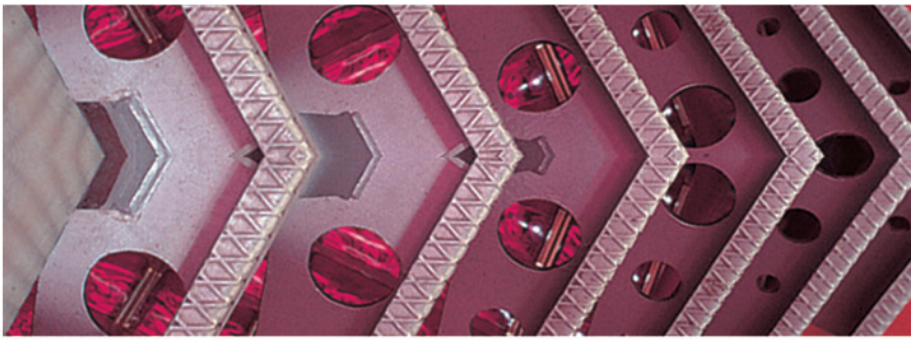
Podstawy ekonometrii i teorii prognozowania

Podręcznik z przykładami i zadaniami

wydanie III

Dorota Witkowska



Oficyna
a Wolters Kluwer business

Podstawy ekonometrii i teorii prognozowania

Podręcznik z przykładami i zadaniami

Dorota Witkowska

wydanie III

Warszawa 2012

Oficyna
a Wolters Kluwer business

Recenzent
Prof. dr hab. Bolesław Borkowski, Szkoła Główna Gospodarstwa Wiejskiego

Wydawca
Joanna Dzwonnik

Redaktor prowadzący
Janina Burek

Opracowanie redakcyjne
Bogumiła Gnypowa

Korekta
Robert Szymański
Joanna Hołdys

Korekta i łamanie
WYDAWNICTWO JAK
www.wydawnictwojak.pl

Projekt graficzny okładki
Barbara Widłak

Zdjęcie wykorzystane na okładce
Piotr Witosławski

Wydanie III

ISBN 978-83-264-3869-1

ISBN PDF-a: 978-83-264-4737-2

Wydane przez:
Wolters Kluwer Polska Sp. z o.o.

Redakcja Książek
01-231 Warszawa, ul. Płocka 5a
tel. 22 535 82 00, fax 22 535 81 35
e-mail: ksiazki@wolterskluwer.pl

www.wolterskluwer.pl
Księgarnia internetowa www.profinfo.pl

Spis treści

O autorce

Prof. zw. dr hab. Dorota Witkowska jest absolwentką Uniwersytetu Łódzkiego. Prowadzone przez nią od ponad 30 lat badania naukowe dotyczą zastosowań metod ilościowych w ekonomii, w tym metod optymalnego sterowania i sztucznych sieci neuronowych. Obecnie jej zainteresowania badawcze koncentrują się na analizach rynków finansowych, porównaniach międzynarodowych oraz tzw. *gender studies*. Opublikowała ponad 180 prac naukowych w kraju i za granicą, w tym w publikatorach z tzw. Listy Filadelfijskiej. Jest autorem, współautorem lub redaktorem 10 monografii i 9 podręczników akademickich. Dwukrotnie była stypendystką Fulbrighta: w Princeton University, gdzie pracowała pod kierunkiem prof. G.C. Chowa, i w Pennsylvania University, w którym prowadziła badania pod kierunkiem laureata Nagrody Nobla prof. L.R. Kleina. Jest członkiem International Atlantic Economic Society (Executive Committee w latach 2006–2009) i uczestniczy w radach naukowych krajowych i międzynarodowych czasopism w Polsce, USA, Portugalii, Kanadzie i Kolumbii. Wypromowała 6 doktorów oraz ponad 100 licencjatów, inżynierów i magistrów na Uniwersytecie Łódzkim, w Politechnice Łódzkiej, Szkole Głównej Gospodarstwa Wiejskiego w Warszawie i na Uniwersytecie Mikołaja Kopernika w Toruniu.

Słowo wstępne

Podręcznik jest przeznaczony dla studentów kierunków ekonomicznych i został przygotowany z myślą o tych, którym przedmioty ilościowe sprawiają pewne problemy. Zawiera omówienie podstawowych zagadnień z zakresu klasycznej ekonometrii, bogato zilustrowane przykładami.

Prezentowany materiał został podzielony na osiem rozdziałów. Rozdział 1 stanowi powtórzenie wiadomości ze statystyki – z zakresu wnioskowania statystycznego i analizy współzależności. W rozdziale 2 zdefiniowano podstawowe pojęcia związane z modelowaniem zjawisk gospodarczych. Omówiono takie zagadnienia, jak pomiar zjawisk gospodarczych oraz specyfikacja i metody doboru zmiennych do modelu, a także przeprowadzono klasyfikację modeli ekonometrycznych i występujących w nich zmiennych. W rozdziale 3 zaprezentowano klasyczną metodę najmniejszych kwadratów.

Rozdziały 4 i 5 zawierają omówienie podstawowych procedur związanych z weryfikacją modeli ekonometrycznych. Przedstawiono w nich najbardziej popularne mierniki i testy statystyczne, umożliwiające określenie jakości oszacowanego modelu oraz jego przydatności w praktyce.

Rozdział 6 stanowi wprowadzenie do analizy szeregów czasowych. Pokazano w nim metody dekompozycji oraz modele szeregów czasowych. Rozdział 7 został poświęcony podstawowym zagadnieniom z zakresu teorii prognozy. Rozdział 8 jest wprowadzeniem do problematyki modeli wielorównaniowych.

Każdy rozdział zamykają pytania kontrolne, a na końcu podręcznika zamieszczono zadania do samodzielnego rozwiązania. Uzupełnieniem podręcznika jest spis literatury cytowanej i zalecanej oraz tablice statystyczne.

Pragnę w tym miejscu podziękować Marioli Chrzanowskiej, Aleksandrze Matuszewskiej, Agnieszce Mazur, Iwonie Staniec, Annie Szmit i Zbigniewowi Szymczakowi za pomoc przy przygotowaniu zadań, które w większości pochodzą ze *Zbioru zadań ze statystyki* (red. D. Witkowska, seria Wydawnictw Dydaktycznych Wydziału Organizacji i Zarządzania Politechniki Łódzkiej, Menadżer, Łódź 2001).

1

Podstawy wnioskowania statystycznego i analizy korelacji

Celem analiz ekonometrycznych jest badanie zależności występujących w zjawiskach gospodarczych. Metody ekonometryczne wywodzą się ze statystycznej analizy współzależności zjawisk oraz metod wnioskowania statystycznego, ponieważ procesy gospodarcze są obserwowane (mierzone) w określonych punktach pomiarowych. Mamy zatem do czynienia z badaniem częściowym, co oznacza, że analizy są prowadzone na podstawie próby statystycznej zawierającej N obserwacji. Aby wnioski wypływające z badań można było uogólnić, zakłada się, że próba ta jest reprezentatywna.

cel, źródło i charakter analiz ekonometrycznych

Wykorzystanie metod wnioskowania statystycznego umożliwia oszacowanie (estymację) parametrów populacji generalnej na podstawie danych z próby. Innym rodzajem wnioskowania statystycznego jest weryfikacja hipotez statystycznych, która pozwala formułować wnioski dotyczące poprawności założeń, twierdzeń i opinii na podstawie wyników pochodzących z próby. Podstawową zaletą wnioskowania statystycznego jest możliwość określenia wielkości błędu, z jakim formułowane są wnioski ogólne.

W niniejszym rozdziale przypomnimy podstawowe zagadnienia z zakresu estymacji nieznanych parametrów, weryfikacji hipotez statystycznych i statystycznej analizy współzależności.

Estymacja nieznanych parametrów

Parametrem zbiorowości generalnej, oznaczanym jako θ, nazywa się miarę opisową, której wartość nie jest znana, na przykład średnią, odchylenie standardowe, parametry strukturalne modelu ekonometrycznego. *Estymatorem* nazywa się dowolną statystykę[1] T_N służącą

parametr zbiorowości generalnej

estymator

[1] *Statystyka* to miara opisowa pochodząca z N-elementowej próby losowej, na przykład średnia arytmetyczna, odchylenie standardowe, wskaźnik struktury.

rozkład estymatora

ocena estymatora

do oszacowania nieznanej wartości parametru w populacji generalnej θ. Inaczej mówiąc, estymatorem jest zmienna losowa, której rozkład zależy od rozkładu szacowanego parametru. Przez *rozkład estymatora* rozumie się rozkład prawdopodobieństwa zmiennej losowej T_N, którego parametrami są: wartość oczekiwana $E(T_N)$ oraz wariancja $D^2(T_N)$. Konkretną wartość liczbową t_N, jaką przyjmuje estymator T_N (parametru θ) dla określonej próby, nazywa się *oceną estymatora parametru* θ. Oznacza to, że wartość t_N będąca oceną estymatora parametru θ jest realizacją zmiennej losowej T_N.

Wartości parametrów zbiorowości generalnej są szacowane z pewnym błędem. Przeciętny błąd szacunku jest odchyleniem standardowym (czyli pierwiastkiem kwadratowym z wariancji) $D(T_N)$ rozkładu estymatora T_N, za pomocą którego szacuje się parametr θ populacji generalnej. Aby szacunek był możliwie precyzyjny, estymator musi mieć pewne własności – musi być zgodny, nieobciążony i efektywny.

estymator zgodny

Estymator T_N parametru θ nazywamy zgodnym, gdy jest on zbieżny według prawdopodobieństwa do szacowanego parametru θ, to znaczy gdy:

$$\bigwedge_{\varepsilon>0} \lim_{N\to\infty} P\left\{|T_N - \theta| > \varepsilon\right\} = 0 \tag{1.1}$$

lub inaczej:

$$\bigwedge_{\varepsilon>0} \lim_{N\to\infty} P\left\{|T_N - \theta| < \varepsilon\right\} = 1 \tag{1.2}$$

Z relacji (1.1) – (1.2) wynika, że w przypadku estymatora zgodnego, dla dostatecznie licznej próby, szansa otrzymania oceny estymatora różnego od parametru jest bliska zeru. Innymi słowy, dla estymatora zgodnego wzrost liczebności próby daje lepsze dopasowanie wyników estymacji do rzeczywistej wartości szacowanego parametru.

estymator nieobciążony

Estymator T_N nazywamy nieobciążonym (estymatorem parametru θ), jeśli jego wartość oczekiwana jest równa szacowanemu parametrowi, czyli jest spełniony warunek:

$$E(T_N) = \theta \tag{1.3}$$

Jeżeli $E(T_N) \neq \theta$, to estymator θ jest obciążony.

Estymator T_N nazywamy asymptotycznie nieobciążonym estymatorem parametru θ, jeśli jest spełniony warunek:

$$\lim_{N\to\infty} E(T_N) = \theta \tag{1.4}$$

Obciążeniem estymatora T_N nazywamy różnicę $o(T_N) = E(T_N) - \theta$.

estymator efektywny

Efektywność estymatora jest własnością względną, to znaczy rozpatruje się ją w odniesieniu do kilku statystyk, porównując ich wariancje. Najbardziej efektywnym estymatorem parametru θ nazywamy ten spośród nieobciążonych estymatorów, który ma najmniejszą wariancję $D^2(T_N)$. Stosowanie estymatora najbardziej efektywnego oznacza, że podczas estymacji popełnia się najmniejszy błąd.

Korzystanie z estymatora T_N zgodnego, nieobciążonego i najbardziej efektywnego pozwala najlepiej oszacować nieznany parametr θ, ponieważ z dużym prawdopodobieństwem można przyjąć, że wyznaczona ocena estymatora T_N jest bliska rzeczywistej wartości θ.

rodzaje estymacji

W literaturze przedmiotu i w praktyce wyróżnia się dwa rodzaje estymacji:

1) *estymację punktową*, czyli metodę szacunku, w której jako wartość parametru zbiorowości generalnej przyjmuje się konkretną wartość, to jest ocenę estymatora wyznaczonego na podstawie *N*-elementowej próby,
2) *estymację przedziałową*, za pomocą której wyznacza się przedział liczbowy, który z ustalonym prawdopodobieństwem pokrywa nieznaną wartość szacowanego parametru zbiorowości generalnej. Prawdopodobieństwo to nosi nazwę *współczynnika ufności* i jest oznaczane jako $(1 - \alpha)$, a znaleziony przedział jest nazywany przedziałem ufności. Innymi słowy, *przedział ufności* informuje, w jakich granicach należy spodziewać się wartości poszukiwanego parametru θ z zadanym z góry prawdopodobieństwem.

współczynnik ufności

przedział ufności

Ogólną postać przedziału ufności można zapisać jako:

$$P\left\{f_1(T_N) \le \theta \le f_2(T_N)\right\} = 1 - \alpha \tag{1.5}$$

gdzie:

- P – symbol prawdopodobieństwa,
- $1 - \alpha$ – współczynnik ufności,
- T_N – estymator,
- θ – szacowany parametr,
- $f_1(T_N)$ i $f_2(T_N)$ – dolna i górna granica przedziału ufności. Funkcje f_1 i f_2 jako funkcje zmiennej losowej T_N są zmiennymi losowymi.

Konstruując przedział ufności, można określić prawdopodobieństwo, z jakim oszacowany przedział zawiera w sobie wartość nieznanego parametru, czego nie uzyskuje się w przypadku estymacji punktowej. Należy przy tym pamiętać, że:

- przy zadanym poziomie ufności $1 - \alpha$, im większa jest liczebność próby, tym krótszy przedział ufności (tzn. zwiększa się precyzja szacunku),

współczynnik ufności a precyzja szacunku

- przy ustalonej liczebności próby wraz ze wzrostem poziomu ufności rośnie rozpiętość przedziału ufności (tzn. zmniejsza się precyzja szacunku).

Długość przedziału ufności jest oznaczana przez $2d$:

$$2d = f_2(T_N) - f_1(T_N) \tag{1.6}$$

Maksymalny błąd szacunku jest równy połowie tej długości, czyli d. W praktyce dużą rolę odgrywa zarówno dokładność oszacowań, jak i prawdopodobieństwo, z jakim wyznaczony przedział ufności będzie pokrywać wartość nieznanego parametru. Zazwyczaj obie wartości, to jest d oraz $1 - \alpha$, są z góry zadane (założone) na początku badania, a na ich podstawie określa się minimalną liczebność próby statystycznej.

etapy procesu estymacji przedziałowej

Proces estymacji przedziałowej nieznanych parametrów charakteryzujących badaną zbiorowość lub zjawisko można zatem podzielić na kilka etapów:

1) określenie parametru (np. średnia, wariancja, parametry modelu ekonometrycznego),
2) wybór estymatora i określenie jego rozkładu prawdopodobieństwa,
3) ustalenie współczynnika ufności oraz wartości maksymalnego błędu szacunku,
4) wyznaczenie minimalnej liczebności próby,
5) zebranie danych statystycznych,
6) wyznaczenie oceny estymatora,
7) budowa przedziału ufności.

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statystycznych ma na celu sprawdzenie sformułowanych hipotez statystycznych, czyli podjęcie określonych decyzji statystycznych. *Hipoteza statystyczna* to sąd (przypuszczenie, założenie) odnoszący się do nieznanego poziomu parametrów lub do nieznanej postaci rozkładu zmiennych losowych w zbiorowości generalnej. Hipotezy dzielimy na dwie grupy:

hipoteza statystyczna

1) *hipotezy parametryczne*, to znaczy sądy dotyczące wartości parametrów rozkładu zmiennej w populacji generalnej,
2) *hipotezy nieparametryczne*, to znaczy przypuszczenia dotyczące kształtu rozkładu populacji generalnej, a także założenia o niezależności zmiennych lub losowości.

hipotezy parametryczne i nieparametryczne

Hipotezy parametryczne i nieparametryczne można formułować zamiennie w zależności od celu prowadzonej analizy statystycznej. Celem weryfikacji hipotez parametrycznych jest sprecyzowanie wartości parametru rozkładu populacji generalnej znanego typu, celem zaś hipotez nieparametrycznych jest sprecyzowanie typu rozkładu zmiennej w określonej populacji generalnej, badanie występowania zależności między zmiennymi oraz weryfikacja założeń o losowości.

Procedura weryfikacji hipotez statystycznych zawsze opiera się na dwóch hipotezach – hipotezie zerowej i hipotezie alternatywnej. *Hipoteza zerowa* (H_0) jest podstawową hipotezą statystyczną, która jest przedmiotem weryfikacji. Proces weryfikacji może doprowadzić bądź do jej odrzucenia, bądź do stwierdzenia, że nie ma podstaw, by ją odrzucić. Formułując hipotezę zerową, zakłada się, że rozkład posiada pewną własność (np. wartość parametru jest równa określonej liczbie, rozkład zmiennej ma określony kształt, dwa rozkłady mają identyczne parametry, zmienne są losowe). *Hipoteza alternatywna* (H_1) to hipoteza konkurencyjna w stosunku do hipotezy zerowej. Jest ona formułowana jako przypuszczenie, że rozkład nie posiada własności określonej w hipotezie zerowej (tzn. posiada ją w innym wariancie). Jeśli odrzuca się hipotezę zerową, to przyjmuje się hipotezę alternatywną. Trzeba przy tym rozróżnić następujące sytuacje:

hipoteza zerowa

hipoteza alternatywna

- jeżeli hipoteza zerowa jest parametryczna, to hipotezy alternatywne można sformułować dwustronnie (tzn. wartość parametru Θ jest różna od ustalonej wartości, czyli $\Theta \neq \hat{\Theta}$, gdzie $\hat{\Theta}$ oznacza konkretną wartość parametru), prawostronnie (tzn. $\Theta > \hat{\Theta}$) lub lewostronnie (tzn. $\Theta < \hat{\Theta}$), przy czym sposób formułowania hipotez zależy nie tylko od celu badania, ale również od rodzaju informacji statystycznych uzyskanych z próby losowej,
- jeżeli hipoteza zerowa jest nieparametryczna, to hipotezy alternatywne są formułowane wyłącznie dwustronnie (tzn. przeciwnie do hipotezy zerowej) i znika problem odmiennego podejścia do weryfikacji hipotezy zerowej.

Hipotezy statystyczne weryfikuje się za pomocą testów statystycznych, przy czym w zależności od rodzaju hipotezy rozróżnia się testy parametryczne (służą do weryfikacji hipotez parametrycznych) i nieparametryczne (służą do

weryfikacji hipotez nieparametrycznych i są to m.in. testy zgodności i testy losowości).

Przy weryfikacji hipotez statystycznych ważne są tzw. poziomy istotności oraz tzw. obszary odrzucenia hipotez zerowych. Oba te pojęcia wiążą się z teorią błędów popełnianych przy podejmowaniu decyzji w procesach weryfikacji hipotez statystycznych. Ponieważ próba nigdy nie daje całkowitej informacji o zbiorowości, z której została pobrana, weryfikując hipotezę zerową, możemy podjąć decyzję poprawną lub popełnić błąd. Rozróżniamy dwa rodzaje błędów, które występują z określonym prawdopodobieństwem:

błąd pierwszego i drugiego rodzaju

1) *błąd pierwszego rodzaju* polega na odrzuceniu hipotezy zerowej wtedy, gdy jest ona w rzeczywistości prawdziwa; jego prawdopodobieństwo będziemy oznaczać przez α,
2) *błąd drugiego rodzaju* polega na przyjęciu hipotezy zerowej wtedy, gdy jest ona w rzeczywistości fałszywa, czyli jeśli jest prawdziwa hipoteza alternatywna; jego prawdopodobieństwo zazwyczaj oznacza się przez β.

Tabela 1.1. Błędy popełniane przy weryfikacji hipotez statystycznych

Decyzja	Hipoteza H_0 jest prawdziwa	Hipoteza H_0 jest fałszywa
Przyjąć weryfikowaną hipotezę H_0	decyzja poprawna	decyzja błędna – błąd drugiego rodzaju
Odrzucić weryfikowaną hipotezę H_0	decyzja błędna – błąd pierwszego rodzaju	decyzja poprawna

Źródło: W. Krysicki, J. Bartos, W. Dyczka, K. Królikowska, M. Wasilewski, *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*, cz. I i II, PWN, Warszawa 1986, s. 79.

poziom istotności

Prawdopodobieństwo popełnienia błędu pierwszego rodzaju (α) jest ustalone z góry jako dowolnie małe prawdopodobieństwo odrzucenia prawdziwej hipotezy zerowej i nazwane *poziomem istotności*. W naukach ekonomicznych prawdopodobieństwo popełnienia błędu pierwszego rodzaju przyjmuje się zazwyczaj z przedziału $<0{,}001; 0{,}1>$, przy czym najczęściej $\alpha = 0{,}05$.

moc testu

Prawdopodobieństwo popełnienia błędu drugiego rodzaju (β) jest wykorzystywane do wyznaczenia tzw. mocy testu. *Moc testu* to prawdopodobieństwo podjęcia poprawnej decyzji, polegającej na odrzuceniu weryfikowanej hipotezy wtedy, gdy jest ona fałszywa, czyli:

$$M = 1 - \beta \tag{1.7}$$

Test statystyczny[2] to reguła postępowania, która na podstawie wyników z próby ma doprowadzić do odrzucenia lub nieodrzucenia postawionej hipotezy statystycznej. Najczęściej używane w praktyce są tzw. *testy istotności*, które umożliwiają odrzucenie hipotezy z ryzykiem równym poziomowi istotności α, bez uwzględniania prawdopodobieństwa popełnienia błędu drugiego rodzaju. Stosując testy istotności, podejmuje się decyzję odnośnie do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej albo decyzję o braku podstaw do odrzucenia hipotezy zerowej, co nie jest równoznaczne z jej przyjęciem, aczkolwiek w praktyce może się do tego sprowadzać. W związku z tym w wypadku testów istotności hipotezę zerową staramy się sformułować w taki sposób (czasem wbrew zdrowemu rozsądkowi), aby można ją było stosunkowo łatwo odrzucić.

testy istotności

Sprawdzian testu (statystyka testu) to zmienna losowa o określonym rozkładzie. Rozkład statystyki testu zależy od przyjętych założeń oraz wiedzy badacza na temat populacji generalnej, od liczebności posiadanej próby statystycznej oraz sformułowanej hipotezy zerowej.

sprawdzian (statystyka) testu

Wartość krytyczna testu to wartość zmiennej losowej o określonym rozkładzie, która przy danym prawdopodobieństwie α stanowi granicę obszaru odrzucenia.

wartość krytyczna testu

Podczas weryfikacji hipotez statystycznych wyznacza się z próby wartość sprawdzianu testu, którą porównuje się z wartościami krytycznymi. Jeżeli wyznaczona wartość sprawdzianu testu znajdzie się w obszarze odrzucenia, to odrzuca się hipotezę zerową i przyjmuje hipotezę alternatywną. W przeciwnym razie nie ma podstaw do odrzucenia hipotezy zerowej.

Wyznaczenie obszaru odrzucenia jest ściśle związane z:

- rodzajem postawionej hipotezy alternatywnej – w przypadku hipotez parametrycznych obszar jednostronny wyznacza się, gdy H_1 jest stawiana jednostronnie, a obszar obustronny – gdy H_1 jest stawiana dwustronnie,
- rozkładem statystyki będącej sprawdzianem testu, na przykład rozkładem normalnym, t-Studenta, χ^2 (chi kwadrat), Fishera-Snedecora, liczby serii,
- poziomem istotności α,
- liczebnością próby i liczbą szacowanych parametrów (różnica dwóch ostatnich określa liczbę *stopni swobody*), które wpływają na wartość krytyczną testu.

obliczanie stopni swobody

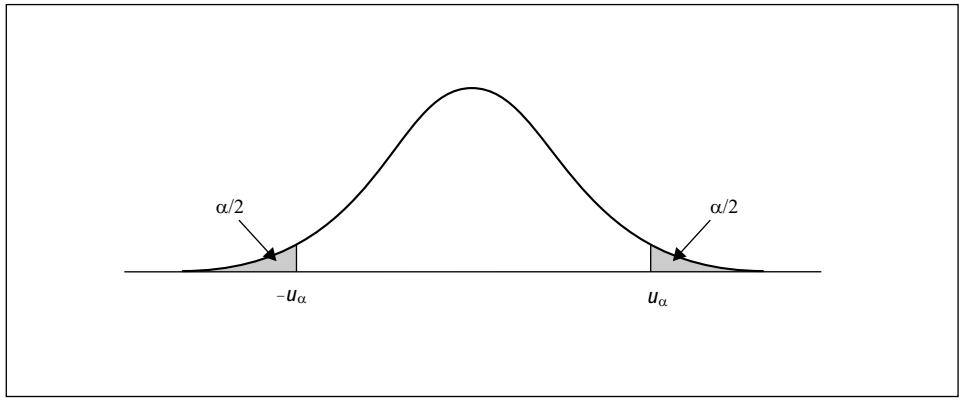
[2] Przedstawiony sposób postępowania przy weryfikacji hipotez statystycznych dotyczy tzw. testów istotności. Opierają się one na założeniu, że jeżeli H_0 jest prawdziwa, to wynik testu wyznaczony na podstawie danych z próby nie powinien istotnie odbiegać od założonego.

Przykłady obszarów odrzucenia dla sprawdzianu testu będącego zmienną losową o rozkładzie normalnym przedstawiono na ilustracjach 1.1, 1.2 i 1.3.

dwustronny obszar odrzucenia

Obustronny (dwustronny) obszar odrzucenia (ilustracja 1.1) to zbiór wszystkich wartości u zmiennej losowej U takich, że $|u| \geq u_\alpha$, gdzie u_α to wartość krytyczna odczytana z tablic rozkładu normalnego dla ustalonego z góry poziomu istotności α, taka, że $P(|u| \geq u_\alpha) = \alpha$.

Ilustracja 1.1. Obustronny obszar odrzucenia skonstruowany na podstawie standaryzowanego rozkładu normalnego dla poziomu istotności α



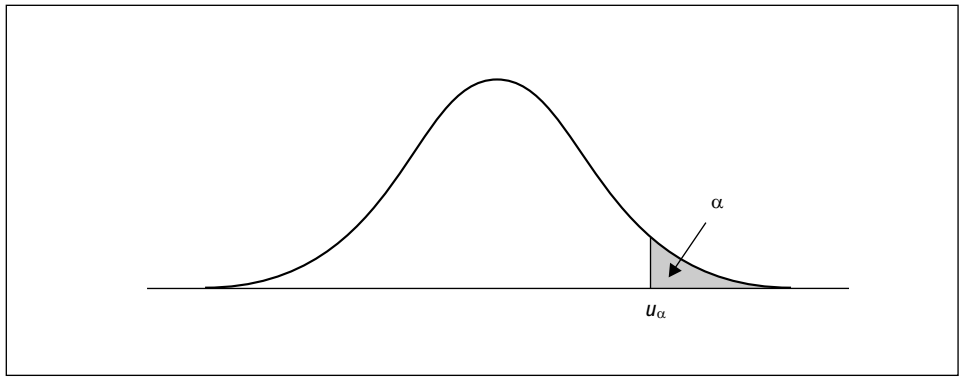
jednostronny obszar odrzucenia

Jednostronny obszar odrzucenia to zbiór wszystkich wartości u zmiennej losowej U takich, że $u \geq u_\alpha$ (mamy wtedy do czynienia z tzw. prawostronnym obszarem odrzucenia, ilustracja 1.2) lub $u \leq -u_\alpha$ (mamy wtedy do czynienia z tzw. lewostronnym obszarem odrzucenia, ilustracja 1.3), gdzie u_α to wartość krytyczna dla z góry zadanego poziomu istotności α, taka, że $P(u \geq u_\alpha) = \alpha$ dla prawostronnego lub $P(u \leq -u_\alpha) = \alpha$ dla lewostronnego obszaru odrzucenia.

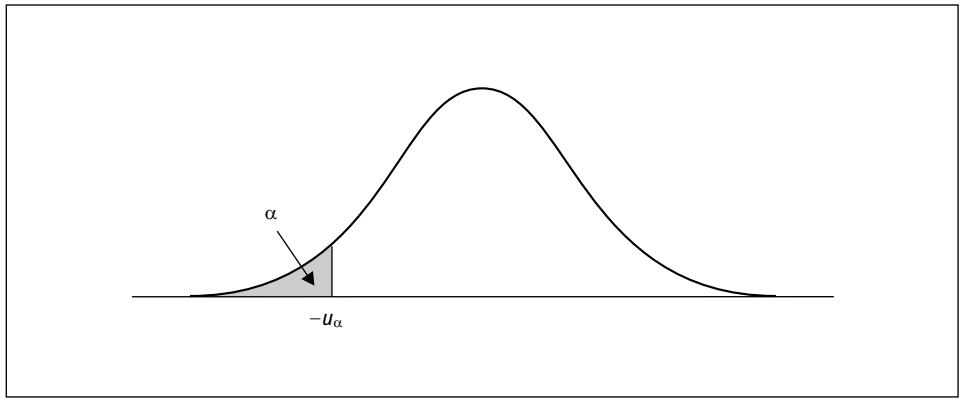
odczytywanie wartości krytycznych z tablic statystycznych

Zauważmy, że odczytywanie wartości krytycznych z tablic statystycznych zależy od sposobu ich konstrukcji. W przypadku standaryzowanego rozkładu normalnego (zob. tablica 2 na końcu książki) wewnątrz tablicy mamy podane wartości dystrybuanty, czyli $P(u \leq u_\alpha) = 1 - \alpha$. Zatem wartość krytyczną dla prawostronnego obszaru odrzucenia odczytujemy bezpośrednio w boczku i w główce tablicy. Przykładowo, dla poziomu istotności $\alpha = 0{,}05$ odczytujemy wartość u_α dla dystrybuanty (skumulowanego prawdopodobieństwa) $P = 1 - \alpha = 0{,}95$, stąd $u_\alpha = 1{,}64$. Dla lewostronnego obszaru

Ilustracja 1.2. Prawostronny obszar odrzucenia skonstruowany na podstawie standaryzowanego rozkładu normalnego dla poziomu istotności α



Ilustracja 1.3. Lewostronny obszar odrzucenia skonstruowany na podstawie standaryzowanego rozkładu normalnego dla poziomu istotności α



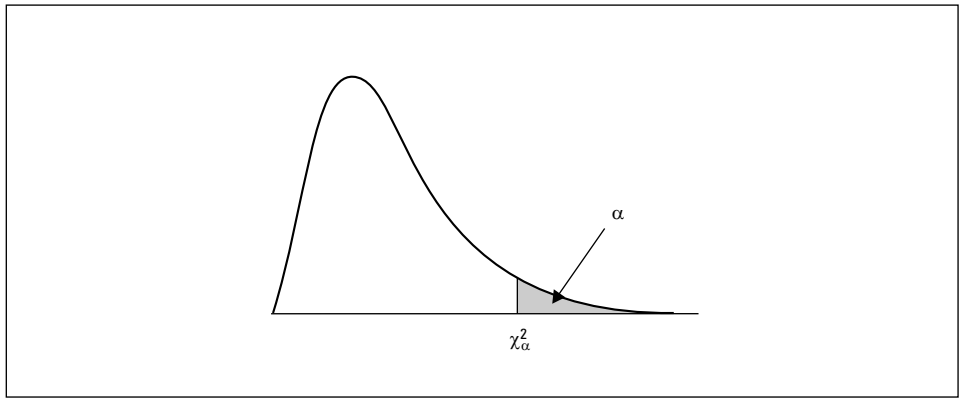
odrzucenia mamy $-u_\alpha = -1{,}64$, ponieważ rozkład normalny jest rozkładem symetrycznym. Z kolei dla obustronnego obszaru odrzucenia obie wartości krytyczne odczytujemy z tablic dla dystrybuanty rozkładu normalnego równej $P = 1 - \alpha/2 = 0{,}975$, stąd $-u_\alpha = -1{,}96$ oraz $u_\alpha = 1{,}96$.

W przypadku rozkładu t-Studenta (zob. tablica 1) tablice statystyczne zawierają wartości krytyczne $P(|t| \le t_{\alpha,k}) = \alpha$. Zatem dla konkretnego poziomu istotności α (w główce tablicy) oraz k stopni swobody (w boczku tablicy) wartość krytyczną t_α odczytuje się wewnątrz tablicy. Odpowiada to sytuacji, w której mamy do czynienia z obustronnym obszarem odrzucenia. Przykładowo, dla poziomu istotności $\alpha = 0{,}05$ i $k = 20$ stopni swobody odczytujemy

wartości $t_\alpha = 2{,}086$ i $-t_\alpha = -2{,}086$. Dla prawostronnego obszaru odrzucenia wartość krytyczną odczytujemy dla $2\alpha = 0{,}1$ i $k = 20$, zatem mamy $t_\alpha = 1{,}725$, a dla lewostronnego obszaru odrzucenia mamy $-t_\alpha = -1{,}725$, ponieważ rozkład t-Studenta jest rozkładem symetrycznym.

Rozkład χ^2 jest rozkładem asymetrycznym. Tablice statystyczne (zob. tablica 3) są tak skonstruowane, że w boczku podana jest liczba stopni swobody k, a w główce – prawdopodobieństwo α. Wewnątrz tablicy są podane wartości krytyczne $P(\chi^2 \le \chi^2_\alpha) = \alpha$. Przykładowo, dla poziomu istotności $\alpha = 0{,}05$ oraz $k = 10$ stopni swobody odczytana z tablic wartość krytyczna wynosi $\chi^2_\alpha = 18{,}3070$.

Ilustracja 1.4. Obszar odrzucenia w teście χ^2 dla poziomu istotności α



Podobny kształt jak przedstawiony na ilustracji 1.4 ma rozkład Fishera-Snedecora[3]. Tablice statystyczne (zob. tablice 4–7) są opracowane dla różnych wartości prawdopodobieństwa α, przy czym w boczku i główce tablic występują dwie liczby oznaczające stopnie swobody. Wewnątrz tablicy odczytuje się wartości krytyczne dla $P(F \ge F_\alpha) = \alpha$. Przykładowo, dla $\alpha = 0{,}05$ oraz $r_1 = 20$ i $r_2 = 10$ otrzymujemy wartość krytyczną $F_\alpha = 2{,}77$.

etapy procesu weryfikacji hipotez statystycznych

Reasumując, w procesie weryfikacji hipotez statystycznych można wyróżnić kilka etapów:

1) sformułowanie hipotezy zerowej (H_0) oraz hipotezy alternatywnej (H_1),

[3] Statystyka Fishera-Snedecora nazywana jest również, zwłaszcza w literaturze anglojęzycznej, statystyką Fishera, dlatego na oznaczenie zmiennej losowej o tym rozkładzie używa się najczęściej symbolu F, chociaż można również spotkać oznaczenie F-S.

2) wybór testu statystycznego służącego do weryfikacji hipotezy zerowej,
3) wyznaczenie wartości sprawdzianu testu,
4) ustalenie poziomu istotności α oraz wyznaczenie obszaru odrzucenia hipotezy zerowej,
5) podjęcie decyzji dotyczącej określonego prawdopodobieństwa popełnienia błędu pierwszego rodzaju.

Jeśli weryfikujemy hipotezy statystyczne za pomocą programów komputerowych, informację o wyniku testu często uzyskujemy nie w postaci wartości sprawdzianu testu, ale w postaci wartości minimalnego prawdopodobieństwa popełnienia błędu pierwszego rodzaju (p), który pozwala na odrzucenie hipotezy zerowej. Przykładowo, informacja $p = 0{,}01$ oznacza, że hipotezę zerową można odrzucić już dla prawdopodobieństwa popełnienia błędu pierwszego rodzaju $\alpha = 0{,}01$. Znaczy to, że przy założonym poziomie istotności większym od tej wartości (np. $\alpha = 0{,}05$) hipotezę zerową odrzucimy. Wartość prawdopodobieństwa $p = 0{,}1$ oznacza, że na przykład dla założonego poziomu istotności $\alpha = 0{,}05$ nie ma podstaw do odrzucenia hipotezy zerowej, można ją bowiem odrzucić dopiero dla $\alpha = 0{,}1$. Innymi słowy, hipotezę zerową można odrzucić, jeżeli poziom prawdopodobieństwa (popełnienia błędu pierwszego rodzaju) p nie jest większy od przyjętego poziomu istotności α, czyli dla $p \leq \alpha$.

Badanie współzależności zjawisk

Procesy gospodarcze są na ogół powiązane ze sobą. Przy badaniu współzależności dwóch zjawisk określa się zwykle jedną z cech (x) jako cechę niezależną, to znaczy taką, której zmienność jest uwarunkowana czynnikami zewnętrznymi, drugą (y) zaś traktuje się jako cechę zależną, to znaczy jej wahania próbuje się wyjaśnić (przynajmniej częściowo) zmiennością cechy niezależnej.

Zależności występujące między zjawiskami mogą być dwojakiego rodzaju:

1) *funkcyjne* (dokładne), kiedy zmiana wartości jednej zmiennej pociąga za sobą konkretną zmianę drugiej, czyli $y = f(x)$, gdzie x, y – wartości zmiennych, f – symbol pewnej funkcji,

zależności funkcyjne (dokładne)

2) *stochastyczne* (probabilistyczne), gdy różnym wariantom jednego zjawiska odpowiadają różne rozkłady drugiego. Jeśli zaś dla różnych wariantów jednej zmiennej (lub cechy) druga ma stale taki sam rozkład, to zmienne są niezależne stochastycznie.

zależności stochastyczne (probabilistyczne)

zależności korelacyjne (statystyczne)

Szczególnym przypadkiem zależności stochastycznej jest *zależność korelacyjna* (statystyczna). Mówimy, że dwie zmienne są zależne korelacyjnie, jeżeli różnym wariantom jednej zmiennej odpowiadają różne średnie warunkowe drugiej.

W naukach społecznych najczęściej spotyka się zależności stochastyczne, do których badania wykorzystuje się analizę korelacji i regresji.

Miary korelacji

Najczęściej stosowane wskaźniki mierzące siłę korelacji cech to: współczynnik korelacji liniowej Pearsona (nazywany również współczynnikiem korelacji prostoliniowej), stosunek korelacji (zwany także współczynnikiem korelacji nieliniowej) i współczynnik korelacji rang (korelacji kolejnościowej) Spearmana.

W przypadku występowania cech ilościowych najbardziej popularnym miernikiem współzależności zmiennych jest współczynnik korelacji liniowej Pearsona. Miara ta jest oparta na pojęciu kowariancji. Dla szeregu prostego (szczegółowego) *kowariancję* wyznacza się według wzoru:

wyznaczanie kowariancji

$$\mathrm{cov}_{xy} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x})(y_n-\bar{y}) \tag{1.8}$$

lub równoważnie:

$$\mathrm{cov}_{xy} = \frac{1}{N}\sum_{n=1}^{N}x_n y_n - \bar{x}\,\bar{y} \tag{1.9}$$

gdzie dla każdego n:

x_n, y_n – obserwacje zmiennych x i y,
$\bar{x}, \bar{y}$ – wartości średnich arytmetycznych wyznaczonych dla zmiennych x i y,
N – liczebność próby statystycznej.

W wypadku szeregu pogrupowanego w tablicy korelacyjnej, w której wyróżniono k wariantów cechy x oraz l wariantów cechy y, ujętych w przedziałach klasowych, kowariancję wyznacza się z relacji:

$$\mathrm{cov}_{xy} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{l}(\dot{x}_i-\bar{x})(\dot{y}_j-\bar{y})N_{ij} \tag{1.10}$$

lub

$$\mathrm{cov}_{xy} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{l}\dot{x}_i\dot{y}_j N_{ij} - \bar{x}\,\bar{y} \tag{1.11}$$

gdzie:

$\dot{x}_i, \dot{y}_j$ – środki przedziałów klasowych wielojednostkowych wyznaczonych dla poszczególnych wariantów zmiennych x_i ($i = 1, 2, ..., k$) oraz y_j ($j = 1, 2, ..., l$); zauważmy, że w przypadku przedziałów klasowych jednojednostkowych środki przedziałów są równoznaczne wyróżnionym wariantom cech, czyli: $\dot{x}_i = x_i$ oraz $\dot{y}_j = y_j$,

N_{ij} – liczba obserwacji charakteryzujących się i-tym wariantem zmiennej x oraz j-tym wariantem zmiennej y; zauważmy, że $N = \sum_{i=1}^{k}\sum_{j=1}^{l} N_{ij}$.

Współczynnik korelacji liniowej Pearsona dany jest wzorem:

$$r_{xy} = r_{yx} = \frac{\text{cov}_{xy}}{S_x S_y} \tag{1.12}$$

gdzie: S_x, S_y – odchylenia standardowe zmiennych x i y.

współczynnik korelacji liniowej Pearsona (współczynnik korelacji prostoliniowej)

Współczynnik korelacji liniowej jest miarą symetryczną i przyjmuje wartości z przedziału <–1, 1>, które informują o sile oraz kierunku korelacji liniowej między zmiennymi. Wartość $r_{xy} = 0$ świadczy o braku korelacji liniowej między badanymi zmiennymi (możliwe, że istnieje między nimi korelacja krzywoliniowa). Jeśli $r_{xy} \neq 0$, to cechy x i y są ze sobą skorelowane. Wartość $r_{xy} > 0$ informuje, że mamy do czynienia z korelacją dodatnią, to znaczy, że wraz ze wzrostem wartości jednej zmiennej wzrasta średnia warunkowa drugiej; dla $r_{xy} < 0$ korelacja jest ujemna, czyli wzrostowi wartości jednej zmiennej towarzyszy spadek średniej warunkowej drugiej. Przy $r_{xy} = 1$ lub -1 występuje liniowa zależność funkcyjna.

Kwadrat współczynnika korelacji liniowej r_{xy}^2 nazywa się współczynnikiem determinacji liniowej i jest oznaczany przez R^2. Informuje on, jaka część zmienności zmiennej zależnej jest wyjaśniona zmiennością zmiennej niezależnej.

stosunek korelacji (współczynnik korelacji nieliniowej)

Inną miarą korelacji jest *stosunek korelacji* (wskaźnik siły korelacji) Pearsona e_{yx} i e_{xy}, nazywany też współczynnikiem korelacji nieliniowej, ponieważ mierzy siłę korelacji zmiennych niezależnie od kształtu tej zależności. Wyznacza się go dla danych pogrupowanych w tablicy korelacyjnej na podstawie wzoru:

$$e_{yx} = \sqrt{\frac{S_{\bar{y}_i}^2}{S_y^2}} \tag{1.13}$$

gdzie:

S_y^2 – wariancja zmiennej y,

$S^2_{\bar{y}_i}$ – wariancja średnich warunkowych zmiennej y, obliczonych dla i-tego ($i = 1, 2, ..., k$) wariantu zmiennej x, czyli:

$$S^2_{\bar{y}_i} = \frac{1}{N}\sum_{i=1}^{k}(\bar{y}_i - \bar{y})^2 N_{i\bullet} \tag{1.14}$$

gdzie:

$N_{i\bullet}$ – liczebności brzegowe wyznaczone dla i-tego wariantu zmiennej x, czyli $N_{i\bullet} = \sum_{j=1}^{l} N_{ij}$

$\bar{y}_i$ – średnia warunkowa[4] zmiennej y wyznaczona dla i-tego wariantu zmiennej x, czyli:

$$\bar{y}_i = \frac{\sum_{j=1}^{l} y_j N_{ij}}{N_{i\bullet}}, \quad i = 1, 2, ..., k \tag{1.15}$$

W podobny sposób oblicza się stosunek korelacji e_{xy} zmiennej x do zmiennej y.

Stosunek korelacji może przyjmować wartości z przedziału <0, 1>. Na ogół jest to miara niesymetryczna, to znaczy

$$e_{xy} \neq e_{yx} \tag{1.16}$$

poza dwoma wypadkami:

1) braku korelacji, wtedy $e_{xy} = e_{yx} = 0$,
2) występowania zależności funkcyjnej między zmiennymi y i x, wówczas $e_{xy} = e_{yx} = 1$.

Między stosunkiem korelacji e_{xy} i współczynnikiem korelacji Pearsona r_{xy} występuje następująca zależność:

$$|r_{xy}| \leq e_{xy} \tag{1.17}$$

Stosunek korelacji może być wyznaczany również wtedy, gdy cecha niezależna jest niemierzalna (jakościowa).

Inną miarą korelacji, wygodną i użyteczną dla niezbyt długich szeregów szczegółowych z dwiema cechami mierzalnymi lub jakościowymi posiadającymi pewien naturalny porządek pozwalający na ustawienie wartości rosnąco

---

[4] Zauważmy, że średnia zmiennej y dana jest w postaci: $\bar{y} = \frac{\sum_{j=1}^{l} y_j N_{\bullet j}}{N} = \frac{\sum_{j=1}^{l}\sum_{i=1}^{k} y_j N_{ij}}{N}$, gdzie $N_{\bullet j}$ to liczebności brzegowe wyznaczone dla j-tego wariantu zmiennej y, czyli: $N_{\bullet j} = \sum_{i=1}^{k} N_{ij}$.

lub malejąco, jest *współczynnik korelacji rang* (korelacji kolejnościowej) *Spearmana* R_{xy}:

współczynnik korelacji rang (korelacji kolejnościowej) Spearmana

$$R_{xy} = R_{yx} = 1 - \frac{6\sum_{n=1}^{N} d_n^2}{N^3 - N} \tag{1.18}$$

gdzie d_n – różnice między kolejnymi numerami (rangami) nadawanymi w kolejności niemalejącej (lub nierosnącej) osobno dla każdej cechy od 1 do N. Jeżeli kilka elementów w szeregu ma taką samą wartość jednej cechy, to nadaje się im rangi będące średnią arytmetyczną rang przypadających na te elementy.

Współczynnik korelacji rang jest miarą symetryczną, wartość R_{xy} należy do przedziału <–1, 1> i mówi o sile oraz kierunku korelacji.

Funkcja regresji

Zależności występujące między dwiema cechami ilościowymi można również mierzyć za pomocą *funkcji regresji*:

funkcja regresji

$$y_n = f(x_n) + \varepsilon_n \tag{1.19}$$

która dla funkcji liniowej f ma postać:

$$y_n = \beta_0 + \beta_1 x_n + \varepsilon_n \tag{1.20}$$

gdzie dla każdego n:
- y_n – obserwacje zmiennej zależnej,
- x_n – obserwacje zmiennej niezależnej,
- ε_n – składnik losowy,
- β_0, β_1 – nieznane parametry funkcji regresji, które należy oszacować.

Do estymacji parametrów liniowej funkcji regresji najczęściej wykorzystuje się metodę najmniejszych kwadratów (MNK), która zostanie omówiona w rozdziale 3. Po oszacowaniu parametrów modelu regresji (1.20) otrzymuje się równanie:

$$\hat{y}_n = b_0 + b_1 x_n \tag{1.21}$$

gdzie dla każdego n:
- $\hat{y}_n$ – wartości teoretyczne zmiennej zależnej, wyznaczone na podstawie równania (1.21),
- x_n – obserwacje zmiennej niezależnej,
- b_0, b_1 – oceny estymatorów parametrów funkcji regresji.

Oszacowany parametr liniowego modelu regresji – b_1 informuje, o ile przeciętnie zmieni się wartość zmiennej zależnej, jeżeli wartość zmiennej niezależnej wzrośnie o jednostkę. Wyraz wolny można interpretować[5] jako średnią wartość zmiennej zależnej, przy zerowych wartościach zmiennej x.

Czasem interesujące może być również badanie zależności odwrotnej, czyli przyjęcie, że zmienna x jest zmienną zależną, a y – niezależną[6]. Wówczas funkcja regresji ma postać:

$$x_n = \alpha_0 + \alpha_1 y_n + \eta_n \tag{1.22}$$

gdzie dla każdego n:

x_n – obserwacje zmiennej zależnej,
y_n – obserwacje zmiennej niezależnej,
η_n – składnik losowy,
α_0, α_1 – nieznane parametry funkcji regresji (1.22).

Po oszacowaniu parametrów równania (1.22) otrzymamy:

$$\hat{x}_n = a_0 + a_1 y_n \tag{1.23}$$

gdzie dla każdego n:

$\hat{x}_n$ – wartości teoretyczne zmiennej zależnej, wyznaczone na podstawie równania (1.23),
y_n – obserwacje zmiennej niezależnej,
a_0, a_1 – oceny estymatorów parametrów funkcji regresji.

Iloczyn współczynników kierunkowych równań regresji (1.21) i (1.23) określa wartość współczynnika determinacji liniowej R^2. Oznacza to, że współczynnik korelacji liniowej Pearsona (1.12) dla analizowanej pary zmiennych x i y można wyznaczyć z następującej relacji:

$$r_{yx} = r_{xy} = \sqrt{R^2} = \sqrt{a_1 b_1} \tag{1.24}$$

Modele regresji mogą być również wykorzystywane do badania relacji występujących między zmienną zależną i k zmiennymi niezależnymi:

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + ... + \beta_k x_{nk} + \varepsilon_n \tag{1.25}$$

gdzie dla każdego n:

y_n – obserwacje zmiennej zależnej,

---

[5] Często nie interpretuje się wyrazu wolnego.

[6] Zauważmy, że ze statystycznego punktu widzenia, jeżeli zmienna y zależy od zmiennej x, to istnieje zależność odwrotna. Jednakże z merytorycznego punktu widzenia może się okazać, że relacja (1.22) nie jest poprawna, na przykład wtedy, gdy zależność (1.21) opisuje związki przyczynowo-skutkowe, czyli zmienna x jest traktowana jako przyczyna zjawiska y.

$x_{n1}, x_{n2}, ..., x_{nk}$ – obserwacje poczynione dla k zmiennych niezależnych x_1, x_2, ..., x_k,
ε_n – składnik losowy,
$\beta_0, \beta_1, \beta_2, ..., \beta_k$ – nieznane parametry funkcji regresji,
n – numer obserwacji ($n = 1, 2, ..., N$).

Mówimy wtedy o *regresji wielorakiej*. Modele regresji postaci (1.25), budowane w celu opisu i wyjaśnienia kształtowania się zjawisk ekonomicznych, legły u podstaw modelowania ekonometrycznego – dały podstawy klasycznej ekonometrii.

modele ekonometryczne – definicja i podział

Modele ekonometryczne są pojedynczymi równaniami lub układami równań. Wśród modeli ekonometrycznych wyróżnia się (por. Gajda 1996, s. 13):

- modele opisowe, kiedy poszczególne równania modelu traktuje się jako opis praw ekonomicznych działających w modelowanym układzie gospodarczym,
- modele strukturalne (modele przyczynowo-skutkowe), kiedy równania modelu mają odzwierciedlać strukturę przyczynowo-skutkową związków występujących między zmiennymi,
- modele (funkcje, równania) regresji, kiedy równania modelu mają podkreślać zastosowania statystycznej koncepcji szacowania parametrów modelu.

Te trzy określenia są pokrewne i bywają używane zamiennie.

Pytania kontrolne

1. Zdefiniuj następujące pojęcia:
   - estymator,
   - przedział ufności,
   - hipoteza statystyczna,
   - błąd pierwszego rodzaju,
   - sprawdzian testu,
   - obszar odrzucenia,
   - wartość krytyczna testu.
2. Jak się oblicza liczbę stopni swobody?
3. Czym różni się zależność funkcyjna od zależności korelacyjnej?
4. Wymień podstawowe miary współzależności cech.
5. Sformułuj funkcję regresji liniowej.

2

Wprowadzenie do modelowania ekonometrycznego

Ekonometria to nauka zajmująca się ustalaniem, za pomocą metod statystycznych, ilościowych prawidłowości zachodzących w życiu gospodarczym. Według G.C. Chowa (1995, s. 15), ekonometria „jest nauką i sztuką stosowania metod statystycznych do mierzenia relacji ekonomicznych". Metody ekonometryczne wykorzystuje się do estymacji parametrów modelu i weryfikacji postawionych hipotez badawczych. Stosuje się je również w praktyce zarządzania, na przykład do konstruowania prognoz.

ekonometria

Model ekonometryczny to formalny zapis prawidłowości (zależności) ekonomicznych występujących w rzeczywistości[1]. W literaturze przedmiotu (por. Gajda 1996, s. 23) wyróżnia się zależności:

model ekonometryczny

- przyczynowe, to znaczy wynikające ze związku przyczynowego, którego istnienie potrafimy wyjaśnić na gruncie teorii,
- symptomatyczne, kiedy mimo że nie znamy teorii potwierdzającej istnienie związku przyczynowego między zmiennymi, możemy znaleźć wspólną przyczynę wpływającą na zmienne, to jest zarówno na zmienną przez model wyjaśnianą, jak i zmienne służące do jej opisu,
- pozorne, kiedy nie umiemy znaleźć przyczyn wpływających na podobne zachowanie się zmiennych.

[1] W szerszym znaczeniu modele ekonometryczne to również takie modele, które pozwalają na wybór optymalnych rozwiązań. W modelach tego rodzaju oprócz zależności funkcyjnych występuje funkcja celu (kryterium), która pozwala wybrać rozwiązanie optymalne spośród rozwiązań dopuszczalnych. W węższym sensie pojęcie modelu ekonometrycznego odnosi się jedynie do modeli klasycznej ekonometrii, czyli takich, które opisują zależności istniejące w gospodarce. Nazywa się je modelami opisowymi (por. *Opisowe modele ekonometryczne*... 1992, s. 7).

Etapy budowy modelu ekonometrycznego

Model ekonometryczny wyraża ilościowe zależności zachodzące między analizowanymi zjawiskami gospodarczymi, co można zapisać jako:

$$\mathbf{y}_n = f(\mathbf{y}_n, \mathbf{x}_n, \Theta, \boldsymbol{\varepsilon}_n) \tag{2.1}$$

gdzie dla każdej obserwacji n:

- $\mathbf{y}_n$ – wektor $[g \times 1]$ zmiennych opisywanych przez poszczególne równania modelu,
- $\mathbf{x}_n$ – wektor $[k \times 1]$ zmiennych, które przyczyniają się do wyjaśnienia analizowanych zjawisk,
- Θ – macierz nieznanych parametrów modelu,
- $\boldsymbol{\varepsilon}_n$ – wektor $[g \times 1]$ składników losowych (zakłócających),
- f – symbol funkcji.

Przedstawiony zapis (2.1) modelu oznacza, że zawiera on g równań stochastycznych, będących funkcjami wyróżnionych zmiennych, parametrów oraz składników losowych. Zadaniem analizy ekonometrycznej jest znalezienie zależności występujących między rozpatrywanymi procesami ekonomicznymi. Innymi słowy, na podstawie posiadanych obserwacji dotyczących kształtowania się badanych zjawisk gospodarczych poszukuje się postaci analitycznej funkcji f oraz wartości nieznanych parametrów Θ.

Analiza ekonometryczna obejmuje kilka etapów:

etapy analizy ekonometrycznej

1) ustalenie celu prowadzonej analizy,
2) specyfikacja modelu ekonometrycznego,
3) zebranie informacji statystycznych,
4) identyfikacja modelu,
5) estymacja parametrów strukturalnych modelu,
6) weryfikacja statystyczna i merytoryczna modelu,
7) wykorzystanie praktyczne.

cele analizy ekonometrycznej

Ustalając cel prowadzonej analizy, należy odpowiedzieć na pytanie, czemu ma ona służyć. Może to być weryfikacja pewnych hipotez dotyczących zależności występujących między analizowanymi zjawiskami i opis tych zależności, budowa prognoz dotyczących przyszłego kształtowania się relacji w gospodarce (np. prognoz popytu) albo prowadzenie symulacji na modelu w celu pogłębienia analizy zależności dynamicznych występujących w rozpatrywanych procesach. Można też budować tzw. modele polityki gospodarczej, na podstawie których