
Spis treści

Wstęp	8
1. Wprowadzenie do R	11
1.1. Instalacja	12
1.1.1. Instalacja oraz pierwsze uruchomienie	12
1.1.2. Instalacja oraz korzystanie z bibliotek	14
1.2. Podstawy obsługi R	16
1.2.1. Struktury danych	16
1.2.2. Grafika	27
1.2.3. Programowanie	35
1.2.4. Dopasowanie środowiska	49
1.3. Zadania	50
2. Co to jest statystyka i skąd się wzięła	57
2.1. Rys historyczny	57
2.2. Co to jest statystyka	63
2.3. Podstawowe pojęcia	64
2.4. Badania statystyczne oraz ich rodzaje	67
2.5. Zadania	71
3. Prezentacja danych statystycznych	73
3.1. Szereg statystyczny	73
3.2. Podstawowe formy prezentacji danych	77
3.3. Zadania	88
4. Analiza struktury	91
4.1. Wskaźniki położenia	91
4.1.1. Średnia arytmetyczna	92
4.1.2. Średnia harmoniczna	95

4.1.3.	Średnia geometryczna	97
4.1.4.	Dominanta	101
4.1.5.	Mediana	103
4.2.	Wskaźniki rozproszenia (zmienności)	109
4.2.1.	Rozstęp.....	109
4.2.2.	Wariancja	109
4.2.3.	Odchylenie standardowe	111
4.2.4.	Współczynnik zmienności.....	113
4.2.5.	Ocena (nota, wynik) standardowa	114
4.2.6.	Nierówność dyspersyjna.....	115
4.3.	Miary asymetrii	115
4.4.	Miary koncentracji	118
4.5.	Zadania	121
5.	Podstawy rachunku prawdopodobieństwa	125
5.1.	Zdarzenia losowe i ich prawdopodobieństwo	127
5.2.	Elementy kombinatoryki.....	132
5.3.	Podstawowe metody rachunku prawdopodobieństwa.....	136
5.3.1.	Prawdopodobieństwo warunkowe oraz niezależność zdarzeń	136
5.3.2.	Prawdopodobieństwo całkowite oraz wzór Bayesa	142
5.4.	Zmienne losowe oraz ich rozkłady prawdopodobieństwa	144
5.4.1.	Dyskretne rozkłady prawdopodobieństwa	148
5.4.2.	Ciągłe rozkłady prawdopodobieństwa	158
5.4.3.	Rozkłady ucięte	173
5.5.	Model	177
5.6.	Zadania	180
6.	Podstawy statystyki matematycznej	187
6.1.	Statystyczna próba losowa	187
6.2.	Estymacja	191
6.2.1.	(*) Estymacja punktowa	195
6.2.2.	Przedział ufności dla wartości oczekiwanej rozkładu normalnego ..	198
6.2.3.	Przedział ufności dla wariancji rozkładu normalnego.....	200
6.2.4.	Przedział ufności dla wskaźnika struktury.....	201
6.2.5.	Minimalna liczebność próby	202
6.3.	Testowanie hipotez statystycznych	206
6.3.1.	Test zgodności χ^2 Pearsona	217

6.3.2.	Test Kołmogorowa-Smirnowa.....	222
6.3.3.	Testy normalności	224
6.3.4.	Testy zgodności dla dwóch populacji	228
6.3.5.	Test istotności dla wartości oczekiwanej (średniej)	231
6.3.6.	Test istotności dla wariancji	234
6.3.7.	Test istotności dla dwóch wariancji	236
6.3.8.	Test istotności dla kilku wariancji	240
6.3.9.	Test istotności dla dwóch wartości oczekiwanych	241
6.3.10.	Test znaków	249
6.3.11.	Test kolejności par Wilcoxon (test znaków rangowanych Wilcoxon)	251
6.3.12.	Analiza wariancji (ANOVA)	253
6.3.13.	Test istotności dla wskaźnika struktury	272
6.3.14.	Test istotności dla dwóch wskaźników struktury	274
6.3.15.	Test serii	279
6.3.16.	Schemat wyboru procedury testowej	283
6.4.	Zadania	283
7.	Podstawy rachunku błędów	297
7.1.	Błędy grube	298
7.2.	Błędy przypadkowe	301
7.3.	Zaokrąglenie pomiarów	303
7.4.	Zadania	304
8.	Związki cech	305
8.1.	Zależność korelacyjna	306
8.1.1.	Test niezależności χ^2 Pearsona	312
8.2.	Miary korelacji dwóch zmiennych	318
8.2.1.	Współczynnik korelacji liniowej Pearsona	318
8.2.2.	Współczynnik korelacji rang Spearmana	324
8.3.	Związek cech niemierzalnych	329
8.3.1.	Współczynnik φ Yule'a	329
8.3.2.	Współczynnik V Craméra	331
8.3.3.	Współczynnik kontyngencji C Pearsona	332
8.3.4.	Współczynnik τ Goodmana i Kruskala	333
8.3.5.	Współczynnik zgodności κ	334
8.4.	Zadania	336

9. Regresja	339
9.1. Regresja liniowa	342
9.1.1. Przedział ufności dla parametrów regresji liniowej	348
9.1.2. Test istotności dla współczynnika regresji	349
9.2. Modelowanie statystyczne	350
9.3. Regresja wielokrotna	357
9.3.1. (*) Regresja a współliniowość zmiennych	362
9.4. Regresja nieliniowa	366
9.4.1. Regresja logistyczna	373
9.5. Analiza przeżycia	378
9.6. Zadania	383
10. Analiza dynamiki zjawisk masowych	387
10.1. Metody badania dynamiki szeregów czasowych	389
10.1.1. Metody podstawowe	389
10.1.2. Indeksy indywidualne i zespołowe	392
10.2. Model wahań w czasie	397
10.2.1. Trend	398
10.2.2. Sezonowość	408
10.3. Zadania	415
11. Metody wizualizacji oraz analizy danych wielowymiarowych	419
11.1. Walidacja modeli	419
11.2. Wstępne przetwarzanie danych	419
11.2.1. Transformacje danych	420
11.2.2. Uzupełnianie (imputacja) wartości brakujących	422
11.3. Redukcja wymiarowości oraz wizualizacja danych wielowymiarowych	423
11.3.1. Analiza składowych głównych	424
11.3.2. Analiza czynnikowa	427
11.3.3. Skalowanie wielowymiarowe	433
11.3.4. Analiza korespondencji	437
11.3.5. Wykresy obrazkowe	439
11.4. Klasyfikacja	442
11.4.1. Analiza skupień	443
11.4.2. Analiza dyskryminacyjna	449
11.5. Zadania	461
12. Użyteczne procedury w R	463

12.1. Zadania	478
Odpowiedzi do zadań	479
Tablice statystyczne	495
Polski indeks pojęć	506
Angielski indeks pojęć	511
Indeks nazwisk	516
Indeks poleceń R	519
Indeks pakietów R	529
Indeks zbiorów danych R	532
Bibliografia	534

Wstęp

W dłuższej perspektywie wkład statystyki w rozwój świata nie zależy zbyt wiele od wykształcenia wielu wysoko wykwalifikowanych statystyków pracujących w przemyśle. Jest to raczej potrzeba stworzenia statystycznie myślącego pokolenia fizyków, chemików, inżynierów i innych, którzy na różne sposoby będą rozwijać gospodarkę jutra.

W.A. Shewhart & W.E. Deming

STATYSTYKA ma długą prehistorię, a krótką historię. Jej pochodzenie można wywodzić od początków ludzkości, ale dopiero w ostatnich czasach okazała się dziedziną o wielkim znaczeniu praktycznym. Czy statystyka jest oddzielną dziedziną wiedzy jak fizyka, chemia, ekonomia czy inne uznane od lat nauki? Nie ma przecież w Polsce tytułów naukowych ze statystyki, można być profesorem matematyki, ale nie statystyki. Matematyk oddaje się dedukowaniu twierdzeń na podstawie danych aksjomatów i reguł wnioskowania. Ekonomista wyjaśnia, co, kiedy i za ile produkować. Fizyk opisuje otaczający nas świat za pomocą praw i reguł nim rządzących, a chemik wyjaśnia zachodzące w świecie reakcje. Każda z tych dziedzin ma swoje potrzeby i własne metody ich rozwiązywania, które nadają im status oddzielnych nauk. Czy istnieją zatem czysto statystyczne problemy, które statystyka mogłaby rozwiązywać? Jeśli nie, to czy jest to rodzaj sztuki, logiki lub techniki stosowanej do rozwiązywania problemów w innych naukach? Przez wiele lat słowo „statystyka” nie było używane ani często, ani poprawnie. Zapatrywano się na nią sceptycznie, niewielu rządowych specjalistów czy pracowników naukowych stosowało jej narzędzia. Inaczej niż dzisiaj, gdy istnieje ogromny rynek pracy dla statystyków – w rządzie, przemyśle, nauce – albowiem niemal zawsze musimy wyciągać wystarczające wnioski z niewystarczających przesłanek. W zasadzie nie ukazują się żadne poważne prace naukowe w wielu dziedzinach, w których nie byłoby metod statystycznych.

Jak przewidywać społeczno-ekonomiczne charakterystyki ludności na podstawie bieżących tendencji? Jak podejmować decyzje sprzyjające wzrostowi dobrobytu społeczeństwa? Czy jutro będzie padać deszcz? Czy ubezpieczyć się na wypadek kłęski neurodzażu, śmierci, katastrofy? Jak łatwo się domyślić, główną przeszkodą w udzieleniu odpowiedzi na te pytania jest niepewność – brak jednoznacznej relacji między przyczyną i skutkiem. Dopiero z początkiem poprzedniego wieku nauczono się wyznaczać niepewność¹. Dane trzeba przetwarzać, aby dowiedzieć się, do jakiego stopnia można z nich usunąć element niepewności. Wiedza o zasobie niepewności zawartej w danych jest kluczem do podjęcia odpowiedniej decyzji. Statystyka jest więc logiką, za pomocą któ-

¹Datuje się to do roku 1933, kiedy KOŁMOGOROW wprowadził aksjomatyczną definicję prawdopodobieństwa.

rej można wspiąć się po drabinie od danych do informacji o jeden szczebel wyżej. Statystyka to pewna metodologia podejmowania decyzji, czyli wnioskowania w warunkach niepewności. Wiedza osoby zajmującej się statystyką (czyli statystyka) pozwala zaufać głoszonemu przez niego sądowi lub podejmowanym decyzjom z uwzględnieniem nieuniknionego ryzyka. Musimy pamiętać, że zawsze istnieje ryzyko porażki, popełnienia błędu, ale zdecydowanie lepiej wiedzieć coś jedynie z pewnym prawdopodobieństwem, niż nie wiedzieć nic z całą pewnością.

Osoby uprawiające statystykę często popełniają różnorakie błędy, spowodowane kilkoma przyczynami:

- Zdecydowana większość ludzi korzystających z metod statystycznych to specjaliści w zupełnie innych dziedzinach, dla których statystyka odgrywa rolę pomocniczą – ekonomiści, biolodzy, chemicy itp.
- Klasyczna teoria statystyki powstawała ponad pół wieku temu i z braku wówczas odpowiednio wydajnych komputerów opiera się na zaawansowanych metodach analitycznych (czytaj: długich i skomplikowanych wzorach) oraz koniecznych do ich wyprowadzenia założeniach, nie zawsze spełnianych w praktyce (raczej rzadko) i często nierozumianych (lub rozumianych błędnie) przez niestatystyków.
- Próba wyjaśnienia tej złożonej teorii na kursie lub w podręczniku dla niestatystyków kończy się zwykle katalogiem przepisów „kiedy stosować który test”. Niestety, żaden katalog nie uwzględni wszystkich przypadków, z którymi możemy mieć do czynienia, i nie zastąpi zrozumienia podstaw.
- Podstawową konsekwencją rozpowszechnienia komputerów jest ułatwienie dostępu do tych skomplikowanych metod: z wczytaniem danych do pakietu statystycznego jakoś sobie poradzimy, potem tylko trzeba „doklikać się” do testu i komputer zawsze „wyrzuci” wynik. Komputer jednak nie przyjmie odpowiedzialności za dobór metody do problemu, poprawne sformułowanie hipotezy oraz wyciągnięcie właściwych wniosków.

Te wszystkie przyczyny popełniania błędów powodują, że w społeczeństwie modne jest powiedzenie, że „statystyka kłamie”. Nie jest to jednak prawda, dużo bliższe prawdy jest stwierdzenie: „Liczby nie kłamią, ale kłamcy liczą”². Książka ta ma służyć jako podręcznik do nauki statystyki dla początkujących, aby nie stali się takimi „kłamcami”. Znajduje się w niej wiele uwag, które w kluczowych momentach przestrzegają przed popełnianiem powszechnych błędów. Do zrozumienia materiału wymagana jest jedynie minimalna wiedza matematyczna, raczej niewykraczająca zakresem poza materiał szkoły średniej, a w wielu miejscach nawet gimnazjum. Część trudniejszych problemów została jedynie zarysowana (pominięto niemal wszystkie wyprowadzenia wzorów) w celu lepszego zrozumienia całości materiału (materiał dodatkowy oznaczono gwiazdką). Wszystkie omawiane techniki są bogato ilustrowane

²Słowa wypowiedziane przez C.H. GROSVENORA (1833–1917), amerykańskiego pułkownika podczas wojny domowej w USA w latach 1860–1865.

przykładami. Szczególną uwagę zwrócono na wizualizację metod statystycznych za pomocą wykresów i tabel.

Ponieważ obecnie większość metod statystycznych jest dość żmudna obliczeniowo, wszystkie przykłady zostały przeprowadzone również z użyciem pakietu statystycznego R³, który zyskał obecnie ogromną popularność na świecie. Wszystkie przytoczone w książce funkcje, pakiety oraz zbiory danych zebrano w odpowiednich indeksach na końcu książki. Wszystkie zbiory opisane są przy pierwszym wystąpieniu, przy kolejnych wzmiankowany jest jedynie pakiet, z którego pochodzą. W dobie Internetu (wyszukiwania informacji) niezbędna jest znajomość języka angielskiego również w statystyce, z tego względu wszystkie istotne pojęcia podano również w tym języku. Wyszukiwanie pojęć ułatwiają indeksy pojęć w języku polskim i angielskim.

Każdy rozdział zakończony jest zadaniami o zróżnicowanym poziomie trudności (trudniejsze zadania oznaczone są gwiazdką), które pozwalają lepiej zrozumieć oraz utrwalić materiał. Część zadań wymaga jedynie kartki oraz czegoś do pisania, natomiast znaczna część została przewidziana do rozwiązania za pomocą R. W przypadku odpowiedzi używane zbiory nie są dołączane (**attach**), natomiast używane są, jakby były dołączone.

Podręcznik ten powstał na podstawie prowadzonych przeze mnie od wielu lat zajęć na Uniwersytecie im. Adama Mickiewicza w Poznaniu.

Na koniec chciałbym podziękować Panu Doktorowi Maciejowi Łuczakowi, bez którego kształt tej książki byłby zupełnie inny.

³<http://www.r-project.org/>

Bibliografia

- [1] Anscombe F.J. (1973), Graphs in Statistical Analysis. *American Statistician*, **27**, 17-21.
- [2] Bernstein P.L. (1997), *Przeciw bogom. Niezwykłe dzieje ryzyka*, WIG-Press.
- [3] Białyński-Birula I., Białyńska-Birula I. (2002), *Modelowanie rzeczywistości*, Prószyński i S-ka.
- [4] Biecek P. (2008), *Przewodnik po pakiecie R*, GiS.
- [5] Box G.E.P., Jenkins G.M. (1983), *Analiza szeregów czasowych*, PWN.
- [6] Brandt S. (1999), *Analiza danych*, PWN.
- [7] Breiman L. (2001), Random Forests. *Machine Learning*, **45 (1)**, 5-32.
- [8] Chen C., Härdle W., Unwin A. (2008), *Handbook of Data Vizualization*, Springer.
- [9] Conover W.J., Johnson M.E., Johnson M.M. (1981), A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, **23**, 351-361.
- [10] Cook D., Swyane D.F. (2007), *Interactive and Dynamic Graphics for Data Analysis With R and GGobi*, Springer.
- [11] Cowpertwait P.S.P., Metcalfe A.V. (2009), *Introductory Time Series with R*, Springer.
- [12] Crawley M.J. (2007), *The R Book*, Wiley.
- [13] Cryer J.D., Chan K.S. (2008), *Time Series Analysis With Applications in R*, Springer.
- [14] Ćwik J., Koronacki J. (2009), *Statystyczne systemy uczące się. Ćwiczenia w oparciu o pakiet R*, Oficyna Wydawnicza Politechniki Warszawskiej.
- [15] Deza M.M., Deza E. (2009), *Encyclopedia of Distances*, Springer.
- [16] Dodge Y., (2008), *The Concise Encyclopedia of Statistics*, Springer.
- [17] Duda R.O., Hart P.E., Stork D.G. (2001), *Pattern Classification*, Wiley.
- [18] Durka P.J. (2003), *Wstęp do współczesnej statystyki*, Adamantan.
- [19] Everitt B.S. (2005), *An R and S-PLUS[®] Companion to Multivariate Analysis*, Springer.
- [20] Everitt B.S. (2008), *Chance Rules*, Springer.

- [21] Feller W. (2006), *Wstęp do rachunku prawdopodobieństwa. Tom I*, PWN.
- [22] Freedman D., Diaconis, P. (1981), On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **57**, 453-476.
- [23] Frey B. (2007), *75 sposobów na statystykę. Jak zmierzyć świat i wygrać z prawdopodobieństwem*, Helion.
- [24] Gatnar E. (2008), *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, PWN.
- [25] Greń J. (1984), *Statystyka matematyczna. Modele i zadania*, PWN.
- [26] Hastie T., Tibshirani R., Friedman J. (2009), *The elements of statistical learning*, Springer.
- [27] Hyndman, R.J., Fan, Y. (1996), Sample quantiles in statistical packages. *American Statistician*, **50**, 361-365.
- [28] Izenman A.J. (2008), *Modern multivariate statistical techniques*, Springer.
- [29] Kala R. (2005), *Statystyka dla przyrodników*, Wydawnictwo AR w Poznaniu.
- [30] Kassyk-Rokicka H. (2001), *Statystyka nie jest trudna – mierniki statystyczne*, PWE.
- [31] Kaufman L., Rousseeuw P.J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley.
- [32] King B.M, Minium E.W. (2009), *Statystyka dla psychologów i pedagogów*, PWN.
- [33] Kopczewska K., Kopczewski T., Wójcik P. (2009), *Metody ilościowe w R. Aplikacje ekonomiczne i finansowe*, CeDeWu.
- [34] Kopczyński M. (2005), *Podstawy statystyki*, Mówią wieki.
- [35] Koronacki J., Mielniczuk J. (2004), *Statystyka dla studentów kierunków technicznych i przyrodniczych*, WNT.
- [36] Kryszko W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. (2000), *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN.
- [37] Krzyśko M. (2004a), *Statystyka matematyczna*, Wydawnictwo Naukowe UAM.
- [38] Krzyśko M. (2004b), *Wykłady z teorii prawdopodobieństwa*, WNT.
- [39] Krzyśko M., Wołyński W., Górecki T., Skorzybut M. (2008), *Systemy uczące się*, WNT.
- [40] Krzyśko M. (2009), *Podstawy wielowymiarowego wnioskowania statystycznego*, Wydawnictwo Naukowe UAM.
- [41] Landis J.R., Koch, G.G. (1977), The measurement of observer agreement for categorical data. *Biometrics* **33**, 159-174
- [42] Lindman H.R. (1974), *Analysis of variance in complex experimental designs*, W.H. Freeman & Co.
- [43] Luszniewicz A. (1997), *Statystyka nie jest trudna – metody wnioskowania statystycznego*, PWE.

- [44] Johnson L., Montgomery D., J. Gardiner (1990), *Forecasting and Time Series Analysis*, McGraw-Hill.
- [45] Magiera R. (2005), *Modele i metody statystyki matematycznej. Część I. Rozkłady i symulacja stochastyczna*, GiS.
- [46] Magiera R. (2007), *Modele i metody statystyki matematycznej. Część II. Wnioskowanie statystyczne*, GiS.
- [47] Martin G. (1984), Munchausen's statistical grid, which makes all trials significant. *Lancet*, ii: 1457.
- [48] Młodinow L. (2009), *Matematyka niepewności*, Prószyński i S-ka.
- [49] Murrell P. (2006), *R Graphics*, Champan & Hall/CRC.
- [50] Pankratz A. (1983), *Forecasting with univariate Box-Jenkins models: Concepts and cases*, Wiley.
- [51] Pfaff B. (2008), *Analysis of Integrated and Cointegrated Time Series with R*, Springer.
- [52] Rao C.R. (1994), *Statystyka i prawda*, PWN.
- [53] Ritz C., Streibig J.C. (2008), *Nonlinear Regression with R*, Springer.
- [54] Robert C.P., Casella G. (2010), *Introducing Monte Carlo Methods with R*, Springer.
- [55] Sarkar D. (2008), *Lattice: Multivariate Data Visualization with R*, Springer.
- [56] Scott D.W. (1979), On optimal and data-based histograms. *Biometrika*, **66**, 605-610.
- [57] Seber G.A.F. (2004), *Multivariate Observations*, Wiley.
- [58] Sękowski T. (1996), *Człowiek i matematyka*, Polska Oficyna Wydawnicza „BGW”.
- [59] Sheather S.J. (2009), *A modern Approach to Regression with R*, Springer.
- [60] Sheskin D.J. (2007), *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall.
- [61] Shiffler R.E. (1987), Bound for the Maximum Z-Score. *Teaching Statistics*, **9**, 80-81.
- [62] Shiffler R.E., Harsha P.D. (1980), Upper and Lower Bounds for the Sample Standard Deviation. *Teaching Statistics*, **2**, 84-86.
- [63] Shumway R.H., Stoffer D.S. (2006), *Time Series Analysis and Its Applications: With R Examples*, Springer.
- [64] Silverman B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.
- [65] Sobczyk M. (1998), *Statystyka*, PWN.
- [66] Specter P. (2008), *Data Manipulation with R*, Springer.
- [67] Steinhaus H. (2010), *Orzeł czy reszka?*, PWN.
- [68] Sturges H. (1926), The choice of a class-interval. *J. Amer. Statist. Assoc.*, **21**, 65-66.

- [69] Szurek M. (2000), *Matematyka dla humanistów*, RTW.
- [70] Tibshirani R. (1996), Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, **48**(1), 267-288.
- [71] Venables W.N., Ripley B.D. (2002), *Modern Applied Statistics with S*, Springer.
- [72] Verzani J. (2005), *Using R for Introductory Statistics*, Champan & Hall/CRC.
- [73] Walesiak M., Gatnar E. (2009), *Statystyczna analiza danych z wykorzystaniem programu R*, PWN.
- [74] Wickham H. (2009), *ggplot: Elegant Graphics for Data Analysis*, Springer.
- [75] Zieliński R., Zieliński W. (1990), *Tablice statystyczne*, PWN.

Polski indeks pojęć

A

analiza
— czynnikowa, 427
— dyskryminacyjna
— — kwadratowa, 449
— — liniowa, 449
— — regularyzowana, 450
— korespondencji, 437
— — wielowymiarowa, 438
— kowariancji, 364
— przeżycia, 168
— składowych
— — głównych, 424
— — niezależnych, 425
— skupień, 443
— współrzędnych głównych,
434
antymoda, 102
autokorelacja, 402
— cząstkowa, 406
autokowariancja, 403

B

bagging, 457
biały szum, 404
biplot, 425
błąd
— niezwiązany z próbą, 68
— próbkowania, 68
— standardowy, 194
— statystyczny, 297
— średniokwadratowy, 192
boosting, 457

C

cecha, 64
centyl, 105

D

decyl, 105
dendrogram, 443
destymulanta, 359
dewiancja, 361
diagram
— Gantta, 467
— korelacyjny, 307
— Venna, 40
— Voronoia, 452
dominanta, 101
drzewa klasyfikacyjne, 454
dystrybuanta empiryczna, 84

E

efekt, 254
eksces, 118
estymator, 191

F

fraktyl, 105
funkcja
— stressu, 433
— wiarogodności, 195
— wiążąca, 372

G

geometryczne odchylenie
standardowe, 111
gęstość spektralna, 409

H

heteroskedastyczność, 347
hipoteza, 207
histogram, 78
homoskedastyczność, 347

I

iloraz szans, 314
inercja, 437

K

kombinacja, 133
kontrast, 261
korelacja
— cząstkowa, 323
— kanoniczna, 323
— wieloraka, 323
korelogram, 406
kowariancja, 318
kryterium Chauveneta, 299
krzywa charakterystyczna,
374
kurtoza, 118
kwantyl, 105
kwartył
— pierwszy, 105
— trzeci, 105
kwartyłowy współczynnik
skośności, 116

L

lasso, 363
lasy losowe, 458

Ł

ładunki, 425

M

mediana, 103
metoda
— clara, 448
— grupowania wokół centro-
idów, 448
— k-średnich, 444
— najbliższego sąsiada, 452
— najmniejszych kwadra-
tów, 342
— wektorów nośnych, 456
moc, 208
moda, 101

N

nierówność Czebyszewa, 167
normalizacja, 420

O

obserwacja, 64
ocena standardowa, 114
odchylenie
— ćwiartkowe, 107
— medianowe, 110
— przeciętne, 110
— standardowe, 111
odległość
— Cooka, 346
— Czekanowskiego, 434
— Jaccarda, 434
— Mahalanobisa, 228

P

p-wartość, 210
paradoks Simpsona, 306
percentyl, 105
periodogram, 409
permutacja, 133
poprawka na ciągłość, 164
populacja, 64
poziom
— istotności, 198, 208

— ufności, 198
 prawo Benforda, 281
 prognoza, 349
 próba, 64
 przycinanie, 454

R

regresja
 — częściowych najmniejszych kwadratów, 363
 — grzbietowa, 363
 — logistyczna, 373
 — lokalnie wielomianowa, 371
 — najbliższych sąsiadów, 371
 — odporna, 362
 — składowych głównych, 362
 — wielokrotna, 357
 reguła
 — kciuka, 166
 — trzech sigm, 166
 rozkład
 — χ^2 , 169
 — *t*-Studenta, 170
 — arcusa sinusa, 159
 — beta, 158
 — Cauchy'ego, 171
 — Cauchy'ego standardowy, 172
 — dwumianowy, 149, 164
 — dwupunktowy, 148
 — empiryczny, 84
 — Erlanga, 160
 — F-Snedecora, 172
 — gamma, 160
 — geometryczny, 152
 — hipergeometryczny, 155
 — jednostajny, 158
 — logarytmiczno-normalny, 164
 — normalny, 161
 — normalny standardowy, 162
 — Pascala, 154
 — Poissona, 151, 164
 — Rayleigha, 169
 — równomierny, 148

— standardowy jednostajny, 158
 — studentyzowanego rozstępu, 259
 — ucięty, 173
 — ujemny dwumianowy, 154
 — Weibulla, 168
 — wielomianowy, 157
 — wielowymiarowy
 — — hipergeometryczny, 156
 — — normalny, 167
 — wykładniczy, 159
 — zero-jedynkowy, 148
 rozstęp, 107
 — międzykwartyłowy, 109

S

sieci neuronowe, 455
 siła dźwigni, 345
 skala
 — ilorazowa, 65
 — nominalna, 65
 — porządkowa, 65
 — przedziałowa, 65
 skalowanie wielowymiarowe, 433
 standaryzacja, 420
 stymulanta, 359
 szereg czasowy, 387

Ś

średnia
 — arytmetyczna, 92
 — chronologiczna, 388
 — generalna, 255
 — geometryczna, 97
 — harmoniczna, 95
 — Herona, 124
 — międzykwartyłowa, 104
 — potęgowa, 98
 — ruchoma, 404
 — ucięta, 104
 — w populacji, 91
 — winsorowska, 104
 — z próby, 91

T

tablica kontyngencji, 312

test

— χ^2

— — niezależności, 312

— — Yatesa, 329

— — zgodności, 217

— analizy

— — kowariancji, 254

— — wariancji, 253

— Andersona-Darlinga, 225

— Ansariiego-Bradleya, 238

— Bartletta, 240

— Bowkera, 277

— Breuscha-Godfrey, 403

— Breuscha-Pagana, 347

— Browna-Forsythe'a, 238

— Cochрана-Mantela-

Haenszela, 314

— Craméra-Smirnowa, 224

— Craméra-von Misesa, 224

— D'Agostino, 225

— Dickeya-Fullera, 406

— Dixona, 299

— dla dwóch wskaźników
struktury, 274

— dla wariancji, 234

— dla wskaźnika struktury,
272

— dokładny, 248

— — Fishera, 317

— Duncana, 259

— Durbina-Watsona, 403

— dwumianowy, 273

— F, 237

— Flignera-Killeene'a, 240

— Friedmana, 265

— G, 218

— Grubbsa, 300

— Hartleya, 237

— HSD Tukeya, 259

— Jarque-Bery, 225

— kolejności par Wilcoxon,
251

— Kolmogorowa-Lillieforsa,
224

— Kolmogorowa-Smirnowa,
222

— Kolmogorowa-Smirnowa,
231

— Kruskala-Wallisa, 264

— Kuipera, 231

— Levene, 237

— Ljunga-Boxa, 403, 414

— LSD Fishera, 258

— Manna-Whitneya, 229

— McNemary, 277

— Mooda, 238

— Mosesa, 238

— Newmana-Keulsa, 259

— NIR Fishera, 258

— permutacyjny, 247

— — losowy, 248

— post hoc, 257

— Scheffého, 259

— serii, 279

— Shapiro-Francia, 224

— Shapiro-Wilka, 224

— Siegela-Tukeya, 238

— t

— — dla dwóch prób, 242

— — dla jednej próby, 232

— — dla prób zależnych, 243

— Welcha, 243

— Woolfa, 314

— z, 231

— znaków, 249

triangulacja Delaunaya, 453

twarze Chernoffa, 439

U

uczenie

— bez nauczyciela, 442

— z nauczycielem, 442

W

wariacja, 133

wariancja, 109

wielkość efektu, 212

wielobok liczebności, 79

- wskaźnik
 - asymetrii, 115
 - asymetrii Pearsona, 116
- współczynnik
 - D Somersa, 326
 - dopasowania, 319
 - gamma Kruskala i Goodmana, 326
 - informacyjny Akaike, 360
 - informacyjny bayesowski, 361
 - Kendalla, 325
 - kontyngencji, 332
 - korelacji
 - — Craméra, 331
 - — Pearsona, 318
 - — Spearmana, 324
 - — Yule'a, 329
 - podobieństwa Gowera, 434
 - Q Yule'a, 327
 - tau Goodmana i Kruskala, 333
 - zarysu, 446
 - zbieżności, 345
 - zgodności
 - — Cohena, 334
 - — Fleissa, 334
 - zmienności, 113
- wykres
 - „łodyga-liście”, 82
 - balonowy, 308
 - bąbelkowy, 472
 - chi, 307
 - efektów, 270
 - gwiazdowy, 440
 - interakcji, 270
 - kołowy, 83
 - koniczyny, 331
 - konturowy, 473
 - kropkowy, 83
 - krzywych Andrewsa, 423
 - kubełkowy, 474
 - kwadratów, 83
 - kwantylowy, 226
 - mozaikowy, 308
 - obrazkowy, 439
 - osypiska, 425
 - pareto, 83
 - paskowy, 81
 - perpektywiczny, 473
 - piramidowy, 83
 - przebiegu, 80
 - pudełkowy, 80
 - punktowo-liniowy, 473
 - radarowy, 440
 - Sheparda, 435
 - skojarzeń, 310
 - skrzypcowy, 81
 - słonecznikowy, 309
 - słupkowy, 83
 - ślad gęstości, 79
 - torbowy, 466
 - trójwymiarowy rozrzutu, 474
 - współrzędnych równoległych, 423
 - zarysu, 446
 - wyniki, 425
- Z**
- zarys, 446
- zmienna
 - pozorna, 54
 - utajona, 427

Angielski indeks pojęć

3

3d scatter plot, 474

A

Akaike's information criterion, 360

analysis of

— covariance, 254

— variance, 253

Anderson-Darling test, 225

Andrews' curves, 423

Ansari-Bradley test, 238

antimode, 102

arc sine distribution, 159

arithmetic mean, 92

association plot, 310

augmented Dickey-Fuller test, 406

autocorrelation, 402

autocovariance, 403

average deviation, 110

B

bagging, 457

bagplot, 466

balloon plot, 308

bar plot, 83

Bartlett's test, 240

Bayesian information criterion, 361

Benford's law, 281

Bernoulli distribution, 148

beta distribution, 158

binomial

— distribution, 149

— test, 273

biplot, 425

boosting, 457

Bowker's test, 277

box-and-whisker plot, 80

boxplot, 80

Breusch-Godfrey test, 403

Breusch-Pagan test, 347

Brown-Forsythe test, 238

bubble plot, 472

C

canonical correlation, 323

Cauchy distribution, 171

centile, 105

Chauvenet's criterion, 299

Chebyshev's inequality, 167

Chernoff's faces, 439

chi-plot, 307

chi-square

— distribution, 169

— test for independence, 312

— test for variance, 234

chronological mean, 388

cluster analysis, 443

clustering large applications, 448

Cochran-Mantel-Haenszel test, 314

coefficient of

- determination, 319
- indetermination, 345
- skewness, 115
- variation, 113

Cohen's Kappa, 334

confidence level, 198

contingency

- coefficient C, 332
- table, 312

contour plot, 473

contrast, 261

Cook's distance, 346

correction for continuity, 164

correlation coefficient, 318

correlogram, 406

correspondence analysis, 437

covariance, 318

Cramér-von Mises test, 224

Cramér's V, 331

D

D'Agostino's K^2 test, 225

decile, 105

decision trees, 454

Delaunay triangulation, 453

dendrogram, 443

density trace, 79

dependent t-test, 243

deviance, 361

discrete uniform distribution, 148

Dixon's test, 299

dotchart, 83

dummy variable, 54

Duncan's test, 259

Durbin-Watson test, 403

E

effect, 254

effect size, 212

empirical

- distribution function, 84
- rule, 166

entity, 64

Erlang distribution, 160

estimator, 191

exact test, 248

excess, 118

exponential distribution, 159

F

F-distribution, 172

F-test, 237

factor analysis, 427

Fisher's exact test, 317

Fleiss' kappa, 334

Fligner-Killeen test, 240

fourfold plot, 331

fractile, 105

frequency polygon, 79

Friedman test, 265

G

G test, 218

gamma distribution, 160

Gantt chart, 467

geometric

- distribution, 152
- mean, 97
- standard deviation, 111

glyphs, 439

Goodman and Kruskal's Gamma, 326

Goodman-Kruskal tau, 333

Gower's similarity coefficient, 434

grand mean, 255

Grubbs' test, 300

H

hanging rootogram, 78

harmonic mean, 95

Hartley's test, 237

hat

- matrix, 345
- value, 345

Heronian mean, 124

heteroskedasticity, 347
 hexbin plot, 474
 hinge, 107
 histogram, 78
 homoscedasticity, 347
 Honestly Significant Differences, 259
 hypergeometric distribution, 155
 hypothesis, 207

I

independent components analysis, 425
 inertia, 437
 interquartile
 — mean, 104
 — range, 107

J

Jarque-Bera test, 225

K

k-means method, 444
 Kendall's τ coefficient, 325
 Kolmogorov-Lilliefors test, 224
 Kolmogorov-Smirnov test, 222
 Kruskal-Wallis test, 264
 Kuiper's test, 231
 kurtosis, 118

L

lasso, 363
 latent variable, 427
 Least Significant Differences, 258
 Levene's test, 237
 leverage, 345
 likelihood function, 195

linear discriminant analysis, 449
 link function, 372
 Ljung-Box test, 403, 414
 loadings, 425
 locally weighted polynomial regression, 371
 log-normal distribution, 164
 logistic regression, 373
 lower quartile, 105

M

Mahalanobis distance, 228
 Mann-Whitney test, 229
 McNemar's test, 277
 mean absolute deviation, 110
 median, 103
 — absolute deviation, 110
 mode, 101
 Mood test, 238
 mosaic plot, 308
 Moses test, 238
 moving average, 404
 multidimensional scaling, 433
 multinomial distribution, 157
 multiple
 — correlation, 323
 — correspondence analysis, 438
 — regression, 357
 multivariate
 — hypergeometric distribution, 156
 — normal distribution, 167

N

nearest neighbor
 — method, 452
 — regression, 371
 negative binomial distribution, 154
 neural networks, 455
 Newman-Keuls test, 259
 non-sampling error, 68
 normal distribution, 161
 normalization, 420

O

observation, 64
odds ratio, 314
one-proportion test, 272
one-sample t-test, 232
ordinary least squares, 342

P

p-value, 210
parallel coordinates plot, 423
Pareto diagram, 83
partial
— autocorrelation, 406
— correlation, 323
— least squares regression, 363
partitioning around medoids, 448
Pascal distribution, 154
Pearson χ^2 test of goodness of fit, 217
Pearson correlation coefficient, 318
Pearsonian coefficient of skewness, 116
percentile, 105
periodogram, 409
permutation test, 247
perspective plot, 473
pie chart, 83
Poisson distribution, 151
population, 64
— mean, 91
— pyramid, 83
post-hoc, 257
power, 208
— mean, 98
principal
— component analysis, 424
— component regression, 362
— coordinates analysis, 434
pruning, 454

Q

Q-Q plot, 226

quadratic discriminant analysis, 449

quartile

— deviation, 107
— skewness coefficient, 116

R

radar plot, 440
random
— error, 297
— forests, 458
— permutation test, 248
range, 109
Rayleigh distribution, 169
receiver operating characteristic, 374
regularized discriminant analysis, 450
ridge regression, 363
robust regression, 362
rule of thumb, 166
run chart, 80
runs test, 279

S

sample, 64
— mean, 91
sampling error, 68
scale
— interval, 65
— nominal, 65
— ordinal, 65
— ratio, 65
scatterplot, 307
scores, 425
scree plot, 425
Shapiro-Francia test, 224
Shapiro-Wilk test, 224
Shepard plot, 435
Siegel-Tukey test, 238
sign test, 249
significance level, 198, 208
silhouette, 446
— coefficient, 446
— plot, 446

Simpson's paradox, 306
Smirnov-Kolmogorov test,
231
Somers' D, 326
Spearman's correlation coefficient, 324
spectral density, 409
squareplot, 83
standard
— Cauchy distribution, 172
— deviation, 111
— normal distribution, 162
— uniform distribution, 158
standardization, 420
star plot, 440
stem and leaf plot, 82
stemplot, 82
stress function, 433
stripchart, 81
Student's t-distribution, 170
Studentized range distribution, 259
sunflower plot, 309
supervised classification, 442
support vector machines, 456
survival analysis, 168
systematic error, 298

T

time series, 387
trimmed mean, 104
truncated
— distribution, 173
— mean, 104
two-proportion test, 274
two-sample t-test, 242

U

uniform distribution, 158
unsupervised classification,
442
upper quartile, 105

V

variable, 64
variance, 109
violin plot, 81
Voronoi diagram, 452

W

Weibull distribution, 168
Welch's t test, 243
white noise, 404
Wilcoxon signed-ranks test,
251
Winsorized mean, 104
wire-frame, 473
Woolf's test, 314

Y

Yates' chi-square test, 329
Yule's correlation coefficients,
329
Yule's Q, 327

Z

z-score, 114
z-test, 231

Indeks nazwisk

A

Achenwall, 57
Akaike, 360
Anchersen, 59
Anderson, 225
Ansari, 238
Arbuthnot, 206

B

Bartlett, 240
Bayes, 143
Benford, 281
Benzécri, 437
Bera, 225
Bernoulli, 62, 207
Bielfeld, 57
Bonferroni, 215
Boole, 214
Bowker, 277
Box, 420
Bradley, 238
Breit, 171
Brown, 238

C

Cardano, 62
Cattell, 185
Cauchy, 171
Chauvenet, 299
Chernoff, 439
Cochran, 314

Cohen, 334
Cook, 346
Cox, 379, 420
Cramér, 224
Cramér, 331
Czebyszew, 167
Czekanowski, 434

D

Darling, 225
DeLaunay, 453
Desgenettes, 61
Diaconis, 74
Dickey, 406
Dirichlet, 452
Dixon, 299
Durbin, 403

E

Erlang, 160

F

Fermat, 62
Fisher, 172, 195, 253, 258, 394
Fleiss, 334
Fligner, 240
Forsythe, 238
Francia, 224
Friedman, 265
Fuller, 406

G

Galton, 318, 337, 339
Gantt, 467
Gauss, 161, 339
Gini, 119
Goodman, 326
Gosset, 170
Gower, 434
Graunt, 59
Grubbs, 300

H

Haenszel, 314
Halley, 60
Hebb, 456
Heron, 124
Hotelling, 323

J

Jaccard, 434
Jarque, 225

K

Kendall, 325
Killeen, 240
Kirgiłow, 59
Kolmogorow, 132, 222, 224,
231
Kruskal, 264, 326
Kuiper, 231

L

Laplace, 61, 130, 207
Laspeyres, 393
Legendre, 339
Levene, 237
Lilliefors, 224
Lorentz, 119, 171

M

Münchhausen, 214

Mahalanobis, 228
Mann, 229
Mantel, 314
McCulloch, 455
McNemar, 277
Mises, 131, 224
Moivre, 130, 161
Mood, 238
Moses, 238

N

Napoleon, 61
Newcomb, 281

P

Paasche, 393
Pareto, 83
Pascal, 62, 154
Pearson, 116, 195, 207, 217,
312, 318, 332
Petty, 59
Pitagoras, 100
Pitts, 455
Poisson, 151

Q

Quételet, 61

R

Rayleigh, 169
Richter, 421

S

Scheffé, 259
Shapiro, 224
Shepard, 435
Sheppard, 110
Siegel, 238
Simpson, 306
Smirnow, 224, 231
Snedecor, 172

Somersa, 326
Spearman, 324, 326, 428
Splawa-Neyman, 62, 191, 207
Staszic, 58
Steinhaus, 62

T

Tippett, 189
Tukey, 107, 238, 259

V

Venn, 40
Voronoi, 452

W

Wald, 279
Wallis, 264
Watson, 403
Weibull, 168
Welch, 243
Whitney, 229
Wigner, 171
Wilcoxon, 229, 251
Wilk, 224
Wold, 363
Wolfowitz, 279
Woolf, 314

Y

Yates, 329
Yule, 327, 329

Indeks poleceń R

.

.First, 49
.Last, 49
.Last.value, 49
.Machine, 49
.Platform, 49
:, 18

A

abline, 29, 228
abs, 41, 112, 301
acf, 414
acos, 41
acosh, 41
ad.test, 228
adaboost.M1, 457
adaptIntegrate, 465
add1, 362
addmargins, 20
addPoints, 453
adf.test, 412
aggregate, 388
agostino.test, 228
AIC, 362, 414
all, 38
ancova, 254
andrews.curves, 423
anova, 257, 270, 381
ansari.test, 239
any, 38
aov, 254, 257, 264
apply, 22
apropos, 13
aq.plot, 301
Arg, 464
args, 14
arima, 414
arima.sim, 408
armasubsets, 407
array, 20, 23, 316
arrows, 29
as.array, 463
as.character, 24, 463
as.complex, 463
as.data.frame, 310, 397, 463
as.Date, 389
as.double, 463
as.EventData, 453
as.expression, 463
as.factor, 376, 463
as.integer, 463
as.list, 463
as.logical, 24, 463
as.matrix, 228, 364, 463
as.numeric, 24, 423, 463
as.table, 331, 334
as.ts, 463
as.vector, 129, 463
asin, 41
asinh, 41
assocplot, 310
atan, 41
atan2, 41
atanh, 41
attach, 23, 40, 371
auto.arima, 414
axis, 29

B

bagging, 457
bagplot, 466
balloonplot, 310
barchart, 470
barplot, 83, 86, 87, 470
bartlett.test, 241
best.nnet, 459
best.randomForest, 459
best.rpart, 459
best.svm, 459
beta, 158
bgtest, 403
binom.test, 274
biplot, 364, 426
Biplots, 426
bmp, 35
box, 159
box.cox.powers, 421
Box.test, 414
boxcox, 421
boxplot, 81, 86, 262, 470
bptest, 357
break, 39
browser, 47
bwplot, 470, 472
by, 20

C

c, 17
ca, 438
calcArea, 453
calcVoronoi, 453
cancor, 324
capabilities, 35
cards, 144
cascadeKM, 445
cat, 45
cbind, 336, 357
cdplot, 376
ce.impute, 423
ceiling, 41
centipede.plot, 262
character, 17
chartr, 16

chiplot, 310
chisq.detail, 313
chisq.test, 219, 221, 222, 313, 316, 438
chol, 21
choose, 41
chplot, 466
chull, 466
clara, 448
class, 45
cloud, 470, 474
clt.examp, 163
clusplot, 448
cm.colors, 34
cmdscale, 435
coef, 357, 376
colMeans, 21, 228
colnames, 25, 432
color2D.matplot, 320
colors, 33
colSums, 21
combn, 41
complete.cases, 423
complex, 17, 464
condense, 454
confint, 316, 357
conflicts, 40
Conj, 464
constrOptim, 465
contour, 79, 470, 473
contourplot, 470
cooks.distance, 357
coplot, 471
cor, 320, 322, 328, 329, 432
cor.test, 322
corresp, 438
corrgram, 364
corrplot, 364
cos, 28, 41
cosh, 41
count.fields, 25
cov, 228
cox.zph, 380
coxph, 381
cumprod, 41
cumsum, 41, 86
curve, 28, 159

cut, 20
cutree, 448
cvm.test, 228
cycle, 388

D

D, 464
daisy, 435
data, 15, 282, 423
data.entry, 18
data.frame, 20, 23, 257, 357,
376, 381, 391, 453
dbeta, 159
dbinom, 149, 150, 165
debug, 46
debugger, 46
decompose, 405
deltat, 388
demo, 13
density, 79, 86, 164, 172, 470
densityplot, 470, 472
descdist, 197
det, 21
detach, 15, 23
dev.off, 35
devAskNewPage, 34
deviance, 260, 362, 369
df.residual, 260
dgeom, 153
dhyper, 156, 317
diag, 21, 167
diff, 41, 86, 122, 222, 391, 405
diffinv, 405
difftime, 389
dim, 20, 317, 331, 334, 397
dimnames, 25, 331, 334, 397,
453
dist, 435, 468
dist.binary, 435
dixon.test, 301
dmultinom, 157
dmvnorm, 167
dnbinom, 155
dnorm, 176, 465
dotchart, 83, 470
dotplot, 470

DOTplot, 81
dpois, 152, 221
drop1, 362
duplicated, 18
durbin.watson, 412

E

eacf, 414
ecdf, 85
ecdfplot, 471
edit, 26
eigen, 21, 432
end, 388
enterTable, 331, 334
equal.count, 470, 472
Error, 264
errorest, 454
eval, 24, 464
evaluate_Weka_classifier, 468
example, 13
exp, 24, 28, 41, 45, 98, 316,
376, 465
expand.grid, 270
expression, 24, 29, 149, 464

F

fa.graph, 432
factanal, 431, 432
factor, 19, 264, 389
factorial, 133
fanny, 448
fastICA, 427
file.choose, 26
filled.contour, 79, 473
filter, 400
find, 13
fisher.test, 317
fitdist, 197
fitdistr, 197
fitted, 357, 369, 456
fivenum, 107
fix, 26
fligner.test, 241
floor, 41

for, 39, 47, 149, 185, 249, 296
forecast, 415
format, 389
fourfoldplot, 331
frequency, 388
friedman, 266
friedman.test, 268
ftable, 20
function, 21, 22, 39, 40, 44, 45,
112, 295, 391, 399,
454, 465, 478

G

g.test, 219
gamma, 160
gantt.chart, 467
garch, 405
garch.sim, 405
geometric.mean, 98
get.segs, 262
getAnywhere, 36
getwd, 27
gini.index, 121
gl, 129, 262
glm, 373, 375, 377
glob2rx, 16
goodfit, 79
granova.lw, 257
grep, 16
grubbs.test, 301
gsub, 16
guiDlgFunction, 478
guiDlgList, 478
gvlma, 357

H

harmonic.mean, 97
hatvalues, 357
hclust, 448
head, 15
heat.colors, 34
heatmap, 320
help, 13
help.search, 13

hexbin, 474
hist, 79, 80, 86, 185, 470
histogram, 470, 472
history, 49
HoltWinters, 400
HSD.test, 260

I

identify, 34, 448
if, 37, 45
if else, 37
ifelse, 37, 164
iidspace, 144
Im, 464
image, 79, 470, 473
importance, 458
index.G1, 445
Inf, 222
influencePlot, 357
install.packages, 15
integer, 17
integrate, 465
interaction.plot, 270
intersect, 40
invisible, 43
ipredknn, 454
IQR, 107
is.array, 463
is.character, 463
is.charcter, 24
is.complex, 463
is.data.frame, 463
is.double, 463
is.element, 40
is.expression, 463
is.factor, 463
is.finite, 19
is.infinite, 19
is.integer, 463
is.list, 463
is.logical, 463
is.matrix, 463
is.na, 19
is.numeric, 24, 463
is.ts, 463
is.vector, 463

isoMDS, 436
isTRUE, 37

J

jarque.bera.test, 228
jitter, 79, 310
jpeg, 35

K

kappa2, 336
kappam.fleiss, 336
kde2d, 80
kmeans, 448
kronecker, 21
kruskal, 266
kruskal.test, 266
ksmooth, 371
kurtosis, 118

L

lapply, 22
lars, 364
layout, 30
lda, 450
LDuncan, 260
legend, 29, 149
length, 17, 41, 86, 98, 112,
239, 241, 249, 260,
266, 295, 296, 301,
391, 399, 412, 453,
458
letters, 129
LETTERS, 129
levelplot, 470
levels, 20, 472
levene.test, 239, 241
library, 14, 15, 228, 239, 251,
440, 478
lillie.test, 228
lines, 29, 80, 86, 172, 369, 371,
400, 436, 466
list, 22, 43, 241, 266, 268, 310,
331, 334, 397, 408,
472, 474

lm, 257, 262, 270, 353, 357,
359, 362, 366, 412
lm.ridge, 364
load, 27, 46
loadingplot, 364
loadings, 432
locator, 34
log, 24, 41, 44, 98, 353, 360,
464
log10, 41
log2, 41
logical, 17
logLik, 369
lorenz.curve, 121
lower.tri, 286
lowess, 371
lqs, 362
ls, 36
LSD.test, 260

M

mad, 124
mahalanobis, 228
make_Weka_classifier, 468
makeCoCo, 331, 334
makeProps, 453
manova, 263
mantelhaen.test, 316
map, 475
map.cities, 475
margin.table, 20, 310
mars, 372
match.arg, 43, 295
matplot, 149, 357
matrix, 20, 31, 44, 149, 268,
296, 313
max, 41, 80, 86
mca, 439
mean, 20, 23, 39, 93, 97, 98,
104, 180, 221, 226,
249, 296, 301, 399
median, 39, 107, 295
memory.limit, 49
memory.size, 49
merge, 23
methods, 36, 44

mfg, 30
 min, 41, 80, 86
 missing, 43, 295
 mnorm, 420
 Mod, 464
 moda, 103
 mode, 17
 model.frame, 295
 month.abb, 129
 month.name, 129
 months, 389
 mood.test, 239
 mosaicplot, 310
 mshapiro.test, 228
 mst, 468
 mtext, 29
 multiedit, 454
 mvr, 363

N

na.omit, 423, 432
 NaiveBayes, 452
 names, 25, 43
 nchar, 16
 nclass.FD, 75
 nclass.scott, 75
 nclass.Sturges, 75
 ncol, 41, 228, 432, 473
 next, 39
 nls, 368
 nlsContourRSS, 369
 nlsResiduals, 369
 nnet, 456
 normarea, 212
 npudens, 80
 nrow, 41, 228, 362, 432, 458,
 473
 nScee, 432
 nsize, 203
 numeric, 17, 226, 391, 399

O

objects, 15
 oddsratio, 316

onet.permutation, 249
 optim, 465
 optimize, 465
 options, 46, 50, 216
 order, 24, 40
 ordered, 20
 outer, 21, 129
 outlier, 301

P

p.adjust, 215
 pacf, 414
 package.skeleton, 47
 pairs, 310, 364, 470
 pairwise.prop.test, 273
 pairwise.t.test, 260
 pairwise.wilcox.test, 266
 palette, 33
 pam, 448
 panel.bwplot, 471
 par, 30, 34, 159, 262, 436
 parallel, 432, 470
 parcoord, 423, 470
 partial.cor, 323
 partialAssociations, 331, 334
 paste, 16, 129, 149
 pbinom, 151, 165
 pcr, 364
 pdf, 35
 performance, 376
 permtest, 249
 persp, 80, 473
 pexp, 161
 pie, 87
 pie3D, 83, 87
 plot, 28, 34, 44, 79, 80, 86,
 164, 176, 310, 357,
 360, 368, 375, 380,
 400, 414, 415, 426,
 432, 435, 438, 439,
 448, 452, 454, 459,
 466, 468, 469, 470,
 472, 474, 476
 plot.design, 270
 plot.new, 172
 plot.window, 172

plotBenfordsLaw, 282
 plotcorr, 320
 plotfit, 369
 plotMap, 453
 plotnScree, 432
 plotuScree, 432
 plsr, 364
 pmax, 41
 pmin, 41
 pmvnorm, 167
 pnbinom, 155
 png, 35
 pnorm, 165, 166, 176, 222, 301
 points, 29, 448
 poly, 29
 polyroot, 465
 postscript, 35
 power.anova.test, 214
 power.prop.test, 214
 power.t.test, 214
 ppoints, 228
 ppois, 151, 166, 221
 ppr, 372
 prcomp, 426, 468
 predict, 357, 376, 400, 414,
 451, 452, 454, 456,
 458
 prediction, 376
 priceIndex, 397
 princomp, 426
 print, 15, 38, 41, 45, 49, 432,
 458
 prob, 144
 prod, 41
 prop.table, 20
 prop.test, 273, 275, 276
 prune.tree, 454
 ptukey, 260
 punif, 159
 pyramid, 84
 pyramid.plot, 84

Q

q, 13
 qbinom, 251
 qchisq, 228

qda, 450
 qdixon, 301
 qgrubbs, 301
 qmvnorm, 167
 qnorm, 176, 212
 qq, 470
 qq.plot, 228
 qqline, 228
 qqmath, 470
 qqnorm, 228
 qqplot, 228, 470
 qr, 21
 qsignrank, 253
 qt, 206
 qtukey, 260
 quantile, 107, 296, 453
 quantityIndex, 397
 quarters, 389
 qunif, 159
 quote, 46
 qwilcox, 230

R

radial.plot, 440
 rainbow, 34, 86
 randomForest, 458
 range, 86
 rank, 324
 rbind, 31
 rbinom, 79, 180
 Rcmdr, 14
 rda, 450
 Re, 464
 read.csv, 25
 read.table, 25
 readShapePoints, 476
 readShapePoly, 476
 recode, 53
 rect, 29
 rect.hclust, 448
 rep, 18, 20, 84, 112, 129, 221,
 222, 239, 257, 260,
 264, 266, 296, 336,
 378, 400
 repeat, 39
 replicate, 22, 23

- require, 15
 - residuals, 357, 405
 - return, 43, 399
 - rev, 24, 41, 476
 - rexp, 185
 - rgb, 34, 453
 - rk4, 464
 - rlm, 362
 - rm, 36
 - rmvnorm, 167
 - RNGkind, 189
 - rnorm, 23, 164, 176, 182, 189, 197, 228, 466, 474
 - rolldie, 144
 - rootogram, 79
 - round, 41, 86, 93, 97, 273, 389, 458
 - row.names, 448
 - rowMeans, 21
 - rownames, 17, 25, 435
 - rowSums, 21
 - rpart, 360
 - rpois, 310
 - Rprof, 47
 - RSiteSearch, 13
 - rstandard, 357
 - rstudent, 357
 - rt, 262
 - rug, 79
 - runif, 172
 - runs.test, 281, 287
- S**
- sammon, 436
 - sample, 129, 164, 249, 295, 296, 451, 458
 - sapply, 22, 23, 43
 - save, 27
 - save.image, 27, 49
 - saveMovie, 477
 - saveSWF, 477
 - scale, 420
 - scan, 24
 - scatter.smooth, 371
 - scatterplot, 310
 - scatterplot.matrix, 310
 - scoreplot, 364
 - sd, 112, 301
 - search, 23
 - segments, 29, 172
 - select, 364
 - seq, 18, 113, 149, 176, 357, 389, 405, 453, 473
 - set.seed, 189
 - setClass, 45
 - setdiff, 40
 - setequal, 40
 - setwd, 27
 - sf.test, 228
 - shapiro.test, 226, 228
 - Shepard, 436
 - shingles, 470
 - SIGN.test, 251
 - silhouette, 448
 - simple.hist.and.boxplot, 81
 - sin, 22, 24, 28, 41
 - sinh, 41
 - skewness, 117
 - smc, 323
 - smooth.spline, 371
 - solve, 21
 - sort, 41, 83, 451, 475
 - source, 42
 - spectrum, 414
 - spline, 371
 - split, 23, 262
 - split.screen, 30
 - splom, 470
 - spplot, 476
 - sqrt, 38, 41, 86, 112, 165, 185, 206, 222, 234, 236, 323, 465
 - squareplot, 83, 87
 - SSasymp, 369
 - SSasympOff, 369
 - SSasympOrig, 369
 - SSbiexp, 369
 - SSfol, 369
 - SSfpl, 369
 - SSgompertz, 369
 - SSlogis, 368, 369
 - SSmicmen, 368, 369
 - SSweibull, 369

stars, 440
 start, 388
 stem, 83
 stem.leaf, 90
 step, 362
 stl, 414
 stop, 44
 str, 24
 stripchart, 81, 470
 stripplot, 470
 strsplit, 16
 strtrim, 16
 sub, 16
 subset, 26, 144
 sum, 21, 41, 47, 112, 148, 150,
 180, 234, 236, 317,
 451, 458
 summary, 41, 257, 264, 357,
 359, 360, 366, 368,
 375, 378, 380, 382,
 412, 426, 454, 456
 summaryBy, 20
 summaryProf, 47
 sunflowerplot, 310
 SuppDists, 177
 suppressWarnings, 44
 supsmu, 371
 Surv, 378, 380
 survdiff, 380
 survfit, 378, 380
 survreg, 382
 svd, 21
 svm, 456
 switch, 38
 symbols, 472
 symbolTable, 32
 Sys.Date, 389
 Sys.sleep, 185
 Sys.time, 389
 system.time, 47

T

t, 21, 453
 t.test, 233, 247, 249, 478
 table, 19, 221, 222, 389, 451,
 452, 456, 458

tail, 15
 tan, 41, 172, 464
 tanh, 41
 tapply, 19
 terms, 352
 terrain.colors, 34, 473
 test.nlsResiduals, 369
 text, 29, 34, 360, 435, 448, 454
 tiff, 35
 time, 388
 title, 29, 34
 tolower, 16
 topo.colors, 34, 476
 tosscoin, 144
 toupper, 16
 tr, 21
 tree, 360, 454
 try, 15, 44
 ts, 388, 414
 ts.intersect, 388
 ts.plot, 388
 ts.union, 388
 tsdia, 414
 tsum.test, 233, 245, 246
 TukeyHSD, 260
 tune.control, 459
 tune.knn, 459
 tune.nnet, 459
 tune.randomForest, 459
 tune.rpart, 459
 tune.svm, 459
 twot.permutation, 249
 typeof, 17

U

undebug, 47
 union, 40
 unique, 18, 389
 uniroot, 465
 unlist, 463
 unsplit, 23
 update, 353
 update.packages, 14
 upper.tri, 286
 urnsamples, 144
 UseMethod, 44

V

v.test, 231
var, 112, 426
var.test, 239
vcov, 357
vector, 17
venn, 40
vif, 363
vioplot, 82

W

warning, 44
weekdays, 389
weighted.mean, 94, 112, 234,
236, 399
Weka.control, 468
which, 18
while, 39
wilcox.test, 230, 253
window, 388
windows, 28
wireframe, 470, 473
with, 23, 270, 472
woolf.test, 316
WOW, 468
write.csv, 26
write.table, 26

X

xyplot, 470, 471, 472

Z

z.test, 233, 478
znorm, 420
zsum.test, 233, 244

Indeks pakietów R

A

adabag, 457
ade4, 435
agricolae, 260, 266
animation, 476
ape, 468
aplpack, 90, 440, 466
— UKDriverDeaths, 90

B

base, 465
BHH2, 249
BiplotGUI, 426
BSDA, 212, 251

C

ca, 438
car, 53, 227, 357, 412, 421
chplot, 466
class, 454
cluster, 435, 448
clusterSim, 445
CoCo, 331
corrgram, 364
cubature, 465

D

DAAG, 249, 363
debug, 46
doBy, 20

dprep, 103, 420, 423

E

e1071, 117, 118, 452, 456, 459
ellipse, 320
elliptic, 465
epicalc, 84

F

faraway, 365
fastICA, 427
fBasics, 32
fitdistrplus, 197
forecast, 414
foreign, 26

G

gap, 465
ggplot2, 31
GPArotation, 431
gplots, 40, 310
granova, 257
grDevices, 466
gvlma, 357

H

hexbin, 474
HH, 254
HiddenMarkov, 465

I

ipred, 454, 457
irr, 336
ISwR, 380

K

kernlab, 456
klaR, 450, 452, 456

L

laercio, 260
lasso2, 21
lattice, 31, 473, 474
latticeExtra, 471
lawstat, 121, 281
linprog, 465
lmtest, 357, 403

M

mapdata, 474
maps, 474
mapproj, 476
MASS, 16, 80, 197, 360, 362,
364, 372, 421, 423,
436, 438, 439, 450
mda, 372, 451
methods, 45
micEcon, 397
mlmRev, 472
mvnormtest, 228
mvoutlier, 301
mvtnorm, 167

N

neural, 456
nFactors, 431
nlrwr, 368
nlstools, 369
nnet, 456
np, 80
NRAIA, 369

O

odesolve, 464
orthopolynom, 465
outliers, 301

P

party, 469
PASWR, 203, 233
PBSmapping, 453
plm, 74
plotrix, 83, 84, 262, 320, 440,
467
pls, 363
pmg, 14
polynom, 465
prob, 144
psych, 97, 98, 323
pwr, 214

Q

quadprog, 465

R

R.matlab, 26
randomForest, 458
rattle, 282
Rcmdr, 14, 323
RcmdrPlugin.IPSUR, 123
— RcmdrTestDrive, 123
rggobi, 34
rgl, 34
ROCR, 376
rootogram, 79
rpanel, 477
rpart, 360, 454
RWeka, 468

S

sem, 430
sp, 74, 476
spdep, 74
survival, 378, 382

svDialogs, 477
svmpath, 456

T

TeachingDemos, 163, 313,
476, 478
tree, 360, 454
truncgof, 231
TSA, 405, 407, 414
tseries, 287, 405, 412

U

UsingR, 81, 83, 371

V

vcd, 79, 316
vegan, 445
vioplot, 82

Indeks zbiorów danych R

A

alr3

— brains, 386

audit

— rattle, 282

B

boot

— bigcity, 386

C

car

— Duncan, 421

— US.pop, 384

D

DAAG

— ais, 478

— head.injury, 461

— leafshape, 461

— litters, 385

— Manitoba.lakes, 89

— monica, 462

— moths, 386

— nsw74psid1, 385

— orings, 89

— possum, 88, 123, 338, 478

— rareplants, 336

— tinting, 50

datasets

— AirPassengers, 400

— cars, 34, 289, 383

— ChickWeight, 386

— eurodist, 434

— faithful, 80

— HairEyeColor, 310

— iris, 308, 319, 323, 324,

423, 451, 453, 454,

456, 457, 458, 466,

474

— islands, 52

— LakeHuron, 53, 296, 416

— longley, 359, 364

— morley, 289

— precip, 196

— pressure, 383

— rainforest, 385

— state.x77, 440

— trees, 472

— UCBA admissions, 307, 338

— USArrests, 426, 435, 447,

448, 468

— volcano, 473

— warpbreaks, 270

Devore5

— ex10.22, 478

dprep

— hepatitis, 422

drc

— heartrate, 384

— RScpetition, 385

F

faraway

— cathedral, 365

I

ISwR

— juul, 89
 — melanom, 380
 — thuesen, 89

L

LearnBayes

— studentdata, 89, 383

M

MASS

— caith, 438
 — cars, 50
 — Cars93, 50, 461
 — cement, 383
 — cpus, 360
 — DDT, 122
 — farms, 439
 — hills, 53, 185, 383
 — Insurance, 50
 — mtcars, 16, 40, 50
 — painters, 461
 — Pima.tr2, 54
 — survey, 89
 — UScereal, 385
 mlmRev
 — Hsb82, 471

N

nlrwr

— L.minor, 368, 384

P

pls

— oliveoil, 363

psych

— bfi, 431

S

schoolmath

— primlist, 52

T

TSA

— gold, 417
 — hours, 417
 — robot, 418
 — wages, 417

U

UsingR

— aid, 122
 — babies, 123, 337, 385
 — blood, 288
 — bumpers, 123
 — cancer, 287
 — carbon, 89
 — cfb, 50
 — chicken, 88
 — chips, 88
 — corn, 288
 — crime, 122
 — emissions, 53, 385
 — exec.pay, 290
 — female.inc, 290
 — firstchi, 123
 — florida, 383
 — galton, 337
 — homedata, 383
 — homeprice, 386
 — homework, 287
 — iq, 123
 — kid.weights, 88
 — math, 123
 — normtemp, 292
 — pi2000, 88
 — primes, 52
 — puerto, 290
 — rat, 290
 — slc, 123
 — smokyph, 286
 — south, 122
 — trees, 123