EKONOMETRIA

PODRĘCZNIK. STUDIA PRZYPADKÓW



ANNA TATARCZAK

20
TOM

Tom 20 Kompetencje, Wiedza, Innowacje – Zintegrowany Program Rozwoju WSEI w Lublinie – Etap II

EKONOMETRIA

Część pierwsza
Podręcznik

Anna Tatarczak

Lublin 2021

WYŻSZA SZKOŁA EKONOMII I INNOWACJI W LUBLINIE
Seria wydawnicza:
Kompetencje, Wiedza, Innowacje – Zintegrowany Program Rozwoju WSEI
w Lublinie – Etap II

Tom 20

EKONOMETRIA
Cz. I - Podręcznik Cz. II - Studia przypadków

Autor:
Anna Tatarczak
ORCID: 0000-0001-8573-5791

Recenzenci:
dr Anna Makarewicz
dr Adam Kiersztyn

Korekta:
Magdalena Grela-Tokarczyk

Skład, łamanie:
Marta Krysińska-Kudlak

Projekt okładki:
Patrycja Kaczmarek

Grafika na okładce:
geralt/Pixabay.com
bedneyimages/Freepik.com

@Copyright by
Innovatio Press, Lublin 2021

Wszelkie prawa zastrzeżone. W sprawie kopiowania lub rozpowszechniania fragmentów albo całości niniejszej pracy należy kontaktować się z wydawcą.

Publikacja oraz recenzje zrealizowano w ramach Programu Operacyjnego Wiedza Edukacja Rozwój 2014-2020 współfinansowanego ze środków Europejskiego Funduszu Społecznego.

Proces wydawniczy (korekta, skład) sfinansowano z dotacji na zadania związane z zapewnieniem osobom niepełnosprawnym warunków do pełnego udziału w procesie kształcenia.

Printed in Poland
Innovatio Press Wydawnictwo Naukowe
Wyższej Szkoły Ekonomii i Innowacji
20-209 Lublin, ul. Projektowa 4
tel.: +48 81 749 17 77, fax: + 48 81 749 32 13
www.wsei.lublin.pl

ISBN – wersja elektroniczna: 978-83-66159-94-5

Spis treści

Wstęp 5

1. Wprowadzenie 7
   1.1. Co to jest ekonometria? 8
   1.2. Model ekonometryczny 9
   1.3. Zmienne modelu 17
   1.4. Szeregi czasowe 18
   1.5. Dobór zmiennych objaśniających 20
      1.5.1. Metoda Hellwiga 24
      1.5.2. Metoda Pawłowskiego 25
      1.5.3. Regresja krokowa 26
   1.6. Postać analityczna modelu 27
      1.6.1. Funkcja liniowa 28
      1.6.2. Funkcja wykładnicza 29
      1.6.3. Funkcje Törnquista 32
   1.7. Oprogramowanie w ekonometrii 37
   1.8. Pytania sprawdzające 43

2. Jednorównaniowy liniowy model ekonometryczny 46
   2.1. Metoda najmniejszych kwadratów 48
   2.2. Weryfikacja modelu 53
   2.3. Ocena stopnia dopasowania modelu 56
   2.4. Ocena normalności rozkładu składnika resztowego 61
   2.5. Autokorelacja 63
   2.6. Heteroskedastyczność 67
   2.7. Pytania sprawdzające 71

3. Prognozowanie 77
   3.1. Rodzaje prognoz 78
   3.2. Metody prognozowania 84

3.3. Etapy prognozowania 86
3.4. Trafność prognozy 89
3.5. Pytania sprawdzające 95

4. Analiza szeregów czasowych 99
4.1. Składowe szeregów czasowych 101
4.2. Model addytywny i multiplikatywny 104
4.3. Sezonowość 107
4.4. Metoda naiwna 110
4.5.Metoda średniej ruchomej 112
4.6. Wygładzanie wykładnicze 114
4.7. Pytania sprawdzające 118

Załączniki 121
Załącznik 1 – tablice rozkładu t-studenta 121
Załącznik 2 – tablice rozkładu chi-kwadrat 122

Literatura 123

Wstęp

Głównym celem podręcznika jest uporządkowanie i opis podstawowych metod i technik stosowanych na wykładzie z ekonometrii. Z wykładem związane są ćwiczenia laboratoryjne, w trakcie których są praktycznie ilustrowane zagadnienia przedstawiane na wykładzie. Podstawowym narzędziem programowym, stosowanym do opisu poszczególnych procedur obliczeniowych, jak i do obliczeń w laboratorium komputerowym jest program EXCEL oraz GRETL.

Podręcznik ten jest głównie przeznaczony dla studentów Wyższej Szkoły Ekonomii i Innowacji poznających podstawy ekonometrii. Mogą z niej również korzystać praktycy prowadzący badania własne lub kierujący pracami badawczymi.

Podręcznik składa się z czterech rozdziałów, jego układ jest dostosowany do programu ekonometrii obowiązującego na kierunkach ekonomicznych. Rozdział pierwszy przedstawia podstawowe pojęcia ekonometrii, dobór zmiennych objaśniających oraz opis najczęściej występujących postaci analitycznych funkcji. Rozdział drugi jest poświęcony metodzie najmniejszych kwadratów. Rozdział trzeci wprowadza czytelnika w pojęcie prognozowania. Rozdział czwarty szeroko opisuje metody analiz szeregów czasowych.

Każdy rozdział pracy zamyka szeroki zestaw pytań do samodzielnego rozwiązania. Dołączona bibliografia przedmiotu umożliwi zainteresowanym pogłębienie wiedzy Czytelnikom, dotarcie do innych podręczników z ekonometrii.

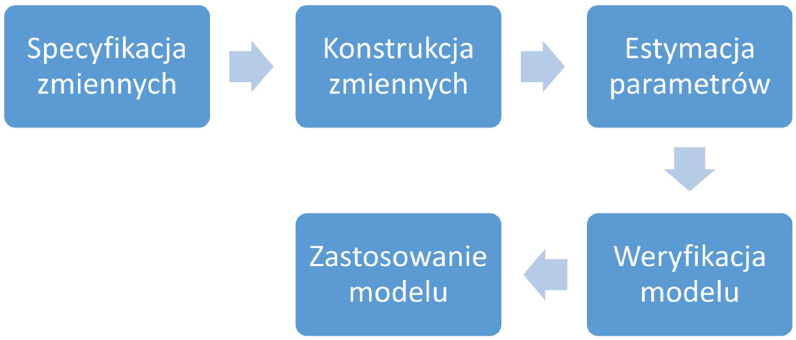
1. Wprowadzenie



1.1. Co to jest ekonometria?

Ekonometria jest to nauka wykorzystująca narzędzia matematyki, informatyki oraz statystyki. Funkcjonuje w ramach ekonomii i pełni wobec niej funkcję pomocniczą. Ekonometria opiera swoje działania na **badaniach ilościowych związków, które zachodzą między zjawiskami i zmiennymi ekonomicznymi**. Głównym założeniem ekonometrii jest to, że wykorzystuje metody, które są opracowywane poza ekonomią. W pewnym sensie **modelowanie ekonometryczne może być rozumiane jako szereg procedur** (patrz Rysunek 1), **których celem jest opracowanie modelu ekonometrycznego.** Procedury te zostaną szczegółowo omówione w dalszej części książki.

Przedmiotem zainteresowań są trzy grupy zagadnień:
- konstrukcja modelu ekonometrycznego,
- estymacja jego parametrów,
- szeroko pojęte wnioskowanie na podstawie modelu.



Rysunek 1. Procedura ekonometryczna.

Celem głównym ekonometrii jest badawcza analiza teorii ekonomicznych.

Możemy wyróżnić trzy cele szczegółowe realizowane przez ekonometrię:
- cel poznawczy: opis mechanizmu kształtowania się zjawisk,
- cel predyktywny: przewidywanie dalszego przebiegu zjawisk,
- cel decyzyjny: sterowanie przebiegiem zjawisk ekonomicznych.

1.2. Model ekonometryczny

Modelem ekonometrycznym nazywamy formalny opis stochastycznej zależności wyróżnionej wielkości, zjawiska lub przebiegu procesu ekonomicznego (zjawisk, procesów) od czynników, które je kształtują, wyrażony w formie pojedynczego równania bądź układu równań.

Strukturę każdego równania określają: zmienna objaśniana, zmienne objaśniające (nielosowe lub losowe) mające ustaloną treść ekonomiczną, parametry strukturalne, zmienna losowa (tradycyjnie nazywana składnikiem losowym) o nieznanej treści oraz określony typ związku funkcyjnego między zmienną objaśnianą a zmiennymi objaśniającymi i składnikiem losowym.

Przykład

Przykładami procesu stochastycznego są:
- produkcja przemysłowa w Polsce w kolejnych latach,
- stopa bezrobocia w kolejnych miesiącach,
- inflacja w kolejnych miesiącach,
- notowania kursów walutowych, akcji na kolejnych sesjach giełdowych (w kolejnych dniach).

Przykład

Dany jest model ekonometryczny

$$Y = \alpha_0 + \alpha_1 X + \varepsilon$$

gdzie:
Y oznacza cenę mieszkania, X jest powierzchnią mieszkania.

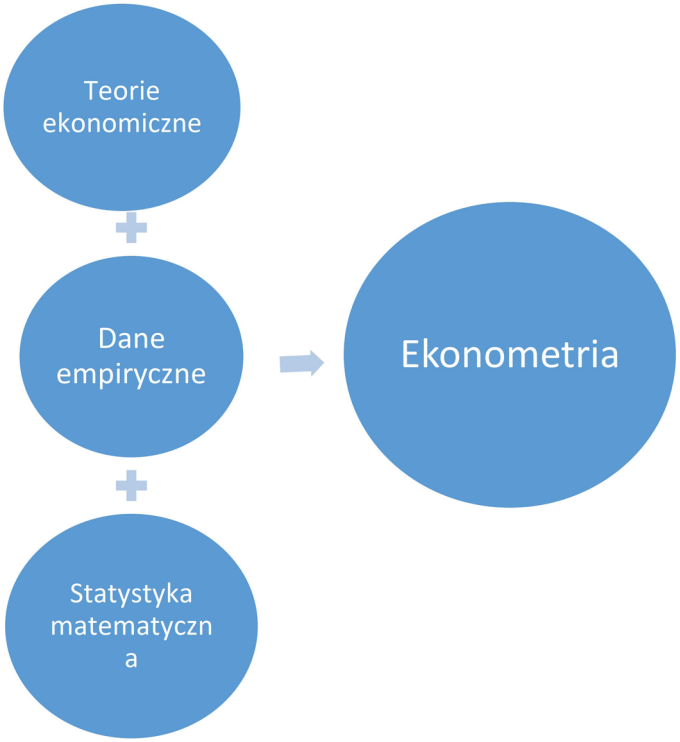
Zmienną Y nazywamy zmienną objaśnianą,
Zmienną X objaśniającą;
α_0, α_1 są nieznanymi parametrami struktury modelu.
Składnik ε wyraża tzw. błąd w równaniu, czyli wpływ na Y czynników nieuwzględnionych w modelu w sposób bezpośredni, takich jak: stan mieszkania, wiek mieszkania itd.

Składnik losowy w modelu wynika z konieczności uwzględnienia:

- wpływu wszystkich czynników mało istotnych niewyspecyfikowanych w równaniu, ale oddziałujących na zmienną endogeniczną,
- różnic pomiędzy postacią analityczną modelu a istniejącą zależnością w rzeczywistości,
- błędów pomiarów zmiennych,
- czynników losowych, wywierających wpływ na zmienną endogeniczną.

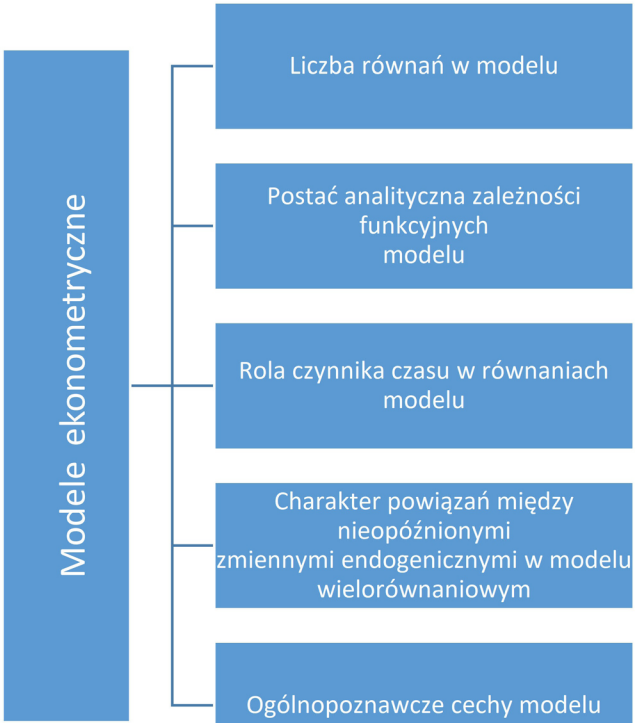
Podejście deterministyczne zakłada, że składnik losowy wyraża efekt oddziaływania na zmienną endogeniczną tych czynników, które nie zostały wyspecyfikowane w modelu (explicite w modelu uwzględnia się tylko zmienne objaśniające najistotniejsze) oraz błędów wynikających z przyjęcia niewłaściwej postaci analitycznej, a także błędów pomiaru wartości zmiennych występujących w modelu (dane statystyczne mogą być obarczone błędami pomiaru). W podejściu indeterministycznym przyczyny występowania składnika losowego upatruje się w tym, że zjawiska losowe są z natury stochastyczne.

Dla zbudowania modelu potrzebna jest znajomość teorii ekonomii, wiedzy z zakresu matematyki i ekonomii oraz ekonomiczna praktyka. Model ekonomiczny cechują wartości poznawcze dla teorii ekonomii oraz znaczenie praktyczne, tak by mógł stać się narzędziem skutecznym w prognozowaniu przyszłych zjawisk. Poza modelem zawiera się w nim również matematyczny opis istniejących prawidłowości ekonomicznych – patrz Rysunek 2.



Rysunek 2. Ekonometria.

Modele ekonometryczne mogą być budowane dla różnych celów. Celami tymi mogą być szczególnie: analiza strukturalna (powiązań strukturalnych), prognozowanie (patrz rozdział Prognozowanie), analizy symulacyjne oraz optymalne sterowanie.



Rysunek 3. Klasyfikacja modeli ekonometrycznych.

Modele ekonometryczne klasyfikujemy ze względu na pięć poniższych kryteriów:

KRYTERIUM 1. Liczba równań w modelu:

- modele jednorównaniowe,
- modele wielorównaniowe, w których każde równanie objaśnia jedną zmienną.

KRYTERIUM 2. Postać analityczna zależności funkcyjnych modelu:

- modele liniowe (wszystkie zależności modelu są liniowe),
- modele nieliniowe, w których chociaż jedna zależność modelu jest nieliniowa.

KRYTERIUM 3. Rola czynnika czasu w równaniach modelu:

- modele statyczne, nie uwzględniają czynnika czasu, wśród zmiennych objaśniających nie występują zmienne opóźnione ani zmienna losowa,
- modele dynamiczne, w których uwzględnia się czynnik czasu. Najlepiej znanym przypadkiem modelu dynamicznego jest model autoregresyjny, w którym wśród zmiennych objaśniających występują jedynie opóźnione w czasie zmienne objaśniane.

KRYTERIUM 4. Ogólnopoznawcze cechy modelu:

- modele przyczynowo-opisowe wyrażające związki przyczynowo-skutkowe między zmiennymi objaśniającymi i objaśnianymi,
- modele symptomatyczne, w których rolę zmiennych objaśniających pełnią zmienne skorelowane z odpowiednimi zmiennymi objaśnianymi, a niewyrażające źródeł zmienności zmiennych objaśnianych.

KRYTERIUM 5. Charakter powiązań między nieopóźnionymi zmiennymi endogenicznymi w modelu wielorównaniowym:

- modele proste,
- modele rekurencyjne,
- modele o równaniach współzależnych.

Przykład

Rozpatrzmy model wzrostu gospodarczego

$$DN_t = aNI_{t-4}^{b} Z_t^{c} \varepsilon_{1t}$$
$$NI_t = dDN_t + \varepsilon_{2t}$$

gdzie:
DN – dochód narodowy,
NI – nakłady inwestycyjne,
Z – zatrudnienie,
a, b, c, d – parametry strukturalne,
ε_1, ε_2 – elementy losowe

Klasyfikacja modelu:
- dwurównaniowy,
- nieliniowy,
- dynamiczny.

Konstruując model ekonometryczny, realizujemy następujące etapy:

1. Określenie celu badania.

Badania ekonometryczne i badanie zjawisk ekonomicznych prowadzi się w celu:
- opisu mechanizmu kształtowania się zjawisk ekonomicznych (cel poznawczy – opisowy),
- oceny zbadanego wcześniej mechanizmu kształtowania się zjawisk ekonomicznych (cel diagnostyczny),
- przewidywania w sensie czasowym i przekrojowym przebiegu zjawisk ekonomicznych (cel predyktywny).

2. Specyfikacja zmiennych modelu.

Ten etap wykonujemy poprzez:
- dobór zmiennych objaśniających,
- zbieranie danych statystycznych,
- wybór zmiennych objaśniających.

Przez dobór zmiennych należy rozumieć merytoryczne propozycje zbioru zmiennych objaśniających. Przy doborze zmiennych do ekonometrycznego modelu związku należy kierować się następującymi względami:
- Zbiór zmiennych objaśniających powinien spełniać przede wszystkim wymagania merytoryczne, które zależą od celu badania. Zmienne powinny być tak dobrane, aby uwzględniały cel badania: opisowy, diagnostyczny czy prognostyczny, gdyż rzadko się zdarza, aby ten sam model wykorzystywany był z uwzględnieniem wszystkich trzech rodzajów celów równocześnie.
- Zmienne objaśniające powinny spełniać rolę przyczyn (co oczywiście nie oznacza, że zawsze powinny być przyczynami w pełnym tego słowa znaczeniu).
- Zmienne objaśniające powinny reprezentować różne aspekty ze sfery przyczyn.
- Zmienne objaśniające powinny być dobrane zarówno ze względu na istniejącą teorię ekonomii, jak i praktykę gospodarczą.

Przez wybór zmiennych objaśniających do modelu ekonometrycznego rozumieć należy taką selekcję (redukcję) zbioru złożonego z „propozycji" zmiennych, aby zbiór zmiennych uwzględnionych w modelu spełniał wymóg przyjętego kryterium formalnego. Do najważniejszych kryteriów formalno-statystycznych stosowanych w metodach wyboru zmiennych należą następujące założenia:
- zmienne występujące w modelu powinny charakteryzować się dużą zmiennością;
- należy zapewnić maksymalne skorelowanie zmiennej objaśnianej ze zmiennymi objaśniającymi;
- zmienne objaśniające nie powinny być istotnie skorelowane między sobą;
- należy dążyć do maksymalnego stopnia dopasowania modelu do rzeczywistych relacji gospodarczych, co wyraża się w maksymalizacji współczynnika determinacji.

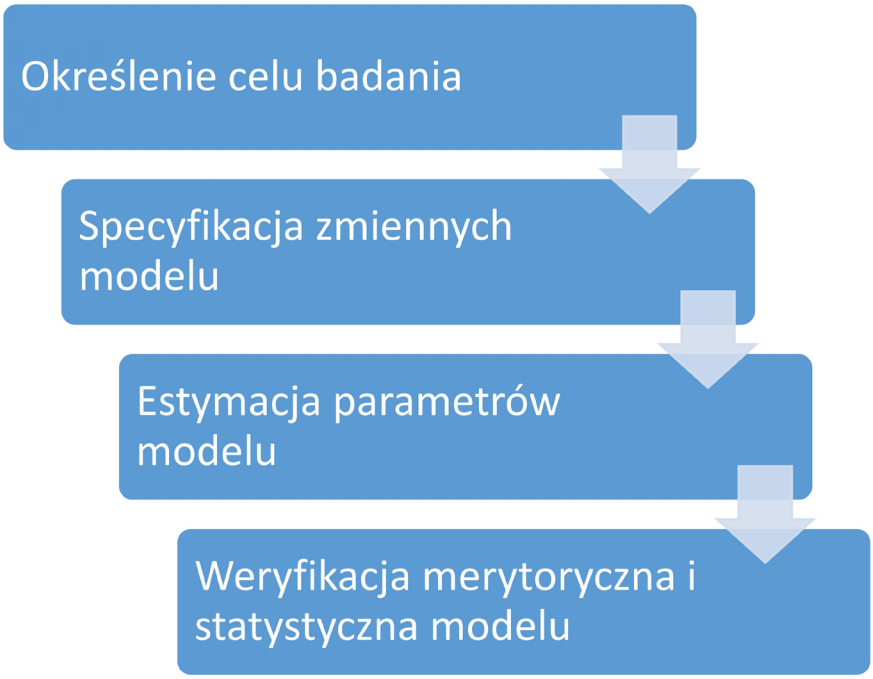
3. Estymacja parametrów modelu.

Ten etap badania sprowadza się zatem do wyboru metody szacunku parametrów modelu ekonometrycznego, a następnie do ich oszacowania.

4. Weryfikacja merytoryczna i statystyczna modelu.

W etapie tym wyróżniamy następujące kroki:

- Badanie dokładności szacunku modelu.
- Badanie stopnia dopasowania modelu do danych empirycznych.
- Badanie statystycznej istotności estymatorów parametrów występujących w modelu.
- Weryfikacja hipotez dotyczących składnika losowego.
- Badania zasadności przyjętej postaci modelu.



Rysunek 4. Etapy budowy modelu ekonometrycznego.



UWAGA

Modele stochastyczne
są to modele uwzględniające składnik losowy, ten typ modeli spotykamy najczęściej w teorii ekonometrii.

Modele deterministyczne
są to modele nieuwzględniające składnika losowego. Przykładem takiego modelu może być wzór konkretnej funkcji, np. funkcji wykładniczej.

1.3. Zmienne modelu

Zmienne modelu możemy podzielić na dwie grupy:

- **zmienne objaśniane** – wyróżnione zjawiska ekonomiczne, które są opisywane (wyjaśniane) przez poszczególne równania modelu;
- **zmienne objaśniające** – zmienne służące do opisu, wyjaśniania zmian zmiennych objaśnianych.

Czasem zdarza się (np. w modelach wielorównaniowych), że zmienne objaśniane pełnią rolę również zmiennych objaśniających. Dlatego stosujemy też inny podział zmiennych.

Zmienne tworzące model ekonomiczny można podzielić na:

- **zmienne endogeniczne** – bieżące i opóźnione, wyjaśniane przez model;
- **zmienne egzogeniczne** – bieżące i opóźnione, których model nie wyjaśnia.

Przykład

Rozpatrzmy model wzrostu gospodarczego

$$DN_t = aNI_{t-4}^{b} Z_t^{c} \varepsilon_{1t}$$
$$NI_t = dDN_t + \varepsilon_{2t}$$

gdzie:
DN – dochód narodowy,
NI – nakłady inwestycyjne,
Z – zatrudnienie,
a, b, c, d – parametry strukturalne,
ε_1, ε_2 – elementy losowe.

Klasyfikacja:
zmienne endogeniczne: DN_t, NI_t, NI_{t-4}
zmienne egzogeniczne: Z_t
zmienne objaśniane: DN_t, NI_t
zmienne objaśniające: NI_{t-4}, Z_t, DN_t
zmienne bieżące: DN_t, Z_t, NI_t
zmienne opóźnione: NI_{t-4}

1.4. Szeregi czasowe

Dane charakteryzujące rozwój zjawisk gospodarczych lub społecznych są odnotowywane i gromadzone wraz z upływem czasu, tworząc ciągi liczbowe określane mianem szeregów czasowych.

Szereg czasowy to uporządkowany zbiór obserwacji statystycznych (zbiór stanów zmiennych), charakteryzujących zmiany poziomu określonego zjawiska w czasie.

Szereg czasowy momentów przedstawia dane charakteryzujące zjawisko (stan obiektu) w określonych **momentach** czasu, którymi są najmniejsze (dalej niepodzielne) jednostki czasu, jak np. kurs euro na zakończenie sesji dla kolejnych dni miesiąca.

Szereg czasowy okresów przedstawia dane charakteryzujące zjawisko (stan obiektu) w określonych okresach czasu. Okres czasu to pewien odcinek, np. dzień, tydzień, miesiąc czy rok. Przykład szeregu czasowego jest podany w Tabeli 1.

Tabela 1. Szereg czasowy okresów.

Miesiąc	Przychód (zł)
Styczeń	23 890
Luty	35 793
Marzec	23 890
Kwiecień	35 793
Maj	13 890
Czerwiec	45 293
Lipiec	50 690
Sierpień	18 793
Wrzesień	21 590
Październik	32 791
Listopad	17 890
Grudzień	35 793

Przykład

Przykładami szeregu są:
- wartości produkcji przemysłowej w latach 2000–2020 w Polsce,
- wartości stopy bezrobocia w Polsce w latach 2010–2020 w kolejnych miesiącach,
- wartości inflacji w latach 2000–2020 w kolejnych miesiącach,
- notowania kursów akcji w Polsce na sesjach roku 2020.

1.5. Dobór zmiennych objaśniających

Przez wybór zmiennych objaśniających do modelu ekonometrycznego rozumieć należy taką selekcję (redukcję) zbioru złożonego z „kandydatek", aby zbiór zmiennych uwzględnionych w modelu spełniał wymóg przyjętego kryterium formalnego. Do modelu powinny zatem wejść takie zmienne, aby miał on sensowną interpretację merytoryczną i aby zapewniał opis zmiennej objaśnianej z założoną z góry dokładnością.

Celem doboru zmiennych i ich selekcji jest identyfikacja zmiennych, których eliminacja poprawia właściwości modelu regresyjnego.

Nadmierna liczba predyktorów jest niekorzystna z uwagi na[1]:
- ryzyko nadmiernej współliniowości predyktorów i związanych z tym problemów,
- wprowadzenie do modelu niepotrzebnej informacji (szumu) i niecelowej utraty stopni swobody, czego skutkiem jest zwiększona wariancja parametrów modelu (pomimo małego obciążenia)[2],

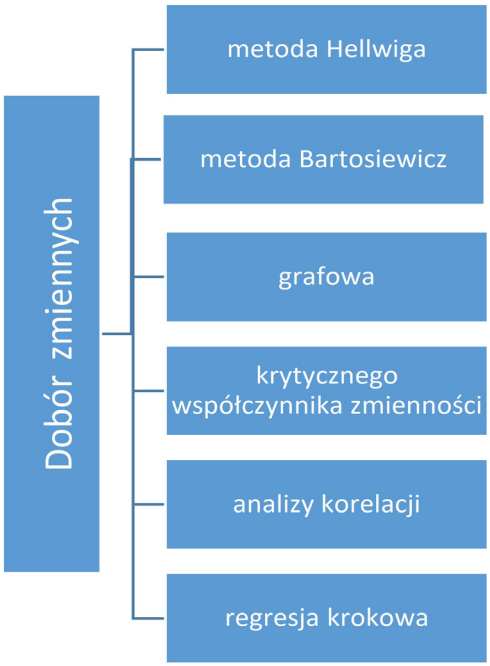
- koszt przygotowania i pozyskania obserwacji rozbudowanych

[1] J.J. Faraway (2002), Practical Regression and Anova using R, http://cran.rproject.org/doc/contrib/Faraway-PRA.pdf (dostęp: 12.10.2012).

[2] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Springer Science+Business Media B.V., New York 2008.

o nadmiarowe predyktory,
- trudności w interpretacji najbardziej znaczącego wpływu predyktorów na zmienną objaśnianą[3].

Zmienne objaśniające do modelu dobieramy według różnych metod, a do najbardziej popularnych należą:
- metoda Hellwiga,
- metoda Bartosiewicz,
- grafowa,
- krytycznego współczynnika zmienności,
- analizy korelacji,
- regresja krokowa w przód i wstecz.



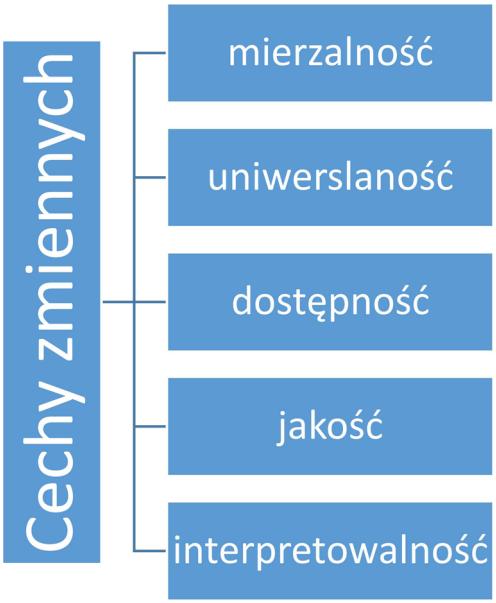
Rysunek 5. Metody doboru zmiennych objaśniających.

Podczas przygotowywania zestawu zmiennych do badania taksonomicznego ważne jest, aby wziąć pod uwagę poniższe kwestie:
- **mierzalność** – zmienne muszą być mierzalne w sposób pośredni

[3] G.S. Maddala, Ekonometria, Wydawnictwo Naukowe PWN, Warszawa 2008.

lub bezpośredni oraz wyrażone przy pomocy wartości względnych albo bezwzględnych,

- **uniwersalność** – każda ze zmiennych powinna mieć powszechnie uznane znaczenie i wagę,
- **dostępność danych liczbowych** – ważne jest to, aby możliwe było uzyskanie kompletnych informacji liczbowych dotyczących wszystkich zmiennych zawartych w badaniu,
- **jakość danych** – należy wziąć pod uwagę, czy dane uwzględnione w badaniu są realne oraz pozbawione błędów przypadkowych, takich jak pomyłki podczas ich zapisu lub niedokładność rachunkowa; dane powinny być dostatecznie dokładne, aby realnie odzwierciedlały wyniki badania,
- **interpretowalność** – wykonując badanie, powinniśmy dobierać takie zmienne, które charakteryzują się odpowiednią wartością merytoryczną. Ich interpretacja powinna być jasno i jednoznacznie określona[4].



Rysunek 6. Cechy zmiennych w badaniu taksonomicznym.

Z formalnego punktu widzenia zmienne powinny odznaczać

[4] A. Zeliaś, Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2000, s. 37.

się następującymi własnościami:

- mieć odpowiednio wysoką zmienność,
- być silnie skorelowane ze zmienną objaśnianą,
- być słabo skorelowane między sobą.

Następnym etapem doboru zmiennych jest eliminacja zmiennych przy wykorzystaniu współczynnika zmienności[5]. **Współczynnik zmienności wyliczamy za pomocą następującego wzoru:**

$$V^k(X_j) = \frac{S(X_j)}{\bar{x}}, j = 1,2, \dots, m,$$

gdzie:

x_i – średnia arytmetyczna wartości j-tej zmiennej, przy czym:

$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$,

$S(x_j)$ – odchylenie standardowe j-tej zmiennej, przy czym:

$$S(x_j) = \left[\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2\right]^{-\frac{1}{2}}.$$

Otrzymany wynik często podawany jest jako wartość procentowa. Najczęściej spotykany w literaturze oraz badaniach poziom krytyczny współczynnika zmienności wynosi 10%. Zmienne poniżej tego poziomu są eliminowane ze zbioru finalnego.

Jedną spośród metod dających możliwość podzielenia zbioru zmiennych jest parametryczna metoda Hellwiga. Podstawą do podjęcia następnego działania eliminacyjnego jest stworzenie macierzy zawierającej współczynniki korelacji liniowej pomiędzy zmiennymi:

[5] M. Gotowska, A. Jakubczak, Zastosowanie wybranych metod do oceny zróżnicowania poziomu życia ludności w Polsce, Artykuł na IX Kongres Ekonomistów Polskich, 2012, s. 2.

$$\boldsymbol{R} = [r_{j,j'}], j, j' = 1,2, \ldots, m,$$

gdzie:
r_{jj} – współczynnik korelacji pomiędzy j-tą i j'-tą dopuszczalnymi zmiennymi diagnostycznymi.

Kolejnym koniecznym działaniem do realizacji metody jest wyznaczenie wartości krytycznej współczynnika korelacji. Przyjmuje się, że zmienne osiągające wartości powyżej poziomu granicznego są między sobą istotnie skorelowane. Wartość progowa współczynnika może być wyznaczana na kilka sposobów. Jej wartość może zostać wyznaczona przez badacza lub wyliczona za pomocą następującego wzoru:

$$r^* = \sqrt{\frac{t_\alpha^2}{t_\alpha^2 + n - 2'}}$$

gdzie:
t_α – wartość krytyczna odczytana z tablic rozkładu t-studenta (patrz Załącznik 1) dla n-2 stopni swobody i dla z góry wyznaczonego poziomu istotności α.[6]

1.5.1. Metoda Hellwiga

Poniżej opisano dokładnie **algorytm metody parametrycznej Hellwiga**. Za pomocą tej metody można wydzielić zmienne wiodące (centralne) i tzw. zmienne satelitarne; prowadzi to do zmniejszenia liczby zmiennych do takich, które w danym analizowanym procesie mają charakter zmiennych wiodących.

1. Wyznaczamy sumy wartości bezwzględnych każdej z kolumn macierzy korelacji:

[6] J. Pociecha, Metody taksonomiczne w badaniach społeczno-ekonomicznych, Państwowe Wydawnictwo Naukowe, Warszawa 1988, s. 141.

$$R_{j'} = \sum_{j=1}^{m} |r_{jj'}| \ , j' = 1,2, \dots, m.$$

2. Znajdujemy kolumnę o największej sumie wartości $R_{j'}$:

$$R_{j_0} = \max_{j'} \{R_{j'}\}$$

3. Z kolumny R_{j0} wyłaniamy elementy o większych wartościach bezwzględnych niż wartość krytyczna współczynnika r*:

$$|r_{jj'_0}| > r^*, j' = 1,2, \dots, m,$$

oraz wyznaczamy wiersze, które odpowiadają tym elementom. Dopuszczalna zmienna, która odpowiada wyróżnionej kolumnie, jest nazywana zmienną centralną, natomiast zmienne, które odpowiadają wyznaczonym wierszom, za jej zmienne satelitarne.

4. Następnie dokonujemy redukcji macierzy korelacji poprzez wykreślenie kolumny oraz wierszy, które odpowiadają wyznaczonej zmiennej centralnej, a także jej satelitarnym zmiennym.

5. Następnie powtarzamy powyższe kroki, posługując się zredukowanymi macierzami korelacji, aż do momentu wyczerpania analizowanego zestawienia zmiennych diagnostycznych[7].

1.5.2. Metoda Pawłowskiego

Metoda Pawłowskiego to metoda maksymalnego wyznacznika macierzy momentów zmiennych objaśniających. W metodzie tej mamy przygotowaną listę m potencjalnych zmiennych objaśnia-

[7] T. Panek, J. Zwierzchowski, Statystyczne metody wielowymiarowej analizy porównawczej: teoria i zastosowania, Oficyna Wydawnicza SGH, Warszawa 2013, s. 23.

jących i z góry ustaloną liczbę k zmiennych, które mają wejść do modelu, przy czym k<m.

Ze zbioru m potencjalnych zmiennych objaśniających rozpatrujemy wszystkie z możliwych kombinacji zmiennych. Kombinacje te będą tworzyć zbiór P_k kombinacji dopuszczalnych, z których szukamy najlepszą kombinację zmiennych objaśniających, tzn. taką, w której zmienne objaśniające są możliwie słabo skorelowane.

W pierwszej kolejności wyznaczamy macierz korelacji R. Następnie rozpatrujemy wyznaczniki poszczególnych podmacierzy otrzymanych z macierzy R, w której zachowano tylko wiersze i kolumny o numerach równych numerom zmiennych w rozpatrywanej kombinacji. Za optymalną kombinację uważa się taką kombinację zmiennych, dla której odpowiedni podwyznacznik jest największy (zmienne X_1, X_2, ..., X_k są tym słabiej skorelowane między sobą, im większy i bliższy jedności jest odpowiedni wyznacznik macierzy korelacji).

1.5.3. Regresja krokowa

Poznamy teraz kolejną metodę wyboru najlepszego podzbioru zmiennych niezależnych – jest to **regresja krokowa**.



Rysunek 7. Metody wyboru podzbioru zmiennych zależnych.

Ze względu na kierunek przeszukiwania modeli wyróżniamy cztery typy tej metody:

1. **Selekcja postępująca:**
   - wychodzimy od modelu tylko z wyrazem wolnym,
   - dodajemy kolejno zmienne do modelu aż do braku możliwości ich dodania (z powodu braku zmiennych spełniających kryterium dodania lub z powodu wyczerpania zmiennych, czyli dojścia do modelu pełnego, tj. ze wszystkimi zmiennymi).

2. **Eliminacja wsteczna:**
   - wychodzimy od modelu pełnego, tj. ze wszystkimi zmiennymi,
   - usuwamy kolejno zmienne z modelu aż do braku możliwości ich usunięcia (z powodu braku zmiennych spełniających kryterium usunięcia lub z powodu wyczerpania zmiennych, czyli dojścia do modelu tylko z wyrazem wolnym).

3. **Regresja krokowa postępująca:**
   - wychodzimy od modelu tylko z wyrazem wolnym,
   - dodajemy jedną zmienną do modelu,
   - usuwamy kolejno z modelu zmienne aż do braku możliwości ich usunięcia,
   - powtarzamy ostatnie dwa kroki aż do braku możliwości dodania zmiennej do modelu (z powodu braku zmiennych spełniających kryterium dodania lub z powodu wyczerpania zmiennych).

4. **Regresja krokowa wsteczna:**
   - wychodzimy od modelu tylko z wyrazem wolnym,
   - usuwamy jedną zmienną do modelu,
   - dodajemy kolejno do modelu zmienne aż do braku możliwości ich dodania,
   - powtarzamy ostatnie dwa kroki aż do braku możliwości usunięcia zmiennej z modelu.

1.6. Postać analityczna modelu

Z poprzedniego rozdziału wiemy już, jak dobrać zmienne objaśniające do modelu. Kolejnym krokiem będzie **wybór postaci modelu ekonometrycznego**. W przypadku gdy mamy do czynienia z jedną zmienną

objaśniającą X, zazwyczaj stosujemy jedną z poniższych postaci modelu. Najczęściej spotykane w praktyce funkcje omówione są poniżej.

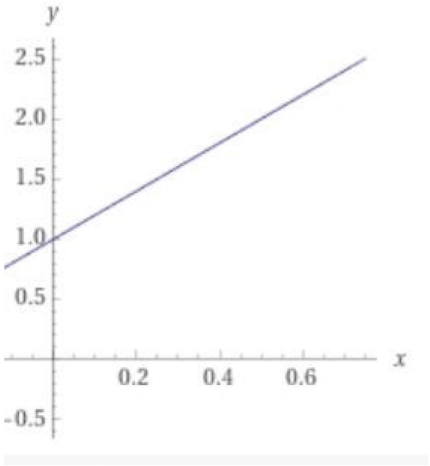
1.6.1. Funkcja liniowa

WAŻNE

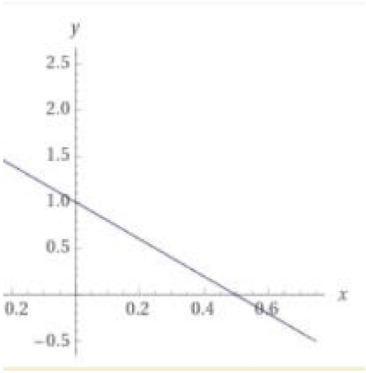
Wzory:

$$Y_t = a \cdot X(t) + c$$
$$a > 0 \textit{ lub } a < 0$$

Parametr c jest interpretowany jako poziom zmiennej objaśnianej Y przy zerowym poziomie zmiennej objaśniającej X. Interpretacja parametru a to: wzrost wartości zmiennej objaśniającej X o jedną jednostkę spowoduje wzrost (a>0) lub spadek (a<0) wartości zmiennej objaśnianej Y o a jednostek.



Rysunek 8. Funkcja liniowa dla a>0.



Rysunek 9. Funkcja liniowa dla a<0.

1.6.2. Funkcja wykładnicza

WAŻNE

Wzory

Funkcja wykładnicza ma postać:

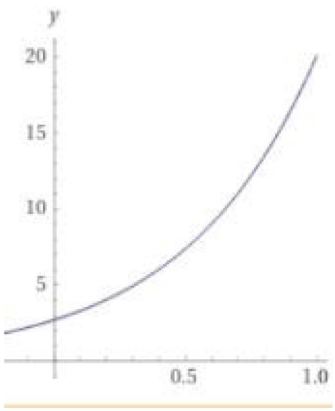
$$Y_t = a^{X(t)} \cdot c$$
$$a > 0$$

Funkcja ta znajduje najczęściej zastosowanie jako **model ekonometryczny, w którym występuje tylko jedna zmienna objaśniająca X.**

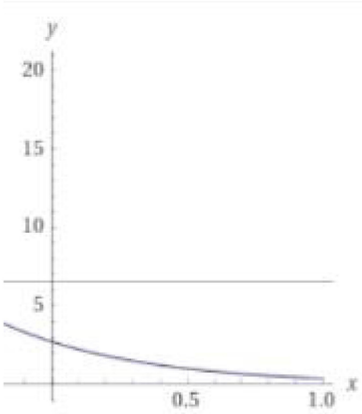
Parametr c jest interpretowany jako poziom zmiennej objaśnianej Y przy zerowym poziomie zmiennej objaśniającej X. Parametr a w modelu funkcji wykładniczej nazywany jest stopą wzrostu. Intepretujemy go w następujący sposób: wzrost wartości zmiennej objaśniającej X o jedną jednostkę powoduje zmianę poziomu zmiennej

objaśnianej Y o (a-1)100%. W praktyce wartość parametru a wynosi zazwyczaj 10 lub jest równa liczbie e (Eulera) i przyjmuje wtedy postać:

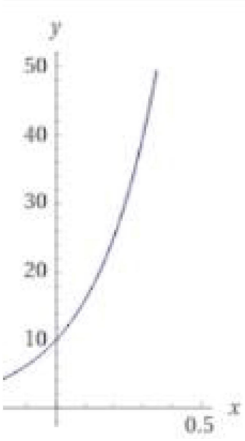
$$Y_t = e^{aX(t)+c} \text{ lub } Y_t = 10^{aX+c}$$



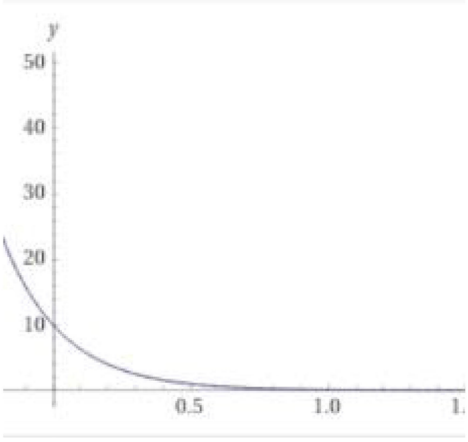
Rysunek 10. Funkcja wykładnicza $Y = e^{aX+c}$ dla a>0.



Rysunek 11. Funkcja wykładnicza $Y = e^{aX+c}$ dla a<0.



Rysunek 12. Funkcja wykładnicza $Y = 10^{aX+c}$ dla a>0.



Rysunek 13. Funkcja wykładnicza $Y = 10^{aX+c}$ dla a<0.

Funkcja wykładnicza znajduje najczęściej zastosowanie jako **model tendencji rozwojowej** (w którym występuje tylko jedna zmienna objaśniająca – zmienna czasowa t):

$$Y = \alpha_0 \alpha_1^t e^{\vartheta} \qquad \text{dla } \alpha_1 > 0$$

- parametr α_0 interpretowany jest jako średni poziom zmiennej Y w roku poprzedzającym badanie (t=0)
- $\alpha_1$nazywane jest średnioroczną stopą wzrostu (spadku) badanego zjawiska w przedziale czasu [1, n], tzn. co roku wartość Y wzrasta (spada) średnio o (α_1 – 1) 100%.
  lub

$$Y = \beta_0 \cdot \beta_1^X \cdot e^{\varepsilon}$$

gdzie:
Y – zmienna objaśniana
X – zmienna objaśniająca
β_α *i* β_1– parametry strukturalne
ε – składnik losowy
e – liczba Nepera

Model trendu (tendencji rozwojowej) należy do szczególnej klasy modeli ekonometrycznych, w których zmienność zmiennej objaśnianej opisywana jest przez specyficzną zmienną objaśniającą, jaką jest czas.

Na ogół modele te nie wyjaśniają mechanizmu kształtowania się rozpatrywanej zmiennej objaśnianej, lecz obrazują kształtowanie się tej zmiennej w czasie.

Wykładnicza linia trendu to linia krzywa, która jest najbardziej użyteczna, gdy dane wzrastają lub spadają według coraz wyższych wartości. Najczęściej zastosowanie znajduje w nauce. Może opisywać populację, która szybko rośnie w kolejnych pokoleniach.

1. **funkcja logistyczna**

$$Y_t = k + a \cdot logt$$
α > 0

2. **funkcje hiperboliczne**

$$Y_t = a \cdot t^{-1} + c$$
α > 0 lub α < 0

t – zmienna objaśniająca, oznaczająca czas (miesiące, kwartały, lata).

Wymienione funkcje znajdują zastosowanie jako model tendencji rozwojowej (np. funkcja wykładnicza oraz funkcja logistyczna).

1.6.3. Funkcje Törnquista

Szwedzki ekonomista Törnquista badał zależność pomiędzy wydatkami na zakup dóbr a wielkością dochodów konsumentów. Zaproponował on wymierny model krzywej popytu jako funkcji dochodu konsumentów. W zależności od rodzaju potrzeb, które są zaspokojone dzięki konsumpcji określonych dóbr, wyróżnia się następujące **funkcje Törnquista**:

- **funkcja Törnquista I rodzaju** – popyt na dobra podstawowe, pierwszej potrzeby, np. wyżywienie, odzież, stałe opłaty,
- **funkcja Törnquista II rodzaju** – popyt na dobra wyższego rzędu,
- **funkcja Törnquista III rodzaju** – popyt na dobra luksusowe.

Funkcje Törnquista należą do rodziny krzywych Engla. Krzywa Engla to wykorzystywana w ekonomii zależność pomiędzy dochodem konsumenta a ilością konsumowanego przez niego dobra, przy założeniu stałych cen wszystkich towarów oraz innych zmiennych (ceteris paribus). Graficznie krzywą Engla zazwyczaj przedstawia się w układzie współrzędnych kartezjańskich z dochodem na osi rzędnych i ilością konsumowanego dobra na osi odciętych.

Funkcja Törnquista I typu wyraża zależność między popytem na dobra podstawowe (Y) a dochodami konsumentów (X).

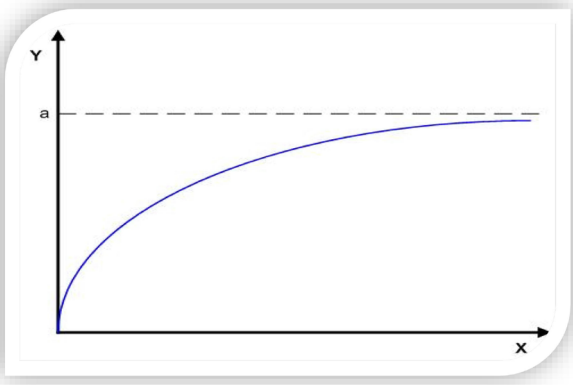
WAŻNE

Wzory

$$T1 : y = \frac{aX}{b+X} \quad X, a\ b > 0$$

gdzie
X – dochód,
Y – wydatki,
parametry a. b są pewnymi stałymi (przy czym parametry te dla każdej funkcji mogą być różne)
a – poziom nasycenia, czyli poziom, do którego wydatki na dane dobro rosną, ale którego nigdy nie przekroczą

Popyt rośnie wraz ze wzrostem dochodów i stabilizuje się na poziomie równym ɑ.Wykres ma asymptotę poziomą. Z drugiej strony popyt spada do zera wraz ze spadkiem dochodów, ale nie znika dla niezerowych dochodów.



Rysunek 14. Wykres funkcji Törnquista I typu.

Dobra pierwszej potrzeby nabywane są już przy najniższych dochodach. Wydatki konsumentów są rosnącą funkcją dochodów, czyli zwiększają się wraz ze wzrostem wielkości dochodów. Jednak wzrost ten mimo wzrostu dochodu jest coraz wolniejszy.

Krzywa Y ma asymptotę poziomą Y=a, co oznacza, że istnieje poziom nasycenia, czyli choćby dochód rósł nieograniczenie, to wydatki nie przekroczą tego poziomu.

Elastyczność popytu względem dochodu wyraża się wzorem:

$$E_X Y = X \cdot \frac{X+b}{aX} \cdot \frac{ab}{(X+b)^2} = \frac{b}{X+b}$$

Jeśli dochód wzrośnie o 1%, to wydatki na dane dobro wzrosną o (E) %.



Rysunek 15. Wykres elastyczności dla funkcji Törnquista I typu.

Z wykresów można zaobserwować, że gdy dochód konsumenta rośnie, to wartość elastyczności maleje do zera.

Popyt dochodowy funkcji ***Y*** jest więc nieelastyczny, ponieważ elastyczność jest zawsze mniejsza od 1. Oznacza to, że wzrost dochodów o 1% powodu wzrost wydatków (popytu) na określone dobro o wartość mniejszą niż 1%.

Przy odpowiednio dużych dochodach wzrost konsumpcji na określone dobro zanika, czyli konsumenci posiadający wyższe dochody

w znacznie większym stopniu zaspokajają dobra podstawowe niż konsumenci o niższych dochodach, u których reakcja na wzrost dochodów jest znacznie silniejsza.

Funkcja Törnquista II typu

Wzory

$$T2: y = \frac{a(x-c)}{x+b},\ a,c,b>0,\ x \geq c,$$

gdzie:
x – dochód
a; b; c – stałe

Wykres omawianej funkcji jest rosnącą funkcją dochodów względem wydatków na dobra wyższego rzędu. Funkcja jest określona dla x > c, co oznacza, że wydatki na dobra wyższego rzędu występują, jeśli zostaną zaspokojone potrzeby na dobra podstawowe.

Elastyczność powyższej funkcji wyraża się wzorem:

$$E_{x0}T_2 = x_0 \cdot \frac{b+c}{(x_0-c)(x_0+b)}, x_0 > c.$$

Elastyczność funkcji T2 jest funkcją malejącą. Mamy:

$$\lim_{x\to\infty} x_0 \cdot \frac{b+c}{(x_0-c)(x_0+b)} = 0$$

a więc wraz ze wzrostem dochodów konsumenta elastyczność funkcji T2 maleje do zera. Zatem procentowy wzrost dochodów powoduje coraz mniejszy procentowy wzrost wydatków (popytu) na dane dobro wyższego rzędu. Możemy też zauważyć, że przy dość dużych dochodach wzrost konsumpcji na dane dobro zanika.

Funkcja Törnquista III typu

Wzory

$$\mathbf{T3}: Y = \frac{a(X-c)}{X+b},\ a > 0;\ b > 0;\ c > 0;\ X > c$$

gdzie:
x – dochód
a – punkt nasycenia
b – szybkość, z jaką popyt osiąga punkt nasycenia
c – minimalny dochód, w którym pojawia się popyt na określone dobro

W przeciwieństwie do funkcji Törnquista I i II rodzaju **popyt na dobra luksusowe rośnie wraz ze wzrostem dochodów bez punktu nasycenia.**

Wraz ze wzrostem dochodów konsumenta popyt zwiększa się coraz wolniej, co sprawia, że dochody osiągają wysoki poziom. Do tego stopnia, że może on pozwolić sobie na zakup dóbr wyższego rzędu. Zabieg ten ma miejsce, kiedy to petent nasyci dobra i usługi niższego rzędu oraz dobra podstawowe (pierwszej potrzeby).

Elastyczność wyraża się wzorem:

$$E_{fT3} = \frac{x^2 + 2abX - bc}{(b+X)\cdot(x-c)}$$

Obliczając granicę E_{fT3} E_{fT3} , otrzymujemy:

$$\lim_{x\to\infty} \frac{X^2 + 2bX - bc}{(X-c)(X+b)} = 1$$

co oznacza, że dopiero dla odpowiednio dużego dochodu konsumenta elastyczność T3 ma wartość blisko, a nawet równą 1. Zatem dla dóbr luksusowych procentowy wzrost dochodów konsumentów o 1% powoduje wzrost wydatków (popytu) o więcej niż 1% na dane dobro.

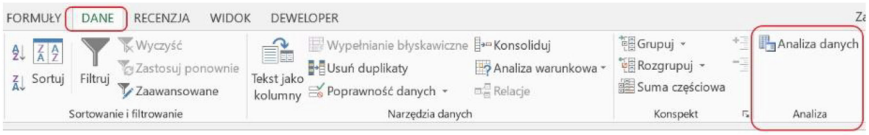
1.7. Oprogramowanie w ekonometrii

Prowadzenie analiz ekonometrycznych, a zwłaszcza opracowanie i analiza dużych zbiorów danych, wymaga wykorzystania komputerów. Zastosowanie programów komputerowych pozwala zredukować czas potrzebny na pracochłonne obliczenia do niezbędnego minimum.

Na szczególną uwagę zasługuje **arkusz kalkulacyjny Microsoft Excel**, który dzięki wbudowanym funkcjom statystycznym oraz opcji „Analiza danych" może znacznie ułatwić analizę statystyczną. Wykorzystanie Excela sygnalizuje możliwości programów komputerowych do obliczeń statystycznych i powinno stanowić pierwszy krok do samodzielnego stosowania profesjonalnych pakietów statystycznych oraz ekonometrycznych (R, Gretl, Statistica, SPSS, SAS i innych). Należy pamiętać, że większość metod ekonometrycznych jest ściśle powiązana z analizami statystycznymi. Dlatego większość oprogramowań statystycznych znajduje zastosowanie w ekonometrii. Pakiety te są stosunkowo proste w obsłudze, jednak wymagają pewnej wiedzy ze statystyki, aby można było poprawnie używać zawarte w nich procedury oraz interpretować uzyskane wyniki. Wymienione oprogramowanie jest omówione poniżej.

Funkcje statystyczne w istotnym stopniu wykorzystywane są w różnych programach komputerowych. Przykładem może być tutaj arkusz kalkulacyjny Excel. Wybór arkusza kalkulacyjnego jako narzędzia prezentacji metod statystycznych jest podyktowany wieloma względami. Znajomość arkuszy kalkulacyjnych jest coraz powszechniejsza ze względu na ich uniwersalność zastosowań. Każda nowsza wersja Excela ma coraz więcej i coraz bardziej rozbudowane funkcje statystyczne.

Arkusze kalkulacyjne mogą zastąpić w wielu przypadkach specjalistyczne pakiety statystyczne. Arkusz kalkulacyjny umożliwia także prezentację metod statystycznych w sposób tradycyjny, za pomocą rozbudowanej tablicy roboczej, z ukazaniem etapów obliczeń. Tej właściwości nie mają profesjonalne pakiety statystyczne. Dużym udogodnieniem dla użytkujących arkusz kalkulacyjny do analizy statystycznej jest **system funkcji, między innymi matematycznych i statystycznych, oraz opcja Analiza danych**. Opcja Analiza danych jest wybierana z zakładki Dane/Analiza – patrz rysunek poniżej.



Rysunek 16. Moduł Analiza danych pakietu Excel.

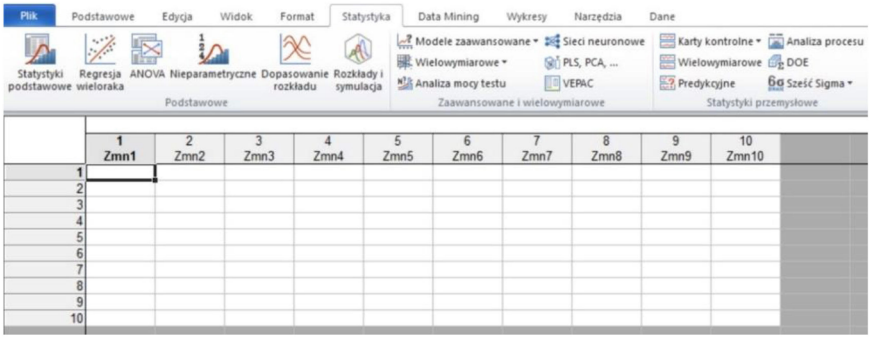
Opcja Analiza danych, może w wielu przypadkach zastąpić profesjonalny pakiet statystyczny. Dzięki niej można m.in. dokonać analizy wariancji, korelacji, regresji, testy statystyczne oraz statystykę opisową. Szczególne często w analizach ekonometrycznych korzystamy z opcji Wygładzanie wykładnicze czy Korelacja.



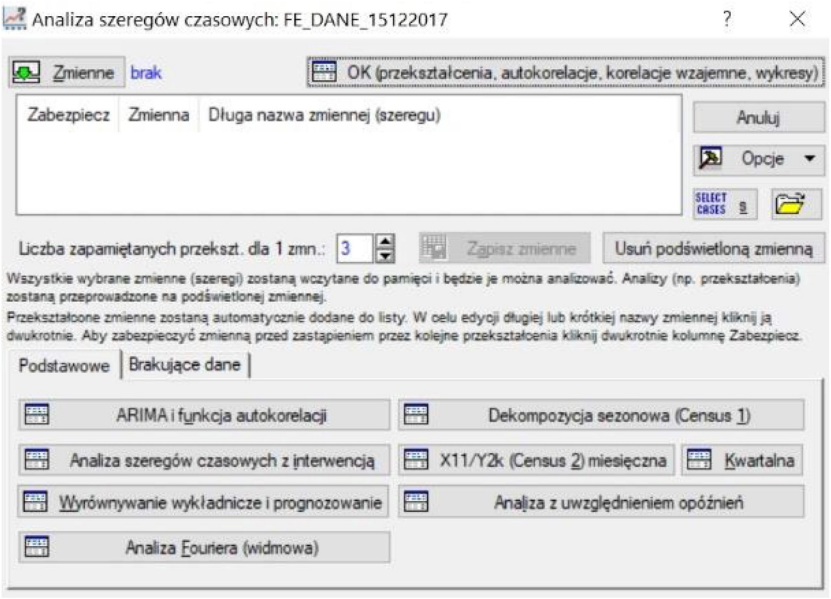
Rysunek 17. Dostępne funkcje Analizy danych.

Znaczenie **komputerowych pakietów statystycznych** jest istotne szczególnie w rozbudowanych firmach przy przeprowadzaniu różnych analiz statystycznych oraz ekonometrycznych. Wówczas korzysta się z takich fachowych pakietów statystycznych jak Statistica.

Statistica to wysokiej jakości system do analizy danych, graficznej prezentacji i zarządzania danymi dla korporacji i pojedynczych analityków. Jest to szeroki wybór najbardziej zaawansowanych metod spośród wszystkich dostępnych na rynku narzędzi do modelowania i prognozowania, z automatyzacją wyboru modeli i interaktywnymi narzędziami wizualizacji. Dodatkowo korzysta się również z tzw. systemów korporacyjnych, czyli zintegrowanych systemów analitycznych.



Rysunek 18. Interfejs programu Statistica.



Rysunek 19. Moduł Analiza szeregów czasowych.

Kolejne **oprogramowanie to SPSS**. Jest to zintegrowana rodzina produktów przydatnych na wszystkich etapach procesu analitycznego – od planowania i zbierania danych po analizowanie, raportowanie i wdrożenie. Podobnie jak Statistica, SPSS jest oprogramowaniem płatnym.

R to nazwa języka programowania, platformy programistycznej jak i całego projektu, w ramach którego rozwijane jest zarówno środowisko, jak i język. R bardzo często jest nazywany pakietem statystycznym, ponieważ dostępnych jest szereg pakietów i funkcji wykorzystywanych do zastosowań statystycznych. Okazuje się, że R znajduje znacznie więcej zastosowań: automatyczne generowanie raportów, wysyłanie maili, renderowanie trójwymiarowych animacji.

R jest projektem opartym o licencję GNU GPL, wyposażonym w bardzo szczegółową dokumentację dostępną w Internecie. Język R jest językiem interpretowanym, korzystanie z niego sprowadza się do podania szeregu komend, które mają zostać kolejno wykonane. Komendy mogą być zebrane w formie skryptu, który zostaje następnie wykonywany krok po kroku.

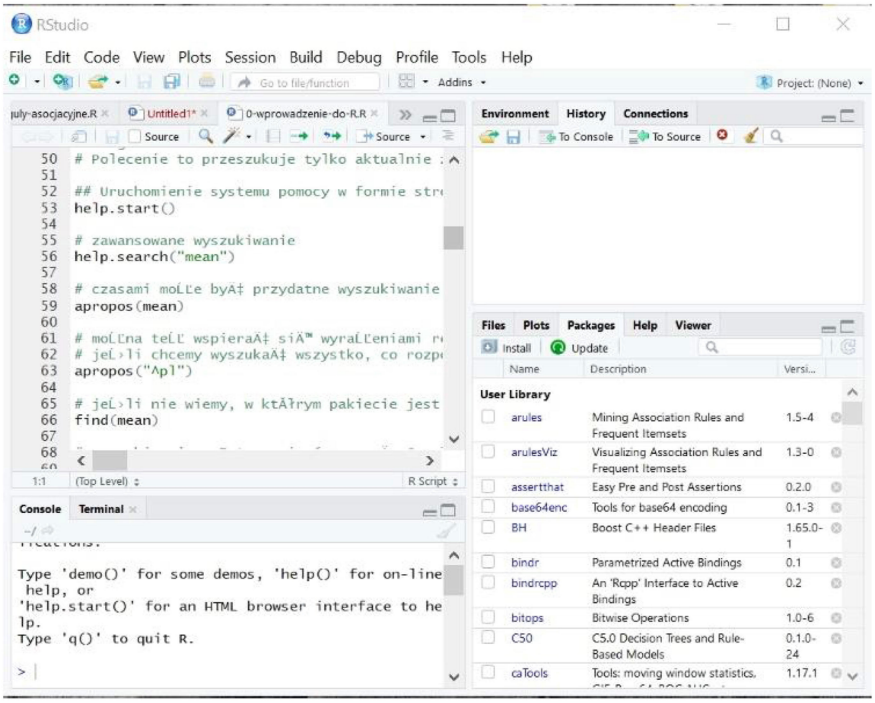
Główne elementy charakterystyczne tego języka programowania:

- Jest to bardzo elastyczne (darmowe) środowisko analityczne o bogatej funkcjonalności, stosowane w wielu badawczych i praktycznych pracach dotyczących analizy danych i odkrywania wiedzy.
- R jako język programowania dostarcza wbudowanych operacji ułatwiających przetwarzanie tabelarycznych zbiorów danych, mechanizmów graficznego opisu danych, bogatych bibliotek funkcji analitycznych, obejmujących szeroki zakres metod statystycznych i metod odkrywania wiedzy oraz – co niezwykle ważne – interaktywny interpreter poleceń i (dla niektórych platform) graficzny interfejs użytkownika.
- Wszystkie niezbędne informacje o języku R znajdziemy na stronach CRAN (The Comprehensive R Archive Network), gdzie dostępne są pakiety źródłowe, dokumentacja oraz obszerne zasoby bibliotek.

Dla większości systemów operacyjnych pakiet R jest dostępny zarówno w postaci źródłowej, jak i skompilowanej. Najlepiej skorzystać ze skompilowanego pliku instalacyjnego, który można

pobrać z jednego z serwerów, których lista znajduje się na stronie: http://cran.r-project.org/mirrors.html.

Do tworzenia i edytowania skryptów w R służą specjalne edytory, przykładem może być **RStudio**. Jego instalacja jest możliwa na systemie Windows, Linux oraz MacOS. Można go pobrać ze strony: www.rstudio.com. Edytor umożliwia wykonywanie pojedynczych komend wpisywanych prosto do konsoli albo wykonywanie całych skryptów przygotowanych uprzednio w oknie skryptowym.



Rysunek 20. Okno RStudio.

Całe okno RStudio podzielone jest na 4 panele główne, które można konfigurować, wybierając z menu: Tools i Global Options. W ten sposób wywołamy okno Options, widoczne na powyższym rysunku. W powyższej konfiguracji prawy górny panel służy do pisania kodu programu. Linijkę kod możemy przekazać do konsoli R (znajdującej się w lewym górnym panelu) za pomocą kombinacji klawiszy: CTRL+Enter. Jeśli chcemy, aby tych linii było więcej, wtedy należy je zaznaczyć i użyć tego skrótu. Z innymi dostępnymi skrótami klawia-

turowymi można się zapoznać, wybierając z menu Help i Keyboard Shortcuts.

Ostatnie omawiane oprogramowanie to **program GRETL**, autorstwa Allina Cottrella z Uniwersytetu Wake Forest w Północnej Karolinie w Stanach Zjednoczonych. Jest rozwijanym od kilku lat pakietem ekonometrycznym. Należy do oprogramowań Powszechnej Licencji Publicznej (GNU), czyli o swobodnym dostępie dla wszystkich użytkowników. Program ten, niezwykle popularny i powszechnie stosowany na świecie, zawiera podstawowe procedury i metody ekonometryczne, przygotowany jest tak, aby można go było wykorzystać na zajęciach z ekonometrii.

1.8. Pytania sprawdzające



1. Co to jest ekonometria?

……
……
……
……
……
……
……

2. Opisz procedurę ekonometryczną.

……
……
……
……
……
……

3. Co to jest zmienna objaśniana i objaśniająca?

……
……
……
……
……
……
……

4. Co to jest zmienna endogeniczna i egzogeniczna?

……
……
……
……
……
……



5. Podaj przykłady procesu stochastycznego.

……
……
……
……
……
……
……
……
……
……

6. Dlaczego w modelach ekonometrycznych występuje składnik losowy?

……
……
……
……
……
……
……
……
……
……

7. Omów klasyfikację modeli ekonometrycznych.

……
……
……
……
……
……
……
……
……
……

8. Wymień kolejne etapy konstrukcji modelu ekonometrycznego.



9. Co to jest szereg czasowy? Podaj przykład szeregu czasowego.

10. Jakie znasz metody doboru zmiennych objaśniających?

1 2 3 4

2. Jednorównaniowy liniowy model ekonometryczny



Przedmiotem tego rozdziału jest ekonometryczny model jednorównaniowy. Jest to model, który wyjaśnia zachowanie jednej zmiennej, zwanej zmienną objaśnianą lub zależną. Rozpatrujemy liniową zależność zmiennej objaśnianej od zmiennych objaśniającychi składnika losowego

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_{2t2} + ... + \beta_k x_{tk} + \xi_{t,}\ t = 1,.....,T$$

czyli:

$$y_t = \beta_0 + \sum_{t=1}^{k} \beta_t x_{ti} + \xi_t$$

wektor obserwacji na zmiennej endogenicznej oraz macierz obserwacji na zmiennych egzogenicznych przyjmują postacie odpowiednio:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ : \\ y_T \\ x_{TK} \end{pmatrix} , \qquad x = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{2k} \\ : & : & : & : & : \\ 1 & x_{T1} & x_{T2} & ... & \end{pmatrix}$$

gdzie:
k – liczba zmiennych objaśniających równania
β – parametr strukturalny modelu
k+1 – liczba parametrów strukturalnych
T – liczba obserwacji

Interpretacja parametru β_i jest następująca: jeżeli zmienna x_t wzrośnie o jednostkę, a pozostałe zmienne objaśniające nie ulegną zmianie, to zmienna endogeniczna zmieni się średnio o β_i jednostek.

Naszym celem jest **oszacowanie parametrów modelu** na podstawie posiadanych informacji statystycznych, dotyczących wartości zmiennych występujących w modelu. Zakładamy, że dysponujemy n-elementowymi szeregami czasowymi obserwacji dla wszystkich

zmiennych modelu. W przypadku danych przekrojowych n oznacza liczbę obiektów. Najlepiej znaną i najczęściej stosowaną w praktyce metodą estymacji nieznanych parametrów strukturalnych modelu jest **metoda najmniejszych kwadratów** (MNK).

Przykład

Niech:
$\mathbf{y}_t$ – miesięczne wydatki na prasę
$\mathbf{x}_t$ – miesięczne dochody

Wtedy równanie

$$y_t = \beta_0 + \beta_1 x_t + \xi_t$$

wyrażą liniową zależność wydatków na prasę w zależności od dochodu.

2.1. Metoda najmniejszych kwadratów

Jest to najprostsza wersja modelu liniowego i chociaż najczęściej nie wystarcza do poprawnego opisu zjawisk ekonomicznych, jego zaletą jest prostota zrozumienia i interpretacji. Metoda ta wykorzystywana jest **do szacowania parametrów modeli liniowych oraz modeli nieliniowych, które można sprowadzić do postaci liniowej**.

Niech będzie dany ogólny model ekonometryczny:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_k X_k + \varepsilon$$

Uwzględniając realizacje zmiennych, często zapisuje się go następująco:

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \cdots + \alpha_k x_{kt} + \varepsilon_t$$

gdzie:

Y, y_t – zmienna objaśniana, zależna, endogeniczna; realizacje zmiennej objaśnianej w okresie t

X_k, X_{kt} – zmienne objaśniające, niezależne; realizacje zmiennych objaśniających w okresie t

ε_t – składnik losowy

t – kolejne realizacje (obserwacje), $t \in 1, 2, \ldots, n$

W zapisie macierzowo-wektorowym model ekonometryczny ma postać:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Zmienne y_i oraz x_i są nam znane (ich wartości zostały zebrane np. w wyniku przeprowadzenia badania ankietowego lub innych dostępnych danych). Wartości parametrów strukturalnych β_k nie są znane, więc naszym zadaniem jest wyznaczenie takich wartości estymatorów parametrów strukturalnych, które będą najbliższe tym parametrom.

Model ekonometryczny po oszacowaniu parametrów strukturalnych a_i będzie miał postać:

$$\hat{y}_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \ ... + \ a_k x_{kt}$$

gdzie: $t \in 1, 2, ... , n$.

W zapisie macierzowym:

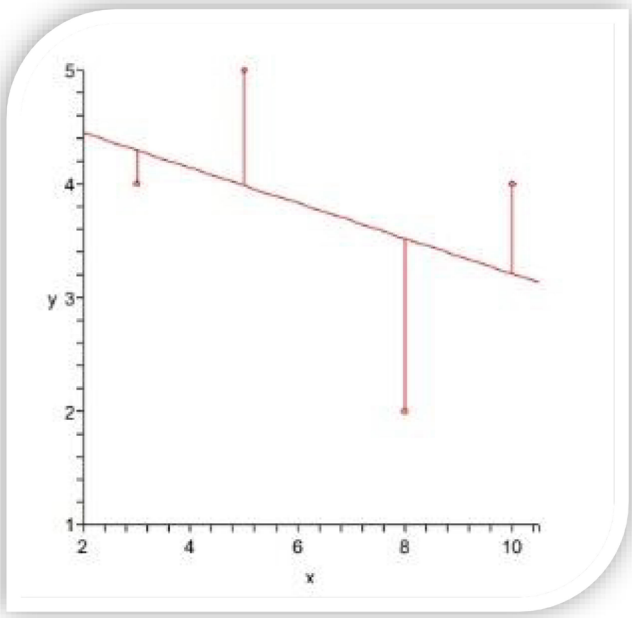
$$\hat{y} = Xa$$

Spośród różnych dostępnych metod niewątpliwie najpopularniejszą jest **minimalizacja sumy kwadratów odległości punktów od prostej**. Stąd też wzięła się jej nazwa – **Metoda Najmniejszych Kwadratów (MNK)**. Używając tej metody, możliwa jest estymacja parametrów modelu stochastycznego. Wzór pozwalający otrzymać oszacowania parametrów równania przy pomocy MNK jest następujący:

$$a = (X^T X)^{-1} X^T Y$$

gdzie:
X – macierz obserwacji na zmiennych objaśniających
y – wektor obserwacji na zmiennej objaśnianej
X^T – macierz transponowana do macierzy X
$(X^T X)^{-1}$ – macierz odwrotna do macierzy $X^T X$

Ideą MNK jest minimalizacja sumy kwadratów reszty.



Rysunek 21. Metoda najmniejszych kwadratów.

Aby można było stosować metodę najmniejszych kwadratów, muszą być spełnione poniższe założenia:

1. Postać funkcjonalna modelu jest liniowa.

2. $\mathrm{E}(\xi\,\mathrm{t}) = 0$, czyli wartość oczekiwana składnika losowego jest równa zero, nie występują wahania przypadkowe, składnik losowy nie zadziała.

3. $\mathrm{Var}(\xi\,\mathrm{t}) = \sigma_{\xi}^{2} = \mathrm{const.}$, czyli wariancja składnika losowego jest stała w czasie; jeżeli wystąpią już wahania w czasie to zawsze o tę samą wartość.

4. $\mathrm{Cov}(\xi\,\mathrm{t},\ \xi\,\mathrm{s}) = 0,$, założenie o braku autokorelacji składnika losowego.

Autokorelacja jest to przenoszenie oddziaływania składnika losowego z okresu t na składnik losowy z okresu s.

5. $\mathrm{Cov}(X_{\mathrm{t}}, \xi_{\mathrm{t}}) = 0$, czyli zmienne objaśniające są zmiennymi nielosowymi.

6. $\xi_{\mathrm{t}} \sim \mathrm{N}(0, \sigma_{\xi}^{2})$, czyli składnik losowy ma rozkład normalny o wartości oczekiwanej równej zero i wariancji stałej w czasie, wynika to z założenia 2 i 3.

Estymator jest BLUE (the Best Linear Unbiased Estimator), czyli najlepszy liniowy nieobciążony estymator.

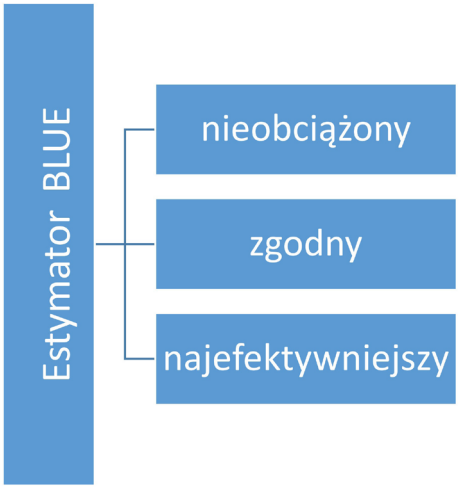
Własności estymatora MNK są następujące:

1. **nieobciążony** – obliczona wartość parametrów strukturalnych jest równa ich wartości rzeczywistej;

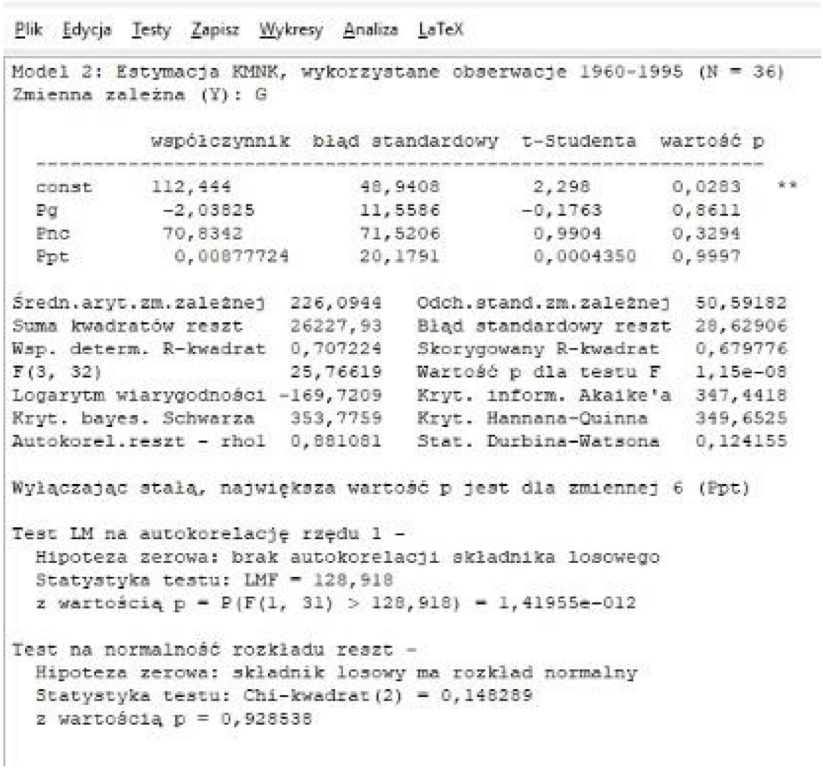
$$E(\hat{\beta}) = \beta$$

2. **zgodny** – to taki, który w miarę wzrostu liczby obserwacji przyjmuje wartości zbliżone szacowanemu parametrowi; dla dowolnie wybranej grupy (nieskończenie wiele obserwacji) zawsze otrzymamy tę samą wartość parametru β;

3. **najefektywniejszy** – to taki, którego wariancja (błąd) jest niewielki; estymator ten ma najmniejszą wariancję spośród wszystkich możliwych do wyznaczenia estymatorów; estymator ma najmniejsze rozproszenie wokół prawdziwej wartości parametru.



Rysunek 22. Własności estymatora BLUE.



Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Model 2: Estymacja KMNK, wykorzystane obserwacje 1960-1995 (N = 36)
Zmienna zależna (Y): G

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	112,444	48,9408	2,298	0,0283	**
Pg	-2,03825	11,5586	-0,1763	0,8611	
Pnc	70,8342	71,5206	0,9904	0,3294	
Ppt	0,00877724	20,1791	0,0004350	0,9997	

Średn.aryt.zm.zależnej	226,0944	Odch.stand.zm.zależnej	50,59182
Suma kwadratów reszt	26227,93	Błąd standardowy reszt	28,62906
Wsp. determ. R-kwadrat	0,707224	Skorygowany R-kwadrat	0,679776
F(3, 32)	25,76619	Wartość p dla testu F	1,15e-08
Logarytm wiarygodności	-169,7209	Kryt. inform. Akaike'a	347,4418
Kryt. bayes. Schwarza	353,7759	Kryt. Hannana-Quinna	349,6525
Autokorel.reszt - rho1	0,881081	Stat. Durbina-Watsona	0,124155

Wyłączając stałą, największa wartość p jest dla zmiennej 6 (Ppt)

Test LM na autokorelację rzędu 1 -
  Hipoteza zerowa: brak autokorelacji składnika losowego
  Statystyka testu: LMF = 128,918
  z wartością p = P(F(1, 31) > 128,918) = 1,41955e-012

Test na normalność rozkładu reszt -
  Hipoteza zerowa: składnik losowy ma rozkład normalny
  Statystyka testu: Chi-kwadrat(2) = 0,148289
  z wartością p = 0,928538

Rysunek 23. Wyniki modelu metodą najmniejszych kwadratów w programie GRETL.

2.2. Weryfikacja modelu

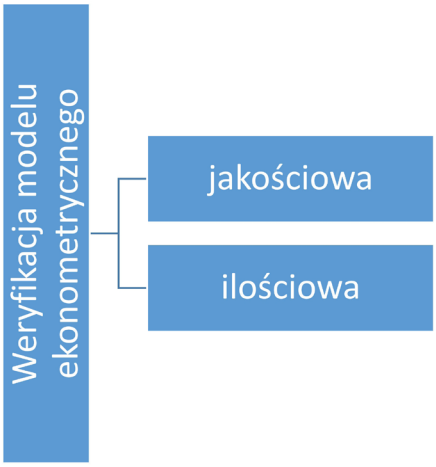
Weryfikacja modelu ekonometrycznego polega na sprawdzeniu, czy przyjęte założenia dotyczące modelu są spełnione w świetle uzyskanych wyników.

Weryfikacja modelu sprowadza się do zbadania:

- czy szacowany model jest zgodny z rzeczywistością,
- czy model jest wystarczająco precyzyjny,
- czy zmienne objaśniające istotnie wpływają na zmienną endogeniczną,
- czy spełnione są założenia metody najmniejszych kwadratów.

Etapy weryfikacji modelu ekonometrycznego:

1. Weryfikacja jakościowa – dotycząca założeń ekonomicznych, sprawdzamy, czy wyznaczone parametry mają sens ekonomiczny.
2. Weryfikacja ilościowa – weryfikacja hipotez statystycznych (wariancja w czasie, autokorelacja, rozkład normalny).

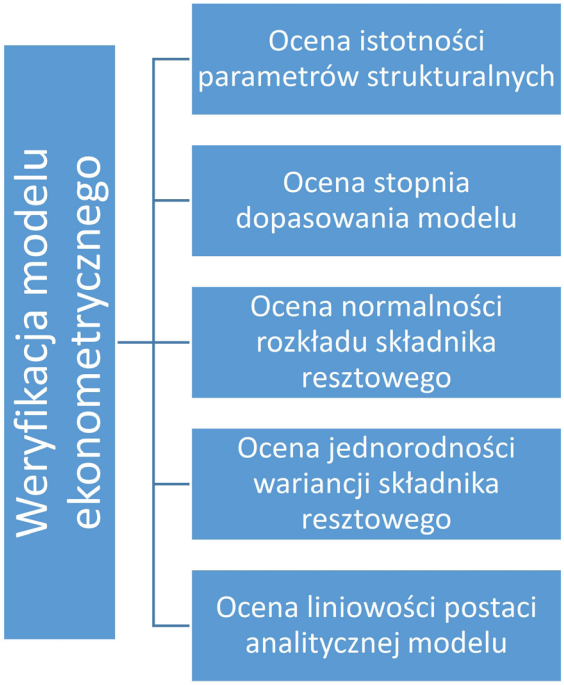


Rysunek 24. Weryfikacja modelu ekonometrycznego.

Weryfikacja ilościowa modelu ekonometrycznego obejmuje:

1. badanie normalności rozkładu składnika losowego,
2. badanie homoskedastyczności rozkładu składnika losowego,
3. badanie autokorelacji składnika losowego,
4. badanie istotności zmiennych objaśniających,
5. analizę syntetycznych miar dopasowania – służą do sprawdzenia, czy model dobrze odzwierciedla kształtowanie się zmiennej endogenicznej.

Oszacowanie parametrów strukturalnych modelu klasyczną metodą najmniejszych kwadratów daje początek analizie weryfikacyjnej, dzięki której można stwierdzić, czy model został oszacowany poprawnie i czy może w rezultacie być stosowany w praktyce.



Rysunek 25. Etapy weryfikacji modelu ekonometrycznego.

2.3. Ocena stopnia dopasowania modelu

Ocenę **dopasowania modelu** dokonujemy na podstawie poniższych miar:



WAŻNE

Współczynnik zgodności:

$$\varphi_y^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

$$\varphi_x^2 = \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2}$$

gdzie:
y_i – rzeczywista wartość zmiennej zależnej
$\hat{y}_i$ – przewidywana wartość zmiennej zależnej
$\bar{y}_i$ – średnia wartość rzeczywistej zmiennej zależnej
x_i – rzeczywista wartość zmiennej niezależnej
$\hat{x}_i$ – przewidywana wartość zmiennej niezależnej
$\bar{x}_i$ – średnia wartość rzeczywistej zmiennej niezależnej

Określa **dopasowanie wyznaczonej linii regresji do punktów empirycznych**. Im wartość współczynnika bliższa zeru, tym wyznaczona linia regresji lepiej opisuje badaną zależność.

Dodatkowo $\varphi^2 \cdot 100\%$ mówi nam, w ilu procentach na zmienność zmiennej zależnej Y wpływają inne czynniki niż zmienna niezależna X.



UWAGA

Przyjmuje się, że:

$\varphi^2 = 0 \varphi^2 = 0$ **dopasowanie doskonałe**
$0 < \varphi^2 \leq 0{,}1$ **dopasowanie jest bardzo dobre**
$0{,}1 < \varphi^2 \leq 0{,}2$ **dopasowanie jest dobre**
$0{,}2 < \varphi^2 \leq 0{,}3$ **dopasowanie jest dość dobre**
$0{,}3 < \varphi^2 \leq 0{,}4$ **dopasowanie jest słabe**
$0{,}4 < \varphi^2 \leq 0{,}5$ **dopasowanie jest bardzo słabe**
$0{,}5 < \varphi^2$ **dopasowanie jest złe**

Kolejna miara dopasowania modelu do rzeczywistości jest **współczynnik determinacji**. Mierzy on stopień dopasowania regresji jako przybliżenia liniowej zależności pomiędzy zmienną celu a zmienną objaśniającą.



WAŻNE

Współczynnik determinacji:

$$R^2 = 1 - \varphi^2$$

Współczynnik determinacji informuje o tym, jaka część zmiennej zależnej Y (skutek) jest wyjaśniana przez zmiany zmiennej niezależnej X (przyczyna). Im współczynnik determinacji jest bliższy 1, tym dopasowanie linii regresji do danych empirycznych jest lepsze.



UWAGA

Przyjmuje się, że:

$0 \leq R^2 < 0{,}5$ dopasowanie jest niezadowalające
$0.5 \leq R^2 < 0{,}6$ dopasowanie jest słabe
$0.6 \leq R^2 < 0{,}8$ dopasowanie jest zadowalające
$0.8 \leq R^2 < 0{,}9$ dopasowanie jest dobre
$0.9 \leq R^2 \leq 1$ dopasowanie jest bardzo dobre

Im prosta regresji jest lepiej dopasowana do danych (czyli im R^2 większe), tym możemy przewidzieć wartości Y dla bardziej oddalonych wartości X od danych, na podstawie których stworzyliśmy prostą regresji liniowej.

Aby funkcję regresji wykorzystać dla celów prognozowania, musi być ona dobrym predyktorem, tzn. być dobrze dopasowana do danych empirycznych. Reszta jest to różnica pomiędzy wartością empiryczną a wartością teoretyczną.

WAŻNE

Wzory – reszty

$$e_i = y_i - \hat{y}_i$$
$$f_i = x_i - \hat{x}_i$$

gdzie:
y_i – rzeczywista wartość zmiennej zależnej
$\hat{y}_i$ – przewidywana wartość zmiennej zależnej
x_i – rzeczywista wartość zmiennej niezależnej
$\hat{x}_i$ – przewidywana wartość zmiennej niezależnej

WAŻNE

Wzory – średni błąd szacunku

$$S_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}$$
$$S_f = \sqrt{\frac{\sum_{i=1}^{n} f_i^2}{n-2}}$$

Średni błąd szacunku – o ile (średnio) wartości empiryczne odchylają się od wartości teoretycznych.

WAŻNE

Wzory – współczynnik zmienności losowej

$$V(e) = \frac{S_e}{\bar{Y}} \cdot 100\%$$

$$V(f) = \frac{S_f}{\bar{X}} \cdot 100\%$$

gdzie:
$\bar{X}$ – średnia arytmetyczna zmiennej niezależnej
$\bar{Y}$ – średnia arytmetyczna zmiennej zależnej

Współczynnik zmienności losowej informuje, jaką część średniej wartości badanego zjawiska stanowi odchylenie standardowe reszt. Mniejsze wartości tego współczynnika wskazują na lepsze dopasowanie modelu do danych empirycznych.

UWAGA

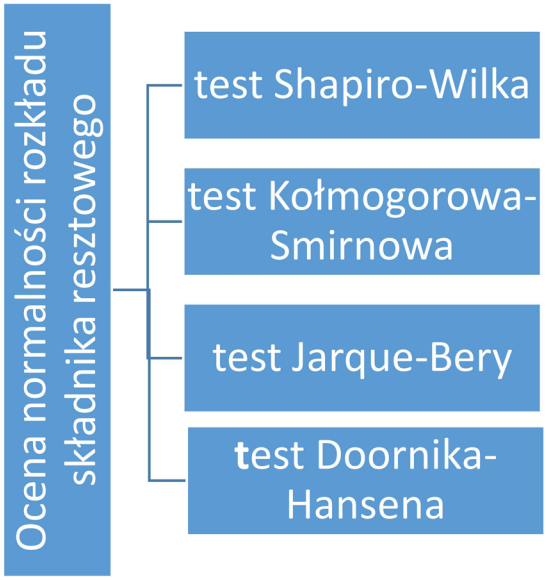
Wartość współczynnika wyrażona jest w procentach, a jej interpretacja zależna jest od wielkości współczynnika:

0 ≤ *V(e)* < 25% mała zmienność
25% ≤ *V(e)* < 45% przeciętna zmienność
45% ≤ *V(e)* < 100% silna zmienność
100% ≤ *V(e)* bardzo silna zmienność

Przyczynami powodującymi otrzymywanie dużych błędów szacunku parametrów mogą być:

- mała liczebność próby losowej wykorzystywanej do estymacji parametrów funkcji regresji,
- nieodpowiednia metoda estymacji,
- niewłaściwa zmienna objaśniająca dla funkcji regresji.

2.4. Ocena normalności rozkładu składnika resztowego



Rysunek 26. Testy oceny normalności rozkładu składnika resztowego.

Jednym z bardziej popularnych testów pozwalających na weryfikację normalności składnika losowego jest **test Shapiro-Wilka**. Test ten weryfikuje hipotezę

H0: składnik losowy modelu $y = \beta X + \varepsilon$ ma rozkład normalny

wobec hipotezy alternatywnej

H1: składnik losowy modelu $y = \beta X + \varepsilon$ nie ma rozkładu normalnego

Test ten jest stosowany przeważnie **do mniejszych grup (N < 100)**.

W przypadku **większych prób** używamy **testu Kołmogorowa-Smirnowa** (jako jeden z przedziałów podaje się **N > 100**).

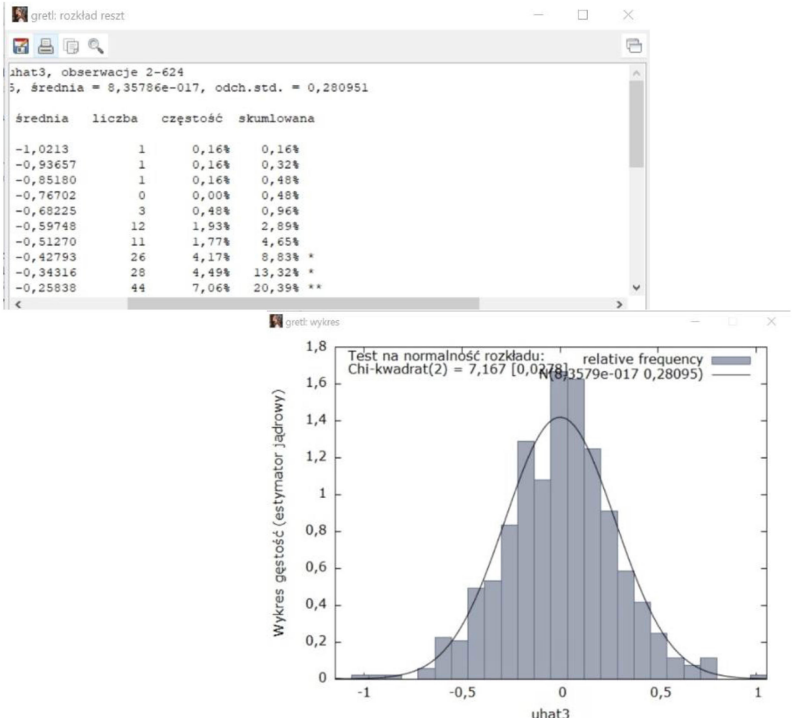
Pakiety statystyczne podają wartość statystyki Shapiro-Wilka oraz wartość p. Jeśli wartość $p \geq 0{,}05$, to nie ma podstaw do odrzucenia hipotezy, że dana zmienna pochodzi z rozkładu normalnego.

Innym testem oceny normalności rozkładu składnika losowego jest **test Jarque-Bery** (JB) oraz **test Doornika_Hansena** (DH). Jeśli składniki losowe mają rozkłady normalne, wtedy statystyki JB oraz DH mają rozkłady chi-kwadrat (patrz Załącznik 2).

Reguła podejmowania decyzji jest następująca:

- jeżeli wartości statystyk JB lub DH nie są większe od chi-kwadrat, to nie ma podstaw do odrzucenia hipotezy zerowej, że składniki zakłócające modelu mają rozkłady normalne,
- jeżeli wartość statystyk JB lub DH są większe od chi-kwadrat, to hipotezę zerową odrzucamy na korzyść alternatywnej, że składniki zakłócające modelu nie mają rozkładów normalnych.

Stosując test DH, możemy skorzystać z programu GRETL, oblicza on wartość statystyki Doornika-Hansena wraz z odpowiadającym tej wartości poziomem p. Jeśli chcemy otrzymać wartość statystyki Jarque-Bery, powinniśmy zapisać reszty z oszacowania modelu w arkuszu kalkulacyjnym programu GRETL i następnie skorzystać z zakładki Zmienna/testy normalności rozkładu – patrz rysunek poniżej.



Rysunek 27. Wyniki testu oceny normalności w programie Gretl.

2.5. Autokorelacja

W ekonometrii skorelowanie zmiennej z tą samą zmienną z innego okresu nazywamy **autokorelacją** lub innymi słowy, gdy wartość oczekiwana błędu losowego zależy od wielkości błędów losowych dla innych obserwacji, co zapisujemy:

$$Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i, \varepsilon_j) \neq 0, i \neq j$$

Autokorelacja składnika losowego jest problemem najczęściej występującym w przypadku **szeregów czasowych** i polega na zależności (skorelowaniu) bieżących wartości składnika losowego od wartości przeszłych.

Jeżeli **w modelu ekonometrycznym występuje autokorelacja**, a tworząc model, problem ten pominiemy, to doprowadzi to do sytuacji, w której będziemy wyciągać wnioski na podstawie błędnego modelu. Wtedy estymator metody najmniejszych kwadratów traci efektywność, choć pozostaje nieobciążony i zgodny. Wobec tego sprawdzenie, czy składniki losowe są skorelowane, jest bardzo ważnym elementem poprzedzającym wnioskowanie statystyczne.

Najpopularniejsze testy na występowanie autokorelacji to test Durbina-Watsona oraz test Breuscha-Godfreya.

Test Durbina-Watsona

W tym teście możliwe są dwa układy hipotez, jeżeli korelacja reszt w próbie jest dodatnia, to:

$$H0 : \rho = 0; H1 : \rho > 0$$

Oznacza to tyle, że **jako hipotezę zerową przyjmujemy brak występowania autokorelacji składników losowych**, gdyż współczynnik autokorelacji reszt przyjmuje wartości bliskie zeru. Weryfikujemy ją na rzecz hipotezy alternatywnej, która zakłada dodatnie skorelowanie w czasie składników losowych, które jest statystycznie istotne. Drugi układ hipotez ma postać:

$$H0 : \rho = 0; H1 : \rho < 0$$

W tym przypadku **hipoteza zerowa również wyklucza występowanie korelacji w wartościach reszt modelu**. Jednakże w tym układzie w hipotezie alternatywnej przyjmuje się występowanie statystycznie istotnej, **ujemnej autokorelacji**. Sprawdzianem hipotez jest statystyka d zdefiniowana wzorem:

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2}$$

gdzie: ε – reszty modelu, n – liczebność próby.

Statystyka Durbina-Watsona przyjmuje wartości od 0 do 4. Im bliżej wartości skrajnych przedziału występowania d, tym współczynnik autokorelacji jest bliższy wartości bezwzględnej z 1. Jeżeli d przyjmuje wartości bliskie 0, to ρ jest bliski jedności, natomiast ρ jest bliski -1, gdy d równa się 4. Brak autokorelacji składników losowych występuje, gdy d jest równe 2.

Test Durbina-Watsona powinien być wykonywany w przypadku, gdy model regresji ma wartość stałą a rozkład składnika losowego błędów predykcji jest zbliżony do rozkładu normalnego.



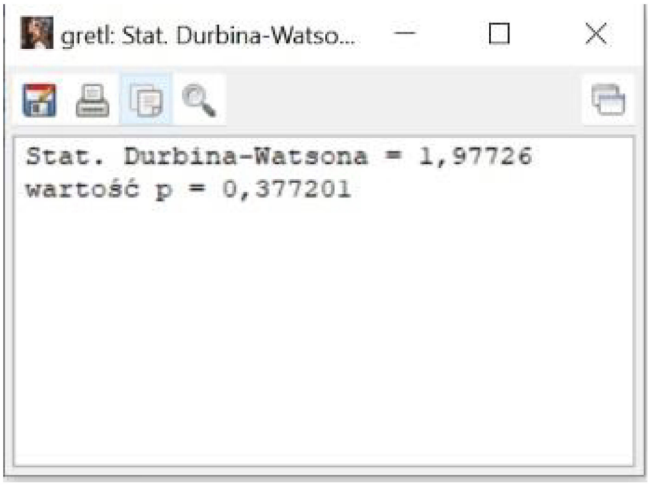
UWAGA

Aby poprawnie stosować test Durbina-Watsona, muszą być spełnione następujące założenia:

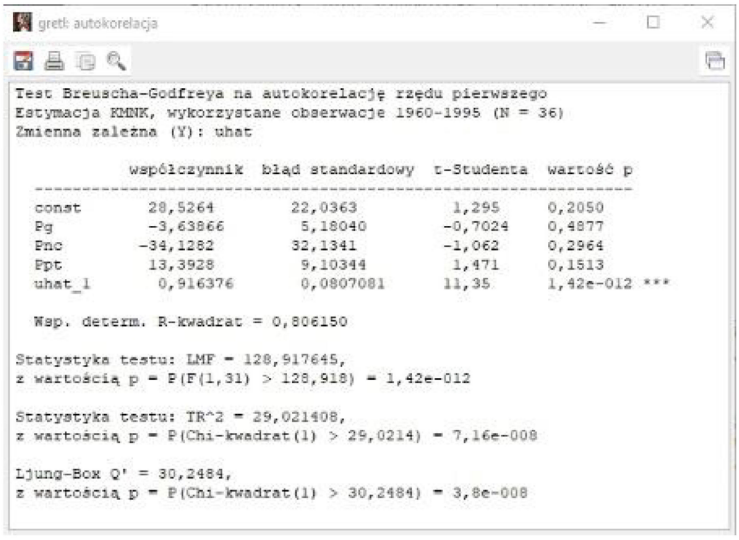
- ***n* > 15,**
- **musi istnieć wyraz wolny w modelu,**
- **składniki resztowe mają rozkład normalny,**
- **w modelu nie występuje opóźniona zmienna zależna w charakterze zmiennej niezależnej.**

W programie GRETL do weryfikacji autokorelacji możemy stosować test Durbina-Watsona oraz test Breuscha-Godfreya.

Test Breuscha-Godfreya jest testem wykorzystującym metodę mnożników Lagrange'a. Ma on przewagę nad testem Durbina-Watsona, ponieważ jest w stanie wykrywać obecność autokorelacji wyższych rzędów.



Rysunek 28. Wyniki testu Durbina-Watsona w programie Gretl.



Rysunek 29. Wyniki testu Breuscha-Godfreya w programie Gretl.

W przypadku stwierdzenia autokorelacji możemy wybrać jedną z trzech opcji postępowania:

- usunąć przyczynę autokorelacji,
- zastosować procedury estymacji w warunkach autokorelacji, np. uogólniona MNK Aitkena lub metoda różniczki zupełnej,
- pogodzić się z mniejszą efektywnością estymatorów.

Przyczyny wywołujące autokorelację składników losowych w modelu ekonometrycznym są następujące:

1. Naturalne – następuje zakłócenie w normalnym przebiegu zjawiska, czyli okres przyjęty za jednostkę obserwacji jest krótszy niż czas działania czynników przypadkowych (tzw. autokorelacja czysta).
2. Błędy w budowie modelu powodujące autokorelację:

dodatnią	ujemną
• przyjęcie niewłaściwej postaci analitycznej (np. zamiast potęgowej szacujemy liniową), • pominięcie jednej lub kilku istotnych zmiennych, • powolne wygasanie efektów składnika losowego, • przyjęcie niewłaściwych opóźnień przy zmiennych, • niesystematyczny rozkład zmiennej losowej.	• przepełnienie modelu zmiennymi, • nieuwzględnienie występowania cyklu dwuokresowego.

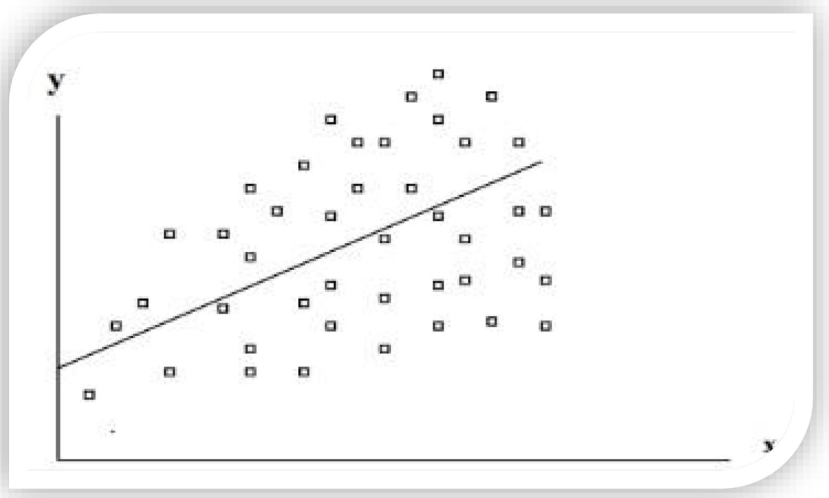
Sposoby eliminacji autokorelacji:

- rozpoznanie przyczyn autokorelacji oraz zmiana konstrukcji modelu,
- zmiana skali zmiennych ekonomicznych,
- dołączenie nowej zmiennej objaśniającej (gdy autokorelacja dodatnia),
- wprowadzenie dynamiki do modelu (gdy autokorelacja dodatnia),
- dołączenie zmiennej zerojedynkowej (gdy autokorelacja dodatnia),
- zmiana postaci analitycznej modelu (gdy autokorelacja dodatnia),
- usunięcie zmiennej (zmiennych) objaśniającej (gdy autokorelacja ujemna),
- dołączenie zmiennej cyklicznej dwuokresowej (gdy autokorelacja ujemna).

2.6. Heteroskedastyczność

Heteroskedastyczność składnika losowego jest drugą formą niespełnienia założenia o sferyczności macierzy wariancji-kowarian-

cji składnika losowego. Zjawisko heteroskedastyczności składnika losowego charakteryzuje przede wszystkim modele oparte o dane przekrojowe. Problem heteroskedastyczności może sygnalizować poważny problem, jak np. problem pominiętych zmiennych.



Rysunek 30. Heteroskedastyczność.

O występowaniu heteroskedastyczności w modelu mówimy, gdy wariancja zmiennej losowej nie jest taka sama dla wszystkich obserwacji:

$$Var(\varepsilon_i) = E(\varepsilon_i^2) \neq const, i \neq j.$$

Heteroskedastyczność można wstępnie zidentyfikować na podstawie wykresu reszt. Jeżeli założenie o heteroskedastyczności lub założenie braku autokorelacji nie jest spełnione, to mówimy o niesferyczności błędów losowych.

Testowania heteroskedastyczności możemy dokonać za pomocą **testu White'a**.

Hipotezy zaproponowane przez White'a są następujące:

$H_0: \sigma_i^2 = \sigma^2$ składnik losowy jest homoscedastyczny, wariacja jest stała,

$H_1: \sigma_i^2 \neq \sigma^2$ składnik losowy jest heteroscedastyczny, wariacja nie jest stała.

Weryfikacja przeprowadzana jest dla regresji pomocniczej w oparciu o test F o rozkładzie Fischera-Snedecora z (k; n - (k + 1)) stopniami swobody. Jeśli wartość testu F dla regresji pomocniczej jest większa od wartości krytycznej, hipotezę zerową odrzucamy. Wtedy wnioskujemy, że rozkład składnika losowego jest heteroskedastyczny, co w konsekwencji prowadzi do odrzucenia modelu.

Program GRETL wyznacza wartość statystyki testu White'a wraz z odpowiadającym tej wartości poziomem p-minimalnego ryzyka koniecznego do przyjęcia, by odrzucić hipotezę zerową. W programie GRETL mamy do wyboru też inne testy dotyczące heteroskedastyczności, wybieramy je z zakładki Testy/testy heteroskedastyczności.

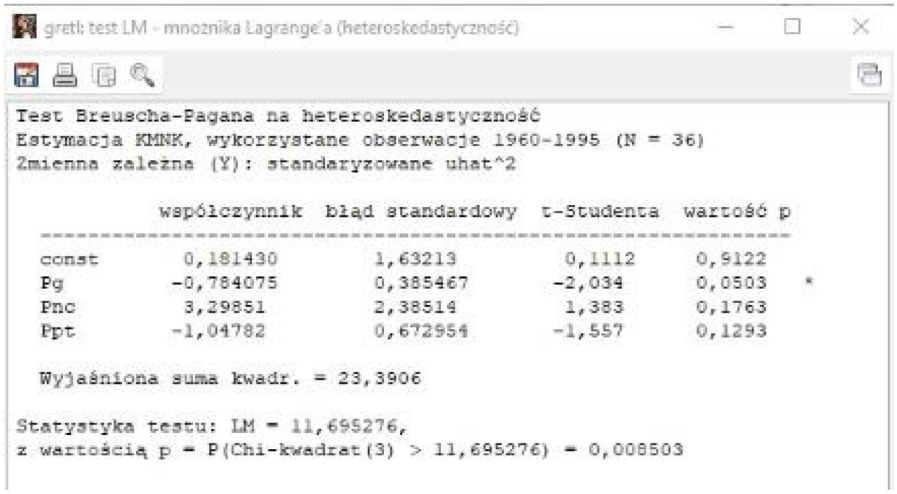
Test White'a na heteroskedastyczność reszt (zmienność wariancji resztowej)
Estymacja KMNK, wykorzystane obserwacje 1960-1995 (N = 36)
Zmienna zależna (Y): uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	-13149,0	13862,5	-0,9485	0,3516
Pg	1730,90	6060,08	0,2856	0,7774
Pnc	19886,6	38882,1	0,5115	0,6133
Ppt	-3186,76	9959,78	-0,3200	0,7516
sq_Pg	1086,48	674,636	1,610	0,1194
X2_X3	-7941,69	8164,89	-0,9727	0,3397
X2_X4	1813,77	2170,14	0,8358	0,4109
sq_Pnc	2508,87	28194,6	0,08898	0,9298
X3_X4	-2639,25	14829,9	-0,1780	0,8601
sq_Ppt	573,207	2004,05	0,2860	0,7771

Wsp. determ. R-kwadrat = 0,665600

Statystyka testu: TR^2 = 23,961592,
z wartością p = P(Chi-kwadrat(9) > 23,961592) = 0,004362

Rysunek 31. Wyniki testu White'a w programie GRETL.



Rysunek 32. Wyniki testu Breuscha-Pagana w programie GRETL.

W przypadku stwierdzenia heteroskedastyczności możemy wybrać jedną z poniższych opcji postępowania:

- zastosować uogólnioną metodę najmniejszych kwadratów lub ważoną metodę najmniejszych kwadratów,
- zastosować odpowiednią transformację np. logarytmiczną, pierwiastkową, odwrotnościową, kwadratową,
- pogodzić się z mniejszą efektywnością estymatorów.

2.7. Pytania sprawdzające



1. Jaka jest idea metody najmniejszych kwadratów?

2. Omów założenia metody najmniejszych kwadratów.

3. Jakie są własności estymatora MNK?



4. Na czym polega weryfikacja modelu ekonometrycznego?

……
……
……
……
……
……
……
……
……
……

5. Jakie są etapy weryfikacji modelu ekonometrycznego?

……
……
……
……
……
……
……
……
……
……

6. Wymień miary stosowane do weryfikacji modelu ekonometrycznego.

……
……
……
……
……
……
……
……
……
……

7. O czym informuje współczynnik determinacji?



8. Co to jest współczynnik zmienności losowej?

9. Jakie są najczęściej powody otrzymania dużych błędów szacunku parametrów modelu ekonometrycznego?



10. W jaki sposób badamy normalność składnika losowego?

11. Co oznacza pojęcie autokorelacji?

12. Jakie znasz testy na występowanie autokorelacji?

13. Jakie warunki muszą być spełnione, żeby można było stosować test Durbina-Watsona?



14. Co należy zrobić w przypadku wystąpienia autokorelacji?

15. Jakie są powody występowania autokorelacji?



16. Jakie są sposoby eliminacji autokorelacji?

17. Omów zjawisko heteroskedastyczności.

18. Omów postępowanie w przypadku stwierdzenia heteroskedastyczności.

1 2 3 4

3. Prognozowanie



Prognozowanie jest rodzajem **wnioskowania ze znanego o nieznanym.** Przy danym stanie wiedzy ma ono dostarczyć jak najlepszego wyobrażenia o przyszłości. Jest to jednak zwykle wyobrażenie bardzo ułomne, z reguły znacznie mniej dokładne niż byśmy tego sobie życzyli.

Nazwą prognozowanie obejmujemy **wnioskowanie oparte na danych czasowych**, czyli prognozowanie o przyszłym biegu zdarzeń. Prognozowanie – podstawowe pojęcia oparte na danych przekrojowych, np. wnioskowanie o dochodach pewnej gminy na podstawie modelu opisującego kształtowanie dochodów innych gmin danego województwa.

Prognoza jest naukowo uzasadnionym sądem o stanie zjawiska w określonym momencie (okresie) należącym do przyszłości. Użycie w powyższym określeniu słowa „sąd" sygnalizuje niepewność prognozy, a odwołanie się do nauki oznacza, że prognoza musi być racjonalnym wnioskowaniem, prowadzącym od przesłanek do wniosków odnoszących się do przyszłości [Cieślak, 1998 s. 91].

Aby można było wnioskować na podstawie modelu ekonometrycznego, muszą być spełnione następujące założenia:

- znajomość modelu zmiennej prognozowanej,
- stabilność parametrów postaci analitycznej,
- stabilność rozkładu odchyleń losowych modelu,
- znajomość wartości zmiennych objaśniających w okresie prognozowania,
- dopuszczalność ekstrapolacji modelu poza próbę statystyczną.

3.1. Rodzaje prognoz

W literaturze występuje wiele podziałów na **rodzaje prognoz**, najczęściej spotyka się jednak podziały ze względu na następujące kryteria:

Kryterium 1. Ze względu na charakter zjawiska

1. **Ilościowe** (gdy stan zmiennej jest wyrażony liczbą):
   - punktowe (gdy podaje się, że zmienna przyjmie jakąś wartość);
   - przedziałowe (gdy podaje się przedział liczbowy, w którym znajdzie się wartość zmiennej);

- skalarne (podawana jest pojedyncza wartość);
- wektorowe (jako prognozę uzyskujemy wektor liczb).

2. **Jakościowe** (prognozowanym zdarzeniem jest stan zmiennej niemierzalnej lub słownie opisana sytuacja dotycząca zmiennej mierzalnej).

3. **Proste** (prognoza dotyczy jednej zmiennej ekonomicznej lub prognozy uzyskane są bez udziału innych prognoz).

4. **Złożone** (sąd dotyczący złożonego zjawiska ekonomicznego, przy konstrukcji prognoz korzystamy z innych zmiennych prognozowanych).

5. **Jednorazowe** (stawiane jednokrotnie i najczęściej dla zmiennych strategicznych):
   - powtarzalne (ciągle poprawiane w miarę dopływu nowych informacji);
   - kompleksowe (prognoza całościowo opisująca przyszłą sytuację zjawiska);
   - sekwencyjne (prognoza zjawiska dla kilku okresów);
     - samorealizujące się (prognoza jest przyczyną zaistnienia zapowiadanego zdarzenia, mimo iż prawdopodobieństwo jej realizacji było niewielkie);
     - samounicestwiające się (prognoza destruktywna powoduje obniżenie szans na realizację zdarzenia).



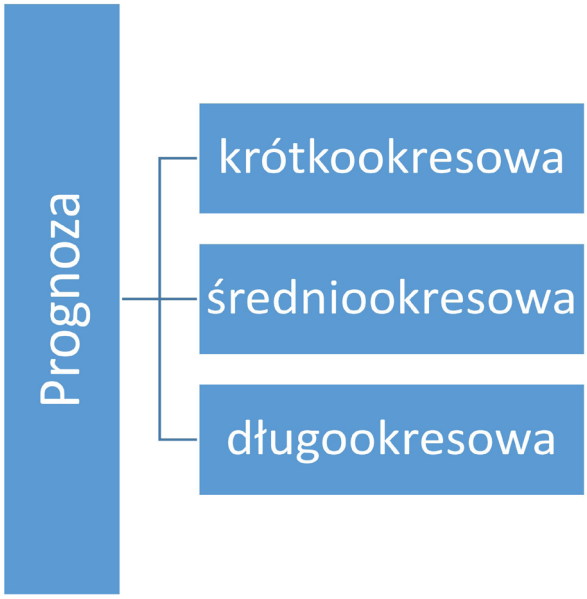
Rysunek 33. Rodzaje prognoz ze względu na charakter zjawiska.

Kryterium 2. Ze względu na okres prognozowania

1. **Krótkookresowa.** Cechą charakterystyczną prognoz krótkookresowych jest niewielki zasięg ekstrapolacji w porównaniu do liczby posiadanych danych, ponadto jest to prognoza jedynie na kilka odcinków czasowych do przodu lub dla konkretnego momentu, dla którego mogą zachodzić zmiany ilościowe badanego zjawiska.

2. **Średniookresowa.** Prognoza średniookresowa wyróżnia się niewielkim zasięgiem ekstrapolacji dla średniej liczby danych. Umożliwia prognozy na średnią liczbę odcinków czasowych oraz prognozy momentu, do którego mogą zachodzić zmiany ilościowe oraz niewielkie (mało istotne) zmiany jakościowe badanego zjawiska.

3. **Długookresowa**. Prognoza długookresowa ma największy zasięg, jest to prognoza na dużą liczbę jednostek czasowych, a do momentu prognozowanego mogą zachodzić zarówno zmiany ilościowe, jak i istotne zmiany jakościowe badanego zjawiska.



Rysunek 34. Rodzaje prognoz ze względu na okres prognozy.

Przykład:

- **Prognozowanie długookresowe sprzedaży** jest potrzebne przede wszystkim dla zaprogramowania zmian organizacyjnych w przedsiębiorstwie, takich na przykład, jak opracowanie nowych form sprzedaży, sposobów wejścia na nowe rynki zbytu, budowy nowych kanałów dystrybucji czy wreszcie organizacji nowych form akwizycji i reklamy.

Jest ono także pomocne przy opracowywaniu polityki wprowadzania nowych produktów na rynek i wycofywania starych. Bez niego nie można prawidłowo zaplanować zakupu nowych maszyn i urządzeń oraz sporządzić budżetu przedsiębiorstwa.

- **Prognozowanie krótkookresowe** wykorzystuje się w planowaniu operacyjnym przedsiębiorstwa. Jest ono wykorzystywane praktycznie w każdym dziale przedsiębiorstwa do planowania bieżących działań. Służy do sporządzania szacunków rozlokowania produkcji na poszczególne rynki w podziale geograficznym i czasowym całego roku.

Stosuje się je do szacowania zmian cen i ich skutków dla wielkości wpływów ze sprzedaży, wydatków na reklamę, a także do testowania skutków wprowadzania zmian w warunkach sprzedaży i określania zadań dla personelu sprzedającego. Mając wielkość sprzedaży, możemy określić planowaną wielkość produkcji, zatrudnienia, wielkości zakupu surowców, materiałów, zaplanować szkolenia pracowników, wydatki administracyjne itd.

Kryterium 3. Ze względu na cel prognozy

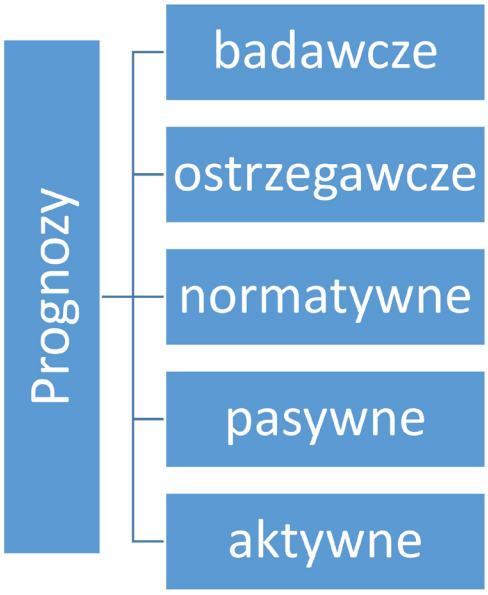
1. **Badawcze** – wszechstronne rozpoznanie przyszłości, zidentyfikowanie przyszłych zdarzeń i ukazanie różnych możliwych ich wariantów; odgrywają ważną rolę – są stymulatorem działań w stosunku do zmiennych sterowanych, akcentują dokładne i obiektywne rozpoznanie przyszłości.

2. **Ostrzegawcze** – podawanie sygnałów zwracających uwagę na niekorzystne dla odbiorcy kształtowania się zjawisk.

3. **Normatywne** – prognozy dotyczą pewnych norm, które obowiązywać będą w przyszłości.

4. **Pasywne** – zniechęcają do podejmowania określonych działań.

5. **Aktywne** – pobudzają do działania.



Rysunek 35. Rodzaje prognoz ze względu na cel prognozy.

Kryterium 4. Ze względu na zasięg

1. **Całościowe (globalne).**

2. **Częściowe (odcinkowe)** – dotyczą tylko pewnego aspektu zjawiska w skali makro.

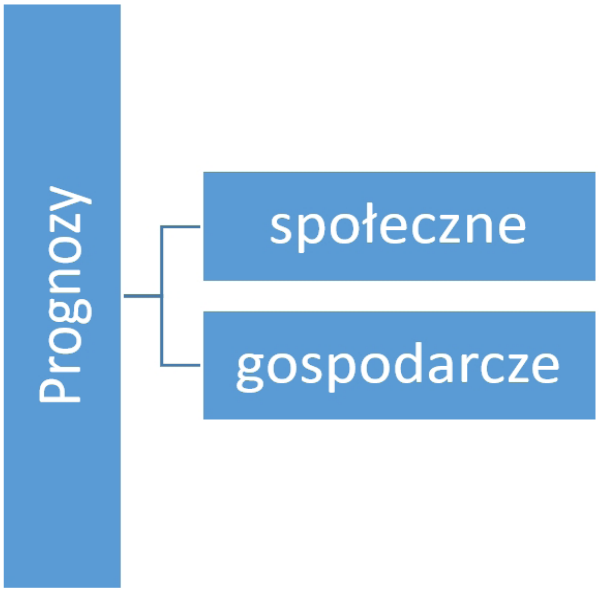
3. **Prognozy makroekonomiczne** (dotyczące całej gospodarki narodowej lub też określonych regionów kraju) w skali mikro.

4. **Prognozy mikroekonomiczne** (dotyczące funkcjonowania pojedynczych jednostek gospodarczych).

Kryterium 5. Z uwagi na rodzaj prognozowanego zjawiska

1. **Gospodarcze** – pełnią rolę informacyjną i ostrzegawczą, wpływają na bieżącą działalność rządu, instytucji finansowych i innych podmiotów gospodarczych, pozwalają odpowiednio wcześnie określić oddziaływanie czynników na prognozowane zjawisko i dostosować swoje strategie do przewidywanych zmian.

2. **Społeczna prognoza** stanowi zjawisko społeczne, które łącznie z innymi może oddziaływać na prognozowane zdarzenie. Czynnik ludzki ma wpływ na prognozowane zdarzenie, w tym także przez sam fakt ogłoszenia prognozy. Prognozy społeczne nie mogą eliminować niepewności co do przyszłego stanu świata, natomiast prognozy te są użyteczne, ponieważ redukują tę niepewność.



Rysunek 36. Prognozy ze względu na rodzaj prognozowanego zjawiska.



UWAGA

Założenia teorii predykcji:

- znany jest oszacowany model ekonometryczny wyjaśniający kształtowanie się zmiennej, dla której tworzymy prognozę,
- struktura opisywanych przez model zjawisk jest stabilna w czasie,
- znane są wartości zmiennych objaśniających w okresie prognozowanym;
- rozkład składnika losowego nie ulega zmianom w czasie,
- dopuszczalna jest ekstrapolacja modelu poza obszar zmienności zmiennych objaśniających obserwowanych w próbie.

3.2. Metody prognozowania

W rozwijającej się obecnie statystycznej teorii prognoz proces przewidywania i oceny przyszłości oparty jest na teoretycznych studiach, analitycznych rozważaniach, logicznych przesłankach oraz praktycznych doświadczeniach.

Powszechnie wykorzystuje się:

- metody ilościowe o charakterze matematyczno-statystycznym,
- pojęcia i aparat analityczny rachunku prawdopodobieństwa,
- budowane modele statystyczno-ekonomiczne.

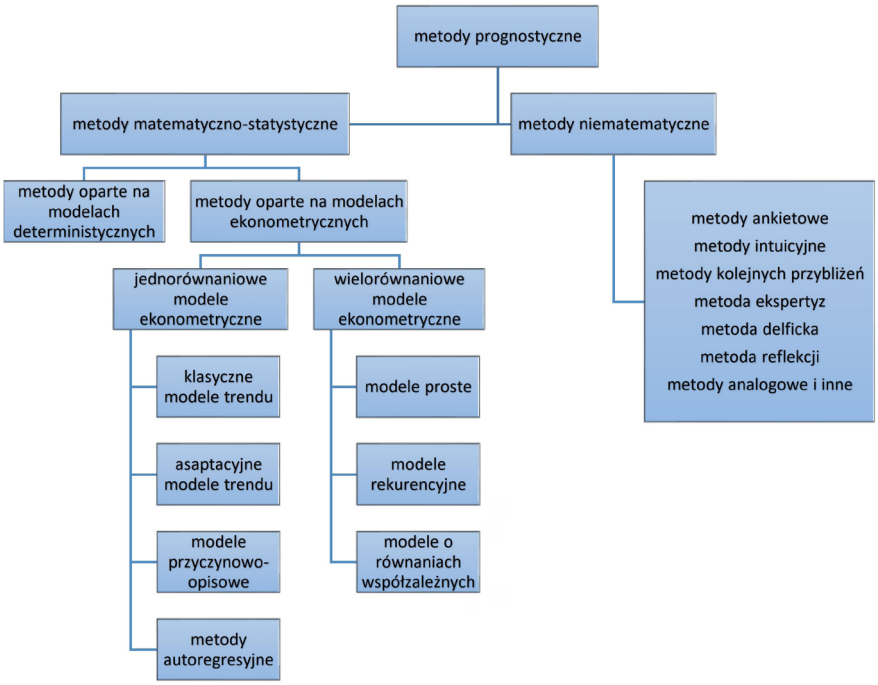
Biorąc pod uwagę rodzaj sporządzanej prognozy oraz charakter badanego zjawiska, stosuje się wybrane metody prognozowania, do których zaliczamy metody matematyczno-statystyczne oraz metody niematematyczne.

Wśród metod niematematycznych wyróżnia się metody prognozowania:

- metody ankietowe,
- intuicyjne,
- kolejnych przybliżeń,
- metodę ekspertyz,
- delficką,
- refleksji,
- metody analogowe i inne.

Niematematyczne, inaczej nazywane **heurystycznymi**, metody prognostyczne wykorzystują opinie wielu ekspertów, którzy dysponują szeroką wiedzą i doświadczeniem.

Do grupy metod heurystycznych należą: **burza mózgów oraz metoda delficka**. Metody te hołdują założeniu, że osąd grupy jest trafniejszy od osądu jednostki. Metody te stosuje się przy prognozowaniu długookresowym, np. w rozwoju nauki i techniki, a także do przewidywania zmian dotychczasowych prawidłowości.



Rysunek 37. Metody prognozowania.

Długość horyzontu prognozy zależy od wielu czynników, główne czynniki to:
- charakter obiektu prognozy,
- wybrany model prognostyczny,
- zastosowana metoda prognozowania.

Prognozy ekonometryczne muszą spełniać następujące konieczne założenia:
- znajomość modelu zmiennej prognozowania,
- stabilność parametrów i postaci analitycznej modelu,
- stabilność rozkładu odchyleń losowych modelu,
- znajomość wartości zmiennych objaśniających w okresie prognozowania,
- dopuszczalność ekstrapolacji modelu poza próbę statystyczną.

3.3. Etapy prognozowania

Proces prognozowania można podzielić na następujące etapy [Cieślak, 1998]:

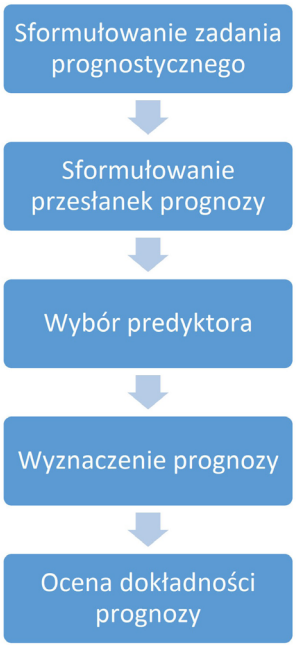
1. **Sformułowanie zadania prognostycznego** – określamy tu:
   - obiekt,
   - zjawisko,
   - prognozowane zmienne,
   - cel wyznaczenia prognozy, pożądany horyzont prognozy,
   - wymagania co do dokładności prognozy.

2. **Sformułowanie przesłanek prognozy** – określamy tu:
   - opis mechanizmu generującego prognozowaną zmienną,
   - listę czynników wpływających na ten mechanizm (w szczególności w modelu ekonometrycznym zaliczymy tu listę zmiennych egzogenicznych wpływających na modelowane zjawiska).

3. **Wybór predyktora** – przepisu, według którego wyznaczamy prognozę oraz zasady prognozowania. Na wybór ten składają się decyzje co do metody prognozowania (na podstawie modelu formalnego strukturalnego lub niestrukturalnego, opinii ekspertów, metodą heurystyczną, analogową itp.).

4. **Wyznaczenie prognozy** – realizacja tego etapu jest związana bezpośrednio z wyborem sposobu i zasady prognozowania. Zastosowanie modelu formalnego oznacza konieczność zbudowania nowego modelu lub wykorzystania (i/lub adaptacji) modelu istniejącego; prognozowanie oparte na opinii ekspertów wymaga określenia grona ekspertów, zapewnienia sobie ich współpracy, określenia sposobu przekształcenia opinii ekspertów (w razie gdyby były zróżnicowana lub wręcz rozbieżne) w jednoznaczną prognozę itd.

5. **Ocena dokładności prognozy** – rozróżniamy tu ocenę ex post opartą na zrealizowanym błędzie prognozy, dokonywaną po zrealizowaniu się zmiennej prognozowanej (tj. gdy prognoza wygaśnie) oraz ocenę ex ante opartą na oczekiwanym błędzie prognozy – wyrażającą przypuszczenia co do dokładności prognozy wyrażone zanim prognoza wygaśnie.



Rysunek 38. Etapy prognozowania ekonometrycznego.

UWAGA

Rozróżniamy dwa rodzaje predykcji (szczegółowo omówione w następnych rozdziałach):

- **punktowa** – polega na wyznaczeniu konkretnej wartości prognozy,
- **przedziałowa** – polega na wyznaczeniu pewnego przedziału liczbowego, któremu można przypisać odpowiednio wysokie prawdopodobieństwo, że rzeczywista realizacja zmiennej prognozowanej znajdzie się w tym przedziale.

Zasady wykonywania prognozy punktowej:

- **Zasada prognozowania według wartości oczekiwanej**, czyli zasada predykcji nieobciążonej, jest to najczęściej stosowana zasada prognozy.
- **Zasada prognozowania według największego prawdopodobieństwa**, wyznaczenie prognozy na podstawie dominanty rozkładu prognozowanej zmiennej; w przypadku spełnienia założenia o normalności składnika losowego pokrywa się ona z poprzednią zasadą.
- **Zasada prognozowania według mediany**, prognozą punktową jest mediana rozkładu (czyli prawdopodobieństwo otrzymania większej lub mniejszej wartości wynosi 50%), rzadko spotykana w praktyce.
- **Zasada prognozowania minimalizującego oczekiwaną stratę**, stosowana, gdy błąd prognozy możemy utożsamić ze stratą, minimalizujemy oczekiwaną stratę.

3.4. Trafność prognozy

O prognozach powiemy, że są:

- **trafne** – gdy okazują się wystarczająco bliskie realizacji prognozowanej zmiennej;
- **nietrafne (chybione)** – gdy rozbieżność prognozy i wielkości prognozowanej okazuje się zbyt wielka jak na nasze potrzeby.

Prognoza jest dopuszczalna, gdy jest obdarzona przez jej odbiorcę stopniem zaufania wystarczającym do tego, by mogła być wykorzystana do celu, dla którego została ustalona. Dopuszczalność prognozy jest określona w tym samym czasie, w którym wyznacza się prognozę.

Dwoma podstawowymi rodzajami mierników dokładności i trafności prognoz są:

- **błędy ex post** – informują o trafności wykonanych prognoz, pokazują odchylenia rzeczywistej wartości zjawiska od jego prognozy;
- **błędy ex ante** – informują o dopuszczalności obliczonych prognoz, pokazują spodziewane odchylenia rzeczywistej wartości zjawiska od jego prognozy.



UWAGA

- **Trafność prognozy określa się po upływie czasu, na który prognoza była wyznaczona.**
- **Stopień trafności prognozy ilościowej mierzy się za pomocą błędów ex post.**
- **Błąd ex post to wartość odchylenia rzeczywistych realizacji zmiennej prognozowanej od obliczonych prognoz.**
- **Błędy ex post można obliczać dla każdego momentu lub okresu należącego do określonego przedziału czasu.**

WAŻNE

Definicja:
Błąd prognoz (percentage error)

$$PE = \frac{y_t - y_t^*}{y_t} \cdot 100\%$$

gdzie:
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

PE jest miarą błędu prognozy dla okresu t. PE określa, jaki procent rzeczywistej realizacji zmiennej prognozowanej wynosi błąd prognozy.

WAŻNE

Definicja:
Współczynnik determinacji

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

gdzie:
n – liczba prognoz
$\hat{y}_t$ – przewidywana wartość zmiennej
y_t – rzeczywista wartość zmiennej
$\bar{y}$ – średnia rzeczywista wartość zmiennej

Współczynnik determinacji informuje nas, na ile nasz model, nasz badany czynnik wyjaśnia zgromadzone dane pomiarowe (zmienną zależną). Analizowany **czynnik** może mieć wpływ na zmienną zależną, pytanie jak mocny jest jego wpływ.

Współczynnik determinacji przyjmuje wartości pomiędzy 0 a 1, jednakże najbardziej popularną formą prezentacji tego współczynnika jest jego procentowa postać.



WAŻNE

Definicja:
Średni błąd predykcji (mean erros – ME)

$$ME=\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

Błąd Me powinien być bliski 0 (wyraźne odchylenia od zera oznaczają, że predykcja jest obciążona), ponieważ błędy in plus są kompensowane błędami in minus.

ME <0 prognozy są przeciętnie przeszacowane
ME = 0 prognozy są przeciętnie równe realizacjom
ME > 0 prognozy są przeciętnie niedoszacowane

Kolejną miarą jest średni błąd bezwzględny MAE (mean absolute error), informuje on, o ile średnio w okresie prognoz będzie wynosić odchylenie od wartości rzeczywistej.



WAŻNE

Definicja:

Średni absolutny błąd predykcji MAE (mean absolute error)

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^*|$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

Porównanie ME i MAE daje informacje o systematycznie wyższych/niższych prognozach niż realizacjach (w przypadku podobnych wartości obu mierników). Różne od siebie MAE i ME oznaczają, że błędy mają odmienne znaki.

Następną miarą, jaką się zajmiemy, jest bliźniacza miara MAE. Średni bezwzględny błąd procentowy MAPE (mean absolute percentage error) informuje o średniej wielkości błędów prognoz dla okresu testowego, wyrażonych w procentach. Wartość MAPE pozwala na porównanie dokładności prognoz różnych modeli.



WAŻNE

Definicja:
Średni absolutny błąd procentowy prognozy MAPE (mean absolute percentage error)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - y_t^*|}{y_t} * 100\%$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej



WAŻNE

Definicja:
Średniokwadratowy predykcji RMSE (root mean square error):

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2}$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

RMSE mierzy, o ile średnio odchylają się realizacje zmiennej prognozowanej od obliczonych prognoz.



WAŻNE

Definicja:

Średni procentowy błąd prognoz (mean percentage error)

$$MPE = 1/n \sum_{t=1}^{n} \frac{|y_t - y_t^*|}{y_t}$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

MPE informuje, jaki procent rzeczywistych realizacji zmiennej prognozowanej stanowią błędy prognozy w okresie predykcji.

3.5. Pytania sprawdzające



1. Co to jest prognozowanie?

2. Omów znane ci rodzaje prognoz.

3. Podaj przykłady prognoz gospodarczych oraz społecznych.



4. Jakie są główne założenia teorii predykcji?

5. Jakie znasz niematematyczne metody prognozowania?

6. Wymień i omów kolejne etapy prognozowania.

7. Czym różni się prognoza punktowa od przedziałowej?



8. Jakie są zasady wykonywania prognozy punktowej?

9. Co to znaczy, że prognoza jest trafna?



10. Co to znaczy, że prognoza jest dopuszczalna?

11. Omów miary stosowane do oceny trafności prognozy.



4. Analiza szeregów czasowych

Dane charakteryzujące rozwój zjawisk gospodarczych lub społecznych są odnotowywane i gromadzone wraz z upływem czasu, tworząc ciągi liczbowe określane mianem szeregów czasowych. **Szereg czasowy** to uporządkowany zbiór obserwacji statystycznych (zbiór stanów zmiennych), charakteryzujących zmiany poziomu określonego zjawiska w czasie.

Przykład

W tabeli poniżej przedstawiono szereg czasowy dotyczący liczby rowerów dziecięcych będących w magazynie sklepu rowerowego, w momencie pierwszego dnia każdego miesiąca.

Data	01.10. 2020	01.11. 2020	01.12. 2020	01.01. 2021	01.02. 2021
Liczba rowerów dziecięcych	20	50	10	30	20

Przykład

W tabeli poniżej przedstawiono szereg czasowy dotyczący liczby rowerów dziecięcych będących w magazynie sklepu rowerowego. W szeregu dynamicznym okresami są poszczególne kwartały 2020 roku, stanowią łącznie pewien przedział czasowy.

Kwartał	I	II	III	IV
Liczba rowerów dziecięcych	221	153	219	230

4.1. Składowe szeregów czasowych

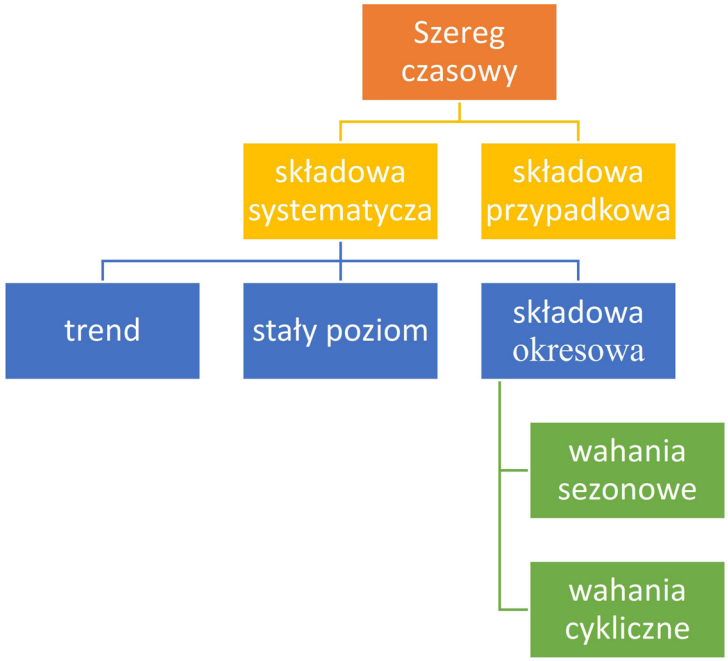
Poziom badanego zjawiska Y_t jest funkcją następujących składników:

- T_t – trendu,
- C_t – wahań cyklicznych,
- S_t – wahań sezonowych,
- I_t - wahań przypadkowych.

Powyższe składowe mogą wystąpić w szeregach czasowych w różnej kombinacji, różnych relacjach oraz mieć różny udział w objaśnianiu zmian występujących w przeszłości. Przyjmuje się jednocześnie założenie (które nie zawsze jest spełnione) o braku powiązania pomiędzy poszczególnymi składowymi[8].

Składowe te mogą mieć charakter zjawiska systematycznego lub ich występowanie może być dziełem przypadku. Dlatego też składowe szeregu czasowego dzielimy na dwie podstawowe grupy: składowe o charakterze systematycznym oraz składowe przypadkowe. Składowe o charakterze systematycznym to: trend, stały poziom oraz składowe okresowe, które mogą mieć charakter sezonowy lub cykliczny. Składowe przypadkowe charakteryzują działanie czynników o charakterze losowym i są nieprzewidywalne co do siły oraz kierunku.

[8] B. Borkowski, Ł. Cygański, D. Czerwińska-Kayzer, J. Florek, M. Hamulczuk, M. Jerzak, A. Wąs, Zarządzanie ryzykiem cenowym a możliwości stabilizowania dochodów producentów rolnych, Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy, Warszawa 2008.



Rysunek 39. Składowe szeregu czasowego.

Trend (tendencja rozwojowa) – długookresowa skłonność do jednokierunkowych zmian (wzrostu lub spadku) wartości zmiennej badanej. Jest konsekwencją działania stałych czynników, np. w przypadku sprzedaży – liczba potencjalnych klientów, ich dochody lub preferencje. Może być wyznaczony, gdy mamy długi ciąg obserwacji.

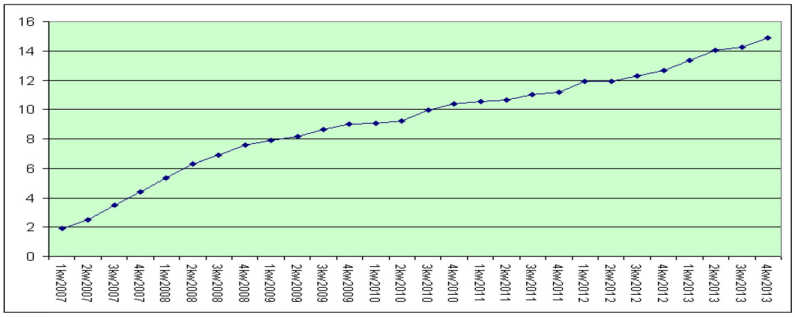
Stały (przeciętny poziom) – występuje, gdy w szeregu czasowym nie ma trendu, zaś wartości badanej zmiennej oscylują wokół pewnego stałego poziomu.

Wahania cykliczne – długookresowe wahania wokół trendu lub stałego poziomu. W ekonomii najczęściej związane z cyklem koniunkturalnym gospodarki.

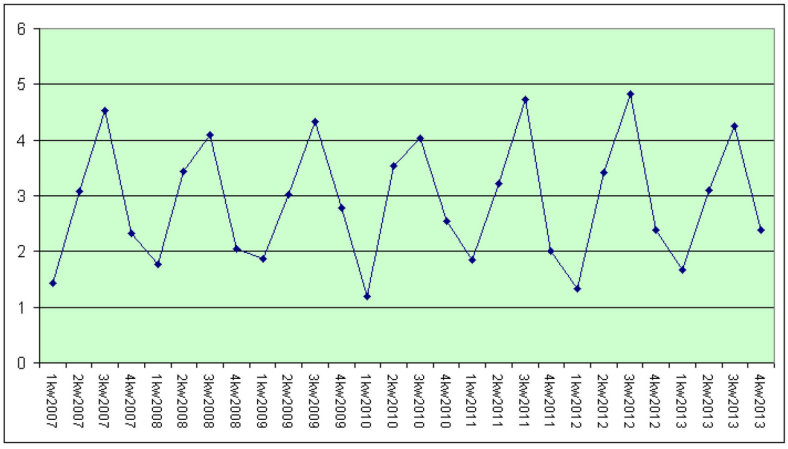
Wahania sezonowe – wahania wokół trendu lub stałego poziomu. Wahania te mają skłonności do powtarzania się w określonym czasie nieprzekraczającym jednego roku, odzwierciedlają wpływ pogody lub kalendarza na działalność gospodarczą.

Dekompozycja szeregu czasowego – wyodrębnienie poszczególnych składowych.

Przykład



Rysunek 40. Trend.



Rysunek 41. Składowa okresowa.

W różnych modelach, addytywnym i multiplikatywnym, stosowane są te same metody wyodrębniania czystego trendu. Najczęściej stosowanymi w praktyce metodami wyodrębniania czystego trendu są metoda mechaniczna (patrz rozdział Metoda średniej

ruchomej) oraz metoda analityczna. Metoda analityczna polega na dopasowaniu odpowiedniej postaci funkcji matematycznej do danych empirycznych.

4.2. Model addytywny i multiplikatywny

Jednym z rodzajów szeregu statystycznego jest szereg czasowy, który można zdefiniować jako ciąg obserwacji jakiegoś zjawiska w kolejnych jednostkach czasu (latach, kwartałach, miesiącach).

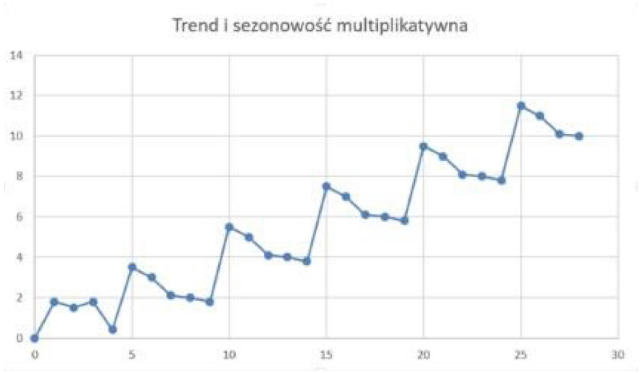
Najczęściej rozważa się cztery czynniki wpływające na rozwój zjawiska w czasie:

- **trend** – długookresowe, systematyczne zmiany, jakim podlega dane zjawisko,
- **wahania sezonowe** – regularne odchylenia od tendencji rozwojowej (trendu) związane np. z porami roku (warunkami klimatycznymi),
- **wahania cykliczne** – związane z cyklem koniunkturalnym,
- **wahania przypadkowe** – nieregularne zmiany.

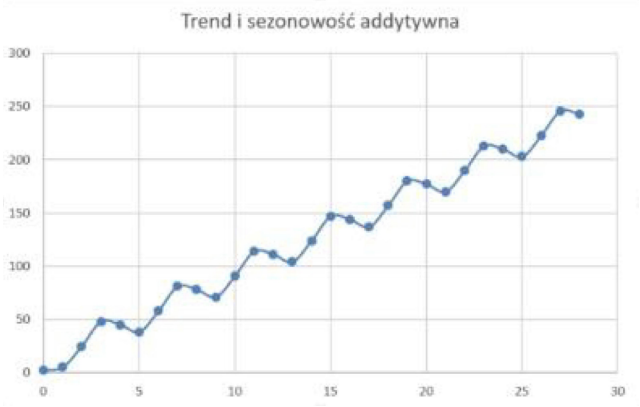


Rysunek 42. Składowe szeregów czasowych.

Przykład


Trend i sezonowość multiplikatywna
14
12
10
8
6
4
2
0
0
5
10
15
20
25
30


Brak trendu i sezonowość multiplikatywna
120
100
80
60
40
20
0
0
5
10
15
20
25
30


Trend i sezonowość addytywna
300
250
200
150
100
50
0
0
5
10
15
20
25
30



Konstruowany model szeregu czasowego może mieć różną postać. Na ogół przyjmuje się addytywną lub multiplikatywną postać modelu.

W modelu addytywnym zakłada się, że obserwowane wartości zmiennej prognozowanej są sumą (wszystkich lub niektórych) składowych szeregu czasowego.

W modelu multiplikatywnym przyjmuje się, że obserwowane wartości zmiennej prognozowanej są iloczynem składowych szeregu czasowego.

WAŻNE

Wzory
Model addytywny

$$y_t = f(t) + g(t) + h(t) + S_t$$

gdzie:
$f(t)$ – funkcja trendu
$g(t)$ – funkcja czasu charakteryzująca wahania sezonowe
$h(t)$ – funkcja czasu charakteryzująca wahania cykliczne
S_t – składnik losowy

WAŻNE

Wzory
Model multiplikatywny

$$y_t = f(t) \cdot g(t) \cdot h(t) \cdot S_t$$

gdzie:
$f(t)$ – funkcja trendu
$g(t)$ – funkcja czasu charakteryzująca wahania sezonowe
$h(t)$ – funkcja czasu charakteryzująca wahania cykliczne
S_t – składnik losowy

4.3. Sezonowość

Wahania sezonowe – wahania wokół trendu lub stałego poziomu. Wahania te mają skłonności do powtarzania się w określonym czasie nieprzekraczającym jednego roku, odzwierciedlają wpływ pogody lub kalendarza na działalność gospodarczą.

Główne przyczyny występowania wahań sezonowych to:
- pogoda, pora roku,
- planowane zjawiska np. święta,
- typowe zachowania społeczne, np. wyjazdy zimowe odbywają się najczęściej w okresie szkolnych ferii zimowych.

Wyróżniamy dwa rodzaje wahań sezonowych:
- multiplikatywne,
- addytywne.

Jedną z najczęściej stosowanych metod przy analizie wahań sezonowych jest **metoda wskaźników**. **Metoda ta składa się z następujących etapów:**
1. wyodrębnienie trendu,
2. eliminację trendu z szeregu czasowego,

3. eliminację wahań sezonowych,
4. wyznaczenie wskaźników sezonowości.

W modelu addytywnym **wskaźniki sezonowości** obliczamy jako różnicę pomiędzy wartościami rzeczywistymi (Y_t) i teoretycznymi ($\hat{Y}_t$) dla poszczególnych okresów (półroczy, kwartałów, miesięcy) zgodnie ze wzorem:

$$WS_t^i = Y_t - \hat{Y}_t$$

Dla każdego kwartału wyznaczamy „średni" surowy wskaźnik sezonowości:

$$WS_I = \frac{ws_I^1 + ws_I^2 + ws_I^3 + \ldots + ws_I^r}{r},$$

$$WS_{II} = \frac{ws_{II}^1 + ws_{II}^2 + ws_{II}^3 + \ldots + ws_{II}^r}{r},$$

$$WS_{III} = \frac{ws_{III}^1 + ws_{III}^2 + ws_{III}^3 + \ldots + ws_{III}^r}{r},$$

$$WS_{IV} = \frac{ws_{IV}^1 + ws_{IV}^2 + ws_{IV}^3 + \ldots + ws_{IV}^r}{r}.$$

Surowe wskaźniki sezonowości zawierają (oprócz wahań sezonowych) wahania przypadkowe. W celu eliminacji wahań przypadkowych dokonujemy korekty surowych wskaźników sezonowości poprzez odjęcie od nich tak zwanego wskaźnika korygującego o postaci:

$$W_K = \frac{\sum_{t=1}^{d} WSt}{d}$$

gdzie: d – to liczba podokresów w cyklu sezonowości.

Wobec tego skorygowany wskaźnik sezonowości obliczamy:

$$WSS_t = WS_t - WSS_k$$

gdzie:
WSS_t – skorygowany wskaźnik sezonowości dla kwartału t,
WS_t – średni surowy wskaźnik sezonowości dla kwartału t,
W_k – wskaźnik korygujący.

Suma skorygowanych wskaźników sezonowości powinna być równa liczbie podokresów w cyklu sezonowości.

W **modelu multiplikatywnym** zamiast różnic pomiędzy wartościami teoretycznymi a modelowymi oblicza się ich iloraz zgodnie ze wzorem:

$$WS_t^i = Y_t / \widehat{Y}_t$$

Następnie dla każdego z analizowanych podokresów (półroczy, kwartałów, miesięcy) oblicza się **surowe wskaźniki sezonowości**, uśredniając wyznaczone wcześniej ilorazy:

$$WS_I = \frac{ws_I^1 + ws_I^2 + ws_I^3 + \ldots + ws_I^r}{r},$$

$$WS_{II} = \frac{ws_{II}^1 + ws_{II}^2 + ws_{II}^3 + \ldots + ws_{II}^r}{r},$$

$$WS_{III} = \frac{ws_{III}^1 + ws_{III}^2 + ws_{III}^3 + \ldots + ws_{III}^r}{r},$$

$$WS_{IV} = \frac{ws_{IV}^1 + ws_{IV}^2 + ws_{IV}^3 + \ldots + ws_{IV}^r}{r}.$$

Ostatnim elementem analizy sezonowości jest **ocena jakości otrzymanego modelu**. W takim przypadku nie wyznaczamy współczynnika. Główną miarą jakości jest odchylenie standardowe składnika resztowego z uwzględnieniem sezonowości, według wzoru:

$$S_u = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n-2}}$$



UWAGA

Sposób postępowania w metodzie wskaźnikowej obejmuje kolejno:

- wyodrębnienie trendu,
- eliminację trendu z szeregu czasowego,
- eliminację wahań przypadkowych,
- obliczenie wskaźników sezonowości.

4.4. Metoda naiwna

Omówimy teraz metodę prognostyczną dotyczącą **analizy szeregów czasowych bez tendencji**.

Metodę naiwną warto stosować tylko przy niewielkich wahaniach przypadkowych. O sile tych wahań informuje wielkość współczynnika zmienności badanego szeregu czasowego:

$$V = \frac{s}{\bar{x}} \cdot 100\%$$

gdzie:
s – odchylenie standardowe,
x – średnia arytmetyczna.

W metodzie tej konstruuje się prognozę zmiennej na moment *t* na poziomie obserwowanej wartości tej zmiennej w momencie *t* – 1

WAŻNE

Definicja

W modelu naiwnym wykorzystywanym w takich przypadkach konstruuje się prognozę zmiennej na moment lub okres T na poziomie zaobserwowanej wartości tej zmiennej w momencie lub okresie t – 1.

$$y^*_t = y_{t-1}$$

gdzie:
y^*_t – prognoza zmiennej y wyznaczona na moment lub okres t
y_{t-1} – wartość zmiennej prognozowanej w momencie lub okresie t-1

Poniżej przedstawiamy kilka innych wersji metody naiwnej:

$$y_t^* = y_{t-1} + C$$

Prognoza równa jest wartości z okresu poprzedniego powiększona o pewną stałą wielkość C ustaloną z góry. Ma zastosowanie, gdy występuje liniowa tendencja rozwojowa i wahania przypadkowe.

$$y_t^* = y_{t-1} \cdot \sqrt[t-2]{\frac{y_{t-1}}{y_{t-2}} \cdot \frac{y_{t-2}}{y_{t-3}} \cdot \ldots \cdot \frac{y_2}{y_1}} = y_{t-1} \cdot \sqrt[t-2]{\frac{y_{t-1}}{y_1}}$$

Prognoza równa jest wartości z okresu poprzedniego powiększona o średni przyrost względny poziomu zjawiska z poprzednich okresów, dla których zgromadzono materiał empiryczny (średniookresowe tempo zmian poziomu zjawiska).

$$y_t^* = y_{t-r} \cdot (1 + C)$$

Prognoza równa jest wartości z tego samego podokresu poprzedniego cyklu sezonowości powiększona o pewien ustalony z góry procent (C).

UWAGA

Zalety metody:
- prosta i łatwa do zrozumienia,
- szybka i tania.

Wady metody:
- niska jakość prognoz,
- brak możliwości oceny ex ante.

4.5.Metoda średniej ruchomej

Druga metoda wyrównywania szeregów czasowych, **metoda średniej ruchomej prostej i ważonej**, wykorzystuje średnie ruchome zarówno do eliminacji wahań przypadkowych, jak i okresowych.

Przy prognozowaniu za pomocą średniej ruchomej zakłada się, że wartość zmiennej prognozowanej w następnym okresie będzie równa średniej arytmetycznej z k ostatnich wartości średniej.

W prognostycznym modelu średniej ruchomej przyjmuje się, że prognoza zmiennej na moment *t* jest:

- **średnią arytmetyczną (w przypadku średniej ruchomej prostej):**

$$y_t^* = \frac{1}{k} \sum_{i=t-k}^{t-1} y_i$$

- **średnią ważoną (w przypadku średniej ruchomej ważonej)**

$$y_t^* = \sum_{i=t-k}^{t-1} y_i \cdot w_i$$

z *k* ostatnich obserwacji w szeregu.

W metodzie średniej ruchomej ważonej **wagi (w_i)** poszczególnych obserwacji nadaje prognosta. Wagi powinny być z przedziału (0,1), a ich suma powinna być równa 1.

Liczba składników średniej arytmetycznej ruchomej powinna odpowiadać wahaniom okresowym lub ich wielokrotności. Dla wyeliminowania wahań okresowych często niezbędne jest wyznaczenie średniej ruchomej z parzystej liczby wyników, jeżeli ta liczba wyników oryginalnego szeregu odpowiada liczbie pomiarów w cyklu wahań.

Użycie średnich ruchomych do **wygładzenia szeregu czasowego** jest jednym z prostszych sposobów wyznaczania głównej tendencji rozwojowej. Metoda ta jest poglądowa i daje wyniki proste do interpretowania, jak i ułatwia prognozowanie. W przypadku występowania w szeregu czasowym trendu liniowego można do konstrukcji prognozy użyć modelu podwójnej średniej ruchomej.

UWAGA

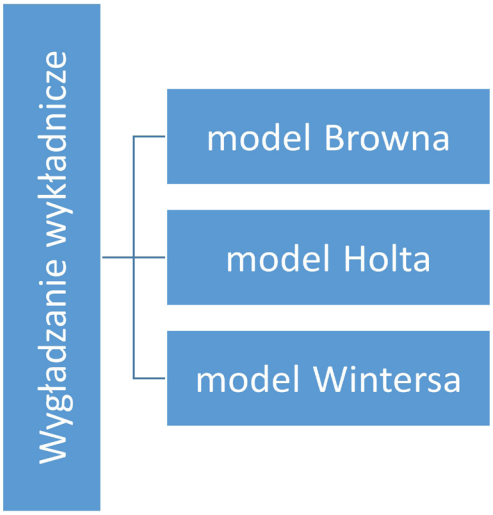
Metoda średniej ruchomej:
- **wygładza dane,**
- **może wyeliminować wahania (sezonowe) i nieregularności,**
- **pomija się część danych (początkowych),**
- **w pewnym stopniu jest wrażliwa na obserwacje nietypowe.**

4.6. Wygładzanie wykładnicze

Kolejną metodą prognozowania zmiennej Y jest **metoda wygładzania wykładniczego,** która polega na wygładzaniu szeregu czasowego przy pomocy średniej ruchomej ważonej, jednakże wagi określanej według prawa wykładniczego.

Do najpowszechniej znanych i stosowanych należą tu:

- model wygładzania wykładniczego Browna,
- model liniowy Holta
- model Wintersa – wersja addytywna i multiplikatywna.



Rysunek 43. Modele wygładzania wykładniczego.

W modelu **wygładzania Browna** prognoza zmiennej na moment *t* równa jest prognozie tej zmiennej na moment *t* – 1 skorygowanej o iloczyn parametru wygładzania α i wartości jej bezwzględnego błędu ex post:

$$y_t^* = \alpha \cdot y_{t-1} + (1 - \alpha) \cdot y_{t-1}^*$$

gdzie:
y^*_t – prognoza w wyznaczonym okresie t
y_{t-1} – wielkość w okresie t-1

Parametr α wybiera się na podstawie kryterium najmniejszego błędu *ex post* prognoz wygasłych.

W przypadku prostego modelu wygładzania wykładniczego niezbędne jest ustalenie wartości początkowej y^*_1. Zwykle przyjmuje się ją jako pierwszą wartość rzeczywistą zmiennej, czyli y_1 lub średnią arytmetyczną rzeczywistych wartości tej zmiennej z próbki wstępnej.

Prosty **model wygładzania wykładniczego Browna** zasadniczo stosujemy w przypadku występowania **prawie stałego poziomu zmiennej prognozowanej oraz niewielkich wahań przypadkowych**. Konstruując prognozę na okres t, przyjmuje się tutaj, że wyniesie ona tyle, ile prognoza wyznaczonej na okres poprzedni t-1, po skorygowaniu jej o pewną część (α – stałą wygładzania) jej bezwzględnego błędu ex post.

Zatem, jeżeli wyznaczona prognoza na okres t-1 była, w porównaniu z rzeczywistą wartością x_{t-1}, zbyt niska to prognoza konstruowana na okres t zwiększa się (porównując z poprzednią), i odwrotnie, jeśli wyznaczona prognoza na okres t-1 była, w porównaniu z prawdziwą wartością zmiennej prognozowanej x_{t-1}, to prognoza konstruowana na okres t zmniejszy się (w porównaniu z poprzedzającą).

Kolejna metoda wygładzania wykładniczego to **model Holta** opisany następującymi równaniami:

$$F_{t-1} = \alpha y_{t-1} + (1-\alpha)(F_{t-2} + S_{t-2})$$
$$S_{t-1} = \beta(F_{t-1} + S_{t-2}) + (1-\alpha)S_{t-2}$$

gdzie:
F_{t-1} – ocena wartości średniej w okresie t-1
S_{t-1} – przyrost trendu w okresie t-1
α, β – parametry modelu z przedziału (0,1]

Natomiast równanie prognozy ma następującą postać:

$$Y_t^* = F_n + (t-n)S_n$$

Do opisu **modelu addytywnego Wintera** służy następujący układ równań:

$$F_t = \alpha(Y_t - C_{t-r}) + (1-\alpha)(F_{t-1} + S_{t-1})$$
$$S_t = \beta(F_t - F_{t-1}) + (1-\beta)S_{t-1}$$
$$C_t = \gamma(Y_t - F_t) + (1-\gamma)C_{t-r}$$

gdzie:
F_t – oszacowanie części stałej szeregu podlegającej wahaniom przypadkowym
S_t – oszacowanie części przyrostowej szeregu podlegającej wahaniom przypadkowym
C_t – oszacowanie części sezonowej trendu podlegającej wahaniom przypadkowym
Y_t – wartości empiryczne szeregu w okresie t
y^*_t – wartości teoretyczne szeregu wygładzonego
α, β, γ – parametry modelu z przedziału (0,1]

Natomiast równanie prognozy ma następującą postać:

$$Y_t^* = F_n + (t-n)S_n + C_{t-r}$$

Do opisu **modelu multiplikatywnego Wintera** służy następujący układ równań:

$$F_t = \alpha(Y_t / C_{t-r}) + (1-\alpha)(F_{t-1} + S_{t-1})$$
$$S_t = \beta(F_t - F_{t-1}) + (1-\beta)S_{t-1}$$
$$C_t = \gamma(Y_t / F_t) + (1-\gamma)C_{t-r}$$

gdzie:
F_t – oszacowanie części stałej szeregu podlegającej wahaniom przypadkowym
S_t – oszacowanie części przyrostowej szeregu podlegającej wahaniom przypadkowym

C_t – oszacowanie części sezonowej trendu podlegającej wahaniom przypadkowym
Y_t – wartości empiryczne szeregu w okresie t
y^*_t – wartości teoretyczne szeregu wygładzonego
α, β, γ – parametry modelu z przedziału (0,1]

Natomiast równanie prognozy ma następującą postać:

$$Y_t^* = (F_n + (t - n)S_n)C_{t-r}$$

Wartości początkowe F i S można obliczyć, wykorzystując średnią ruchomą 3-elementową.



4.7. Pytania sprawdzające

1. Co to jest szereg czasowy?

……
……
……
……
……
……
……
……
……

2. Jakie są składowe szeregów czasowych?

……
……
……
……
……
……
……
……
……

3. Co to jest trend, wahania cykliczne i dekompozycja szeregu?

……
……
……
……
……
……
……
……
……

4. Omów model addytywny i multiplikatywny szeregu.



5. Jaką znasz metodę stosowaną do analizy wahań sezonowych? Omów tę metodę.

6. Wymień znane ci metody prognostyczne.



7. Jakie wady/zalety ma metoda naiwna?

8. Omów metodę średniej ruchomej.

9. Jakie znasz metody wygładzania wykładniczego? Omów je.

Załączniki

Załącznik 1 – tablice rozkładu t-studenta

r	α													r
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001	
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	53,657	636,619	1
2	,142	,289	,445	,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	6,925	31,598	2
3	,137	,277	,424	,584	,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941	3
4	,134	,271	,414	,569	,741	,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610	4
5	,132	,267	,408	,559	,727	,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859	5
6	,131	,265	,404	,553	,718	,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959	6
7	,130	,263	,402	,549	,711	,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405	7
8	,130	,262	,399	,546	,706	,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041	8
9	,129	,261	,398	,543	,703	,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781	9
10	,129	,260	,397	,542	,700	,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587	10
11	,129	,260	,396	,540	,697	,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437	11
12	,128	,259	,395	,539	,695	,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318	12
13	,128	,259	,394	,538	,694	,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221	13
14	,128	,258	,393	,537	,692	,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140	14
15	,128	,258	,393	,536	,691	,866	1,074	1,341	1,753	2,181	2,602	2,947	4,073	15
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015	16
17	,128	,257	,392	,534	,689	,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965	17
18	,127	,257	,392	,534	,688	,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922	18
19	,127	,257	,391	,533	,688	,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883	19
20	,127	,257	,391	,533	,687	,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850	20
21	,127	,257	,391	,532	,686	,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819	21
22	,127	,256	,390	,532	,686	,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792	22
23	,127	,256	,390	,532	,685	,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767	23
24	,127	,256	,390	,531	,685	,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745	24
25	,127	,256	,390	,531	,684	,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725	25
26	,127	,256	,390	,531	,684	,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707	26
27	,127	,256	,389	,531	,684	,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690	27
28	,127	,256	,389	,530	,683	,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674	28
29	,127	,256	,389	,530	,683	,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659	29
30	,127	,256	,389	,530	,683	,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646	30
40	,126	,255	,388	,529	,681	,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551	40
60	,126	,254	,387	,527	,679	,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460	
120	,126	,254	,386	,526	,677	,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373	120
∞	,126	,253	,385	,524	,674	,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291	∞

Załącznik 2 – tablice rozkładu chi-kwadrat

r	α														
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	$0,^{3}157$	$0,^{3}628$	0,0039	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,0201	0,0404	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,268
4	0,297	0,429	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,465
5	0,554	0,752	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,517
6	0,872	1,134	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	1,239	1,564	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	1,646	2,032	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
9	2,088	2,532	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	2,558	3,059	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	3,053	3,609	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	3,571	4,178	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	4,107	4,765	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	4,660	5,368	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	5,229	5,985	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	5,812	6,614	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	6,408	7,255	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	7,015	7,906	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	7,633	8,567	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	8,260	9,237	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	8,897	9,915	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,797
22	9,542	10,600	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	10,196	11,293	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,728
24	10,856	11,992	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	11,524	12,697	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,620
26	12,198	13,409	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,052
27	12,879	14,125	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,476
28	13,565	14,847	16,928	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278	56,893
29	14,256	15,574	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588	58,302
30	14,953	16,306	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892	59,703

Literatura

1. Barczak, A. St., Biolik, J. (1998). Podstawy ekonometrii. Katowice: Wydawnictwo Akademii Ekonomicznej w Katowicach.
2. Barczak, A. S. (2007). Ekonometria i nauki o zarządzaniu. Katowice: Prace Naukowe/Akademia Ekonomiczna w Katowicach, s. 433-437.
3. Borkowski, B., Dudek, H., Szczesny, W. (2003). Ekonometria. Wybrane zagadnienia. Warszawa: Wydawnictwo Naukowe PWN.
4. Cieślak, M. (Ed.). (2008). Prognozowanie gospodarcze: metody i zastosowania. Warszawa: Wydawnictwo Naukowe PWN.
5. Dziechciarz, J. (2003). Ekonometria, Metody, przykłady, zadania. Wrocław: Wydawnictwo AE we Wrocławiu.
6. Górecki, B. R. (2010). Ekonometria. Podstawy teorii i praktyki, Warszawa: Key Text.
7. Gruszczyński, M., Podgórka, M. (2000). Ekonometria. Warszawa: Oficyna Wydawnicza SGH.
8. Kopczewska, K. (2019). Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN. Warszawa: CeDeWu.
9. Kufel, T. (2013). Ekonometria: rozwiązywanie problemów z wykorzystaniem programu GRETL. Warszawa: Wydawnictwo Naukowe PWN.
10. Nowak, E. (2002). Zarys metod ekonometrii. Zbiór zadań. Warszawa: Wydawnictwo Naukowe PWN.
11. Nowak, E. (Ed.). (1998). Prognozowanie gospodarcze: metody, modele, zastosowania, przykłady: praca zbiorowa. Warszawa: Agencja Wydawnicza Placet.
12. Snarska, A. (2011). Statystyka, ekonometria, prognozowanie: ćwiczenia z Excelem 2007. Warszawa: Wydawnictwo Placet.
13. Welfe, A. (2018). Ekonometria. Warszawa: PWE.

Grafika użyta do rozdzielania poszczególnych rozdziałów:

1. Ad. 1 – janoon028/Freepik.com
2. Ad. 2 – pressfoto/Freepik.com
3. Ad. 3 – rawpixel.com/Freepik.com
4. Ad. 4 – Mediamodifier/Pixabay.com

EKONOMETRIA

Część druga
Studia przypadków

Anna Tatarczak

Lublin 2021

Spis treści

Wstęp 127

Studium przypadku 1. Prezentacja graficzna danych 128
Przypadek 1 129
Przypadek 2 134
Przypadek 3 138
Przypadek 4 144
Zadania 149

Studium przypadku 2. Modele ekonometryczne 151
Przypadek 1 154
Przypadek 2 158
Przypadek 3 165
Przypadek 4 170
Zadania 176

Studium przypadku 3. Szeregi czasowe – prognozowanie 177
Przypadek 1 182
Przypadek 2 187
Przypadek 3 190
Zadania 196

Załączniki 198
Załącznik 1 – instalacja dodatku Analiza danych 198
Załącznik 2 – oprogramowanie GRETL 200
Załącznik 3 – funkcje macierzowe EXCELA 203
Załącznik 4 – tablice rozkładu normalnego 204
Załącznik 5 – tablice rozkładu t-Studenta 205

Bibliografia 206

Wstęp

Głównym celem opracowania jest uporządkowanie i opis podstawowych metod i technik stosowanych na wykładzie z ekonometrii. Z wykładem związane są ćwiczenia laboratoryjne, w trakcie których są praktycznie ilustrowane zagadnienia przedstawiane na wykładzie. Podstawowymi narzędziami programowymi, stosowanymi do opisu poszczególnych procedur obliczeniowych, jak i do obliczeń w laboratorium komputerowym są programy EXCEL, Power BI, GRETL oraz STATISTICA.

Opracowanie jest głównie przeznaczone dla studentów Wyższej Szkoły Ekonomii i Innowacji poznających podstawy ekonometrii. Mogą z niego również korzystać praktycy prowadzący badania własne lub kierujący pracami badawczymi.

Zostały opisane trzy główne case-study (każde z podziałem na mini-casy), które są najczęściej wykorzystywane w pracy ekonometryka: prezentacja graficzna, tworzenie modeli i predykcja.

Wszystkie rozdziały mają podobną strukturę: na początku przedstawiony jest cel ćwiczeń oraz problem, który będzie analizowany. Następnie czytelnik zapozna się z krótkim wstępem teoretycznym. Główna treść każdego rozdziału to rozwiązane przykłady praktyczne ze szczegółowymi opisami w postaci komentarzy. Ponieważ najlepszą nauką ekonometrii jest praktyka, więc każdy z case'ów kończy się ćwiczeniami do wykonania.

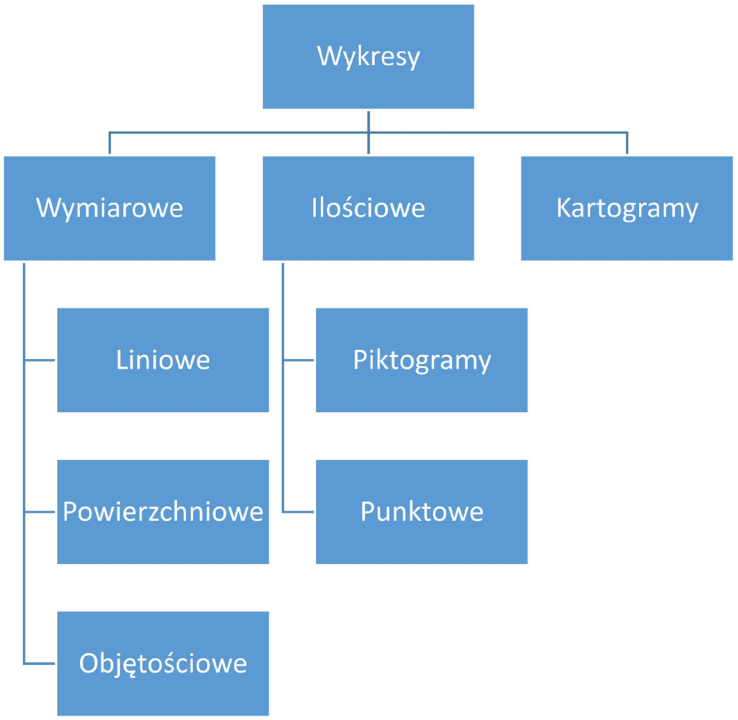
Dołączona bibliografia przedmiotu umożliwi zainteresowanym pogłębieniem wiedzy Czytelnikom dotarcie do innych podręczników z ekonometrii.

Studium przypadku 1.
Prezentacja graficzna danych

Wstęp teoretyczny

Ważną rolę w ekonometrii odgrywa **graficzna prezentacja danych statystycznych**. Dzięki wykresom można przeprowadzić nie tylko wstępną analizę zjawiska czy zbiorowości, ale także dokonać doboru odpowiednich parametrów opisu rozkładu cechy. Graficzna prezentacja danych jest mniej dokładna niż prezentacja tabelaryczna, jednak pozwala ona na szybką wizualizację dotyczącą kształtowania się poziomu danego zjawiska.

Istnieje wiele różnych **typów wykresów**. W literaturze przedmiotu można spotkać podział na **wykresy wymiarowe, wykresy ilościowe i kartogramy**. Podział ten został przedstawiony schematycznie na rysunku poniżej.



Wykres statystyczny służy do prezentacji zebranego materiału statystycznego za pomocą obrazu graficznego, tj. wielkości, kształtu lub barwy.

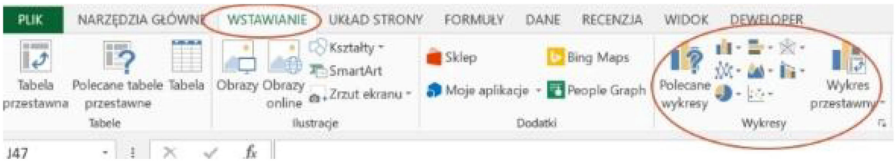
Poprawnie wykonany wykres powinien składać się z następujących elementów:

- tytułu,
- pola wykresu,
- skali,
- legendy,
- źródła danych statystycznych i ewentualnych uwag wyjaśniających.

W programie EXCEL dostępnych jest wiele typów wykresów.

Aby wstawić jeden z nich, należy:

1. Zaznaczyć tabelę z danymi.
2. Wybrać w kategorii WSTAWIANIE z sekcji WYKRESY ikonkę wykresu kołowego.
3. Wybrać odpowiedni podtyp wykresu (kołowy, pierścieniowy, 3-W).



Do prezentacji graficznej oprócz programu EXCEL będziemy wykorzystywać również **oprogramowanie GRETL** oraz **POWER BI**.

Przypadek 1

Polecenie:

Dane są obserwacje dotyczące zysku pewnego sklepu spożywczego (w tys. zł) w 20 kolejnych dniach roboczych:

16,4; 12,5; 14,4; 13,0; 14,4; 14,4; 13,9; 11,1; 16,5; 15,8; 13,0; 15,8; 18,5; 10,5; 16,4; 15,7; 16,5; 16,9; 11,1; 16,4

Zaproponuj graficzną prezentację powyższych danych z wykorzystaniem programu EXCEL.

Rozwiązanie:

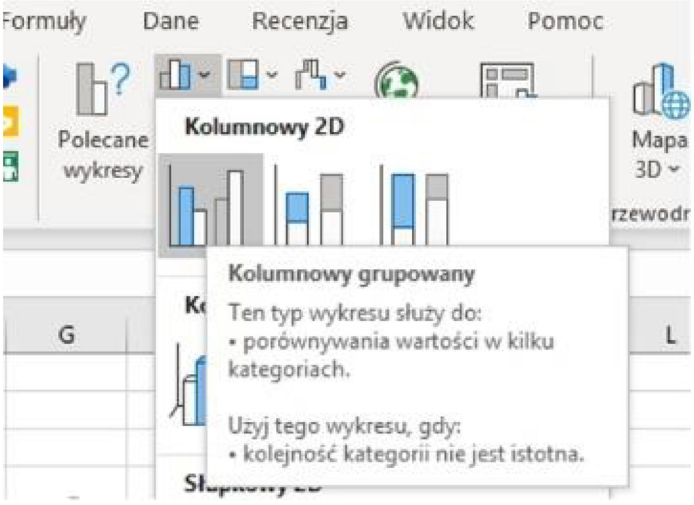
Krok 1.

Po przeniesieniu danych do arkusza excel klikamy lewym przyciskiem myszy na A w celu zaznaczenia całej kolumny.

Krok 2.

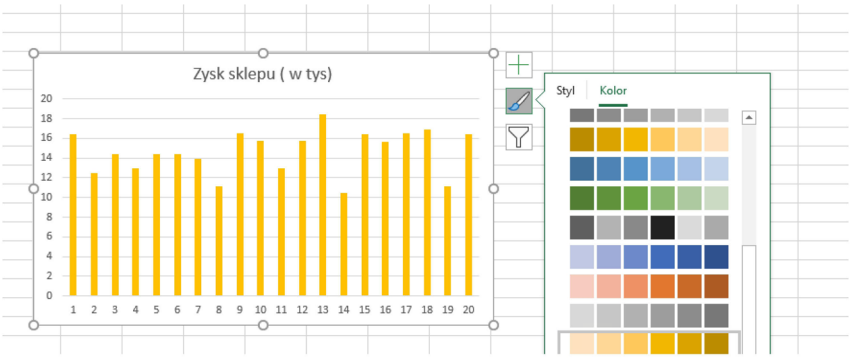
Następnie, gdy dane są już zaznaczone, wybieramy *Wstawianie/ Wykresy*. Spośród proponowanych wykresów wybieramy wykres kolumnowy grupowany.





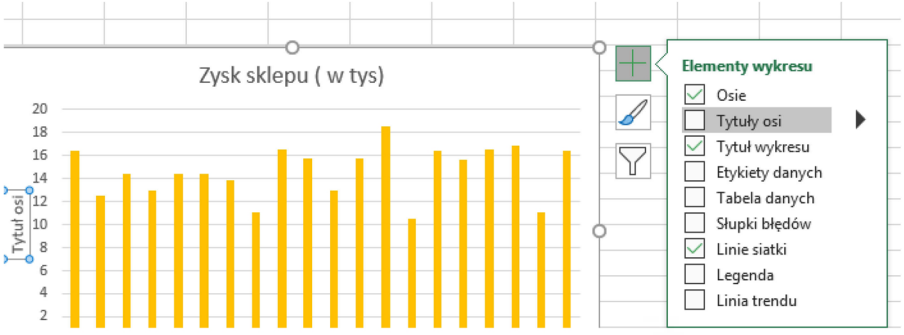
Krok 3.

W celu zmiany tytułu wykresu wybieramy *Tytuł wykresu* widoczny nad wykresem i wpisujemy odpowiedni tytuł zgodny z treścią zadania. W celu zmiany koloru wykresu klikamy odpowiednio na *Style wykresu* kolor i wybieramy kolor np. niebieski.



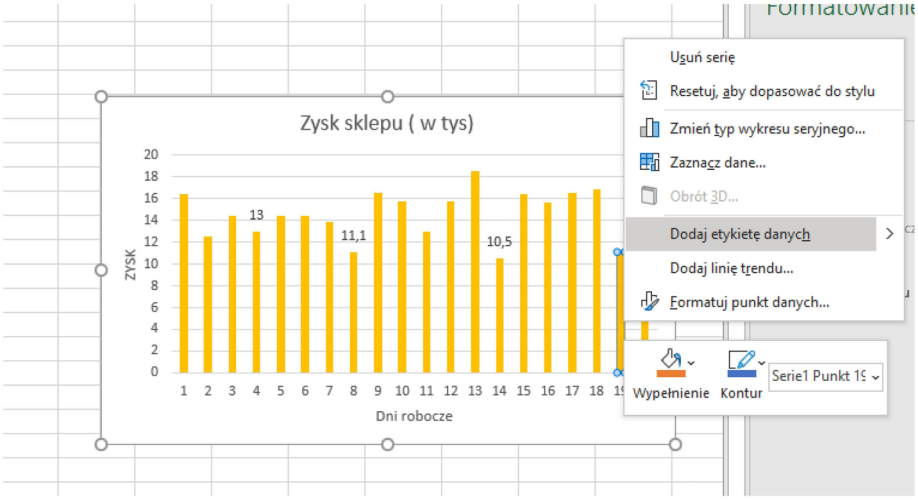
Krok 4.

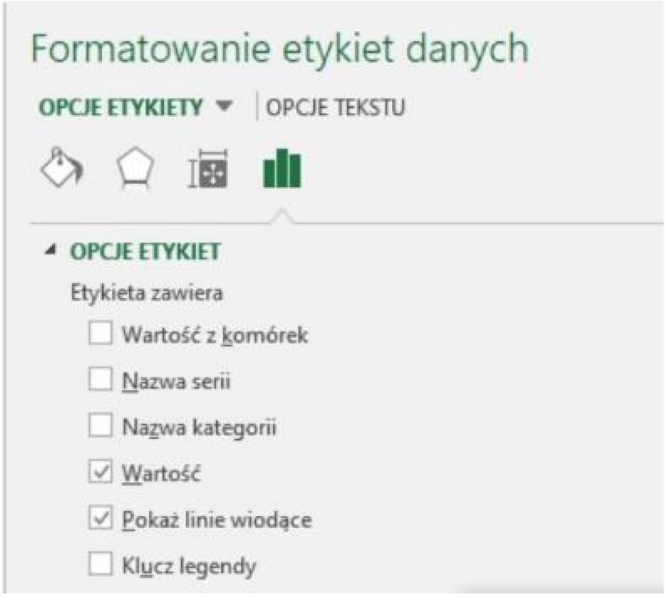
W celu dodania tytułów osi klikamy na znak plusa przy wykresie i zaznaczamy *tytuły osi*.



Krok 5.

Jeżeli chcemy, aby na wykresie były widoczne dokładne dane (zysk z danego dnia), klikamy prawym przyciskiem myszy na dany „słupek”, a następnie dodajemy etykiety danych.





Krok 6.

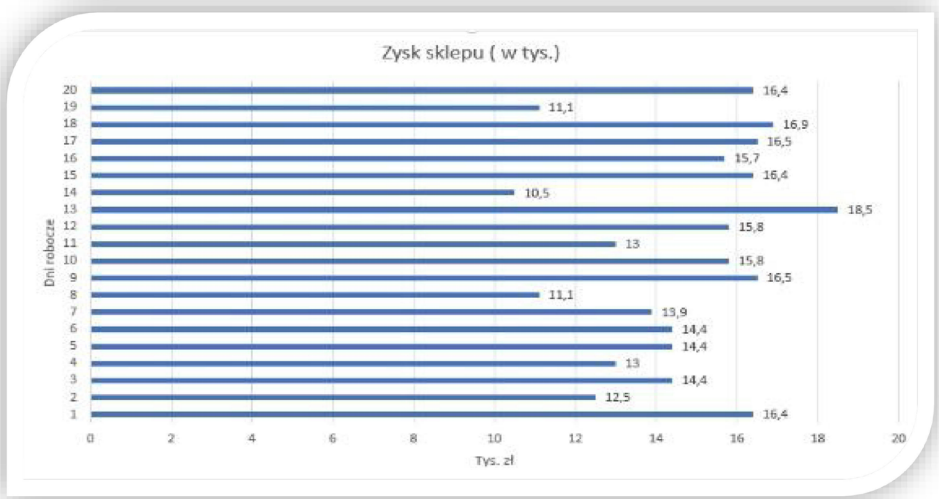
Finalnie reprezentacja graficzna danych powinna wyglądać podobnie jak na rysunku poniżej.



Krok 7.

Klikając prawym przyciskiem myszy w otrzymany wykres, wybieramy opcję *Zmień typ wykresu…* Teraz będziemy mogli przećwiczyć różne typy wykresu dla danych z zadania. Wybieramy kolejno: *wykres liniowy, słupkowy*





Przypadek 2

Polecenie:

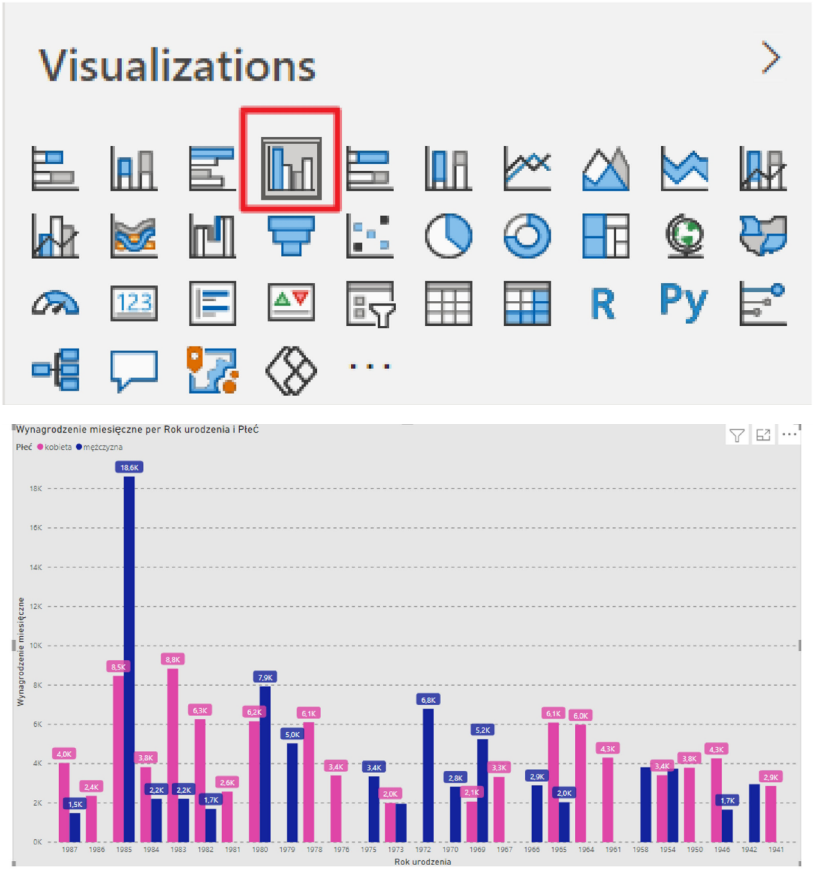
Dla danych z pliku Załącznik1.xlsx dokonaj prezentacji graficznej danych z wykorzystaniem programu Power BI.

	A	B	C	D	E	F	G	H
1	lp	płeć	rok urodzenia	wykształcenie	ulubiony typ filmu	wynagrodzenie miesięczne	liczba dzieci	czas dojazdu do pracy [min]
2	1	kobieta	1978	podstawowe	horror	3169	0	23
3	2	mężczyzna	1979	zawodowe	przyrodniczy	3061	1	45
4	3	kobieta	1965	średnie	kryminał	2748	2	65
5	4	kobieta	1964	licencjat	film akcji	3763	1	92
6	5	mężczyzna	1980	inżynierskie	dramat	1585	2	31
7	6	kobieta	1964	magisterskie	komedia	2219	0	15
8	7	mężczyzna	1985	podstawowe	romans	2874	1	63
9	8	kobieta	1986	zawodowe	melodramat	2359	2	14
10	9	kobieta	1987	licencjat	horror	1007	2	88
11	10	mężczyzna	1985	licencjat	kryminał	2919	1	66
12	11	mężczyzna	1982	inżynierskie	przyrodniczy	1699	2	34

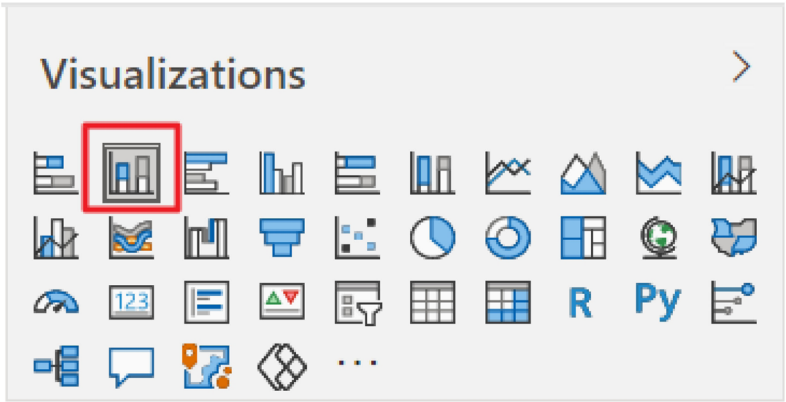
Rozwiązanie:

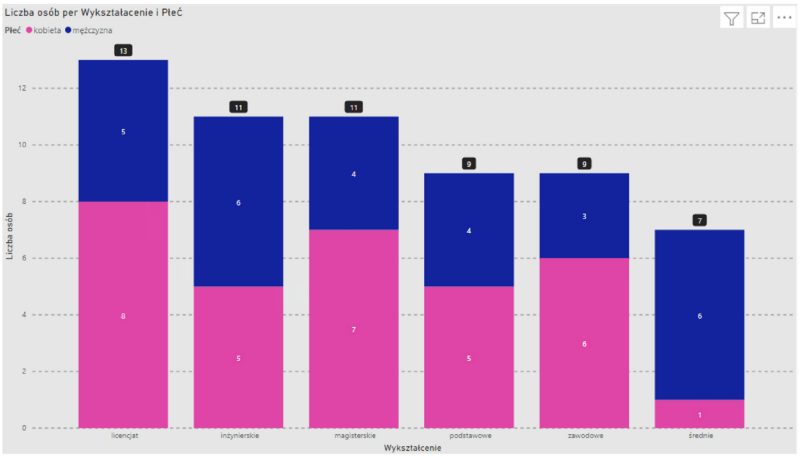
Wybierając wskazane poniżej metody prezentacji danych, możemy dokonać przykładowych wizualizacji danych.

Krok 1.

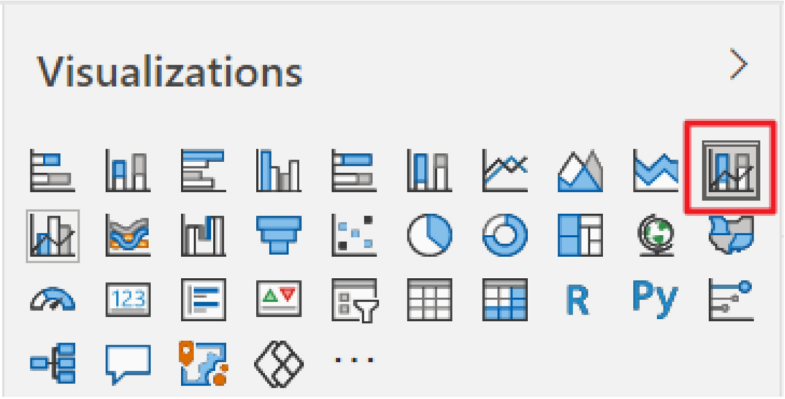

Visualizations
R
Py
Wynagrodzenie miesięczne per Rok urodzenia i Płeć
Płeć kobieta mężczyzna
Wynagrodzenie miesięczne
Rok urodzenia

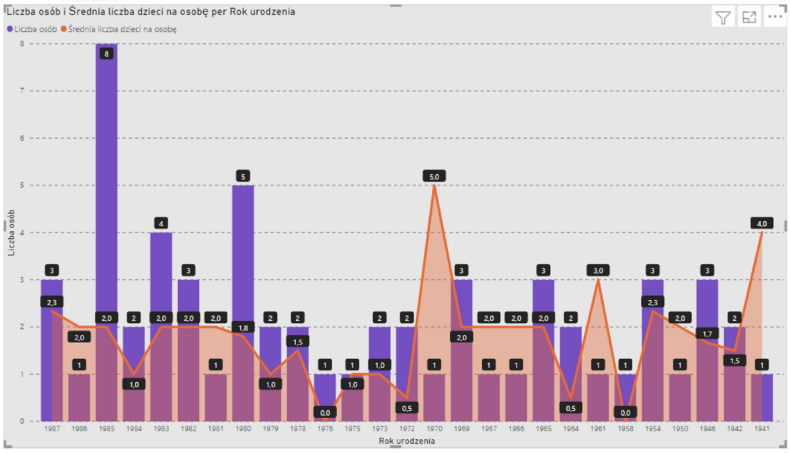
Krok 2.


Visualizations
R
Py

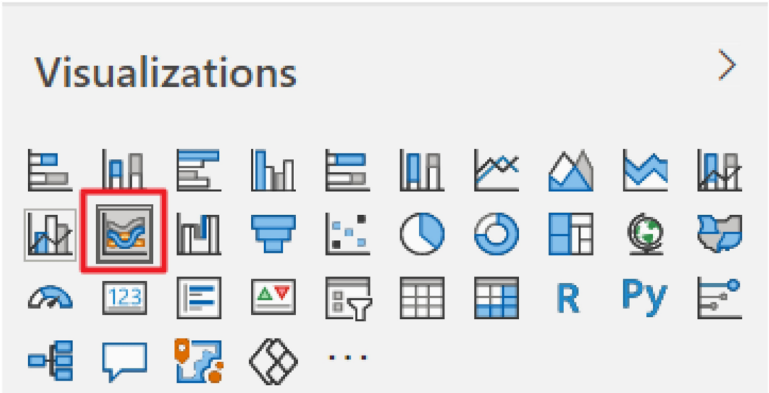

Liczba osób per Wykształcenie i Płeć
Płeć kobieta mężczyzna
Liczba osób
licencjat
inżynierskie
magisterskie
podstawowe
zawodowe
średnie
Wykształcenie

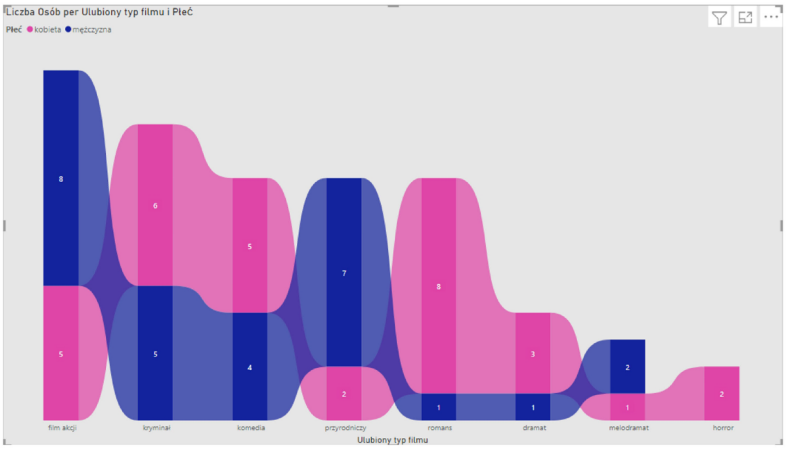
Krok 3.


Visualizations
R
Py

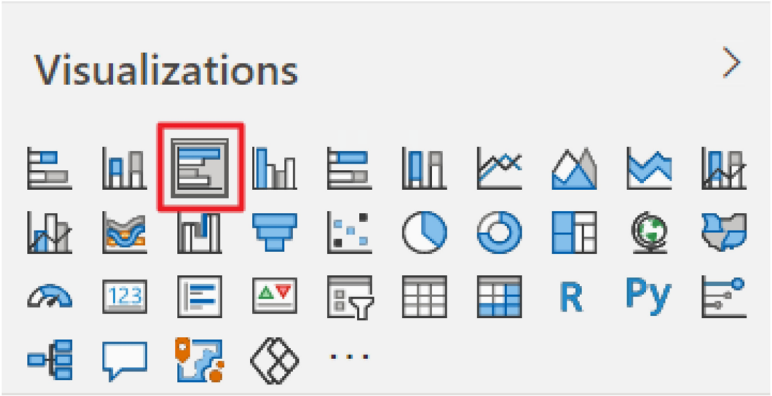

Liczba osób i Średnia liczba dzieci na osobę per Rok urodzenia
Liczba osób Średnia liczba dzieci na osobę
Liczba osób
Rok urodzenia

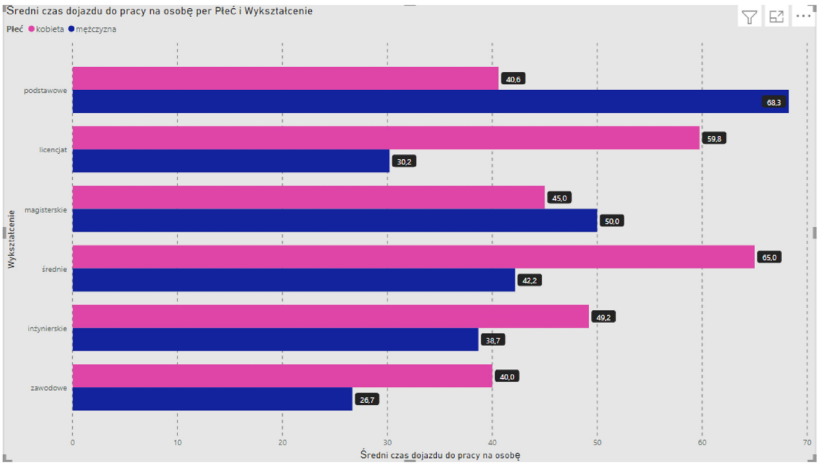
Krok 4.


Visualizations
R
Py


Liczba Osób per Ulubiony typ filmu i Płeć
Płeć kobieta mężczyzna
film akcji
kryminał
komedia
przyrodniczy
romans
dramat
melodramat
horror
Ulubiony typ filmu

Krok 5.


Visualizations
R
Py



Przypadek 3

Polecenie:

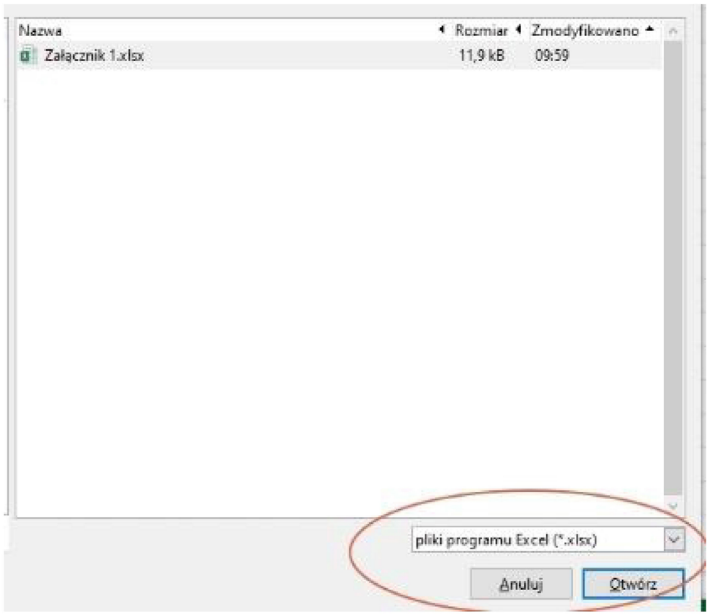
Dla danych z pliku Załącznik1.xlsx dokonaj prezentacji graficznej danych z wykorzystaniem programu GRETL.

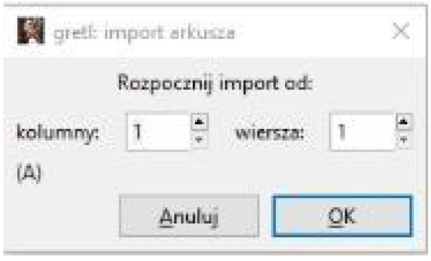
	A	B	C	D	E	F	G	H
1	lp	płeć	rok urodzenia	wykształcenie	ulubiony typ filmu	wynagrodzenie miesięczne	liczba dzieci	czas dojazdu do pracy [min]
2	1	kobieta	1978	podstawowe	horror	3169	0	23
3	2	mężczyzna	1979	zawodowe	przyrodniczy	3061	1	45
4	3	kobieta	1965	średnie	kryminał	2748	2	65
5	4	kobieta	1964	licencjat	film akcji	3763	1	92
6	5	mężczyzna	1980	inżynierskie	dramat	1585	2	31
7	6	kobieta	1964	magisterskie	komedia	2219	0	15
8	7	mężczyzna	1985	podstawowe	romans	2874	1	63
9	8	kobieta	1986	zawodowe	melodramat	2359	2	14
10	9	kobieta	1987	licencjat	horror	1007	2	88
11	10	mężczyzna	1985	licencjat	kryminał	2919	1	66
12	11	mężczyzna	1982	inżynierskie	przyrodniczy	1699	2	34

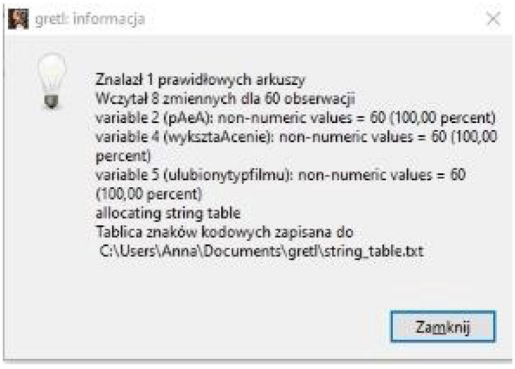
Rozwiązanie:

Krok 1.

Wczytujemy plik z danymi, wybierając *Plik/Otwórz dane/Pliki użytkowników*.

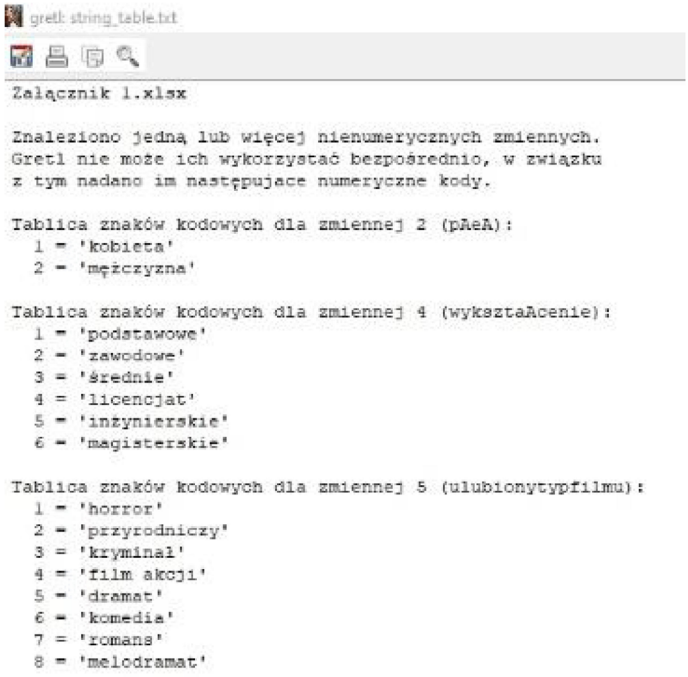

Nazwa
Rozmiar
Zmodyfikowano
Załącznik 1.xlsx
11,9 kB
09:59
pliki programu Excel (*.xlsx)
Anuluj
Otwórz


gretl: import arkusza
Rozpocznij import od:
kolumny:
1
wiersza:
1
(A)
Anuluj
OK


gretl: informacja
Znalazł 1 prawidłowych arkuszy
Wczytał 8 zmiennych dla 60 obserwacji
variable 2 (pAeA): non-numeric values = 60 (100,00 percent)
variable 4 (wyksztaAcenie): non-numeric values = 60 (100,00 percent)
variable 5 (ulubionytypfilmu): non-numeric values = 60 (100,00 percent)
allocating string table
Tablica znaków kodowych zapisana do
C:\Users\Anna\Documents\gretl\string_table.txt
Zamknij

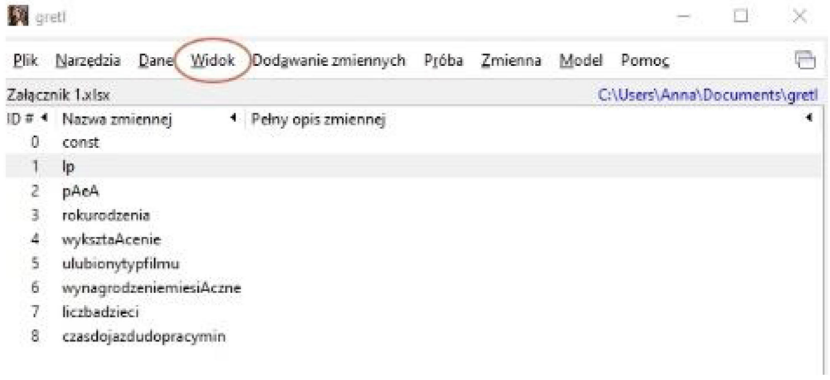
Krok 2.

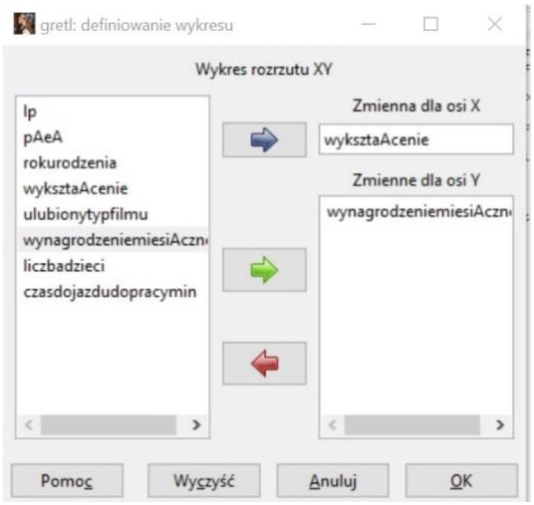
Po wczytaniu danych pojawia się poniższe okno informacyjne dotyczące kodowania zmiennych.

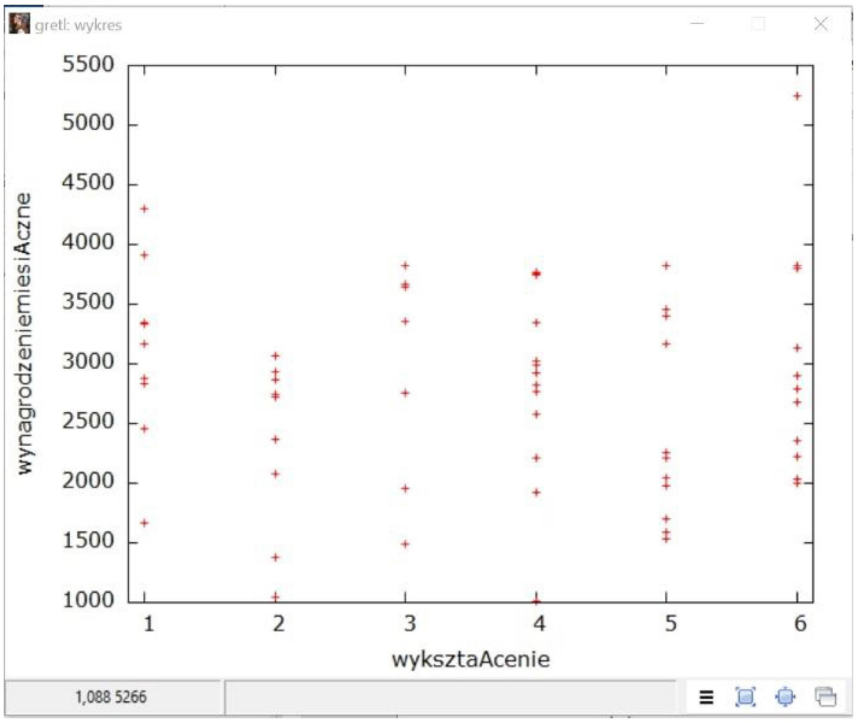


Krok 3.

Wybieramy *Widok/Wykres zmiennych/Wykres rozrzutu X-Y*.

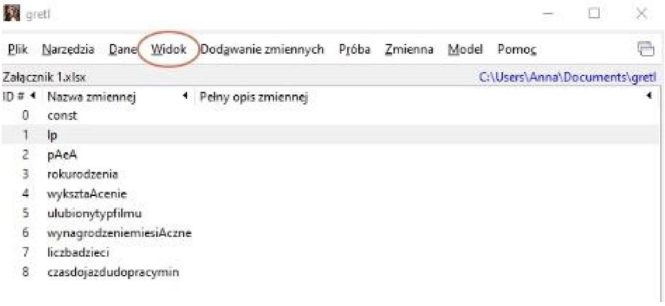


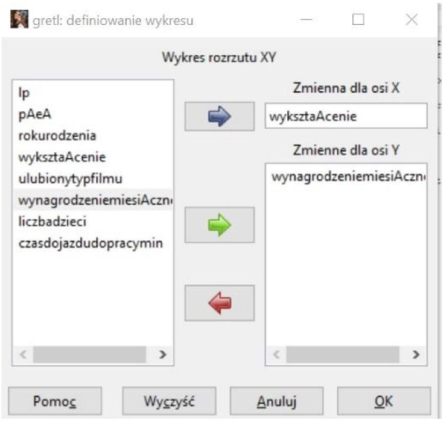

gretl: definiowanie wykresu
Wykres rozrzutu XY
lp
pAeA
rokurodzenia
wyksztaAcenie
ulubionytypfilmu
wynagrodzeniemiesiAczn
liczbadzieci
czasdojazdudopracymin
Zmienna dla osi X
wyksztaAcenie
Zmienne dla osi Y
wynagrodzeniemiesiAczn
Pomoc
Wyczyść
Anuluj
OK

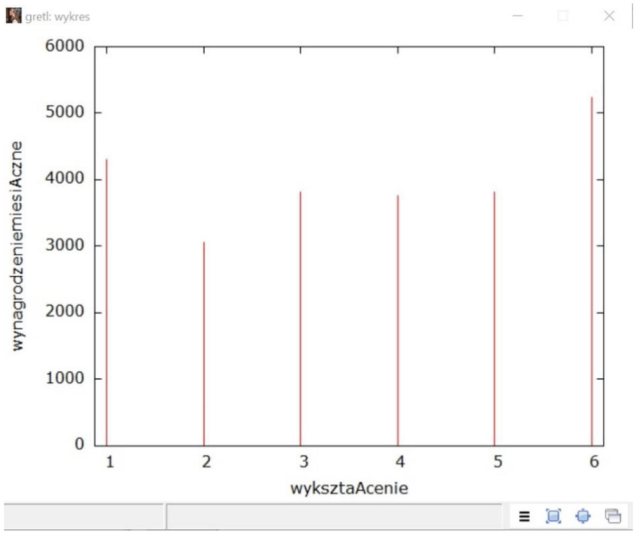

gretl: wykres
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
wynagrodzeniemiesiAczne
1
2
3
4
5
6
wyksztaAcenie
1,088 5266

Krok 4.

Wybieramy *Widok/Wykres zmiennych/Wykres rozrzutu X-Y z impulsami*.

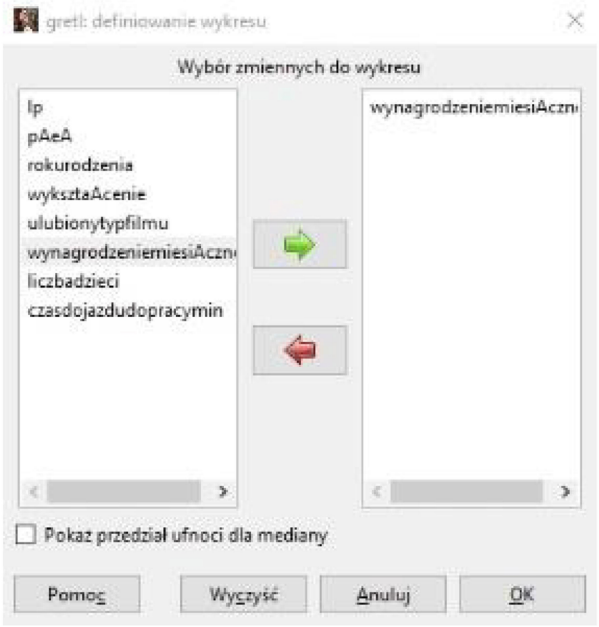


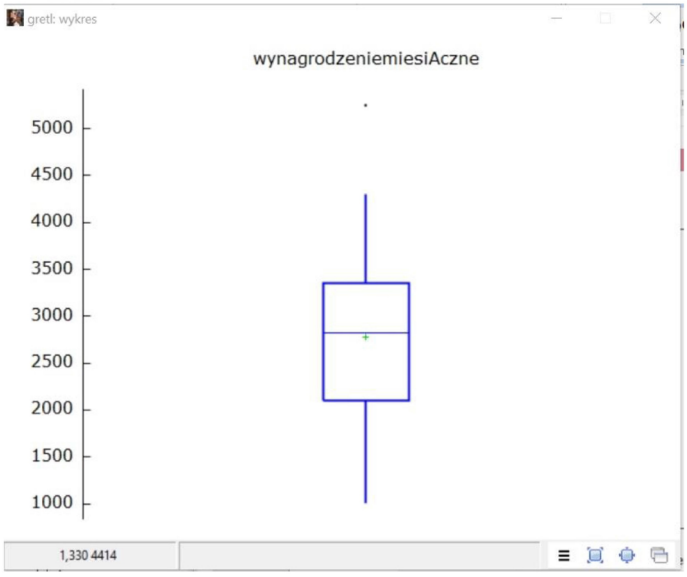




Krok 5.

Wybieramy *Widok/Wykres zmiennych/Wykres pudełkowy*.





Przypadek 4

Polecenie:

Dla danych z pliku Załącznik1.xlsx dokonaj prezentacji graficznej danych z wykorzystaniem programu EXCEL.

	A	B	C	D	E	F	G	H
1	lp	płeć	rok urodzenia	wykształcenie	ulubiony typ filmu	wynagrodzenie miesięczne	liczba dzieci	czas dojazdu do pracy [min]
2	1	kobieta	1978	podstawowe	horror	3169	0	23
3	2	mężczyzna	1979	zawodowe	przyrodniczy	3061	1	45
4	3	kobieta	1965	średnie	kryminał	2748	2	65
5	4	kobieta	1964	licencjat	film akcji	3763	1	92
6	5	mężczyzna	1980	inżynierskie	dramat	1585	2	31
7	6	kobieta	1964	magisterskie	komedia	2219	0	15
8	7	mężczyzna	1985	podstawowe	romans	2874	1	63
9	8	kobieta	1986	zawodowe	melodramat	2359	2	14
10	9	kobieta	1987	licencjat	horror	1007	2	88
11	10	mężczyzna	1985	licencjat	kryminał	2919	1	66
12	11	mężczyzna	1982	inżynierskie	przyrodniczy	1699	2	34

Rozwiązanie:

Krok 1.

Rozpoczniemy od zliczenia liczby osób w podziale na płeć.
W komórce K3 piszemy =LICZ.JEŻELI(B2:B61;”kobieta”).
W komórce K4 wpisujemy =LICZ.JEŻELI(B2:B62;”mężczyzna”).

Ostatecznie mamy:

kobiety	32
mężczyźni	28

Krok 2.

Wykonamy podobne obliczenia jak w Kroku 1 w zakresie wykształcenia oraz płci. Konstruujemy w tym celu tabelę pomocniczą.

W komórce K8 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$8;$D2:$D61;K$7)
W komórce L8 wpisujemy
= LICZ.WARUNKI($B$2:$B$61;$J$8;$D2:$D61;L$7)
W komórce M8 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$8;$D2:$D61;M$7)
W komórce N8 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$8;$D2:$D61;N$7)
W komórce O8 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$8;$D2:$D61;O$7)
W komórce P8 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$8;$D2:$D61;P$7)
W komórce K9 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$9;$D2:$D61;K$7)
W komórce L9 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$9;$D2:$D61;L$7)
W komórce M9 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$9;$D2:$D61;M$7)
W komórce N9 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$9;$D2:$D61;N$7)
W komórce O9 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$9;$D2:$D61;O$7)
W komórce P9 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$9;$D2:$D61;P$7)

	podstawowe	zawodowe	średnie	licencjat	inżynierskie	magisterskie
kobieta	5	6	1	8	5	7
mężczyzna	4	3	6	5	6	4

Krok 3.

Wykonamy podobne obliczenia jak w Kroku 1 w zakresie ulubionego typu filmu oraz płci. Konstruujemy w tym celu tabelę pomocniczą.

	horror	przyrodniczy	kryminał	film akcji	dramat	komedia	romans
kobieta	2	2	6	5	3	5	8
mężczyzna	0	7	5	8	1	4	1

W komórce K13 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$13;$E2:$E61;K$12)
W komórce L13 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$13;$E2:$E61;L$12)
W komórce M13 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$13;$E2:$E61;M$12)
W komórce N13 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$13;$E2:$E61;N$12)
W komórce O13 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$13;$E2:$E61;O$12)
W komórce P13 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$13;$E2:$E61;P$12)
W komórce Q13 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$13;$E2:$E61;Q$12)
W komórce K14 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$14;$E2:$E61;K$12)
W komórce L14 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$14;$E2:$E61;L$12)
W komórce M14 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$14;$E2:$E61;M$12)
W komórce N14 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$14;$E2:$E61;N$12)
W komórce O14 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$14;$E2:$E61;O$12)
W komórce P14 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$14;$E2:$E61;P$12)
W komórce Q14 wpisujemy =LICZ.WARUNKI($B$2:$B$61;$J$14;$E2:$E61;Q$12)

Krok 4.

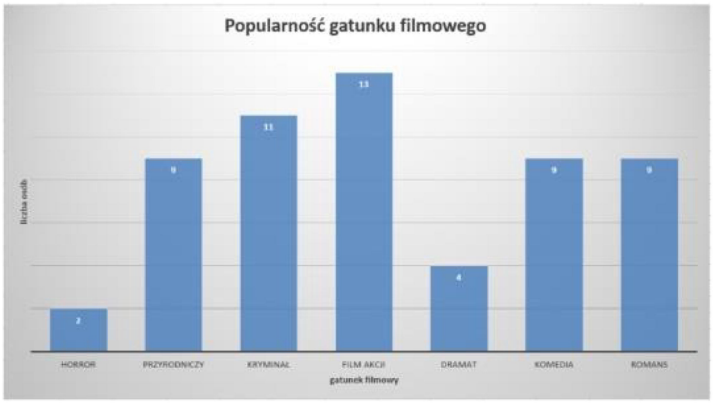
Podsumowując (obliczając sumę) tabele uzyskane w Kroku 2 oraz Kroku 3, mamy teraz informacje w poniższym zakresie.

horror	przyrodniczy	kryminał	film akcji	dramat	komedia	romans
2	9	11	13	4	9	9

podstawowe	zawodowe	średnie	licencjat	inżynierskie	magisterskie
9	9	7	13	11	11

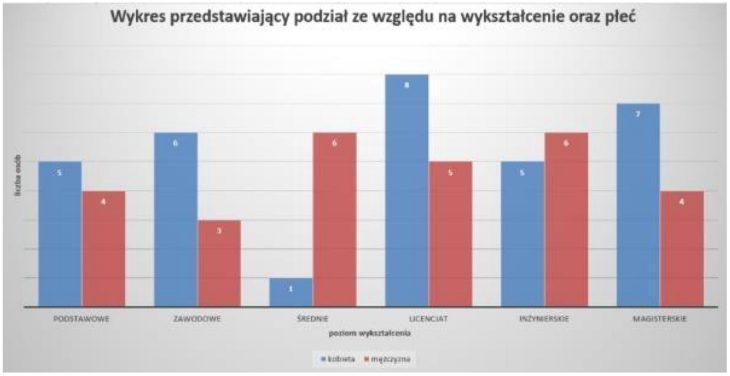
Krok 5.

Zaznaczamy dane dotyczące gatunku filmu z tabeli uzyskanej w Kroku 4. Wybieramy *wykres kolumnowy.*



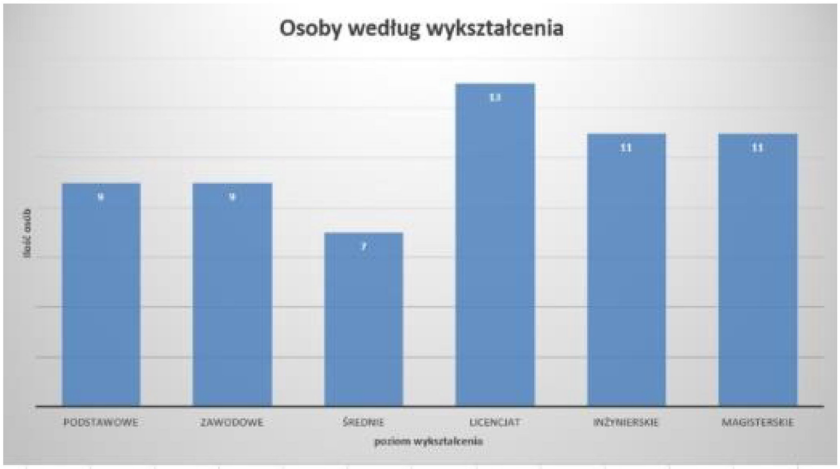
Krok 6.

Korzystając z tabeli uzyskanej w Kroku 2, otrzymujemy poniższy wykres.



Krok 7.

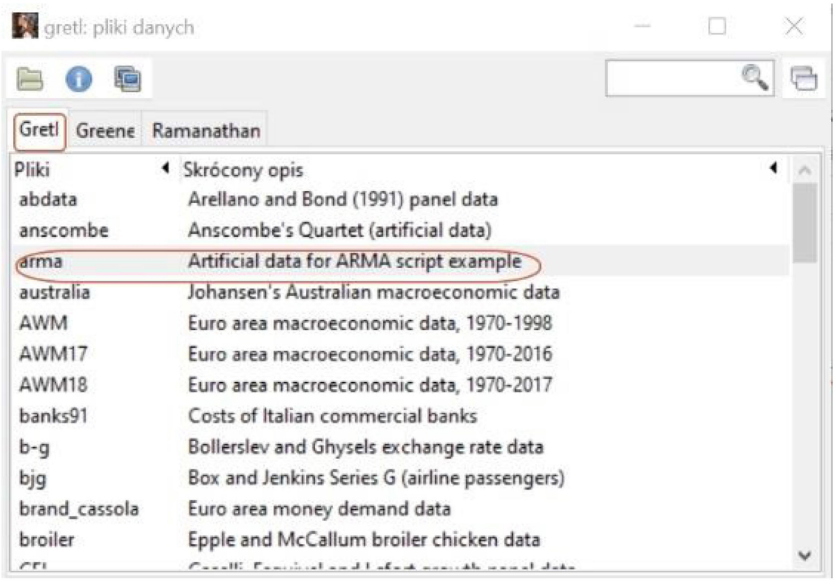
Zaznaczamy dane dotyczące wykształcenia z tabeli uzyskanej w Kroku 4. Wybieramy *wykres kolumnowy.*



Zadania

Zadanie 1.

Otwórz plik z przykładami programu Gretl, plik arma. Następnie dokonaj ilustracji graficznej danych.



.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Zadanie 2.

Dla danych z tabeli dokonaj wizualizacji danych z wykorzystaniem oprogramowania EXCEL, STATISTICA oraz POWER BI.

Województwo	Produkcja sprzedana przemysłu	Inwest.	Wartość Brutto Środków trwałych	Płace	Stopa Bezrobocia Rejestrow. %	Podmioty gospodarki
	Na mieszkańca W tys. zł		W tys. zł			
Dolnośląskie	35,22	4,63	140 713	3,83	10,67	30,27
Kujawsko-pomorskie	23,66	2,96	66 028	2,25	15,04	17,60
Lubelskie	13,99	1,94	47 076	3,40	12,24	15,24
Lubuskie	26,26	2,84	37 636	3,26	12,93	10,08
Łódzkie	26,20	3,52	95 203	3,43	11,16	21,39
Małopolskie	22,78	2,76	107 796	3,55	9,09	32,03
Mazowieckie	43,45	6,01	395 519	4,72	8,81	61,95
Opolskie	22,31	3,72	37 223	3,46	11,64	7,09
Podkarpackie	17,33	2,45	56 803	3,21	14,14	14,40
Podlaskie	17,01	2,10	29 497	3,36	12,37	8,86
Pomorskie	33,63	3,90	88 496	3,79	10,42	25,85
Śląskie	42,67	3,95	220 977	3,89	8,84	37,42
Świętokrzyskie	18,64	2,25	34 817	3,29	13,8	9,22
Warmińsko-mazurskie	18,57	2,17	34 327	3,20	18,14	11,56
Wielkopolskie	37,63	3,60	147 190	3,45	7,56	34,73
Zachodnio-pomorskie	19,54	2,82	58 055	3,48	15	19,03

.....
.....
.....
.....
.....
.....
.....
.....
.....

Studium przypadku 2.
Modele ekonometryczne

Wstęp teoretyczny

Do oszacowania parametrów liniowego modelu ekonometrycznego:

$$Y = \alpha_0 + \alpha_1 \cdot t + \varepsilon_t$$

wykorzystuje się następujący wzór:

$$a = (T^T T)^{-1} T^T y$$

Zależność między funkcją trendu z okresami sumującymi się do 0 i numerowanymi tradycyjnie (1, 2, ..., *n*) przedstawia się następująco:

$$\hat{y}_t = a_0 + a_1 \cdot t \quad (\sum t = 0)$$
$$\hat{y}_t = a_0 + a_1 \cdot (t - \bar{t}) \quad (\sum t \neq 0)$$

Prognoza punktowa na podstawie modelu wahań w czasie

Model addytywny ma postać:

$$y^*_{T(I,\ldots,d)} = (a_0 + a_1 \cdot T) + g_{T(I,\ldots,d)}$$

Model multiplikatywny ma postać:

$$y^*_{T(I,\ldots,d)} = (a_0 + a_1 \cdot T) \cdot s_{T(I,\ldots,d)}$$

Bezwzględny błąd prognozy punktowej w modelu addytywnym *ex ante*:

$$D_{T(I,\ldots,d)} = S_z \sqrt{1 + \frac{1}{n} + \frac{(T - \bar{t})^2}{\sum_{t=1}^{n} (t - \bar{t})^2}}$$

Prognoza przedziałowa na podstawie modelu wahań w czasie

Model addytywny:

W przypadku gdy odchylenia losowe modelu mają rozkład normalny, prognozy wyznacza się w następujący sposób:

$$P\{y_T^* - t_\alpha \cdot D_{T(I,\ldots,d)} < Y_{T(I,\ldots,d)} < y_T^* + t_\alpha \cdot D_{T(1,\ldots,d)}\} = 1 - \alpha$$

gdy jest mała próba ($n \leq 30$) t_α odczytuje się z tablic t-Studenta (patrz Załącznik 5) dla
$n - d -$ 1 stopni swobody

$$P\{y_T^* - u_\alpha \cdot D_{T(I,\ldots,d)} < Y_{T(I,\ldots,d)} < y_T^* + u_\alpha \cdot D_{T(I,\ldots,d)}\} = 1 - \alpha$$

gdy jest duża próba ($n > 30$), u_α odczytuje się z tablic dystrybuanty rozkładu normalnego (patrz Załącznik 4).

W przypadku, gdy odchylenia losowe nie mają rozkładu normalnego (nie weryfikowano hipotezy lub została odrzucona) prognozy wyznacza się w następujący sposób:

$$P\{y_T^* - u \cdot D_{T(I,\ldots,d)} < Y_{T(I,\ldots,d)} < y_T^* + u \cdot D_{T(I,\ldots,d)}\} = 1 - \alpha$$

u wyznacza się z nierówności Czebyszewa:

$$u = \sqrt{\frac{1}{1 - p}}$$

Bezwzględny błąd prognozy przedziałowej *ex ante*:

$$D'_{T(I,\ldots,d)} = t_\alpha \cdot D_{T(I,\ldots,d)}$$
$$D'_{T(I,\ldots,d)} = u_\alpha \cdot D_{T(I,\ldots,d)}$$
$$D'_{T(I,\ldots,d)} = u \cdot D_{T(I,\ldots,d)}$$

Względny błąd prognozy przedziałowej *ex ante*:

$$V_T' = \frac{D_T'}{y_T^*} \cdot 100\%$$

WAŻNE

Wzór
Współczynnika determinacji:

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

gdzie:
n – liczba prognoz
$\hat{y}_t$ – przewidywana wartość zmiennej
y_t – rzeczywista wartość zmiennej
$\bar{y}$ – średnia rzeczywista wartość zmiennej

Współczynnik determinacji informuje nas, na ile nasz model, nasz badany czynnik wyjaśnia zgromadzone dane pomiarowe (zmienną zależną). Współczynnik determinacji przyjmuje wartości pomiędzy 0 a 1, jednakże najbardziej popularną formą prezentacji tego współczynnika jest jego procentowa postać.

Przypadek 1

Polecenie:

Sprzedaż przedsiębiorstwa *X* w tys. w kolejnych kwartałach roku 2019 oraz 2020 kształtowała się następująco.

Lata	2019				2020			
Kwartały	I	II	III	IV	I	II	III	IV
Sprzedaż (tys. zł)	30	28	25	23	21	20	18	12

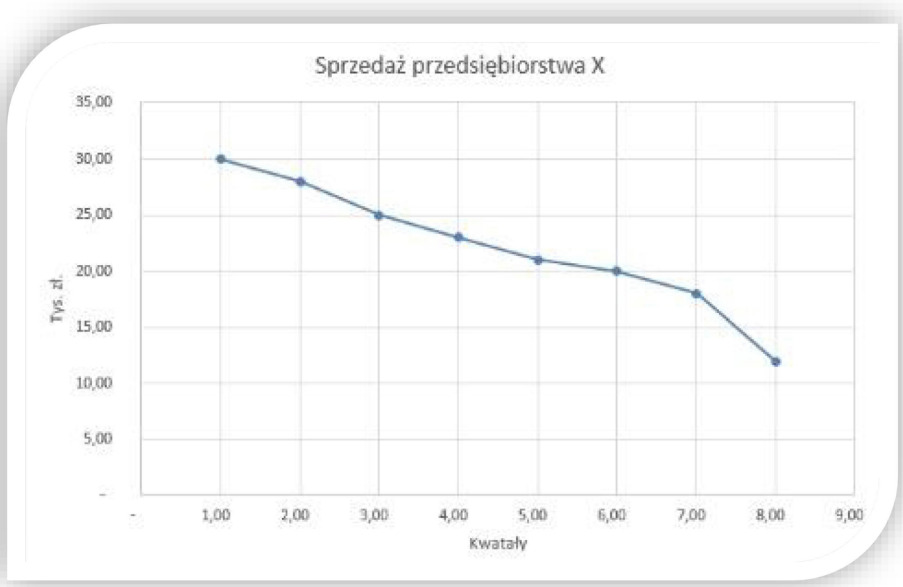
Przyjmując wiarygodność prognozy równą 0,95 (i zakładając rozkład normalny odchyleń losowych modelu), wyznacz prognozę punktową i przedziałową sprzedaży w dwóch pierwszych kwartałach roku 2001 r. Zadanie rozwiąż, stosując wzory opisane we Wstępie teoretycznym oraz program EXCEL.

Rozwiązanie:

Krok 1.

Wprowadzamy dane do arkusza.

	A	B	C
1			
2	kwartały	t	Y
3	1	1,00	30,00
4	2	2,00	28,00
5	3	3,00	25,00
6	4	4,00	23,00
7	1	5,00	21,00
8	2	6,00	20,00
9	3	7,00	18,00
10	4	8,00	12,00



W kolejnych krokach oszacujemy parametry modelu i mierniki dopasowania modelu do danych.

Krok 2.

Do oszacowania modelu definiujemy macierz X oraz macierz transponowaną do X.

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}.$$

15									
16				1	1,00				
17				1	2,00				
18				1	3,00				
19				1	4,00				
20			X	1	5,00				
21				1	6,00				
22				1	7,00				
23				1	8,00				
24									
25	XT								
26		1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
27		1,00	2,00	3,00	4,00	5,00	6,00	7,00	8,00

Krok 3.

Stosując funkcje MACIERZ.ILOCZYN, obliczamy iloczyn macierzy z Kroku 2.

28			
29	XT*X		
30		8,00	36,00
31		36,00	204,00

Krok 4.

Wpisując w komórce B34 wartość =WYZNACZNIK.MACIERZY(B30:C31), otrzymujemy wyznacznik macierzy

32		
33	det(XT*X)	
34		336
35		

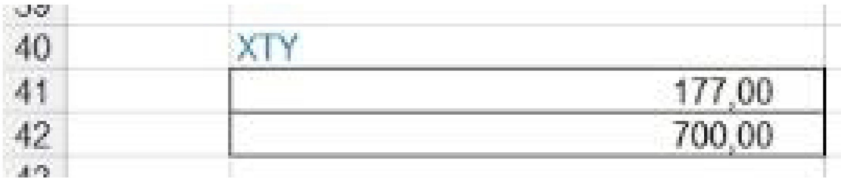
Krok 5.

Korzystając z funkcji MACIERZ.ODW, obliczamy

35			
36	XT*X^(-1)		
37		0,61	-0,11
38		-0,11	0,02

Krok 6.

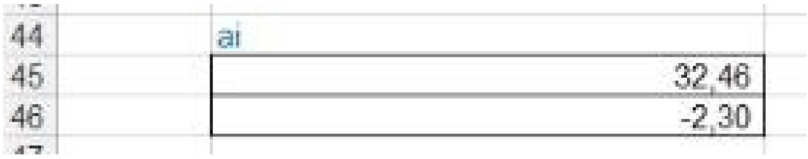
Mamy $Y = \begin{bmatrix} 30 \\ 28 \\ 25 \\ 23 \\ 21 \\ 20 \\ 18 \\ 12 \end{bmatrix}$, korzystając z funkcji MACIERZ.ILOCZYN



40	XTY	
41		177,00
42		700,00

Krok 7.

Korzystając z funkcji MACIERZ.ILOCZYN, obliczamy współczynniki.



44	ai	
45		32,46
46		-2,30

Interpretacja:

32,46 – teoretyczny poziom zjawiska (sprzedaż) w roku zerowym (I kwartał 2019) – 30,46 tys. zł

-2,30 – informuje o tym, że sprzedaż z okresu na okres w analizowanych kwartałach spadła o 2,30.

Model oszacowany ma postać:

$$\hat{y}_t = 32{,}46 - 2{,}3 \cdot t$$

Krok 8.

Dokonując prognozy na kolejne dwa kwartały, wpisujemy kolejno:

W komórce C68 wartość =B46+B47*B68+D57
W komórce C69 wartość = B46+B47*B69+E57
W komórce D68 wartość = $C$63*((1+1/8+(((B68-$B$13)^2) /$F$12)))^0,5
W komórce D69 wartość = $C$63*((1+1/8+(((B69-$B$13)^2) /$F$12)))^0,5
W komórce E68 wartość = D68/C68
W komórce E69 wartość = D69/C69
W komórce F68 wartość = C68-D68*$C$64
W komórce F69 wartość = C69-D69*$C$64
W komórce G68 wartość = C68+D68*$C$64
W komórce G69 wartość = C69+D69*$C$64

65							
66			Prognoza	Błąd bezwzgledny	Błąd względny	Prognoza przedziałowa	
67	Kwartały	Okres prognozy T	Y*	DT	VT	Dolna	Górna
68	1	9	11,71	2,95	0,25	4,41	19,02
69	2	10	10,21	3,11	0,30	2,51	17,92
70							

Przypadek 2

Polecenie:

Sprzedaż przedsiębiorstwa *X* w tys. w kolejnych kwartałach roku 2019 oraz 2020 kształtowała się następująco.

Lata	2019				2020			
Kwartały	I	II	III	IV	I	II	III	IV
Sprzedaż (tys. zł)	30	28	25	23	21	20	18	12

Przyjmując wiarygodność prognozy równą 0,95 (i zakładając rozkład normalny odchyleń losowych modelu), wyznacz prognozę punktową i przedziałową sprzedaży w dwóch pierwszych kwartałach roku 2001 r. Zadanie rozwiąż, stosując moduł Regresja programu EXCEL.

Rozwiązanie:

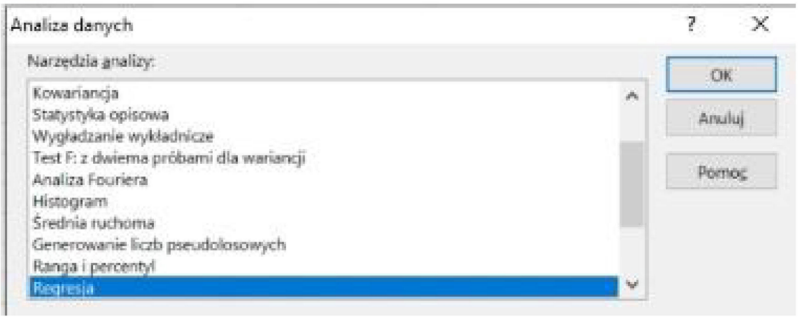
Krok 1.

Wczytaj dane do nowego arkusza EXCEL.

	A	B	C
1	kwartały	t	Sprzedaż
2	1	1,00	30,00
3	2	2,00	28,00
4	3	3,00	25,00
5	4	4,00	23,00
6	1	5,00	21,00
7	2	6,00	20,00
8	3	7,00	18,00
9	4	8,00	12,00

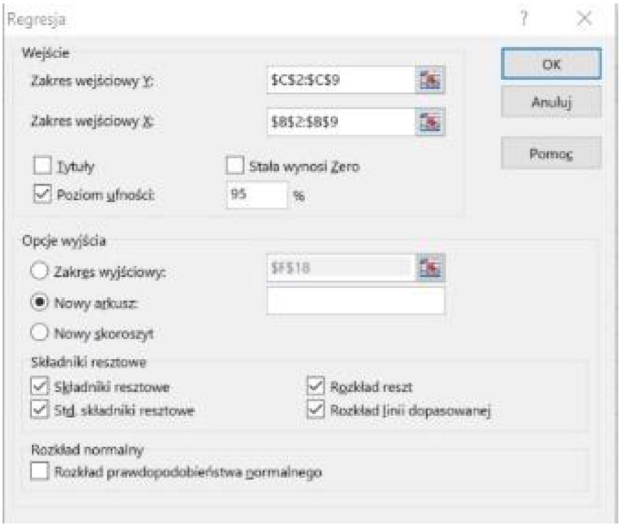
Krok 2.

Wybierz z menu *Dane/Analiza Danych/Regresja*.



Krok 3.

W oknie *Regresja* wskaż zakresy wejściowe *Y* oraz *X*. Zaznacz pola wyboru, tak jak pokazano na rysunku poniżej. Program wyświetla wyniki w nowym arkuszu.



Krok 4.

Otrzymaliśmy następujące wyniki w zakresie statystyki regresji:

- Współczynnik korelacji — *Wielokrotność R = 0,98*
- Współczynnik determinacji — *R kwadrat = 0,96*
- Współczynnik determinacji dla małej liczby danych — *Dopasowany R kwadrat = 0,95*
- Błąd standardowy = 1,24

2		
3	*Statystyki regresji*	
4	Wielokrotność R	0,98
5	R kwadrat	0,96
6	Dopasowany R kwadrat	0,95
7	Błąd standardowy	1,24
8	Obserwacje	8
9		

Krok 5.

Z analizy wariancji wynika, że liniowość równania regresji jest istotna.

10	ANALIZA WARIANCJI					
11		*df*	*SS*	*MS*	*F*	*Istotność F*
12	Regresja	1	221,72	221,72	145,31	0,00
13	Resztkowy	6	9,15	1,53		
14	Razem	7	230,88			

Krok 6.

Równanie regresji ma postać .

15									
16		*Współczynniki*	*Błąd standardowy*	*t Stat*	*Wartość-p*	*Dolne 95%*	*Górne 95%*	*Dolne 95,0%*	*Górne 95,0%*
17	Przecięcie	32,46	0,96	33,73	0,00	30,11	34,82	30,11	34,82
18	Zmienna X 1	-2,30	0,19	-12,05	0,00	-2,76	-1,83	-2,76	-1,83
19									

Krok 7.

Wartość-p dla testów t jest mniejsza niż 0,05, a więc parametry i są istotne.

15									
16		*Współczynniki*	*Błąd standardowy*	*t Stat*	*Wartość-p*	*Dolne 95%*	*Górne 95%*	*Dolne 95,0%*	*Górne 95,0%*
17	Przecięcie	32,46	0,96	33,73	0,00	30,11	34,82	30,11	34,82
18	Zmienna X 1	-2,30	0,19	-12,05	0,00	-2,76	-1,83	-2,76	-1,83
19									

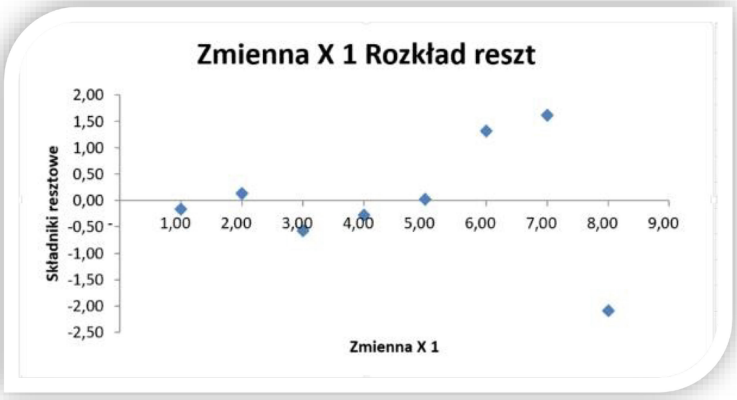
Krok 8

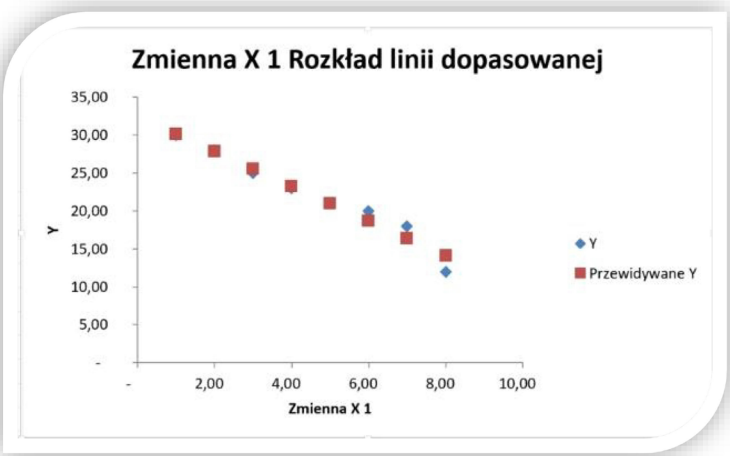
Program wyświetla składniki resztowe, czyli reszty oraz reszty standaryzowane.

23				
24	*Obserwacja*	*Przewidywane Y*	*Składniki resztowe*	*Std. składniki resztowe*
25	1	30,17	-0,17	-0,15
26	2	27,87	0,13	0,11
27	3	25,57	-0,57	-0,50
28	4	23,27	-0,27	-0,24
29	5	20,98	0,02	0,02
30	6	18,68	1,32	1,16
31	7	16,38	1,62	1,42
32	8	14,08	-2,08	-1,82
33				

Krok 9.

Program tworzy wykres rozkładu reszt oraz rozkład linii dopasowanej. Z wykresu rozrzutu reszt wynika, że reszty mają rozkład losowy i brak jest autokorelacji reszt.





Interpretacja:

Można przyjąć, że model dobrze opisuje badane zjawisko. Błąd standardowy estymacji wynosi 1,24. Oznacza to, że przewidywane wartości sprzedaży różnią się od wartości empirycznych średnio o 0,012 (‰).

Z analizy wariancji wynika, że liniowość równania regresji jest istotna. Testy t wskazują, że obliczone parametry są istotne. Z wykresu rozrzutu reszt wynika, że spełnione są założenia dotyczące losowości odchyleń i braku autokorelacji. Wartość oczekiwana reszt jest równa 0.

Krok 10.

Wykonamy teraz prognozy punktowe. Korzystając z otrzymanego modelu, w komórce C10 wpisujemy wartość =32,46-2,3*B10, a w komórce C11 wartość =32,46-2,3*B11. Przewidywana wielkość sprzedaży wynosi w kolejnych kwartałach 1,76 oraz 9,46.

	A	B	C
1	kwartały	t	Sprzedaż
2	1	1,00	30,00
3	2	2,00	28,00
4	3	3,00	25,00
5	4	4,00	23,00
6	1	5,00	21,00
7	2	6,00	20,00
8	3	7,00	18,00
9	4	8,00	12,00
10		9,00	11,76
11		10,00	9,46

Krok 11.

Zajmiemy się teraz prognozą przedziałową. Korzystając z poniższych wyników:

15									
16		*Współczynniki*	*Błąd standardowy*	*t Stat*	*Wartość-p*	*Dolne 95%*	*Górne 95%*	*Dolne 95,0%*	*Górne 95,0%*
17	Przecięcie	32,46	0,96	33,73	0,00	30,11	34,82	30,11	34,82
18	Zmienna X 1	-2,30	0,19	-12,05	0,00	-2,76	-1,83	-2,76	-1,83
19									

W komórce D10 wpisujemy wartość =30,11-2,76*B10
W komórce D11 wpisujemy wartość =30,11-2,76*B11
W komórce E10 wpisujemy wartość =34,82-1,83*B10
W komórce E11 wpisujemy wartość =34,82-1,83*B11

	A	B	C	D	E
1	kwartały	t	Sprzedaż		
2	1	1,00	30,00		
3	2	2,00	28,00		
4	3	3,00	25,00		
5	4	4,00	23,00		
6	1	5,00	21,00		
7	2	6,00	20,00		
8	3	7,00	18,00	Prognoza przedziałowa	
9	4	8,00	12,00	Dolne	Górne
10		9,00	11,76	5,27	18,35
11		10,00	9,46	2,51	16,52

Prognoza przedziałowa dla następnego kwartału to [5,27; 18,35], a dla jeszcze następnego to [2,51; 16,52].

Przypadek 3

Polecenie:

Dokonano pomiaru kosztu wytworzenia pewnego towaru przy różnej wielkości produkcji tego wyrobu. Na podstawie poniższych danych oszacuj parametry funkcji regresji typu wykładniczego. Zadanie wykonaj w programie EXCEL.

Produkcja (tys. sztuk)	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Koszt (tys. zł)	1,6	2,1	2,4	2,9	3,5	4,4	5,5	6,8	8,5	10,2

Rozwiązanie:

Krok 1.

Wczytaj dane.

	A	B
1	Produkcja (tys. sztuk)	Koszt (tys. zł)
2	0,5	1,6
3	1	2,1
4	1,5	2,4
5	2	2,9
6	2,5	3,5
7	3	4,4
8	3,5	5,5
9	4	6,8
10	4,5	8,5
11	5	10,2

Krok 2.

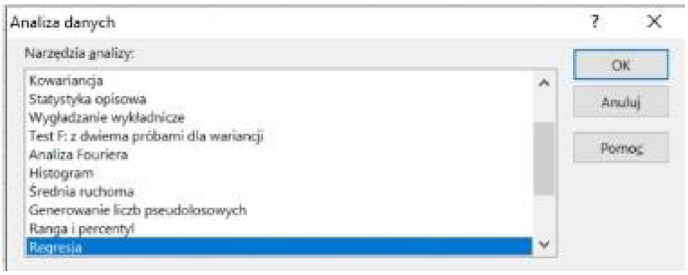
W komórce C2 wpisz =LN(B2) – skopiuj wartość na całą kolumnę.

W komórce D2 wpisz =$G$2*EXP($G$1*A2) – skopiuj wartość na całą kolumnę.

Komórki G2 oraz G1 będą zawierały informacje dotyczące parametrów funkcji – obliczymy je w dalszych krokach.

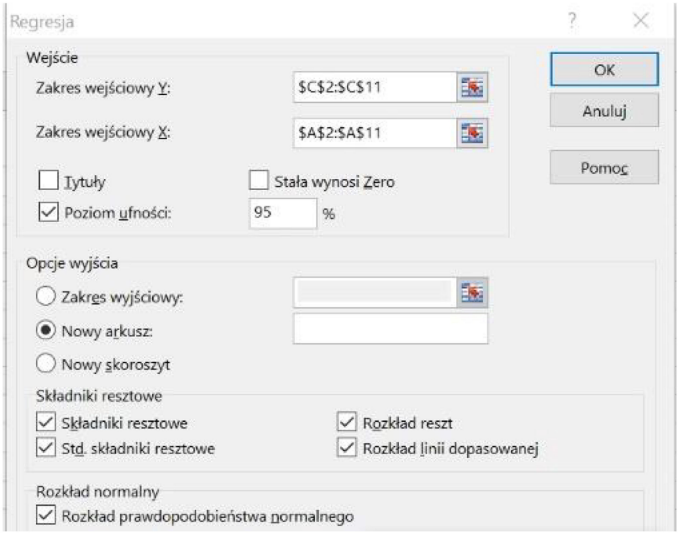
Krok 3.

Z menu wybierz *Dane/Analiza danych/Regresja*.



Krok 4.

Uzupełnij pola według poniższego schematu.



Kliknij *OK*. Program tworzy nowy arkusz z wynikami obliczeń i wykresami.

Krok 5.

Współczynnik a=0,41 natomiast współczynnik b = exp(0,27).

	Współczynniki	*Błąd standardowy*	*t Stat*	*Wartość-p*	*Dolne 95%*	*Górne 95%*	*Dolne 95,0%*	*Górne 95,0%*
Przecięcie	0,27	0,02	13,00	0,00	0,22	0,32	0,22	0,32
Zmienna X 1	0,41	0,01	61,23	0,00	0,39	0,43	0,39	0,43

Otrzymane wartości wpisujemy w komórkach G1 oraz G2 arkusza.

	F	G
	a	0,41
	b	1,31

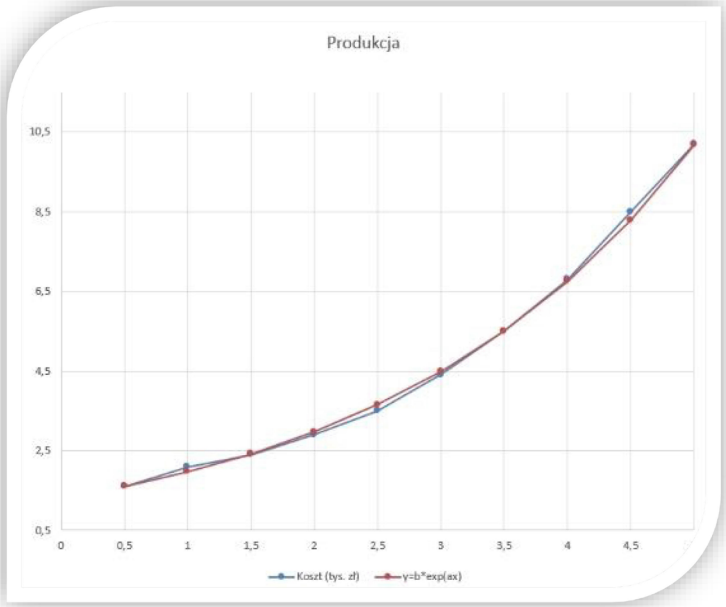
Krok 6.

W kolumnie D otrzymujemy wzór funkcji wykładniczej.

	A	B	C	D
1	Produkcja (tys. sztuk)	Koszt (tys. zł)	lny	y=b*exp(ax)
2	0,5	1,6	0,47	1,61
3	1	2,1	0,74	1,97
4	1,5	2,4	0,88	2,42
5	2	2,9	1,06	2,97
6	2,5	3,5	1,25	3,65
7	3	4,4	1,48	4,48
8	3,5	5,5	1,70	5,50
9	4	6,8	1,92	6,75
10	4,5	8,5	2,14	8,29
11	5	10,2	2,32	10,18

Krok 7.

Tworzymy wykres dopasowania wartości empirycznych do teoretycznych (zaznaczając dane z kolumn A, B oraz D).



Krok 8.

Interpretacja modelu:

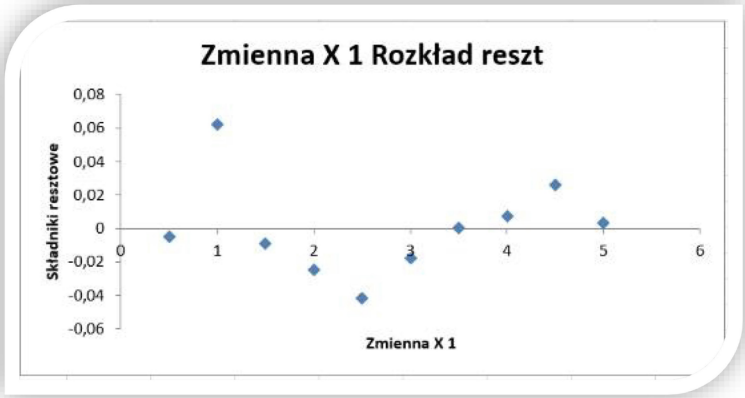
Równanie regresji i współczynniki regresji są istotne. Współczynniki korelacji R = 0,998, co świadczy o bardzo dobrym dopasowaniu.

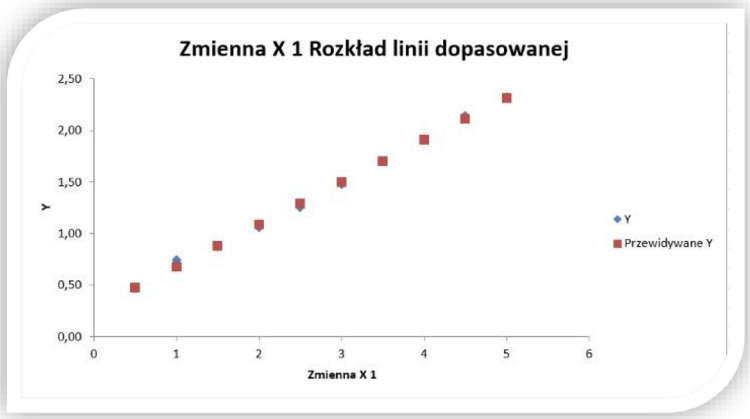
Współczynnik determinacji 0,997 oznacza, że model w pełni wyjaśnia poniesione koszty w zależności od produkcji wyrobu.

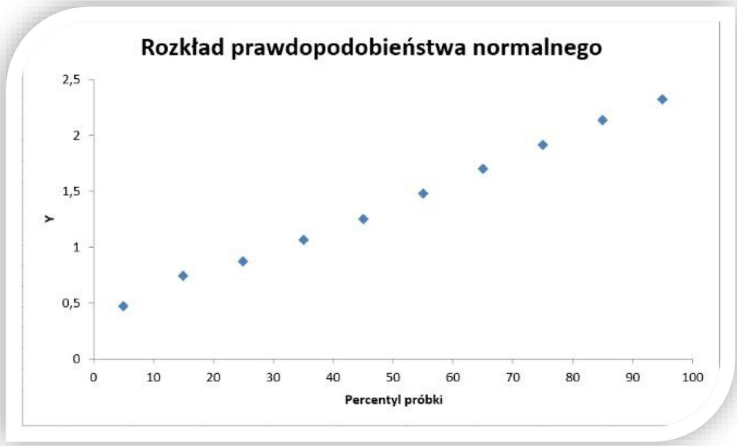
Z otrzymanych wykresów wynika, że reszty mają rozkład normalny, a odchylenia punktów od dopasowanej linii są losowe.

Można przyjąć, że ten model jest poprawny i może być stosowany do predykcji.

2		
3	*Statystyki regresji*	
4	Wielokrotność R	0,998934771
5	R kwadrat	0,997870676
6	Dopasowany R kwadrat	0,997604511
7	Błąd standardowy	0,030400855
8	Obserwacje	10







Przypadek 4

Polecenie:

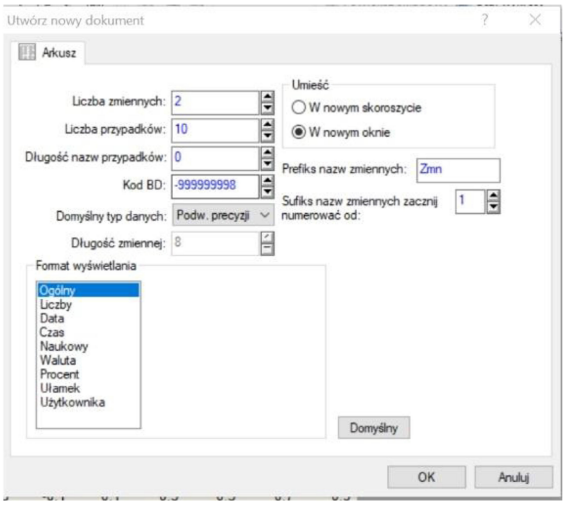
Dokonano pomiaru kosztu wytworzenia pewnego towaru przy różnej wielkości produkcji tego wyrobu. Na podstawie poniższych danych oszacuj parametry funkcji regresji typu wykładniczego. Zadanie wykonaj w programie STATISTICA.

Produkcja (tys. sztuk)	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Koszt (tys. zł)	1,6	2,1	2,4	2,9	3,5	4,4	5,5	6,8	8,5	10,2

Rozwiązanie:

Krok 1.

Otwórz nowy arkusz dla 2 zmiennych i 10 danych.



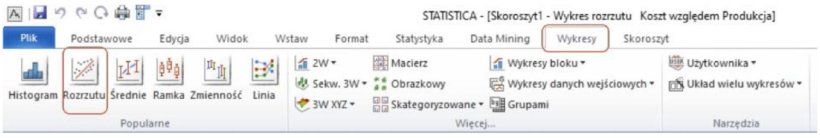
Krok 2.

W nagłówkach kolumn wpisz nazwy zmiennych.

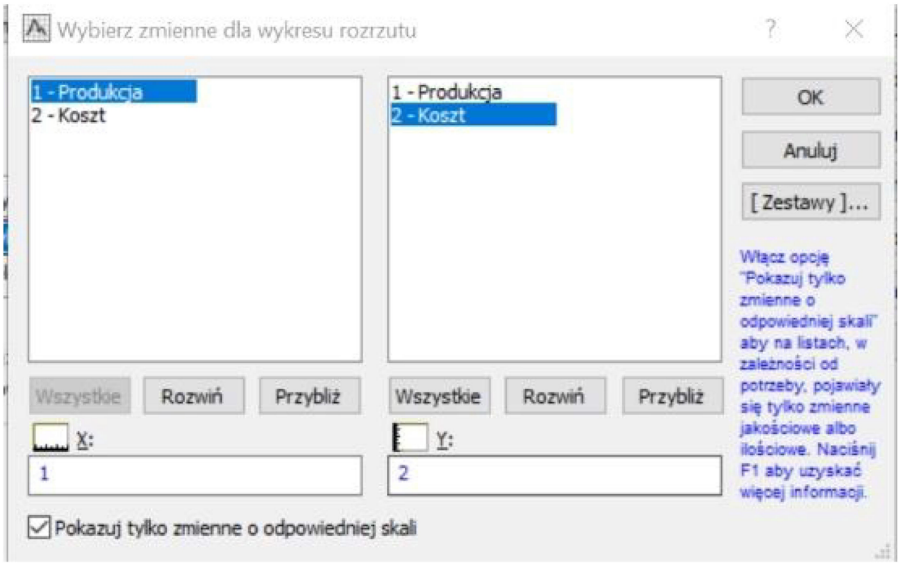
	1 Produkcja	2 Koszt
1	0,5	1,6
2	1	2,1
3	1,5	2,4
4	2	2,9
5	2,5	3,5
6	3	4,4
7	3,5	5,5
8	4	6,8
9	4,5	8,5
10	5	10,2

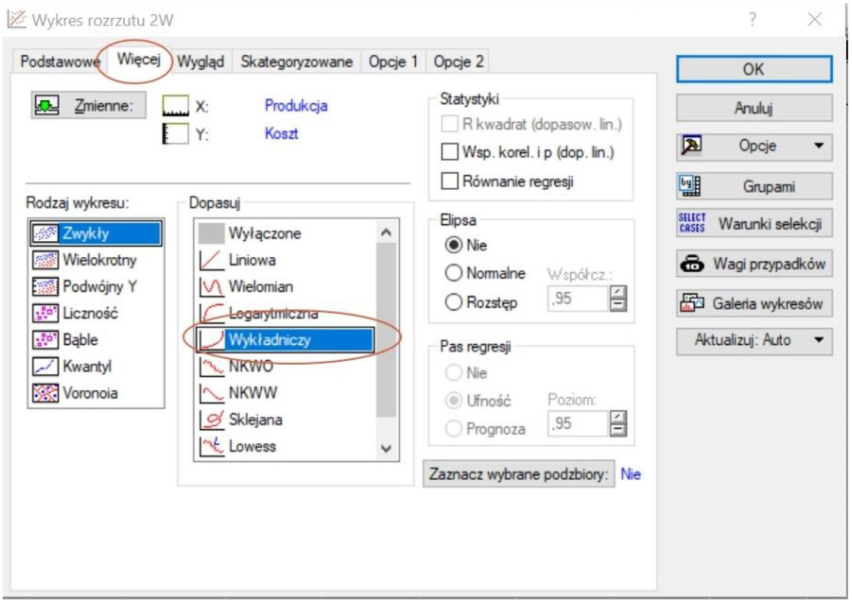
Krok 3.

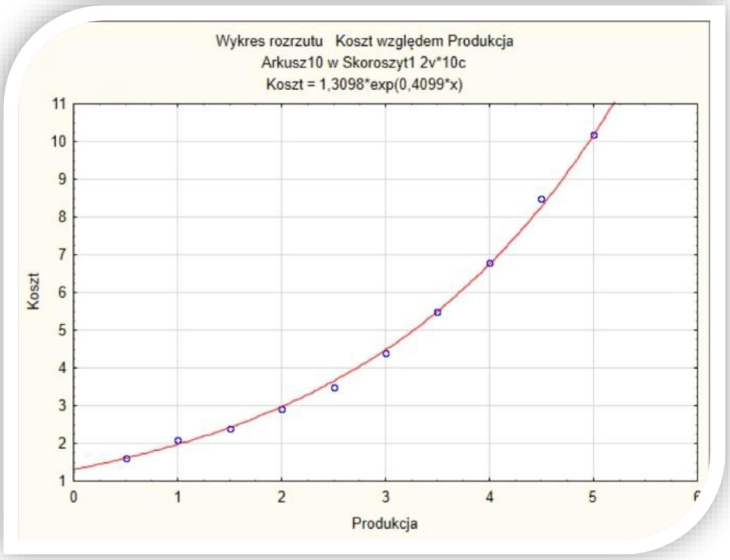
Utwórz wykres rozrzutu.



W oknie *Dopasuj* wybierz *Wykładnicza*. Jako zmienną X podaj Produkcję, a zmienną Y – koszt.



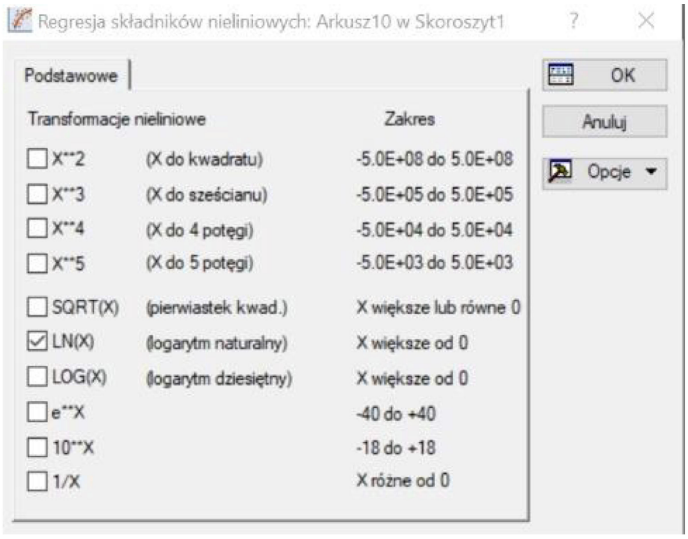




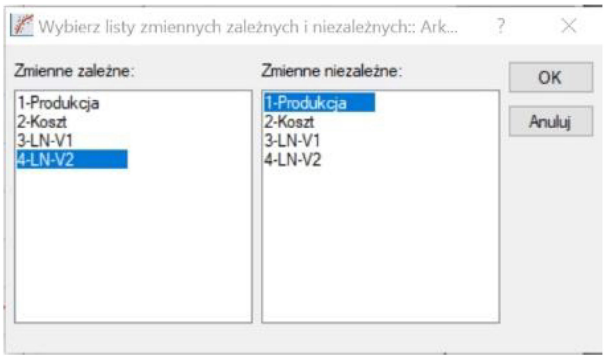
Uzyskaliśmy równanie funkcji

Krok 4.

Wybierz z menu: *Statystyka/Modele zaawansowane*. Następnie *Linearyzowana regresja nieliniowa*. Wybierz zmienne. Kliknij *OK*. Dalej wybierz transformacje nieliniowe: *LN(X)*.



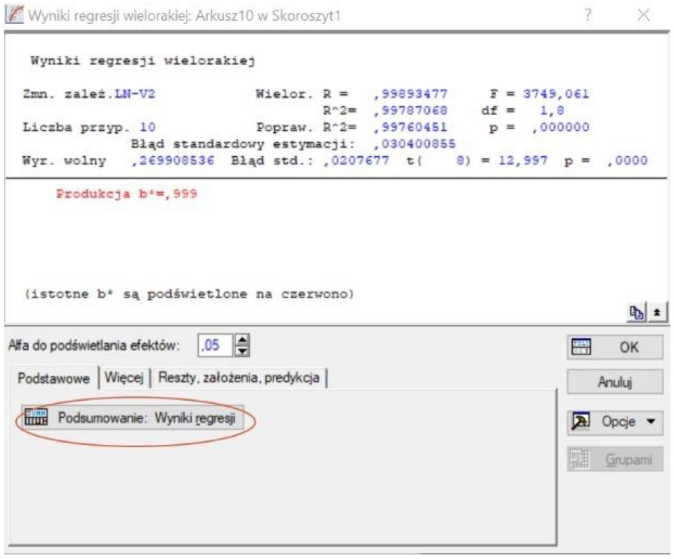
Wybierz zmienne tak jak na poniższym rysunku.



Kliknij *OK*. Program wyświetla okno do dalszych analiz.

Krok 5.

W zakładce *Podstawowe* wybierz przycisk *Podsumowanie: Wyniki regresji*. Program tworzy arkusz z wynikami.



N=10	Podsumowanie regresji zmiennej zależnej: LN-V2 (Arkusz10 w Skoroszyt1) R= ,99893477 R^2= ,99787068 Popraw. R2= ,99760451 F(1,8)=3749,1 p<,00000 Błąd std. estymacji: ,03040					
	b*	Bł. std. z b*	b	Bł. std. z b	t(8)	p
W. wolny			0,269909	0,020768	12,99653	0,000001
Produkcja	0,998935	0,016315	0,409874	0,006694	61,22958	0,000000

statystyka	Stat.podsum.; Zmn. zal.:LN-V2 **Wartość**
R wielorakie	0,998934771
Wielorakie R2	0,997870676
Skorygowane R2	0,997604511
F(1,8)	3749,06097
p	0,000000000000562594199
Błąd std. estymacji	0,0304008549

Otrzymane równanie ma postać:

$$y = 0{,}41e^{1{,}31x} \pm 0{,}02$$

gdzie
parametr b = exp(0,2699)=1,3098.

Równanie regresji i współczynniki regresji są istotne. Współczynnik korelacji R = 0,9976, co świadczy o bardzo dobrym dopasowaniu. Można sprawdzić, że reszty mają rozkład normalny, a odchylenia punktów od dopasowanej linii są losowe.

Zadania

Zadanie 1.

Wartość sprzedanych (w tys. zł) produktów w poszczególnych półroczach lat 2010–2014 kształtowała się następująco:

10 8 16 16 20 14 21 18 24 19

Wyznacz prognozę punktową i przedziałową w poszczególnych półroczach 2015 r. Oceń wielkość błędów prognozy (poziom wiarygodności prognozy 0,95).

.....
.....
.....
.....
.....
.....

Zadanie 2.

Na podstawie danych z kolejnych lat opisujących rozwój sieci sprzedaży pewnego przedsiębiorstwa (w szt.):

2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
42	21	52	36	34	38	40	40	54	48	50

oszacuj parametry liniowej funkcji trendu przy założeniu (t=1,2,...n) oraz wyznacz prognozę punktową liczby punktów handlowych tego przedsiębiorstwa w 2011 roku oraz bezwzględny i względny błąd prognozy *ex ante*.

.....
.....
.....
.....
.....
.....

Studium przypadku 3.
Szeregi czasowe – prognozowanie

Wstęp teoretyczny

W prognostycznym modelu średniej ruchomej przyjmuje się, że prognoza zmiennej na moment *t* jest:

- **średnią arytmetyczną (w przypadku średniej ruchomej prostej):**

$$y_t^* = \frac{1}{k}\sum_{i=t-k}^{t-1} y_i$$

- **średnią ważoną (w przypadku średniej ruchomej ważonej)**

$$y_t^* = \sum_{i=t-k}^{t-1} y_i \cdot w_i$$

z *k* ostatnich obserwacji w szeregu.

W modelu **wygładzania Browna** prognoza zmiennej na moment równa jest prognozie tej zmiennej na momentskorygowanej o iloczyn parametru wygładzania *α* i wartości jej bezwzględnego błędu *ex post:*

$$y_t^* = \alpha \cdot y_{t-1} + (1-\alpha) \cdot y_{t-1}^*$$

gdzie:
y^*_t – prognoza w wyznaczonym okresie t
y_{t-1} – wielkość w okresie t-1

Parametr α wybiera się na podstawie kryterium najmniejszego błędu *ex post* prognoz wygasłych.

Kolejna metoda wygładzania wykładniczego to **model Holta** opisany następującymi równaniami:

$$F_{t-1} = \alpha y_{t-1} + (1-\alpha)(F_{t-2} + S_{t-2})$$
$$S_{t-1} = \beta(F_{t-1} + S_{t-2}) + (1-\alpha)S_{t-2}$$

gdzie:
F_{t-1} – ocena wartości średniej w okresie t-1
S_{t-1} – przyrost trendu w okresie t-1
α, β – parametry modelu z przedziału (0,1]

Natomiast **równanie prognozy** ma następującą postać:

$$Y_t^* = F_n + (t-n)S_n$$

Do opisu **modelu addytywnego Wintera** służy następujący układ równań:

$$F_t = \alpha(Y_t - C_{t-r}) + (1-\alpha)(F_{t-1} + S_{t-1})$$
$$S_t = \beta(F_t - F_{t-1}) + (1-\beta)S_{t-1}$$
$$C_t = \gamma(Y_t - F_t) + (1-\gamma)C_{t-r}$$

gdzie:
F_t – oszacowanie części stałej szeregu podlegającej wahaniom przypadkowym
S_t – oszacowanie części przyrostowej szeregu podlegającej wahaniom przypadkowym
C_t – oszacowanie części sezonowej trendu podlegającej wahaniom przypadkowym
Y_t – wartości empiryczne szeregu w okresie t
y^*_t – wartości teoretyczne szeregu wygładzonego
α, β, γ – parametry modelu z przedziału (0,1]

Natomiast **równanie prognozy** ma następującą postać:

$$Y_t^* = F_n + (t-n)S_n + C_{t-r}$$

Do opisu **modelu multiplikatywnego Wintera** służy następujący układ równań:

$$F_t = \alpha(Y_t/C_{t-r}) + (1-\alpha)(F_{t-1} + S_{t-1})$$
$$S_t = \beta(F_t - F_{t-1}) + (1-\beta)S_{t-1}$$
$$C_t = \gamma(Y_t/F_t) + (1-\gamma)C_{t-r}$$

gdzie:
F_t – oszacowanie części stałej szeregu podlegającej wahaniom przypadkowym
S_t – oszacowanie części przyrostowej szeregu podlegającej wahaniom przypadkowym

C_t – oszacowanie części sezonowej trendu podlegającej wahaniom przypadkowym
Y_t – wartości empiryczne szeregu w okresie t
y^*_t – wartości teoretyczne szeregu wygładzonego
α, β, γ – parametry modelu z przedziału (0,1]

Natomiast **równanie prognozy** ma następującą postać:

$$Y_t^* = (F_n + (t-n)S_n)C_{t-r}$$

Wartości początkowe F i S można obliczyć, wykorzystując średnią ruchomą 3-elementową.

Definicja:

Średni błąd predykcji (mean erros – ME)

$$\mathbf{ME}=\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

Definicja:

Średni absolutny błąd predykcji MAE
(mean absolute error)

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}|y_i - y_i^*|$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

Definicja:

Średni absolutny błąd procentowy prognozy MAPE (mean absolute percentage error)

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - y_t^*|}{y_t} * 100\%$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

Definicja:

Średniokwadratowy predykcji RMSE (root mean square error):

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^*)^2}$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

Definicja:

Średni procentowy błąd prognoz (mean percentage error)

$$MPE = 1/n \sum_{t=1}^{n} \frac{|y_t - y_t^*|}{y_t}$$

gdzie:
n – liczba prognoz
y_t – rzeczywista wartość zmiennej
y^*_t – przewidywana wartość zmiennej

Przypadek 1

Polecenie:

Liczbę sprzedanych komputerów w jednym z lubelskich sklepów w ciągu dziesięciu kolejnych tygodni przedstawia poniższa tabela.

a) Przedstaw graficznie dane.
b) Zaprognozuj metodą średniej ruchomej (3-okresowej oraz 4-okresowej) sprzedaż komputerów w jedenastym tygodniu.
c) Oblicz błędy prognozy.

Tydzień	1	2	3	4	5	6	7	8	9	10
Liczba komputerów	28	27	26	27	24	20	22	21	23	22

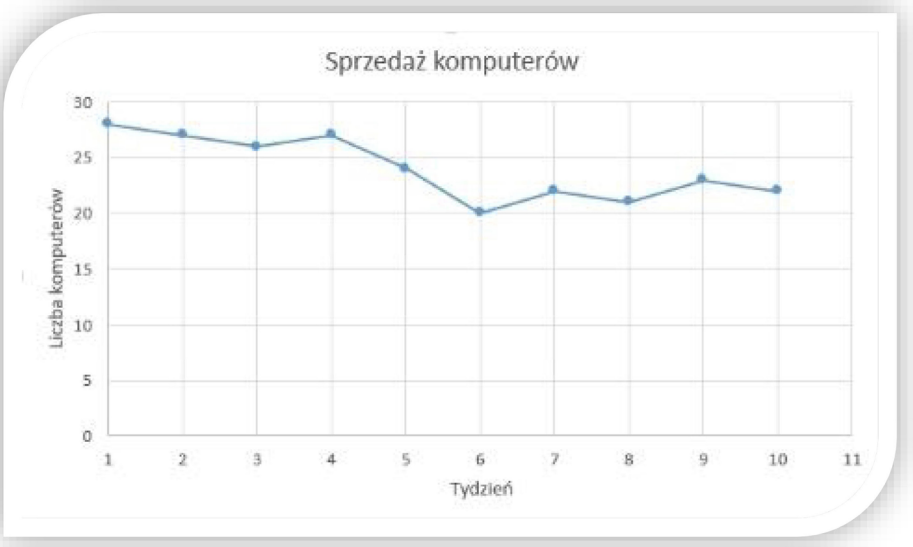
Rozwiązanie:

Krok 1.

Wprowadzamy dane w dwóch kolumnach. Zaznaczamy dane, następnie z menu *Wstawianie/Wykresy* wybieramy typ wykresu *Punktowy z prostymi liniami i znacznikami*.



Otrzymany wykres formatujemy, dodając tytuły osi i tytuł wykresu.



Krok 2.

Obliczymy prognozę metodą 4-okresowej średniej ruchomej. W komórce C6 wpisujemy =(B2+B3+B4+B5)/4, następnie wyrażenie kopiujemy na całą kolumnę danych. Uzyskana prognoza wynosi 22.

	A	B	C
1	Tydzień	Liczba komputerów	średnia 4 okresowa - y*_i
2	1	28	
3	2	27	
4	3	26	
5	4	27	
6	5	24	27
7	6	20	26
8	7	22	24,25
9	8	21	23,25
10	9	23	21,75
11	10	22	21,5
12	11		22

Krok 3.

Teraz obliczymy błędy otrzymanej prognozy. Dodajemy dwie kolumny do tabeli danych. W komórce D6 wpisujemy =(B6-C6) i kopiujemy wartość na całą kolumnę D. W komórce E6 wpisujemy =D6^2 i kopiujemy wartość na całą kolumnę E. Sumujemy wartości otrzymane w kolumnach C, D, E.

	A	B	C	D	E
1	Tydzień	Liczba komputerów	średnia 4 okresowa - y*_i	blad(yi-y*_i)	(yi-y_i*)^2
2	1	28			
3	2	27			
4	3	26			
5	4	27			
6	5	24	27	-3	9,00
7	6	20	26	-6	36,00
8	7	22	24,25	-2,25	5,06
9	8	21	23,25	-2,25	5,06
10	9	23	21,75	1,25	1,56
11	10	22	21,5	0,5	0,25
12	11		22		
13			165,75	-12	56,94

Krok 4.

Obliczamy średni błąd kwadratowy, średni błąd procentowy oraz błąd bezwzględny procentowy dla otrzymanej prognozy 4-okresowej.

W komórce I2 wpisujemy =PIERWIASTEK(E13/6)
W komórce I3 wpisujemy =D13/C13*100%
W komórce I4 wpisujemy =E13/C13*100%

Otrzymujemy następujące wyniki z zakresu błędów prognozy.

	H	I
	średni błąd kwadratowy	3,08
	średni błąd procentowy	-7%
	błąd bezwzględny procentowy	34%

Krok 5.

Tydzień	Liczba komputerów	średnia 3 okresowa - y*_i	błąd(yi-y*_i)	(yi-y_i*)^2
1	28			
2	27			
3	26			
4	27	27,00	0,00	0,00
5	24	26,67	-2,67	7,11
6	20	25,67	-5,67	32,11
7	22	23,67	-1,67	2,78
8	21	22,00	-1,00	1,00
9	23	21,00	2,00	4,00
10	22	22,00	0,00	0,00
11		22		
		190,00	-9,00	47,00

Kopiujemy dane dotyczące liczby sprzedanych komputerów do obszaru A18:B22. Obliczymy prognozę metodą 3-okresowej średniej ruchomej. W komórce C22 wpisujemy =(B19+B20+B21)/3, następnie wyrażenie kopiujemy na całą kolumnę danych. Uzyskana prognoza wynosi 22.

17			
18	Tydzień	Liczba komputerów	średnia 3 okresowa - y*_i
19	1	28	
20	2	27	
21	3	26	
22	4	27	27,00
23	5	24	26,67
24	6	20	25,67
25	7	22	23,67
26	8	21	22,00
27	9	23	21,00
28	10	22	22,00
29	11		22

Krok 6.

Teraz obliczymy błędy otrzymanej prognozy. Dodajemy dwie kolumny do tabeli danych. W komórce D22 wpisujemy =(B22-C22) i kopiujemy wartość na całą kolumnę D. W komórce E22 wpisujemy =D22^2 i kopiujemy wartość na całą kolumnę E. Sumujemy wartości otrzymane w kolumnach C, D, E.

średni blad kwadratowy prognoza	2,59
redni błąd procentowy	-5%

Przypadek 2

Polecenie:

Liczbę sprzedanych telewizorów w jednym z lubelskich sklepów w ciągu piętnastu kolejnych tygodni przedstawia poniższa tabela.

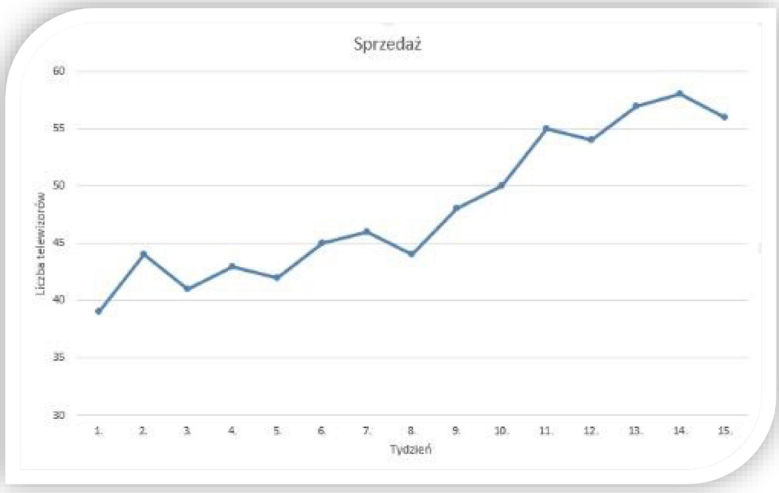
a) Przedstaw graficznie dane.
b) Zaprognozuj metodą Holta sprzedaż telewizorów w szesnastym tygodniu.
c) Oblicz błędy prognozy.

Tydzień	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Liczba telewizorów	39	44	41	43	42	45	46	44	48	50	55	54	57	58	56

Rozwiązanie:

Krok 1.

Wprowadź dane w kolumnach *A* i *B*. Utwórz wykres, wybierając *Wstawianie/Wykresy*.



Na podstawie wykresu widzimy, że w szeregu występują wahania przypadkowe oraz trend liniowy.

Krok 2

Dokonamy prognozy metodą Holta. Najpierw należy ustalić wartości stałych alpha oraz beta. Wielkości te wpisujemy w komórkach B20 oraz B21.

19		
20	Alpha	0,3
21	Beta	0,75

Krok 3.

W komórce C2 wpisujemy wartość =B2, a w komórce C3 wartość =$B$20*B2+(1-$B$20)*(C2+D2). Następnie kopiujemy wartość z komórki C3 na kolumnę C.
W komórce D2 wpisujemy wartość =B3-B2, a w komórce D3 wartość =$B$21*(C3-C2)+(1-$B$21)*D2. Następnie kopiujemy wartość z komórki D3 na kolumnę D.
W komórce E2 wpisujemy wartość =C2+D2. Następnie kopiujemy wartość z komórki E2 na kolumnę E.

	A	B	C	D	E
1	Tydzień	TV y_t	F_t	S_t	Prognoza y_t*
2	1.	39	39	5	44
3	2.	44	42,50	3,88	46,38
4	3.	41	45,66	3,34	49,00
5	4.	43	46,60	1,54	48,14
6	5.	42	46,60	0,38	46,98
7	6.	45	45,49	-0,74	44,75
8	7.	46	44,82	-0,68	44,14
9	8.	44	44,70	-0,26	44,44
10	9.	48	44,31	-0,36	43,94
11	10.	50	45,16	0,55	45,71
12	11.	55	47,00	1,52	48,51
13	12.	54	50,46	2,98	53,43
14	13.	57	53,60	3,10	56,71
15	14.	58	56,79	3,17	59,96
16	15.	56	59,37	2,73	62,10
17	16.		60,27	1,35	62

Otrzymana prognoza to 62 sztuki telewizorów.

Krok 4.

W wierszu 18 obliczamy sumy odpowiednich kolumn.

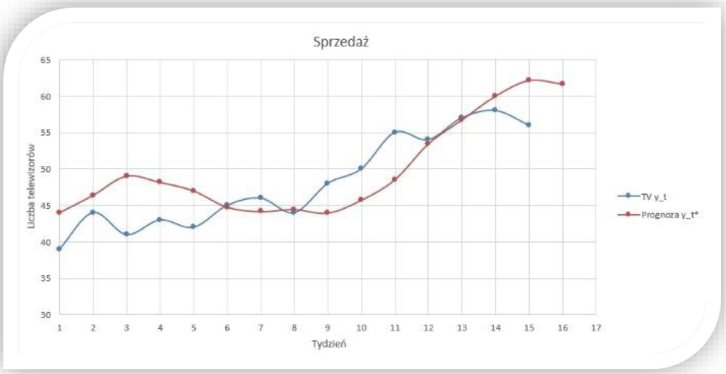
15	14.	58	56,79	3,17	59,96	-1,96	3,85	-0,03	0,03	1,96
16	15.	56	59,37	2,73	62,10	-6,10	37,22	-0,11	0,11	6,10
17	16.		60,27	1,35	62					
18					Suma	-16,20	268,07	-0,42	1,12	51,80
19										

Krok 5

Obliczamy błędy, wpisując kolejno w komórki
D23 wartość =PIERWIASTEK(G18/15)
D25 wartość =F18/15
D27 wartość =H18/15
D29 wartość =J18/15
D31 wartość = I18
D33 wartość =D31/15

22		
23	**średni błąd kwadratowy SSE**	4,23
24		
25	**błąd średni progn.**	-1,08
26		
27	**średni błąd % SPE**	-3%
28		
29	**średni błąd bezwz.**	3,45
30		
31	**błąd bezwz. %**	112%
32		
33	**średni bezwz. błąd %**	7%
34		

Krok 6.



Przypadek 3

Polecenie:

Dochody miesięczne w jednym z lubelskich sklepów w ciągu dwunastu kolejnych tygodni przedstawia poniższa tabela.

Tydzień	1	2	3	4	5	6	7	8	9	10	11	12
Dochody (tys. zł)	18,3	20,5	19	20,7	23,1	21,1	22,4	24,5	22,8	24,2	25,7	27,3

a) Przedstaw graficznie dane.
b) Zaprognozuj metodą wygładzania wykładniczego (dla alpha=0,1 oraz alpha =0,05) dochody sklepu w trzynastym tygodniu.
c) Oblicz błędy prognozy.

Rozwiązanie:

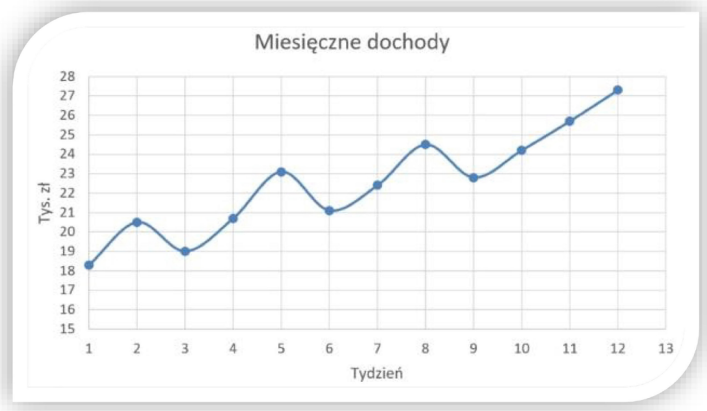
Krok 1.

Wprowadzamy dane do arkusza kalkulacyjnego do kolumny A oraz kolumny B, w komórce C17 wpisujemy wartość alpha równą 0,1.

	A	B
1	**Tydzień**	**Dochody yt**
2	1.	18,3
3	2.	20,5
4	3.	19
5	4.	20,7
6	5.	23,1
7	6.	21,1
8	7.	22,4
9	8.	24,5
10	9.	22,8
11	10.	24,2
12	11.	25,7
13	12.	27,3
14	13.	

Krok 2.

Zaznaczamy dane i korzystając w wykresu punktowego, dokonujemy graficznej prezentacji danych.



Krok 3.

W kolejnej kolumnie C wpisujemy prognozę zgodnie ze wzorami na prognozowanie metodą wykładniczą. W komórce C2 wpisujemy wartość równą B2, dalej w komórce C3 wpisujemy =$C$17*B2+(-1-$C$17)*C2. Kopiując wpisaną wartość na dalsze komórki w kolumnie C, otrzymujemy szukaną prognozę.

	A	B	C
1	**Tydzień**	**Dochody yt**	**Prognoza yt***
2	1.	18,3	18,3
3	2.	20,5	18,3
4	3.	19	18,52
5	4.	20,7	18,57
6	5.	23,1	18,78
7	6.	21,1	19,21
8	7.	22,4	19,40
9	8.	24,5	19,70
10	9.	22,8	20,18
11	10.	24,2	20,44
12	11.	25,7	20,82
13	12.	27,3	21,31
14	13.		21,91

Krok 4.

Obliczając błędy otrzymanej prognozy, wpisujemy kolejno w komórce D2 wyrażenie =B2-C2, w komórce E2 wyrażenie =D2^2, w komórce F2 wyrażenie =D2/B2, w komórce G2 wyrażenie =MODUŁ. LICZBY(D2), w komórce H2 wyrażenie =G2/B2. Wszystkie wpisane wyrażenia kopiujemy na całe kolumny. W ostatnim wierszu obliczamy sumy.

	A	B	C	D	E	F	G	H
1	**Tydzień**	**Dochody yt**	**Prognoza yt***	**yt-yt***	**(yt-yt*)^2**	**(yt-yt*)/yt**	**|yt-yt*|**	**|yt-yt*|/yt**
2	1.	18,3	18,3	0,00	0,00	0,00	0,00	0,00
3	2.	20,5	18,3	2,20	4,84	0,11	2,20	0,11
4	3.	19	18,52	0,48	0,23	0,03	0,48	0,03
5	4.	20,7	18,57	2,13	4,55	0,10	2,13	0,10
6	5.	23,1	18,78	4,32	18,65	0,19	4,32	0,19
7	6.	21,1	19,21	1,89	3,56	0,09	1,89	0,09
8	7.	22,4	19,40	3,00	8,99	0,13	3,00	0,13
9	8.	24,5	19,70	4,80	23,02	0,20	4,80	0,20
10	9.	22,8	20,18	2,62	6,86	0,11	2,62	0,11
11	10.	24,2	20,44	3,76	14,11	0,16	3,76	0,16
12	11.	25,7	20,82	4,88	23,82	0,19	4,88	0,19
13	12.	27,3	21,31	5,99	35,92	0,22	5,99	0,22
14	13.		21,91					
15			Suma	36,06	144,55	1,52	36,06	1,52

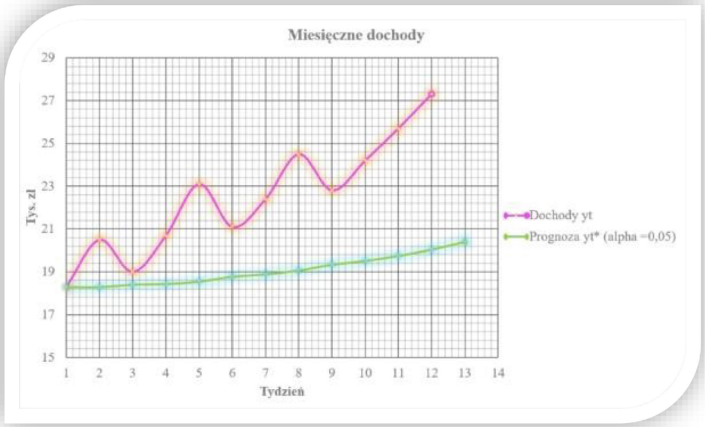
Krok 5.

Obliczając błędy, wpisujemy kolejno
w komórce F17 wyrażenie =PIERWIASTEK(E13/12)
w komórce F19 wyrażenie =D15/12
w komórce F21 wyrażenie =F15/12
w komórce F23 wyrażenie =G15/12
w komórce F25 wyrażenie =H15
w komórce F27 wyrażenie =F25/12

średni kwadratowy błąd prognozy, SSE	1,73
błąd średni prognozy, SP	3,01
średni (względny) błąd %, SPE	12,68%
średni błąd bezwzględny prognozy	3,01
błąd bezwzględny %, PE	152,12%
średni bezwzględny błąd %, MPE	=F25/12

Krok 6.

Dokonujemy jeszcze raz graficznej prezentacji danych. Na jednym wykresie umieszczamy dane rzeczywiste oraz otrzymane prognozy.



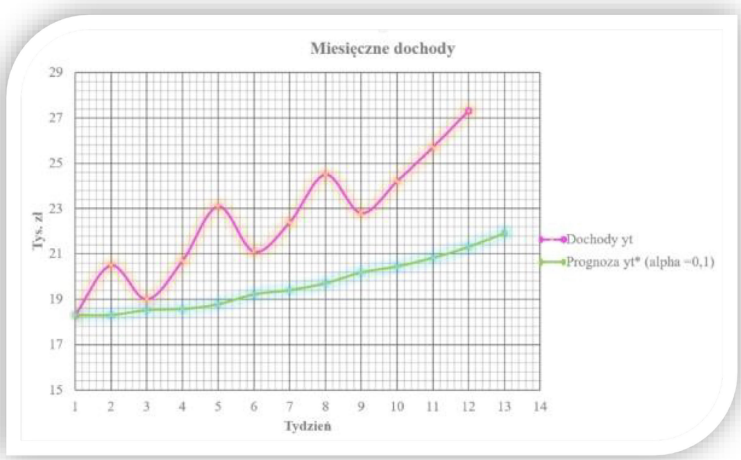
Krok 7.

Powtarzamy obliczenia dla aplha wynoszącego 0,05.

	A	B	C	D	E	F	G	H
1	**Tydzień**	**Dochody yt**	**Prognoza yt***	**yt-yt***	**(yt-yt*)^2**	**(yt-yt*)/yt**	**|yt-yt*|**	**|yt-yt*|/yt**
2	1.	18,3	18,3	0,00	0,00	0,00	0,00	0,00
3	2.	20,5	18,3	2,20	4,84	0,11	2,20	0,11
4	3.	19	18,41	0,59	0,35	0,03	0,59	0,03
5	4.	20,7	18,44	2,26	5,11	0,11	2,26	0,11
6	5.	23,1	18,55	4,55	20,68	0,20	4,55	0,20
7	6.	21,1	18,78	2,32	5,38	0,11	2,32	0,11
8	7.	22,4	18,90	3,50	12,28	0,16	3,50	0,16
9	8.	24,5	19,07	5,43	29,47	0,22	5,43	0,22
10	9.	22,8	19,34	3,46	11,95	0,15	3,46	0,15
11	10.	24,2	19,52	4,68	21,95	0,19	4,68	0,19
12	11.	25,7	19,75	5,95	35,41	0,23	5,95	0,23
13	12.	27,3	20,05	7,25	52,60	0,27	7,25	0,27
14	13.		20,41					
15			Suma	42,20	200,02	1,77	42,20	1,77

Poniżej podsumowanie otrzymanych błędów.

średni kwadratowy błąd prognozy, SSE	2,09
błąd średni prognozy, SP	3,52
średni (względny) błąd %, SPE	14,79%
średni błąd bezwzględny prognozy	3,52
błąd bezwzględny %, PE	177,48%
średni bezwzględny błąd %, MPE	14,79%



Zadania

Zadanie 1.

Ilość sprzedanych książek (w sztukach) w pewnym sklepie w kolejnych 12 tygodniach kształtowała się następująco:

100	110	118	209	195	137	144	107	233	128	165	176

Wyznacz prognozę ilości sprzedanych książek w kolejnym tygodniu.

a) Wyznacz prognozę metodą średniej ruchomej prostej ze stałą wygładzania $k = 3$ i $k = 4$.
b) Wyznacz prognozę metodą średniej ruchomej ważonej ze stałą wygładzania $k = 3$ ($w_1 = 0{,}2$; $w_2 = 0{,}3$; $w_3 = 0{,}5$).
c) Oceń, która z poprzednich metod daje najlepsze wyniki?
d) Oceń trafność prognozy dla najlepszej metody, wiedząc, że w 13 tygodniu sprzedano 215 książek.

.....
.....
.....
.....
.....

Zadanie 2.

Liczba studentów na pewnej lubelskiej uczelni w latach 2010–2020 (w tys. osób) była następująca:

10,2	12,6	11,9	12,5	12,3	14,2	12,0	11,3	10,9	13,2

Wyznacz prognozę liczby studentów w 2021 roku.

a) Wyznacz prognozę metodą średniej ruchomej prostej ze stałą wygładzania $k = 3$ i $k = 4$.
b) Wyznacz prognozę metodą średniej ruchomej ważonej ze stałą wygładzania $k = 3$ ($w_1 = 0{,}1$; $w_2 = 0{,}1$; $w_3 = 0{,}8$).

c) Oceń, która z poprzednich metod daje najlepsze wyniki?
d) Oceń trafność prognozy dla najlepszej metody, wiedząc, że w 2021 roku było 12,1 tys. studentów.

.....
.....
.....
.....
.....
.....
.....
.....

Zadanie 3.

Na podstawie danych z lat 2007–2014 dotyczących wartości sprzedaży w tys. zł pewnego produktu:

Rok	2007	2008	2009	2010	2011	2012	2013	2014
Sprzedaż	50	51	59	68	51	73	60	61

a) Wyznacz prognozę sprzedaży w roku 2015. Zastosuj kilka metod prognozowania.
b) Oceń trafność prognoz.

.....
.....
.....
.....
.....
.....
.....
.....
.....

Załączniki

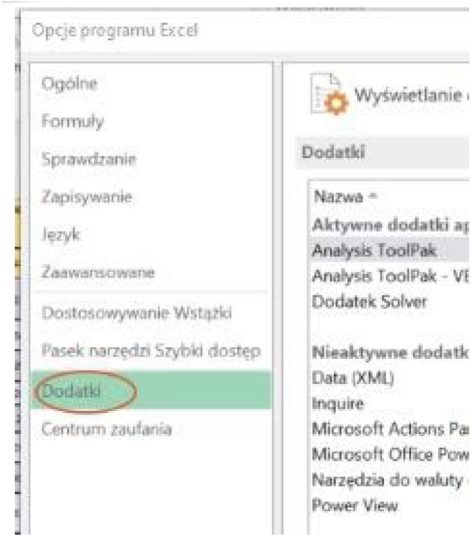
Załącznik 1 – instalacja dodatku Analiza danych

Podczas rozwiązywania zadań ze skryptu skorzystamy z dodatku *Analiza danych*. Jeśli nie masz zainstalowanego tego dodatku (jest widoczny na karcie *Dane*), to wykonaj poniższe kroki.

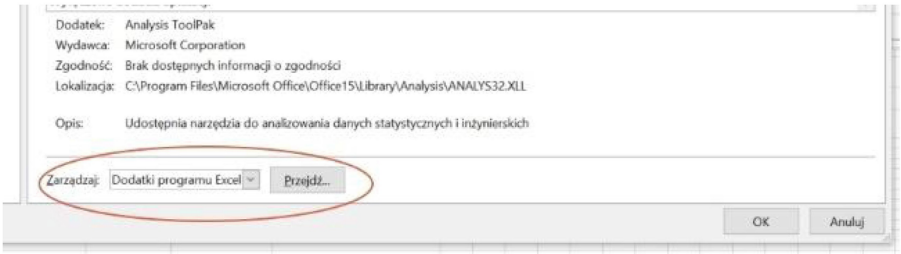
1. Kliknij kartę *Plik* i przycisk *Opcje* w jej pasku nawigacyjnym.



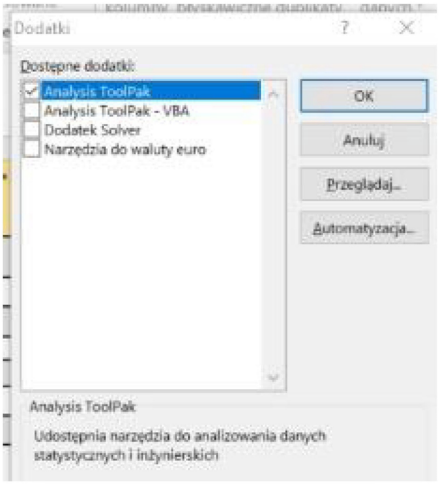
2. Otworzy się okno *Opcje programu Excel*. Kliknij zakładkę *Dodatki* w pasku nawigacyjnym.



3. W razie potrzeby wybierz pozycję *Dodatki programu Excel* z rozwijanej listy *Zarządzaj* i kliknij przycisk *Przejdź*.



4. Pojawi się okno dialogowe *Dodatki*. Jeżeli zobaczysz na liście dodatek *Analysis ToolPak*, upewnij się, że odpowiadające mu pole wyboru jest zaznaczone (*Analysis ToolPak* to stare określenie tego dodatku). Kliknij przycisk *OK.* Przycisk *Analiza danych* powinien być teraz widoczny w grupie *Analiza* na karcie *Dane*.



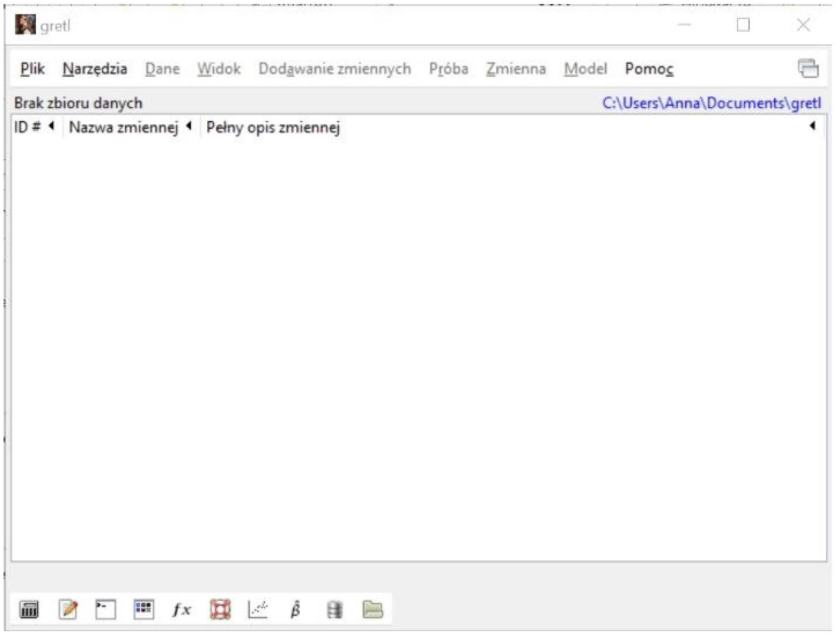


Załącznik 2 – oprogramowanie GRETL

Jednym z oprogramowań używanych w skrypcie jest program GRETL, autorstwa Allina Cottrella z Uniwersytetu Wake Forest w Północnej Karolinie w Stanach Zjednoczonych. Należy do oprogramowań Powszechnej Licencji Publicznej (GNU), czyli o swobodnym dostępie dla wszystkich użytkowników.

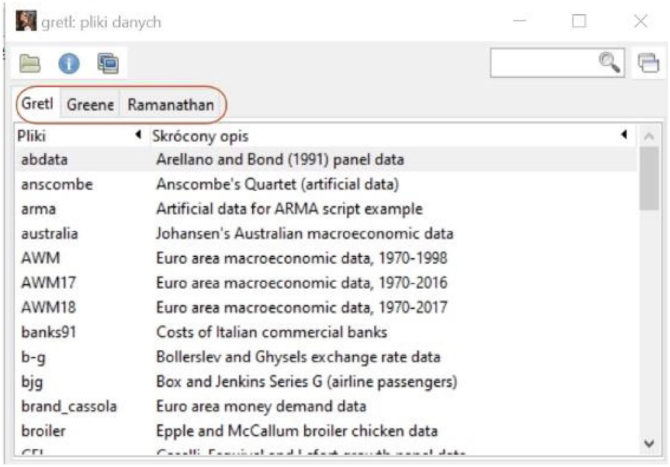
Program ten, niezwykle popularny i powszechnie stosowany na świecie, zawiera podstawowe procedury i metody ekonometryczne, przygotowany jest tak, aby można go było wykorzystać na zajęciach z ekonometrii.

Niewątpliwą zaletą tego oprogramowania jest fakt, że jest on darmowy. Plik instalacyjny należy pobrać ze strony http://www.kufel.torun.pl/. Ekran startowy programu wygląda następująco:



Aby skorzystać z istniejącej bazy danych, korzystamy z opcji *Plik/Otwórz dane/Pliki użytkowników* (alternatywnie można nacisnąć Ctrl+O) oraz podajemy lokalizację pliku z bazą. Istnieje również możliwość wykorzystania jednej z dostępnych w ramach Gretla baz

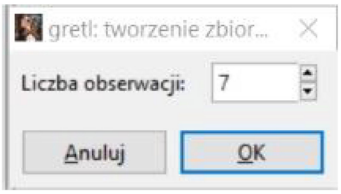
danych. Wybierzemy wtedy *Plik/Otwórz dane/Pliki z przykładami*. Wyświetli się okno z zakładkami odpowiadającymi zainstalowanym bazom danych (standardowo są trzy zakładki – Ramanathan, Green i Gretl). W każdej zakładce jest od kilkunastu do kilkudziesięciu różnych baz danych.

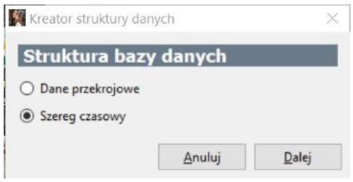


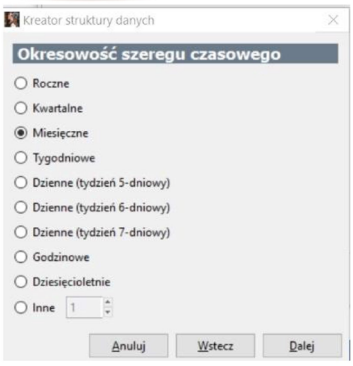
Poniżej przedstawiamy krótkie ćwiczenie ilustrujące wprowadzanie danych (w postaci poniższej tabeli) do programu GRETL.

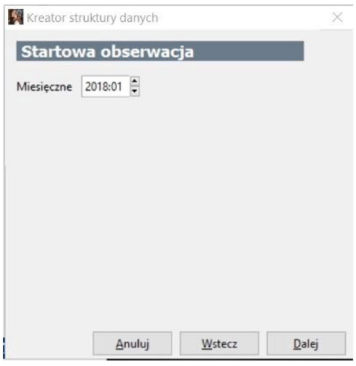
Miesiąc	Sprzedaż
Styczeń	1000
Luty	1200
Marzec	1300
Kwiecień	1250
Maj	1400
Czer-wiec	1500
Lipiec	1600

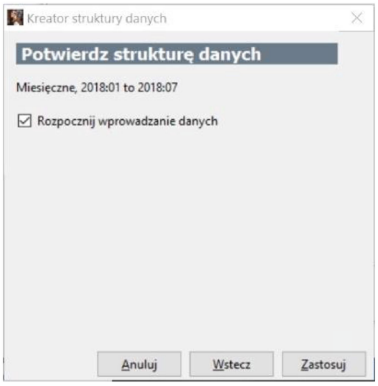
Wybieramy z Menu: Plik/Nowy zbiór danych,
liczba obserwacji: 7,
struktura: szeregi czasowe,
okresowość: miesięczne,
obserwacja startowa: 2018,
zaznacz „Rozpocznij wprowadzanie danych”.

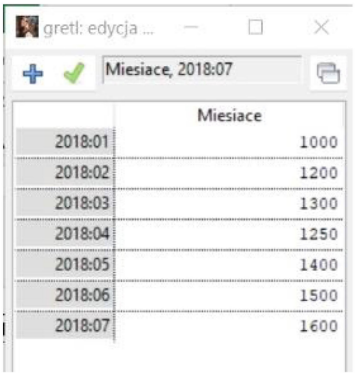












Każdą kolejną zmienną dodajemy, posługując się ikoną „+” (dodaj zmienną). Po zakończeniu potwierdzamy, klikając Zastosuj i zamykamy okno wprowadzania danych, klikając „x”.

Po zapisaniu zmian oraz po powrocie do głównego okna programu możemy zawsze wrócić do edytora danych przez Dane/Edycja wartości. Klikając w głównym oknie prawym przyciskiem myszy na zmiennej, możemy zobaczyć wartości zmiennej, jej statystyki opisowe itd., możemy edytować atrybuty zmiennej, a więc np. opisać ją lub zmienić jej nazwę.

Załącznik 3 – funkcje macierzowe EXCELA

W problemach opisanych w rozdziale drugim – Modele ekonometryczne korzystamy z funkcji macierzowych programu EXCEL. Załącznik ten zawiera krótki opis używanych funkcji tablicowych i macierzowych.

Transponowanie macierzy

A – dowolna macierz, zapisana w komórkach arkusza.

- lewym klawiszem myszy zaznaczamy komórki przeznaczone do zapisania wyniku,
- wpisujemy funkcję TRANSPONUJ(adres elementów macierzy A),
- wciskamy klawisze Shift-Ctrl-Enter (Enter na końcu).

Iloczyn macierzy AB

A, B – macierze odpowiedniego rozmiaru (liczba kolumn macierzy A jest równa liczbie wierszy macierzy B), zapisane w komórkach arkusza.

- lewym klawiszem myszy zaznaczamy komórki przeznaczone do zapisania wyniku,
- wpisujemy lub wstawiamy funkcję MACIERZ.ILOCZYN(adres elementów macierzy A; adres elementów macierzy B),
- wciskamy klawisze Shift-Ctrl-Enter (Enter na końcu).

Wyznacznik macierzy detA

A – macierz kwadratowa, zapisana w komórkach arkusza.

- lewym klawiszem myszy klikamy w komórkę przeznaczoną do zapisania wyniku,
- będąc w tej komórce, wpisujemy lub wstawiamy funkcję WYZNACZNIK.MACIERZY(adres elementów macierzy A),
- wciskamy klawisz Enter.

Wyznaczanie macierzy odwrotnej

A – macierz kwadratowa nieosobliwa, zapisana w komórkach arkusza.

- lewym klawiszem myszy zaznaczamy komórki przeznaczone do zapisania wyniku,
- wpisujemy lub wstawiamy funkcję MACIERZ.ODW(adres elementów macierzy A),
- wciskamy klawisze Shift-Ctrl-Enter (Enter na końcu).

Załącznik 4 – tablice rozkładu normalnego

Dystrybuanta standardowego rozkładu normalnego

X – zmienna losowa, f(x) – funkcja gęstości, F(x) – dystrybuanta

$X \sim N(0, 1)$, $\quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $\quad F(x) = \int_{-\infty}^{x} f(t)\,dt$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91308	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997

Załącznik 5 – tablice rozkładu t-Studenta

r	α													r
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001	
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	53,657	636,619	1
2	,142	,289	,445	,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	6,925	31,598	2
3	,137	,277	,424	,584	,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941	3
4	,134	,271	,414	,569	,741	,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610	4
5	,132	,267	,408	,559	,727	,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859	5
6	,131	,265	,404	,553	,718	,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959	6
7	,130	,263	,402	,549	,711	,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405	7
8	,130	,262	,399	,546	,706	,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041	8
9	,129	,261	,398	,543	,703	,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781	9
10	,129	,260	,397	,542	,700	,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587	10
11	,129	,260	,396	,540	,697	,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437	11
12	,128	,259	,395	,539	,695	,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318	12
13	,128	,259	,394	,538	,694	,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221	13
14	,128	,258	,393	,537	,692	,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140	14
15	,128	,258	,393	,536	,691	,866	1,074	1,341	1,753	2,181	2,602	2,947	4,073	15
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015	16
17	,128	,257	,392	,534	,689	,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965	17
18	,127	,257	,392	,534	,688	,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922	18
19	,127	,257	,391	,533	,688	,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883	19
20	,127	,257	,391	,533	,687	,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850	20
21	,127	,257	,391	,532	,686	,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819	21
22	,127	,256	,390	,532	,686	,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792	22
23	,127	,256	,390	,532	,685	,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767	23
24	,127	,256	,390	,531	,685	,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745	24
25	,127	,256	,390	,531	,684	,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725	25
26	,127	,256	,390	,531	,684	,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707	26
27	,127	,256	,389	,531	,684	,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690	27
28	,127	,256	,389	,530	,683	,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674	28
29	,127	,256	,389	,530	,683	,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659	29
30	,127	,256	,389	,530	,683	,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646	30
40	,126	,255	,388	,529	,681	,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551	40
60	,126	,254	,387	,527	,679	,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460	
120	,126	,254	,386	,526	,677	,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373	120
∞	,126	,253	,385	,524	,674	,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291	∞

Bibliografia

1. Barczak, A. St., Biolik, J. (1998). Podstawy ekonometrii. Katowice: Wydawnictwo Akademii Ekonomicznej w Katowicach.
2. Barczak, A. S. (2007). Ekonometria i nauki o zarządzaniu. *Prace Naukowe/Akademia Ekonomiczna w Katowicach*, s. 433-437.
3. Borkowski, B., Dudek, H., Szczesny, W. (2003). Ekonometria. Wybrane zagadnienia. Warszawa: Wydawnictwo Naukowe PWN.
4. Cieślak, M. (Ed.). (2008). Prognozowanie gospodarcze: metody i zastosowania. Warszawa: Wydawnictwo Naukowe PWN.
5. Dziechciarz, J. (2003). Ekonometria, Metody, przykłady, zadania. Wrocław: Wydawnictwo AE we Wrocławiu.
6. Górecki, B. R. (2010). Ekonometria. Podstawy teorii i praktyki. Warszawa: Key Text.
7. Gruszczyński, M., Podgórka M. (2000). Ekonometria. Warszawa: Oficyna Wydawnicza SGH.
8. Kopczewska, K. (2019). Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN. Warszawa: CeDeWu. Pl. Wydawnictwa Fachowe.
9. Kufel, T. (2013). Ekonometria: rozwiązywanie problemów z wykorzystaniem programu GRETL. Warszawa: Wydawnictwo Naukowe PWN.
10. Nowak, E. (2002). Zarys metod ekonometrii. Zbiór zadań. Warszawa: Wydawnictwo Naukowe PWN.
11. Nowak, E. (Ed.). (1998). Prognozowanie gospodarcze: metody, modele, zastosowania, przykłady: praca zbiorowa. Warszawa: Agencja Wydawnicza Placet.
12. Snarska, A. (2011). Statystyka, ekonometria, prognozowanie: ćwiczenia z Excelem 2007. Warszawa: Wydawnictwo Placet.
13. Welfe, A. (2018). Ekonometria. Warszawa: Polskie Wydawnictwo Ekonomiczne.

Grafika użyta do rozdzielania poszczególnych rozdziałów:

1. Ad. 1 – janoon028/Freepik.com
2. Ad. 2 – pressfoto/Freepik.com
3. Ad. 3 – rawpixel.com/Freepik.com
4. Ad. 4 – Mediamodifier/Pixabay.com

Z recenzji:

Ekonometria, podobnie jak statystyka, jest dziedziną nauki, która charakteryzuje się specyficznym językiem. (…) Innowacyjność oraz oryginalność podręcznika polega na zaprezentowaniu podstawowych narzędzi komputerowych, tj. Excela, Power Bi, Gretl oraz Statistica, wspomagających ekonometrię. Dzięki takiemu podejściu możliwe jest przeprowadzenie bardziej złożonych analiz i budowanie skomplikowanych modeli ekonometrycznych (…) z wykorzystaniem dużej liczby czytelnych grafik, obrazujących omawiane pojęcia.

dr Adam Kiersztyn

Rozwój gospodarki rynkowej w Polsce sprawia, iż metody ekonometryczne oraz związane z nimi prognozowanie i symulacje gospodarcze cieszą się coraz większym uznaniem i zapotrzebowaniem społecznym, znajdując liczne zastosowania w skali zarówno mikro-, jak i makrogospodarczej. Stanowi to istotną przesłankę do powstawania nowych opracowań, mających ułatwić percepcję niezbędnej wiedzy z zakresu ekonometrii. (…) Podręcznik łączy wykład współczesnej ekonometrii z tematyką badań operacyjnych i modelowaniem.

dr Anna Makarewicz

Wyższa Szkoła Ekonomii i Innowacji w Lublinie
20-209 Lublin, ul. Projektowa 4
tel.: +48 81 749 17 77
www.wsei.lublin.pl

ISBN – wersja elektroniczna: 978-83-66159-94-5

WSEI LUBLIN
WYDAWNICTWO NAUKOWE
INNOVATIO PRESS

KOMPETENCJE WIEDZA INNOWACJE
Zintegrowany program rozwoju WSEI III ETAP

Fundusze Europejskie Program Regionalny

Rzeczpospolita Polska

Unia Europejska Europejski Fundusz Społeczny