

W PROSTOCIE TKWI SIŁA



wydanie II

# Statystyka

dla  
**bystrzaków**



Zrozum koncepcje  
statystyczne, techniki,  
wzory i obliczenia

Interpretuj  
diagramy i wykresy  
i określaj  
prawdopodobieństwo

Analizuj dane  
przy użyciu  
wielu technik

**Deborah J. Rumsey**

profesor na Wydziale Statystyki  
Uniwersytetu Stanowego Ohio

Tytuł oryginału: Statistics For Dummies, 2nd Edition

Tłumaczenie: Leszek Sielicki

ISBN: 978-83-8322-583-8

Original English language edition Copyright © 2011 by Wiley Publishing, Inc., Indianapolis, Indiana

All rights reserved including the right of reproduction in whole or in part in any form.

This translation published by arrangement with John Wiley & Sons, Inc.

Oryginalne angielskie wydanie Copyright © 2011 by Wiley Publishing, Inc., Indianapolis, Indiana

Wszelkie prawa, włączając prawo do reprodukcji całości lub części w jakiegokolwiek formie, zarezerwowane. Tłumaczenie opublikowane na mocy porozumienia z John Wiley & Sons, Inc.

Translation copyright © 2016, 2023 by Helion S.A.

Wiley, the Wiley Publishing Logo, For Dummies, Dla Bystrzaków, the Dummies Man logo, A Reference for the Rest of Us!, The Dummies Way, Dummies Daily, The Fun and Easy Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and/or other countries. Used by permission.

Wiley, the Wiley Publishing Logo, For Dummies, Dla Bystrzaków, the Dummies Man logo, A Reference for the Rest of Us!, The Dummies Way, Dummies Daily, The Fun and Easy Way, Dummies.com, Making Everything Easier, i związana z tym szata graficzna są markami handlowymi John Wiley and Sons, Inc. i/lub firm stowarzyszonych w Stanach Zjednoczonych i/lub innych krajach. Wykorzystywane na podstawie licencji.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the Publisher.

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione. Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie praw autorskich niniejszej publikacji.

Autor oraz wydawca dołożyli wszelkich starań, by zawarte w tej książce informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych lub autorskich. Autor oraz wydawca nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres

<https://dlabystrzakow.pl/user/opinie/stab2v>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Helion S.A.

ul. Kościuszki 1c, 44-100 Gliwice

tel. 32 230 98 63

e-mail: [dlabystrzakow@dlabystrzakow.pl](mailto:dlabystrzakow@dlabystrzakow.pl)

WWW: <https://dlabystrzakow.pl>

Printed in Poland.

- Kup książkę
- Poleć książkę
- Oceń książkę

- Księgarnia internetowa
- Lubię to! » Nasza społeczność

# Spis treści

---

<b>O autorce .....</b>	<b>15</b>
<b>Podziękowania od autorki .....</b>	<b>17</b>
<b>Wstęp .....</b>	<b>19</b>
O książce .....	20
Konwencje zastosowane w książce .....	20
Czego nie czytać .....	21
Naiwne założenia .....	21
Jak podzielona jest książka .....	21
Część I: Istotne statystyki dotyczące statystyki .....	21
Część II: Podstawy wykonywania obliczeń .....	22
Część III: Rozkłady i centralne twierdzenie graniczne .....	22
Część IV: Przewidywanie i hipotetyzowanie z przekonaniem .....	22
Część V: Badania statystyczne i pogoń za znaczącą relacją .....	23
Część VI: Dekalogi .....	23
Ikony użyte w książce .....	23
Co dalej .....	24

## ***Część I: Istotne statystyki dotyczące statystyki .....*** 25

<b>Rozdział 1: Statystyka w pigułce .....</b>	<b>27</b>
Powodzenie w statystycznym świecie .....	27
Projektowanie odpowiednich badań .....	28
Sondáže .....	29
Eksperymenty .....	29
Gromadzenie wartościowych danych .....	30
Dobór właściwej próby .....	30
Unikanie tendencyjności danych .....	31
Tworzenie skutecznych podsumowań .....	32
Statystyki opisowe .....	32
Wykresy i grafy .....	33
Ustalanie rozkładów .....	34
Prowadzenie właściwych analiz .....	34
Błąd statystyczny i przedziały ufności .....	35
Testowanie statystyczne .....	36
Korelacja, regresja i tabele krzyżowe .....	37

## 6 Statystyka dla bystrzaków ---

Wyciąganie wiarygodnych wniosków .....	38
Podawanie zawyżonych wyników .....	38
Kwestionowanie twierzeń o przyczynie i skutku .....	38
Jak być detektywem, a nie sceptykiem .....	39

### **Rozdział 2: Statystyka w codziennym życiu ..... 41**

Statystyka i media — więcej pytań niż odpowiedzi? .....	41
Sondowanie problemów z popcornem .....	42
Walka z wirusami .....	42
Ustalanie przyczyn wypadków .....	43
Rozważania nad pomyłkami lekarskimi .....	44
Kwestie utraty gruntów .....	44
Kontrolowanie szkół .....	44
Problematyka sportowa .....	45
Wiadomości ekonomiczne .....	46
Informacje turystyczne .....	46
Badanie statystyk seksualnych .....	47
Dywagacje nad prognozami pogody .....	47
Rozmyślanie o filmach .....	48
Prześwietlanie horoskopów .....	48
Wykorzystywanie statystyki w pracy .....	48
Odbieranie porodów — oraz udzielanie informacji .....	49
Pozowanie do zdjęć .....	49
Gromadzenie danych o pizzy .....	50
Statystyki w biurze .....	50

### **Rozdział 3: Przejmujemy kontrolę — tak wiele liczb i tak mało czasu ..... 51**

Wykrywanie błędów, przesady i zwykłych kłamstw .....	52
Sprawdzanie obliczeń .....	52
Wykrywanie mylących statystyk .....	53
Poszukiwanie kłamstw we właściwych miejscach .....	59
Skutki mylących statystyk .....	59

### **Rozdział 4: Narzędzia pracy ..... 61**

Statystyka — więcej niż tylko liczby .....	61
Poznajemy podstawy statystycznego żargonu .....	63
Dane .....	63
Zbiór danych .....	64
Zmienna .....	65
Populacja .....	65
Próba losowa czy inna? .....	66
Statystyka .....	67
Parametr .....	68
Tendencyjność .....	68
Średnia (przeciętna) .....	69
Mediana .....	69
Odchylenie standardowe .....	69
Percentyl .....	70

Wynik standaryzowany .....	70
Rozkład i rozkład normalny .....	71
Centralne twierdzenie graniczne .....	72
Wartości z .....	73
Eksperymenty .....	73
Sondaże (ankiety) .....	74
Błąd statystyczny .....	75
Przedział ufności .....	76
Testowanie statystyczne .....	77
Wartości p .....	77
Istotność statystyczna .....	78
Korelacja kontra przyczynowość .....	79

## **Część II: Podstawy wykonywania obliczeń ..... 81**

### **Rozdział 5: Średnie, mediany i nie tylko ..... 83**

Podsumowywanie danych z wykorzystaniem statystyk opisowych .....	83
Obliczenia danych katerycznych — tabele i wartości procentowe .....	84
Badanie środka za pomocą średniej i mediany .....	86
Uśrednianie do średniej .....	87
Podział danych według mediany .....	88
Porównywanie średnich i median — histogramy .....	89
Pojęcie rozproszenia .....	91
Ustalanie odchylenia standardowego .....	91
Poza zakresem .....	95
Badanie reguły trzech sigm .....	95
Pomiar pozycji względnej za pomocą percentyli .....	97
Obliczanie percentyli .....	98
Interpretacja percentyli .....	99
Zestawianie podsumowania pięcioliczbowego .....	101
Badanie rozstępu ćwiartkowego .....	103

### **Rozdział 6: Właściwy obraz — wykresy danych katerycznych ..... 105**

Kawalki układanki — wykresy kołowe .....	105
Podział wydatków osobistych .....	106
Dochody z loterii .....	107
Jedzenie z dostawą do domu .....	108
Analiza trendów struktury wiekowej .....	109
Podnoszenie poprzeczki na wykresach słupkowych .....	110
Śledzenie kosztów transportu .....	110
Zyski z loterii .....	112
Kwestia skali na wykresie słupkowym .....	113
Rozmyślenia nad najbardziej irytującymi sytuacjami w miejscu pracy .....	114

**Rozdział 7: Krok po kroku — wykresy danych liczbowych ..... 117**

Histogramy .....	117
Tworzenie histogramu .....	117
Interpretacja histogramu .....	120
Obrazy i liczby .....	124
Wykrywanie mylących histogramów .....	125
Badanie wykresów pudełkowych .....	128
Tworzenie wykresu pudełkowego .....	128
Interpretacja wykresu pudełkowego .....	130
Tworzenie wykresów czasowych .....	135
Interpretacja wykresów czasowych .....	135
Kwestia zróżnicowania — wykresy czasowe kontra histogramy .....	136
Wyszukiwanie mylących wykresów czasowych .....	136

**Część III: Rozkłady i centralne  
twierdzenie graniczne ..... 141****Rozdział 8: Zmienne losowe i rozkład dwumianowy ..... 143**

Definiowanie zmiennej losowej .....	143
Zmienne dyskretne i ciągłe .....	144
Rozkłady prawdopodobieństw .....	145
Średnia i wariancja dyskretnej zmiennej losowej .....	146
Identyfikacja dwumianu .....	146
Sprawdzanie warunków rozkładu dwumianowego krok po kroku .....	147
Brak ustalonej liczby prób .....	148
Więcej niż sukces lub porażka .....	148
Próby nie są niezależne .....	148
Prawdopodobieństwo sukcesu ( $p$ ) ulega zmianie .....	149
Znajdowanie prawdopodobieństw dwumianowych z wykorzystaniem wzoru .....	149
Znajdowanie prawdopodobieństw dwumianowych z wykorzystaniem tablicy dwumianowej .....	151
Znajdowanie prawdopodobieństw dla określonych wartości $X$ .....	152
Znajdowanie prawdopodobieństw dla $X$ większego lub mniejszego od określonej wartości oraz dla znajdującego się pomiędzy dwiema wartościami .....	152
Sprawdzanie średniej i odchylenia standardowego zmiennej o rozkładzie dwumianowym .....	153

**Rozdział 9: Rozkład normalny ..... 155**

Badanie podstaw rozkładu normalnego .....	155
Spotkanie ze standardowym rozkładem normalnym (rozkładem $Z$ ) .....	158
Sprawdzanie $Z$ .....	158
Standaryzacja $X$ do $Z$ .....	159
Znajdowanie prawdopodobieństw dla $Z$ z wykorzystaniem tablicy $Z$ .....	160
Znajdowanie prawdopodobieństw dla rozkładu normalnego .....	161
Znajdowanie $X$ , gdy znamy wartość procentową .....	164
Ustalanie percentyli dla rozkładu normalnego .....	164
Podchwytliwe sformułowania w zadaniach z percentylami .....	166
Aproksymacja rozkładu dwumianowego rozkładem normalnym .....	167

<b>Rozdział 10: Rozkład t-Studenta .....</b>	<b>169</b>
Podstawy rozkładu t-Studenta .....	169
Porównanie rozkładów Z i t-Studenta .....	169
Ustalanie wpływu zróżnicowania na rozkłady t-Studenta .....	170
Korzystanie z tablicy t-Studenta .....	171
Znajdowanie prawdopodobieństw z wykorzystaniem tablicy rozkładu t-Studenta .....	172
Ustalanie percentyli dla rozkładu t-Studenta .....	172
Określanie wartości $t^*$ dla przedziałów ufności .....	173
Badanie zachowań z wykorzystaniem tablicy t-Studenta .....	173
<b>Rozdział 11: Rozkłady z prób i centralne twierdzenie graniczne .....</b>	<b>175</b>
Definiowanie rozkładu z próby .....	175
Średnia rozkładu z próby .....	177
Pomiar błędu standardowego .....	177
Liczebność próby a błąd standardowy .....	178
Odchylenie standardowe a błąd standardowy dla populacji .....	179
Kształt rozkładu z próby .....	181
Przypadek 1.: rozkład X jest normalny .....	181
Przypadek 2.: rozkład X nie jest normalny — pojawia się centralne twierdzenie graniczne .....	181
Znajdowanie prawdopodobieństw dla średniej z próby .....	184
Rozkład z próby dla odsetka z próby .....	185
Znajdowanie prawdopodobieństw dla odsetka z próby .....	187
 <b>Część IV: Przewidywanie i hipotetyzowanie z przekonaniem .....</b>	 <b>189</b>
<b>Rozdział 12: Miejsce na błąd statystyczny .....</b>	<b>191</b>
Istotność plusa i minusa .....	191
Ustalanie błędu statystycznego — wzór ogólny .....	193
Pomiar zróżnicowania próby .....	193
Obliczanie błędu statystycznego dla odsetka z próby .....	194
Ogłaszanie wyników .....	196
Obliczanie błędu statystycznego dla średniej z próby .....	196
Przekonanie, że mamy rację .....	198
Ustalanie znaczenia liczebności próby .....	198
Liczebność próby a błąd statystyczny .....	198
Więcej nie zawsze znaczy (o wiele) lepiej! .....	199
Utrzymanie błędu statystycznego w odpowiedniej perspektywie .....	199
<b>Rozdział 13: Przedziały ufności — trafne oszacowania .....</b>	<b>201</b>
Nie wszystkie oszacowania są sobie równe .....	201
Łączenie statystyki z parametrem .....	202
Kwestie nazewnictwa .....	203
Interpretacja wyników .....	204
Problematyka zakresu .....	204

Wybór poziomu ufności .....	205
Uwzględnianie liczebności próby .....	207
Ocena zróżnicowania populacji .....	208
Obliczanie przedziału ufności dla średniej z populacji .....	209
Przypadek 1.: odchylenie standardowe populacji jest znane .....	209
Przypadek 2.: odchylenie standardowe populacji nie jest znane i/lub wartość $n$ jest niewielka .....	210
Ustalanie pożądanej wartości liczebności próby .....	211
Ustalanie przedziału ufności dla pojedynczego odsetka populacji .....	213
Określanie przedziału ufności dla różnicy dwóch średnich .....	214
Przypadek 1.: odchylenia standardowe populacji są znane .....	215
Przypadek 2.: odchylenia standardowe populacji są nieznanne i/lub liczebności prób są niewielkie .....	216
Szacowanie różnicy dwóch odsetków .....	217
Znajdowanie mylących przedziałów ufności .....	219
<b>Rozdział 14: Twierdzenia, testy i wnioski .....</b>	<b>221</b>
Formułowanie hipotez .....	222
Definiowanie zera .....	222
Jaka jest alternatywa? .....	222
Gromadzenie dowodów (danych) .....	223
Zestawianie dowodów — statystyka testowa .....	224
Gromadzenie statystyk z prób .....	224
Pomiar rozproszenia z wykorzystaniem błędów standardowych .....	224
Pojęcie wyników standaryzowanych .....	225
Obliczanie i interpretacja statystyki testowej .....	225
Ocena dowodów i podejmowanie decyzji — wartości $p$ .....	226
Łączenie statystyk testowych i wartości $p$ .....	226
Definiowanie wartości $p$ .....	227
Obliczanie wartości $p$ .....	227
Wyciąganie wniosków .....	228
Określanie granic dla odrzucenia $H_0$ .....	229
Testowanie zylaków .....	230
Ocena prawdopodobieństwa złej decyzji .....	230
Fałszywe alarmy — błędy pierwszego rodzaju .....	231
Brak wykrycia — błędy drugiego rodzaju .....	231
<b>Rozdział 15: Popularne testy statystyczne — wzory i przykłady .....</b>	<b>233</b>
Testowanie jednej średniej z populacji .....	233
Małe próby i nieznanne odchylenia standardowe — test t-Studenta .....	235
Stosowanie testu t-Studenta .....	236
Związek pomiędzy $t$ i $Z$ .....	237
Postępowanie z ujemnymi wartościami $t$ .....	237
Badanie opcji „nie równa się” .....	238
Testowanie pojedynczego odsetka z populacji .....	238
Porównywanie dwóch (niezależnych) średnich z populacji .....	240
Testowanie średniej z różnicy (test t-Studenta dla zmiennych zależnych) .....	242
Porównywanie dwóch odsetków z populacji .....	245



## ***Część V: Badania statystyczne i pogoń za znaczącą relacją ..... 249***

### **Rozdział 16: Sondáže, sondáže i jeszcze raz sondáže .....251**

Znaczenie sondáže .....	252
Docieranie do źródła .....	252
Gorące tematy sondáže .....	254
Wpływ sondáže na ludzi .....	254
Za kulisami — tajniki sondáže .....	256
Planowanie i projektowanie sondáže .....	256
Wybór próby .....	260
Prowadzenie sondáže .....	262
Interpretacja wyników i wyszukiwanie problemów .....	264

### **Rozdział 17: Eksperymenty — przełomy w medycynie czy mylące wyniki? .....267**

Sprowadzanie badań do podstaw .....	268
Podstawowe terminy badawczego żargonu .....	268
Obserwacja badań obserwacyjnych .....	269
Badanie eksperymentów .....	269
Projektowanie dobrego eksperymentu .....	270
Projektowanie eksperymentu, aby możliwe było dokonywanie porównań .....	270
Wybór liczebności próby .....	273
Wybór obiektów .....	274
Przypisywanie losowe .....	274
Kontrola zmiennych zakłócających .....	275
Poszanowanie kwestii etycznych .....	277
Gromadzenie dobrych danych .....	278
Prawidłowa analiza danych .....	280
Wyciąganie właściwych wniosków .....	280
Podejmowanie świadomych decyzji .....	282

### **Rozdział 18: Szukanie powiązań — korelacja i regresja .....285**

Przedstawianie zależności za pomocą wykresu rozrzutu .....	286
Tworzenie wykresu rozrzutu .....	287
Interpretacja wykresu rozrzutu .....	287
Ustalanie istnienia zależności liniowych za pomocą korelacji .....	288
Obliczanie korelacji .....	289
Interpretacja korelacji .....	290
Badanie właściwości korelacji .....	292
Pojęcie regresji liniowej .....	292
Ustalanie, która zmienna to X, a która Y .....	292
Sprawdzanie warunków .....	293
Poszukiwanie linii regresji .....	293
Interpretacja linii regresji .....	295
Zestawienie wszystkiego razem — linia regresji dla świerszczy .....	296
Dokonywanie właściwych prognoz .....	297
Wjaśnienie zależności — korelacja kontra przyczyna i skutek .....	298

## **Rozdział 19: Tabele krzyżowe i niezależność ..... 301**

Tworzenie tabeli krzyżowej .....	302
Budowa komórek .....	303
Ustalanie sum .....	303
Interpretacja tabel krzyżowych .....	304
Wydrebnianie zmiennych za pomocą rozkładów brzegowych .....	305
Badanie wszystkich grup — rozkład łączny .....	308
Porównywanie grup za pomocą rozkładów warunkowych .....	310
Sprawdzanie niezależności i opis zależności .....	313
Sprawdzanie istnienia niezależności .....	313
Opis relacji zależnej .....	316
Ostrożna interpretacja wyników .....	317
Sprawdzanie istnienia uzasadnionego związku przyczyny i skutku .....	317
Przełożenie z próby na populację .....	318
Tworzenie rozważnych prognoz .....	319
Opieranie się pokusie wyciągnięcia pochopnych wniosków .....	320

## **Część VI: Dekalogi ..... 321**

### **Rozdział 20: Dziesięć porad dla statystycznych detektywów ..... 323**

Wskazywanie mylących wykresów .....	323
Wykresy kołowe .....	325
Wykresy słupkowe .....	325
Wykresy czasowe .....	326
Histogramy .....	326
Wykrywanie tendencyjnych danych .....	327
Poszukiwanie błędu statystycznego .....	328
Identyfikowanie prób nielosowych .....	328
Ustalanie brakujących liczebności prób .....	329
Wykrywanie błędnie interpretowanych korelacji .....	330
Ujawnianie zmiennych zakłócających .....	331
Sprawdzanie liczb .....	332
Selektywne raportowanie .....	332
Kwestia anegdoty .....	333

### **Rozdział 21: Dziesięć niezawodnych porad egzaminacyjnych ..... 335**

Dowiedz się, czego nie wiesz, i zrób coś z tym .....	336
Unikanie pułapek „tak-tak” .....	337
Pułapka „tak-tak” nr 1 .....	337
Pułapka „tak-tak” nr 2 .....	338
Zaprzyjżnij się z wzorami .....	339
Zrób listę „jeżeli-to-jak” .....	341
Ustal, o co chodzi w poleceniu .....	342
Oznacz dane .....	343

Zrób rysunek .....	345
Stwórz powiązanie i rozwiąż zadanie .....	346
Wykonaj obliczenia dwa razy .....	347
Analizuj swoje odpowiedzi .....	347
<b>Dodatek A: Tablice statystyczne .....</b>	<b>349</b>
Tablica Z .....	349
Tablica t-Studenta .....	352
Tablica dwumianowa .....	354
<b>Skorowidz .....</b>	<b>359</b>



## Rozdział 1

# Statystyka w pigułce

.....

### *W tym rozdziale:*

- ▶ Ustalamy, o co chodzi w procesie statystycznym.
  - ▶ Odnosimy sukcesy statystyczne w codziennym życiu, w pracy i w szkole.
- .....

**W**spółczesny świat tonie w danych do takiego stopnia, że każdy (nawet ja!) może się czuć przytłoczony. Nie winię Cię, jeśli do tej pory z niedowierzaniem podchodziłeś do statystyk prezentowanych w mediach — mnie też czasem się to zdarza. Mam jednak dobrą wiadomość: chociaż czyha na Ciebie bardzo wiele mylących i błędnych informacji, to powstaje także wiele świetnych rzeczy; na przykład, wiele różnorodnych badań i technik związanych z danymi pomaga poprawiać jakość naszego życia. Twoim zadaniem jest umiejętnie oddzielać ziarno od plew i być pewnym swoich umiejętności w tym zakresie. Dzięki dogłębnemu zrozumieniu statystyki i procedur statystycznych zaczniesz pewnie radzić sobie z liczbami w życiu codziennym, w pracy czy w szkole. O tym właśnie opowiada ta książka.

W tym rozdziale przyglądam się roli, jaką statystyka odgrywa we współczesnym społeczeństwie, opartym na danych. Wyjaśniam, co należy robić, aby nie tylko w nim przetrwać, ale móc się rozwijać. Abyś mógł zyskać szersze spojrzenie na statystykę, wspólnie zastosujemy metodę naukową — projektowanie efektywnych obserwacji, gromadzenie użytecznych danych, organizowanie i analizowanie informacji, interpretację wyników oraz wyciąganie odpowiednich wniosków. (A Ty myślałeś, że statystyka to tylko obliczenia!).

## *Powodzenie w statystycznym świecie*

Trudno jest ujarzmić powódź statystyk, które w mniejszym lub większym stopniu wpływają na nasze codzienne życie. Wszystko zaczyna się w chwili, gdy budzisz się rano, sprawdzasz wiadomości i słuchasz pogodyнки przedstawiającej prognozę opartą na analizach statystycznych danych historycznych i aktualnych warunków pogodowych. Podczas śniadania studiujesz informacje o wartościach odżywczych na pudełku płatków zbożowych. W pracy wykorzystujesz liczby z wykresów i tabel, wprowadzasz dane do arkusza kalkulacyjnego, przeprowadzasz diagnostykę, wykonujesz pomiary, prowadzisz obliczenia, szacujesz wydatki, podejmujesz decyzje, wykorzystując statystyczne punkty odniesienia, oraz zamawiasz towary na podstawie danych historycznych dotyczących poziomu sprzedaży.

Na obiad idziesz do restauracji, która zajęła pierwsze miejsce w ankiecie przeprowadzonej wśród 500 osób. Spożywasz posiłek, którego cenę ustalono na podstawie danych marketingowych. Idziesz do lekarza, mierzą Ci ciśnienie, temperaturę, wagą i robią badanie krwi; po zebraniu wszystkich informacji otrzymujesz swoje wyniki i informację, jak wypadasz na podstawie norm statystycznych.

Jedziesz do domu samochodem, który jest obsługiwany przez komputer prowadzący diagnostykę statystyczną. Po powrocie do domu włączasz wiadomości i poznajesz najnowsze statystyki dotyczące przestępczości, sprawdzasz, co się działo na giełdzie, i dowiadujesz się, ile osób odwiedziło w zeszłym tygodniu zoo.

Wieczorem myjesz zęby pastą, która, co zostało statystycznie dowiedzione, redukuje ubytki, czytasz kilka stron książki (uznanej za bestseller na podstawie statystycznych szacunków sprzedaży) i idziesz spać — a następnego ranka zaczynasz wszystko od nowa. Czy jednak możesz mieć pewność, że wszystkie statystyki, które każdego dnia napotykasz i na których polegasz, są prawidłowe? W rozdziale 2. omawiam bardziej szczegółowo kilka przykładów tego, w jaki sposób statystyka uczestniczy w naszym życiu i pracy, jaki wpływ na nas wywiera i jak należy być bardziej świadomym jej obecności.



Istnieją dane statystyczne, które są niejasne, niewłaściwe albo po prostu błędne. Musisz bardziej świadomie postrzegać statystyki, które napotykasz każdego dnia, i tak wytrenować swój umysł, abyś potrafił się zatrzymać, powiedzieć „momencik!”, zweryfikować informacje, zadać pytania i podnieść czerwoną flagę, jeżeli coś będzie nie tak. Z rozdziału 3. dowiesz się, w jaki sposób można zostać wyprowadzonym w pole przez nieprawidłowe statystyki, i rozwiniesz w sobie umiejętności krytycznego myślenia i identyfikowania problemów, zanim bez refleksji uwierzysz w wyniki.

Statystyka, jak każda inna dziedzina, posiada swój charakterystyczny żargon, którego szereg najczęściej stosowanych terminów przedstawiam i wyjaśniam w rozdziale 4. Znajomość fachowego słownictwa zwiększa zdolność rozumienia kwestii statystycznych i komunikowania się z innymi na wyższym poziomie i bez poczucia zastraszenia. To, że posługujesz się precyzyjną terminologią, opisując, co jest nie tak z danym wynikiem statystycznym (i dlaczego), podnosi Twoją wiarygodność. A Twoje prezentacje z wykorzystaniem tablic statystycznych, grafów, wykresów i analiz także będą zawierać więcej informacji i będą skuteczniejsze. (Kurczę, jeśli nie ma nawet żadnego innego powodu, to ten żargon będzie Ci potrzebny, bo ja się nim posługuję w tej książce; ale nie przejmuj się, wszystko wyjaśniam).

W kolejnych podrozdziałach przekonasz się, w jaki sposób statystyka bierze udział we wszystkich etapach metody naukowej.

## *Projektowanie odpowiednich badań*

Pytania zadają wszyscy: firmy farmaceutyczne i biologowie; analitycy marketingowi i rząd. I w ostatecznym rozrachunku wszyscy wykorzystują statystykę, aby uzyskać odpowiedzi na swoje pytania. Zwłaszcza wiele badań medycznych i psychologicznych prowadzi się, bo ktoś chce poznać odpowiedź na jakieś pytanie, na przykład:

- ✓ Czy za sprawą tej szczepionki zapobieganie grypie będzie skuteczne?
- ✓ Jakie jest zdanie obywateli o stanie gospodarki?
- ✓ Czy wzrost poziomu wykorzystywania sieci społecznościowych powoduje depresję u nastolatków?

Pierwszym krokiem po sformułowaniu pytania badawczego jest zaprojektowanie efektywnej obserwacji w celu zebrania danych, które pomogą odpowiedzieć na postawione pytanie. Oznacza to ustalenie, jakiego procesu należy użyć, aby uzyskać potrzebne dane. W tym podrozdziale prezentuję przegląd dwóch głównych typów badań — sondaży i eksperymentów — i sprawdzam, dlaczego tak ważne jest zweryfikowanie sposobu, w jaki badanie zostało zaprojektowane, zanim damy wiarę jego wynikom.

## Sondaże

**Badanie obserwacyjne** to badanie, w którym dane od konkretnych osób są gromadzone bez oddziaływania na te osoby. Najczęstszym badaniem obserwacyjnym jest sondaż.

**Sondaże** to kwestionariusze przekazywane osobom, które zostały wybrane z interesującej badacza populacji. Sondaże mogą przybierać różne formy: sondaży w formie papierowej przesyłanych za pośrednictwem poczty, kwestionariuszy na stronach internetowych, głosowań audiotele prowadzonych przez stacje telewizyjne, ankiet telefonicznych itd.



Prawidłowo przeprowadzone sondaże mogą być bardzo użytecznym narzędziem do pozyskiwania informacji. Jeżeli jednak prowadzone są niewłaściwie, mogą doprowadzić do zgromadzenia informacji nieprawdziwych. Problemem mogą tu być niewłaściwie sformułowane, wprowadzające w błąd pytania, brak odpowiedzi od osób wybranych do udziału w sondażu albo nieuwzględnienie reprezentatywnej grupy należącej do danej populacji. Te potencjalne problemy oznaczają, że zanim sondaż zostanie przeprowadzony, należy go dobrze przemyśleć.



Wielu badaczy poświęca sporo czasu i pieniędzy, aby prowadzone przez nich sondaże były użyteczne, a Ty (na podstawie kryteriów, które omawiam w rozdziale 16.) będziesz wiedział, że można im zaufać. Miewając jednak do czynienia z wieloma różnymi rodzajami sondaży, obecnymi w mediach, w pracy czy na różnych zajęciach w szkole, sam musisz być w stanie szybko sprawdzać i weryfikować sposoby projektowania i prowadzenia sondaży, a także umieć w dobrze przemyślany sposób wskazywać konkretne problemy. Narzędzia potrzebne do oceny ankiet znajdziesz w rozdziale 16.

## Eksperymenty

**Eksperyment** zakłada poddanie jego uczestników jednemu lub większej liczbie działań w taki sposób, aby można było dokonać jednoznacznych porównań.

Po przeprowadzeniu działań rejestrowane są reakcje. Aby na przykład zbadać wpływ określonej dawki leku na ciśnienie krwi, jednej grupie uczestników możemy podać 10 mg leku, a drugiej 20 mg. Zazwyczaj istnieje także grupa kontrolna, której członkowie otrzymują placebo (np. pigułki z cukru) albo leczeni są z wykorzystaniem leków standardowych, a nie eksperymentalnych (jak w przypadku podawania już istniejących leków pacjentom z AIDS).



Skuteczne i wiarygodne eksperymenty projektowane są tak, aby zminimalizować tendencyjność, zebrać wiele użytecznych danych i dokonać odpowiednich porównań (grupy doświadczalnej z grupą kontrolną). Potencjalne problemy związane z eksperymentami to między innymi sytuacje, w których badacze i/lub badani wiedzą, jakiego rodzaju kuracji zostali poddani, czynniki nieuwzględnione w badaniu, które mają wpływ na wynik (takie jak waga uczestnika podczas badania poziomu dawek leku), lub brak grupy kontrolnej (brak punktu odniesienia, z którym można porównać wyniki).

Gdy jednak eksperyment zostanie zaprojektowany poprawnie, to może pomóc badaczowi w ustaleniu związku przyczynowo-skutkowego, jeśli różnica w reakcji pomiędzy grupą eksperymentalną i grupą kontrolną jest statystycznie istotna (czyli jest mało prawdopodobne, że zaistniała wyłącznie przez przypadek).



Eksperymenty są stosowane w celu wspomaganie tworzenia i testowania leków, ustalania dobrych praktyk w zakresie wytwarzania i przygotowywania żywności oraz badania, czy jakaś nowa kuracja może wyleczyć chorobę albo przynajmniej ograniczyć jej skutki. Dzięki przeprowadzaniu dobrze zaprojektowanych eksperymentów jakość naszego życia z pewnością staje się coraz wyższa. Nie wszystkie eksperymenty bywają jednak dobrze zaprojektowane, a zdolność do określenia, które wyniki są wiarygodne, a które niewiarygodne (gra słów całkowicie zamierzona), ma kluczowe znaczenie, zwłaszcza gdy wyniki są dla nas bardzo istotne. Wszystkie niezbędne informacje o eksperymentach i sposobach ich oceny znajdziesz w rozdziale 17.

## Gromadzenie wartościowych danych

Po zaprojektowaniu badania — niezależnie od tego, czy jest to sondaż, czy eksperyment — należy wybrać uczestników i przeprowadzić obserwację statystyczną, aby zgromadzić dane. Ten etap procesu ma kluczowe znaczenie dla pozyskania wiarygodnych danych, a w tym podrozdziale zajmiemy się jego najistotniejszymi szczegółami.

### Dobór właściwej próby

Statystycy mają takie powiedzenie: „Śmieci na wejściu — śmieci na wyjściu”. Jeśli wybierzesz swoje *obiekty* (osoby, które będą uczestniczyć w badaniu) w sposób *tendencyjny* — czyli faworyzując pewne osoby lub grupy osób — wtedy wyniki będą tendencyjne. To proste jak drut.

Załóżmy, że Bolek chce poznać opinie mieszkańców miasta na temat planowanej budowy kasyna. Idzie do centrum handlowego z kartką i prosi przechodzące osoby o wyrażanie opinii. Co tu jest nie tak? No cóż, Bolek pozyska wyłącznie opinie:  
a) osób, które robią zakupy w tym centrum handlowym; b) w tym konkretnym dniu;  
c) w tej konkretnej chwili; d) tylko tych, które zechcą odpowiedzieć.

Te okoliczności są zbyt restrykcyjne — napotykanymi ludźmi nie reprezentują przekroju społeczności miasta. Podobnie, Bolek mógłby zorganizować ankietę internetową i poprosić o głosy. Danych dostarczą mu jednak tylko osoby, które będą wiedziały o stronie z ankietą i zechcą odpowiedzieć, a taki trud zadają sobie zwykle wyłącznie





posiadacze skrajnych opinii. Ostatecznie Bolek będzie dysponował tylko pakietem tendencyjnych danych od osób całkowicie niereprezentatywnych z punktu widzenia społeczności miasta.



Aby zminimalizować tendencyjność w sondażu, kluczowe znaczenie ma *losowość*. Musisz wybrać próbę jednostek losowo — czyli w procesie, który można określić jako „wyciąganie nazwisk z kapelusza”. Badacze wykorzystują różne metody wybierania uczestników losowo, a jak to robią, dowiesz się z rozdziału 16.

Zauważ, że podczas projektowania eksperymentu zestawianie losowej próby i zapraszanie do udziału często bywa nieetyczne, bo eksperyment oznacza określone oddziaływanie na obiekty. Zamiast tego należy ogłosić nabór ochotników. Następnie powinniśmy się upewnić, że ochotnicy wybrani z grupy reprezentują populację będącą przedmiotem naszego zainteresowania oraz że dane dotyczące tych osób zostały zebrane prawidłowo, więc można będzie dokonać projekcji wyników na większą grupę. Jak to się robi, przekonasz się w rozdziale 17.

Po przeczytaniu rozdziałów 16. i 17. będziesz wiedział, jak należy analizować metody doboru prób. Nauczysz się także opracowywać plan, którym sam będziesz mógł się posłużyć do przeprowadzenia wyboru próby. Wtedy dowiesz się także, kiedy można stwierdzić, że mamy „śmieci na wejściu — śmieci na wyjściu”.

## Unikanie tendencyjności danych

**Tendencyjność** to systematyczne faworyzowanie pewnych osób lub reakcji. Tendencyjność to zbrodnia statystyków, którzy robią wszystko, co mogą, aby ją zminimalizować. Mam podać przykład tendencyjności? Załóżmy, że prowadzisz telefoniczną ankietę dotyczącą zadowolenia z pracy; jeśli będziesz dzwonił na numery domowe w ciągu dnia, między godziną 8:00 a 16:00, to pominiesz wszystkie osoby pracujące właśnie w dzień. A może osoby pracujące w dzień są bardziej zadowolone niż pracujące w nocy?

Na tendencyjność trzeba uważać podczas gromadzenia danych do sondaży. Niektóre sondaże bywają na przykład zbyt długie — co będzie, jeśli ktoś przestanie odpowiadać na pytania w połowie? A co się stanie, jeśli ktoś zechce podać nieprawdziwe informacje i odpowie, że zarabia rocznie 100 000 zł, a nie 45 000? Co zrobisz, jeśli otrzymasz odpowiedzi, które nie znajdują się na Twojej liście możliwych odpowiedzi? Podczas gromadzenia danych do sondażu mogą wystąpić liczne problemy, a Ty musisz być w stanie je zidentyfikować.



W kwestii tendencyjności i gromadzenia danych eksperymenty mogą czasem stanowić nawet większe wyzwanie. Załóżmy, że chcemy przeprowadzić test ciśnienia krwi; co będzie, jeżeli instrument, którego używasz, zepsuje się w trakcie eksperymentu? Co zrobisz, jeżeli ktoś zrezygnuje w połowie eksperymentu? A jeżeli podczas eksperymentu stanie się coś, co rozproszy obiekt lub badacz? Albo nie będzie można znaleźć żyły, a badanie krwi powinno zostać przeprowadzone dokładnie godzinę po podaniu dawki leku? Te problemy to tylko kilka przykładowych kwestii, które mogą się pojawić podczas gromadzenia danych do eksperymentów i które trzeba umieć dostrzegać.

Po przeczytaniu rozdziału 16. (o próbach i sondażach) i rozdziału 17. (o eksperymentach) będziesz potrafił dobierać próby i gromadzić dane w wolny od tendencyjności sposób, poświęcając uwagę drobiazgom, które mogą mieć faktyczny wpływ na wyniki. Nabędziesz także umiejętność oceny wiarygodności wyników statystycznych i zdobędziesz posłuch, ponieważ będziesz wiedział, o czym mówisz.

# Tworzenie skutecznych podsumowań

Po zgromadzeniu prawidłowych danych kolejnym krokiem jest podsumowanie ich w celu uzyskania możliwości spojrzenia z szerszej perspektywy. Statystycy opisują dane na dwa podstawowe sposoby: liczbowo (tzw. *statystyki opisowe*) oraz za pomocą rysunków (czyli wykresów i grafów).

## Statystyki opisowe

**Statystyki opisowe** to liczby opisujące zbiór danych pod względem jego istotnych cech:

- ✓ Jeżeli dane mają charakter **kategoryczny** (gdy jednostki są umieszczane w grupach, takich jak płeć czy przynależność polityczna), ich podsumowaniem jest zazwyczaj liczba osób w każdej z grup (zwana **liczebnością**) lub odsetek osób w każdej z grup (zwany **częstością**).
- ✓ **Dane liczbowe** oznaczają pomiary lub zliczenia, w ramach których konkretne liczby coś oznaczają (na przykład wzrost i wagę). Z danych liczbowych można pozyskać więcej informacji niż tylko liczba lub odsetek w każdej z grup. Niektóre z tych informacji to:
  - Miary średnie (innymi słowy, gdzie jest „środek” danych?).
  - Miary rozproszenia (na ile różnorodne lub na ile skoncentrowane są dane wokół środka?).
  - Jeżeli jest to potrzebne, liczby, które mierzą wartość relacji pomiędzy dwiema zmiennymi (takimi jak wzrost i waga).

Pewne statystyki opisowe są w niektórych sytuacjach bardziej odpowiednie niż inne; na przykład średnia nie zawsze bywa najlepszą miarą środka zbioru danych; często lepszym wyborem jest mediana. Także odchylenie standardowe nie jest jedyną istniejącą miarą zmienności; rozstęp ćwiartkowy ma też doskonałe właściwości. Musisz być w stanie rozpoznawać, interpretować i oceniać rodzaje statystyk opisowych, z którymi codziennie się stykasz, i wiedzieć, które są w określonych sytuacjach najlepsze.

Interpretacje, porównania i oceny najczęściej spotykanych statystyk opisowych znajdziesz w rozdziale 5. Powszechnie stosowane statystyki opisowe obejmują liczebności i częstości (zliczenia i odsetki) dla danych kategorycznych oraz średnie, mediany, odchylenie standardowe, percentyle i ich kombinacje dla danych liczbowych.



## Wykresy i grafy

Dane podsumowuje się w sposób wizualny za pomocą wykresów i/lub grafów. Są to prezentacje zorganizowane w ten sposób, aby w jednej chwili umożliwić nam szersze spojrzenie na dane i/lub przybliżyć konkretny wynik, który znaleziono. W dzisiejszym świecie szybkiej informacji i miniplików dźwiękowych grafy i wykresy występują powszechnie. Większość z nich prezentuje to, czego dotyczy, jednoznacznie, skutecznie i rzetelnie; może się jednak zdarzać, że pozostawiają one nieco zbyt wiele miejsca na *licentia poetica* i w konsekwencji narażają nas na błędne interpretacje zawartych w nich informacji.



W rozdziałach 6. i 7. omawiam główne typy grafów i wykresów wykorzystywanych do podsumowywania danych, zarówno kategorycznych, jak i liczbowych (patrz poprzedni podrozdział na temat tych rodzajów danych). Dowiesz się, jak je tworzyć, jakie są ich cele oraz jak interpretować ich wyniki. Wskazuję także wiele możliwych sposobów tworzenia grafów i wykresów tak, aby wprowadzały one odbiorcę w błąd, i pomagam szybko dostrzegać tego rodzaju problemy. Chodzi o stwierdzenie: „Chwileczkę! Tu chyba jest błąd!”, gdy wiemy, skąd się on wziął. Oto najistotniejsze kwestie:

- ✓ Jednymi z podstawowych wykresów stosowanych dla danych kategorycznych są wykresy kołowe oraz wykresy słupkowe, które rozkładają takie zmienne jak płeć lub aplikacje wykorzystywane przez nastolatków w telefonach komórkowych. Wykres słupkowy może na przykład prezentować opinie na temat określonego zagadnienia z wykorzystaniem pięciu słupków oznaczonych w kolejności od „zdecydowanie się nie zgadzam” po „zdecydowanie się zgadzam”. W rozdziale 6. znajdziesz wszystkie istotne informacje na temat tworzenia, interpretowania i, co najważniejsze, oceny wykresów i grafów z punktu widzenia ich wiarygodności. Będziesz zaskoczony, widząc, ile można zrobić błędów w prostym wykresie słupkowym.
- ✓ W odniesieniu do danych liczbowych, takich jak wzrost, waga, czas lub ilość, konieczne jest stosowanie innego rodzaju wykresów. Grafy zwane histogramami i wykresami pudełkowymi służą do podsumowywania danych liczbowych i mogą być niezwykle pouczające, doskonale zestawiając wszystkie informacje dotyczące zbioru danych. Oczywiście, one także mogą być mylące, przez przypadek, a nawet celowo. (Więcej na ten temat w rozdziale 7.).



Wykresy i grafy widzimy codziennie — wystarczy, że otworzysz gazetę, a najprawdopodobniej znajdziesz co najmniej kilka wykresów, nawet bez szczególnie długich poszukiwań. Niezbędne jest tutaj posiadanie statystycznego „szkła powiększającego”, które pomaga zinterpretować informacje tak, abyś mógł dostrzec wprowadzające w błąd wykresy, zanim wyciągniesz błędne wnioski i być może zaczniesz działać na ich podstawie. Wszystkie niezbędne do tego narzędzia znajdziesz w rozdziale 6. (dla danych kategorycznych) i w rozdziale 7. (dla danych liczbowych).

## Ustalanie rozkładów

**Zmienna** to cecha, którą można obliczyć, zmierzyć lub przypisać do określonej kategorii. Przykładami mogą być płeć, wiek, wzrost, waga czy liczba posiadanych zwierząt. **Rozkład** to zestawienie możliwych wartości zmiennej (lub przedziałów wartości) i tego, jak często (lub z jaką gęstością) one występują. Na przykład, w Stanach Zjednoczonych rozkład płci przy urodzeniu szacuje się na 52,4% mężczyzn i 47,6% kobiet.

Dla różnych zmiennych istnieją różne rodzaje rozkładów. Trzy rozkłady przedstawione poniżej są najczęściej spotykane w ramach kursu wprowadzającego do zagadnień statystyki. Mają także wiele zastosowań w świecie rzeczywistym:

- ✓ Jeżeli zmienna jest zliczeniem liczby sukcesów w pewnej liczbie prób (jak liczba osób, które wyzdrowiały, przyjmując określony lek), to ma ona rozkład **dwumianowy**.
- ✓ Jeżeli zmienna przyjmuje wartości występujące zgodnie z „krzywą w kształcie dzwonu”, jak wyniki egzaminów maturalnych, to zmienna ta ma rozkład **normalny**.
- ✓ Jeżeli zmienna opiera się na średnich z prób i dysponujemy ograniczonymi danymi, jak w teście ograniczonym do dziesięciu obiektów, którego celem jest sprawdzenie skuteczności określonej diety, odpowiedni może być rozkład **t-Studenta**.

W przypadku rozkładów powinieneś wiedzieć, jak ustalić rodzaj rozkładu przyjmowanego przez konkretną zmienną, jak znaleźć jego prawdopodobieństwa oraz jak określić średnią i odchylenie standardowe wyników. Aby Ci to ułatwić, służę trzema rozdziałami, po jednym dla każdego rozkładu: rozdział 8. traktuje o rozkładzie dwumianowym, rozdział 9. o normalnym, a rozdział 10. poświęcam rozkładowi t-Studenta.

Osoby uczestniczące w kursach wprowadzających do statystyki (lub w ogóle dowolnych kursach statystyki) wiedzą, że jednym z najtrudniejszych do zrozumienia zagadnień są rozkłady z prób i centralne twierdzenie graniczne (te dwie kwestie idą w parze). W rozdziale 11. poznasz te kwestie krok po kroku, co pozwoli Ci zrozumieć, czym jest rozkład z próby, do czego służy i w jaki sposób tworzy podstawę takich analiz danych jak testowanie hipotez i przedziały ufności (więcej informacji na temat analizowania danych znajdziesz w kolejnym podrozdziale). Gdy zrozumiesz centralne twierdzenie graniczne, będziesz w stanie rozwiązywać skomplikowane problemy znacznie łatwiej, a wszystkie kluczowe informacje na ten temat przedstawiam w rozdziale 11.

## Prowadzenie właściwych analiz

Po zgromadzeniu i opisanie danych za pomocą liczb i obrazów zaczyna się zabawa: znajdowanie drogi w ciemnym pomieszczeniu o nazwie *analiza statystyczna*. Jeżeli badanie zostało właściwie zaprojektowane, to można odpowiedzieć na postawione na początku pytania, stosując odpowiednią analizę — kluczowy jest tutaj wyraz *odpowiednia*.





Istnieje wiele rodzajów analiz, a kluczowe znaczenie mają wybór odpowiedniej analizy dla konkretnej sytuacji, prawidłowa interpretacja jej wyników i znajomość ograniczeń. Podobnie istotna jest umiejętność oceny wyboru analiz przez innych oraz wniosków, jakie na ich podstawie wyciągnęli.

W tej książce znajdziesz wszelkie informacje i narzędzia niezbędne do analizy danych przy użyciu najpowszechniejszych metod statystycznych: przedziałów ufności, testów hipotez, korelacji i regresji oraz analizy tabel krzyżowych. W tym podrozdziale prezentuję podstawowy przegląd tych metod.

## ***Błąd statystyczny i przedziały ufności***

Często obserwujemy statystyki, które próbują szacować wartości odnoszące się do całej populacji; prawdę mówiąc, widujemy je niemal codziennie w postaci wyników sondaży. Media informują nas, jaka jest średnia cena benzyny, co obywatele sądzą o działaniach prezydenta lub ile godzin tygodniowo spędzamy w internecie. Nikt jednak nie może podać określonego konkretną liczbą wyniku i stwierdzić, że jest to dokładne oszacowanie dla całej populacji, chyba że zbierze dane na temat każdego członka społeczeństwa. Mógłbyś na przykład usłyszeć, że 60% obywateli popiera stanowisko prezydenta dotyczące opieki zdrowotnej, ale wiesz, że Ciebie nie pytano, więc jak to możliwe, że zapytano wszystkich? A ponieważ nie zapytano wszystkich, wiesz, że odpowiedź określona jedną liczbą nie będzie prawidłowa.

Fakty są takie, że dane gromadzi się na podstawie próby z populacji (na przykład Instytut Gallupa stosuje próby składające się z 2500 losowo wybieranych osób), wyniki z tej próby są analizowane i na podstawie wyników dla próby wyciąga się wnioski odnoszące się do całej populacji (na przykład do wszystkich Amerykanów).



Najważniejsze jest to, że wyniki z prób różnią się w zależności od próby i wartość tej różnicy powinna być ogłaszana (często jednak się tego nie robi). Statystyka wykorzystywana do pomiaru i raportowania poziomu precyzji w wynikach czyichś prób nazywana jest **błędem statystycznym**. W tym kontekście słowo *błąd* nie oznacza, że ktoś popełnił pomyłkę; chodzi o to, że skoro nie próbkowaliśmy całej populacji, to pojawi się rozdzźwięk między naszymi wynikami a rzeczywistą wartością, którą próbujemy oszacować dla populacji.

Ktoś może na przykład twierdzić, że 60% z 1200 ankietowanych osób popiera podejście prezydenta do opieki zdrowotnej, i podaje wyniki z błędem statystycznym plus minus 2%. Wynik końcowy, za pomocą którego przedstawiamy nasze ustalenia jako prawdopodobne wartości w przedziale pomiędzy 58% a 62%, nazywany jest **przedziałem ufności**.



Każdy z nas miewa do czynienia z wynikami zawierającymi błąd statystyczny i przedziały ufności, a przy współczesnym natłoku danych wiele osób wykorzystuje je także w pracy. Musisz wiedzieć, jakie czynniki wpływają na błąd statystyczny (np. liczebność próby), jakie cechy posiada prawidłowo określony przedział ufności i jak je rozpoznać. Powinieneś także umieć znajdować przedziały ufności, gdy będzie to konieczne.

Z rozdziału 12. dowiesz się wszystkiego, co powinieneś wiedzieć o błędzie statystycznym. Poznasz wszystkie jego składniki, dowiesz się, co mierzy, a czego nie i jak go obliczać w różnych sytuacjach. W rozdziale 13. poznasz krok po kroku wzory, obliczenia i interpretacje przedziałów ufności dla średniej z populacji, odsetka z populacji oraz różnicy między dwiema średnimi i odsetkami.

### Testowanie statystyczne

Jeden z głównych obszarów badań naukowych nosi nazwę testowania statystycznego. **Test statystyczny** jest techniką wykorzystywania danych, aby przyjąć lub odrzucić twierdzenie dotyczące populacji. Pewien polityk mógłby na przykład twierdzić, że 80% społeczeństwa się z nim zgadza — ale czy to prawda? Jakaś firma mogłaby twierdzić, że dostarcza pizzę w czasie najwyżej 30 minut, ale czy tak jest rzeczywiście? Naukowcy prowadzący badania medyczne stale wykorzystują testy statystyczne, aby sprawdzać, czy dany lek jest skuteczny, porównywać nowe leki z istniejącymi pod względem skutków ubocznych albo badać, jaka dieta jest najbardziej efektywna dla określonych grup ludzi.



Najczęściej testowanymi cechami populacji są:

- ✓ Średnia z populacji (Czy przeciętny czas dostawy wynoszący 30 minut naprawdę taki jest?).
- ✓ Odsetek z populacji (Czy to prawda, że 80% wyborców popiera tego kandydata, czy też jest ich mniej?).
- ✓ Różnica między dwiema średnimi lub odsetkami z populacji (Czy to prawda, że przeciętny spadek masy ciała w ramach nowej diety jest o 10 kilogramów większy niż w przypadku najpopularniejszego programu odchudzania? Albo czy to prawda, że nowy lek obniża ciśnienie krwi o 10% skuteczniej niż obecnie stosowany?).



Testy statystyczne stosuje się w wielu dziedzinach, które mają wpływ na nasze codzienne życie, takich jak badania medyczne, reklama czy dane wyborcze, i praktycznie wszędzie, gdzie dokonuje się porównań na podstawie średnich lub odsetków. Jeżeli zaś chodzi o konkretne branże, testy statystyczne są intensywnie wykorzystywane w takich branżach jak marketing, gdy chcemy ustalić, czy dana reklama jest skuteczna lub czy pewna grupa osób nabywa obecnie mniej, czy więcej naszych produktów w porównaniu z zeszłym rokiem.

Często poznajemy wyłącznie wnioski z testów statystycznych (na przykład: „Ten lek jest znacznie bardziej skuteczny i ma mniej skutków ubocznych niż lek, który stosujemy obecnie”); nie mamy jednak dostępu do metod, które doprowadziły do wyciągnięcia tych wniosków. W rozdziale 14. analizuję podstawy i szczegóły testów statystycznych, abyś mógł przeprowadzać je i weryfikować z przekonaniem. W rozdziale 15. przechodzę do sedna sprawy i prezentuję szczegółowe instrukcje dotyczące konfigurowania i przeprowadzania testów statystycznych w wielu konkretnych sytuacjach (jedna średnia z populacji, jeden odsetek z populacji, różnica dwóch średnich z populacji itd.).

Po przeczytaniu rozdziałów 14. i 15. poczujesz się znacznie pewniej, gdy będziesz musiał się dowiedzieć, do której grupy klientów należy skierować działania marketingowe związane z określonym produktem, jakiej marki opony są najtrwalsze albo czy określona dieta jest skuteczna. Będziesz mógł także odpowiedzieć na poważniejsze pytania, na przykład jaki zabieg chirurgiczny należy wybrać.

## Korelacja, regresja i tabele krzyżowe

Jednym z najczęstszych celów badań jest znajdowanie związków między zmiennymi, na przykład:

- ✓ Jakie zachowania związane ze stylem życia zwiększają lub zmniejszają ryzyko zachorowania na raka?
- ✓ Jakie działania niepożądane wywołuje nowy lek?
- ✓ Czy nowy ziołowy suplement diety obniża poziom cholesterolu?
- ✓ Czy spędzanie znacznej ilości czasu w internecie może prowadzić do przybierania na wadze?

Znajdowanie powiązań między zmiennymi pomaga naukowcom zajmującym się medycyną projektować lepsze leki i kuracje, dostarcza specjalistom od marketingu informacji o tym, kto chętnie kupi ich produkty, oraz informuje polityków, jakimi powinni posługiwać się argumentami, aby poprzeć określone działania lub się im sprzeciwić.



W sferze wielkiego biznesu związanego z poszukiwaniem zależności między zmiennymi można znaleźć niewiarygodną liczbę wyników statystycznych — ale czy można stwierdzić, które są prawidłowe, a które nie? Na podstawie takich badań podejmuje się wiele istotnych decyzji i ważne jest, aby mieć świadomość, jakie warunki muszą zostać spełnione, żeby ich wyniki można było uznać za wiarygodne. Dotyczy to zwłaszcza sytuacji, gdy stwierdza się istnienie związku przyczynowo-skutkowego.

W rozdziale 18. opisuję szczegóły i niuanse analizy danych pochodzących z zestawień dwóch zmiennych liczbowych (takich jak poziom dawkowania leku i ciśnienie krwi), omawiam kwestie znajdowania i interpretowania **korelacji** (siły i kierunku relacji liniowej pomiędzy zmiennymi  $x$  i  $y$ ), prezentuję sposoby znajdowania równania prostej, która najlepiej dopasowuje się do danych (i informuję, kiedy można to robić), oraz przedstawiam sposoby wykorzystywania uzyskanych wyników w celu prognozowania wartości danej zmiennej na podstawie innej (co nazywa się **regresją**). Przedstawiam także narzędzia umożliwiające ustalenie, kiedy linia jest dobrze dopasowana do danych, a kiedy nie i jakie wnioski można wyciągnąć (a jakich nie można) w sytuacjach, gdy linia rzeczywiście pasuje.

Metody wykorzystywane do poszukiwania i opisywania powiązań pomiędzy dwiema zmiennymi kategorycznymi (takimi jak liczba dawek leku przyjmowanych dziennie i obecność lub brak nudności) opisuję szczegółowo w rozdziale 19. Prezentuję także informacje o gromadzeniu i zestawianiu danych w **tabelach krzyżowych** (w których możliwe wartości jednej zmiennej tworzą wiersze, a możliwe wartości drugiej zmiennej tworzą kolumny), o interpretacji wyników, analizie danych z tabel krzyżowych w poszukiwaniu zależności, a także o sprawdzaniu wyników pod kątem niezależności. Dodatkowo, podobnie jak w całej książce, prezentuję strategię krytycznej oceny wyników tego rodzaju analiz z punktu widzenia ich wiarygodności.

## Wyciąganie wiarygodnych wniosków



Do przeprowadzania analiz statystycznych badacze używają oprogramowania statystycznego, które opiera się na wzorach. Wzory nie wiedzą jednak, czy są wykorzystywane właściwie, i nie ostrzegą Cię, gdy wyniki będą nieprawidłowe. Co więcej, komputery nie są w stanie stwierdzić, co oznaczają wyniki; sam musisz to ustalić. Z tej książki dowiesz się, jakiego rodzaju wnioski możesz, a jakich nie możesz wyciągać po przeprowadzeniu analizy. Poniższe podrozdziały stanowią wstęp do wyciągnięcia odpowiednich wniosków.

### Podawanie zawyżonych wyników

Jednym z najczęstszych błędów popełnianych podczas wnioskowania jest zawyżanie wyników lub uogólnianie ich dla większej grupy niż rzeczywiście reprezentowana w badaniu. Pani profesor mogłaby na przykład chcieć ustalić, które reklamy nadawane podczas finału Super Bowl najbardziej podobają się widzom. Gromadzi więc setkę swoich studentów w finałową niedzielę i prosi ich, aby oceniali pojawiające się reklamy. Studenci tworzą listę pięciu najlepszych i pani profesor stwierdza, że wszystkim widzom Super Bowl te pięć reklam podobało się najbardziej. Ale wie przecież tylko, które z nich najbardziej podobały się *jej studentom* — nie badała innych grup, więc nie może wyciągać wniosków dotyczących wszystkich widzów.

### Kwestionowanie twierdzeń o przyczynie i skutku

Łatwo wyciągnąć nieprawidłowe wnioski w sytuacji, gdy badacze stwierdzają, że dwie zmienne są ze sobą powiązane (za pośrednictwem analizy takiej jak regresja; zob. wcześniejszy podrozdział „Korelacja, regresja i tabele krzyżowe”, aby uzyskać więcej informacji), a następnie automatycznie dochodzą do wniosku, że te dwie zmienne łączą związek przyczynowo-skutkowy.

Zalóżmy na przykład, że naukowiec przeprowadził ankietę dotyczącą zdrowia i odkrył, że osoby, które przyjmowały codziennie witaminę C, zgłaszają mniej przeziębień niż takie, które tej witaminy codziennie nie przyjmowały. Po ustaleniu wyników, wykorzystując te dane jako dowód, badacz pisze referat i publikuje informację prasową, zgodnie z którą witamina C zapobiega przeziębieniom.

Otóż, chociaż może być prawdą, że witamina C zapobiega przeziębieniom, badanie tego naukowca nie może tego stwierdzić. Badanie było obserwacyjne, co oznacza, że nie kontrolowano żadnych innych czynników, które mogą być związane zarówno z witaminą C, jak i przeziębieniami. Być może na przykład osoby, które przyjmują codziennie witaminę C, są ogólnie bardziej świadome w kwestiach zdrowotnych, częściej myją ręce, więcej ćwiczą i lepiej się odżywiają; wszystkie te zachowania mogą być pomocne w zmniejszaniu ich skłonności do przeziębień.

Dopóki nie przeprowadzimy kontrolowanego eksperymentu, nie możemy wyciągać wniosków przyczynowo-skutkowych na podstawie relacji, które znaleźliśmy. (Eksperymenty omawiam bardziej szczegółowo we wcześniejszej części tego rozdziału).





## Jak być detektywem, a nie sceptykiem

Statystyka to znacznie więcej niż liczby. Aby naprawdę „załapać” statystykę, musisz zrozumieć, jak wyciągać odpowiednie wnioski na podstawie badania danych i mieć na tyle oleju w głowie, aby nie wierzyć we wszystko, co usłyszysz lub przeczytasz, zanim nie ustalisz, skąd wzięły się dane informacje, co z nimi robiono i w jaki sposób wyciągano wnioski. Omawiam te kwestie w całej książce, ale koncentruję się na nich w rozdziale 20., w którym przedstawiam dziesięć sposobów umożliwiających nabycie umiejętności detektywa w sferze statystyki dzięki rozpoznawaniu typowych błędów popełnianych przez naukowców i media.



Informacja dla uczniów i studentów: w rozdziale 21. omawiam przydatne na egzaminie praktyki statystyczne i doradzam, jak poprawić swoje wyniki. Większość moich porad opiera się na zrozumieniu ogólnych założeń omawianych zagadnień oraz szczegółów związanych z prawidłowym rozwiązywaniem problemów statystycznych.



Zajmowanie sceptycznego czy też cynicznego stanowiska wobec statystyki jest bardzo łatwe, zwłaszcza po tym, gdy dowiemy się, co ma miejsce za kulisami badań; nie pozwól jednak, żeby przytrafiło się to Tobie. Dzięki statystykom można znaleźć wiele przydatnych informacji, które mogą w pozytywny sposób wpłynąć na nasze życie. Skanalizuj swój sceptycyzm, stawiając sobie dwa cele:

- ✓ Zostań dobrze poinformowanym konsumentem informacji statystycznych, z którymi stykasz się na co dzień.
- ✓ Ustal swoją pozycję w pracy jako kogoś, kogo inni radzą się w kwestiach statystycznych i kto wie, kiedy i jak może pomóc innym, a kiedy powinien poszukać statystyka.

Czytając i wykorzystując informacje zawarte w tej książce, zyskasz pewność, że potrafisz podejmować prawidłowe decyzje dotyczące wyników statystycznych. Nauczysz się w rzetelny sposób prowadzić własne badania statystyczne i będziesz gotów zająć się kolejnym projektem w pracy, krytycznie ocenić irytującą reklamę polityczną lub z wyróżnieniem zdać kolejny egzamin!



# Skorowidz

---

## A

analiza  
  danych, 280  
  statystyczna, 34  
  trendów, 109  
ankieta, 74  
aproksymacja rozkładu dwumianowego, 167

## B

badanie  
  eksperymentów, 269  
  obserwacyjne, 29, 268  
  podstaw rozkładu normalnego, 155  
  środką danych, 86  
  właściwości korelacji, 292  
  wykresów pudełkowych, 128, 132  
  wyników testu ACT, 100  
  zachowań, 173  
  rozstępu ćwiartkowego, 103  
błąd  
  standardowy, 177, 178, 193  
  dla populacji, 179  
  statystyczny, 35, 75, 191, 328  
  dla odsetka z próby, 194  
  dla średniej z próby, 196  
błędne korelacje, 330  
błędy  
  drugiego rodzaju, 231  
  pierwszego rodzaju, 231  
  próbkiowania, 75

## C

cechy populacji, 36  
cel sondażu, 257  
centralne twierdzenie graniczne, 72, 181, 185

centrum, 132  
ciągłe zmienne losowe, 144  
częstość, 32  
czynnik, 268

## D

dane, 63  
  ciągłe, 64  
  dwuwymiarowe, 286  
  dyskretne, 64  
  ilościowe, 64  
  jakościowe, 64, 84  
  kategoryczne, 64, 84, 105  
  liczbowe, 32, 64, 117  
  obiektywne, 279  
  poprawne, 278  
  porządkowe, 64  
  symetryczne, 90  
  tendencyjne, 327  
  wiarygodne, 278  
definiowanie  
  liczebności próby, 273  
  populacji docelowej, 257  
  rozkładu z próby, 175  
  wartości  $p$ , 227  
dobór właściwej próby, 30  
dwumian, 146  
dyskretne zmienne losowe, 144

## E

efekt  
  marginalny, 247  
  placebo, 74, 271  
eksperymenty, 29, 268, 270  
  podwójnie ślepe, 74  
  ślepe, 74

ekstrapolacja, 298  
etapy  
  badania, 62  
  znajdowania prawdopodobieństwa, 161

## F

formułowanie hipotez, 222

## G

grafy, 33  
gromadzenie  
  dobrych danych, 278  
  dowodów, 223  
  statystyk z prób, 224  
  wartościowych danych, 30  
grupa, 125  
  doświadczalna, 74  
  kontrolna, 74

## H

hipoteza, 222  
  alternatywna, 77  
  zerowa, 77, 222, 234  
histogram, 89, 117, 326  
  lewoskośny, 124  
  mylący, 125  
  prawoskośny, 124  
  zbliżony do symetrycznego, 124

## I

identyfikowanie  
  dwumianu, 146  
  prób nielosowych, 328  
interpretacja  
  histogramu, 120  
  korelacji, 290  
  linii regresji, 295  
  nachylenia, 295  
  odchylenia standardowego, 93  
  percentyli, 99  
  punktu przecięcia, 296  
  statystyki testowej, 225  
  tabel krzyżowych, 304  
  wykresu czasowego, 135

  wykresu pudełkowego, 130, 134  
  wykresu rozrzutu, 287  
  wyników, 204, 317  
istotność  
  statystyczna, 78  
  testu, 228

## K

kontrola zmiennych zakłócających, 275  
korelacja, 37, 79, 288, 292  
krzywa dzwonowa, 71  
kształt danych, 120

## L

liczba  
  grup, 125  
  prób, 148  
liczebność próby, 32, 57, 178, 198, 207  
  mała, 236  
  ustalanie, 329  
  ustalanie pożądanej wartości, 211  
  wybór, 273  
linia regresji, 292, 293  
losowe generowanie numerów, 67  
losowość, 66, 143

## M

małe próby, 235  
mediana, 69, 88, 125  
metoda najmniejszych kwadratów, 294  
moc testu, 232  
mylące  
  histogramy, 125  
  przedziały ufności, 219  
  statystyki, 59  
  wykresy, 323  
  wykresy czasowe, 136  
  wyniki, 267

## N

nadmiar danych, 137  
niezależność próby, 148, 315  
nieznane odchylenia standardowe, 235

## 0

obiekty, 268  
 obliczanie, 81
 

- błędu statystycznego, 194
  - dla średniej z próby, 196
- danych kategorycznych, 84
- korelacji, 289
- liczebności próby, 212
- odchylenia standardowego, 91
- percentyli, 98
- prawdopodobieństwa, 149
- przedziału ufności, 209
- rozkładów
  - brzegowych, 305
  - łącznych, 308
  - warunkowych, 310
- statystyki testowej, 225
- wartości p, 227

 obserwacja
 

- badania obserwacyjnych, 269
- skali, 127, 136

 ocena
 

- dowodów, 226
- prawdopodobieństwa złej decyzji, 230

 odchylenie standardowe, 69, 91, 179
 

- interpretacja, 93
- populacji
  - nieznane, 209, 216
  - znane, 210, 215
- właściwości, 93
- zmiennej o rozkładzie
  - dwumianowym, 153

 odrzucanie hipotezy zerowej, 229
 odsetek
 

- z populacji, 36, 185
- z próby, 185, 187

 ogłaszanie wyników, 196
 określenie celu sondażu, 257
 opis
 

- relacji zależnej, 316
- zależności, 313

## P

parametr, 68, 202  
 percentyle, 70, 97, 165
 

- dla rozkładu normalnego, 164

planowanie sondażu, 256  
 podejmowanie decyzji, 226, 282  
 podsumowanie pięcioliczbowe, 101  
 pomiar
 

- błędu standardowego, 177
- pozycji względnej, 97
- rozproszenia, 224
- zróznicowania próby, 132, 193

 populacja, 36, 65
 

- docelowa, 256, 257

 porady, 323, 335  
 porażka, 148  
 porównywanie
 

- dwóch niezależnych średnich, 240
- dwóch odsetków z populacji, 245
- grup, 310
- rozkładów, 169, 315

 poszukiwania błędu statystycznego, 328  
 poziom, 268
 

- alfa, 228
- istotności testu, 228
- realizacji próby, 263
- ufności, 193, 203

 pozycja względna, 97  
 prawdopodobieństwo, 145
 

- dwumianowe, 149
- porażki, 149
- sukcesu, 149
- testowe, 226
- złej decyzji, 230

 prawidłowa analiza danych, 280  
 prognoza, 297, 319  
 projektowanie
 

- eksperymentu, 270
- sondażu, 256

 prowadzenie sondażu, 262  
 próba, 260
 

- kliniczna, 277
- losowa, 66, 260
- nielosowa, 66, 328
- podwójnie ślepa, 279

 przedział ufności, 35, 72, 76, 173, 201
 

- dla różnicy dwóch średnich, 214
- dla odsetka z populacji, 213
- dla różnicy dwóch odsetków z populacji, 217
- dla różnicy średnich, 216
- dla średniej z populacji, 209

 przyczyna i skutek, 38, 269, 298, 317

przyczynowość, 79  
przypisywanie losowe, 274  
pułapka tak-tak, 337  
punkt  
końcowy, 127  
początkowy, 127  
przebieg, 71

## R

reakcja, 268  
regresja, 37  
liniowa, 292  
reguła trzech sigm, 70, 95  
relacja zależna, 316  
reszta, 298  
rozkład, 34, 71, 175, 304  
brzegowy, 305  
dwumianowy, 34, 147, 149  
odchylenie standardowe, 153  
średnia, 153  
wariancja, 153  
łączny, 308  
normalny, 34, 70, 71, 155  
prawdopodobieństwa, 145, 151  
t-Studenta, 34, 169  
warunkowy, 310  
X  
normalny, 181  
nie jest normalny, 181  
Z, 71, 158  
z próby, 175, 181, 185  
rozproszenie, 91  
rozstęp ćwiartkowy, 132

## S

selektywne raportowanie, 332  
skała, 55  
na wykresie, 113  
skośność  
lewostronna, 90  
prawostronna, 90  
sondaże, 29, 74, 251  
formułowanie pytań, 259  
gromadzenie danych, 262  
interpretacja wyników, 264  
określenie celu, 257  
poziom realizacji próby, 263

wybór próby, 260  
wybór rodzaju, 258  
spis powszechny, 68  
sprawdzanie  
kształtu danych, 120  
liczb, 332  
niezależności, 313, 315  
obliczeń, 52  
źródeł, 57  
standardowy rozkład normalny, 71, 73, 158  
standaryzacja, 159, 163  
X do Z, 159  
statystyka, 27, 67, 202  
myląca, 59  
opisowa, 32, 83  
testowa, 224  
dla pojedynczego odsetka, 238  
w miejscu pracy, 48  
w życiu codziennym, 41  
stopnie swobody, 170  
stosowanie testu t-Studenta, 236  
sukces, 148  
suma  
kwadratów błędów, 294  
łączna, 304  
sumy brzegowe  
kolumn, 303  
wierszy, 303  
szacowanie różnicy dwóch odsetków, 217  
szacunki, 68

## Ś

średnia, 87, 125  
z populacji, 36, 69  
z próby, 69, 177, 184  
zmiennej losowej, 146

## T

tabela krzyżowa, 37, 84, 302  
tablica  
dwumianowa, 151, 354  
t-Studenta, 172, 236, 352  
Z, 160, 349  
tendycyjność  
danych, 31, 68, 327  
odpowiedzi, 262

- test  
 ACT, 100  
 statystyczny, 36, 77, 233  
 t-Studenta, 235, 236  
 dla zmiennych zależnych, 242
- testowanie  
 jednej średniej z populacji, 233  
 pojedynczego odsetka z populacji, 238  
 statystyczne, 221  
 średniej z różnicy, 242
- trend, 109
- tworzenie  
 grupy, 119  
 histogramu, 117  
 rozważnej prognozy, 319  
 tabeli krzyżowej, 302  
 wykresu czasowego, 135  
 wykresu pudełkowego, 128  
 wykresu rozrzutu, 287
- U**
- upraszczanie nadmiaru danych, 137
- ustalanie  
 błędu statystycznego, 193  
 brakujących liczebności prób, 329  
 istnienia zależności liniowych, 288  
 percentyli dla rozkładu t-Studenta, 172  
 położenia centrum, 132  
 pożądanej wartości liczebności próby,  
 211  
 przedziału ufności, 213  
 znaczenia liczebności próby, 198
- uśrednianie do średniej, 87
- uwzględnianie liczebności próby, 207
- W**
- wariancja zmiennej losowej, 146
- wartości  
 graniczne, 120  
 odstające, 133  
 percentyli  
 niskie, 165  
 wysokie, 165  
 ujemne t, 237
- wartość  
 p, 77, 226, 227  
 t, 236
- $t^*$ , 173  
 z, 73, 158  
 $z^*$ , 194, 206
- weryfikacja hipotezy, 222
- właściwości  
 odchylenia standardowego, 93  
 współczynnika korelacji, 292
- wnioski, 228, 280, 320
- wskazywanie mylących wykresów, 323
- wskaźnik, 54
- współczynnik korelacji, 289, 292  
 Pearsona, 289
- wybór  
 liczebności próby, 273  
 obiektów, 274  
 poziomu ufności, 205  
 próby, 260
- wyciąganie wniosków, 38, 228, 280
- wykres, 33  
 czasowy, 135, 326  
 kołowy, 105, 325  
 pudełkowy, 128  
 słupkowy, 110, 325  
 słupkowy skumulowany, 311
- wykresy  
 danych kategorycznych, 105  
 danych liczbowych, 117  
 rozkładów  
 brzegowych, 306  
 łącznych, 309  
 warunkowych, 311  
 rozrzutu, 286, 287  
 rozrzutu z korelacjami, 291
- wykrywanie  
 błędów, 52  
 mylących statystyk, 53  
 tendencyjnych danych, 327
- wynik  
 standaryzowany, 70, 225  
 statystycznie istotny, 78, 2069
- wyniki sondażu, 264
- wyodrębnianie zmiennych, 305
- wzór, 339  
 konwersji, 184  
 na błąd statystyczny  
 dla odsetka z próby, 194  
 dla średniej z próby, 196  
 na liczebność próby, 211

wzór

na prawdopodobieństwo dwumianowe,  
149

na przedział ufności dla różnicy dwóch  
średnich, 215

na standaryzację wartości, 159

na średnią rozkładu dwumianowego, 153

na wariancję rozkładu dwumianowego,  
153

na wartość  $t$ , 236

## Z

zabieg, 268

zadania z percentylami, 166

zakres, 95, 204

zależność, 316

liniowa, 288

zbiór

bimodalny, 126

danych, 64

zestawianie dowodów, 224

zmienna, 34, 65

kategoryczna, 301

niezależna, 313

losowa, 143, 175

ciągła, 144

dyskretna, 144

objaśniająca, 293

odpowiedzi, 293

zakłócająca, 276, 331

znaczenie sondaży, 252

znajdowanie

linii regresji, 293

mylących przedziałów ufności, 219

prawdopodobieństw

dla odsetka z próby, 187

dla rozkładu normalnego, 161

dla średniej z próby, 184

dla  $Z$ , 160

dla  $X$  o dowolnej wartości, 152

dla  $X$  o określonej wartości, 152

dwumianowych, 149, 151

$X$ , 164

zróżnicowanie, 170

populacji, 208

próby, 193

związek

pomiędzy  $t$  i  $Z$ , 237

przyczynowo-skutkowy, 38, 269, 298, 317

## Ź

źródła informacji

medialne, 51

pierwotne, 51



# PROGRAM PARTNERSKI

— GRUPY HELION —



1. ZAREJESTRUJ SIĘ
2. PREZENTUJ KSIĄŻKI
3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW w działający bankomat!

**Dowiedz się więcej i dołącz już dzisiaj!**

<http://program-partnerski.helion.pl>

GRUPA  
**Helion** 

# Prosty sposób na statystykę

Czujesz się bezradny w obliczu statystyki? Bez obaw! Z tego przystępnego przewodnika dowiesz się, jak należy gromadzić i weryfikować dane, a także prezentować je na wykresach. Ponadto nauczysz się rozszyfrowywać rozkłady, obliczać przedziały ufności, weryfikować hipotezy, analizować dane z wykorzystaniem korelacji, regresji, tabel krzyżowych — i wiele więcej!

## W książce:

- proste objaśnienia terminów statystycznych
- informacje o zestawianiu i weryfikacji danych
- zmienne losowe i rozkłady
- przeprowadzanie, interpretowanie i weryfikacja sondży i eksperymentów

## Deborah J. Rumsey

jest profesorem na Wydziale Statystyki i specjalistą do spraw nauczania statystyki na Uniwersytecie Stanowym Ohio. Autorka książek *Statistics Workbook For Dummies*, *Statistics II For Dummies* i *Probability For Dummies*.

Cena: 49,90 zł

ISBN 978-83-8322-583-8



9 788383 225838

dla  
**bystrzaków**