

Modele regresji od teorii do praktyki

Dominik Krężolek

Agnieszka Orwat-Acedańska

Grażyna Trzpiot

Fragment książki
materiał promocyjny

Od klasycznych modeli liniowych po nowoczesne podejścia
nieparametryczne i addytywne

print pdf epub mobi

e_{edu-Libri}



Modele regresji od teorii do praktyki

Dominik Krężołek
Agnieszka Orwat-Acedańska
Grażyna Trzpiot

Modele regresji od teorii do praktyki

e.edu-Libri

Kraków–Legionowo 2025

© edu-Libri s.c. 2025

Redakcja merytoryczna i korekta: Danuta Kamińska-Hass

Projekt okładki i stron tytułowych: GRAFOS

Ilustracja na okładce: Dominik Krężolek

Recenzent:

Prof. dr hab. Józef Stawicki

Publikacja dofinansowana przez Uniwersytet Ekonomiczny w Katowicach

Wydawnictwo edu-Libri

ul. Zalesie 15, 30-384 Kraków

e-mail: edu-libri@edu-libri.pl

Skład i łamanie: GRAFOS

Druk i oprawa: OSDW Azymut Sp. z o.o.

Łódź ul. Senatorska 31

ISBN druk 978-83-66395-99-2

ISBN pdf 978-83-68441-00-0

ISBN epub 978-83-68441-01-7

ISBN mobi 978-83-68441-02-4

Spis treści

1. Wprowadzenie	7
2. Modele regresji	11
3. Modele regresji liniowej	17
3.1. Standardowy model regresji liniowej	17
3.2. Klasyczny model regresji wielorakiej	22
3.3. Model logitowy dla binarnej zmiennej objaśnianej	30
3.4. Mieszany liniowy model regresji.....	33
4. Klasyczne rozszerzenia modeli regresji	35
4.1. Nieparametryczny model regresji	35
4.2. Standardowy addytywny model regresji	40
4.3. Addytywny model regresji Poissona	46
4.4. Przestrzenny model regresji.....	56
5. Analiza danych a modele regresji	65
5.1. Regresja krokowa	65
5.2. Regresja składowych głównych	69
5.3. Uogólnione modele regresji	79
5.3.1. Binarne modele regresji	80
5.3.2. Log-liniowy model Poissona	82
5.3.3. Uogólniony liniowy model	82
5.4. Regulowana regresja	85
5.4.1. Regresja grzbietowa	85
5.4.2. Regresja LASSO.....	86
Uwagi końcowe	93
Zakończenie	95
Bibliografia	103
Indeks	105

1. Wprowadzenie

Rzeczywistość nauk statystycznych, stanowiących podwaliny analizy danych, datuje się na przełom wieków XVIII i XIX. Rozwój statystyki jest paralelny do rozwoju matematyki (rys. 1.1). Najczęściej u podwalin rozwoju metod statystycznych wymieniamy twierdzenia rachunku prawdopodobieństwa, badanie korelacji oraz twierdzenie o rozkładach warunkowych sir Thomasa Bayes'a. Różnica mierzona w latach pomiędzy publikacją twierdzeń o rozkładzie dwumianowym a publikacją twierdzenia o prawdopodobieństwie całkowitym to blisko 150 lat. Wówczas zapisano dowody analityczne sformułowanych i ogłoszonych twierdzeń [Trzpiot, Krężolek, 2018, s. 112–132].



Rysunek 1.1. Rozwój nauk statystycznych na osi czasu

Źródło: <https://www.slideshare.net/capgemini/impact-of-big-data-on-analytics>

Kolejnym krokiem milowym w rozwoju nauk statystycznych było określenie zasad modelu regresji. Najwcześniejszą formą regresji była metoda najmniejszych kwadratów opublikowana przez Legendre'a w 1805 r. [Legendre, 1805, s. 72–80] oraz Gaussa [Gauss, 1809, s. 205–224]. Obaj naukowcy zastosowali tę metodę do ustalenia, na podstawie obserwacji astronomicznych, orbit ciał wokół Słońca, przede wszystkim komet, ale także później nowo odkrytych mniejszych planet. Gauss opublikował kolejny etap rozwoju teorii najmniejszych kwadratów w 1821 roku [Gauss, 1821], w tym wersję twierdzenia Gaussa-Markowa.

Termin regresja został wprowadzony przez Francisa Galtona w XIX wieku w celu opisanego pewnego zjawiska biologicznego polegającego na tym, że wzrosty potomków wysokich przodków miały tendencję do regresu w kierunku średniej (zjawisko znane również jako regres w kierunku średniej) [Galton, 1889, s. 80–86]. Dla Galtona regresja miała tylko znaczenie w odniesieniu do zjawiska biologicznego, niemniej jednak jego prace zostały później rozszerzone m.in. przez Yule'a i Pearsona [Pearson, Yule, 1903, s. 211–236] oraz Fishera [Fisher, 1922, s. 597–612] na bardziej ogólny kontekst statystyczny.

W modelowaniu statystycznym analiza regresji jest zbiorem procesów statystycznych służących do szacowania zależności między zmiennymi. Obejmuje wiele technik modelowania i analizowania zmiennych, gdy nacisk położony jest na związek między zmienną zależną i jedną (regresja prosta) lub większą liczbą zmiennych niezależnych (regresja wieloraka). Mówiąc dokładniej, analiza regresji pomaga zrozumieć, w jaki sposób zmienia się wartość typowa zmiennej zależnej (zwanej predykcją), gdy zmienia się dowolna ze zmiennych niezależnych (predyktorów), przy zachowaniu zasady *ceteris paribus*.

Metody regresji nadal stanowią obszar aktywnych badań. W ostatnich dziesięcioleciach opracowano nowe metody regresji stabilnej, regresji uwzględniającej skorelowane odpowiedzi, takie jak szeregi czasowe i krzywe wzrostu, w której zmienną predykcją lub zmienną odpowiedzi są krzywe, obrazy, wykresy lub inne złożone obiekty danych. Warto również podkreślić, że metody regresji stanowią fundament ekonometrii – dziedziny zajmującej się ilościowym opisem, modelowaniem i weryfikacją zależności ekonomicznych. Modele regresji są w ekonomii powszechnie wykorzystywane do analizy wpływu czynników makroekonomicznych, badania zależności między zmiennymi finansowymi czy prognozowania zjawisk gospodarczych. W dalszej części książki, omawiając przykłady dotyczące danych ekonomicznych, wielokrotnie będziemy odwoływać się właśnie do regresji jako podstawowego narzędzia analizy ekonometrycznej [Gruszczyński i in., 2009; Borkowski i in., 2003]. Ekonometryczny model regresji różni się od statystycznego modelu regresji tym, że jest osadzony w kontekście ekonomicznym i podlega dodatkowym wymogom interpretacyjnym i diagnostycznym [Pitera, 2019; Tatarczak, 2022; Gładysz, 2013]. W modelu ekonometrycznym szczególną uwagę przykładą się do doboru odpowiedniej postaci funkcji, weryfikacji założeń modelu (np. homoskedastyczności, braku autokorelacji, normalności składnika losowego)

oraz testowania hipotez ekonomicznych, co wymaga zastosowania rygorystycznych procedur statystycznych [Gładysz, 2013; Tatarczak, 2022].

Powstały kolejne metody regresji uwzględniające różne typy brakujących danych, regresja nieparametryczna, bayesowskie metody regresji, regresja, w której zmienne predykcyjne są mierzone z błędem, regresja z większą liczbą zmiennych predykcyjnych niż obserwacje i wnioskowanie przyczynowe z regresją. Analiza regresji jest szeroko wykorzystywana do przewidywania i prognozowania, gdzie jej zastosowanie w znacznym stopniu pokrywa się z dziedziną uczenia maszynowego. Analiza regresji służy również do zrozumienia, które spośród zmiennych niezależnych są powiązane ze zmienną zależną i do zbadania form tych zależności. W warunkach ograniczonych analizę regresji można wykorzystać do wnioskowania o zależnościach przyczynowych między zmiennymi niezależnymi i zależnymi. Jednak może to prowadzić do złudzeń lub fałszywych relacji, więc należy zachować ostrożność (przykładowo korelacja nie dowodzi związku przyczynowego) [Armstrong, 2012, s. 689].

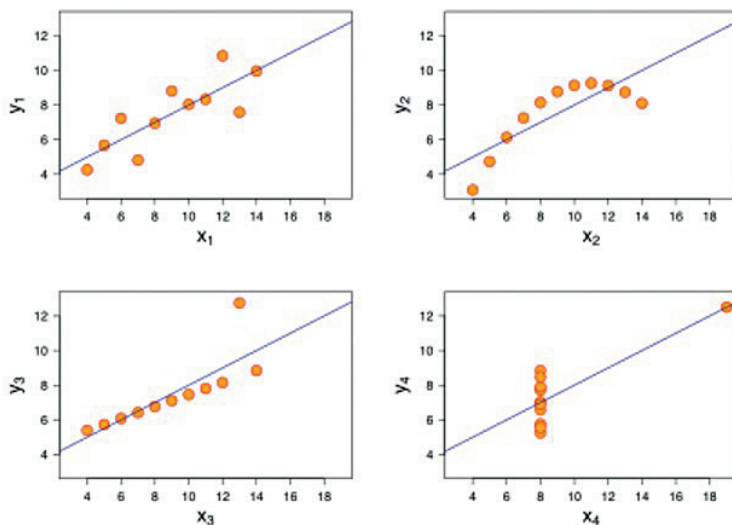
Efektywność metod analizy regresji w praktyce zależy od formy procesu generowania danych i od tego, w jaki sposób odnosi się do zastosowanego podejścia regresyjnego. Ponieważ prawdziwa forma procesu generowania danych nie jest ogólnie znana, analiza regresji często zależy w pewnym stopniu od przyjęcia założeń na temat tego procesu. Te założenia są czasami sprawdzalne, jeśli dostępna jest wystarczająca ilość danych. Modele regresji są często przydatne, nawet jeśli założenia są umiarkowanie naruszone, chociaż mogą nie być optymalne. Jednak w wielu zastosowaniach, zwłaszcza przy niewielkich efektach lub kwestiach przyczynowości opartych na danych obserwacyjnych, metody regresji mogą dawać mylące wyniki [Bishop, 2006, s. 3].

2. Modele regresji

Termin regresja oznacza metodę pozwalającą na zbadanie związku pomiędzy zmiennymi i wykorzystanie tej wiedzy do przewidywania nieznanymi wartości jednych wielkości na podstawie znajomości wartości innych. Związek ten może być dalej wykorzystywany do prognozowania wartości zmiennej objaśnianej. Celem modelowania jest opis wpływu zadanego zbioru zmiennych wyjaśniających $X_1, \dots, X_k$ na zmienną objaśnianą Y . Zmienna objaśniana Y nazywana jest również zmienna zależną, zmienne wyjaśniające nazywane są również zmiennymi niezależnymi. Poszczególne modele różnią się rodzajem zmiennej objaśnianej (zmienna ciągła lub skokowa: binarna, skategoryzowana albo przeliczalna) oraz rodzajem zmiennych wyjaśniających (które mogą być również ciągłe lub skokowe, w tym binarne lub skategoryzowane). W bardziej złożonych przypadkach jest także możliwe włączenie przedziałów czasowych, zmienne można opisać przestrzennie (np. poprzez ich rozmieszczenie lub położenie geograficzne) albo można użyć wskaźników dla grupy.

Główną cechą modeli regresji jest to, że związek między zmienną objaśnianą Y i zmiennymi wyjaśniającymi $X_1, \dots, X_k$ nie jest deterministyczny, nie jest funkcją $f(x_1, \dots, x_k)$ wartości $x_1, \dots, x_k$, a funkcją opisującą związek zmiennych losowych, zatem obserwujemy błędy losowe. Oznacza to, że zmienna objaśniana Y jest zmienną losową, której rozkład zależy od zmiennych objaśniających. Chcemy odpowiedzieć sobie na pytanie, jak zmienia się wartość oczekiwana zmiennej Y w zależności od wartości $x = x_1, \dots, x_k$ zmiennej $X = X_1, \dots, X_k$, czyli $E(Y) = f(x_1, \dots, x_k)$, gdzie $f(x_1, \dots, x_k)$ jest funkcją regresji opisującą poszukiwany związek.

W celu wstępnej oceny zależności najczęściej konstruuje się diagramy korelacyjne. Ich wagę doskonale uwypuklił Anscombe [Anscombe, 1973, s. 17–21], który skonstruował 4 zbiory danych o identycznych podstawowych charakterystykach, ale diametralnie różnych diagramach korelacyjnych. Średnia dla każdej zmiennej x_i wynosiła 9, dla zmiennej y_i 10; wariancja dla x_i była równa 7,5, a dla y_i wynosiła 2,75; współczynnik korelacji liniowej wynosił 0,816 dla każdego zbioru oraz prosta regresji miała postać $y = 3 + 0,5x$.



Rysunek 2.1. Cztery zestawy danych składające się na kwartet Anscombe'a

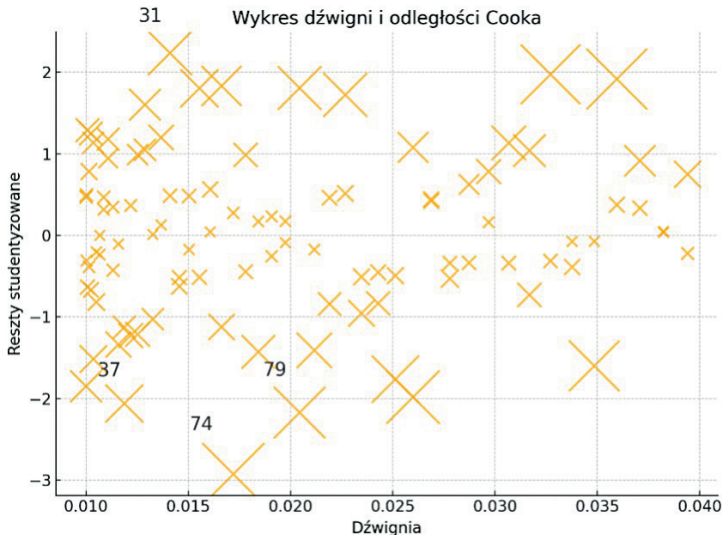
Źródło: https://en.wikipedia.org/wiki/Anscombe%27s_quartet, dostęp 15.03.2025.

Pierwszy wykres punktowy (u góry po lewej) wydaje się przedstawiać prostą zależność liniową odpowiadającą dwóm skorelowanym zmiennym, gdzie y można modelować jako rozkład gaussowski ze średnią liniowo zależną od x . W przypadku drugiego wykresu (prawy górny róg) związek między dwiema zmiennymi jest oczywisty, jednak nie jest on liniowy, a współczynnik korelacji Pearsona nie jest istotny. W tym przypadku bardziej odpowiednia byłaby regresja nieliniowa i odpowiadający jej współczynnik determinacji. Na trzecim wykresie (w lewym dolnym rogu) modelowana zależność jest liniowa, ale powinna mieć inną linię regresji (wymagana byłaby odporna regresja). Obliczona regresja jest równoważona przez jedną wartość odstającą, która wywiera wystarczający wpływ, aby obniżyć współczynnik korelacji z 1 do 0,816. Czwarty wykres (na dole po prawej) pokazuje przykład, w którym jeden punkt o dużej dźwigni wystarcza do uzyskania wysokiego współczynnika korelacji, mimo że inne punkty danych nie wskazują na żaden związek między zmiennymi. Zatem każda omówiona sytuacja powinna być analizowana z wykorzystaniem innego modelu regresji¹.

W analizie modeli regresji wykorzystujemy również liczne wizualizacje – tzw. wykresy diagnostyczne. Przykładowo:

¹ Dane wykorzystane do analiz zaprezentowanych w książce są dostępne na platformie GitHub pod adresem: <https://github.com/dominikkrezolek/Modele-regresji-od-teroi-do-praktyki>. Zawarte w książce analizy empiryczne zostały przeprowadzone w programach RStudio, Statistica oraz w arkuszu kalkulacyjnym Excel.

- **Wykres dźwigni** – wykorzystywany do zbadania, czy występują obserwacje odstające. Dla każdej reszty jest obliczana tzw. siła dźwigni zwana również wpływem (miara wpływu obserwacji na oceny estymatorów regresji). W modelu adekwatnym siła dźwigni nie powinna być zbyt duża, gdyż oznacza to, że pojedyncza obserwacja ma duży wpływ na oceny parametrów. Przyjmuje się, że obserwacja jest wpływowa, jeśli przekracza dwukrotność średniej siły dźwigni. Inną podobną miarą wpływu obserwacji na model jest **odległość Cooka**. Pokazuje ona różnicę między wyznaczonymi wartościami współczynników a wartościami obliczonymi przy wyłączeniu danego przypadku z obliczeń. Wszystkie odległości powinny być tego samego rzędu. Jeśli nie są, to można przypuszczać, że dany przypadek (przypadki) miał istotny wpływ na obciążenie współczynników równania regresji. Często preferuje się analizę odległości Cooka zamiast analizy wartości wpływowych. Powinna ona być mniejsza od 1, jeśli chcemy uznać model za adekwatny. Graficzna prezentacja na rys. 2.2.



Rysunek 2.2. Wykres dźwigni i odległości Cooka

Źródło: opracowanie własne.

- **Wykres reszt** – wykres przedstawiający na jednej osi wartości estymowane, a na drugiej reszty lub standaryzowane reszty. Powszechną praktyką jest uznanie, że obserwacja jest odstająca, jeżeli jej reszta standaryzowana (lub studentyzowana) jest większa co do wartości bezwzględnej od 2. Dla modelu adekwatnego średnia wartość reszt nie powinna zależeć od wartości dopasowania (powinniśmy w wyniku dostać pas punktów losowo rozmieszczonych wokół prostej $y = 0$). Przykładowy wykres reszt pokazano na rys. 2.3.

Bibliografia

- Anscombe F.J. [1973], *Graphs in Statistical Analysis*, „American Statistician”, 27(1).
- Armstrong J.S. [2012], *Illusions in Regression Analysis*, „International Journal of Forecasting”, 28(3).
- Bishop Ch.M. [2006], *Pattern Recognition and Machine Learning*, New York.
- Borkowski, B., Dudek, H., Szczesny, W. [2003], *Ekonometria. Wybrane zagadnienia*, Warszawa.
- Dickey D.A., Fuller W.A. [1979], *Distribution of the estimators for autoregressive time series with a unit root*, „Journal of the American Statistical Association”, 74(366), 427–431.
- Durbin J., Watson G.S. [1950], *Testing for serial correlation in least squares regression. I*, „Biometrika”, 37, 3–4.
- Fahrmeir L., Kneib T., Lang S., Marx B. [2013], *Regression. Models, Methods and Applications*, Springer, New York.
- Fisher R.A. [1922], *The goodness of fit of regression formulae, and the distribution of regression coefficients*, „Journal of the Royal Statistical Society”, 85(4).
- Friedman J.H., Stuetzle W. [1981], *Projection Pursuit Regression*, „Journal of the American Statistical Association”, 76.
- Galton F. [1889], *Kinship and Correlation*, „Statistical Science. Institute of Mathematical Statistics”, 4 (2).
- Gauss C.F. [1809], *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*, Hamburg.
- Gładysz B. [2013], *Ekonometria*, Wydawnictwo Naukowe PWN.
- Goldfeld S.M., Quandt R.E. [1965], *Some Tests for Homoscedasticity*, „Journal of the American Statistical Association”, 60(310).
- Greene W.H. [2000], *Econometric analysis*, wyd. 4, London.
- Gruszczyński M., Kuszewski T., Podgórska M. (red.) [2009], *Ekonometria i badania operacyjne: podręcznik dla studiów licencjackich*, Warszawa.
- Härdle W. [1992], *Applied Nonparametric Regression*, Berlin.
- Hastie T., Tibshirani R. [1990], *Generalized Additive Models*, New York.
- Kopczewska K., Elhorst P. [2024], *New developments in spatial econometric modelling*, „Spatial Economic Analysis”, 19(1), 1–7.
- Kopczewska K. [2005], *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R*, wyd. 2, CeDe-Wu, Warszawa.
- Kossowski T. [2018], *Teoretyczne aspekty modelowania przestrzennego w badaniach regionalnych*, „Rozwój Regionalny i Polityka Regionalna”, (42), 9–26.
- Krężolek D. [2023], *Wybrane metody i techniki analizy danych z programem Statistica*, Katowice.
- Legendre A.-M. [1805], *Nouvelles méthodes pour la détermination des orbites des comètes*, Paris.
- Moran P.A.P. [1950], *Notes on continuous stochastic phenomena*, „Biometrika”, 37(1/2), 17–23.
- Orwat-Acedańska A. [2019], *Dynamic spatial panel data models in identifying socio-economic factors affecting the level of health in selected European countries*, „European Spatial Research and Policy”, 26(1), 195–211.

- Osińska M. [2007], *Ekonometria współczesna*, Toruń.
- Panek T., Zwierchowski J. [2013], *Statystyczne metody wielowymiarowej analizy porównawczej. Teoria i zastosowania*, Warszawa.
- Pearson K., Yule G.U. [1903], *The Law of Ancestral Heredity*, „Biometrika”, 2(2).
- Pitera, K. [2019], *Statystyka dla ekonomistów*, Wydawnictwo Uniwersytetu Ekonomicznego.
- Ptak-Chmielewska A. [2013], *Uogólnione modele liniowe*, Warszawa.
- Shapiro S.S., Wilk M.B. [1965], *An analysis of variance test for normality (complete samples)*, „Biometrika”, 52(3–4).
- Suchecki B. (red.) [2010], *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, C.H. Beck, Warszawa.
- Suchecki B. (red.) [2012], *Ekonometria przestrzenna II. Modele zaawansowane*, C.H. Beck, Warszawa.
- Suchecki B. [2023], *Metody stochastyczne w ekonometrii przestrzennej – nowoczesna perspektywa*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Śliwicki D. [2016], *Estymacja jądrowa w analizie ekonometrycznej*, Toruń.
- Tatańczak A. [2022], *Modele ekonometryczne w praktyce*, Wydawnictwo Uniwersytetu Warszawskiego.
- Tibshirani R. [1996], *Regression Shrinkage and Selection via the Lasso*, „Journal of the Royal Statistical Society”, 58(1).
- Trzęsiok J. [2004], *Wybrane nieparametryczne metody regresji i ich zastosowania*, „Prace Naukowe Akademii Ekonomicznej im. Oskara Langego we Wrocławiu”, nr 1022.
- Trzpiot G., Krężolek D. [2018], *Od analizy danych do innowacyjnych technologii*, [w:] *Innowacyjna gospodarka. Innowacyjne organizacje. Innowacyjni ludzie.*, C.M. Olszak, G. Głód (red.), Katowice.
- White H. [1980], *A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity*, „Econometrica”, 48.
- Wood, S.N. [2006], *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC, Boca Raton, ISBN 978-1584884743.

Indeks

Analiza głównych składowych (PCA) 69–72, 76, 96

– regresji 8–9

Autokorelacja 93, 95

– przestrzenna 59–60

– – składnika losowego SEM 57

– reszt 24, 59

– składników losowych 30

Błąd standardowy 67, 75, 78

Dewiancja 44–45, 49, 51, 55

Dźwignia 12

–, siła 13

–, wykres 13

Funkcja kary 85–86

Heteroskedastyczność 18–19, 59, 77, 93, 95

Homoskedastyczność 8, 18, 26–27, 30, 79

Iloraz szans 32, 95

Kryterium AIC 49, 51, 54–55, 60–61, 63

– BIC 60, 65

Liczba referencyjna stopni swobody 42

Liniowy model uogólniony 82

Metoda bootstrap 88, 91

– głównych składowych 76

– LASSO 86–88, 96

– najmniejszych kwadratów 8, 23, 86

– – –, uogólniona 26

– – –, podwójna 93

– – –, – pośrednia 93

– największej wiarygodności 31, 80

Model błędu przestrzennego (SEM) 60–61, 63

– klasyczny regresji wielorakiej 22, 24, 27

– LASSO 89, 91

– liniowy klasyczny 93

– logitowy 30–31, 80, 84, 95

– log-liniowy 84

– log-liniowy Poissona 82

– opóźnienia przestrzennego (SLM) 57, 60–63

– regresji binarny 80

– – głównych składowych 69

– – mieszany liniowy 33

– – nieparametryczny 35–36, 38

– – – Poissona addytywny 46–47, 51

– – – uogólniony addytywny 48

– – przestrzenny 56–57

– – standardowy addytywny 40–41, 44

Modele nieparametryczne 36

– regresji uogólnione 79

Obserwacje odstające 13

Odległość Cooka 13

Przedział ufności 88–89, 91

Przeuczenie modelu (*overfitting*) 92

Regresja głównych składowych 71, 75, 78–79

– grzbietowa 85–86, 96

– krokowa 65–66, 96

– – postępująca 66

– LASSO 89

– regulowana 85

– składowych głównych 69

Reszty standaryzowane 13

–, wykres 13

Rodzina rozkładów wykładniczych 83

Rozkład dwumianowy 7, 79

– gamma 79

– geometryczny 79

– normalny 79

– Poissona 79

– wielomianowy 79

– wykładniczy 79

Rzeczywista liczba stopni swobody (EDF)

42, 44, 50, 54

Skorygowany R^2 44–45, 51, 55

Statystyka Walda 62

Test Breuscha-Pagana 22, 30

– Dickeya-Fullera 28

– Durbina-Watsona 24, 30

– Goldfelda-Quandt 19, 22

– LR 62

– Shapiro-Wilka 19–20, 22, 26, 29, 75, 79

– Walda 32, 95,

– White’a 22, 26, 30

– – mnożników Lagrange’a 75, 79

Walidacja krzyżowa 86–87, 90

Wektor głównych składowych 70–71

Współczynnik determinacji 12, 21, 26, 29,

37, 39, 68, 75, 78, 89, 92, 95

– – R^2 66

– informacyjny Akaike (AIC) 65

Współczynniki głównych składowych

73–74, 77

– LASSO 90

Współliniowość 23, 36, 69, 71–72, 80

Wykres kwantyl-kwantyl 14

Zmienna binarna 11, 30, 80

– ciągła 11

– skokowa 11

Książka „Modele regresji – od teorii do praktyki” to kompleksowy przewodnik po różnych typach modeli regresyjnych – od klasycznych modeli liniowych po nowoczesne podejścia nieparametryczne i addytywne.

Publikacja łączy teorię z praktycznymi przykładami, co czyni ją bardzo przystępną zarówno dla studentów, jak i praktyków analizy danych. Każde zagadnienie teoretyczne jest poparte przykładami z danych rzeczywistych i interpretacją wyników. Autorzy szczegółowo pokazują, jak ocenić jakość modelu. Książka nie ogranicza się do regresji liniowej, ale uwzględnia również inne rodzaje, jak regresja logitowa, nieparametryczna, mieszana, GAM oraz inne. Analiza danych z różnych rynków sprawia, że czytelnik lepiej rozumie wartość modeli regresji w praktyce. Dzięki temu stanowi nowoczesne i aktualne źródło wiedzy, które może być cennym uzupełnieniem nawet dla bardziej zaawansowanych analityków.

W przeciwieństwie do bardziej formalnych akademickich opracowań, książka napisana jest przystępnym językiem, z myślą o czytelniku na różnych poziomach zaawansowania. Autorzy wyjaśniają każdy krok analizy, co sprawia, że publikacja może służyć zarówno studentom, jak i praktykom analizy danych.

Autorzy są pracownikami Katedry Demografii i Statystyki Ekonomicznej Uniwersytetu Ekonomicznego w Katowicach

www.edu-libri.pl

Wydawnictwo edu-Libri jest oficyną wydawniczą publikacji naukowych i edukacyjnych.

Współpracujemy z profesjonalnymi redaktorami merytorycznymi i technicznymi z dużym doświadczeniem w przygotowywaniu publikacji specjalistycznych. Stawiamy na jakość łączoną z nowoczesnością, a najważniejsze dla nas są przyjemność współtworzenia i satysfakcja z dobrze wykonanego zadania.

Nasze publikacje (drukowane i elektroniczne) są dostępne w księgarniach stacjonarnych i internetowych oraz w czytelniach on-line ibuk.pl, osbi.pl i nasbi.pl - szczegóły na stronie wydawnictwa.

