: 2y_1 + 4y_2 \geq 4,$$

$$C_3: 4y_1 + 2y_2 \geq 10,$$

$$y_1, y_2 \geq 0.$$

3. Określ tablicę bazową oraz kolejną drugą tablicę simpleksową do zadania dualnego według sposobu elementu rozwiązującego:

$$FC_D: 30y_1 + 36y_2 \rightarrow \min,$$

$$C_1: 2y_1 + 2y_2 \geq 6,$$

$$C_2: 2y_1 + 4y_2 \geq 4,$$

$$C_3: 4y_1 + 2y_2 \geq 10,$$

$$y_1, y_2 \geq 0.$$

4. Sformułuj zadania dualne do wymienionych zadań prymalnych:

a)

$$FC_P: 2x_1 + 5x_2 + x_3 \rightarrow \min,$$

$$C_1: 2x_1 + x_2 \geq 70,$$

$$C_2: 3x_1 + 2x_2 + x_3 \geq 140,$$

$$C_3: x_1 + 3x_2 \geq 105,$$

$$x_1, x_2, x_3 \geq 0.$$

b)

$$FC_P: 2x_1 + 5x_2 + x_3 \rightarrow \max,$$

$$C_1: 2x_1 + x_2 \leq 70,$$

$$C_2: 3x_1 + 2x_2 + x_3 \leq 140,$$

$$C_3: x_1 + 3x_2 \leq 105,$$

$$x_1, x_2, x_3 \geq 0.$$

c)

$$FCP: 2x_1 + 2x_2 \rightarrow \min,$$

$$C_1: 2x_1 + x_2 \geq 20,$$

$$C_2: x_1 + 2,5x_2 \geq 20,$$

$$C_3: 5x_1 + x_2 \geq 20,$$

$$x_1, x_2 \geq 0.$$

d)

$$FCP: 2x_1 + 5x_2 + x_3 \rightarrow \max,$$

$$C_1: 2x_1 + x_2 + 2x_3 \leq 70,$$

$$C_2: 3x_1 + 2x_2 + x_3 \leq 140,$$

$$x_1, x_2, x_3 \geq 0.$$

5. Po zamianie przytoczonych zadań na zadania dualne rozwiąż przykład c) z użyciem algorytmu simpleks, stosując rachunek macierzowy. Porównaj otrzymane wyniki z uzyskanymi przy zastosowaniu modułu (LP) programu WinQSB w ostatniej 3. iteracji ($y_1^* = 0,75$; $y_2^* = 0,5$).

Tabela 2.8.8

III tablica simpleksowa (Simplex Tableau – Iteration 3)

		Y1	Y2	Y3	Slack_C 1	Slack_C 2		
x_b	$C(j)$	20	20	20	0	0	RHS	Ratio
Y1	20	1	0	2,88	0,63	-0,25	0,75	
Y2	20	0	1	-0,75	-0,25	0,5	0,5	
	$C(j)-Z(j)$	0	0	-22,5	-7,5	-5	25	

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

2.9. Zagadnienia nieliniowe

By lepiej opisać rzeczywistość, jaka występuje w procesach gospodarczych, wprowadzono modele nieliniowe. Zadanie programowania nieliniowego formułujemy jako¹⁶⁷:

$$\begin{aligned} f(x) = f(x_1, x_2, \dots, x_n) \rightarrow \min \text{ (lub max)}, & \quad g_i(x) = g_i(x_1, x_2, \dots, x_n) \leq 0 \text{ lub } \geq 0, \\ (i = 1, 2, \dots, r), & \quad x_1, x_2, \dots, x_n \geq 0. \end{aligned}$$

Założenie: funkcje f i g są ciągłe i przynajmniej jedna z nich nie jest liniowa. Program nieliniowy może mieć postać:

- kanoniczną (wszystkie warunki ograniczające są równościami),
- standardową (wszystkie warunki ograniczające są nierównościami).

Nie występuje uniwersalna metoda rozwiązywania programów nieliniowych. Rozpatrzmy metodę mnożników Lagrange'a oraz metodę polegającą na wprowadzeniu zmiennych nieistotnych.

2.9.1. Metoda mnożników Lagrange'a

To metoda stosowana do programów nieliniowych i nie gwarantuje jednak uzyskania nieujemnych zmiennych decyzyjnych, co jest jej wadą. Obejmuje etapy:

- 1) Sprawdzenie, czy $f(x)$ ma ekstremum bezwarunkowe, a jeśli tak, to czy jest też ekstremum warunkowym, czyli spełnia ograniczenia. Wpierw wyznaczamy pochodne cząstkowe funkcji celu $f(x)$ względem zmiennych decyzyjnych i przyrównujemy do zera, a następnie określamy wyznacznik¹⁶⁸:

¹⁶⁷ K. Kukuła (red.), *Badania operacyjne w przykładach i zadaniach...*, op.cit., rozdz. 6.

¹⁶⁸ Ibidem, s. 214.

$$\det \mathbf{B} = \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{vmatrix} > 0.$$

Sprawdzamy też, czy wartości minorów głównych macierzy są dodatnie. Jeśli ekstremum bezwarunkowe spełnia warunki, to przyjmujemy jako rozwiązanie minimum wartości x_j z pierwszych pochodnych.

2) Przekształcenie funkcji celu w funkcję Lagrange'a (gdy ekstremum bezwarunkowe nie spełnia ograniczeń) poprzez wprowadzenie nieoznaczonych mnożników λ_i .

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^r \lambda_i g_i(\mathbf{x}). \quad (2.9.1)$$

3) Obliczamy pochodne cząstkowe funkcji Lagrange'a względem x_j , a także λ_i , przyrównując je do zera. Rozwiązanie otrzymanego układu równań jest przeważnie rozwiązaniem optymalnym.

Przykład

Zakład gazowniczy rozsyła gaz do dwóch ciepłowni. Funkcja miesięcznych kosztów eksploatacji sieci przesyłowej gazu w zł określona jest jako:

$$f(x_1, x_2) = 7x_1^2 - 5x_1x_2 + 6x_2^2 - 10x_1 - 6x_2 + 50.$$

Niech x_1 określa wielkość przesyłu gazu do ciepłowni pierwszej, a x_2 do drugiej. Chcemy zminimalizować koszty przesyłu gazu, którego miesięczna produkcja wynosi 60 tys. m³.

Procedura jest następująca:

a) Szukamy minimum funkcji przy ograniczeniu $g_1(x_1, x_2) = x_1 + x_2 = 60$ i warunkach brzegowych $x_1, x_2 \geq 0$:

$$\frac{\partial f}{\partial x_1} = 14x_1 - 5x_2 - 10 = 0,$$

$$\frac{\partial f}{\partial x_2} = -5x_1 + 12x_2 - 6 = 0.$$

W rozwiązaniu tego układu równań skorzystajmy ze wzorów Cramera jako ilorazów wyznaczników macierzy¹⁶⁹:

$$|\mathbf{A}| = \begin{vmatrix} 14 & -5 \\ -5 & 12 \end{vmatrix} = 143, \quad |\mathbf{A}_1| = \begin{vmatrix} 10 & -5 \\ 6 & 12 \end{vmatrix} = 150, \quad |\mathbf{A}_2| = \begin{vmatrix} 14 & 10 \\ -5 & 6 \end{vmatrix} = 134,$$

$$x_1 = \frac{|\mathbf{A}_1|}{|\mathbf{A}|}, \quad x_1 = \frac{150}{143} = 1,05, \quad x_2 = \frac{|\mathbf{A}_2|}{|\mathbf{A}|}, \quad x_2 = \frac{134}{143} = 0,94.$$

Punkt ekstremum bezwarunkowego: $(x_1, x_2) = (1,05; 0,94)$.

b) Określamy macierz drugich pochodnych i obliczamy wyznacznik:

$$\mathbf{B} = \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix} = \begin{vmatrix} 14 & -5 \\ -5 & 12 \end{vmatrix}, \quad |\mathbf{B}| = 143 > 0.$$

Ponieważ minory główne tej macierzy $(14; 12) > 0$, funkcja celu f osiąga minimum w $(x_1, x_2) = (1,05; 0,94)$.

Mamy jednak warunek ograniczający $g_1(x_1, x_2) = x_1 + x_2 = 60$, do którego podstawiamy współrzędne punktu i sprawdzamy ekstremum warunkowe:

$$1,05 + 0,94 = 1,99 < 60.$$

Zatem punkt ten nie jest ekstremum warunkowym funkcji celu $f(x_1, x_2)$ dla warunku:

$$g_1(x_1, x_2) = 60.$$

c) Tworzymy funkcję Lagrange'a:

$$L(x_1, x_2, \lambda) = 7x_1^2 - 5x_1x_2 + 6x_2^2 - 10x_1 - 6x_2 + 50 + \lambda(x_1 + x_2 - 60).$$

d) Określamy pochodne cząstkowe funkcji L , otrzymując układ trzech równań o niewiadomych x_1, x_2, λ :

$$\frac{\partial L}{\partial x_1} = 14x_1 - 5x_2 - 10 + \lambda = 0, \quad \frac{\partial L}{\partial x_2} = -5x_1 + 12x_2 - 6 + \lambda = 0, \quad \frac{\partial L}{\partial \lambda} = x_1 + x_2 - 60 = 0.$$

Po odjęciu dwóch pierwszych równań otrzymujemy:

$$19x_1 - 17x_2 = 4,$$

$$x_1 + x_2 = 60, \text{ z tego równania } x_1 = 60 - x_2.$$

Po podstawieniu do równania $19x_1 - 17x_2 = 4$ otrzymamy $x_2 = 31,56$ tys. m^3 . Wobec tego:

$$x_1 = 60 - 31,56 = 28,44 \text{ tys. } \text{m}^3.$$

¹⁶⁹ T. Jurlewicz, Z. Skoczylas, *Algebra liniowa...*, op.cit, s. 84.

e) Obliczmy miesięczne koszty eksploatacji sieci przesyłowej gazu do dwóch ciepłowni, podstawiając wyniki do funkcji celu:

$$f(x_1, x_2) = 7 \cdot 28,44^2 - 5 \cdot 28,44 \cdot 31,56 + 6 \cdot 31,56^2 - 10 \cdot 28,44 - 6 \cdot 31,56 + 50 = 6676,45 \text{ zł}$$

P r z y k ł a d

Rozwiążmy zadanie programowania nieliniowego, uwzględniając pomysł przykładowo w publikacji pod redakcją M. Lipiec-Zajchowskiej¹⁷⁰, lecz do rozwiązania zastosujemy moduł (NLP) (*Nonlinear Programming*) programu WinQSB.

Nasze zadanie dotyczy wyboru powtórzeń reklamy firmy. Do wyboru mamy $n = 4$ czasopisma, w których chcemy te reklamy umieścić. Zasięg reklamy zależy od nakładu danego czasopisma oraz częstotliwości pojawienia się w nim reklamy. Chcemy, aby ten zasięg był maksymalny. Interesuje nas, ile zamieszczeń reklamy powinno się znaleźć w każdym z czasopism, aby dotrzeć do jak największej liczby czytelników, biorąc pod uwagę ograniczony fundusz, jaki na ten cel posiadamy.

Oznaczmy:

A – fundusz przeznaczony na reklamę – 100 tys. zł,

a_j – koszt zamieszczenia jednej reklamy w j -tym czasopiśmie,

b_j – minimalna liczba zamieszczeń reklamy w j -tym czasopiśmie, d_j – maksymalna,

X_j – zmienna decyzyjna – liczba zamieszczeń reklamy w j -tym czasopiśmie,

$f_j(x_j)$ – zasięg reklamy (liczba czytelników j -tego czasopisma):

$$f_j(x_j) = c_j \cdot x_j - r_j \cdot x_j^2,$$

c_j – maksymalny zasięg reklamy w j -tym czasopiśmie, gdy pojawiła się ona tylko raz,

r_j – tempo spadku przeciętnego zasięgu reklamy w przypadku powtórzeń.

T a b e l a 2.9.1

Zasięg reklamy w czasopismach

Czasopismo j	a_j [tys. zł]	c_j [tys. czytelników]	r_j	b_j	d_j
1	5	50	5	0	5
2	8	100	8	0	5
3	15	120	10	0	5
4	20	180	10	0	5

Ź r ó ł o: Opracowanie własne na podstawie: M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003, tabl. 3.6.

¹⁷⁰ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne...*, op.cit. rozdz. 3.3.

Sformułowanie zadania decyzyjnego programowania nieliniowego – funkcja celu i ograniczenia:

$$FC: Z = \sum_{j=1}^n f_j(x_j) \rightarrow \max, \quad \sum_{j=1}^n a_j \cdot x_j \leq A, \quad b_j \leq x_j \leq d_j, \quad x_j - \text{całkowite}, \quad j=1, 2, \dots, n.$$

$$FC: Z = 50 \cdot x_1 - 5 \cdot x_1^2 + 100 \cdot x_2 - 8 \cdot x_2^2 + 120 \cdot x_3 - 10 \cdot x_3^2 + 180 \cdot x_4 - 10 \cdot x_4^2,$$

$$C_1: 5 \cdot x_1 + 8 \cdot x_2 + 15 \cdot x_3 + 20 \cdot x_4 \leq 100.$$

Jest to zatem zadanie programowania nieliniowego całkowitoliczbowego. Sformułujmy teraz zadanie decyzyjne z zastosowaniem modułu (NLP).

Tabela 2.9.2

Sformułowanie nieliniowego zadania decyzyjnego według rotacji modułu (NLP)

NLP	REKLAMA	4	1		
	OBJ / Constraint / Variable Bound				
Maximize	50*X1-5*X1^2+100*X2-8*X2^2+120*X3-10*X3^2+180*X4-10*X4^2				
C1	5*X1+8*X2+15*X3+20*X4<=100				
X1	>=0, <=5				
X2	>=0, <=5				
X3	>=0, <=5				
X4	>=0, <=5				

Źródło: Opracowanie własne w (NLP), program WinQSB.

W module (NLP) nie ma opcji ustawienia całkowitoliczbowości. Zatem wynik z rozwiązania sumarycznego zaokrąglimy do liczby całkowitej (tab. 2.9.3).

Tabela 2.9.3

Rozwiązanie sumaryczne

09-21-2007	Decision Variable	Solution Value	
1	X1	4,6352	5
2	X2	4,7745	5
3	X3	2,4999	2
4	X4	0,0532	0
Maximized	Objective Function =	666,4647	

Źródło: Opracowanie własne, moduł (NLP).

Zasięg reklamy (bez uwzględnienia zaokrąglenia częstości publikowania) wynosi 666 tys. czytelników. Sprawdźmy jeszcze nasze ograniczenie:

$$C_1: 5 \cdot x_1 + 8 \cdot x_2 + 15 \cdot x_3 + 20 \cdot x_4 = 5 \cdot 5 + 8 \cdot 5 + 15 \cdot 2 + 0 = 95 \leq 100.$$

Przykład

Zastosujmy teraz dodatek Solver programu Excel do rozwiązania poprzedniego przykładu z programowania nieliniowego, lecz z określeniem całkowitości zmiennych X_1, X_2, X_3, X_4 .

Kroki postępowania:

a) Sformułowanie zadania w arkuszu kalkulacyjnym (zainicjowanie komórek zmiennych jako wartości zerowe, w komórkach obliczanych wpisujemy wyrażenia jak w tab. 2.9.4).

Tabela 2.9.4

Sformułowanie zadania decyzyjnego w arkuszu kalkulacyjnym

0	=(50*\$B\$2-5*\$B\$2^2)+(100*\$B\$3-8*\$B\$3^2)+(120*\$B\$4-10*\$B\$4^2)+(180*\$B\$5-10*\$B\$5^2)										1
X1	0										2
X2	0										3
X3	0										4
X4	0										5
C1	=5*\$B\$2+8*\$B\$3+15*\$B\$4+20*\$B\$5										6
C2	PRAWDA =\$B\$6<=100										7
Parametry w Solver:											8
Komórka celu:			\$A\$1								9
Równa:			Maks								10
Komórki zmienne:			\$B\$2:\$B\$5								11
Komórki ograniczające:			B6		fundusz na reklamę						12
Dla komórek B2, B3, B4, B5 po 3 ograniczenia np. \$B\$2całkowita; \$B\$2>=0; \$B\$2<=5											13
Dla spełnienia warunku całkowitości zmiennych program dodaje symbol <i>int</i>											14
A	B	C	D	E	F	G	H	I	J	15	

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

b) Rozwiązanie z zastosowaniem dodatku Solver.

Przy prowadzeniu kryterium całkowitości otrzymujemy inną liczbę powtórzeń w czasopismach i wartość funkcji celu. Zasięg reklamy w stosunku do liczby czytelników jest jednak większy (tab. 2.9.5).

Tabela 2.9.5

Raport wyników z realizacji nieliniowego zadania decyzyjnego

Komórka celu (Maks)

Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$A\$1		0	793

Komórki decyzyjne

Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$B\$2	X1	0	1
\$B\$3	X2	0	3
\$B\$4	X3	0	2
\$B\$5	X4	0	2

Warunki ograniczające

Komórka	Nazwa	Wartość komórki	Formuła	Status	Luz
\$B\$6	C1	99	\$B\$6<=100	Niewiążące	1
\$B\$4	X3	2	\$B\$4=całkowita	Wiążące	0
\$B\$2	X1	1	\$B\$2=całkowita	Wiążące	0
\$B\$2	X1	1	\$B\$2>=0	Niewiążące	1
\$B\$2	X1	1	\$B\$2<=5	Niewiążące	4
\$B\$3	X2	3	\$B\$3>=0	Niewiążące	3
\$B\$3	X2	3	\$B\$3=całkowita	Wiążące	0
\$B\$3	X2	3	\$B\$3<=5	Niewiążące	2
\$B\$4	X3	2	\$B\$4>=0	Niewiążące	2
\$B\$4	X3	2	\$B\$4<=5	Niewiążące	3
\$B\$5	X4	2	\$B\$5<=5	Niewiążące	3
\$B\$5	X4	2	\$B\$5=całkowita	Wiążące	0
\$B\$5	X4	2	\$B\$5>=0	Niewiążące	2

Raport wyników Arkusz: [S284.xls]S284

2007-09-21 18:15:04

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

2.9.2. Wprowadzenie zmiennych nieistotnych

Metodę nieoznaczonych mnożników Lagrange'a można stosować także do programów nieliniowych o postaci standardowej, w których wszystkie ograniczenia są nierównościami. Zamieniamy nierówności na równości poprzez wprowadzenie zmiennych nieistotnych u^2 .

Przykład

Zaprezentujemy zagadnienie na przykładzie wcześniejszego zadania decyzyjnego, zamieniając ograniczenie równości na $x_1 + x_2 \leq 60$, co z kolei doprowadzamy do postaci równości poprzez wprowadzenie zmiennej nieistotnej u^2 :

$$x_1 + x_2 + u^2 - 60 = 0.$$

Procedura:

a) Tworzymy funkcję Lagrange'a dążącą do minimum:

$$L(x_1, x_2, \lambda, u) = 7x_1^2 - 5x_1x_2 + 6x_2^2 - 10x_1 - 6x_2 + 50 + \lambda(x_1 + x_2 + u^2 - 60).$$

Określamy pochodne cząstkowe dla funkcji L , otrzymując układ czterech równań o niewiadomych x_1, x_2, λ, u^2 .

$$(1) \frac{\partial L}{\partial x_1} = 14x_1 - 5x_2 - 10 + \lambda = 0, \quad (2) \frac{\partial L}{\partial x_2} = -5x_1 + 12x_2 - 6 + \lambda = 0,$$

$$(3) \frac{\partial L}{\partial \lambda} = x_1 + x_2 + u^2 - 60 = 0, \quad (4) \frac{\partial L}{\partial u} = 2\lambda u = 0.$$

Równanie $2\lambda u = 0$ jest spełnione, gdy: $\lambda = 0$ lub $u = 0$. Przyjmując $u = 0$, otrzymamy układ trzech równań:

$$(1) 14x_1 - 5x_2 + \lambda = 10,$$

$$(2) -5x_1 + 12x_2 + \lambda = 6,$$

$$(3) x_1 + x_2 = 60.$$

Skorzystajmy ze wzorów Cramera, obliczając wcześniej w Excelu wyznaczniki wyrażeniem [=WYZNACZNIK.MACIERZY(...)]:

$$|\mathbf{A}| = \begin{vmatrix} 14 & -5 & 1 \\ -5 & 12 & 1 \\ 1 & 1 & 0 \end{vmatrix} = -36, \quad |\mathbf{A}_1| = \begin{vmatrix} 10 & -5 & 1 \\ 6 & 12 & 1 \\ 60 & 1 & 0 \end{vmatrix} = -1024,$$

$$|\mathbf{A}_2| = \begin{vmatrix} 14 & 10 & 1 \\ -5 & 6 & 1 \\ 1 & 60 & 0 \end{vmatrix} = -1139, \quad |\mathbf{A}_3| = \begin{vmatrix} 14 & -5 & 10 \\ -5 & 12 & 6 \\ 1 & 1 & 60 \end{vmatrix} = 8296,$$

$$x_1 = \frac{|\mathbf{A}_1|}{|\mathbf{A}|} = \frac{-1024}{-36} = 28,44, \quad x_2 = \frac{|\mathbf{A}_2|}{|\mathbf{A}|} = \frac{-1136}{-36} = 31,56, \quad \lambda = \frac{|\mathbf{A}_3|}{|\mathbf{A}|} = \frac{8296}{-36} = -230,44,$$

$$f(x_1, x_2, \lambda) = 7x_1^2 - 5x_1x_2 + 6x_2^2 - 10x_1 - 6x_2 + 50 + \lambda(x_1 + x_2 - 60).$$

Po podstawieniu danych do równania otrzymujemy:

$$f(x_1, x_2, \lambda) = 6676,45 \text{ zł.}$$

Uzyskaliśmy takie same wartości zmiennych decyzyjnych oraz funkcji celu, jak w warunku równościowym.

Sprawdźmy teraz warunek $\lambda = 0$ przy podstawieniu $u^2 = u'$ w układzie równań:

$$(1) 14x_1 - 5x_2 + 0u' = 10,$$

$$(2) -5x_1 + 12x_2 + 0u' = 6,$$

$$(3) x_1 + x_2 + 1u' = 60.$$

Obliczenia przeprowadzone w Excelu na obszarach arkusza kalkulacyjnego podano w tab. 2.9.6:

T a b e l a 2.9.6

Macierze, wyznaczniki oraz wartości zmiennych

	Macierze	Wyznacznik	Wektor												
$\mathbf{A} =$	<table border="1"> <tr><td>14</td><td>-5</td><td>0</td></tr> <tr><td>-5</td><td>12</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </table>	14	-5	0	-5	12	0	1	1	1	$ \mathbf{A} = 143$	<table border="1"> <tr><td>10</td></tr> <tr><td>6</td></tr> <tr><td>60</td></tr> </table>	10	6	60
14	-5	0													
-5	12	0													
1	1	1													
10															
6															
60															
$\mathbf{A}_1 =$	<table border="1"> <tr><td>10</td><td>-5</td><td>0</td></tr> <tr><td>6</td><td>12</td><td>0</td></tr> <tr><td>60</td><td>1</td><td>1</td></tr> </table>	10	-5	0	6	12	0	60	1	1	$ \mathbf{A}_1 = 150$	Zmienne: $x_1 = \frac{ \mathbf{A}_1 }{ \mathbf{A} } = 1,05$			
10	-5	0													
6	12	0													
60	1	1													
$\mathbf{A}_2 =$	<table border="1"> <tr><td>14</td><td>10</td><td>0</td></tr> <tr><td>-5</td><td>6</td><td>0</td></tr> <tr><td>1</td><td>60</td><td>1</td></tr> </table>	14	10	0	-5	6	0	1	60	1	$ \mathbf{A}_2 = 134$	$x_2 = \frac{ \mathbf{A}_2 }{ \mathbf{A} } = 0,94$			
14	10	0													
-5	6	0													
1	60	1													
$\mathbf{A}_3 =$	<table border="1"> <tr><td>14</td><td>-5</td><td>10</td></tr> <tr><td>-5</td><td>12</td><td>6</td></tr> <tr><td>1</td><td>1</td><td>60</td></tr> </table>	14	-5	10	-5	12	6	1	1	60	$ \mathbf{A}_3 = 8296$	$u' = \frac{ \mathbf{A}_3 }{ \mathbf{A} } = 58,01$			
14	-5	10													
-5	12	6													
1	1	60													

Ź r ó d ł o: Opracowanie własne z użyciem funkcji WYZNACZNIK. MACIERZY(...) Excela.

Wektor to wyrazy wolne układu równań.

Gdy $\lambda = 0$, to otrzymaliśmy wyniki $x_1 = 1,05$ oraz $x_2 = 0,94$, jak do wcześniej obliczonego ekstremum bezwarunkowego nieuwzględniającego ograniczenia $x_1 + x_2 = 60$.

P r z y k ł a d

Nieliniowe programy decyzyjne o postaci standardowej, gdy wszystkie ograniczenia są nierównościami, możemy również rozwiązywać, stosując twierdzenia Kuhna-Tuckera¹⁷¹.

Mamy następujące zadanie decyzyjne:

$$f(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1 + 4x_2 + 8 \rightarrow \min, \text{ przy warunku: } g(x_1, x_2) = x_1 + 2x_2 \leq 4, \\ \text{oraz } x_1, x_2 \geq 0.$$

funkcja Lagrange'a przybiera postać:

$$L(x_1, x_2, \lambda) = 2x_1^2 + x_2^2 - 2x_1 - 4x_2 + 8 + \lambda(x_1 + 2x_2 - 4),$$

¹⁷¹ K. Kukuła (red.), *Badania operacyjne w przykładach i zadaniach...*, op.cit., rozdz. 6.1.2.

a sześć warunków Kuhna-Tuckera są następujące:

$$(1) \begin{cases} \frac{\partial L}{\partial x_1} = 4x_1 - 2 + \lambda - \nu_1 = 0, \\ \frac{\partial L}{\partial x_2} = 2x_2 - 4 + 2\lambda - \nu_2 = 0, \end{cases}$$

$$(2) \nu_1 x_1 + \nu_2 x_2 = 0, \quad (3) x_1 + 2x_2 - 4 + w = 0, \quad (4) \lambda w = 0, \quad (5) x_1, x_2 \geq 0, \quad (6) \lambda \geq 0.$$

Warunek drugi ma cztery kombinacje, np. $\nu_1 = 0$ oraz $\nu_2 = 0$.

Warunek czwarty: $w = 0$ lub $\lambda = 0$.

Możemy rozwiązywać układy równań w poszczególnych kombinacjach, otrzymując w rozwiązaniu x_1, x_2, λ , i obliczać funkcję $f(x_1, x_2, \lambda)$. Z rozwiązań wybieramy te, w których funkcja celu przyjmuje wartość ekstremalną.

Przyjmijmy na przykład kombinację: $\nu_1 = 0, \nu_2 = 0, w = 0$, przy której uzyskuje się rozwiązanie optymalne¹⁷², co daje z pochodnych (1) i ograniczenia (3) równania:

$$4x_1 + 0x_2 + 1\lambda = 2, \quad 0x_1 + 2x_2 + 2\lambda = 4, \quad 1x_1 + 2x_2 + 0\lambda = 4.$$

Rozwiążmy te równania jako układ równań, stosując dobrze nam już znane wzory Cramera i korzystając pomocniczo z funkcji Exela:

T a b e l a 2.9.7

Macierze i obliczone wyznaczniki

Macierz:	Wyznacznik:	Wektor			
$\mathbf{A} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 2 \\ 1 & 2 & 0 \end{bmatrix}$	$ \mathbf{A} = -18$	<table border="1"> <tr><td>2</td></tr> <tr><td>4</td></tr> <tr><td>4</td></tr> </table>	2	4	4
2					
4					
4					
$\mathbf{A}_1 = \begin{bmatrix} 2 & 0 & 1 \\ 4 & 2 & 2 \\ 4 & 2 & 0 \end{bmatrix}$	$ \mathbf{A}_1 = -8$				
$\mathbf{A}_2 = \begin{bmatrix} 4 & 2 & 1 \\ 0 & 4 & 2 \\ 1 & 4 & 0 \end{bmatrix}$	$ \mathbf{A}_2 = -32$				
$\mathbf{A}_3 = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 4 \\ 1 & 2 & 4 \end{bmatrix}$	$ \mathbf{A}_3 = -4$				

Ź r ó d ł o: Opracowanie własne funkcjami Excela.

¹⁷² Ibidem, s. 220.

$$x_1 = \frac{|A_1|}{|A|} = \frac{-8}{-18} = 0,44, \quad x_2 = \frac{|A_2|}{|A|} = \frac{-32}{-18} = 1,78, \quad \lambda = \frac{|A_3|}{|A|} = \frac{-4}{-18} = 0,22,$$

$$f(x_1, x_2, \lambda) = 2x_1^2 + x_2^2 - 2x_1 - 4x_2 + 8 + \lambda(x_1 + 2x_2 - 4),$$

$$f(x_1, x_2, \lambda) = 2 \cdot 0,44^2 + 1,78^2 - 2 \cdot 0,44 - 4 \cdot 1,78 + 8 + 0,22 \cdot (0,44 + 2 \cdot 1,78 - 4) = 3,56.$$

Jeżeli w zadaniu decyzyjnym nie wszystkie warunki ograniczające mają ten sam zwrot nierówności, to zmieniamy je, mnożąc przez $(-1)^{173}$. Gdy $f(x) \rightarrow \max$ oraz $g_i(x) \geq 0$, to w podanych na przykładzie warunkach (1) oraz (3) Kuhna-Tuckera zmieniamy kierunki nierówności, czyli dodajemy v_i oraz odejmujemy w .

2.9.3. Pytania

1. Wymień i objaśnij etapy metody mnożników Lagrange'a, gdy program nieliniowy ma postać kanoniczną.
2. Podaj procedurę korzystania z modułu (NLP) programu WinQSB.
3. Dane jest rozwiązanie nieliniowego całkowitoliczbowego zadania decyzyjnego w dodatku Solver programu Excel oraz raport wyników. Przeanalizuj informacje oraz sformułuj funkcję celu, ograniczenie i warunki brzegowe dla zmiennych.

Tabela 2.9.8

Zdefiniowanie nieliniowego zadania decyzyjnego

15870	=5*\$B\$2^2+10*\$B\$3^2-5*\$B\$2-20*\$B\$3+30							1
X1	32							2
X2	34							3
C1	200	=2*\$B\$2+4*\$B\$3						4
C2	PRAWDA	=\$B\$4>=150						5
Parametry w Solver:								6
Komórka celu:	\$A\$1						7	
Równa:	Min.						8	
Komórki zmieniane:	B2:B3						9	
Komórki ograniczające:	B4						10	
Dla komórek: B2, B3 po 3 ograniczenia brzegowe np. \$B\$2<=100; \$B\$2=całkowita (int); \$B\$2>=								11
A	B	C	D	E	F	G	H	13

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

¹⁷³ Ibidem, s. 218.

Tabela 2.9.9

Raport wynikowy

Komórka celu (Min)

Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$A\$1		30	15 870

Komórki decyzyjne

Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$B\$2	X1	0	32
\$B\$3	X2	0	34

Warunki ograniczające

Komórka	Nazwa	Wartość komórki	Formuła	Status	Luz
\$B\$4	C1	200	\$B\$4>=200	Wiążące	0
\$B\$3	X2	34	\$B\$3>=0	Niewiążące	34
\$B\$2	X1	32	\$B\$2=całkowita	Wiążące	0
\$B\$2	X1	32	\$B\$2<=100	Niewiążące	68
\$B\$2	X1	32	\$B\$2>=0	Niewiążące	32
\$B\$3	X2	34	\$B\$3=całkowita	Wiążące	0
\$B\$3	X2	34	\$B\$3<=100	Niewiążące	66

Microsoft Excel 8.0 Raport wyników

Arkusz: [S284.xls]S284-Problem (3)

Raport utworzony: 2007-09-23 12:07:23

Źródło: Opracowanie własne z zastosowaniem dodatku Solver programu Excel.

4. Napisz funkcję Lagrange'a dla podanego programu nieliniowego o funkcji celu dążącej do minimum i ograniczeniu, które jest nierównością, wprowadzając zmienną nieistotną.

$$FC: 5x_1^2 + 2x_1x_2 - 4x_2^2 - 8x_1 + 4x_2 + 10 \rightarrow \min,$$

$$C_1: x_1 + 4x_2 \leq 80,$$

$$x_1, x_2 \geq 0.$$

2.9.4. Zadania

1. Znajdź metodą mnożników Lagrange'a rozwiązanie optymalne zadania programowania nieliniowego całkowitoliczbowego, które określone jest następująco:

$$FC: 5x_1^2 + 10x_2^2 - 5x_1 - 20x_2 + 30 \rightarrow \max,$$

$$C_1: 2x_1 + 4x_2 = 200.$$

Warunki brzegowe:

$$x_1 \leq 100,$$

$$x_2 \leq 100,$$

$$x_1, x_2 \geq 0.$$

2. Rozwiąż zadanie pierwsze z zastosowaniem dodatku Solver programu Excel. Sprawdź swoje wyniki z raportem (tab. 2.9.10).

T a b e l a 2.9.10

Raport z realizacji zadania programu nieliniowego

Komórka celu (Maks)

Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$A\$1		30	24030

Komórki decyzyjne

Komórka	Nazwa	Wartość początkowa	Wartość końcowa
\$B\$2	X1	0	0
\$B\$3	X2	0	50

Warunki ograniczające

Komórka	Nazwa	Wartość komórki	Formuła	Status	Luz
\$B\$4	C 1	200	\$B\$4=200	Wiążące	0
\$B\$3	X2	50	\$B\$3>=0	Niewiążące	50
\$B\$2	X1	0	\$B\$2=całkowita	Wiążące	0
\$B\$2	X1	0	\$B\$2<=100	Niewiążące	100
\$B\$2	X1	0	\$B\$2>=0	Wiążące	0
\$B\$3	X2	50	\$B\$3=całkowita	Wiążące	0
\$B\$3	X2	50	\$B\$3<=100	Niewiążące	50

Microsoft Excel 8.0 Raport wyników

Arkusz: [S284.xls]S284-Proble m (4)

Raport utworzony: 2007-09-23 12:34:48

Ź r ó d ł o: Opracowanie własne dodatkiem Solver programu Excel.

3. Zastosuj funkcję Lagrange'a oraz warunki Kuhna-Tuckera w rozwiązaniu manualnym podanych zadań decyzyjnych. Porównaj wyniki z zamieszczonymi rozwiązaniami Solvera.

a) $f(x_1, x_2) = x_1^2 + 2x_2^2 - 8x_1 + 4x_2 + 2 \rightarrow \min,$ $g(x_1, x_2) = 2x_1 + x_2 \leq 8,$ $x_1, x_2 \geq 0.$

Tabela 2.9.11

Zdefiniowanie programu nieliniowego

	4	=B\$2^2+2*B\$3^2-8*B\$2+4*B\$3+20					1
X1		4					2
X2		0					3
C1		8	=2*B\$2+B\$3				4
C2		PRAWDA	=B\$4<=8				5
Parametry w Solver:							
Komórka celu:			\$A\$1				7
Równa:			Min				8
Komórki zmieniane:			B2:B3				9
Komórki ograniczające:			B4				10
Dla komórek: B2, B3 ograniczenia brzegowe, np. B\$2>=0							
A	B	C	D	E	F		13

Źródło: Opracowanie własne dodatkiem Solver programu Excel.

Rozwiązanie: $FC = 4$, $x_1^* = 4$, $x_2^* = 0$.

b) $f(x_1, x_2) = x_1^2 + 2x_2^2 - 8x_1 + 4x_2 + 20 \rightarrow \max$, $g(x_1, x_2) = 2x_1 + x_2 \leq 80$,
 $x_1, x_2 \geq 0$.

Tabela 2.9.12

Zdefiniowanie programu nieliniowego

	13140	=B\$2^2+2*B\$3^2-8*B\$2+4*B\$3+20					1
X1		0					2
X2		80					3
C1		80	=2*B\$2+B\$3				4
C2		PRAWDA	=B\$4<=80				5
Parametry w Solver:							
Komórka celu:			\$A\$1				7
Równa:			Max				8
Komórki zmieniane:			B2:B3				9
Komórki ograniczające:			B4				10
Dla komórek: B2, B3 ograniczenia brzegowe, np. B\$2>=0							
A	B	C	D	E	F		13

Źródło: Opracowanie własne, dodatek Solver programu Excel.

Rozwiązanie: $FC = 1340$, $x_1^* = 0$, $x_2^* = 80$.

2.10. Metody rozwiązywania zagadnień transportowych

2.10.1. Zamknięte zagadnienie transportowe (ZZT) – zastosowanie WinQSB

W danym zagadnieniu transportowym określamy:

- warunki dla dostawców,
- warunki dla odbiorców,
- warunki brzegowe,
- funkcję celu.

Algorytm transportowy jest procedurą iteracyjną, rozpoczyna się bowiem od początkowego rozwiązania dopuszczalnego, które następnie jest poprawiane.

Wyznaczenie początkowego rozwiązania dopuszczalnego realizujemy między innymi metodami:

- a) Metodą kąta północno-zachodniego – starszą i pracochłonną.

W tej metodzie realizacja początkowego rozwiązania dopuszczalnego jest następująca¹⁷⁴:

– rozpoczęcie wypełniania macierzy przewozów (x_{ij}) od klatki w lewym górnym rogu, wpisując do niej mniejszą ilość (A_i , B_j) odpowiadającą tej klatce,

– przesunięcie w prawo, gdy pierwszy towar nie został całkowicie rozdysponowany lub w dół, gdy całą podaż tego dostawcy rozdzielono odbiorcom a następnie:

obliczenie kosztu transportu,

poprawianie macierzy przewozów w kolejnych iteracjach z punktu widzenia celu, np. maksymalizacji lub minimalizacji kosztów.

- b) Metodą minimalnego elementu macierzy, np. kosztów:

doprowadzamy, aby w każdej kolumnie i w każdym wierszu było zero poprzez odejmowanie minimalnych wartości w poszczególnych wierszach a następnie kolumnach,

¹⁷⁴ Ibidem, rozdz. 2.1.1.

rozmieszczamy przewozy na trasach o koszcie najniższym, czyli zerowym, rozpoczynając od klatki z jednym zerem.

Gdy przewozy są w klatkach z zerami, mamy już rozwiązanie optymalne, a jeśli nie, stosujemy dalej algorytm transportowy.

P r z y k ł a d

Dokonyjemy rozmieszczenia przewozów, stosując metodę minimalnego elementu macierzy – kosztu [zł/t]¹⁷⁵.

$$FC: K(x_{ij}) = 40x_{11} + \dots + 30x_{34} \rightarrow \min.$$

T a b e l a 2.10.1

Koszty jednostkowe, potrzeby cukierni oraz zapasy hurtowni

Hurtownie	Cukiernie				A_i
	C_1	C_2	C_3	C_4	
H_1	40	30	40	10	80
H_2	20	50	50	20	60
H_3	20	30	50	30	100
B_j	50	50	70	70	240

Ź r ó d ł o: Opracowanie własne; dane umowne. Komórki inicjujące zaznaczono kursywą.

T a b e l a 2.10.2

Odjęcie od wartości w klatkach wierszy kosztu najniższego w danym wierszu

Hurtownie	Cukiernie				A_i
	C_1	C_2	C_3	C_4	
H_1	30	20	30	0	80
H_2	0	30	30	0	60
H_3	0	10	30	10	100
B_j	50	50	70	70	240

Ź r ó d ł o: Opracowanie własne; dane umowne. Komórki inicjujące zaznaczono kursywą.

¹⁷⁵ Ibidem, rozdz. 2.1.1.

Tabela 2.10.3

Odjęcie od wartości w klatkach kolumn kosztu najniższego w danej kolumnie

Hurtownie	Cukiernie				A_i
	C_1	C_2	C_3	C_4	
H_1	30	10	0	0	80
H_2	0	20	0	0	60
H_3	0	0	0	10	100
B_j	50	50	70	70	240

Źródło: Opracowanie własne.

Tabela 2.10.4

Rozmieszczenie przewozów, rozpoczynając od wierszy/kolumn z jednym zerem

Hurtownie	Cukiernie				A_i
	C_1	C_2	C_3	C_4	
H_1			20	60	80
H_2	25		25	10	60
H_3	25	50	25		100
B_j	50	50	70	70	240

Źródło: Opracowanie własne.

Rozwiązanie jest już optymalne, ponieważ wszystkie przewozy są w klatkach z zerami.

Tabela 2.10.5

Elementy obliczenia kosztu transportu

		$40 \cdot 20$	$10 \cdot 60$
$20 \cdot 25$		$50 \cdot 25$	$20 \cdot 10$
$20 \cdot 25$	$30 \cdot 50$	$50 \cdot 25$	

Źródło: Opracowanie własne.

Koszty transportu: $K_{ij} = \sum$ kosztów przewozów.

$$K_{ij} = 40 \cdot 20 + 10 \cdot 60 + 20 \cdot 25 + 50 \cdot 25 + 20 \cdot 10 + 20 \cdot 25 + 30 \cdot 50 + 50 \cdot 25 = 6600 \text{ zł.}$$

P r z y k ł a d

Zastosowanie programu WinQSB do rozwiązania zagadnienia transportowego.

W programie tym zmienne zapisywane są jako np. X11, a typ zmiennej (Var. T) może być następujący: C – continuous (rzeczywiste), I – integer (całkowitoliczbowe), B – binarne.

Tabela 2.10.6

Koszty jednostkowe, potrzeby sklepów oraz możliwości składnic

(i)	Sklep 1	Sklep 2	Sklep 3	Sklep 4	Sklep 5	Podaż a_i
Składnica 1	4	5	4	8	3	120
Składnica 2	2	6	4	7	4	100
Składnica 3	6	3	3	5	4	90
Popyt b_j	40	60	50	80	80	310

Źródło: Opracowanie własne.

Funkcja celu: $FC: 4X_{11} + \dots + 4X_{35} \rightarrow \min.$

Warunki – podaż składnicy:

$$C_1: X_{11} + X_{12} + X_{13} + X_{14} + X_{15} = 120,$$

$$C_2: X_{21} + X_{22} + X_{23} + X_{24} + X_{25} = 100,$$

$$C_3: X_{31} + X_{32} + X_{33} + X_{34} + X_{35} = 90.$$

Warunki – popyt sklepów:

$$C_4: X_{11} + X_{21} + X_{31} = 40,$$

$$C_5: X_{12} + X_{22} + X_{32} = 60,$$

$$C_6: X_{13} + X_{23} + X_{33} = 50,$$

$$C_7: X_{14} + X_{24} + X_{34} = 80,$$

$$C_8: X_{15} + X_{25} + X_{35} = 80,$$

$$X_{ij} \geq 0 \quad (i = 1, 2, 3) \quad (j = 1, 2, 3, 4, 5) \quad \Sigma b_j = 310 \quad \Sigma a_i = 310.$$

Tabela 2.10.7

Zapis zagadnienia transportowego w programie WinQSB

Var.	X11	X12	X13	X14	X15	X21	X22	X23	X24	X25	X31	X32	X33	X34	X35	D	RHS
Min.	4	5	4	8	3	2	6	4	7	4	6	3	3	5	4		
C1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	=	120
C2	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	=	100
C3	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	=	90
C4	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	=	40
C5	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	=	60
C6	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	=	50
C7	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	=	80
C8	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	=	80
Low.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Upp.	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M		
Var. T.	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I		

D – znak ograniczenia; Low. (LowerBound) – poziom dolny; Upp. (UpperBound) – poziom górny.

Źródło: Opracowanie własne, wstępnie formatowanie danych w arkuszu kalkulacyjnym programu Excel, a następnie import do programu WinQSB.

Tabela 2.10.8

Rozwiązanie sumaryczne zagadnienia transportowego w programie WinQSB

12-05-2006 11:21:11	Decision Variable	Solution Value	Unit Cost or Profit C_{ij}	Total Contribution	Reduced Cost	Basis Status
1	X11	0	4	0	2	At bound
2	X12	40	5	200	0	Basis
3	X13	0	4	0	0	At bound
4	X14	0	8	0	1	At bound
5	X15	80	3	240	0	Basis
6	X21	40	2	80	0	Basis
7	X22	0	6	0	1	At bound
8	X23	50	4	200	0	Basis
9	X24	10	7	70	0	Basis
10	X25	0	4	0	1	At bound
11	X31	0	6	0	6	At bound
12	X32	20	3	60	0	Basis
13	X33	0	3	0	1	At bound
14	X34	70	5	350	0	Basis
15	X35	0	4	0	3	At bound

gdzie:

Decision Variable – zmienne decyzyjne,

Solution Value – wartość rozwiązania,

Unit Cost or Profit C_{ij} – koszt jednostkowy,

Total Contribution – iloczyn wartości i kosztu jednostkowego,

(Suma wartości w kol. Total Contribution stanowi funkcję celu $FC = 1200$ zł)

Reduced Cost oraz Basis Status – pojęcia z danej iteracji algorytmu transportowego.

Źródło: Opracowanie własne w programie WinQSB.

Przykład

Zastosowanie metody kar dla tras niedopuszczalnych (nawiązanie do danych w tabeli 2.10.6)

Trasy niedopuszczalne – $M = 50$

$$FC: 4X_{11} + 5X_{12} + 4X_{13} + 8X_{14} + 3X_{15} + 2X_{21} + 6X_{22} + 4X_{23} + 7X_{24} + 4X_{25} + MX_{31} + MX_{32} + 3X_{33} + 5X_{34} + 4X_{35} \rightarrow \min$$
Warunki brzegowe: $X_{ij} \geq 0$ (LowerBound) – minimalny popyt, $X_{ij} \leq 80$ – UpperBound – maksymalny popyt.

Tabela 2.10.9

Rozwiązanie sumaryczne dla $M = 50$

Summary	Decision Variable	Solution Value	Unit Cost or Profit $C(j)$	Total Contribution	Reduced Cost	Basis Status
1	X11	0	4	0	3	At bound
2	X12	40	5	200	0	Basis
3	X13	0	4	0	1	At bound
4	X14	0	8	0	3	At bound
5	X15	80	3	240	0	Basis
6	X21	40	2	80	0	Basis
7	X22	20	6	120	0	Basis
8	X23	40	4	160	0	Basis
9	X24	0	7	0	1	At bound
10	X25	0	4	0	0	At bound
11	X31	0	50	0	49	At bound
12	X32	0	50	0	45	At bound
13	X33	10	3	30	0	Basis
14	X34	80	5	400	0	Basis
15	X35	0	4	0	1	At bound
FC: Objective Function (Min)				1230 zł		

Źródło: Opracowanie własne w programie WinQSB.

2.10.2. Przykład zagadnienia transportowego rozwiązane w dodatku Solver

Tabela 2.10.10

Sformułowanie oraz rozwiązanie zagadnienia transportowego

A	B	C	D	E	F	G	H
Zakłady	Razem	Katowice	Bydgoszcz	Wrocław	Gdańsk	Warszawa	7
Z1	300	0	0	0	80	220	8
Z2	260	0	0	180	80	0	9
Z3	280	180	80	20	0	0	10
A	B	C	D	E	F	G	11
Razem		180	80	200	160	220	12
							13
	Popyt	180	80	200	160	220	14
Zakłady	Podaż	Koszty jednostkowe przewozu (c_{ij})					15
Z1	310	10	8	6	5	4	16
Z2	260	6	5	4	3	6	17
Z3	280	3	4	5	5	9	18
Σ	83	19	17	15	13	19	19
Koszt:	3200	540 (180 · 3)	320	820	640	880	20
Komórka docelowa B20			Minimalizacja kosztów wysyłki				
Komórki zmieniane C8:G10			Liczba wyrobów wysyłanych z zakładów do hurtowni w miastach				
Ograniczenia		B8:B10<=B16:B18		Liczba wysłanych <= produkcji podaży			

Źródło: Opracowanie według przykładu w dodatku Solver programu Excel.

Ponadto program Excel-Solver umożliwia uzyskanie raportów:

- tabela przewozów,
- warunki ograniczające,
- raport wrażliwości,
- raport granic.

Raport „Tabela przewozów” zawiera tab. 2.10.11.

Tabela 2.10.11

Tabela przewozów

Komórka celu B\$20	Nazwa Koszt przewozu: podaż	Wartość początkowa 83	Wartość końcowa funkcji celu 3 200
Komórki decyzyjne	Nazwa	Wartość początkowa (zainicjowana)	Wartość końcowa
\$C\$8	Z1 Katowice	1	0
\$D\$8	Z1 Bydgoszcz	1	0
\$E\$8	Z1 Wrocław	1	0
\$F\$8	Z1 Gdańsk	1	80
\$G\$8	Z1 Warszawa	1	220
\$C\$9	Z2 Katowice	1	0
\$D\$9	Z2 Bydgoszcz	1	0
\$E\$9	Z2 Wrocław	1	180
\$F\$9	Z2 Gdańsk	1	80
\$G\$9	Z2 Warszawa	1	0
\$C\$10	Z3 Katowice	1	180
\$D\$10	Z3 Bydgoszcz	1	80
\$E\$10	Z3 Wrocław	1	20
\$F\$10	Z3 Gdańsk	1	0
\$G\$10	Z3 Warszawa	1	0

Wartość początkowa stanowi sumę kosztów dla ilości zainicjowanych.

Źródło: Opracowanie według przykładu w dodatku Solver programu Excel.

2.10.3. Otwarte zagadnienie transportowe – OZT

Zagadnienie transportowe otwarte sprowadzamy do zamkniętego poprzez wprowadzenie fikcyjnego odbiorcy, którego zapotrzebowanie jest równe nadwyżce podaży nad popytem. W praktyce oznacza to najczęściej, że nadwyżka materiałów pozostanie w magazynach dostawców, co wymaga poniesienia dodatkowego kosztu magazynowania materiału. W funkcji celu minimalizuje się łączne koszty transportu i magazynowania, a więc rozpatrujemy zagadnienie transportowo-magazynowe.

Przykład

Opracowanie manualne otwartego zagadnienia transportowego (OZT)¹⁷⁶ dotyczącego przewozu cukru z hurtowni do cukierni.

Tabela 2.10.12

Podaż, popyt oraz koszty jednostkowe

Hurtownie	Jednostkowe koszty magazynowania c_{ij}	Cukiernie				A_i
		C_1	C_2	C_3	C_4	
H_1	2	40	30	40	10	150(120)
H_2	5	20	50	50	20	60
H_3	4	20	30	50	30	60
B_j		50	50	70	70	270(240)

Źródło: Opracowanie własne.

Warunki ograniczające dostawców:

$$\sum x_{1j} \leq 150, \quad \sum x_{2j} \leq 60, \quad \sum x_{3j} \leq 60, \quad j = 1, 2, 3, 4.$$

Tabela 2.10.13

Wprowadzenie fikcyjnego odbiorcy różnicy towaru (270 – 240 = 30) w celu zamiany zagadnienia OZT na ZZT

Hurtownie	Cukiernie					A_i
	C_1	C_2	C_3	C_4	F (x_{i5})	
H_1	40	30	40	10	2	150
H_2	20	50	50	20	5	60
H_3	20	30	50	30	4	60
B_j	50	50	70	70	30	270

Źródło: Opracowanie własne.

Warunki ograniczające odbiorców:

$$\sum x_{i1} = 50, \quad \sum x_{i2} = 50, \quad \sum x_{i3} = 70, \quad \sum x_{i4} = 70, \quad i = 1, 2, 3.$$

Dodatkowo warunek dla odbiorcy fikcyjnego:

$$x_{15} + x_{25} + x_{35} = 30,$$

$$K(x_{ij}) = 40x_{11} + \dots + 4x_{35} \rightarrow \min,$$

x_{i5} – ilość cukru, jaka pozostanie w magazynach dostawców.

¹⁷⁶ Ibidem, rozdz. 2.1.2.

Tabela 2.10.14

Przekształcona macierz kosztów – po odjęciu najmniejszego kosztu
w wierszach i kolumnach

Hurtownie	Cukiernie					A_i
	C_1	C_2	C_3	C_4	F	
H_1	23	2	0	0	0	150
H_2	0	19	7	7	0	60
H_3	1	0	8	18	0	60
B_j	50	50	70	70	30	270

Źródło: Opracowanie własne.

Tabela 2.10.15

Początkowe rozwiązanie dopuszczalne
(wprowadzenie przewozów w klatkach z zerami)

Hurtownie	Cukiernie					A_i
	C_1	C_2	C_3	C_4	$F(x_{is})$	
H_1			70	70	10	150
H_2	50				10	60
H_3		50			10	60
B_j	50	50	70	70	30	270

Źródło: Opracowanie własne.

Rozwiązanie jest optymalne: wszystkie przewozy w klatkach z zerami.

$$K = 40 \cdot 70 + 10 \cdot 70 + 2 \cdot 10 + 20 \cdot 50 + 5 \cdot 10 + 30 \cdot 50 + 4 \cdot 10 = 6110 \text{ zł}$$

2.10.4. Zamknięte zagadnienie transportowe – o niebilansujących się ilościach w klatkach zerowych

Przykład

Znajdujemy rozwiązanie optymalne¹⁷⁷:

Tabela 2.10.16

Koszty jednostkowe, zapotrzebowania wydziałów oraz zapasy magazynowe

Magazyn	Wydział			A_i
	W_1	W_2	W_3	
M_1	3	4	6	30
M_2	2	5	5	40
M_3	4	1	8	25
B_j	30	30	35	95

Źródło: Opracowanie własne; dane umowne.

¹⁷⁷ W. Sikora (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii...*, op.cit., rozdz. 2.1.

Tabela 2.10.17

Przekształcenie macierzy*

Magazyn	Wydział			A_i
	W_1	W_2	W_3	
M_1	0	1	0	30
M_2	0	3	0	40
M_3	3	0	4	25
B_j	30	30	35	95

* Odjęcie od wierszy a następnie od kolumn najmniejszego kosztu przewozu w danym wierszu/kolumnie.

Źródło: Opracowanie własne.

Tabela 2.10.18

Początkowe rozwiązanie dopuszczalne*

Magazyn	Wydział			A_i
	W_1	W_2	W_3	
M_1	10	5	15	30
M_2	20		20	40
M_3		25		25
B_j	30	30	35	95

* Ulokowanie przewozów w klatkach z zerami; rozpoczynamy od kolumny/wiersza z jednym zerem.

Źródło: Opracowanie własne.

Nie wszystkie ilości są w klatkach zerowych, czyli nie mamy pewności, że dana wersja rozwiązania jest optymalna. Skorzystajmy więc z algorytmu simpleks modułu (LP) programu WinQSB.

Tabela 2.10.19

Rozwiązanie sumaryczne zamkniętego zagadnienia transportowego

12-13-2007 14:27:10	Decision Variable	Solution Value	Unit Cost or Profit $C(j)$	Total Contribution	Reduced Cost	Basis Status
1	X11	(0)	25	3	75	0 Basis
2	X12	(5)	5	4	20	0 Basis
3	X13	(26)	0	6	0	0 At bound
4	X21	(30)	5	2	10	0 Basis
5	X22	(0)	0	5	0	2 At bound
6	X23	(10)	35	5	175	0 Basis
7	X31	(0)	0	4	0	4 At bound
8	X32	(25)	25	1	25	0 Basis
9	X33	(0)	0	8	0	5 At bound
	Objective	Function	(Min.) =		305	

Uwaga: występuje rozwiązanie alternatywne.

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

W tabeli 2.10.19 podano rozwiązanie komputerowe przy założeniu równości zaspokojenia popytu trzech wydziałów oraz równości podaży dostaw z trzech magazynów. Program sygnalizuje istnienie rozwiązania alternatywnego (wyniki w nawiasach) o funkcji celu wynoszącej 305. Rozwiążmy zagadnienie transportowe manualnie, korzystając z metody potencjałów.

Metoda potencjałów (opieramy się na danych z tab. 2.10.16)¹⁷⁸.

Tabela 2.10.20

*Rozwiązanie metodą minimalnego elementu macierzy**

Magazyn	Wydział			A_i
	W_1	W_2	W_3	
M_1	3(25)	4(5)	6(X)	X
M_2	2(5)	5(X)	5(35)	X
M_3	4(X)	1(25)	8(X)	X
B_j	X	X	X	

*Rozpoczynamy od minimalnego kosztu jednostkowego w macierzy kosztów, czyli od przewozu od M_3 do W_2 .

Źródło: Opracowanie własne.

$$K_1 = 3 \cdot 25 + 4 \cdot 5 + 2 \cdot 5 + 5 \cdot 35 + 1 \cdot 25 = 305.$$

Sprawdzenie: czy otrzymane rozwiązanie jest optymalne.

Tabela 2.10.21

Zmienne dualne dla tras bazowych $\alpha_i + \beta_j = c_{ij}$

Magazyn	Wydział			α_i
	W_1	W_2	W_3	
M_1	3	4	6	0
M_2	2	5	5	-1
M_3	4	1	8	-3
β_j	3	4	6	

Źródło: Opracowanie własne w Excelu.

Kryteria optymalności:

zero (X) dla tras bazowych; dla pozostałych $\Delta_{ij} = c_{ij} - \alpha_i - \beta_j$ szukamy kryteriów ujemnych dla komórek niebazowych

X	X	0
X	2	X
4	X	3

¹⁷⁸ Ibidem.

W naszym przypadku nie występują kryteria ujemne, zatem mamy już rozwiązanie optymalne. Gdyby wystąpiły, to przesuamy w pętli wartości wcześniejszego rozwiązania bazowego w kierunku kryterium ujemnego, tak aby suma w wierszach i kolumnach nie uległa zmianie.

Przykład
Występowanie kryterium ujemnego

Tabela 2.10.22

Koszty jednostkowe, potrzeby wydziałów oraz zapasy magazynowe

Magazyn	Wydział			A_i
	W_1	W_2	W_3	
M_1	7	2	6	40
M_2	5	4	1	50
M_3	3	4	5	30
B_j	20	30	70	

Źródło: Opracowanie własne; dane umowne.

Przekształcenie macierzy (według metody minimalnego elementu macierzy w wierszach oraz kolumnach)

5	0	4
4	3	0
0	1	2

Tabela 2.10.23

Wstępne rozmieszczenie przewozów w klatkach zerowych

Magazyn	W_1	W_2	W_3	A_i
M_1	5 10	0 30	4	40
M_2	4	3	0 50	50
M_3	0 10	1	2 20	30
B_j	20 20	30 30	70 70	

Źródło: Opracowanie własne; dane umowne.

Obliczamy koszt według kosztów jednostkowych macierzy przekształconej:

$$K_1 = 5 \cdot 10 + 0 \cdot 30 + 0 \cdot 50 + 0 \cdot 10 + 2 \cdot 20 = 90.$$

Sprawdzenie optymalności:

Tabela 2.10.24

Zmienne dualne dla klatek bazowych według: $c_{ij} = \alpha_i + \beta_j$

						α_i	
	5	10	0	30	4	X	0
	4	X	3	X	0	50	-7
	0	10	1	X	2	20	-5
β_j	5		0		7		

Źródło: Opracowanie własne w Excelu.

Tabela 2.10.25

Kryteria dla klatek niebazowych

							α_i
	5	X	0	X	4	-3	0
	4	6	3	10	0	X	-7
	0	X	1	6	2	X	-5
β_j	5		0		7		

Źródło: Opracowanie własne w Excelu.

Obliczenie kryterium według wzoru: $\Delta_{ij} = c_{ij} - \alpha_i - \beta_j$, np. trasy $M_1 - W_3$: $\Delta_{ij} = 4 - 0 - 7 = -3$.

Pojawiła się klatka z kryterium ujemnym. Przesuwamy wartość 10 z klatki zmiennej X_{11} (na przecięciu M_1 i W_1) w kierunku kryterium (-3) i następnie odejmujemy 10 od wartości 20 w klatce zmiennej X_{33} , dodając 10 w klatce zmiennej X_{31} .

Tabela 2.10.26

Przesunięte wartości przewozów

Magazyn	W_1	W_2	W_3	A_i			
M_1	5	X	0	30	4	10	40
M_2	4	X	3	X	0	50	50
M_3	0	20	1	X	2	10	20
B_j	20		20		70		

Źródło: Opracowanie własne w Excelu.

Ponownie sprawdzenie optymalności:

Tabela 2.10.27

Zmienne dualne dla klatek bazowych według: $c_{ij} = \alpha_i + \beta_j$

					α_i			
	5	X	0	30	4	10	0	
	4	X	3	X	0	50	-4	
	0		20	1	X	2	10	-2
β_j		2		0		4		

Źródło: Opracowanie własne w Excelu.

Tabela 2.10.28

Kryteria dla klatek niebazowych według: $\Delta_{ij} = c_{ij} - \alpha_i - \beta_j$

					α_i		
	5	3	0	X	4	X	0
	4	6	3	7	0	X	-4
	0	X	1	3	2	X	-2
β_j		2		0		4	

Źródło: Opracowanie własne w Excelu.

Kryteria są dodatnie, tak więc rozwiązanie jest optymalne. Koszt według macierzy przekształconej:

$$K_2 = 0 \cdot 30 + 4 \cdot 10 + 0 \cdot 50 + 0 \cdot 20 + 2 \cdot 10 = 60.$$

Koszt według macierzy bazowej (danych wejściowych) – tabela 2.10.22:

$$K = 2 \cdot 30 + 6 \cdot 10 + 1 \cdot 50 + 3 \cdot 20 + 5 \cdot 10 = 280.$$

Program WinQSB rozwiązuje powyższe zagadnienie transportowe w sześciu iteracjach. Kolejne rozwiązania dla zmiennych oznaczonych jako X_{11} , ..., X_{33} podano w tab. 2.10.29.

Tabela 2.10.29

Zestawienie rozwiązań w kolejnych iteracjach

Iteracja	X_{11}	X_{12}	X_{13}	X_{21}	X_{22}	X_{23}	X_{31}	X_{32}	X_{33}
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	50	0	0	0
3	0	30	0	0	0	50	0	0	0
4	0	30	0	0	0	50	20	0	0
5	0	30	0	0	0	50	20	0	10
6	0	30	10	0	0	50	20	0	10

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Otrzymane komputerowo rozwiązanie w iteracji szóstej jest identyczne z uzyskanym manualnie (tab. 2.10.26).

P r z y k ł a d

Określmy rozwiązanie optymalne metodą potencjałów. Porównajmy wyniki z otrzymanymi z WinQSB

T a b e l a 2.10.30

Koszty jednostkowe transportu, zapotrzebowania montowni oraz możliwości magazynów

Magazyn	Montownie				A_i
	M_1	M_2	M_3	M_4	
M_1	7	15	16	0	100
M_2	12	11	8	4	200
M_3	5	13	9	10	100
B_j	120	70	80	130	

Ź r ó d ł o: Opracowanie własne; dane umowne.

Rozwiązanie manualne – przekształcenie macierzy:

według wierszy:

7	15	16	0
8	7	4	0
0	8	4	5

według kolumn:

7	8	12	0
8	0	0	0
0	1	0	5

T a b e l a 2.10.31

Rozmieszczenie przewozów

20	X	X	80	100
7	8	12	0	
X	70	80	50	200
8	0	0	0	
100	X	X	X	100
0	1	0	5	
120	70	80	130	

Ź r ó d ł o: Opracowanie własne; dane umowne.

Tabela 2.10.32

Sprawdzenie optymalności*

					α_i
	20	X	X	80	0
7		8	12	0	
	X	70	80	50	0
8		0	0	0	
	100	X	X	X	-7
0		1	0	5	
β_i	7	0	0	0	

* Zmienne dualne dla klatek bazowych $c_{ij} = \alpha_i + \beta_j$.

Źródło: Opracowanie własne w Excelu.

Tabela 2.10.33

Kryteria dla klatek niebazowych według: $\Delta_{ij} = c_{ij} - \alpha_i - \beta_j$

		1	2	3	4	α_i
1	X	X	8	12	X	0
	7		8	12	0	
2	11	X	X	X	X	0
	8		0	0	0	
3	X	X	8	7	12	-7
	0		1	0	5	
β_i	7	0	0	0	0	

Źródło: Opracowanie własne w Excelu, przykładowo: $\Delta_{21} = 8 - (0 + 7)$.

Kryteria są dodatnie, zatem rozwiązanie jest optymalne. Koszt według klatek bazowych (zob. tab. 2.10.30) i rozmieszczenia przewozów (tab. 2.10.31):

$$K_1 = 7 \cdot 20 + 0 \cdot 80 + 11 \cdot 70 + 8 \cdot 80 + 4 \cdot 50 + 5 \cdot 100 = 2250 \text{ zł.}$$

Tabela 2.10.34

Komputerowe rozwiązanie sumaryczne

2006-12-05	Variable	Solution	Unit Cost $C(j)$	Total
1	X_{11}	20	7	140
2	X_{12}	0	15	0
3	X_{13}	0	16	0
4	X_{14}	80	0	0
5	X_{21}	0	12	0
6	X_{22}	70	11	770
7	X_{23}	80	8	640
8	X_{24}	50	4	200
9	X_{31}	100	5	500
10	X_{32}	0	13	0
11	X_{33}	0	9	0
12	X_{34}	0	10	0

Źródło: Opracowanie własne, moduł (LP), program WinQSB.

Widzimy, że wyniki rozwiązania komputerowego pokrywają się z rozwiązaniem manualnym.

Przykład

Rozwiąż zagadnienie transportowe dostawy paszy przez hurtownie do składnic rolniczych.

Tabela 2.10.35

Koszty jednostkowe, zapasy hurtowni oraz zapotrzebowanie składnic rolniczych

Hurtownia	Składnica rolnicza			A_i	Koszt magazynowania
	S_1	S_2	S_3		
H_1	5	4	5	700	1
H_2	4	8	7	500	5
H_3	6	4	7	800	6
B_j	400	600	500	2 000(1 500)	

Źródło: Opracowanie własne; dane umowne.

$$FC: 5x_{11} + 4x_{12} + 5x_{13} + 4x_{21} + 8x_{22} + 7x_{23} + 6x_{31} + 4x_{32} + 7x_{33} \rightarrow \min.$$

Warunki ograniczające dla dostawców:

$$\begin{aligned} \sum x_{1j} &\leq 700, & \sum x_{2j} &\leq 500, & \sum x_{3j} &\leq 800, \\ x_{11} + x_{12} + x_{13} &\leq 700, & x_{21} + x_{22} + x_{23} &\leq 500, & x_{31} + x_{32} + x_{33} &\leq 800. \end{aligned}$$

Warunki dla odbiorców:

$$\begin{aligned} x_{11} + x_{21} + x_{31} &= 400, \\ x_{12} + x_{22} + x_{32} &= 600, \\ x_{13} + x_{23} + x_{33} &= 500. \end{aligned}$$

T a b e l a 2.10.36

Wprowadzenie fikcyjnego odbiorcy

Hurtownia	Składnica rolnicza				A_i
	S_1	S_2	S_3	$F(x_{i4})$	
H_1	5	4	5	1	700
H_2	4	8	7	5	500
H_3	6	4	7	6	800
B_j	400	600	500	500	

Ź r ó d ł o: Opracowanie własne; dane umowne.

Dodatkowo warunek dla odbiorcy fikcyjnego:

$$x_{14} + x_{24} + x_{34} = 500.$$

$$\begin{aligned} K_{(x_{ij})} &= 5x_{11} + 4x_{12} + 5x_{13} + x_{14} + 4x_{21} + 8x_{22} + 7x_{23} + 5x_{24} + 6x_{31} + 4x_{32} + 7x_{33} + \\ &+ 6x_{34} \rightarrow \min, \\ x_{i4} &- \text{ilość paszy jaka pozostanie w hurtowniach.} \end{aligned}$$

T a b e l a 2.10.37

Przekształcona macierz kosztów (odjęcie minimum od wierszy)

Hurtownia	Składnica rolnicza				A_i
	S_1	S_2	S_3	$F(x_{i4})$	
H_1	4	3	4	0	700
H_2	0	4	3	1	500
H_3	2	0	3	2	800
B_j	400	600	500	500	2 000

Ź r ó d ł o: Opracowanie własne w Excelu.

T a b e l a 2.10.38

Przekształcona macierz kosztów (odjęcie minimum od kolumn)

Hurtownia	Składnica rolnicza				A_i
	S_1	S_2	S_3	$F(x_{i4})$	
H_1	4	3	1	0	700
H_2	0	4	0	1	500
H_3	2	0	0	2	800
B_j	400	600	500	500	2 000

Ź r ó ł o: Opracowanie własne w Excelu.

T a b e l a 2.10.39

*Początkowe rozwiązanie dopuszczalne
(wprowadzenie przewozów w klatkach zerowych)*

Hurtownia	Składnica rolnicza				A_i
	S_1	S_2	S_3	$F(x_{i4})$	
H_1			200	500	
H_2	400		100		
H_3		600	200		
B_j					

Ź r ó ł o: Opracowanie własne; dane umowne.

Wstępnie rozwiązanie nie jest optymalne, bo przewóz $H_1 - S_3$ nie jest w klatce zerowej.

$$K_1 = 5 \cdot 200 + 1 \cdot 500 + 4 \cdot 400 + 7 \cdot 100 + 4 \cdot 600 + 7 \cdot 200 = 7600.$$

Sprawdzenie analityczne: czy rozwiązanie jest jednak optymalne.

T a b e l a 2.10.40

Zmienne dualne $\alpha_i + \beta_j = c_{ij}$

Hurtownia	Składnica rolnicza				α_i
	S_1	S_2	S_3	$F(x_{i4})$	
H_1	5 (x)	4 (x)	5 (200)	1 (500)	0
H_2	4 (400)	8 (x)	7 (100)	5 (x)	2
H_3	6 (x)	4 (600)	7 (200)	6 (x)	2
β_j	2	2	5	1	

Ź r ó ł o: Opracowanie własne; dane umowne.

Kryteria optymalności:

zero dla tras bazowych x

dla niebazowych obliczamy $\Delta_{ij} = c_{ij} - \alpha_i - \beta_j$ i szukamy kryteriów ujemnych.

3	2	x	x
x	4	x	2
2	x	x	3

Rozwiązanie jest jednak optymalne, gdyż nie ma kryteriów ujemnych.

P r z y k ł a d

Określmy minimalny koszt transportu towarów z hurtowni do składnic rolniczych, gdy popyt równa się podaży.

T a b e l a 2.10.41

*Koszty jednostkowe, zapasy hurtowni i potrzeby składnic rolniczych
gdy $B_j = A_i$*

Hurtownia	Składnica rolnicza			A_i
	S_1	S_2	S_3	
H_1	5	4	5	200
H_2	4	8	7	500
H_3	6	4	7	800
B_j	400	600	500	1 500

Ź r ó d ł o: Opracowanie własne; dane umowne.

$$FC: 5x_{11} + 4x_{12} + 5x_{13} + 4x_{21} + 8x_{22} + 7x_{23} + 6x_{31} + 4x_{32} + 7x_{33} \rightarrow \min.$$

Warunki ograniczające dostawców:

$$x_{11} + x_{12} + x_{13} = 200,$$

$$x_{21} + x_{22} + x_{23} = 500,$$

$$x_{31} + x_{32} + x_{33} = 800.$$

Warunki dla odbiorców:

$$x_{11} + x_{21} + x_{31} = 400,$$

$$x_{12} + x_{22} + x_{32} = 600,$$

$$x_{13} + x_{23} + x_{33} = 500.$$

Przekształcona macierz kosztów:

T a b e l a 2.10.42

Odjęcie minimalnej wartości w wierszach

Hurtownia	Składnica rolnicza			A_i
	S_1	S_2	S_3	
H_1	1	0	1	200
H_2	0	4	3	500
H_3	2	0	3	800
B_j	400	600	500	1 500

Ź r ó d ł o: Opracowanie własne w Excelu.

T a b e l a 2.10.43

Odjęcie minimalnej wartości w kolumnach

Hurtownia	Składnica rolnicza			A_i
	S_1	S_2	S_3	
H_1	1	0	0	200
H_2	0	4	2	500
H_3	2	0	2	800
B_j	400	600	500	1500

Ź r ó d ł o: Opracowanie własne w Excelu.

T a b e l a 2.10.44

Początkowe rozwiązanie dopuszczalne

Hurtownia	Składnica rolnicza			A_i
	S_1	S_2	S_3	
H_1			200	
H_2	400	100		
H_3		500	300	
B_j				

Ź r ó d ł o: Opracowanie własne; dane umowne.

Wstępne rozwiązanie nie jest optymalne, ponieważ nie wszystkie przewozy są w klatkach zerowych.

$$K_1 = 5 \cdot 200 + 4 \cdot 400 + 8 \cdot 100 + 4 \cdot 500 + 7 \cdot 300 = 7500 \text{ zł.}$$

Sprawdzenie analityczne, czy rozwiązanie jest optymalne według komórek bazowych (z podaną ilością przewozów).

Tabela 2.10.45

Zmienne dualne $\alpha_i + \beta_j = c_{ij}$

Hurtownia	Składnica rolnicza			α_i
	S_1	S_2	S_3	
H_1	5 (x)	4 (x)	5 (200)	0
H_2	4 (400)	8 (100)	7 (x)	6
H_3	6 (x)	4 (500)	7 (300)	2
β_j	-2	2	5	

Źródło: Opracowanie własne w Excelu.

Kryteria optymalności:

zero dla tras bazowych, które oznaczamy jako x ,dla pozostałych $\Delta_{ij} = c_{ij} - \alpha_i - \beta_j$. Szukamy kryteriów ujemnych dla komórek niebazowych.

7	2	x
x	x	-4
	(100)	
8	x	x

Tabela 2.10.46

Poprawa rozwiązania bazowego według wskazania strzałką kryterium ujemnego (-4), przesuwanie ilość „100” do (H_2-S_3) i dalej korygując ilości w pętli

Hurtownia	Składnica rolnicza			α_i
	S_1	S_2	S_3	
H_1	5 (x)	4 (x)	5 (200)	
H_2	4 (400)	8 (x)	7 (100)	
H_3	6 (x)	4 (600)	7 (200)	
β_j				

Źródło: Opracowanie własne; dane umowne.

Liczenie ponowne zmiennych dualnych dla skorygowanych ilości $\alpha_i + \beta_j = c_{ij}$.

$$8 \cdot 100 + 7 \cdot 0 + 4 \cdot 500 + 7 \cdot 300 = 4900,$$

$$8 \cdot 0 + 7 \cdot 100 + 4 \cdot 600 + 7 \cdot 200 = 4500,$$

$$4900 - 4500 = 400.$$

Tabela 2.10.47

Zmienne dualne

Hurtownia	Składnica rolnicza			α_i
	S_1	S_2	S_3	
H_1	5 (x)	4 (x)	5 (200)	0
H_2	4 (400)	8 (x)	7 (100)	2
H_3	6 (x)	4 (600)	7 (200)	2
β_j	2	2	5	

Źródło: Opracowanie własne; dane umowne.

Obliczenie $\Delta_{ij} = c_{ij} - \alpha_i - \epsilon_j$

3	2	x
x	4	x
2	x	x

Kryteria dodatnie – rozwiązanie jest optymalne. Koszty przewozu obliczono według:

– sposób 1, bazując na rozwiązaniu podanym w tabelach 2.10.46:

$$5 \cdot 200 + 4 \cdot 400 + 7 \cdot 100 + 4 \cdot 600 = 7100 \text{ zł,}$$

– sposób 2, z uwzględnieniem wzoru: $\Delta K = \Delta_{ij} \cdot x_{ij}$; $\Delta_{23} = -4$; $x_{22} = 100$:

$$K_2 = K_1 + \Delta K = 7500 - 4 \cdot 100 = 7100 \text{ zł.}$$

$$\text{Ponieważ: } 8 \cdot 100 + 7 \cdot 0 + 4 \cdot 500 + 7 \cdot 300 = 4900,$$

$$8 \cdot 0 + 7 \cdot 600 + 7 \cdot 200 = 4500,$$

$$\text{zatem: } 4500 - 4900 = -400.$$

Przykład

Podajmy optymalne przewozy węgla w tys. t z kopalń do składnic regionalnych i zastosujmy metodę minimalnego elementu macierzy oraz sprawdzenie optymalności iteracji według kryterium ze zmiennymi dualnymi.

Tabela 2.10.48

Koszty jednostkowe, podaż kopalń i popyt składnic regionalnych

Kopalnia	Składnica regionalna			A_i	Koszt magazynowania
	S_1	S_2	S_3		
K_1	1	3	6	30	5
K_2	2	5	4	20	6
K_3	2	2	7	40	5
B_j	28	25	17	90 (70)	

Źródło: Opracowanie własne; dane umowne.

$$FC: x_{11} + 3x_{12} + 6x_{13} + 2x_{21} + 5x_{22} + 4x_{23} + 2x_{31} + 2x_{32} + 7x_{33} \rightarrow \min.$$

Warunki ograniczające dostawców:

$$\begin{aligned} \sum x_{1j} &\leq 30, & \sum x_{2j} &\leq 20, & \sum x_{3j} &\leq 40, \\ x_{11} + x_{12} + x_{13} &\leq 30, & x_{21} + x_{22} + x_{23} &\leq 20, & x_{31} + x_{32} + x_{33} &\leq 40. \end{aligned}$$

Warunki dla odbiorców:

$$\begin{aligned} x_{11} + x_{21} + x_{31} &= 28, \\ x_{12} + x_{22} + x_{32} &= 25, \\ x_{13} + x_{23} + x_{33} &= 17. \end{aligned}$$

Wprowadzenie fikcyjnego odbiorcy w celu przejścia z OZT do ZZT.

T a b e l a 2.10.49.

Koszt jednostkowy, podaż kopalń i popyt składnic oraz fikcyjny odbiorca

Kopalnia	Składnica regionalna				A_i
	S_1	S_2	S_3	$F(x_{i4})$	
K_1	1	3	6	5	30
K_2	2	5	4	6	20
K_3	2	2	7	5	40
B_j	28	25	17	20	

Ź r ó d ł o: Opracowanie własne; dane umowne.

Dodatkowo warunek dla odbiorcy fikcyjnego:

$$x_{14} + x_{24} + x_{34} = 20$$

$$K(x_{ij}) = x_{11} + 3x_{12} + 6x_{13} + 2x_{21} + 5x_{22} + 4x_{23} + 2x_{31} + 2x_{32} + 7x_{33} + 5x_{14} + 6x_{24} + 5x_{34} \rightarrow \min,$$

x_{i4} – ilość węgla, jaka pozostanie w kopalniach

Przekształcona macierz kosztów:

T a b e l a 2.10.50

Odjęcie wartości minimalnych w wierszach

Kopalnia	Składnica regionalna				A_i
	S_1	S_2	S_3	$F(x_{i4})$	
K_1	0	2	5	4	30
K_2	0	3	2	4	20
K_3	0	0	5	3	40
B_j	28	25	17	20	

Ź r ó d ł o: Opracowanie własne; dane umowne.

Tabela 2.10.51

Odjęcie minimalnej wartości w kolumnach

Kopalnia	Składnica regionalna				A_i
	S_1	S_2	S_3	$F(x_{i4})$	
K_1	0	2	3	1	30
K_2	0	3	0	1	20
K_3	0	0	3	0	40
B_j	28	25	17	20	

Źródło: Opracowanie własne; dane umowne.

Tabela 2.10.52

Początkowe rozwiązanie dopuszczalne
(lokowanie przewozów w klatkach zerowych)

Kopalnia	Składnica regionalna				A_i
	S_1	S_2	S_3	$F(x_{i4})$	
K_1	25			5	
K_2	3		17		
K_3		25		15	
B_j					

Źródło: Opracowanie własne; dane umowne.

Wstępnie rozwiązanie nie jest optymalne, gdyż przewóz fikcyjny ($K_1 - F$) nie jest w klatce zerowej.

$$K_1 = 1 \cdot 25 + 5 \cdot 5 + 2 \cdot 3 + 4 \cdot 17 + 2 \cdot 25 + 5 \cdot 15 = 249.$$

Sprawdzenie analityczne, czy rozwiązanie jest mimo to optymalne:

Tabela 2.10.53

Zmienne dualne dla klatek bazowych $\alpha_i + \beta_j = c_{ij}$

Kopalnia	Składnica regionalna				α_i
	S_1	S_2	S_3	$F(x_{i4})$	
K_1	1 (25)	3 (x)	6 (x)	5 (5)	0
K_2	2 (3)	5 (x)	4 (17)	6 (x)	1
K_3	2 (x)	2 (25)	7 (x)	5 (15)	0
β_j	1	2	3	5	

Źródło: Opracowanie własne w Excelu.

Kryteria optymalności:

zero dla tras bazowych x .

dla pozostałych $\Delta_{ij} = c_{ij} - \alpha_i - \beta_j$. Szukamy kryteriów ujemnych dla komórek niebazowych

x	1	3	x
x	2	x	1
1	x	4	x

Są kryteria dodatnie, zatem rozwiązanie jest optymalne.

2.10.5. Pytania

1. Opisz na podanym przykładzie procedurę iteracyjną zamkniętego zagadnienia transportowego według metody minimalnego elementu macierzy.

Tabela 2.10.54

Koszty jednostkowe, zapotrzebowanie firmy budowlanej i możliwości hurtowni

Hurtownia	Firma			A_i
	F_1	F_2	F_3	
H_1	50	40	50	70
H_2	40	80	70	50
H_3	60	40	70	80
B_j	40	60	100	200

Źródło: Opracowanie własne; dane umowne.

2. Na czym polega:

- metoda kar,
- doprowadzenie otwartego zagadnienia transportowego do zamkniętego,
- utworzenie w macierzy kosztów klatek z zerami.

3. Sprawdź według kryterium ze zmiennymi dualnymi, czy podana iteracja jest optymalna.

T a b e l a 2.10.55

Rozwiązanie n -tej iteracji

Dostawca	Odbiorca		
	O_1	O_2	O_3
D_1	2 (x)	5 (5)	4 (20)
D_2	1 (30)	3 (8)	6 (x)
D_3	2 (x)	2 (10)	7 (x)

Ź r ó d ł o: Opracowanie własne; dane umowne.

W nawiasach ilość dostawy; gdy x, to nie ma dostawy. Przed nawiasem podano koszt jednostkowy przewozu.

2.10.6. Zadania

1. Rozwiąż zagadnienie transportowe dostawy nawozów sztucznych przez hurtownie do gospodarstw rolniczych.

T a b e l a 2.10.56

Koszty jednostkowe, zapotrzebowanie gospodarstw oraz możliwości hurtowni

Hurtownia	Gospodarstwo			A_i	Koszt magazynowania
	S_1	S_2	S_3		
H_1	4	8	15	800 (A_1)	2
H_2	5	4	10	500 (A_2)	5
H_3	6	10	8	900 (A_3)	8
B_j	500	800	400		

Ź r ó d ł o: Opracowanie własne; dane umowne.

2. Określ koszt optymalny transportu według danych tab. 2.10.56, zastępując $A_1 = 800$, ilością $A_1 = 200$.

3. Podaj optymalne przewozy kruszywa w tys. t ze żwirowni na plac budowy. Zastosuj metodę minimalnego elementu macierzy oraz sprawdzanie optymalności iteracji kryterium ze zmiennymi dualnymi.

Tabela 2.10.57

Koszty jednostkowe, podaż żwirowni i potrzeby placów budowy

Żwirownia	Plac budowy			A_i	Koszt magazynowania
	P_1	P_2	P_3		
Z_1	2	4	8	40	6
Z_2	2	6	4	20	4
Z_3	4	2	10	60	8
B_j	30	35	20		

Źródło: Opracowanie własne; dane umowne.

2.11. Zastosowanie WinQSB do programowania dynamicznego

2.11.1. Problem wyboru najkrótszej trasy (komiwojażera)

Moduł programowanie dynamiczne (*Dynamic Programming*) programu WinQSB składa się z trzech algorytmów problemowych:

- komiwojażera – najkrótszej trasy (*Stagecoach Problem*),
- załadunku (Knapsack Problem),
- sterowania produkcją i zapasami (*Production and Inventory Control Problem*).

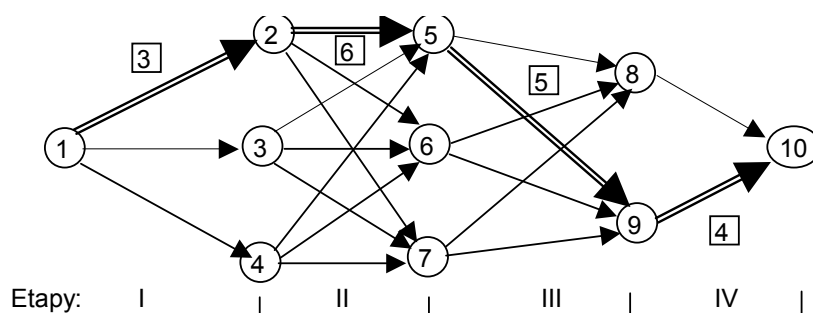
Danymi potrzebnymi do rozwiązania problemu komiwojażera są: liczba etapów, liczba stanów w każdym etapie, odległości/koszty pomiędzy stanami.

Problem załadunku wymaga wprowadzenia następujących danych: całkowita ładowność, objętość danego towaru, oczekiwany zysk ze sprzedaży towaru.

Problem sterowania produkcją i zapasami wymaga danych, takich jak: zapotrzebowanie na produkt, pojemność magazynu wraz z jego jednostkowym kosztem magazynowania, zdolność produkcyjna wraz z jednostkowym kosztem produkcji, koszt uruchomienia produkcji.

P r z y k ł a d

Wybór najkrótszej trasy z miasta 1 do miasta 10¹⁷⁹.



Ź r ó d ł o: Opracowanie własne; dane umowne.

Wykres 2.11.1. Sieć powiązań między miastami

¹⁷⁹ M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne...*, op.cit., rozdz. 4.2.

Najkrótsza trasa między miastami: 1–2–5–9–10 = 18 jednostek odległości.

Tabela 2.11.1

*Odległości między miastami
według węzłów*

Etap I z/do:	1	2	3
1		3	5
1		4	4
Etap II z/do:	2	5	6
2		6	5
2		7	7
3		5	7
3		6	8
3		7	6
4		5	8
4		6	6
4		7	5
Etap III z/do:	5	8	6
5		9	5
6		8	10
6		9	11
7		9	15
7		8	12
Etap IV z/do:	8	10	5
9		10	4

Źródło: Opracowanie własne na podstawie: M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003, Aneks, rozdz. 1.21.

Z menu modułu programowanie dynamiczne (DP) wybieramy opcję (*Stagecoach-Shortest Route Problem*), podając nazwę problemu, liczbę węzłów (*Numer of Nodes*).

Tabela 2.11.2

*Definicja sieci połączeń pomiędzy węzłami–miastami
(problem wyboru najkrótszej trasy – SC)*

From \ To	Node1	Node2	Node3	Node4	Node5	Node6	Node7	Node8	Node9	Node10
Node1		3	5	4						
Node2					6	5	7			
Node3					7	8	6			
Node4					8	6	5			
Node5								6	5	
Node6								10	11	
Node7								12	15	
Node8										5
Node9										4
Node10										

Źródło: Opracowanie własne z zastosowaniem modułu (DP).

W kolejnym kroku definiujemy węzeł początkowy i końcowy (1, 10). Wybór opcji rozwiąż (*Solve*) pozwala na uzyskanie najkrótszej trasy (tab. 2.11.3).

Tabela 2.11.3

Rozwiązanie problemu wyboru najkrótszej trasy między miastami

09-07-2004 Stage	From Input State	To Output State	Distance	Cumulative Distance	Distance to Node10
1	Node1	Node2	3	3	18
2	Node2	Node5	6	9	15
3	Node5	Node9	5	14	9
4	Node9	Node10	6	18	4
	From Node1	To Node10	Min. Distance	18	

Źródło: Opracowanie własne z zastosowaniem modułu (DP).

Wybór opcji (*Solve and Display Steps*) umożliwia tabelaryczną prezentację całej sieci z zaznaczoną najkrótszą trasą. Najkrótsza trasa z miasta 1 do miasta 10 wynosi 18 jednostek odległości i prowadzi przez miasta 1–2–5–9–10.

Tabela 2.11.4

Sieć z zaznaczoną najkrótszą trasą (Optimal)

09-07-2004 09:42:17	Stage	From Input State	To Output State	Distance	Distance to Node10	Status
1	1	Node1	Node3	3	18	Optimal
2	2	Node2	Node5	6	15	Optimal
3	2	Node3	Node5	7	16	
4	2	Node4	Node5	8	17	
5	3	Node5	Node9	5	9	Optimal
6	3	Node6	Node9	11	15	
7	3	Node7	Node8	12	17	
8	4	Node8	Node10	5	5	Optimal
9	4	Node9	Node10	4	4	
	From Node1	To Node10	Minimum	Distance =	18	

Źródło: Opracowanie własne z zastosowaniem modułu (DP).

Moduł (DP) umożliwia ponadto przeprowadzenie analizy „co, jeżeli” po wywołaniu opcji (*Perform What if Analysis*). Polega ona na wyświetleniu lub wydruku najkrótszej trasy do węzła końcowego od wskazanego przez użytkownika; na przykład od węzła drugiego do końcowego nr 10. Według tego modułu jest to również trasa 2–5–9–10, lecz o długości 15 jednostek.

2.11.2. Załadunek

Mamy więc wyznaczoną najkrótszą trasę z miasta 1 do miasta 10. Nasz komiwojażer ma teraz problem załadunku. Do swego samochodu dostawczego, o ładowności 30 jednostek, ma zapakować pięć różnych typów mikrokomputerów K_1, \dots, K_5 . Bierze pod uwagę następujące dane dotyczące towarów do przewiezienia:

- liczbę sztuk poszczególnych typów towarów,
- zysk ze sprzedaży każdego z nich wyrażony w tys. zł,
- kubaturę w m^3 , jaką zajmą w samochodzie.

Tabela 2.11.5
Parametry towarów do załadunku

Towar	Liczba sztuk	Kubatura	Zysk
K_1	10	2	10
K_2	5	1	8
K_3	9	2	5
K_4	4	3	10
K_5	5	6	8

Źródło: Opracowanie własne na podstawie: M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych...*, op.cit., rozdz. 4.2, rozdz. 1.2.2, tabl. 1.17.

Komiwojażer ma prawo wyboru towaru i zakłada, że może wszystko sprzedać, co zabierze. Jakie towary powinien zabrać w podróż, aby osiągnąć maksymalny zysk?

Z menu modułu (DP) wybieramy opcję (*Knapsack Problem*), określając jednocześnie liczbę rodzaju towarów (*Number of Items*). W tabeli 2.11.6 podano definicję problemu załadunku. Wielkość x w definicji funkcji zysku (*Return Function*) oznacza liczbę jednostek poszczególnych towarów.

Tabela 2.11.6

Zdefiniowanie problemu

Pozycja (Stage)	Nazwa (Item)	Liczba sztuk (Available)	Kubatura (Capacity Required)	Funkcja zysku (Return Function)
1	K_1	10	2	10x
2	K_2	5	1	8x
3	K_3	9	2	5x
4	K_4	4	3	10x
5	K_5	5	6	8x
Knapsack	Capacity =	30		

Źródło: Opracowanie własne z zastosowaniem modułu (DP).

Rozwiązanie zadania następuje po wywołaniu opcji (*Solve the Problem*).

Tabela 2.11.7

Rozwiązanie problemu załadunku

09-07-2004 Stage	Item Name	Decision Quantity (x)	Return Function	Total Item Return Value	Capacity Left (poziom załadowania)
1	K1	10	10x	100	10
2	K2	5	8x	40	5
3	K3	1	5x	5	3
4	K4	1	10x	10	0
5	K5	0	8x	0	0
	Total	Return	Value = 155		CPU = 0,04 sec

Źródło: Opracowanie własne z zastosowaniem modułu (DP).

Z rozwiązania komputerowego trwającego 0,04 sekundy wynika, że komiwojazer powinien zabrać: 10 sztuk komputerów K1, 5 sztuk K2, 1 sztukę K3, 1 sztukę K4 i nie zabierać komputerów K5.

2.11.3. Zarządzanie produkcją i zapasami

Przykład

Firma wytwarzająca belki stalowe typu B zawarła umowę z developerem na dostarczenie na budowę w ciągu pięciu miesięcy liczby belek podanych w tab. 2.11.8. Obecny zapas wyrobu B u producenta wynosi 5 sztuk.

Tabela 2.11.8

Określenie problemu produkcji belek

Miesiąc	Popyt	Zdolności produkcyjne	Pojemność magazynu	Koszt uruchomienia produkcji	Jednostkowy koszt produktu	Jednostkowy koszt magazynowania
Lipiec	6	8	5	400	200	80
Sierpień	4	6	4	300	300	90
Wrzesień	7	8	3	400	350	80
Październik	3	7	6	700	300	100
Listopad	5	6	5	300	400	100

Źródło: Opracowanie własne na podstawie: M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych...*, op.cit., rozdz. 4.2, rozdz. 1.2.2, s. 138, tabl. 1.22.

Jaki jest harmonogram produkcji, który przy zapasie wyrobu $B = 5$ zapewni najkorzystniejszą realizację umowy?

Z modułu (DP) programu WinQSB wybieramy opcję (*Production and Inventory Scheduling*). Dodatkowo określamy liczbę analizowanych okresów (*Number of Periods*) wynoszącą 5.

Tabela 2.11.9

Komputerowa definicja problemu

Stage	Period	Demand	Production Capacity	Storage Capacity	Production Setup Cost	Variable Cost Function
1	Lipiec	6	8	5	400	$200P+80H$
2	Sierpień	4	6	4	300	$300P+90H$
3	Wrzesień	7	8	3	400	$350P+80H$
4	Październik	3	7	6	700	$300P+100H$
5	Listopad	5	6	5	300	$400P+100H$

Stage – etap,
 Period – okres,
 Demand – popyt,
 Production Capacity – wielkość produkcji,
 Storage Capacity – wielkość magazynu,
 Production Setup Cost – koszt uruchomienia produkcji,
 Variable Cost Function – koszty produkcji i magazynowania.

Źródło: Opracowanie własne, moduł (DP).

Tabela 2.11.10

Rozwiązanie problemu

09-07-2004	Period	Net	Starting	Production	Ending	Setup	Variable Cost	Variable	Total
Stage	Description	Demand	Inventory	Quantity (P)	Inventory (H)	Cost	Function (P,H)	Cost	Cost
1	Lipiec	1	0	5	4	400,00 zł	$200P+80H$	1 320,00 zł	1 720,00 zł
2	Sierpień	4	4	0	0	0,00 zł	$300P+90H$	0,00 zł	0,00 zł
3	Wrzesień	7		8	1	400,00 zł	$350P+80H$	2 880,00 zł	3 280,00 zł
4	Październik	3	1	7	5	700,00 zł	$300P+100H$	2 600,00 zł	3 300,00 zł
5	Listopad	5	5	0	0	0,00 zł	$400P+100H$	0,00 zł	0,00 zł
Total		20	10	20	10	1 500,00 zł		6 800,00 zł	8 300,00 zł

Stan magazynu na początku (Starting Inventory) wynosi zero, gdyż wykorzystuje się pełny zapas do sprzedaży.

Źródło: Opracowanie własne, moduł (DP).

Z rozwiązania wynika, że firma, aby wywiązać się z umowy, powinna produkować w miesiącach od lipca do listopada belki stalowe w liczbie podanej w kolumnie Production Quantity (P). Minimalny łączny koszt produkcji i magazynowania w ciągu pięciu miesięcy wyniesie 8300 zł.

2.11.4. Pytania

1. Jakie dane wejściowe są potrzebne do rozwiązania problemów algorytmami:

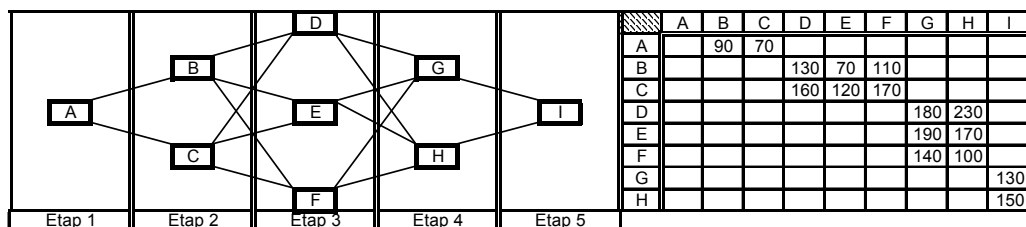
- komiwojażera (najkrótszej trasy),
- załadunku,
- sterowania produkcją i zapasami.

2. Określ tabelę definiowania problemu załadunku wprowadzaną do komputera opcją (*Knapsack Problem*) w celu osiągnięcia maksymalnego zysku ze sprzedaży pięciu towarów *A, B, C, D, E*.

3. Przeanalizuj tab. 2.11.8 definiowania problemu zarządzania produkcją i zapasami wyrobu *B*. Określ, które informacje są potrzebne do zastosowania komputerowej opcji produkcja i harmonogramowanie zapasów (*Production and Inventory Scheduling*).

2.11.5. Zadania

1. Zaznacz na podanej sieci odległości najkrótszą trasę samochodu ciężarowego pokonującego pięć etapów do przewiezienia załadunku między miejscowościami A–I, analizując najkrótsze odległości od ostatniego etapu. Porównaj swoje rozwiązanie z trasą otrzymaną z modułu (DP) opcją (*Stagecoach Problem*) pakietu WinQSB (tab. 2.11.11).



Źródło: Opracowanie własne; dane umowne.

Wykres 2.11.2. Odległości między miejscowościami

Tabela 2.11.11

Komputerowe określenie najkrótszej trasy

09-25-2007	From	To	Distance	Cumulative	Distance to I
Stage	Input State	Output State		Distance	
1	A	B	90	90	450
2	B	F	110	200	360
3	F	H	100	300	250
4	H	I	150	450	150
	From A	To I	Min. Distance	450	

Źródło: Opracowanie własne, moduł (DP).

Rozwiązanie optymalne: trasa A–B–F–H–I. Długość trasy (90 + 110 + 100 + 150) = 450 jednostek.

2. Biorąc za podstawę omawianą sieć połączeń między etapami (A–I), określ najkrótszą trasę między miejscowościami B–I. Sprawdź swoje rozwiązanie z podaną odpowiedzią modułu (DP) opcja (*What if Analysis*) – co, jeżeli?:

Best route from B to I:
B–F–H–I Total distance = 360.

3. Po określeniu optymalnej trasy przewozu z miejscowości A do I wyznacz, jakie ilości towarów powinien zabrać kierowca do swego samochodu o ładowności 30 jednostek. Zdefiniuj problem załadunku, bazując na danych tab. 2.11.12 oraz rozwiąż zadanie, stosując opcję problem załadunku (*Knapsack Problem*) moduł (DP). Porównaj swoje wyniki z raportem w tab. 2.11.13.

T a b e l a 2.11.12

Problem załadunku

Nazwa towaru	Liczba sztuk	Kubatura jednostki	Zysk [zł]
A	10	1	50
B	15	2	100
C	10	1,5	200
D	5	0,5	300
E	20	1	100

Ź r ó d ł o: Opracowanie własne; dane umowne.

T a b e l a 2.11.13

Rozwiązanie problemu załadunku

09-25-2007 Stage	Item Name	Decision Quantity (X)	Return Function	Total Item Return Value	Capacity Left
1	A	0	50X	0	30
2	B	0	100X	0	30
3	C	9	200X	1800	16,5 *
4	D	5	300X	1500	14
5	E	14	100X	1400	0
Total		Return	Value = 4700		CPU = 0,18

* Przykładowo: $16,5 = 30 - (9 \cdot 1,5)$

Ź r ó d ł o: Opracowanie własne; moduł (DP)

4. Zakład stolarski produkujący pudełka drewniane *P* zawarł umowę na ich dostawę w kwartałach roku, nie mając zapasu. Jaki powinien być harmo-

nogram produkcji pudełek przy minimalizacji kosztu produkcji i magazynowania? W tabeli 2.11.14 podano: liczbę sztuk pudełek, zdolność produkcyjną, pojemność magazynowania w sztukach, oraz koszty uruchomienia produkcji, koszt jednostkowy produktu, koszt jednostkowy magazynowania wyrażone w zł. Do rozwiązania zastosuj opcję (*Production and Inventory Scheduling*) modułu (DP). Porównaj swój wynik z rozwiązaniem podanym w tab. 2.11.15.

Tabela 2.11.14

Informacje o produkcji i magazynowaniu drewnianych pudełek

Kwartał	Liczba sztuk	Zdolność	Pojemność	Koszt uruchomienia	Koszt produktu (P)	Koszt magazynowania (H)
1	5	5	6	10	5	2
2	8	10	8	20	5	2
3	4	6	6	30	6	3
4	7	6	6	20	6	3

Źródło: Opracowanie własne; dane umowne.

Tabela 2.11.15

Raport z realizacji komputerowej problemu produkcji i magazynowania

09-25 -2007 Stage	Period Description	Net Demand	Starting Inventory	Production Quantity	Ending Inventory	Setup Cost	Variable Cost Function (P,H)	Variable Cost	Total Cost
1	Kwartał 1	5	0	5	0	10,00 zł	5P+2H	25,00 zł	35,00 zł
2	Kwartał 2	8	0	8	0	20,00 zł	5P+2H	40,00 zł	60,00 zł
3	Kwartał 3	4	0	5	1	30,00 zł	6P+3H	33,00 zł	63,00 zł
4	Kwartał 4	7	1	6	0	20,00 zł	6P+3H	36,00 zł	56,00 zł
Total		24	1	24	1	80,00 zł		134,00 zł	214,00 zł

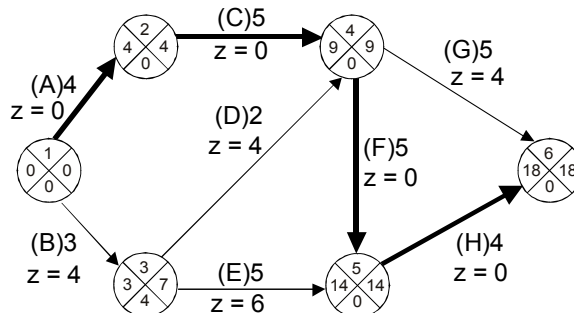
Źródło: Opracowanie własne, moduł (DP).

2.12. Analiza przedsięwzięć

2.12.1. Metoda ścieżki krytycznej CPM¹⁸⁰

Przykład

Realizacja manualna (węzły 1–6)



Źródło: Opracowanie własne.

Wykres 2.12.1. Sieć zależności i wyniki analizy

Realizacja komputerowa, moduł (PERT-CPM), programu WinQSB.

Tabela 2.12.1

Zdefiniowanie problemu

Activity Number	Activity Name	Immediate Predecessor	Normal Time
1	A		4
2	B		3
3	C	A	5
4	D	B	2
5	E	B	5
6	F	C,D	5
7	G	C,D	5
8	H	E,F	4

Activity Number – numer czynności,

Immediate Predecessor – czynności poprzedzające.

Źródło: Opracowanie własne; dane umowne.

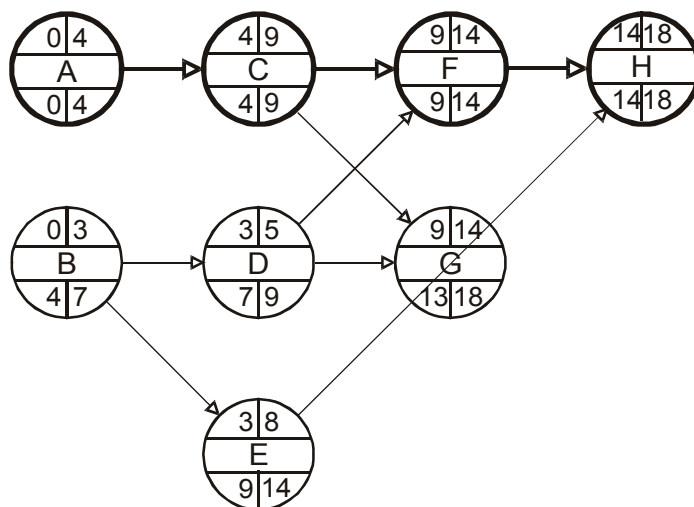
¹⁸⁰ Ibidem, rozdz. 5.2.

Tabela 2.12.2

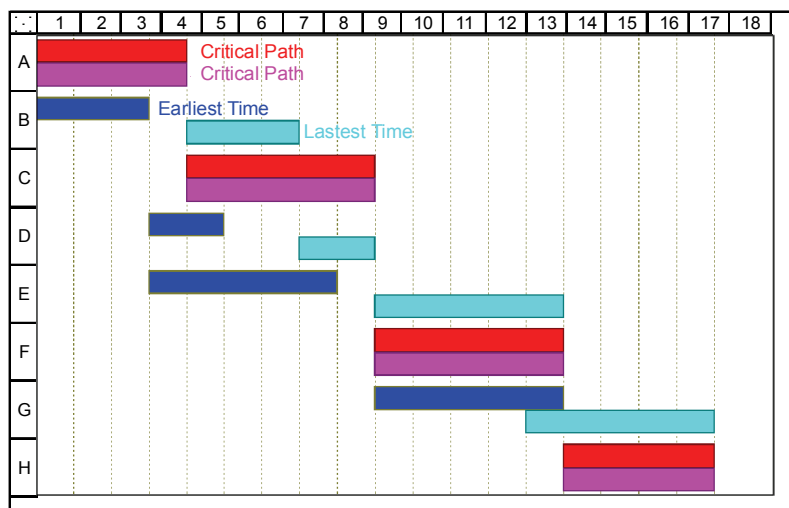
Analiza czynności

Activity Number	Activity Name	On Critical Path	Activity Time	Earliest Start (ES)	Earliest Finish (EF)	Latest Start (LS)	Latest Finish (LF)	Slack (LS-ES)
1	A	yes	4	0	4	0	4	0
2	B	no	3	0	3	4	7	4
3	C	yes	5	4	9	4	9	0
4	D	no	2	3	5	7	9	4
5	E	no	5	3	8	9	14	6
6	F	yes	5	9	14	9	14	0
7	G	no	5	9	14	13	18	4
8	H	yes	4	14	18	14	18	0
Project Completion		Time	=	18	weeks			
Number of Critical Path(s)		=	1					
Termin czynności:				ES, EF – najwcześniejszy (start, finish), LS, LF – najpóźniejszy (start, finish).				

Źródło: Opracowanie własne; dane umowne.



Źródło: Opracowanie własne z zastosowaniem modułu (PERT-CPM).
Wykres 2.12.2. Prezentowanie przez moduł (PERT-CPM) węzłów jako czynności



Critical Path – ścieżka krytyczna,
 Earliest Time – czas najwcześniejszy,
 Lastest Time – czas najpóźniejszy,
 A–H czynności od 1 do 18 tygodni.

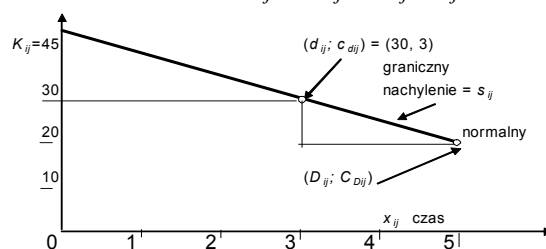
Źródło: Opracowanie własne z zastosowaniem modułu (PERT-CPM).

Wykres 2.12.3. Prezentowanie sieci zależności wraz ze ścieżką krytyczną jako wykres belkowy Gantta

2.12.2. Wersja kosztowa (CPM)

Wersja ta uwzględnia czas i koszt realizacji czynności. Krzywa czasowo-kosztowa każdej czynności aproksymowana jest funkcją liniową (punkty: normalny, graniczny)¹⁸¹. Bezpośredni koszt czynności (i, j):

$$K_b = a_0 + a_1 x_{ij} = K_{ij} + S_{ij} \cdot x_{ij}.$$



Źródło: Opracowanie własne w Excelu.

Wykres 2.12.4. Krzywa czasowo-kosztowa czynności (przykład)

¹⁸¹ Ibidem, rozdz. 5.2.3.

Tabela 2.12.3

Opis czynności (na przykładzie czynności A)

Czynność/ nazwa	Początek	Koniec	Czas trwania		Koszty	
	<i>i</i>	<i>j</i>	normalny (<i>D_{ij}</i>)	graniczny (<i>d_{ij}</i>)	normalny (<i>CD_{ij}</i>)	graniczny (<i>cd_{ij}</i>)
A	1	2	5	3	20	30

Źródło: Opracowanie własne.

Na podstawie wykresu 2.12.4:

$$S_{ij} = \frac{C_{D_{ij}} - c_{d_{ij}}}{D_{ij} - d_{ij}}, \text{ stąd: } \frac{30 - K_{ij}}{3 - 0} = -5, \quad K_{ij} = 45,$$

$$\text{lub nachylenie } S_{ij} = (20 - 30) / (5 - 3) = -5$$

W ten sposób, posługując się Excelem, możemy obliczyć nachylenia S_{ij} wszystkich czynności. Zatem całkowity koszt projektu (przedsięwzięcia) można określić wzorem:

$$K_p = \sum(K_{ij} + S_{ij} \cdot x_{ij}).$$

2.12.3. Zadanie programowania liniowego według metody simpleks

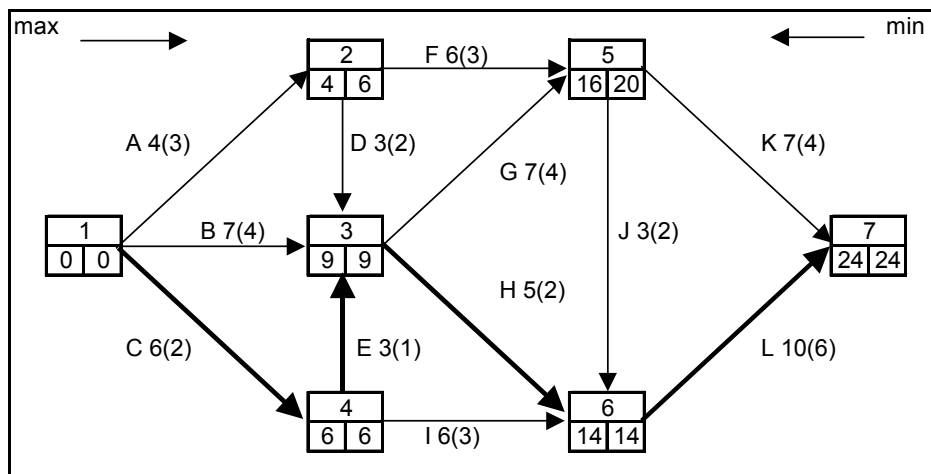
Określenia kroków postępowania¹⁸²:

- 1) Założenie czasu zakończenia projektu T .
- 2) Dobór x_{ij} i określenie funkcji celu:

$$\sum (K_{ij} + S_{ij} \cdot x_{ij}) \rightarrow \min.$$

- 3) Określenie dodatkowej zmiennej pomocniczej y_k dla najwcześniejszego czasu wystąpienia zdarzenia k .

¹⁸² Ibidem, przykład 5.3.



Ź r ó d ł o: Opracowanie własne na podstawie: M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003, rys. 5.8.

Wykres 2.12.5. Sieć realizacji projektu

Na schemacie wymieniono czynności A–L oraz ich czasy trwania (normalny i graniczny – w zdarzeniach). Przykładowe określenie najwcześniejszego wystąpienia zdarzenia 3¹⁸³:

$$y_3 = \max[y_2 + x_{23}, y_1 + x_{13}, y_4 + x_{43}],$$

gdzie:

$$y_2 + x_{23} \leq y_3,$$

$$y_1 + x_{13} \leq y_3,$$

$$y_4 + x_{43} \leq y_3.$$

Podobnie można wprowadzić warunki ograniczające dla wszystkich zdarzeń, przy czym:

$$y_i + x_{ij} \leq y_j, \quad \text{lub w postaci} \quad y_i + x_{ij} - y_j = 0.$$

4) $y_1 = 0$, $y_n =$ czas zakończenia projektu.

5) W funkcji celu suma kosztów $\sum K_{ij}$ jest wartością stałą i można ją pominąć, gdyż pochodna stałej wynosi zero.

Minimalizacja całkowitego kosztu bezpośredniego projektu jest tożsama z maksymalizacją funkcji: $\sum (-S_{ij}) \cdot x_{ij}$

¹⁸³ Ibidem, s. 113.

Funkcja celu z pominięciem $\sum K_{ij}$ będzie następująca:

$$FC: Z = \sum (-S_{ij}) \cdot x_{ij} \rightarrow \max.$$

Warunki ograniczające dla wszystkich czynności (i,j) :

$$x_{ij} \geq d_{ij}, \quad x_{ij} \leq D_{ij}, \quad y_i + x_{ij} - y_j = 0, \quad y_n \leq T.$$

Warunki brzegowe: $x_{ij} \geq 0, y_k \geq 0$.

Przyjmujemy założenie, że każda ścieżka może być krytyczna dla danego założonego czasu realizacji T .

6) Charakterystyka terminów zdarzeń w sieci według czasów normalnych.

Tabela 2.12.4

Charakterystyki projektu

Czynność/ nazwa	Początek	Koniec	Czas trwania		Koszty		-S _{ij}	Ścieżka krytyczna Czas normalny
	i	j	normalny	graniczny	normalny	graniczny		
A	1	2	4	3	20	30	10	
B	1	3	7	4	30	40	3,3	
C	1	4	6	2	40	50	2,5	40
D	2	3	3	2	30	30	0	
E	4	3	3	1	20	30	5	20
F	2	5	6	3	30	60	10	
G	3	5	7	4	40	50	3,3	
H	3	6	5	2	40	50	3,3	40
I	4	6	6	3	40	60	6,7	
J	6	5	3	2	40	40	0	
K	5	7	7	4	60	70	3,3	
L	6	7	10	6	80	100	5	80
				Suma:	470	610		180

Źródło: Opracowanie własne w Excelu na podstawie: M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003, przykład 5.3.

7. Ustalenie najwcześniejszych i najpóźniejszych terminów wystąpienia zdarzeń w czasach normalnych według sieci, poruszając się odpowiednio od zdarzenia 1. do 7. i odwrotnie.

8. Określenie zapasu na czynnościach, przy czym na ścieżce krytycznej obejmującej czynności C, E, H, L zapasy wynoszą zero.

T a b e l a 2.12.5

Określenie zapasów na ścieżce krytycznej

Czynność (i,j)	Zapas = $z_{ij} = T_j^P - T_j^W - t_{ij}$
C(1,4)	6 - 0 - 6 = 0
E(4,3)	9 - 6 - 3 = 0
H(3,6)	14 - 9 - 5 = 0
L(6,7)	24 - 14 - 10 = 0

T_j^P – najwcześniejszy termin zajścia danego zdarzenia,

T_j^W – najpóźniejszy termin zajścia danego zdarzenia,

$t_{ij} = x_{ij}$ – czas trwania czynności.

Ź r ó d ł o: Opracowanie własne w Excelu na podstawie:
M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003, przykład 5.3.

9) Podsumowanie:

w czasach normalnych: ścieżka krytyczna C–E–H–L; $T = 24$; koszt przedsięwzięcia $K_p = 470$,

$$K_{\text{ścieżka}} = 180,$$

w czasach granicznych trzy ścieżki A–D–G–K, A–D–H–J–K, A–D–H–L (180; 220; 210), przy czym minimum to 180; $T = 13$; $K_p = 610$.

A teraz sformułujemy zadanie decyzyjne z funkcją celu maksimum (korzystając z wersji kosztowej CPM w WinQSB)

Na ekranie definiowania zadania decyzyjnego w module programowania liniowego funkcję celu zapisujemy obok komórki (*Maximize*), pamiętając, aby dane były z kropką dziesiętną. Jak już wcześniej wskazano, funkcja celu bazuje na określeniu:

$$Z = \sum (-S_{ij}) \cdot x_{ij} \rightarrow \max.$$

Wymagało to obliczenia nachyleń S_{ij} (tab. 2.12.4) dla wszystkich czynności w sieci przedsięwzięć. Przykładowo $(-S_{ij})$ dla czynności A między węzłami 1–2 wynosi:

$$(-S_{12}) = -\frac{20 - 30}{4 - 3} = 5.$$

Funkcją celu według zapisu w module (LP) programu WinQSB przedstawimy jako:

$$FC: 10X_{12}+3.3X_{13}+2.5X_{14}+0X_{23}+5X_{43}+10X_{25}+3.3X_{35}+3.3X_{36}+6.7X_{46}+0X_{65}+3.3X_{57}+5X_{67}.$$

W zadaniu występuje dwanaście zmiennych X_{ij} w wymienionej funkcji celu, zmienne pomocnicze Y_1 – Y_7 oraz 38 ograniczeń dla siedmiu węzłów. Zdefiniowanie funkcji celu, ograniczeń i warunków brzegowych zaprezentowano również w postaci normalnej (nierówności).

T a b e l a 2.12.6

Zdefiniowanie problemu wersji kosztowej CPM

	OBJ/Constraint/Bound		
Maximize		Y2Y3	Y2+X23-Y3=0
C13G	X12>=3	Y4Y3	Y4+X43-Y3=0
C12N	X12<=4	Y2Y5	Y2+X25-Y5=0
C13G	X13>=4	Y3Y5	Y3+X35-Y5=0
C13N	X13<=7	Y3Y6	Y3+X36-Y6=0
C14G	X14>=2	Y4Y6	Y4+X46-Y6=0
C14N	X14<=6	Y5Y7	Y5+X57-Y7=0
C23G	X23>=2	Y6Y5	Y6+X65-Y5=0
C23N	X23<=3	Y6Y7	Y6+X67-Y7=0
C43G	X43>=1	Integer:	
C43N	X43<=3	Binary:	
C25G	X25>=3	Unrestricted:	
C25N	X25<=6	X12	>=0, <=M
C35G	X35>=4	X13	>=0, <=M
C35N	X35<=7	X14	>=0, <=M
C36G	X36>=2	X23	>=0, <=M
C36N	X36<=5	X43	>=0, <=M
C46G	X46>=3	X25	>=0, <=M
C46N	X46<=6	X35	>=0, <=M
Y1	Y1=0	X36	>=0, <=M
Y7	Y7<=24	X46	>=0, <=M
C65G	X65>=2	X65	>=0, <=M
C65N	X65<=3	X57	>=0, <=M
C57G	X57>=4	X67	>=0, <=M
C57N	X57<=7	Y1	>=0, <=M
C67G	X67>=6	Y2	>=0, <=M
C67N	X67<=10	Y3	>=0, <=M
Y1Y2	Y1+X12-Y2=0	Y4	>=0, <=M
Y1Y3	Y1+X13-Y3=0	Y5	>=0, <=M
Y1Y4	Y1+X14-Y4=0	Y6	>=0, <=M
		Y7	>=0, <=M

Ź r ó ł o: Opracowanie własne, moduł (LP) programu WinQSB.

Rozwiązanie zaprezentowane w tab. 2.12.7 uzyskane zostało w 32 iteracjach w zaledwie w 0,07 s pracy jednostki centralnej mikrokomputera. W rozwiązaniu komputerowym podane są czasy trwania poszczególnych

czynności dobrane z przedziału normalno-granicznego, a ponadto przy założeniu $Y_1 = 0$ czasu dla zmiennych pomocniczych Y_2 – Y_6 .

Tabela 2.12.7

Rozwiązanie sumaryczne w wersji kosztowej CPM

12-19-2006 21:17:57	Decision Variable	Solution Value	Unit Cost or Profit C(j)	Total Contribution
1	X12	4	10	40
2	X13	6	3,3	19,8
3	X14	3	2,5	7,5
4	X23	2	0	0
5	X43	3	5	15
6	X25	6	10	60
7	X35	4	3,3	13,2
8	X36	2	3,3	6,6
9	X46	5	6,7	33,5
10	X65	2	0	0
11	X57	7	3,3	23,1
12	X67	9	5	45
13	Y1	0	0	0
14	Y2	4	0	0
15	Y3	6	0	0
16	Y4	3	0	0
17	Y5	10	0	0
18	Y6	8	0	0
19	Y7	17	0	0
Objective Function			(Max.) =	263,7

Wartość funkcji celu (Objective Function) Max = 263,7

Źródło: Opracowanie własne, moduł (LP) programu WinQSB.

Fragment pierwszej tabeli simpleksowej – bazowej (tab. 2.12.8) po wywołaniu opcji (*Solve and Display Steps*) zawiera w kolumnach tylko dziewiętnaście zmiennych decyzyjnych (X_{12} – Y_7). W wektorze zmiennych bazowych pierwszej iteracji zamiany postaci standardowej (nierówności) w postać kanoniczną, a następnie w postać zwaną „uzupełnioną” uwzględniającą macierz jednostkową, mikrokomputer wygenerował odpowiednie zmienne swobodne oraz sztuczne.

W wierszu „Basis” wpisane są współczynniki funkcji celu, a w kolejnych wierszach współczynniki zmiennych decyzyjnych, swobodnych oraz sztucznych poszczególnych ograniczeń zadania decyzyjnego. Ponieważ funkcja celu dąży do maksimum, współczynniki funkcji celu zmiennych swobodnych mają wartość zero, a dla zmiennych sztucznych ($-M$).

W wierszu k_j obliczono kryteria stanowiące różnicę między współczynnikiem funkcji celu c_j a sumą iloczynów cen bazowych i współczynników macierzy danej zmiennej.

Tabela 2.12.8

Fragment tabeli bazowej I iteracji

Simplex Tableau -Iteration 1																				
		X12	X13	X14	X23	X43	X25	X35	X36	X46	X65	X57	X67	Y1	Y2	Y3	Y4	Y5	Y6	Y7
Basis	$C(j)$	10	3,3	2,5	0	5	10	3,3	3,3	6,7	0	3,3	5	0	0	0	0	0	0	0
Artificial_C13G	-M	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Slack_C12N	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Artificial_C13G	-M	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Slack_C13N	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Artificial_C14G	-M	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Slack_C14N	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Artificial_C23G	-M	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Slack_C23N	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Artificial_C43G	-M	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Slack_C43N	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Artificial_C25G	-M	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Slack_C25N	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Artificial_C35G	-M	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Slack_C35N	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Artificial_C36G	-M	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Slack_C36N	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Artificial_C46G	-M	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Slack_C46N	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Artificial_Y1	-M	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Slack_Y7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Artificial_C65G	-M	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Slack_C65N	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Artificial_C57G	-M	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Slack_C57N	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Artificial_C67G	-M	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Slack_C67N	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Artificial_Y1Y2	-M	1	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0
Artificial_Y1Y3	-M	0	1	0	0	0	0	0	0	0	0	0	0	1	0	-1	0	0	0	0
Artificial_Y1Y4	-M	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	-1	0	0	0
Artificial_Y2Y3	-M	0	0	0	1	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0
Artificial_Y4Y3	-M	0	0	0	0	1	0	0	0	0	0	0	0	0	0	-1	1	0	0	0
Artificial_Y2Y5	-M	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	-1	0	0
Artificial_Y3Y5	-M	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	-1	0	0
Artificial_Y3Y6	-M	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	-1	0
Artificial_Y4Y6	-M	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	-1	0
Artificial_Y5Y7	-M	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	-1
Artificial_Y6Y5	-M	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	-1	1	0
Artificial_Y6Y7	-M	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	-1
$k_j =$	$C(j)-Z(j)$	10	3,3	2,5	0	5	10	3,3	3,3	6,7	0	3,3	5	0	0	0	0	0	0	0
	* Big M	2	2	2	2	2	2	2	2	2	2	2	2	4	1	-1	1	-2	0	-2

Źródło: Opracowanie własne, moduł (LP) programu WinQSB.

2.12.4. Metoda PERT

W omówieniu tego tematu z zastosowaniem programu WinQSB posłużono się siecią zależności podaną na wykresie 2.12.1.

W odróżnieniu do metody CPM w metodzie PERT zakłada się losowość czasu realizacji, a czas określa się symbolem T_e ¹⁸⁴ i rozpatrujemy problem jako probabilistyczny. W obliczaniu T_e bazujemy na oszacowanych czasach: optymistycznym a , najbardziej prawdopodobnym m , pesymistycznym b . Interpretacja czasów najwcześniejszych oraz najpóźniejszych jest analogiczna do metody CPM, lecz w kategorii czasów średnich ustalanych według wzoru:

$$T_e = (a + 4m + b) / 6 \quad (2.12.1)$$

Na przykład dla czynności A podanej na wykresie 2.12.1 o czasie $t = m = 4$ i oszacowaniu czasu $a = 3$, a czasu $b = 5$:

$$T_e = (3 + 4 \cdot 4 + 5) / 6 = 4.$$

Tak określony czas stanowi podstawę do wyznaczenia ścieżki krytycznej według wcześniej omówionej metody CPM.

Pierwiastek z wariancji czasu oczekiwanego między węzłami $i-j$ danej czynności σ_{i-j}^2 jest miarą rozbieżności między oceną optymistyczną a pesymistyczną czasu trwania czynności.

$$\sigma_{i-j}^2 = \left(\frac{b-a}{6} \right)^2 \quad (2.12.2)$$

Przykładowo dla czynności A: $\sigma_{1-2}^2 = \sigma_{1-2}^2 = \left(\frac{5-3}{6} \right)^2 = \frac{1}{9}$. Zatem pierwiastek z wariancji czasu oczekiwania wynosi $1/3$.

T a b e l a 2.12.9

Czasy realizacji czynności

Czynność	Czas optymistyczny a	Czas najbardziej prawdopodobny m	Czas pesymistyczny b	T_e
A	3	4	5	4
B	1	3	6	3,2
C	4	5	7	5,2
D	1	2	2	1,8
E	3	5	6	4,8
F	4	5	7	5,2
G	3	5	8	5,2
H	2	4	5	3,8

Ź r ó d ł o: Opracowanie własne w Excelu.

¹⁸⁴ K. Kukuła (red.), *Badania operacyjne w przykładach i zadaniach...* op.cit., rozdz. 5.3.

Tabela 2.12.10

Definicja sieci

Activity Number	Activity Name	Immediate Predecessor	Optimistic time a	Most likely time m	Pessimistic time b
1	A		3	4	5
2	B		1	3	6
3	C	A	4	5	7
4	D	B	1	2	2
5	E	B	3	5	9
6	F	C,D	4	5	7
7	G	C,D	3	5	8
8	H	E,F	2	4	5

Źródło: Opracowanie własne, moduł (PERT-CPM).

Tabela 2.12.11

Rozwiązanie komputerowe problemu probabilistycznego

Activity Number	Activity Name	On Critical Path	Activity Mean Time T_e	Earliest Start	Earliest Finish	Latest Start	Latest Finish	Slack (LS-ES)	Standard Deviation
1	A	Yes	4	0	4	0	4	0	0,33
2	B	no	3,17	0	3,17	4,17	7,33	4,17	0,83
3	C	Yes	5,17	4	9,17	4	9,17	0	0,5
4	D	no	1,83	3,17	5	7,33	9,17	4,17	0,17
5	E	no	5,33	3,17	8,5	9	14,33	5,83	1
6	F	Yes	5,17	9,17	14,33	9,17	14,33	0	0,5
7	G	no	5,17	9,17	14,33	13	18,17	3,83	0,83
8	H	Yes	3,83	14,33	18,33	14,33	18,17	0	0,5
Project Completion Time =				18,17	weeks				
Number of Critical Path(s) =				1					

Activity Mean Time – oczekiwany czas realizacji czynności,
Standard Deviation – odchylenie standardowe dla czasów realizacji,
weeks – tygodnie.

Project Completion Time – całkowity czas realizacji

Number of Critical Path(s) – ilość ścieżek krytycznych

Źródło: Opracowanie własne, moduł (PERT-CPM), problem (probabilistic PERT), zaokrąglenie do dwóch miejsc po przecinku w Excelu.

Czas realizacji projektu: 18, 17 tygodni. Występuje jedna ścieżka krytyczna.

Przykład

Ekolog planuje wykonanie zadaszonego drewnianego poidła dla zwierząt leśnych. Opracował następujący plan realizacji z oszacowaniem czasu trwania czynności od-do w godzinach.

Tabela 2.12.12

Czasy realizacji czynności

Symbol	Nazwa czynności	Czynności poprzedzające	Czas trwania			$T_e = \frac{a+4m+b}{6}$
			a	m	b	
A	Sporządzenie rysunku		10	15	23	15,5
B	Zakup materiałów	A	5	10	15	10
C	Cięcie elementów	B	15	23	30	22,8
D	Montaż elementów podestu	C	12	19	26	19
E	Montaż elementów dachu	C	12	19	26	19
F	Wykonanie konstrukcji połączenia dachu z podestem	D,E	7	12	20	12,5
G	Malowanie poidła	F	12	17	25	17,5
H	Schnięcie	G	10	10	10	10
I	Umieszczenie poidła na polanie	H	7	7	7	7

Źródło: Opracowanie własne w Excelu.

Przyjmijmy założenie: czas trwania przedsięwzięcia $t_d = 112$ godzin. Interesuje nas, jakie jest prawdopodobieństwo, że zbudujemy poidło w czasie krótszym niż 112 godzin.

$$P(X \leq 112).$$

Pozostawia się Czytelnikowi narysowanie sieci czynności, korzystając z modułu (PERT-CPM) programu WinQSB. Po zdefiniowaniu sieci w podanym module uzyskujemy czas realizacji $t_w = 114,33$ godz. oraz odchylenie standardowe dla przedsięwzięcia:

$$\sigma_w = 5,34 \text{ godz.}$$

W rozwiązaniu komputerowym występuje ścieżka krytyczna A-B-C-E-F-G-H-I.

Wariancja przedsięwzięcia σ_w^2 jest sumą wariancji czynności na ścieżce krytycznej. Program określa odchylenia standardowe σ_w dla poszczególnych czynności po wywołaniu funkcji rozwiązywania (*Solve and Analyze*), które wynoszą:

A	B	C	D	E	F	G	H	I
2,1667	1,6667	2,5	1,6667	2,3333	2,1667	2,1667	0	0

$$\begin{aligned} \sigma_{tw}^2 &= \sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_E^2 + \sigma_F^2 + \sigma_G^2 + \sigma_H^2 + \sigma_I^2 = 2,1667^2 + 1,6667^2 + 2,5^2 + \\ &+ 2,3333^2 + 2,1667^2 + 2,1667^2 = 28,5894. \end{aligned}$$

Pierwiastek z tej wariancji jest odchyleniem między ocenami optymistycznymi a pesymistycznymi:

$$\sigma_{tW} = \sqrt{\sigma_{tW}^2} = \sqrt{28,5894} = 5,34,$$

czyli tyle, ile podaje moduł (PERT-CPM) programu WinQSB.

Jak już wiadomo, odchylenie standardowe σ_w czasów realizacji poszczególnych czynności obliczyć można według wzoru 2.12.2 i tak np. dla czynności A:

$$\sigma^2_A = \left(\frac{b-a}{6}\right)^2 = \left(\frac{23-10}{6}\right)^2 = 4,6946; \quad \sigma_A = \sqrt{4,6946} = 2,1667.$$

Przy założeniu rozkładu normalnego oraz dokonaniu standaryzacji zmiennej X prawdopodobieństwo, że czas realizacji przedsięwzięcia będzie mniejszy od 112 godz. wynosi:

$$X \approx N(t_w, \sigma_{tW}), \quad X \approx N(114,33; 5,34).$$

$$P(X \leq 112) = P\left(\frac{X-114,33}{5,34} \leq \frac{112-114,33}{5,34}\right) = P(U \leq -0,44) = \varphi(-0,44) = \\ = 1 - \varphi(0,44) = 1 - 0,67 = 0,33$$

Zatem prawdopodobieństwo wykonania poidła dla zwierząt leśnych w czasie mniejszym niż 112 godzin wynosi 0,33.

2.12.5. Budowa sieci zależności programem Project

Program Project pracuje w środowisku Windows. Służy do projektowania sieci przedsięwzięć. W przykładzie zademonstrowano zastosowanie wersji Microsoft Project for Windows 95, którą można zainstalować z siedmiu dyskietek, wkładając kolejno do napędu A:/. Po włożeniu pierwszej dyskietki inicjujemy procedurę instalacji poprzez wywołanie pliku (*Setup*) i naciśnięcie klawisza *Enter*. Następnie postępujemy zgodnie z instrukcją na ekranie.

Procedura instalacji dotyczy wersji dostępnej autorom niniejszej pracy.

P r z y k ł a d

Sporządzenie sieci zależności do tematu „Budowa zagłówek przez rodzinę w celu spędzenia urlopu”.

Na potrzeby zdefiniowania tematu oraz rozwiązania problemu z użyciem programu Project wykonujemy zestawienie czynności (*Task Name*) obejmujące:

- czasy trwania czynności (*Duration*) w dniach (d),
- dni tygodnia rozpoczęcia oraz zakończenia (*Start, Finish*),
- czynności poprzedzające (*Predecessor*),
- zasoby – wykonawcy (*Resource Names*).

Tabela 2.12.13

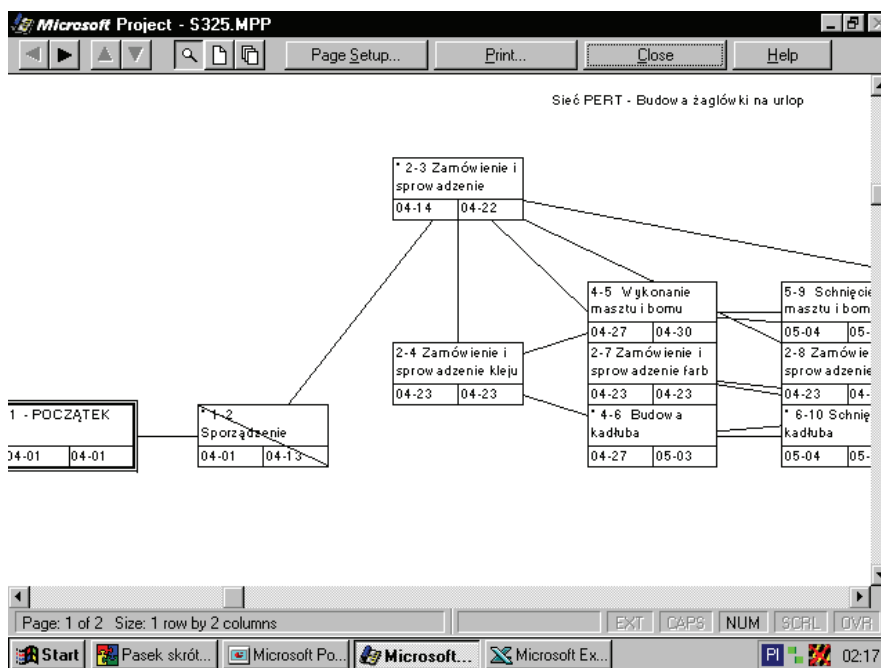
Sformułowanie przedsięwzięcia według wymagań programu Project

ID	Task Name	Duration	Start	Finish	Predecessor	Resource Names
1	1 – początek	0 d	Thu 04-04-01	Thu 04-04-01		
2	1-2: sporządzenie planów i wykonanie szablonów	8 d	Thu 04-04-01	Tue 04-04-13	1	Ojciec; matka
3	2-3: zamówienie i sprowadzenie drewna	7 d	Wed 04-04-14	Thu 04-04-22	2	Syn
4	2-4: zamówienie i sprowadzenie kleju żywicznego i gwoździ	1 d	Fri 04-04-23	Fri 04-04-23	3	Syn
5	2-7: zamówienie i sprowadzenie farb i lakierów	1 d	Fri 04-04-23	Fri 04-04-23	3	Syn
6	2-8: zamówienie i sprowadzenie płótna, nici i linek	2 d	Fri 04-04-23	Mon 04-04-26	3	Syn
7	2-11: zamówienie i sprowadzenie okuć i stalówek	4 d	Tue 04-04-27	Fri 04-04-30	6	Syn
8	2-12: zamówienie i sprowadzenie wyposażenia wnętrza	5 d	Mon 04-05-03	Fri 04-05-07	7	Syn
9	3-4: przycięcie drewna	3 d	Thu 04-04-22	Tue 04-04-27	3	Ojciec; syn
10	4-5: wykonanie masztu i bomu	2 d	Tue 04-04-27	Fri 04-04-30	4;9	Ojciec
11	4-6: budowa kadłuba	4 d	Tue 04-04-27	Mon 04-05-03	9;4	Ojciec; syn
12	5-9: schnięcie masztu i bomu	1 d	Tue 04-05-04	Tue 04-05-04	10	
13	6-10: schnięcie kadłuba	2 d	Tue 04-05-04	Wed 04-05-05	11	
14	7-9: ślepe	0 d	Fri 04-04-23	Fri 04-04-23	5	
15	7-10: ślepe	0 d	Fri 04-04-23	Fri 04-04-23	5	
16	8-12: skrojenie i uszycie żagli	3 d	Tue 04-04-27	Thu 04-04-29	6	Matka
17	9-11: lakierowanie masztu i bomu	1 d	Wed 04-05-05	Wed 04-05-05	14;12	Matka
18	10-11: uszczelnianie i malowanie kadłuba	1,13 d	Wed 04-05-05	Fri 04-05-07	15;13	Ojciec
19	11-12: wyposażenie pokładu	2 d	Mon 04-05-10	Thu 04-05-13	17;18;7	Ojciec; matka
20	12-13: wyposażenie wnętrza	2 d	Fri 04-05-14	Mon 04-05-17	8;16;19	Matka
21	12-14: trymowanie żagli	0,13 d	Fri 04-05-14	Fri 04-05-14	8;16;19	Ojciec
22	13-14: ślepe	0 d	Mon 04-05-17	Mon 04-05-17	20	
23	14-15: wykończenie i sprawdzenie łodzi	1 d	Tue 04-05-18	Wed 04-05-19	22;21	Ojciec; matka; syn
24	Koniec	0 d	Wed 04-05-19	Wed 04-05-19	23	

W tabeli terminy rozpoczęcia oraz zakończenia danej czynności poprzedzone są skrótem dnia tygodnia w języku angielskim, następnie: rok, miesiąc, dzień.

Źródło: Opracowanie własne z zastosowaniem programu Project.

Na podstawie sformułowanego przedsięwzięcia program Project buduje sieć zależności czynności w układzie czasu kalendarzowego. W zamieszczonym fragmencie sieci określono nazwy czynności oraz dni rozpoczęcia i zakończenia. Umożliwiono także pokazanie ścieżki krytycznej.



Źródło: Opracowanie własne z zastosowaniem programu Project.

Wykres 2.12.6. Fragment sieci zależności przykładowego przedsięwzięcia „Budowa żagłówki na urlop”

2.12.6. Zadania różne

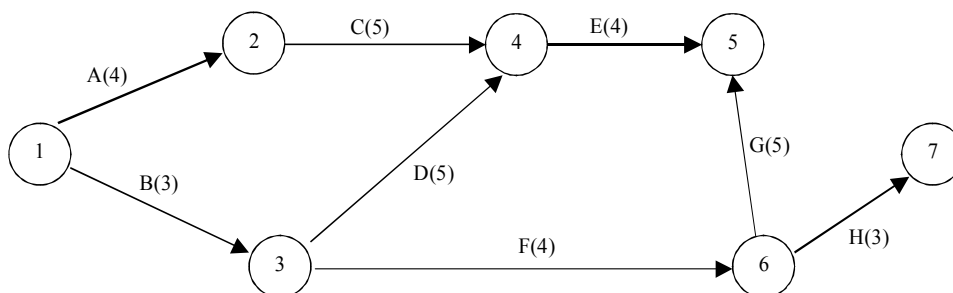
1. Zbuduj sieć zależności metodą ścieżki krytycznej CPM, mając daną charakterystykę tabelaryczną oraz zainicjowaną sieć (wykres 2.12.7).

Tabela 2.12.14

Sformułowanie problemu

Nr (rozpoczęcia, zakończenia)	Nazwa czynności	Poprzedzanie danej czynności	Czas realizacji (deterministyczny)
1-2	A		4
1-3	B		3
2-4	C	A	5
3-4	D	B	5
4-5	E	C, D	4
3-6	F	B	4
5-6	G	E	5
6-7	H	F, G	3

Źródło: Opracowanie własne, dane umowne.



Źródło: Opracowanie własne.

Wykres 2.12.7. Zainicjowanie sieci czynności

Wyznacz ścieżkę krytyczną oraz parametry (czas najwcześniejszy, czas najpóźniejszy, zapas czasu), mając zainicjowaną sieć.

2. Dana jest charakterystyka przedsięwzięcia (projektu) z danymi czasami i kosztami normalnymi oraz granicznymi. Sformułuj model programowania liniowego w celu minimalizacji całkowitego kosztu bezpośredniego projektu.

Tabela 2.12.15

Czasy realizacji czynności

Czynność/nazwa	Początek	Koniec	Czas trwania		Koszty	
	<i>i</i>	<i>j</i>	normalny	graniczny	normalny	graniczny
A	1	2	4	2	10	21
B	1	3	6	4	25	50
C	1	4	5	3	32	48
D	2	3	1	1	20	20
E	4	3	2	1	15	30
F	2	5	7	4	30	58
G	3	5	5	3	45	50
H	3	6	4	3	38	40
I	4	6	5	4	42	54
J	6	5	1	1	30	30
K	5	7	6	3	50	60
L	6	7	8	5	87	98

Źródło: Opracowanie własne, dane umowne.

Całkowity koszt projektu, czyli funkcja celu: $\sum(K_{ij} + S_{ij} X_{ij}) \rightarrow \min$.

Założenie: Koszt stanowi liniową funkcję X_{ij} – czasu trwania czynności między zdarzeniami $i-j$ (początku i końca czynności). Ponieważ K_{ij} jest stała, jak wiadomo z podrozdziału 2.12.3, to dla optymalizacji istotne jest tylko $S_{ij} X_{ij}$, gdzie S_{ij} oznacza nachylenie linii kosztu czynności. Możemy rozpatrywać zagadnienie jako maksymalizację ($-S_{ij} X_{ij}$), przy czym:

$$S_{ij} = (C_{Dij} - c_{dij}) / (D_{ij} - d_{ij}).$$

Bazując na tym wzorze obliczamy nachylenie – współczynniki – (zob. tab. 2.12.16) i wstawiamy do funkcji celu:

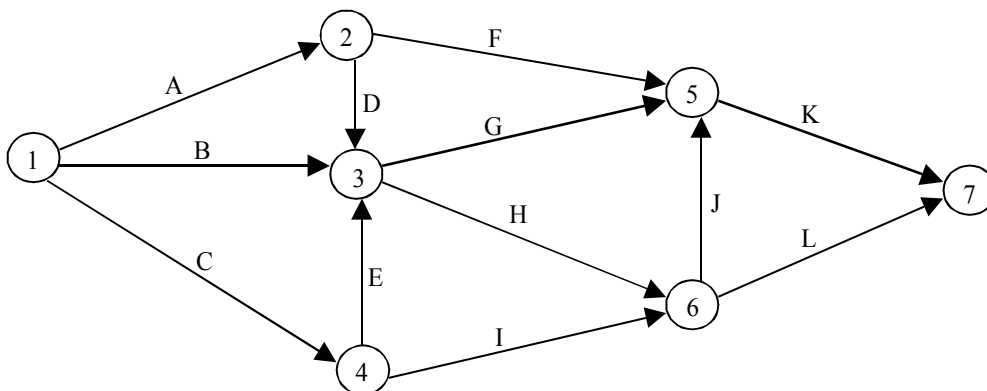
$$FC: 5,5X_{12} + 12,5X_{13} + 8X_{14} + 0X_{23} + 15X_{43} + 9,33X_{25} + 2,5X_{35} + 2X_{36} + 12X_{46} + 0X_{65} + 3,33X_{57} + 3,66X_{67} \rightarrow \max.$$

Warunki ograniczające (według oznaczeń w podrozdziale 2.12.3):

$x_{ij} \geq d_{ij}$, $x_{ij} \leq D_{ij}$, $y_i + x_{ij} - y_j = 0$, $y_1 = 0$, $y_n \leq T$ (czas realizacji całego projektu).

Warunki brzegowe: $x_{ij} \geq 0$, $y_k \geq 0$ (dla zdarzeń kończących czynności).

Czytelnikowi pozostawia się dokończenie zainicjowanej sieci poprzez wpisanie czynności, czasów trwania i określenie ścieżki krytycznej.



Źródło: Opracowanie własne.

Wykres 2.12.8. Zainicjowana sieć czynności

Nachylenie linii kosztu ($-S_{ij}$) czynności między zdarzeniami $i-j$ określono na podstawie kosztów i czasów granicznych dla poszczególnych czynności.

Tabela 2.12.16

Nachylenie oraz oczekiwany czas realizacji

Zdarzenia	Nazwa czynności	$-S_{ij}$	X_{ij}
1-2	A	5,5	4
1-3	B	12,5	5
1-4	C	8	4
2-3	D	0	1
4-3	E	15	1
2-5	F	9,33	6
3-5	G	2,5	5
3-6	H	2	4
4-6	I	12	5
6-5	J	0	0
5-7	K	3,33	6
6-7	L	3,66	7

Źródło: Opracowanie własne w Excelu; wartość X_{ij} według obliczenia w module (PERT-CPM).

Sprawdzenie w Excelu poprawności wyników obliczeń w kolumnach ($-S_{ij}$) oraz zdefiniowanie zadania decyzyjnego w module programowania liniowego (LP) programu WinQSB, a następnie wykonanie raportu sumarycznego z wynikami X_{ij} pozostawia się również do wykonania Czytelnikowi.

3. Oblicz oczekiwane czasy realizacji T_e i wykreśl ścieżkę krytyczną, mając daną sieć czynności w formie tabelarycznej.

Tabela 2.12.17

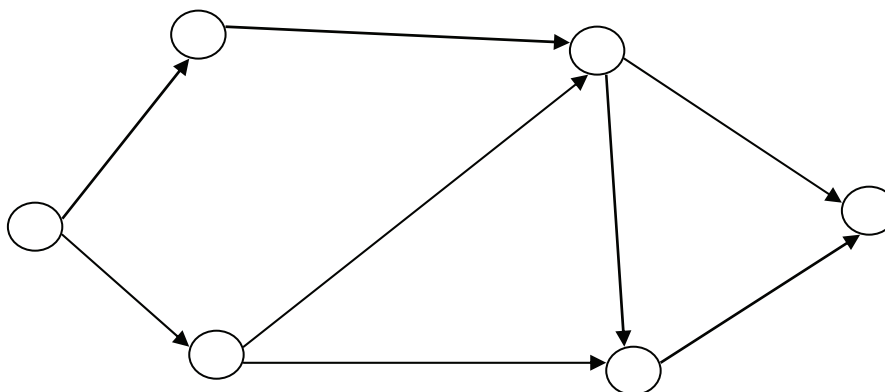
Czasy realizacji przedsięwzięcia

Nr	Czynność	Poprzedzenie	a	m	b
1	A		1	2	5
2	B		1	2	4
3	C	A	3	3	5
4	D	B	1	3	3
5	E	B	1	2	8
6	F	C, D	1	2	4
7	G	C, D	3	3	3
8	H	E, F	1	2	3

a – czas optymistyczny,
 m – czas najbardziej prawdopodobny,
 b – czas pesymistyczny.

Źródło: Opracowanie własne w Excelu; dane umowne.

4. Należy dokończyć opis sieci zależności na podstawie czasów realizacji z tab. 2.12.18.



Źródło: Opracowanie własne.

Wykres 2.12.9. Zainicjowana sieć przedsięwzięcia

Tabela 2.12.18

Czasy realizacji „t”

Zdarzenia	Nazwa czynności	t
1-2	A	14/6
1-3	B	13/6
2-4	C	20/6
3-4	D	16/6
3-5	E	17/6
4-5	F	13/6
4-6	G	3
5-6	H	2

Źródło: Opracowanie własne, dane umowne.

2.12.7. Pytania

1. Opisz zasady budowy sieci przedsięwzięć metodą sieci krytycznej CPM.
2. Przedstaw wersję kosztową metody CPM.
3. Omów metodę PERT z szczególnym uwzględnieniem losowości trwania czynności.

2.12.8. Zadania

1. Zbuduj sieć zależności metodą ścieżki krytycznej CPM, mając daną charakterystykę tabelaryczną.

Tabela 2.12.19

Charakterystyka sieci

Nr (rozpoczęcia, zakończenia)	Nazwa czynności	Poprzedzanie danej czynności	Czas realizacji (deterministyczny)
1-2	A		5
1-3	B		6
2-4	C	A	4
3-5	D	B	5
4-5	E	C, D	8
5-6	F	D, E	4

Źródło: Opracowanie własne, dane umowne.

Wyznacz ścieżkę krytyczną oraz parametry (czas najwcześniejszy, czas najpóźniejszy, zapas czasu).

2. Dana jest charakterystyka projektu z czasami i kosztami normalnymi oraz granicznymi. Sformułuj model programowania liniowego w celu minimalizacji całkowitego kosztu projektu, stosując moduł (LP) programu WinQSB.

Tabela 2.12.20

Charakterystyka sieci w formie tabelarycznej

Czynność/nazwa	Początek	Koniec	Czas trwania		Koszty	
	<i>i</i>	<i>j</i>	normalny	graniczny	normalny	graniczny
A	1	2	8	4	20	30
B	1	3	5	3	24	35
C	1	4	3	1	26	38
D	2	3	9	5	23	29
E	4	3	4	2	28	34
F	2	5	6	3	22	27
G	3	5	2	1	30	39
H	3	6	4	1	34	42
I	4	6	7	3	40	48
J	6	5	9	2	37	46
K	5	7	5	2	45	50
L	6	7	1	1	50	56

Źródło: Opracowanie własne, dane umowne.

3. Wykreśl ścieżkę krytyczną metodą PERT do problemu probabilistycznego, mając daną sieć czynności w formie tabelarycznej.

T a b e l a 2.12.21

Czasy realizacji dla metody PERT

Nr	Czynność	Poprzedzenie	<i>a</i>	<i>m</i>	<i>b</i>
1	A		2	3	6
2	B		1	2	4
3	C	A	4	5	7
4	D	B	3	4	6
5	E	B	5	6	8
6	F	C, D	1	2	4
7	G	C, D	3	3	3
8	H	E, F	6	8	10

Ź r ó ł o: Opracowanie własne, dane umowne.

ANEKS

Wyrażenia Excela Zastosowane wzory Tablice statystyczne

$$R_{xy} = \frac{\sum_{n=1}^N (y_n - \bar{y})(x_n - \bar{x})}{N \cdot S(y) \cdot S(x)}$$

$$M_R = \sqrt{\frac{2}{m \cdot (m-1)} \sum_{j=2}^m \sum_{l < j} r_{jl}^2}$$

$$THEIL = \frac{\sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} (\hat{Y}_n - Y_n)^2}}{\sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} \hat{Y}_n^2 + \frac{1}{1+h} \sum_{n=S}^{S+h} Y_n^2}}$$

Wyrażenia Excela

W opracowaniu dominuje zastosowanie wyrażień, zwanych też formułami, i funkcji arkusza kalkulacyjnego Excel.

W formułowaniu wyrażień stosowane są operatory matematyczne:

+ dodawanie,

– odejmowanie,

* mnożenie,

/ dzielenie,

% procent,

^ potęgowanie oraz pierwiastkowanie, np. $A2^{0,5}$

= równe,

< mniejsze,

<= mniejsze równe,

> większe,

>= większe równe,

◇ nierówne.

Działania wykonywane są na adresach komórek lub ich obszarach. Zarówno adresom jak i obszarom można nadać nazwy, i następnie posługiwać się nimi w działaniach.

Procedura nadania nazwy, posługując się funkcjami menu i podmenu, jest następująca:

– zaznaczenie komórki lub obszaru,

– (*Wstaw, Nazwa, Definiuj*),

– podanie nazwy,

– potwierdzenie klawiszem *Enter*.

Każdy zapis działania w Excelu, zwany wyrażeniem, zaczyna się od znaku „=” . Wyrażenia mogą być proste lub złożone zawierające w sobie inne działania na adresach czy też obszarach komórek arkusza kalkulacyjnego. Każdy adres obejmuje symbol literowy kolumny oraz numer wiersza arkusza, np. A5. Do pomocy w formułowaniu wyrażień zamieszczono w niniejszej części przykłady ich zastosowania stanowiące opracowanie własne w Excelu¹⁸⁵.

¹⁸⁵ *Opis funkcji Microsoft Excel*, Microsoft: Corporation, Ireland 1993.

Przykład
Pomocnicze obliczenia statystyczne

Tabela 1

Przykładowe wyrażenia statystyki opisowej

A/1	B	C	D	E	F	G	H	J	K		Sortowane	n	
2	1345	1301	1368	1322	1310	1370	1318	1350	1303				
3										minimum	1301	1	
4	10	=ILELICZB(B2:K2)										1303	2
5	10	=ILENIEPUSTYCH(B2:K2)										1310	3
6	1370	=MAX(B2:K2)										1318	4
7	1322	=MEDIANA(B2:K2)									mediana	1322	5
8	1301	=MIN(B2:K5)										1345	6
9	26,96	=ODCH.STANDARDOWE(B2:K2)										1350	7
10	726,90	=WARIANCJA(B2:K2) – wariancja z próbki									maksimum	1370	9

Źródło: Opracowanie własne w Excelu.

Przykład
Obliczanie potęg, iloczynów, silni oraz sum kolumnowych

Duże wartości danych w komórkach arkusza kalkulacyjnego Excela przedstawiane są w postaci inżynierskiej, np. 4,037912,E+06, przy czym (E+06) to 10 do potęgi szóstej. Komórki o niezmiennym adresie biorące udział w kolejnych wyrażeniach zapisywane są z dolarem np. \$B\$5.

Tabela 2

Przykłady wyrażen na potęgi, iloczyny oraz sumy (ze skopiowaniem wyrażen w wierszach następnych: w wierszu 4. przykładowe wyrażeni)

1/A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
2														
3														
4	=B5*B3	X=C5+1	Y=D5+2	Z=E5+3	=X^2	=Y^2	=Z^2	Suma kwadratów =E5+F5+H5	=C6^2+D6^2+E6^2	=C5-D5-E5	=X^2-Y^2-Z^2	=X*Y*Z	=X^2*Y^2*Z^2	=SILNIA(C6)
5	1	1	2	3										
6	2	2	4	6	4	16	36	56	56	-8	-48	48	2304	2
7	4	3	6	9	9	36	81	126	126	-12	-108	162	26244	6
8	8	4	8	12	16	64	144	224	224	-16	-192	384	147456	24
9	16	5	10	15	25	100	225	350	350	-20	-300	750	562500	120
10	32	6	12	18	36	144	324	504	504	-24	-432	1296	1679616	720
11	64	7	14	21	49	196	441	686	686	-28	-588	2058	4235364	5040
12	128	8	16	24	64	256	576	896	896	-32	-768	3072	9437184	40320
13	256	9	18	27	81	324	729	1134	1134	-36	-972	4374	19131876	362880
14	512	10	20	30	100	400	900	1400	1400	-40	-1200	6000	36000000	3628800
SUMY:	1023	55	110	165	384	1536	3456	5376	5376	-216	-4608	18144	71222544	4037912
	a)	b)	c)	d)										4,037912,E+06

- a) = \$B\$5 * (\$B\$3^10 - 1)/(\$B\$3 - 1)
- b) = 10 * (\$C\$5 + \$C\$14)/2
- c) = 10 * (\$D\$5 + \$D\$14)/2
- d) = 10 * (\$E\$5 + \$E\$14)/2

Źródło: Opracowanie własne w Excelu.

P r z y k ł a d

Formułowanie wyrażeń złożonych na przykładzie obliczania figur płaskich.

Oprócz wyrażeń arytmetycznych zastosowano funkcję PIERWIĄSTEK(...) oraz przywołano wartości PI().

Tabela 3

Wyrażenia Excela na przykładzie obliczania powierzchni figur płaskich

A	B	C	D	E	F	G	H	I	J	K	L/1	
		Wyrażenia na obliczanie figur płaskich (pola, obwody)										2
p.	Figura	Kształt	Parametry						Pole	Obwód		3
1	Kwadrat		a	b	h	c	d	r	=D5^2	=4*D5		4
				2						4	8	5
2	Prostokąt								=D7*E7	=2*D7+2*E7		6
			4	2						8	12	7
3	Trójkąt					=D10^2+F10^2 =PIERWIĄSTEK(G9)			=0,5*D10*E10	=D10+F10+G10		8
						25						9
			4		3	5				6	12	10
4	Trapez								=(D12+E12)/2*F12	=D12+E12+G12+H12		11
			5	3	2,5	3	3,5			10	14,5	12
5	Sześciokąt					=15^2-(0,5*115)^2 =PIERWIĄSTEK(F14)			=12*0,5*0,5*115*F15	=6*115		13
					18,75							14
					4,33			5	64,95	30		15
6	Koło								=PI()*117^2	=2*PI()*117		16
								5	78,54	32,42		17

Ź r ó d ł o: Opracowanie własne w Excelu.

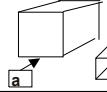
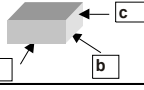
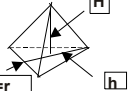
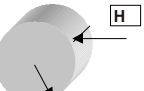
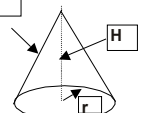
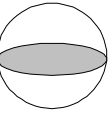
P r z y k ł a d

Zastosowanie wyrażeń złożonych na przykładzie obliczania brył

Zwróćmy uwagę na przywołanie wartości z poprzedniego arkusza tego samego skoroszytu, np. [=FIGURY!F15], oraz podnoszenie do potęgi trzeciej – [=I16^3].

Tabela 4

Przykładowe wyrażenia na obliczanie objętości oraz powierzchni brył

A	B	C	D	E	F	G	H	I	J	K	L	1		
Wyrażenia na obliczanie brył (objętości, powierzchni)												2		
Lp.	Bryła	Kształt	Parametry							Objętość	Powierzchnia	3		
1	Sześcian		a	b	h	c	H	r	l	=D5^3	=6*D5^2	4		
			3									54		
2	Prostopadłościan		a	b		c				=D7*E7*G7	=2*D7*E7+2*E7*G7+2*D7*G7	6		
			4	2		1,5						34		
3	Ostrosłup									=FIGURY! F15	=1/3*0,5*F9*H9	=4*0,5*F9	8	
			5		4,33		4,5					8,66		
4	Walec obrotowy												10	
							2,5	5					11	
5	Stożek obrotowy									=H14^2+I14^2 =PIERWIASTEK (J13)	=1/3*PI()*I14^2* H14	=PI()*I14*(I14+J1 4)	12	
									94,25				13	
							8	6	9,71	253,42	262,78	14		
6	Kula												15	
												4188,79	1256,64	16

Źródło: Opracowanie własne w Excelu.

Przykład

Zastosowanie zaokrąglenia do liczby całkowitej i badania logicznego.

Tabela 5

Przykładowe obliczanie arytmetyczne, zaokrąglenia, badanie logiczne

A1	B	C	D	E
3	brutto		1 000,00 zł	
5	netto		909,09 zł	=D3-D5
6	premia		90,91 zł	=0,1*D3
8			909 zł	= ZAOKR.DO.CAŁK(D3-D5)
9			91 zł	=0,1*D6
10			PRAWDA	=CZY.LOGICZNA(D3-(SUMA(D5+D6))=0)

Źródło: Opracowanie własne w Excelu.

Zastosowane wzory (z podaniem podrozdziałów)

(1.1.2) Ogólna postać modelu liniowego

$$\begin{aligned} X_1 &= \sum_{j \neq 1}^m b_{1j} X_j + c_1 \\ X_2 &= \sum_{j \neq 1}^m b_{2j} X_j + c_2 \\ &\dots\dots\dots \\ X_h &= \sum_{j \neq 1}^m b_{hj} X_m + c_h \end{aligned} \quad (1.1.1)$$

(1.1.4.) Funkcja liniowa zmiennej objaśnianej

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + \varepsilon \quad (1.1.2)$$

(1.1.4) Model regresji liniowej wielowymiarowej

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_k X_k \quad (1.1.3)$$

(1.1.5.1) Empiryczny (próbkowy) współczynnik korelacji

$$R_{xy} = \frac{\sum_{n=1}^N (y_n - \bar{y})(x_n - \bar{x})}{N \cdot S(y) \cdot S(x)} \quad (1.1.4)$$

(1.1.5.1) Statystyka określenia istotności współczynnika korelacji

$$t_{obl} = R_{xy} \sqrt{\frac{N-2}{1-R_{xy}^2}} \quad (1.1.5)$$

(1.1.5.2) Statystyka określenia istotności współczynnika korelacji wielorakiej

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k} \quad (1.1.6)$$

(1.1.5.2) Współczynnik determinacji

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (1.1.7)$$

(1.1.5.3) Statystyka t -Studenta określenia istotności parametrów strukturalnych

$$I_i = \frac{|a_i|}{S(a_i)} \quad (1.1.8)$$

(1.1.6.2) Standaryzacja reszt

$$u_t = \frac{e_t - \bar{e}}{\hat{S}_e} \quad (1.1.9)$$

(1.1.6.2) Odchylenie standardowe reszt

$$\hat{S}_e = \sqrt{\frac{\sum_{t=1}^n (e_t - \bar{e})^2}{n}} \quad (1.1.10)$$

(1.1.6.2) Formuła testu Jarque'a-Bery badania normalności rozkładu odchyłeń losowych

$$JB = \frac{N}{6} \left[\beta_1 + \frac{1}{4} (\beta_2 - 3)^2 \right] \quad (1.1.11)$$

(1.1.6.2) Współczynnik skośności rozkładu reszt

$$S = \sqrt{\beta_1} = \frac{1}{N} \sum_{i=1}^N \frac{e_i^3}{S^3(e_i)} \quad (1.1.12)$$

(1.1.6.2) Błąd standardowy reszt

$$S(e_i) = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (1.1.13)$$

(1.1.6.2) Współczynnik kurtozy

$$\beta_2 = \frac{1}{N} \sum_{i=1}^N \frac{e_i^4}{S^4(e_i)} \quad (1.1.14)$$

(1.1.15) Alternatywna formuła testu Jarque'a-Bery

$$JB = \frac{N-k}{6} \left[S^2 + \frac{1}{4}(K-3)^2 \right] \quad (1.1.15)$$

(1.1.6.3) Statystyka badania nieobciążoności odchyłeń losowych

$$I = \frac{|\bar{e}| \sqrt{n-1}}{\hat{S}_e} \quad (1.1.16)$$

(1.1.6.4) Współczynnik autokorelacji reszt

$$r_\tau = \frac{\sum_{t=\tau+1}^n (e_t - \bar{e}_t)(e_{t-\tau} - \bar{e}_{t-\tau})}{\sqrt{\sum_{t=\tau+1}^n (e_t - \bar{e}_t)^2 \sum_{t=\tau+1}^n (e_{t-\tau} - \bar{e}_{t-\tau})^2}} \quad (1.1.17)$$

(1.1.6.4) Współczynnik autokorelacji reszt pierwszego rzędu

$$r_1 = \frac{\sum_{t=2}^n e_t \bar{e}_{t-1}}{\sqrt{\sum_{t=2}^n e_t^2 \sum_{t=2}^n e_{t-1}^2}} \approx \frac{\sum_{t=2}^n e_t \bar{e}_{t-1}}{\sum_{t=2}^n e_t^2} \quad (1.1.18)$$

(1.1.6.4) Statystyka testu Durбина-Watsona badania autokorelacji reszt

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})}{\sum_{t=1}^n e_t^2} \quad (1.1.19)$$

(1.1.6.4) Statystyka testu istotności współczynnika autokorelacji dowolnego rzędu

$$I_\tau = \frac{|r_\tau| \sqrt{n-\tau-2}}{\sqrt{1-r_\tau^2}} \quad (1.1.20)$$

(1.1.6.5) Statystyka testu F stałości wariancji

$$F = \frac{S_{e,2}^2}{S_{e,1}^2} \quad (1.1.21)$$

(1.1.6.5) Wariancja resztowa n_1 pierwszych obserwacji

$$S_{e,1}^2 = \frac{\sum_{t=1}^n (e_t - \bar{e}_1)^2}{n_1 - k - 1} \quad (1.1.22)$$

(1.3.2.1) Średnia arytmetyczna zmiennej X

$$\bar{x} = \frac{\sum_{i=1}^K x_i}{n} \quad (1.3.1)$$

(1.3.2.1) Średnia arytmetyczna ważona zmiennej X

$$\bar{x} = \frac{\sum_{i=1}^K n_i \cdot x_i}{n} \quad (1.3.2)$$

(1.3.2.1) Średnia harmoniczna zmiennej X

$$H_x = \frac{n}{\sum_{i=1}^n (1/x_i)} \quad (1.3.3)$$

(1.3.2.1) Średnia harmoniczna ważona zmiennej X

$$H_x = \frac{n}{\sum_{i=1}^K (n_i/x_i)} \quad (1.3.4)$$

(1.3.2.1) Średnia geometryczna zmiennej X szeregu szczegółowego

$$G_x = \sqrt[n]{\prod_{i=1}^n x_i} \quad (1.3.5)$$

(1.3.2.1) Średnia geometryczna zmiennej X szeregu rozdzielczego

$$G_x = \sqrt[n]{\dot{x}_1^{n_1} \cdot \dot{x}_2^{n_2} \dots \dot{x}_k^{n_k}} \quad (1.3.6)$$

(1.3.2.1) Średnia geometryczna iloczynu dwóch zmiennych

$$G_{xy} = G \cdot G_y \quad (1.3.7)$$

(1.3.2.2) Mediana parzystej liczby obserwacji

$$M_x = \frac{x_{n/2} + x_{n/2+1}}{2} \quad (1.3.8)$$

(1.3.2.2) Mediana szeregu rozdzielczego

$$M_x = x_0 + h \cdot \frac{(n+1)/2 - n_s}{n_0} \quad (1.3.9)$$

(1.3.2.2) Kwartyl szeregu rozdzielczego

$$Q_x = x_0 + h \cdot \frac{n/4 - n_s}{n_0} \quad (1.3.10)$$

(1.3.3.2) Przeciętna zmiennej

$$D_x = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (1.3.11)$$

(1.3.3.2) Przeciętna zmiennej szeregu rozdzielczego

$$D_x = \frac{\sum_{i=1}^K |\dot{x}_i - \bar{x}| \cdot n_i}{n} \quad (1.3.12)$$

(1.3.3.2) Wariancja populacji

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (1.3.13)$$

(1.3.3.2) Wariancja populacji szeregu rozdzielczego

$$S_x^2 = \frac{\sum_{i=1}^K (\dot{x}_i - \bar{x})^2 \cdot n_i}{n} \quad (1.3.14)$$

(1.3.3.3) Współczynnik zmienności

$$V_x = \frac{S_x}{\bar{x}} \quad (1.3.15)$$

(1.3.3.3) Współczynnik zmienności względem miar pozycyjnych

$$V_x = \frac{Q_x}{M_x} \quad (1.3.16)$$

(1.3.4) Klasyczny współczynnik asymetrii

$$A_s = \frac{(1/n) \cdot \sum_{i=1}^n (x_i - \bar{x})^3}{S_x^3} \quad (1.3.17)$$

(1.3.4) Pozycyjny współczynnik asymetrii

$$A_Q = \frac{Q_3 - 2 \cdot M_x + Q_1}{Q_3 - Q_1} \quad (1.3.18)$$

(1.4.1) Kowariancja między zmiennymi X_j, X_l

$$\text{cov}_{jl} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l) \quad (1.4.1)$$

(1.4.1) Współczynnik korelacji między zmiennymi X_j, X_l

$$r_{jl} = \frac{\text{cov}_{jl}}{\sqrt{\text{var}_{jj} \cdot \text{var}_{ll}}} \quad (1.4.2)$$

(1.4.1) Syntetyczna miara współzależności liniowej

$$M_R = \sqrt{\frac{2}{m \cdot (m-1)} \sum_{j=2}^m \sum_{l < j} r_{jl}^2} \quad (1.4.3)$$

(1.4.2.1) Standaryzacja zmiennej X_j dla i -tej obserwacji

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1.4.4)$$

(1.4.2.1) Unitaryzacja zmiennej X_j dla i -tej obserwacji

$$z_{ij} = \frac{x_{ij} - \min x_{ij}}{O_j} \quad (1.4.5)$$

(1.4.2.2) Średnia arytmetyczna zmiennej zunitaryzowanej

$$\bar{z}_j = \frac{\bar{x}_j - \min x_{ij}}{O_j} \quad (1.4.6)$$

(1.4.2.3) Odległość euklidesowa między obserwacjami

$$d_{il} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{lj})^2} \quad (1.4.7)$$

(1.4.2.3) Odległość miejska między obserwacjami

$$d_{il} = \sum_{j=1}^m |z_{ij} - z_{lj}| \quad (1.4.8)$$

(1.4.2.4) Określenie składowej wektora miary położenia

$$v_{kj} = \frac{1}{n_k} \sum_{i \in C_k} z_{ij} \quad (1.4.9)$$

(1.4.2.4) Odległość krytyczna wiązań grafu spójnego

$$d^* = \bar{d} + k \cdot s_d \quad (1.4.10)$$

(1.4.2.5) Odległość euklidesowa obserwacji od wzorca rozwoju

$$d_{i0} = \sqrt{\sum_{j=1}^m (z_{ij} - z_{0j})^2} \quad (1.4.11)$$

(1.4.2.5) Miara rozwoju obserwacji według metody wzorca rozwoju

$$m_i = 1 - \frac{d_{i0}}{d_0} \quad (1.4.12)$$

(1.4.2.5) Miara rozwoju według metody sum standaryzowanych

$$m_i = \frac{p_i - p_0}{p_0 - p_{-0}} \quad (1.4.13)$$

(1.5.1) Oszacowanie parametru a modelu regresji liniowej $\hat{Y} = aX + b$

$$a = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (1.5.1)$$

(1.5.2) Oszacowanie parametru b modelu regresji liniowej $\hat{Y} = aX + b$

$$b = \bar{y} - a\bar{x} \quad (1.5.2)$$

(1.5.2) Błąd standardowy wartości Y przy zmiennej objaśniającej X

$$S_{y-x} = \sqrt{\left[\frac{1}{n(n-2)} \right] \left[n\sum y^2 - (\sum y)^2 - \frac{[n\sum xy - (\sum x)(\sum y)]^2}{n\sum x^2 - (\sum x)^2} \right]} \quad (1.5.3)$$

(1.5.3) Wariancja próbki

$$S^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{X})^2}{n-1} \quad (1.5.4)$$

(1.5.6.1) Trend liniowy

$$\hat{Y} = a_0 + a_1 t \quad (1.5.5)$$

(1.5.6.3) Trend potęgowy

$$\hat{Y} = a_0 t^{a_1} \quad (1.5.6)$$

(1.5.6.4) Trend wykładniczy

$$\hat{Y} = a_0 a_1^t \quad (1.5.7)$$

(1.5.6.5) Trend logarytmiczny

$$\hat{Y} = a_0 + a_1 \ln t \quad (1.5.8)$$

(1.5.6.6) Trend hiperboliczny – postać 1.

$$\hat{Y} = a_0 + \frac{a_1}{t} \quad (1.5.9)$$

(1.5.6.6) Trend hiperboliczny – postać 2.

$$\hat{Y} = \frac{a_0 \cdot t}{t + a_1} \quad (1.5.10)$$

(1.5.6.7) Trend logistyczny

$$\hat{Y} = \frac{a_0}{1 + a_1 \cdot e^{-t}} \quad (1.5.11)$$

(1.6.2) Indywidualna pojemność nośnika informacji

$$h^1_{Li} = \frac{(r^1_i)^2}{1 + \sum_{j \neq i} |r^1_{ij}|} \quad (1.6.1)$$

(1.6.2) Pojemność integralna kombinacji

$$H^1_L = \sum_{i=1}^{m_L} h^1_{Li} \quad (1.6.2)$$

(1.6.3) Bezwzględny błąd prognozy

$$v_T = \left[\frac{(T - \bar{t})^2}{\sum_{t=1}^n (t - \bar{t})^2} + \frac{1}{n} + 1 \right]^{0.5} \cdot S_e \quad (1.6.3)$$

(1.6.3) Odchylenie standardowe reszt do obliczania bezwzględnego błędu prognozy

$$S_e = \left[\frac{1}{n - m - 1} \sum_{t=1}^n (y_t - \hat{y})^2 \right]^{0.5} \quad (1.6.4)$$

(1.6.3) Średni procentowy względny błąd prognozy

$$\eta_T = \frac{y_T}{\hat{y}_T} \cdot 100 \quad (1.6.5)$$

(1.6.4) Współczynnik korelacji wielokrotnej

$$R = R_k = \sqrt{1 - \frac{\det R^*}{\det R}} \quad (1.6.6)$$

(1.6.6) Współczynnik zmienności oparty na rozstępie

$$W_j = 2 \cdot \frac{\max_i x_{ij} - \min_i x_{ij}}{\max_i x_{ij} + \min_i x_{ij}} \quad (1.6.7)$$

(1.6.8) Statystyka istotności współczynnika korelacji

$$t = \frac{|r_{xy}|}{\sqrt{1 - r_{xy}^2}} \cdot \sqrt{n - 2} \quad (1.6.8)$$

(1.6.9) Krytyczna wartość współczynnika korelacji

$$r^* = \sqrt{\frac{t_\alpha^2}{t_\alpha^2 + n - 2}} \quad (1.6.9)$$

(1.6.7) Test Akaike'a

$$AK = n \ln(\Sigma z_n^2) + 2k \quad (1.6.10)$$

(1.6.7) Test Schwarza

$$SCH = n \ln(\Sigma z_n^2) + k \ln(n) \quad (1.6.11)$$

(1.6.10) Współczynnik korelacji między zmienną objaśnianą a zmienną objaśniającą

$$r_{xy} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1.6.12)$$

(1.7.1) Równanie macierzowe oszacowania parametrów regresji liniowej wielowymiarowej

$$\mathbf{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (1.7.1)$$

(1.7.1) Macierz wariancji i kowariancji

$$\mathbf{D}^2(a) = S_e^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (1.7.2)$$

(1.7.1) Współczynnik zbieżności

$$\Phi^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1.7.3)$$

(1.7.1) Współczynnik wyrazistości

$$W = (S_e / \bar{y}) \cdot 100 \quad (1.7.4)$$

(1.7.1) Wskaźnik średniego poziomu reszt

$$p = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{\hat{y}_i} \quad (1.7.5)$$

(1.7.2) Współczynnik kurtozy zmiennej X

$$K = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_j - \bar{x}}{S} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (1.7.6)$$

(1.7.2) Współczynnik skośności zmiennej X

$$S_k = \frac{n(n+1)}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{S} \right)^3 \quad (1.7.7)$$

(1.7.2) Skorygowany (dopasowany) współczynnik determinacji

$$\bar{R}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (1.7.8)$$

(1.7.2) Estymowana suma kwadratów

$$ESS = \sum_{i=1}^n (\hat{y}_i - \hat{y})^2 \quad (1.7.9)$$

(1.7.2) Resztowa suma kwadratów

$$RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2 \quad (1.7.10)$$

(1.7.2) Standaryzacja reszt

$$z_i = \frac{(e_i - \bar{e})}{S_{ei}} \quad (1.7.11)$$

(1.7.3) Test Kołomogorowa o statystyce maksymalnej różnicy

$$D_n = \sup_x |F_n(x) - F_0(x)| \quad (1.7.12)$$

(1.7.3) Wartość obliczeniowa statystyki

$$\lambda = \sqrt{n}D_n \quad (1.7.13)$$

(1.7.3) Macierz rzutowania

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \quad (1.7.14)$$

(1.7.3) Kwadrat odległości Mahalonobisa

$$MD_i^2 = (n-1) \left(h_i - \frac{1}{n} \right) \quad (1.7.15)$$

(1.8.2) Wariancja resztowa – wersja macierzowa

$$S^2(e) = \frac{\mathbf{y}^T \mathbf{y} - \mathbf{a}^T \mathbf{X}^T \mathbf{y}}{n - m - 1} \quad (1.8.1)$$

(1.10) Współczynnik autokorelacji zmiennej

$$\rho_k = \frac{\sum_{t=1+k}^n (Y_t - \bar{Y}) \cdot (Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (1.10.1)$$

(1.10) Funkcja autokorelacji cząstkowej

$$\rho_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \rho_{kk-1} \cdot \rho_{k-1}}{1 - \sum_{j=1}^{k-1} \rho_{kk-1} \rho_j} \quad (1.10.2)$$

(1.11.1) Wektor ocen parametrów strukturalnych według UMNG

$$\mathbf{a} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} \quad (1.11.1)$$

(1.11.1) Wariancja odchyłeń losowych w UMNK szacowana resztami

$$S_e^2 = \frac{1}{n-k-1} \mathbf{e}^T \mathbf{V}^{-1} \mathbf{e} \quad (1.11.2)$$

(1.11.1) Kowariancja przy zastosowaniu UMNK

$$\mathbf{D}^2(a) = S_e^2 (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \quad (1.11.3)$$

(1.11.2) Liniowy model prawdopodobieństwa (LPM)

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \quad (1.11.4)$$

(1.11.3) Logit

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_1 + \beta_2 X_i \quad (1.11.5)$$

(1.11.4) Probit

$$\text{Probit}_i = F^{-1} P_i = \beta_1 + \beta_2 X_i + \varepsilon_i \quad (1.11.6)$$

(1.11.4) Zależność między logitem a probitem

$$\text{Logit} / \text{Probit} = \frac{\pi}{\sqrt{3}} = 1,8 \quad (1.11.7)$$

(1.11.4) Oszacowanie parametrów strukturalnych modelu probitowego

$$\mathbf{a} = (\mathbf{X}^T \cdot \mathbf{V}^{-1} \mathbf{X})^{-1} \cdot \mathbf{X}^T \mathbf{V}^{-1} \cdot \text{Pr} \quad (1.11.8)$$

(1.11.4) Oszacowanie parametrów strukturalnych modelu logitowego

$$\mathbf{a} = (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \cdot \mathbf{X}^T \mathbf{W}^{-1} \mathbf{L} \quad (1.11.9)$$

(1.12) Pierwiastek średniego kwadratu błędu

$$RMSE = \sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} (\hat{Y}_n - Y_n)^2} \quad (1.12.1)$$

(1.12) Średni absolutny błąd

$$MAE = \frac{1}{1+h} \sum_{n=S}^{S+h} |\hat{Y}_n - Y_n| \quad (1.12.2)$$

(1.12) Średni procentowy absolutny błąd

$$MAPE = \frac{1}{1+h} \sum_{n=S}^{S+h} \left| \frac{\hat{Y}_n - Y_n}{Y_n} \right| \cdot 100 \quad (1.12.3)$$

(1.12) Współczynnik Thiela

$$THEIL = \frac{\sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} (\hat{Y}_n - Y_n)^2}}{\sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} \hat{Y}_n^2} + \sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} Y_n^2}} \quad (1.12.4)$$

(1.12.1) Współczynnik Janusowy

$$J^2 = \frac{\frac{1}{T-n} \sum_{t=n+1}^T (y_t - y_{tP})^2}{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (1.12.5)$$

(1.12.1) Skorygowany współczynnik determinacji, gdy znamy R^2

$$\bar{R}^2 = 1 - \frac{n-1}{n-k} \cdot (1 - R^2) \quad (1.12.6)$$

(1.12.9.3) Prognoza w modelu prostego wyrównywania wykładniczego

$$F_n = \alpha \cdot Y_{n-1} + (1-\alpha) \cdot F_{n-1} \quad (1.12.7)$$

(1.12.9.3) Zależność między liczbą okresów w a parametrem wyrównywania wykładniczego α

$$w = (2/\alpha - 1) \quad (1.12.8)$$

(1.12.9.4) Sygnał adaptacyjny

$$TST_n = \left| \frac{SAD_n}{MAD_n} \right| \quad (1.12.9)$$

(1.12.9.4) Prognoza w modelu ARRES

$$F_n = F_{n-1} + TST_{n-1} \cdot (Y_{n-1} - F_{n-1}) \quad (1.12.10)$$

(1.12.9.5) Oszacowanie trendu w modelu podwójnych średnich ruchomych

$$trend = \frac{2}{W-1} (MA(W)_n - MA(MxW)_n) \quad (1.12.11)$$

(1.12.9.5) Prognoza w modelu podwójnej średniej ruchomej

$$F_{n+m} = a_n + b_n \cdot m \quad (1.12.12)$$

(1.13.1) Prognoza w modelu multiplikatywnym Holta-Wintersa

$$F_{n+m} = (S_n + b_n \cdot m) \cdot I_{n-L+1} \quad (1.13.1)$$

(1.13.1) Prognoza w modelu addytywnym Holta-Wintersa

$$F_{n+m} = (S_n + b_n \cdot m) + I_{n-L+1} \quad (1.13.2)$$

(2.2) Zapis macierzowy równań Leontiewa

$$\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{y} \quad (2.2.1)$$

(2.4) Kryterium simpleksowe

$$k_j = c_j - \left(\sum c_{bi} a_{ij} \right) \quad (2.4.1)$$

(2.5) Wielkość pomocnicza do obliczenia kryterium simpleksowego

$$z_j = \sum_{i=0}^m a_{ij} c_{bi} \quad (2.5.1)$$

(2.9) Funkcja Lagrange'a

$$L(x, \lambda) = f(x) + \sum_{i=1}^r \lambda_i g_i(x) \quad (2.9.1)$$

(2.12) Oczekiwany czas trwania czynności

$$T_e = (a + 4m + b) / 6 \quad (2.12.1)$$

(2.12) Wariancja czasu oczekiwanego w metodzie PERT

$$\sigma_{i-j}^2 = \left(\frac{b-a}{6} \right)^2 \quad (2.12.2)$$

Tablice statystyczne

I	Wartości krytyczne $t(P, n)$ rozkładu t Studenta
II	Wartości krytyczne $\chi^2(\alpha, r)$ rozkładu χ^2
III	Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$
IV	Wartości krytyczne rozkładu Durbina-Watsona dla $\alpha = 0$
V	Wartości krytyczne $IF(\alpha, r_1, r_2)$ rozkładu F
VI	Dystrybuanta rozkładu normalnego
VII	Test Hellwiga

Źródła:

Tablice: I, II, III, V, VI zob.: R. Zieliński, *Tablice statystyczne*, PWN, Warszawa 1972; IV: M. Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003, s. 219; VII: J. Dziechciarz (red.), *Ekonometria. Metody, przykłady, zadania*, Akademia Ekonomiczna, Wrocław 2003, s. 342.

Wartości krytyczne $t(P, n)$ rozkładu t Studenta
 $P\{|t_n| > t(P, n)\} = P$

Tablica I

n/P	0,5	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	1,0000	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	318,3088	636,6192
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	22,3271	31,5991
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	10,2145	12,9240
4	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	7,1732	8,6103
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	5,8934	6,8688
6	0,7176	1,5398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588
7	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	4,7853	5,4079
8	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	4,5008	5,0413
9	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,2969	4,7809
10	0,6998	1,3732	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5869
11	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370
12	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	3,9296	4,3178
13	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	3,8520	4,2208
14	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	3,7884	4,1405
15	0,6912	1,3406	1,7530	2,1314	2,6025	2,9457	3,2860	3,7328	4,0728
16	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150
17	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,6458	3,9651
18	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,6105	3,9216
19	0,6876	1,3277	1,7291	2,0930	2,5395	2,9609	3,1737	3,5794	3,8834
20	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,5518	3,8495
21	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5272	3,8193
22	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,5050	3,7921
23	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676
24	0,6848	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,4668	3,7455
25	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251
26	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7066
27	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6896
28	0,6834	1,3125	1,7911	2,0484	2,4671	2,7633	3,0469	3,4082	3,6739
29	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,3962	3,6594
30	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460
32	0,6822	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,3653	3,6218
34	0,6818	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,3479	3,6007
36	0,6814	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,3326	3,5821
38	0,6810	1,3042	1,6890	2,0244	2,4286	2,7116	2,9803	3,3190	3,5657
40	0,6807	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,3069	3,5510
42	0,6804	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,2960	3,5377
44	0,6801	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,2861	3,5258
46	0,6799	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,2771	3,5150
48	0,6796	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,2689	3,5051
50	0,6794	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,2614	3,4960
55	0,6790	1,2971	1,6730	2,0040	2,3961	2,6682	2,9247	3,2561	3,4764
60	0,6786	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,2317	3,4602
65	0,6783	1,2947	1,6686	1,9971	2,3851	2,6535	2,9060	3,2204	3,4466
70	0,6780	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	3,2108	3,4350
80	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,1953	3,4163
90	0,6772	1,2910	1,6620	1,9867	2,3685	2,6316	2,8779	3,1833	3,4019
100	0,6770	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,1737	3,3905
120	0,6765	1,2886	1,6577	1,9799	2,3578	2,6174	2,8599	3,1595	3,3735
150	0,6761	1,2872	1,6551	1,9759	2,3515	2,6090	2,8492	3,1455	3,3566
200	0,6757	1,2858	1,6525	1,9719	2,3451	2,6006	2,8385	3,1315	3,3398
250	0,6755	1,2849	1,6510	1,9695	2,3414	2,5956	2,8322	3,1232	3,3299
300	0,6753	1,2844	1,6499	1,9679	2,3388	2,5923	2,8279	3,1176	3,3233
400	0,6751	1,2837	1,6487	1,9659	2,3357	2,5882	2,8227	3,1107	3,3150
500	0,6750	1,2832	1,6479	1,9647	2,3338	2,5877	2,8195	3,1066	3,3101
∞	0,6745	1,2816	1,6449	1,9600	2,3263	2,5758	2,8070	3,0902	3,2905

Wartości krytyczne χ^2 rozkładu χ^2 ; $P\{\chi^2_r \geq \chi^2(\alpha, r)\} = \alpha$

Tablica II

r/α	0,9995	0,999	0,995	0,99	0,975	0,95	0,90	0,80	0,70	0,60
1	0,0 ⁶ 393	0,0 ³ 157	0,0 ⁴ 393	0,0 ⁴ 157	0,0 ⁵ 982	0,0 ² 393	0,0158	0,0642	0,148	0,275
2	0,0 ² 100	0,0 ² 200	0,0100	0,0201	0,0506	0,103	0,211	0,446	0,713	1,022
3	0,0153	0,0243	0,0717	0,115	0,216	0,352	0,584	1,005	1,424	1,869
4	0,0639	0,0908	0,207	0,297	0,484	0,711	1,064	1,649	2,195	2,753
5	0,158	0,210	0,412	0,554	0,831	1,145	1,610	2,343	3,000	3,655
6	0,299	0,381	0,676	0,872	1,237	1,635	2,204	3,070	3,828	4,570
7	0,485	0,598	0,989	1,239	1,690	2,167	2,833	3,822	4,671	5,493
8	0,710	0,857	1,344	1,645	2,180	2,773	3,490	4,594	5,527	6,423
9	0,972	1,153	1,735	2,088	2,700	3,325	4,168	5,380	6,393	7,357
10	1,265	1,479	2,156	2,558	3,247	3,940	4,865	6,179	7,267	8,295
11	1,587	1,834	2,603	3,053	3,816	4,575	5,578	6,989	8,148	9,237
12	1,934	2,214	3,074	3,571	4,404	5,226	6,304	7,807	9,034	10,182
13	2,305	2,617	3,565	4,107	5,009	5,892	7,042	8,634	9,926	11,129
14	2,697	3,041	4,075	4,660	5,629	6,571	7,790	9,467	10,821	12,079
15	3,108	3,483	4,601	5,229	6,262	7,261	8,547	10,307	11,721	13,030
16	3,536	3,942	5,142	5,812	6,908	7,962	9,312	11,152	12,624	13,983
17	3,980	4,416	5,697	6,408	7,564	8,672	10,085	12,002	13,531	14,937
18	4,439	4,905	6,265	7,015	8,231	9,390	10,865	12,857	14,440	15,893
19	4,912	5,407	6,844	7,633	8,907	10,117	11,651	13,716	15,352	16,850
20	5,398	5,921	7,434	8,260	9,591	10,851	12,443	14,578	16,266	17,809
21	5,896	6,447	8,034	8,897	10,283	11,591	13,240	15,445	17,182	18,768
22	6,404	6,983	8,643	9,542	10,982	12,338	14,041	16,314	18,101	19,729
23	6,924	7,529	9,260	10,196	11,688	13,091	14,848	17,187	19,021	20,690
24	7,453	8,085	9,886	10,856	12,401	13,845	15,659	18,062	19,943	21,652
25	7,991	8,649	10,520	11,524	13,120	14,611	16,473	18,940	20,867	22,616
26	8,538	9,222	11,160	12,198	13,844	15,379	17,292	19,820	21,792	23,579
27	9,093	9,803	11,808	12,879	14,573	16,151	18,114	20,703	22,719	24,544
28	9,656	10,391	12,461	13,565	15,308	16,928	18,939	21,588	23,647	25,509
29	10,227	10,986	13,121	14,256	16,047	17,708	19,768	22,475	24,577	26,475
30	10,804	11,588	13,787	14,953	16,971	18,493	20,599	23,364	25,508	27,442
31	11,389	12,196	14,458	15,655	17,539	19,281	21,434	24,255	26,440	28,409
32	11,979	12,811	15,134	16,362	18,291	20,072	22,271	25,148	27,373	29,376
33	12,576	13,431	15,815	17,073	19,047	20,867	23,110	26,042	28,307	30,344
34	13,176	14,057	16,501	17,789	19,806	21,664	23,952	26,938	29,242	31,313
35	13,788	14,688	17,192	18,509	20,569	22,465	24,797	27,836	30,178	32,282
36	14,401	15,324	17,887	19,233	21,336	23,269	25,643	28,735	31,115	33,252
37	15,020	15,965	18,586	19,960	22,106	24,075	26,492	29,635	32,053	34,222
38	15,644	16,611	19,289	20,691	22,878	24,884	27,343	30,537	32,992	35,192
39	16,273	17,262	19,996	21,426	23,654	25,695	28,196	31,441	33,932	36,163
40	16,906	17,916	20,707	22,164	24,433	26,509	29,051	32,345	34,872	37,134
41	17,544	18,575	21,421	22,906	25,215	27,326	29,907	33,251	35,813	38,105
42	18,186	19,238	22,138	23,650	25,909	28,144	30,765	34,157	36,775	39,077
43	18,832	19,905	22,859	24,398	26,785	28,965	31,625	35,065	37,698	40,050
44	19,482	20,576	23,584	25,148	27,575	29,787	32,487	35,974	38,641	41,022
45	20,137	21,251	24,311	25,901	28,366	30,612	33,350	36,884	39,585	41,955
46	20,794	21,929	25,041	26,657	29,160	31,439	34,215	37,795	40,529	42,968
47	21,456	22,610	25,775	27,416	29,959	32,268	35,081	38,708	41,474	43,942
48	22,121	23,295	26,511	28,177	30,755	33,098	35,949	39,621	42,420	44,915
49	22,789	23,983	27,249	28,491	31,555	33,930	36,818	40,534	43,366	45,889
50	23,461	24,674	27,991	29,707	32,357	34,764	37,689	41,449	44,313	46,864
51	24,136	25,368	28,735	30,475	33,162	35,600	38,560	42,365	45,261	47,838
52	24,814	26,065	29,481	31,246	33,968	36,437	39,433	43,281	46,209	48,813
53	25,495	26,765	30,230	32,018	34,776	37,276	40,308	44,199	47,157	49,788
54	26,179	27,468	30,981	32,793	35,586	38,116	41,183	45,117	48,106	50,764
55	26,866	28,173	31,735	33,570	36,398	38,958	42,060	46,036	49,054	51,739

Wartości krytyczne χ^2 rozkładu χ^2 ; $P\{\chi^2_r \geq \chi^2(\alpha, r)\} = \alpha$

Tablica II (cd.)

r/α	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
1	0,455	0,708	1,074	1,642	2,706	3,841	5,024	6,636	7,879	10,828	12,116
2	1,386	1,833	2,408	3,219	4,605	5,991	7,378	9,210	10,597	13,816	15,202
3	2,366	2,946	3,665	4,642	6,251	7,815	9,348	11,345	12,838	16,266	17,730
4	3,357	4,045	4,878	5,989	7,779	9,488	11,143	13,277	14,860	18,467	19,997
5	4,351	5,132	6,064	7,289	9,236	11,070	12,832	15,086	16,750	20,515	22,105
6	5,345	6,211	7,231	8,558	10,645	12,592	14,449	16,812	18,548	22,458	24,103
7	6,346	7,283	8,383	9,803	12,017	14,067	16,013	18,475	20,278	24,322	26,018
8	7,344	8,351	9,524	11,030	13,362	15,507	17,535	20,090	21,955	26,125	27,868
9	8,343	9,414	10,656	12,242	14,684	16,919	19,023	21,666	23,589	27,877	29,666
10	9,342	10,473	11,781	13,442	15,987	18,307	20,483	23,209	25,188	29,588	31,420
11	10,341	11,530	12,899	14,631	17,275	19,675	21,920	24,725	26,757	31,264	33,136
12	11,340	12,584	14,011	15,812	18,549	21,026	23,336	26,217	28,300	32,909	34,821
13	12,340	13,636	15,119	16,985	19,812	22,362	24,735	27,688	29,819	34,528	36,478
14	13,339	14,685	16,222	18,151	21,064	23,685	26,119	29,141	31,319	36,123	38,109
15	14,339	15,733	17,322	19,311	22,307	24,996	27,488	30,578	32,801	37,697	39,719
16	15,338	16,780	18,418	20,465	23,542	26,296	28,845	32,000	34,267	39,252	41,308
17	16,338	17,824	19,511	21,615	24,769	27,587	30,191	33,409	35,718	40,790	42,879
18	17,338	18,868	20,601	22,760	25,989	28,869	31,526	34,805	37,156	42,312	44,434
19	18,338	19,910	21,689	23,900	27,204	30,144	32,852	36,191	38,582	43,820	45,973
20	19,337	20,951	22,775	25,038	28,412	31,410	34,170	37,566	39,997	45,315	47,498
21	20,337	21,991	23,858	26,171	29,615	32,671	35,479	38,932	41,401	46,797	49,010
22	21,337	23,031	24,939	27,301	30,813	33,924	36,781	40,289	42,796	48,268	50,511
23	22,337	24,069	26,018	28,429	32,007	35,172	38,086	41,638	44,181	49,728	52,000
24	23,337	25,106	27,096	29,553	33,196	36,415	39,364	42,980	45,558	51,179	53,479
25	24,337	26,143	28,172	30,675	34,382	37,652	40,646	44,314	46,928	52,618	54,947
26	25,336	27,179	29,246	31,795	35,563	38,885	41,923	45,642	48,290	54,052	56,407
27	26,336	28,214	30,319	32,912	36,741	40,113	43,194	46,963	49,645	55,476	57,858
28	27,336	29,249	31,391	34,027	37,916	41,337	44,461	48,278	50,993	56,892	59,300
29	28,336	30,283	32,461	35,139	39,087	42,557	45,722	49,588	52,336	58,301	60,735
30	29,336	31,316	33,530	36,250	40,256	43,773	46,979	50,892	53,672	59,703	62,162
31	30,336	32,349	34,598	37,359	41,422	44,985	48,232	52,191	55,003	61,098	63,582
32	31,336	33,381	35,665	38,466	42,585	46,194	49,480	53,486	56,328	62,487	64,995
33	32,336	34,413	36,731	39,572	43,745	47,400	50,725	54,776	57,648	63,870	66,402
34	33,336	35,444	37,795	40,676	44,903	48,602	51,966	56,061	58,964	65,247	67,803
35	34,336	36,475	38,859	41,778	46,059	49,802	53,203	57,342	60,275	66,619	69,199
36	35,336	37,505	39,922	42,879	47,212	50,998	54,437	58,619	61,581	67,985	70,588
37	36,336	38,535	40,984	43,978	48,363	52,192	55,668	59,892	62,882	69,346	71,972
38	37,335	39,564	42,045	45,076	49,513	53,384	56,895	61,162	64,181	70,703	73,351
39	38,335	40,593	43,105	46,173	50,660	54,572	58,120	62,428	65,476	72,055	74,725
40	39,335	41,622	44,165	47,269	51,805	55,768	59,342	63,691	66,766	73,402	76,095
41	40,335	42,651	45,224	48,363	52,949	56,942	60,561	64,950	68,053	74,745	77,459
42	41,335	43,679	46,282	49,456	54,090	58,124	61,777	66,206	69,336	76,084	78,820
43	42,335	44,706	47,339	50,548	55,230	59,304	62,990	67,459	70,616	77,419	80,176
44	43,335	45,734	48,396	51,639	56,369	60,481	64,201	68,709	71,893	78,749	81,528
45	44,335	46,761	49,452	52,729	57,505	61,656	65,410	69,957	73,166	80,077	82,876
46	45,335	47,787	50,507	53,818	58,641	62,830	66,617	71,201	74,437	81,400	84,220
47	46,335	48,814	51,562	54,906	59,774	64,001	67,821	72,443	75,704	82,720	85,560
48	47,335	49,840	52,616	55,993	60,907	65,171	69,023	73,683	76,969	84,037	86,897
49	48,335	50,866	53,670	57,079	62,038	66,339	70,222	74,919	78,231	85,351	88,231
50	49,335	51,892	54,723	58,164	63,167	67,505	71,420	76,154	79,490	86,661	89,561
51	50,335	52,917	55,775	59,248	64,295	68,669	72,616	77,386	80,747	87,968	90,881
52	51,335	53,942	56,827	60,332	65,422	69,832	73,810	78,616	82,001	89,272	92,211
53	52,335	54,967	57,879	61,414	66,548	70,993	75,002	79,843	83,253	90,573	93,531
54	53,335	55,992	58,930	62,496	67,673	72,153	76,192	81,069	84,502	91,872	94,849
55	54,335	57,016	59,980	63,577	68,796	73,311	77,380	82,292	85,749	93,167	96,163

Wartości krytyczne χ^2 rozkładu χ^2

Tablica II (cd.)

r/a	0,9995	0,999	0,995	0,99	0,975	0,95	0,90	0,80	0,70	0,60
56	27,555	28,881	32,490	34,350	37,212	39,801	42,937	46,955	50,005	52,715
57	28,248	29,591	33,248	35,131	38,027	40,646	43,816	47,876	50,956	53,691
58	28,943	30,305	34,008	35,913	38,844	41,492	44,696	48,797	51,909	54,667
59	29,640	31,020	34,770	36,698	39,662	42,339	45,577	49,718	52,857	55,643
60	30,340	31,738	35,535	37,485	40,482	43,188	46,459	50,641	53,809	56,620
61	31,043	32,459	36,301	38,273	41,303	44,038	47,342	51,564	54,761	57,597
62	31,748	33,181	37,068	39,063	42,126	44,889	48,226	52,487	55,714	58,574
63	32,455	33,906	37,838	39,855	42,950	45,741	49,111	53,412	56,666	59,551
64	33,165	34,633	38,610	40,649	43,776	46,595	49,996	54,336	57,619	60,528
65	33,877	35,362	39,383	41,444	44,603	47,450	50,883	55,262	58,573	61,506
66	34,591	36,093	40,158	42,240	45,431	48,305	51,770	56,188	59,527	62,484
67	35,307	36,826	40,935	43,038	46,261	49,162	52,659	57,115	60,481	63,461
68	36,025	37,561	41,713	43,838	47,092	50,020	53,548	58,042	61,436	64,440
69	36,745	38,298	42,494	44,639	47,924	50,879	54,438	58,970	62,391	65,418
70	37,467	39,036	43,275	45,442	48,758	51,739	55,329	59,898	63,346	66,396
71	38,192	39,777	44,058	46,246	49,592	52,600	56,221	60,827	64,302	67,375
72	38,918	40,520	44,843	47,051	50,428	53,462	57,113	61,756	65,258	68,353
73	39,646	41,262	45,629	47,858	51,265	54,325	58,006	62,686	66,214	69,332
74	40,376	42,010	46,417	48,666	52,103	55,189	58,900	63,616	67,170	70,311
75	41,107	42,757	47,206	49,475	52,942	56,054	59,795	64,547	68,127	71,290
76	41,841	43,507	47,997	50,286	53,782	56,920	60,690	65,478	69,084	72,270
77	42,576	44,258	48,788	51,097	54,623	57,786	61,586	66,409	70,042	73,249
78	43,313	45,010	49,582	51,910	55,466	58,654	62,483	67,341	70,999	74,229
79	44,051	45,764	50,376	52,725	56,309	59,522	63,308	68,274	71,957	75,208
80	44,791	46,520	51,172	53,540	57,153	60,391	64,278	69,207	72,915	76,188
81	45,533	47,277	51,969	54,357	57,998	61,261	65,176	70,140	73,874	77,168
82	46,276	48,036	52,767	55,174	58,845	62,132	66,076	71,074	74,833	78,148
83	47,021	48,796	53,567	55,993	59,692	63,004	66,976	72,008	75,792	79,128
84	47,767	49,557	54,368	56,813	60,540	63,876	67,876	72,943	76,751	80,108
85	48,515	50,320	55,170	57,634	61,389	64,749	68,777	73,878	77,710	81,089
86	49,264	51,085	55,973	58,456	62,239	65,623	69,679	74,813	78,670	82,069
87	50,015	51,850	56,777	59,279	63,089	66,498	70,581	75,749	79,630	83,050
88	50,767	52,617	57,582	60,103	63,941	67,373	71,484	76,685	80,590	84,031
89	51,521	53,386	58,389	60,928	64,793	68,249	72,387	77,622	81,550	85,012
90	52,276	54,155	59,196	61,754	65,647	69,126	73,291	78,558	82,511	85,993
91	53,032	54,926	60,005	62,581	66,501	70,003	74,196	79,496	83,472	86,974
92	53,790	55,698	60,815	63,409	67,356	70,882	75,101	80,433	84,433	87,955
93	54,549	56,472	61,625	64,238	68,211	71,760	76,006	81,371	85,394	88,936
94	55,309	57,246	62,437	65,068	69,068	72,640	76,912	82,309	86,356	89,917
95	56,070	58,022	63,250	65,898	69,925	73,520	77,818	83,248	87,317	90,899
96	56,833	58,799	64,063	66,730	70,783	74,400	78,725	84,187	88,279	91,881
97	57,597	59,577	64,878	67,562	71,642	75,282	79,633	85,126	89,241	92,862
98	58,362	60,356	65,694	68,396	72,501	76,164	80,541	86,065	90,204	93,844
99	59,128	61,136	66,510	69,230	73,361	77,046	81,449	87,005	91,166	94,826
100	59,896	61,918	67,328	70,075	74,222	77,929	82,358	87,945	92,129	95,808

Funkcje pomocnicze $p_1(\alpha)$ i $p_2(\alpha, \nu)$ ($\nu = 1/\sqrt{2r}$)

ν/α	0,9995	0,999	0,995	0,99	0,975	0,950	0,90	0,80	0,70	0,60
0,00	6,652	5,700	3,757	2,941	1,894	1,137	0,428	-0,194	-0,483	-0,624
0,01	6,537	5,691	3,757	2,945	1,901	1,145	0,436	-0,189	-0,487	-0,622
0,02	6,522	5,681	3,758	2,949	1,908	1,153	0,443	-0,183	-0,491	-0,620
0,03	6,506	5,670	3,758	2,953	1,914	1,160	0,451	-0,177	-0,495	-0,618
0,04	6,489	5,659	3,758	2,956	1,921	1,168	0,459	-0,171	-0,499	-0,616
0,05	6,471	5,647	3,758	2,959	1,927	1,176	0,466	-0,165	-0,503	-0,614
0,06	6,452	5,635	3,757	2,962	1,933	1,183	0,474	-0,159	-0,507	-0,612
0,07	6,432	5,621	3,755	2,964	1,940	1,191	0,481	-0,153	-0,511	-0,609
0,08	6,411	5,607	3,754	2,967	1,946	1,198	0,489	-0,147	-0,516	-0,607
$p_1(\alpha)$	-3,2905 2673	-3,0902 3231	-2,5758 2930	-2,3263 4787	-1,9599 6398	-1,6448 5363	-1,2815 5157	-0,8416 2123	-0,5244 0051	-0,2533 4710

Wartości krytyczne χ^2 rozkładu χ^2

Tablica II (cd.)

r/α	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
56	55,335	58,040	61,031	64,658	69,918	74,468	78,567	83,513	86,994	94,460	97,475
57	56,335	59,064	62,080	65,737	71,040	75,624	79,752	84,733	88,236	95,751	98,784
58	57,335	60,088	63,129	66,816	72,160	76,778	80,936	85,950	89,477	97,039	100,090
59	58,335	61,111	64,178	67,894	73,279	77,931	82,117	87,166	90,715	98,324	101,394
60	59,335	62,135	65,226	68,972	74,397	79,082	83,298	88,379	91,952	99,607	102,695
61	60,335	63,158	66,274	70,049	75,514	80,232	84,476	89,591	93,186	100,888	103,993
62	61,335	64,181	67,322	71,125	76,630	81,381	85,654	90,802	94,419	102,166	105,289
63	62,335	65,204	68,369	72,201	77,745	82,529	86,830	92,010	95,649	103,442	106,583
64	63,335	66,226	69,416	73,276	78,860	83,675	88,004	93,217	96,878	104,716	107,875
65	64,335	67,249	70,462	74,351	79,971	84,821	89,177	94,422	98,105	105,988	109,164
66	65,335	68,271	71,508	75,425	81,086	85,965	90,349	95,626	99,330	107,258	110,451
67	66,335	69,293	72,554	76,498	82,197	87,108	91,519	96,828	100,554	108,526	111,736
68	67,334	70,315	73,600	77,571	83,308	88,250	92,688	98,028	101,776	109,791	113,018
69	68,334	71,337	74,645	78,643	84,418	89,391	93,856	99,227	102,996	111,055	114,299
70	69,334	72,358	75,689	79,715	85,527	90,531	95,023	100,425	104,215	112,317	115,578
71	70,334	73,380	76,734	80,786	86,635	91,670	96,189	101,621	105,432	113,577	116,854
72	71,334	74,401	77,778	81,857	87,743	92,808	97,353	102,816	106,648	114,835	118,129
73	72,334	75,422	78,822	82,927	88,850	93,945	98,516	104,010	107,862	116,092	119,402
74	73,334	76,443	79,865	83,997	89,956	95,081	99,678	105,202	109,074	117,346	120,673
75	74,334	77,464	80,908	85,066	91,061	96,217	100,839	106,393	110,286	118,599	121,942
76	75,334	78,485	81,951	86,135	92,166	97,351	101,999	107,582	111,495	119,850	123,209
77	76,334	79,505	82,994	87,203	93,270	98,484	103,158	108,771	112,704	121,100	124,475
78	77,334	80,526	84,036	88,271	94,374	99,617	104,316	109,958	113,911	122,348	125,739
79	78,334	81,546	85,078	89,338	95,476	100,749	105,473	111,144	115,117	123,594	127,001
80	79,334	82,566	86,120	90,405	96,578	101,879	106,629	112,329	116,321	124,839	128,261
81	80,334	83,586	87,161	91,472	97,608	103,010	107,783	113,512	117,524	126,082	129,520
82	81,334	84,606	88,202	92,538	98,780	104,139	108,937	114,695	118,726	127,324	130,778
83	82,334	85,626	89,243	93,604	99,880	105,267	110,090	115,876	119,927	128,565	132,033
84	83,334	86,646	90,284	94,669	100,980	106,395	111,242	117,057	121,126	129,804	133,288
85	84,334	87,665	91,325	95,734	102,079	107,522	112,393	118,236	122,325	131,041	134,540
86	85,334	88,685	92,365	96,799	103,177	108,648	113,544	119,414	123,522	132,277	135,792
87	86,334	89,704	93,405	97,863	104,275	109,773	114,693	120,591	124,718	133,512	137,041
88	87,334	90,723	94,445	98,927	105,372	110,898	115,841	121,767	125,913	134,745	138,290
89	88,334	91,742	95,484	99,991	106,469	112,022	116,989	122,942	127,106	135,978	139,537
90	89,334	92,761	96,524	101,054	107,565	113,145	118,136	124,116	128,299	137,208	140,782
91	90,334	93,780	97,563	102,117	108,661	114,268	119,282	125,289	129,491	138,438	142,027
92	91,334	94,799	98,602	103,179	109,756	115,390	120,427	126,462	130,681	139,666	143,269
93	92,334	95,818	99,641	104,241	110,850	116,511	121,571	127,633	131,871	140,893	144,511
94	93,334	96,836	100,679	105,303	111,944	117,632	122,715	128,803	133,059	142,119	145,751
95	94,334	97,855	101,717	106,364	113,038	118,752	123,858	129,973	134,247	143,344	146,990
96	95,334	98,873	102,755	107,425	114,131	119,871	125,000	131,141	135,433	144,567	148,228
97	96,334	99,892	103,793	108,486	115,223	120,990	126,141	132,309	136,619	145,789	149,465
98	97,334	100,910	104,831	109,547	116,315	122,108	127,282	133,476	137,803	147,010	150,700
99	98,334	101,928	105,868	110,607	117,407	123,225	128,422	134,642	138,987	148,230	151,934
100	99,334	102,946	106,906	111,667	118,498	124,342	129,561	135,807	140,169	149,449	153,167

Funkcje pomocnicze $p_1(\alpha)$ i $p_2(\alpha, \nu)$ ($\nu = 1/\sqrt{2r}$)

ν/α	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
0,00	-0,667	-0,624	-0,483	-0,194	0,428	1,137	1,894	2,941	3,757	5,700	6,552
0,01	-0,667	-0,626	-0,479	-0,200	0,421	1,129	1,887	2,937	3,755	5,708	6,565
0,02	-0,667	-0,628	-0,476	-0,206	0,413	1,121	1,880	2,933	3,754	5,716	6,578
0,03	-0,667	-0,630	-0,472	-0,212	0,405	1,113	1,873	2,928	3,752	5,723	6,590
0,04	-0,666	-0,631	-0,468	-0,218	0,398	1,105	1,866	2,923	3,750	5,730	6,601
0,05	-0,666	-0,633	-0,464	-0,224	0,390	1,097	1,859	2,918	3,748	5,736	6,612
0,06	-0,666	-0,635	-0,460	-0,229	0,382	1,089	1,851	2,913	3,745	5,741	6,622
0,07	-0,666	-0,637	-0,457	-0,235	0,375	1,081	1,844	2,908	3,742	5,747	6,631
0,08	-0,666	-0,639	-0,453	-0,241	0,367	1,073	1,836	2,902	3,739	5,741	6,640
$p_1(\alpha)$	0,0000 0000	0,2533 4710	0,5244 0051	0,8416 2123	1,2815 5157	1,6448 5363	1,9599 6398	2,3263 4787	2,5758 2930	3,0902 3231	3,2905 2673

Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

Tablica III

n_1	n_2	α							
		0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995
2	2	-	-	-	-	4	4	4	4
	3	-	-	-	-	5	5	5	5
	4	-	-	-	-	5	5	5	5
	5	-	-	-	-	5	5	5	5
	6	-	-	-	-	5	5	5	5
	7	-	-	-	-	5	5	5	5
	8	-	-	-	2	5	5	5	5
	9	-	-	-	2	5	5	5	5
	10	-	-	-	2	5	5	5	5
	11	-	-	-	2	5	5	5	5
	12	-	-	2	2	5	5	5	5
	13	-	-	2	2	5	5	5	5
	14	-	-	2	2	5	5	5	5
	15	-	-	2	2	5	5	5	5
	16	-	-	2	2	5	5	5	5
	17	-	-	2	2	5	5	5	5
	18	-	-	2	2	5	5	5	5
	19	-	2	2	2	5	5	5	5
	20	-	2	2	2	5	5	5	5
	3	3	-	-	-	-	6	6	6
4		-	-	-	-	6	7	7	7
5		-	-	-	2	7	7	7	7
6		-	-	2	2	7	7	7	7
7		-	-	2	2	7	7	7	7
8		-	-	2	2	7	7	7	7
9		-	2	2	2	7	7	7	7
10		-	2	2	3	7	7	7	7
11		-	2	2	3	7	7	7	7
12		2	2	2	3	7	7	7	7
13		2	2	2	3	7	7	7	7
14	2	2	2	3	7	7	7	7	
15	2	2	3	3	7	7	7	7	
16	2	2	3	3	7	7	7	7	
17	2	2	3	3	7	7	7	7	
18	2	2	3	3	7	7	7	7	
19	2	2	3	3	7	7	7	7	
20	2	2	3	3	7	7	7	7	
4	4	-	-	-	2	7	8	8	8
	5	-	-	2	2	8	8	8	9
	6	-	2	2	3	8	8	9	9
	7	-	2	2	3	8	9	9	9
	8	2	2	3	3	9	9	9	9
	9	2	2	3	3	9	9	9	9
	10	2	2	3	3	9	9	9	9
	11	2	2	3	3	9	9	9	9
	12	2	3	3	4	9	9	9	9
	13	2	3	3	4	9	9	9	9
	14	2	3	3	4	9	9	9	9
15	3	3	3	4	9	9	9	9	
16	3	3	4	4	9	9	9	9	
17	3	3	4	4	9	9	9	9	
18	3	3	4	4	9	9	9	9	
19	3	3	4	4	9	9	9	9	
20	3	3	4	4	9	9	9	9	

Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

Tablica III cd.

n_1	n_2	α							
		0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995
5	5	-	2	2	3	8	9	9	10
	6	2	2	3	3	9	9	10	10
	7	2	2	3	3	9	10	10	11
	8	2	2	3	3	9	10	11	11
	9	2	3	3	4	10	11	11	11
	10	3	3	3	4	10	11	11	11
	11	3	3	4	4	11	11	11	11
	12	3	3	4	4	11	11	11	11
	13	3	3	4	4	11	11	11	11
	14	3	3	4	5	11	11	11	11
	15	3	4	4	5	11	11	11	11
	16	3	4	4	5	11	11	11	11
17	3	4	4	5	11	11	11	11	
18	4	4	5	5	11	11	11	11	
19	4	4	5	5	11	11	11	11	
20	4	4	5	5	11	11	11	11	
6	6	2	2	3	3	10	10	11	11
	7	2	3	3	4	10	11	11	12
	8	3	3	3	4	11	11	12	12
	9	3	3	3	4	11	11	12	12
	10	3	3	4	5	11	12	13	13
	11	3	4	4	5	12	12	13	13
	12	3	4	4	5	12	12	13	13
	13	3	4	5	5	12	13	13	13
	14	4	4	5	5	12	13	13	13
	15	4	4	5	6	13	13	13	13
	16	4	4	5	6	13	13	13	13
	17	4	5	5	6	13	13	13	13
18	4	5	5	6	13	13	13	13	
19	4	5	6	6	13	13	13	13	
20	4	5	6	6	13	13	13	13	
7	7	3	3	3	4	11	12	12	12
	8	3	3	4	4	12	12	13	13
	9	3	4	4	5	12	13	13	14
	10	3	4	5	5	12	13	14	14
	11	4	4	5	5	13	13	14	14
	12	4	4	5	6	13	13	14	15
	13	4	5	5	6	13	14	15	15
	14	4	5	5	6	13	14	15	15
	15	4	5	6	6	14	14	15	15
	16	5	5	6	6	14	15	15	15
	17	5	5	6	7	14	15	15	15
	18	5	5	6	7	14	15	15	15
19	5	5	6	7	14	15	15	15	
20	5	6	6	7	14	15	15	15	
8	8	3	4	4	5	12	13	13	14
	9	3	4	5	5	13	13	14	14
	10	4	4	5	6	13	14	14	15
	11	4	5	5	6	14	14	15	15
	12	4	5	6	6	14	15	15	16
	13	5	5	6	6	14	15	16	16
	14	5	5	6	7	15	15	16	16
	15	5	5	6	7	15	15	16	17
	16	5	6	6	7	15	16	16	17
	17	5	6	7	7	15	16	17	17

Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

Tablica III cd.

n_1	n_2	α							
		0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995
	18	6	6	7	8	15	16	17	17
	19	6	6	7	8	15	16	17	17
	20	6	6	7	8	16	16	17	17
9	9	4	4	5	6	13	14	15	15
	10	4	5	5	6	14	15	15	16
	11	5	5	6	6	14	15	16	16
	12	5	5	6	7	15	15	16	17
	13	5	6	6	7	15	16	17	17
	14	5	6	7	7	16	16	17	17
	15	5	6	7	8	16	17	17	18
	16	6	6	7	8	16	17	17	18
	17	6	7	7	8	16	17	18	18
	18	6	7	8	8	17	17	18	19
	19	6	7	8	8	17	17	18	19
	20	7	7	8	9	17	17	18	19
10	10	5	5	6	6	15	15	16	16
	11	5	5	6	7	15	16	17	17
	12	5	6	7	7	16	16	17	18
	13	5	6	7	8	16	17	18	18
	14	6	6	7	8	16	17	18	18
	15	6	7	7	8	17	17	18	19
	16	6	7	8	8	17	18	19	19
	17	7	7	8	9	17	18	19	19
	18	7	7	8	9	18	18	19	20
	19	7	8	8	9	18	19	19	20
	20	7	8	9	9	18	19	19	20
11	11	5	6	7	7	16	16	17	18
	12	6	6	7	8	16	17	18	18
	13	6	6	7	8	17	18	18	19
	14	6	7	8	8	17	18	19	19
	15	7	7	8	9	18	18	19	20
	16	7	7	8	9	18	19	20	20
	17	7	8	9	9	18	19	20	21
	18	7	8	9	10	19	19	20	21
	20	8	8	9	10	19	20	21	21
12	12	6	7	7	8	17	18	18	19
	13	6	7	7	8	17	18	19	20
	14	7	7	8	9	18	19	20	20
	15	7	8	8	9	18	19	20	21
	16	7	8	9	10	19	20	21	21
	17	8	8	9	10	19	20	21	21
	18	8	8	9	10	20	20	21	22
	19	8	9	10	10	20	21	22	22
	20	8	9	10	11	20	21	22	22
13	13	7	7	8	9	18	19	20	20
	14	7	7	9	9	19	19	20	21
	15	7	8	9	10	19	20	21	21
	16	8	8	9	10	20	20	21	22
	17	8	9	10	10	20	21	22	22
	18	8	9	10	11	20	21	22	23
	19	9	9	10	11	21	22	23	23
	20	9	10	10	11	21	22	23	23
14	14	7	8	9	10	19	20	21	22
	15	8	8	9	10	20	21	22	22
	16	8	9	10	11	20	21	22	23

Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

Tablica III cd.

n_1	n_2	α							
		0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995
	17	8	9	10	11	21	22	23	23
	18	9	9	10	11	21	22	23	24
	19	9	10	11	12	22	22	23	24
	20	9	10	11	12	22	23	24	24
15	15	8	9	10	11	20	21	22	23
	16	9	9	10	11	21	22	23	23
	17	9	10	11	11	21	22	23	24
	18	9	10	11	12	22	23	24	24
	19	10	10	11	12	22	23	24	25
	20	10	11	12	12	23	24	25	25
16	16	9	10	11	11	22	22	23	24
	17	9	10	11	12	22	23	24	25
	18	10	10	11	12	23	24	25	25
	19	10	11	12	13	23	24	25	26
	20	10	11	12	13	24	24	25	26
17	17	10	10	11	12	23	24	25	25
	18	10	11	12	13	23	24	25	26
	19	10	11	12	13	24	25	26	26
	20	11	11	13	13	24	25	26	27
18	18	11	11	12	13	24	25	26	26
	19	11	12	13	14	24	25	27	27
	20	11	12	13	14	25	26	27	28
19	19	11	12	13	14	25	26	27	28
	20	12	12	13	14	26	26	28	28
20	20	12	13	14	15	26	27	27	29

Wartości krytyczne rozkładu Durbina-Watsona dla $\alpha = 0,05$

Tablica IV

T	$k = 1$		$k = 2$		$k = 3$	
	dL	dU	dL	dU	dL	dU
6	0,610	1,400				
7	0,700	1,356	0,467	1,896		
8	0,763	1,332	0,559	1,777	0,368	2,287
9	0,824	1,320	0,629	1,699	0,455	2,128
10	0,879	1,320	0,697	1,641	0,525	2,016
11	0,927	1,324	0,758	1,604	0,595	1,928
12	0,971	1,331	0,812	1,579	0,658	1,864
13	1,010	1,340	0,861	1,562	0,715	1,816
14	1,045	1,350	0,905	1,551	0,767	1,779
15	1,077	1,361	0,946	1,543	0,814	1,750
16	1,106	1,371	0,982	1,539	0,857	1,728
17	1,133	1,381	1,015	1,536	0,897	1,710
18	1,158	1,391	1,046	1,535	0,933	1,696
19	1,180	1,401	1,074	1,536	0,967	1,685
20	1,201	1,411	1,100	1,537	0,998	1,676
21	1,221	1,420	1,125	1,538	1,026	1,667
22	1,239	1,429	1,147	1,541	1,053	1,664
23	1,257	1,437	1,168	1,543	1,078	1,660
24	1,273	1,446	1,188	1,546	1,101	1,656
25	1,288	1,454	1,206	1,550	1,123	1,654
26	1,302	1,461	1,224	1,553	1,143	1,652
27	1,316	1,469	1,240	1,556	1,162	1,651
28	1,328	1,476	1,255	1,560	1,181	1,650
29	1,341	1,483	1,270	1,563	1,198	1,650
30	1,352	1,489	1,284	1,567	1,214	1,650
31	1,363	1,496	1,297	1,570	1,229	1,650
32	1,373	1,502	1,309	1,574	1,244	1,650
33	1,383	1,508	1,321	1,577	1,258	1,651
34	1,393	1,514	1,333	1,580	1,271	1,652
35	1,402	1,519	1,343	1,584	1,283	1,653
36	1,411	1,525	1,354	1,587	1,295	1,654
37	1,419	1,530	1,364	1,590	1,307	1,655

Wartości krytyczne $F(\alpha, r_1, r_2)$ rozkładu F dla $\alpha = 0,05$

Tablica V

r_1/r_2	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	161	200	216	225	230	234	237	239	241	242	244	245	246	247	248
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,71	8,69	8,67	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,87	5,84	5,82	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,64	4,60	4,58	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,96	3,92	3,90	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,53	3,49	3,47	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,24	3,20	3,17	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,03	2,99	2,96	2,94
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,86	2,83	2,80	2,77
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,74	2,70	2,67	2,65
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,64	2,60	2,57	2,54
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,55	2,51	2,48	2,46
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,48	2,44	2,41	2,39
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,42	2,38	2,35	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,37	2,33	2,30	2,28
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,33	2,29	2,26	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,29	2,25	2,22	2,19
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,26	2,21	2,18	2,16
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,22	2,18	2,15	2,12
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,20	2,16	2,12	2,10
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,17	2,13	2,10	2,07
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,15	2,11	2,07	2,05
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,13	2,09	2,05	2,03
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,11	2,07	2,04	2,01
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,09	2,05	2,02	1,99
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,08	2,04	2,00	1,97
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,06	2,02	1,99	1,96
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,05	2,01	1,97	1,94
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	1,99	1,96	1,93
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,07	2,01	1,97	1,94	1,91
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,05	1,99	1,95	1,92	1,89
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,03	1,98	1,93	1,90	1,87
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,02	1,96	1,92	1,88	1,85
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,95	1,90	1,87	1,84
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	1,99	1,93	1,89	1,86	1,83
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	1,98	1,92	1,88	1,84	1,81
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	1,97	1,91	1,87	1,83	1,80
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,96	1,90	1,86	1,82	1,79
50	4,03	3,18	2,79	2,56	2,40	2,28	2,20	2,13	2,07	2,03	1,95	1,89	1,85	1,81	1,78
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,86	1,82	1,78	1,75
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,88	1,82	1,77	1,73	1,70
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,79	1,75	1,71	1,68
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,83	1,77	1,72	1,69	1,65
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,82	1,76	1,71	1,67	1,64
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,80	1,74	1,69	1,66	1,62
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,78	1,72	1,68	1,64	1,61
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,77	1,71	1,66	1,62	1,59
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,76	1,70	1,65	1,61	1,58
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,69	1,64	1,60	1,57

Wartości krytyczne $F(\alpha, r_1, r_2)$ rozkładu F dla $\alpha = 0,05$

Tablica V cd.

r_1/r_2	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞
1	249	249	249	250	250	251	251	251	252	252	252	253	254	254	254
2	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5
3	8,65	8,64	8,63	8,62	8,62	8,60	8,59	8,59	8,58	8,57	8,56	8,55	8,54	8,53	8,53
4	5,79	5,77	5,76	5,75	5,75	5,73	5,72	5,71	5,70	5,69	5,67	5,66	5,65	5,64	5,63
5	4,54	4,53	4,52	4,50	4,50	4,48	4,46	4,45	4,44	4,43	4,41	4,41	4,39	4,37	4,37
6	3,86	3,84	3,83	3,82	3,81	3,79	3,77	3,76	3,75	3,74	3,72	3,71	3,69	3,68	3,67
7	3,43	3,41	3,40	3,39	3,38	3,36	3,34	3,33	3,32	3,30	3,29	3,27	3,25	3,24	3,23
8	3,13	3,12	3,10	3,09	3,08	3,06	3,04	3,03	3,02	3,01	2,99	2,97	2,95	2,94	2,93
9	2,92	2,90	2,89	2,87	2,86	2,84	2,83	2,81	2,80	2,79	2,77	2,76	2,73	2,72	2,71
10	2,75	2,74	2,72	2,71	2,70	2,68	2,66	2,65	2,64	2,62	2,60	2,59	2,56	2,55	2,54
11	2,63	2,61	2,59	2,58	2,57	2,55	2,53	2,52	2,51	2,49	2,47	2,46	2,43	2,42	2,40
12	2,52	2,51	2,49	2,48	2,47	2,44	2,43	2,41	2,40	2,38	2,36	2,35	2,32	2,31	2,30
13	2,44	2,42	2,41	2,39	2,38	2,36	2,34	2,33	2,31	2,30	2,27	2,26	2,23	2,22	2,21
14	2,37	2,35	2,33	2,32	2,31	2,28	2,27	2,25	2,24	2,22	2,20	2,19	2,16	2,14	2,13
15	2,31	2,29	2,27	2,26	2,25	2,22	2,20	2,19	2,18	2,16	2,14	2,12	2,10	2,08	2,07
16	2,25	2,24	2,22	2,21	2,19	2,17	2,15	2,14	2,12	2,11	2,08	2,07	2,04	2,02	2,01
17	2,21	2,19	2,17	2,16	2,15	2,12	2,10	2,09	2,08	2,06	2,03	2,02	1,99	1,97	1,96
18	2,17	2,15	2,13	2,12	2,11	2,08	2,06	2,05	2,04	2,02	1,99	1,98	1,95	1,93	1,92
19	2,13	2,11	2,10	2,08	2,07	2,05	2,03	2,01	2,00	1,98	1,96	1,94	1,91	1,89	1,88
20	2,10	2,08	2,07	2,05	2,04	2,01	1,99	1,98	1,97	1,95	1,92	1,91	1,88	1,86	1,84
21	2,07	2,05	2,04	2,02	2,01	1,98	1,96	1,95	1,94	1,92	1,89	1,88	1,84	1,82	1,81
22	2,05	2,03	2,01	2,00	1,98	1,96	1,94	1,92	1,91	1,89	1,86	1,85	1,82	1,80	1,78
23	2,02	2,00	1,99	1,97	1,96	1,93	1,91	1,90	1,88	1,86	1,84	1,82	1,79	1,77	1,76
24	2,00	1,98	1,97	1,95	1,94	1,91	1,89	1,88	1,86	1,84	1,82	1,80	1,77	1,75	1,73
25	1,98	1,96	1,95	1,93	1,92	1,89	1,87	1,86	1,84	1,82	1,80	1,78	1,75	1,73	1,71
26	1,97	1,95	1,93	1,91	1,90	1,87	1,85	1,84	1,82	1,80	1,78	1,76	1,73	1,71	1,69
27	1,95	1,93	1,91	1,90	1,88	1,86	1,84	1,82	1,81	1,79	1,76	1,74	1,71	1,69	1,67
28	1,93	1,91	1,90	1,88	1,87	1,84	1,82	1,80	1,79	1,77	1,74	1,73	1,69	1,67	1,65
29	1,92	1,90	1,88	1,87	1,85	1,83	1,81	1,79	1,77	1,75	1,73	1,71	1,67	1,65	1,64
30	1,91	1,89	1,87	1,85	1,84	1,81	1,79	1,77	1,76	1,74	1,71	1,70	1,66	1,64	1,62
32	1,88	1,86	1,85	1,83	1,82	1,79	1,77	1,75	1,74	1,71	1,69	1,67	1,63	1,61	1,59
34	1,86	1,84	1,82	1,80	1,80	1,77	1,75	1,73	1,71	1,69	1,66	1,65	1,61	1,59	1,57
36	1,85	1,82	1,81	1,79	1,78	1,75	1,73	1,71	1,69	1,67	1,64	1,62	1,59	1,56	1,55
38	1,83	1,81	1,79	1,77	1,76	1,73	1,71	1,69	1,68	1,65	1,62	1,61	1,57	1,54	1,53
40	1,81	1,79	1,77	1,76	1,74	1,72	1,69	1,67	1,66	1,64	1,61	1,59	1,55	1,53	1,51
42	1,80	1,78	1,76	1,74	1,73	1,70	1,68	1,66	1,65	1,62	1,59	1,57	1,53	1,51	1,49
44	1,79	1,77	1,75	1,73	1,72	1,69	1,67	1,65	1,63	1,61	1,58	1,56	1,52	1,49	1,48
46	1,78	1,76	1,74	1,72	1,71	1,68	1,65	1,64	1,62	1,60	1,57	1,55	1,51	1,48	1,46
48	1,77	1,75	1,73	1,71	1,70	1,67	1,64	1,62	1,61	1,59	1,56	1,54	1,49	1,47	1,45
50	1,76	1,74	1,72	1,70	1,69	1,66	1,63	1,61	1,60	1,58	1,54	1,52	1,48	1,46	1,44
60	1,72	1,70	1,68	1,66	1,65	1,62	1,59	1,57	1,56	1,53	1,50	1,48	1,44	1,41	1,39
80	1,68	1,65	1,63	1,62	1,60	1,57	1,54	1,52	1,51	1,48	1,45	1,43	1,38	1,35	1,32
100	1,65	1,63	1,61	1,59	1,57	1,54	1,52	1,49	1,48	1,45	1,41	1,39	1,34	1,31	1,28
125	1,63	1,60	1,58	1,57	1,55	1,52	1,49	1,47	1,45	1,42	1,39	1,36	1,31	1,27	1,25
150	1,61	1,59	1,57	1,55	1,53	1,50	1,48	1,45	1,44	1,41	1,37	1,34	1,29	1,25	1,22
200	1,60	1,57	1,55	1,53	1,52	1,48	1,46	1,43	1,41	1,39	1,35	1,32	1,26	1,22	1,19
300	1,58	1,55	1,53	1,51	1,50	1,46	1,43	1,41	1,39	1,36	1,32	1,30	1,23	1,19	1,15
500	1,56	1,54	1,52	1,50	1,48	1,45	1,42	1,40	1,38	1,34	1,30	1,28	1,21	1,16	1,11
1000	1,55	1,53	1,51	1,49	1,47	1,44	1,41	1,38	1,36	1,33	1,29	1,26	1,19	1,13	1,08
∞	1,54	1,52	1,50	1,48	1,46	1,42	1,39	1,37	1,35	1,32	1,27	1,24	1,17	1,11	1,00

Dystrybuanta rozkładu normalnego

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-u^2/2} du$$

Tablica VI

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500 000	0,503 989	0,507 978	0,511 966	0,515 953	0,519 939	0,523 922	0,527 903	0,531 881	0,535 856
0,1	0,539 828	0,543 795	0,547 758	0,551 717	0,555 670	0,559 618	0,563 559	0,567 495	0,571 424	0,575 345
0,2	0,579 260	0,583 166	0,587 064	0,590 954	0,594 835	0,598 706	0,602 568	0,606 420	0,610 261	0,614 092
0,3	0,617 911	0,621 720	0,625 516	0,629 300	0,633 072	0,636 831	0,640 576	0,644 309	0,648 027	0,651 732
0,4	0,655 422	0,659 097	0,662 757	0,666 402	0,670 031	0,673 645	0,677 242	0,680 822	0,684 386	0,687 933
0,5	0,691 462	0,694 974	0,698 468	0,701 944	0,705 401	0,708 840	0,712 260	0,715 661	0,719 043	0,722 405
0,6	0,725 747	0,729 069	0,732 371	0,735 653	0,738 914	0,742 154	0,745 373	0,748 571	0,751 748	0,754 903
0,7	0,758 036	0,761 148	0,764 238	0,767 305	0,770 350	0,773 373	0,776 373	0,779 350	0,782 305	0,785 236
0,8	0,788 145	0,791 030	0,793 892	0,796 731	0,799 546	0,802 337	0,805 105	0,807 850	0,810 570	0,813 267
0,9	0,815 940	0,818 589	0,821 214	0,823 814	0,826 391	0,828 944	0,831 472	0,833 977	0,836 457	0,838 913
1,0	0,841 345	0,843 752	0,846 136	0,848 495	0,850 830	0,853 141	0,855 428	0,857 690	0,859 929	0,862 143
1,1	0,864 334	0,866 500	0,868 643	0,870 762	0,872 857	0,874 928	0,876 976	0,879 000	0,881 000	0,882 977
1,2	0,884 930	0,886 861	0,888 768	0,890 651	0,892 512	0,894 350	0,896 165	0,897 958	0,899 727	0,901 475
1,3	0,903 200	0,904 902	0,906 582	0,908 241	0,908 877	0,911 492	0,913 085	0,914 657	0,916 207	0,917 736
1,4	0,919 243	0,920 730	0,922 196	0,923 641	0,925 066	0,926 471	0,927 855	0,929 219	0,930 563	0,931 888
1,5	0,933 193	0,934 478	0,935 745	0,936 992	0,938 220	0,939 429	0,940 620	0,941 792	0,942 947	0,944 083
1,6	0,945 201	0,946 301	0,947 384	0,948 449	0,949 497	0,950 529	0,951 543	0,952 540	0,953 521	0,954 486
1,7	0,955 435	0,956 367	0,957 284	0,958 185	0,957 070	0,959 941	0,960 796	0,961 636	0,962 462	0,963 273
1,8	0,964 070	0,964 852	0,965 620	0,966 375	0,967 116	0,967 843	0,968 557	0,969 258	0,969 946	0,970 621
1,9	0,971 283	0,971 933	0,972 571	0,973 197	0,973 810	0,974 412	0,975 002	0,975 581	0,976 148	0,976 705
2,0	0,977 250	0,977 784	0,978 308	0,978 822	0,979 325	0,979 818	0,980 301	0,980 774	0,981 237	0,981 691
2,1	0,982 136	0,982 571	0,982 997	0,983 414	0,983 823	0,984 222	0,984 614	0,984 997	0,985 371	0,985 738
2,2	0,986 097	0,986 447	0,986 791	0,987 126	0,987 455	0,987 776	0,988 089	0,988 396	0,988 696	0,988 989
2,3	0,989 276	0,989 556	0,989 830	0,990 097	0,990 358	0,990 613	0,990 863	0,991 106	0,991 344	0,991 576
2,4	0,991 802	0,992 024	0,992 240	0,992 451	0,992 656	0,992 857	0,993 053	0,993 244	0,993 431	0,993 613
2,5	0,993 790	0,993 963	0,994 132	0,994 297	0,994 457	0,994 614	0,994 766	0,994 915	0,995 060	0,995 201
2,6	0,995 339	0,995 473	0,995 604	0,995 731	0,995 855	0,995 975	0,996 093	0,996 207	0,996 319	0,996 427
2,7	0,996 533	0,996 636	0,996 736	0,996 833	0,996 926	0,997 020	0,997 110	0,997 197	0,997 282	0,997 365
2,8	0,997 445	0,997 523	0,997 599	0,997 673	0,997 744	0,997 814	0,997 882	0,997 948	0,998 012	0,998 074
2,9	0,998 134	0,998 193	0,998 250	0,998 305	0,998 359	0,998 411	0,998 462	0,998 511	0,998 559	0,998 605
3,0	0,998 650	0,998 694	0,998 736	0,998 777	0,998 817	0,998 856	0,998 893	0,998 930	0,998 966	0,998 999
3,1	0,9 ⁰ 324	0,9 ⁰ 646	0,9 ⁰ 957	0,9 ¹ 260	0,9 ¹ 553	0,9 ¹ 836	0,9 ² 112	0,9 ² 378	0,9 ² 636	0,9 ² 886
3,2	0,9 ³ 129	0,9 ³ 363	0,9 ³ 590	0,9 ³ 810	0,9 ⁴ 024	0,9 ⁴ 230	0,9 ⁴ 429	0,9 ⁴ 623	0,9 ⁴ 810	0,9 ⁴ 991
3,3	0,9 ⁵ 166	0,9 ⁵ 335	0,9 ⁵ 499	0,9 ⁵ 658	0,9 ⁵ 811	0,9 ⁵ 959	0,9 ⁶ 103	0,9 ⁶ 242	0,9 ⁶ 376	0,9 ⁶ 505
3,4	0,9 ⁶ 631	0,9 ⁶ 752	0,9 ⁶ 869	0,9 ⁶ 982	0,9 ⁷ 091	0,9 ⁷ 197	0,9 ⁷ 299	0,9 ⁷ 398	0,9 ⁷ 493	0,9 ⁷ 585
3,5	0,9 ⁷ 674	0,9 ⁷ 759	0,9 ⁷ 842	0,9 ⁷ 922	0,9 ⁷ 999	0,9 ⁸ 074	0,9 ⁸ 146	0,9 ⁸ 215	0,9 ⁸ 282	0,9 ⁸ 347
3,6	0,9 ⁸ 409	0,9 ⁸ 469	0,9 ⁸ 527	0,9 ⁸ 583	0,9 ⁸ 637	0,9 ⁸ 689	0,9 ⁸ 739	0,9 ⁸ 787	0,9 ⁸ 834	0,9 ⁸ 879
3,7	0,9 ⁸ 922	0,9 ⁸ 964	0,9 ⁹ 039	0,9 ⁹ 046	0,9 ⁹ 079	0,9 ⁹ 158	0,9 ⁹ 1504	0,9 ⁹ 1858	0,9 ⁹ 2159	0,9 ⁹ 2468
3,8	0,9 ² 765	0,9 ⁴ 052	0,9 ⁴ 327	0,9 ⁴ 593	0,9 ⁴ 848	0,9 ⁴ 094	0,9 ⁴ 331	0,9 ⁴ 558	0,9 ⁴ 777	0,9 ⁴ 988
3,9	0,9 ⁵ 190	0,9 ⁵ 385	0,9 ⁵ 573	0,9 ⁵ 753	0,9 ⁵ 926	0,9 ⁶ 092	0,9 ⁶ 253	0,9 ⁶ 406	0,9 ⁶ 554	0,9 ⁶ 696
4,0	0,9 ⁶ 833	0,9 ⁶ 964	0,9 ⁷ 090	0,9 ⁷ 211	0,9 ⁷ 327	0,9 ⁷ 439	0,9 ⁷ 546	0,9 ⁷ 649	0,9 ⁷ 748	0,9 ⁷ 843
4,1	0,9 ⁷ 934	0,9 ⁸ 022	0,9 ⁸ 106	0,9 ⁸ 186	0,9 ⁸ 263	0,9 ⁸ 338	0,9 ⁸ 409	0,9 ⁸ 477	0,9 ⁸ 542	0,9 ⁸ 605
4,2	0,9 ⁸ 665	0,9 ⁸ 723	0,9 ⁸ 778	0,9 ⁸ 832	0,9 ⁸ 882	0,9 ⁸ 931	0,9 ⁸ 978	0,9 ⁹ 0226	0,9 ⁹ 0655	0,9 ⁹ 1066
4,3	0,9 ⁵ 1460	0,9 ⁵ 1837	0,9 ⁵ 2199	0,9 ⁵ 2545	0,9 ⁵ 2876	0,9 ⁵ 3193	0,9 ⁵ 3497	0,9 ⁵ 3788	0,9 ⁵ 4066	0,9 ⁵ 4332
4,4	0,9 ⁴ 587	0,9 ⁵ 4831	0,9 ⁵ 5065	0,9 ⁵ 5288	0,9 ⁵ 5502	0,9 ⁵ 5706	0,9 ⁵ 5902	0,9 ⁵ 6089	0,9 ⁵ 6268	0,9 ⁵ 6439
4,5	0,9 ⁶ 602	0,9 ⁶ 759	0,9 ⁶ 908	0,9 ⁷ 051	0,9 ⁷ 187	0,9 ⁷ 318	0,9 ⁷ 442	0,9 ⁷ 561	0,9 ⁷ 675	0,9 ⁷ 784
4,6	0,9 ⁷ 888	0,9 ⁷ 987	0,9 ⁸ 081	0,9 ⁸ 172	0,9 ⁸ 258	0,9 ⁸ 340	0,9 ⁸ 419	0,9 ⁸ 494	0,9 ⁸ 566	0,9 ⁸ 634
4,7	0,9 ⁸ 699	0,9 ⁸ 761	0,9 ⁸ 821	0,9 ⁸ 877	0,9 ⁸ 931	0,9 ⁸ 983	0,9 ⁹ 0320	0,9 ⁹ 0789	0,9 ⁹ 1255	0,9 ⁹ 1661
4,8	0,9 ² 067	0,9 ² 453	0,9 ² 822	0,9 ⁶ 3173	0,9 ⁶ 3508	0,9 ⁶ 3827	0,9 ⁶ 4131	0,9 ⁶ 4420	0,9 ⁶ 4696	0,9 ⁶ 4958
4,9	0,9 ⁶ 5208	0,9 ⁶ 5446	0,9 ⁶ 5673	0,9 ⁶ 5889	0,9 ⁶ 6094	0,9 ⁶ 6289	0,9 ⁶ 6475	0,9 ⁶ 6652	0,9 ⁶ 6821	0,9 ⁶ 6981

Test Hellwiga

Tablica VII

Test Hellwiga

<i>n</i>	<i>a</i>							
	0,10		0,05		0,01		0,005	
2								
3								
4	0	2		2				
5	0	2	0	3		3		3
6	0	3	0	3		4		4
7	0	3	0	4	0	4		4
8	1	4	0	4	0	5	0	5
9	1	4	1	5	0	5	0	5
10	1	5	1	5	0	6	0	6
11	2	5	1	6	1	6	0	6
12	2	6	2	6	1	7	1	7
13	2	6	2	6	1	7	1	7
14	2	6	2	7	1	8	1	8
15	3	7	2	7	2	8	1	8
16	6	7	3	8	2	9	2	9
17	3	8	3	8	2	9	2	9
18	4	8	3	9	2	9	2	10
19	4	9	4	9	3	10	2	10
20	4	9	4	9	3	10	3	11
21	5	9	4	10	3	11	3	11
22	5	10	5	10	4	11	3	12
23	5	10	5	11	4	12	4	12
24	6	11	5	11	4	12	4	13
25	6	11	5	12	4	13	4	13
26	6	11	6	12	5	13	4	13
27	7	12	6	12	5	13	4	14
28	7	12	6	13	5	14	5	14
29	7	13	6	13	6	14	5	15
30	8	13	7	14	6	15	6	15

Bibliografia

- Balcerak A., Kwaśnicki W. (red.), *Symulacja systemów społeczno-gospodarczych*, Politechnika Wrocławska, Wrocław 2005.
- Bronsztejn I., Siemieniaj K., *Poradnik encyklopedyczny. Matematyka*, PWN, Warszawa 1959.
- Burdzy J., *Tablice statystyczne*, Politechnika Lubelska, Lublin 1999.
- Cieślak M. (red.), *Prognozowanie gospodarcze*, Wydawnictwo Naukowe PWN, Warszawa 2005.
- Czerwiński Z., *Matematyka na usługach ekonomii*, PWN, Warszawa 1984.
- Czyżewski R., Hundert M., Klósk R., *Wybrane zagadnienia z prognozowania*, Economicus, Szczecin 2006.
- Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania*, Akademia Ekonomiczna, Wrocław 2003.
- Freund J.E., *Podstawy nowoczesnej statystyki*, PWE, Warszawa 1971.
- Gajda J.B., *Wielorównaniowe modele ekonometryczne*, PWN, Warszawa 1988.
- Gajda J., *Ekonometria. Wykład i łatwe obliczenia w programie komputerowym*, Wydawnictwo C.H. Beck, Warszawa 2004.
- Gale D., *Teoria liniowych modeli ekonomicznych*, PWN, Warszawa 1969.
- Gedynin O., *Optymalne sterowanie procesami gospodarczymi*, PWN, Warszawa 1977.
- Grabiński T., Wydymus S., Zeliaś A., *Metody doboru zmiennych w modelach ekonometrycznych*, PWN, Warszawa 1982.
- Idźkiewicz A.Z., *PERT. Metody analizy sieciowej*, PWN, Warszawa 1967.
- Ignasiak E. (red.), *Badania operacyjne*, PWE, Warszawa 2001.
- Jajuga K. (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Akademia Ekonomiczna, Wrocław 2002.
- Jurlewicz T., Skoczylas Z., *Algebra liniowa 1. Definicje, twierdzenia i wzory*, Oficyna Wydawnicza GiS, Wrocław 2003.
- Jurlewicz T., Skoczylas Z., *Algebra liniowa 1. Przykłady i zadania*, Oficyna Wydawnicza GiS, Wrocław 2003.
- Klein L.R., *Wykłady z ekonometrii*, PWE, Warszawa 1982.
- Korn G.A., Korn T.M., *Matematyka dla pracowników naukowych i inżynierów*, PWN, Warszawa 1983.
- Kot S., *Ekonometryczne modele dobrobytu*, Wydawnictwo Naukowe, PWN, Warszawa 2000.

- K u d r y c k a I., *Problemy i metody modelowania ekonometrycznego*, PWN, Warszawa 1984.
- K u f e l T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2007.
- Kukuła K. (red.), *Badania operacyjne w przykładach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa 2004.
- K u k u ł a K., *Wprowadzenie do ekonometrii*, Wydawnictwo Naukowe PWN, Warszawa 2004.
- L a n g e O., *Wstęp do ekonometrii*, PWN, Warszawa 1967.
- Leksykon naukowo-techniczny*, WNT, Warszawa 1972.
- Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 1: *Statystyka*, Wydawnictwo C.H. Beck, Warszawa 2003.
- Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003.
- Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 3: *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003.
- M a d d a l a G.S., *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2006.
- „Małe Roczniki Statystyczne Polski”, GUS, Warszawa 1999–2006.
- Martos B., *Programowanie nieliniowe*, PWN, Warszawa 1983.
- Microsoft: Excel. Opis funkcji*, Microsoft Corporation Ireland 1993.
- Microsoft: Excel. Podręcznik użytkownika t. 2: Analiza arkusza, wymiana danych, automatyzacja prac*, Microsoft Corporation Ireland 1993.
- N o w a k E., *Zarys metod ekonometrii. Zbiór zadań*, Wydawnictwo Naukowe PWN, Warszawa 2002.
- N y k o w s k i I., *Programowanie liniowe*, PWE, Warszawa 1984.
- Osińska M. (red.), *Ekonometria współczesna*, TNOiK „Dom Organizatora”, Toruń 2007.
- O s t a s i e w i c z S., R u s n a k Z., S i e d l e c k a U., *Statystyka. Elementy teorii i zadania*, Akademia Ekonomiczna, Wrocław 2006.
- P l u c i ń s k a A., P l u c i ń s k i E., *Zadania z rachunku prawdopodobieństwa i statystyki matematycznej*, PWN, Warszawa 1970.
- Praca z pakietem Microsoft Office 97. Konkretnie rozwiązania w codziennej pracy*, Microsoft Corporation, Ireland 1995.
- Radzikowska B. (red.), *Metody prognozowania. Zbiór zadań*, Akademia Ekonomiczna, Wrocław 2004.
- Sikora W. (red.), *Przykłady i zadania z badań operacyjnych i ekonometrii*, Akademia Ekonomiczna, Poznań 2003.
- S t r a h l D., S o b c z a k E., M a r k o w s k a M., B a l - D o m a ń s k a B., *Modelowanie ekonometryczne z Excelem*, Akademia Ekonomiczna, Wrocław 2002.
- Z i e l i ń s k i R., *Tablice statystyczne*, PWN, Warszawa 1972.
- „Roczniki Statystyczne RP”, GUS, Warszawa 2000–2006.

Indeks rzeczowy

Uwaga: Pauzy oznaczają powtórzenie wyrazów poprzedniego hasła

- Algorytm simpleks 390
- Analiza korelacji 141
 - wrażliwości 455
- Antywzorzec 96
- Asymetria prawostronna 68
- Autokorelacja odchyłeń losowych 178

- Bezwzględny błąd prognozy ex ante** 266

- Całkowity koszt projektu 536
- Całościowa ocena przydatności modelu 192
- Ciąg reszt 27
- Cykl sezonowości 332
- Czas graniczny 535
 - najbardziej prawdopodobny 542
 - normalny 535
 - optymistyczny 542
 - pesymistyczny 542
 - trwania przedsięwzięcia 544
- Częstość 173

- Destymulanta** 94
- Drzewo celów 372
 - decyzyjne 376

- Ekonomiczne gry symulacyjne** 51
- Ekstremum bezwarunkowe 480
 - warunkowe 480
- Element rozwiązujący 425
- Estymacja parametrów 184
- EViews 39

- Excel 44
 - Solver 365

- Funkcja celu** 300
 - Lagrange’a 481
 - REGLINP 44
 - regresji liniowej 171
 - trendu 117
 - hiperbolicznego 128
 - liniowego 117
 - logarymicznego 127
 - logistycznego 131
 - potęgowego 121
 - wykładniczego 124
 - Excela 104

- Graf** 92
- Gretl 37

- Heteroscedastyczność** 179
- Histogram 68
- Hybryda 50

- Identyfikowalność równań** 217
- Ilość kombinacji zmiennych objaśniających 149
- Indeks sezonowości 332
- Indywidualna pojemność nośników informacji 138
- Istotność parametrów strukturalnych 189
- Izolinie 405

- Klasy informacji** 88

- Klasyczna metoda najmniejszych kwadratów 103
- Klasyczny współczynnik asymetrii 74
- Klatki zerowe 495
- Kowariancja 80
- Kryterium Akeike'a 291
 - Bayesa 381
 - Hurwicza 381
 - Savage 381
 - Schwarza 291
 - simpleks 392
 - Walda 381
- Krytyczna wartość współczynnika korelacji 24
- Kwadrat odległości Mahalonobisa 184
- Kwartyl 64

- Liniowy model ekonometryczny** 103
 - prawdopodobieństwa 247
- Losowość reszt 167

- Macierz diagonalna** 246
 - iloczynu 185
 - korelacji 80
 - kowariancji 80
 - kwadratowa 88
 - nieosobliwa 211
 - odwrotna 166
 - rozszerzona 165
 - rzutowania 183
 - symetryczna współczynników kowariancji 80
- Mediana 62
- Metoda CPM 532
 - geometryczna 405
 - grafowa 145
 - iteracyjna środków ciężkości 89
 - kar dla tras niedopuszczalnych 498
 - PERT 532
 - potencjałów 509
 - sum standaryzowanych 98
 - wzorca rozwoju 96
- Miary błędów *ex ante* 266
 - rozwoju 97
- Model adaptacyjny 262
 - addytywny 332
 - autoregresyjny 315
 - Boxa-Jenkinsa 230
 - Browna 309
 - ekonometryczny 17
 - Holta-Wintersa 329
 - jednorównaniowy 164
 - Kleina 353
 - logitowy 248
 - multiplikatywny 332
 - nierównowagi 20
 - o równaniach łącznie współzależnych 216
 - probitowy 249
 - prostego wyrównywania wykładniczego 302
 - prosty 211
 - rekurencyjny 212
 - symulacyjny 49
 - trendu kwadratowego 349
- Model wielomianu 348
 - wielorównaniowy 212
 - podwójnych średnich ruchomych 305
 - prostych średnich ruchomych 298
 - ważonych średnich ruchomych 300
- Moduł programowania liniowego i całkowitoliczbowego 365
- Nachylenie linii czasowo-kosztowej 535
- Nominanta 94
- Normalność rozkładu reszt 180

- Obserwacja nietypowa** 182
- Obszar rozwiązań dopuszczalnych 405
- Ocena parametrów 25
- Oczekiwany czas realizacji czynności 542
- Odbiorca fikcyjny 518

- Odchylenie ćwiartkowe 70
 - losowe 22
 - przeciętne 70
 - reszt 28
 - standardowe 73
 - dla przedsięwzięcia 544
- Odległość euklidesowa 87
 - miejska 87
 - między skupieniami 93
- Otwarte zagadnienie transportowe 501
- Parametr wyrównywania wykładniczego 302
 - modelu regresji 171
- Pierwiastek średniego kwadratu błędu 265
- Podwójna metoda najmniejszych kwadratów 218
- Pojemność integralna dla kombinacji 139
- Pojęcie ekonometrii 17
 - modelu ekonometrycznego 17
- Porządkowanie liniowe zbioru informacji 94
- Postać kanoniczna 390
 - zredukowana modelu 218
- Pozycyjny współczynnik asymetrii 75
- Problem komiwojażera 523
 - sterowania produkcją i zapasami 527
 - załadunku 526
- Programowanie całkowitoliczbowe 367
 - dynamiczne 523
- Regresja krokowa 148
- Rozkład reszt 167
- Rozstęp 70
- Rozwiązanie optymalne 405
- Ruchoma próba 139
- Segment 119
- Składnik losowy 167
- Skorygowany współczynnik determinacji 107
- Skupienia 91
- Stacjonarność szeregu czasowego 230
- Standardowy błąd oceny poszczególnych parametrów 165
- Standaryzacja reszt 176
 - zmiennej 83
- Statystyka Durбина-Watsona 178
 - empiryczna testu *t*-Studenta 25
 - opisowa 74
 - regresji 107
- Stymulanta 94
- Sygnal adaptacyjny 303
- Symulacja 49
- Syntetyczna miara współzależności liniowej 82
- Sytuacja decyzyjna 378
- Szereg stacjonarny 230
 - statystyczny 55
 - szczegółowy 55
- Sztuczna baza danych 445
- Ścieżka krytyczna 532
- Średni bezwzględny błąd prognozy *ex ante* 265
 - procentowy absolutny błąd 265
- Średnia arytmetyczna 57
 - geometryczna 60
 - harmoniczna 60
- Tablica simpleksowa 474
- Taksonomia wrocławska 91
- TESS-2 52
- Test Durбина-Watsona 32
 - Fishera-Snedecora 33
 - istotności współczynnika korelacji wielorakiej 24
 - Jargue’a-Bery 28
 - nieliniowości White’a 196
 - serii 27
 - zgodności Hellwiga 27
- Transformacja liniowa 123

Transpozycja macierzy 211
 Trend 332
 – pelzający 119

 Układ równań normalnych 283
 Unitaryzacja 84
 Uogólniona metoda najmniejszych kwadratów 246

 Wahania cykliczne 332
 – przypadkowe 332
 Wariancja populacji 72
 – próby 73
 – resztowa 34
 Wartość średnia trendu pelzającego 121
 – krytyczna współczynnika korelacji 326
 Warunki Kuhna-Tuckera 486
 Wektor miar położenia 82
 – współczynników korelacji 165
 – zmiennych 264
 Wersja kosztowa CPM 534
 Węzły 377
 Wielkości statystyczne zmiennych 111
 WinQSB 42
 Wskaźnik średniego względnego poziomu reszt 168
 – względnej ważności i -tego elementu 374
 Współczynnik autokorelacji 232
 – cząstkowej 148
 – reszt 31
 – błędu kowariancji 266
 – wariancji 265
 – determinacji 166
 – Janusowy 267
 Współczynnik korelacji 23
 – wielorakiej 144
 – kurtozy 29
 – skośności 29
 – średniego błędu 265
 – Thiela 265
 – wyrazistości 166
 – zbieżności 166
 – zmienności 73
 Wykres Gantta 534
 – korelacyjny 78
 – pudełkowy 67
 Wyrażenia 148
 Wyznacznik macierzy 156
 Względny błąd prognozy *ex ante* 266
 Wzory Cramera 283
 Wzorzec 96

 Zadanie decyzyjne 367
 – dualne 470
 – programowania nieliniowego całkowitoliczbowego 484
 – prymalne 470
 Zamknięte zagadnienie transportowe 494
 Zmienna dychotomiczna 246
 – objaśniana 315
 Zmienne bazowe 436
 – dualne 457
 – egzogeniczne 218
 – endogeniczne 211
 – nieistotne 486
 – objaśniające 212

$$\mathbf{B}_4^{-1} = \frac{1}{4} \begin{bmatrix} -2 & 4 & 0 \\ -6 & 6 & 4 \\ 2 & -2 & 0 \end{bmatrix} = \begin{bmatrix} 0,5 & 1 & 0 \\ -1,5 & 1,5 & 1 \\ 0,5 & -0,5 & 0 \end{bmatrix},$$

- iloczyny z macierzą odwrotną:

$$\mathbf{B}_4^{-1} \mathbf{A} = \begin{bmatrix} -0,5 & 1 & 0 \\ -1,5 & 1,5 & 1 \\ 0,5 & -0,5 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 2 \\ 2 & 2 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix},$$

$$\mathbf{B}_3^{-1} \mathbf{b} = \begin{bmatrix} -0,5 & 1 & 0 \\ -1,5 & 1,5 & 1 \\ 0,5 & -0,5 & 0 \end{bmatrix} \cdot \begin{bmatrix} 2000 \\ 1400 \\ 1200 \end{bmatrix} = \begin{bmatrix} 400 \\ 300 \\ 300 \end{bmatrix},$$

$$\mathbf{B}_4^{-1} \mathbf{A} = [30 \ 0 \ 40] \cdot \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} = [40 \ 30],$$

$$\mathbf{c}_b^T \mathbf{B}_4^{-1} = [30 \ 0 \ 40] \cdot \begin{bmatrix} -0,5 & 1 & 0 \\ -1,5 & 1,5 & 1 \\ 0,5 & -0,5 & 0 \end{bmatrix} = [5 \ 10 \ 0],$$

funkcja celu:

$$\mathbf{c}_b^T \mathbf{B}_4^{-1} \mathbf{b} = [30 \ 0 \ 40] \cdot \begin{bmatrix} 400 \\ 300 \\ 300 \end{bmatrix} = 2400.$$

żenia Excela (na przykładowych obszarach arkusza kalkulacyjnego):

\mathbf{B}_i^{-1}	= MACIERZ.ODW(F194:H196)
$\mathbf{B}_i^{-1} \mathbf{A}$	= MACIERZ.ILOCZYN(F187:H189;B194:C196)
$\mathbf{B}_i^{-1} \mathbf{b}$	= MACIERZ.ILOCZYN(F187:H189;J194:J196)
$\mathbf{c}_b^T \mathbf{B}_i^{-1}$	= MACIERZ.ILOCZYN(B192:D192;F187:H189)
$\mathbf{c}_b^T \mathbf{B}_i^{-1} \mathbf{A}$	= MACIERZ.ILOCZYN(B192:D192;D187:E189)
$\mathbf{c}_b^T \mathbf{B}_i^{-1} \mathbf{b}$	= MACIERZ.ILOCZYN(B192:D192;CI187:I189)

w kolejnych iteracjach: