

## Wizualizacja danych

## 1

## Pierwsze wrażenie

## Czy masz problem ze zrozumieniem faktów zapisanych w danych?

Statystyki opisowe pomogą Ci zrozumieć znaczenie złożonych zbiorów danych. Dzięki nim **bardzo złożone kwestie** staną się **zupełnie proste**. Gdy już je zrozumiesz i będziesz chciał **podzielić się tą wiedzą z innymi**, z pewnością zainteresuje Cię potencjał licznych technik **wizualizacji danych**, jakie oferuje statystyka. Jeśli chcesz więc dobrać optymalny wykres dla swoich danych, weź w rękę swój płaszcz, spakuj ulubiony suwak logarytmiczny i ruszaj z nami po przygodę do Statsville.



Teraz widzisz,  
że miałam rację: zyski  
firmy są co miesiąc  
takie same.



Nieprawda,  
są wręcz  
zdmiewające.  
Spójrz, jak szybuja  
w górę!



|                                              |    |
|----------------------------------------------|----|
| Statystyki są wszędzie                       | 40 |
| Co Ci dadzą statystyki?                      | 41 |
| Jak to z wykresami było                      | 42 |
| Prosty, lecz bardzo użyteczny wykres kołowy  | 46 |
| Wykres słupkowy jest bardziej precyzyjny     | 48 |
| Wykres kolumnowy                             | 48 |
| Wykres wierszowy                             | 49 |
| Wszystko jest kwestią odpowiedniej skali     | 50 |
| Wykorzystanie skali bezwzględnej             | 51 |
| Praca z bardziej złożonymi zbiorami danych   | 52 |
| Kategorie a liczby                           | 56 |
| Praca z danymi zgrupowanymi                  | 57 |
| Konstruujemy histogram                       | 58 |
| Krok 1: Określ szerokość słupków histogramu  | 64 |
| Krok 2: Dostosuj wysokość słupków histogramu | 65 |
| Krok 3: Wykreśl swój histogram               | 66 |
| Czym są częstości skumulowane                | 72 |
| Jak wykreślić częstości skumulowane          | 73 |
| Jak wybrać odpowiedni typ wykresu            | 77 |

## Miary tendencji centralnej

## 2 Droga Środka

**Niekiedy trzeba po prostu dotrzeć do sedna sprawy.** Czasami trudno ogarnąć ogrom informacji ukrytych w gąszczu danych. Pierwszym krokiem może być wówczas wyznaczenie średnich. Statystycy nazywają je miarami tendencji centralnej. Dzięki nim potrafią szybko wskazać wartości reprezentatywne dla danego zbioru danych i na tej podstawie wyciągnąć ważne wnioski. W tym rozdziale nauczysz się wyznaczać wartości kilku najważniejszych i najpopularniejszych statystyk — średniej, mediany i dominanty. Zobaczysz, jak łatwo i efektywnie można za ich pomocą dokonać **opisu danych**.

The Health Club  
Statsville's Premier Spa



20 lat



21 lat



20 lat



19 lat



|                                                     |     |
|-----------------------------------------------------|-----|
| Witamy w ośrodku odnowy                             | 84  |
| Najpopularniejszą średnią jest średnia arytmetyczna | 85  |
| W świecie symboli                                   | 86  |
| Jak sobie radzić z niewiadomymi                     | 87  |
| Wracamy do średniej                                 | 88  |
| Wróćmy do naszego klubu                             | 91  |
| Każdy ćwiczył kiedyś kung-fu                        | 92  |
| W naszych danych są wartości nietypowe              | 95  |
| Czym są dane asymetryczne                           | 96  |
| Rozmowa przy dystrybutorze                          | 98  |
| Z pomocą przychodzi nam mediana                     | 99  |
| Jak znaleźć medianę w trzech prostych krokach       | 100 |
| Nasz interes kwitnie                                | 103 |
| Nauka pływania dla... najmłodszych                  | 104 |
| Dlaczego średnia i mediana nie są miarodajne?       | 107 |
| Jak możemy sobie radzić z danymi tego typu?         | 107 |
| Cała prawda o średniej arytmetycznej                | 109 |
| Prezentujemy dominantę (modę)                       | 111 |
| Jak znaleźć dominantę w trzech prostych krokach     | 112 |

## Miary zróżnicowania

## 3 Potęgą zmienności

**Nie wszystkiemu można wierzyć, ale jak się o tym przekonać?**

Średnie pozwalają nam poznać typową wartość dla naszych danych, ale **nie mówią nam wszystkiego**. Umiemy już znajdować wartości centralne zbioru danych, ale średnia arytmetyczna, mediana czy dominanta nie zawsze wystarczają do wyciągnięcia głębszych wniosków. W tym rozdziale poszerzymy naszą wiedzę o narzędzia, dzięki którym będziemy mogli coś powiedzieć o **zróżnicowaniu** naszych danych.



Wszyscy trzej mają tę samą średnią punktów, więc potrzebuję innej wskazówki, która pomoże mi wybrać najlepszego. Czy masz jakiś pomysł?



|                                                              |     |
|--------------------------------------------------------------|-----|
| W poszukiwaniu zawodnika                                     | 122 |
| Musimy porównać wyniki kandydatów                            | 123 |
| O czym mówi rozstęp                                          | 124 |
| Obserwacje nietypowe rodzą pewien problem                    | 127 |
| Musimy znaleźć sposób na pozbycie się obserwacji nietypowych | 129 |
| Na ratunek spieszą kwartyle                                  | 130 |
| Rozstęp międzykwartylowy wyklucza obserwacje nietypowe       | 131 |
| Anatomia kwartyli                                            | 132 |
| Nie musimy ograniczać się tylko do kwartyli                  | 136 |
| Czym są percentyle?                                          | 137 |
| Wykres pudełkowy dobrze prezentuje rozproszenie danych       | 138 |
| Zmienność to coś więcej niż tylko rozstęp                    | 142 |
| Jak obliczyć odchylenie od średniej                          | 143 |
| Zmienność możemy zmierzyć za pomocą wariancji...             | 144 |
| ...ale odchylenie standardowe jest miarą bardziej intuicyjną | 145 |
| Cała prawda o odchyleniu standardowym                        | 146 |
| Szybszy sposób na wariancję                                  | 151 |
| A gdybyśmy potrzebowali punktu odniesienia dla porównań?     | 156 |
| Standaryzacja danych sposobem na ich porównywanie            | 157 |
| Jak interpretować dane wystandaryzowane                      | 158 |
| Nasza drużyna mistrzem!                                      | 163 |

## Prawdopodobieństwo zdarzeń

4

## Natura ryzyka

**Życie pełne jest niepewności.** Czasami trudno jest nawet przewidzieć, co wydarzy się w ciągu najbliższych paru minut. Szanse zajścia pewnych zdarzeń są jednak większe niż innych, czego uczy nas **rachunek prawdopodobieństwa**. Szacowanie prawdopodobieństwa zdarzeń ułatwia **przewidywanie przyszłości**, ponieważ pozwala ocenić, jak duże są szanse ich wystąpienia. A to pozwala podejmować bardziej **świadome wybory**. W tym rozdziale dowiesz się, czym jest prawdopodobieństwo zdarzeń i jak może Ci ono pomóc zapanować nad przyszłością!

[illegible]

|                                                                          |     |
|--------------------------------------------------------------------------|-----|
| Wielki Szlem                                                             | 166 |
| Wejdź do gry!                                                            | 167 |
| Jakie są moje szanse?                                                    | 170 |
| Znajdujemy prawdopodobieństwo wygranej w ruletkę                         | 173 |
| Do czego przydają się diagramy Venna                                     | 174 |
| Możesz także dodać prawdopodobieństwa                                    | 180 |
| Zdarzenia rozłączne                                                      | 185 |
| Gdy część wspólna sprawia problem                                        | 186 |
| Trochę notacji                                                           | 187 |
| Znowu nieudany obrót...                                                  | 193 |
| Prawdopodobieństwo warunkowe                                             | 194 |
| Obliczamy prawdopodobieństwa warunkowe                                   | 195 |
| Prawdopodobieństwa warunkowe można przedstawić na drzewie stochastycznym | 196 |
| Drzewa są pomocne w obliczaniu prawdopodobieństw                         | 197 |
| Krok 1: Znajdujemy $P(\text{czarne} \cap \text{parzyste})$               | 205 |
| Krok 2: Znajdujemy $P(\text{parzyste})$                                  | 207 |
| Krok 3: Znajdujemy $P(\text{czarne}   \text{parzyste})$                  | 208 |
| Wykorzystaj prawdopodobieństwo całkowite, by znaleźć $P(B)$              | 210 |
| Twierdzenie Bayesa                                                       | 211 |
| Gdy zdarzenia wpływają na siebie, są zdarzeniami zależnymi               | 219 |
| Jeśli zdarzenia nie wpływają na siebie, są niezależne                    | 220 |
| Kilka słów o liczeniu prawdopodobieństw dla zdarzeń niezależnych         | 221 |





## Dyskretne rozkłady prawdopodobieństwa

## 5

## Zarządzamy oczekiwaniami

**Zdarzenia mało prawdopodobne czasem się zdarzają, ale jakie są ich konsekwencje?**

W poprzednim rozdziale przekonaliśmy się, jak rachunek prawdopodobieństwa może pomóc nam ocenić szanse zajścia pewnych zdarzeń. Jednak nie powie nam on nic na temat **wpływu** tych zdarzeń na nasze życie. Choć przy stole ruletki czasem pada spora wygrana, to jednak — czy jest ona warta tych wszystkich pieniędzy, jakie przy okazji można stracić? W tym rozdziale pokażemy Ci, jak można posłużyć się rachunkiem prawdopodobieństwa do **przewidywania skutków powtarzalnych zdarzeń w długich seriach**, a także jak można **ocenić dokładność** takich prognoz.



|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| Wracamy do kasyna Dana                                                  | 236 |
| Tworzymy rozkład prawdopodobieństwa wygranej na automacie               | 239 |
| Wartość oczekiwana pozwala przewidzieć wynik...                         | 242 |
| ...a wariancja mówi o tym, jak bardzo jest on zmienny                   | 243 |
| Wariancja a rozkład prawdopodobieństwa                                  | 244 |
| Obliczamy wariancję dla naszego przykładu                               | 245 |
| Gdy ceny idą w górę                                                     | 250 |
| Miedzy $E(X)$ i $E(Y)$ istnieje związek liniowy                         | 255 |
| Podsumujmy nasze rozważania                                             | 256 |
| Ogólne wzory na przekształcenia liniowe                                 | 257 |
| Każde pociągnięcie dźwigni jest niezależnym zdarzeniem                  | 260 |
| Przydatne skróty                                                        | 261 |
| Nowe automaty wchodzą do gry!                                           | 267 |
| Dodaj $E(X)$ do $E(Y)$ , by uzyskać $E(X+Y)$ ...                        | 268 |
| ...lub odejmij $E(Y)$ od $E(X)$ , by uzyskać $E(X-Y)$                   | 269 |
| Podobne operacje możesz wykonywać na zmiennych przekształcanych liniowo | 270 |
| Rozbiłeś bank!                                                          | 276 |

## Podstawy kombinatoryki

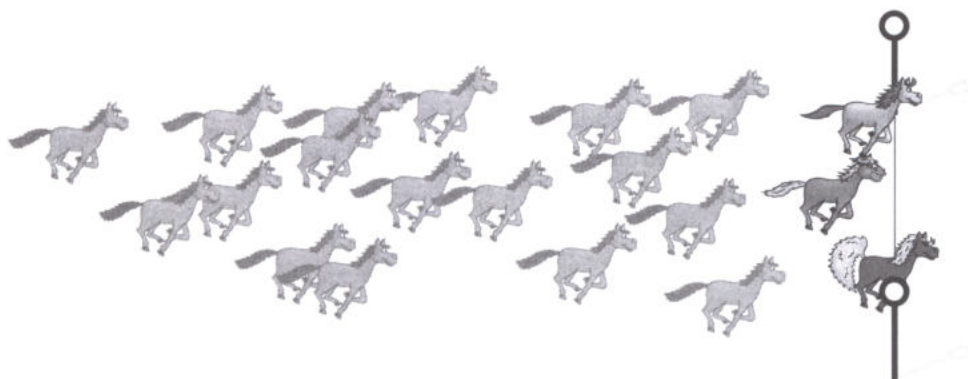
## 6

## Porządkujemy obiekty

**Czasami kolejność ma znaczenie.** Policzenie **wszystkich możliwych sposobów** grupowania czy porządkowania pewnego zbioru obiektów może być niezwykle pracochłonne. Często jednak nie mamy wyjścia, bo takie informacje są **kluczowe** dla rachunku prawdopodobieństwa. W tym rozdziale poznamy **szybki i efektywny** sposób na zdobycie tego rodzaju informacji, który nie wymaga od nas znajomości wszystkich możliwych wyników doświadczenia losowego. Zostań więc z nami, a przekonasz się, jak łatwo można **zliczać wszystkie możliwości grupowania i porządkowania obiektów**.



|                                                                  |     |
|------------------------------------------------------------------|-----|
| Derby Statsville                                                 | 280 |
| Wyścig trójki koni                                               | 281 |
| Na ile sposobów konie mogą przekroczyć linię mety?               | 283 |
| Zliczamy możliwe ustawienia zwycięzców                           | 284 |
| Ustawiamy obiekty w okrąg                                        | 285 |
| Czas na wyścig nowicjuszy                                        | 289 |
| Porządkowanie klas to coś innego niż porządkowanie ich elementów | 290 |
| Porządkujemy zwierzęta według gatunku                            | 291 |
| Ogólna formuła na liczbę uporządkowań w przypadku powtórzeń      | 292 |
| Czas na wyścig dwudziestu koni                                   | 295 |
| Na ile sposobów możemy zapłacić trzy miejsca medalowe?           | 296 |
| Obliczamy wariacje                                               | 297 |
| Gdy kolejność nie ma znaczenia                                   | 298 |
| Liczmy kombinacje                                                | 299 |
| Cała prawda o kombinacjach                                       | 300 |
| To już koniec zawodów                                            | 306 |



## Poznajemy rozkłady: geometryczny, dwumianowy i Poissona

## 7 Stajemy się dyskretni

**Wyznaczanie rozkładów prawdopodobieństwa zabiera sporo czasu.**

Wiemy już, jak wyznaczać rozkłady prawdopodobieństwa, choć nie zawsze jest to łatwe. Pewnie więc zgodzisz się z nami, że byłoby dobrze, gdyby istniały ogólne, bardziej uniwersalne rozkłady prawdopodobieństwa, którymi **łatwo** można by się posługiwać na co dzień. W tym rozdziale zaprezentujemy kilka **standardowych rozkładów prawdopodobieństwa**, o ściśle określonych charakterystykach. Gdy zrozumiesz, jak to działa, będziesz w stanie **w rekordowo szybkim czasie obliczać prawdopodobieństwa, wartości oczekiwane i wariancje**. Przygotuj się więc na poznanie rozkładów: geometrycznego, dwumianowego i Poissona. Miłej lektury.

|                                                                           |     |
|---------------------------------------------------------------------------|-----|
| Znajdujemy rozkład prawdopodobieństwa dla osiągnięć Chada                 | 311 |
| Istnieje rozkład prawdopodobieństwa,<br>który dobrze opisuje nasz problem | 312 |
| Prawdopodobieństwo możemy przedstawić za pomocą wzoru                     | 315 |
| Rozkład geometryczny pozwala operować także<br>na nierównościach          | 317 |
| Wartość oczekiwana dla rozkładu geometrycznego                            | 318 |
| Wartość oczekiwana wynosi $1/p$                                           | 319 |
| Wariancja dla rozkładu geometrycznego                                     | 321 |
| Krótki przewodnik po rozkładzie geometrycznym                             | 322 |
| Właśnie poznałeś rozkład geometryczny                                     | 325 |
| Arcyfrajerzy                                                              | 327 |
| Lepiej grać czy jednak zrezygnować?                                       | 329 |
| Uogólniamy rozkład na więcej niż trzy przypadki                           | 331 |
| Uogólniamy nasze wzory jeszcze bardziej                                   | 334 |
| Ile wynosi wartość oczekiwana i wariancja dla tego rozkładu               | 336 |
| Wartość oczekiwana i wariancja rozkładu dwumianowego                      | 339 |
| Krótki przewodnik po rozkładzie dwumianowym                               | 340 |
| Wartość oczekiwana i wariancja dla rozkładu Poissona                      | 346 |
| Jaki jest więc rozkład prawdopodobieństwa?                                | 350 |
| Liczymy prawdopodobieństwa zdarzeń złożonych<br>dla rozkładu Poissona     | 351 |
| Rozkład Poissona w przebraniu                                             | 354 |
| Krótki przewodnik po rozkładzie Poissona                                  | 357 |

Automat  
do popcornuAutomat  
do napojów



## Poznajemy rozkład normalny

## 8

## Wybieramy normalność

**Dyskretny rozkłady prawdopodobieństwa nie w każdej sytuacji się sprawdzają.**

W poprzednim rozdziale poznaliśmy trzy rozkłady prawdopodobieństwa, dla których byliśmy w stanie wymienić wszystkie wartości, jakie może przyjąć zmienna losowa. Nie zawsze jednak jest to możliwe.

Niekiedy posiadane przez nas dane w ogóle **nie przystają** do żadnego z tych trzech rozkładów.

W tym rozdziale dowiemy się o istnieniu rozkładów innego typu, tak zwanych **ciągłych rozkładów prawdopodobieństwa**, oraz poznamy jeden z najważniejszych rozkładów w statystyce — **rozkład normalny**.



|                                                                   |     |
|-------------------------------------------------------------------|-----|
| Zmienne dyskretne przyjmują wybrane wartości...                   | 364 |
| ...ale nie wszystkie zmienne muszą być dyskretne                  | 365 |
| Ile będzie czekać Julie?                                          | 366 |
| Musimy znaleźć rozkład prawdopodobieństwa dla danych ciągłych     | 367 |
| Dla zmiennych ciągłych możemy wyznaczyć funkcję gęstości          | 368 |
| Prawdopodobieństwo = pole powierzchni                             | 369 |
| Aby obliczyć prawdopodobieństwo, znajdź najpierw $f(x)$ ...       | 370 |
| ...a następnie oblicz prawdopodobieństwo, wyznaczając pole        | 371 |
| Znaleźliśmy szukane prawdopodobieństwo                            | 375 |
| Szukanie bratniej duszy                                           | 376 |
| Modelujemy wzrost mężczyzn                                        | 377 |
| Rozkład normalny stanowi „idealny” model opisu danych ciągłych    | 378 |
| Jak znajdować prawdopodobieństwa dla rozkładu normalnego?         | 379 |
| Liczmy prawdopodobieństwo w trzech krokach                        | 380 |
| Krok 1: Wyznacz parametry definiujące rozkład                     | 381 |
| Krok 2: Dokonaj standaryzacji, by otrzymać $N(0, 1)$              | 382 |
| Aby dokonać standaryzacji, najpierw przesuwamy środek rozkładu... | 383 |
| ...a następnie zmieniamy jego szerokość                           | 383 |
| Obliczamy Z, dla której będziemy odczytywać prawdopodobieństwo    | 384 |
| Krok 3: Odczytaj prawdopodobieństwo z tabeli                      | 387 |

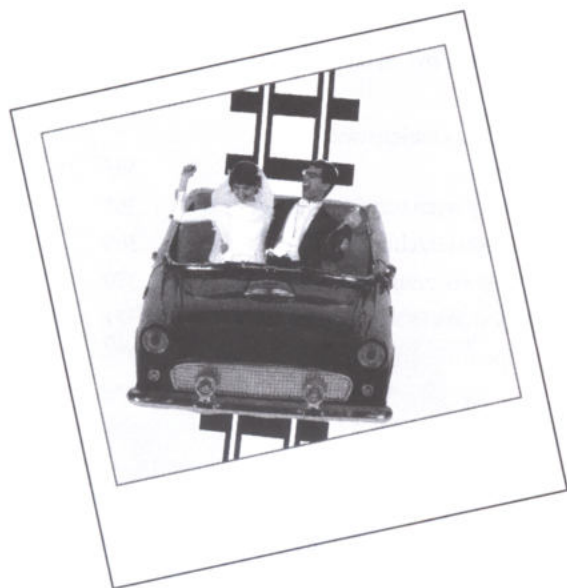




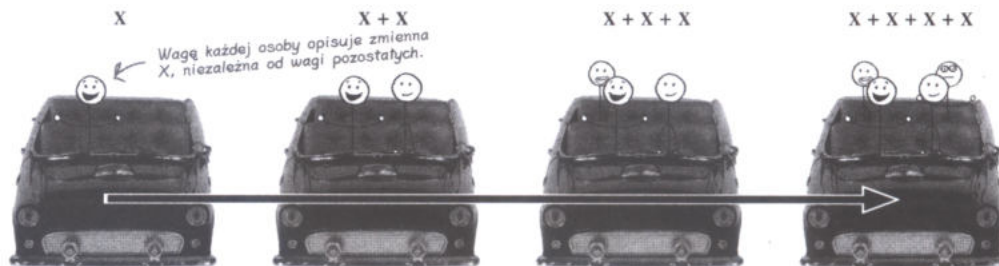
## Poznajemy rozkład normalny (cd.)

## 9 Więcej niż normalność

**Gdyby tak istniał jedynie rozkład normalny...** Życie byłoby *o wiele prostsze*, gdyby wszystko dało się opisać rozkładem normalnym. Nie trzeba by poświęcać mnóstwa czasu na wyznaczanie różnych prawdopodobieństw w tak wielu rozkładach — można by przeznaczyć go na dużo przyjemniejsze rzeczy, na przykład na rozrywkę. Na szczęście są sposoby na to, by **najbardziej złożone problemy** rozwiązywać równie łatwo, jak w przypadku rozkładu normalnego. W tym rozdziale dowiesz się, kiedy jest możliwe **zastąpienie innego rozkładu** przez rozkład normalny i jak się to robi w praktyce.



|                                                                          |     |
|--------------------------------------------------------------------------|-----|
| Wszyscy na pokład Kolejki Miłości!                                       | 401 |
| Sumujemy zmienne o rozkładzie normalnym                                  | 402 |
| Nadal jest to jednak waga                                                | 403 |
| Jaki jest więc rozkład wagi młodej pary?                                 | 405 |
| Znajdujemy prawdopodobieństwo                                            | 408 |
| Więcej ludzi chce skorzystać z Kolejki Miłości                           | 413 |
| Przekształcenia liniowe odnoszą się do zmian wartości...                 | 414 |
| ...zmienne niezależne mówią o tym,<br>ile różnych wartości posiadasz     | 415 |
| Wartość oczekiwana i wariancja<br>dla niezależnych zmiennych losowych    | 416 |
| Wejść do gry czy zrezygnować?                                            | 421 |
| Rozkład normalny przychodzi nam z pomocą                                 | 424 |
| Kiedy stosować przybliżenie rozkładu<br>dwumianowego rozkładem normalnym | 427 |
| Ponowny rzut oka na problem aproksymacji                                 | 432 |
| Rozkład dwumianowy jest rozkładem<br>dyskretnym, a normalny — ciągłym    | 433 |
| Bierzemy poprawkę na ciągłość                                            | 434 |
| Cała prawda o rozkładzie normalnym                                       | 442 |
| Wszyscy na pokład!                                                       | 443 |
| Kiedy można aproksymować rozkład<br>Poissona rozkładem normalnym         | 445 |
| Olbrzymi sukces!                                                         | 451 |



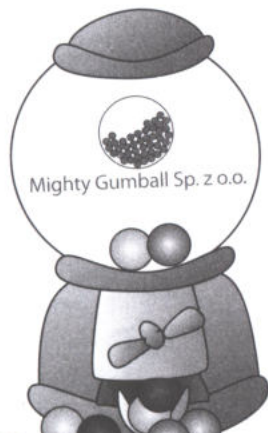
## Przeprowadzamy losowanie

## 10

## Pobieramy próbkę

**Statystyka zajmuje się analizą danych, ale skąd właściwie bierze dane?**

Czasami zebranie danych potrzebnych do analiz nie jest wcale trudne. Tak było wtedy, gdy potrzebowaliśmy informacji o przeciętnym wieku klientów klubu odnowy czy też danych o wielkości sprzedaży gier komputerowych. Ale co w sytuacji, gdy potrzebne nam dane nie są ogólnie dostępne i trzeba je w jakiś sposób zdobyć? Czasami ilość różnych informacji, jakie są nam potrzebne, jest na tyle duża, że nie wiemy nawet, jak się do ich gromadzenia zabrać. Od czego zacząć? W tym rozdziale dowiesz się, jak **gromadzić rzeczywiste dane w sposób efektywny**, odpowiedni do potrzeb i możliwie jak najniższym kosztem. Witamy w świecie losowań!



|                                                        |     |
|--------------------------------------------------------|-----|
| Wielki test produktów Mighty Gumball                   | 454 |
| Firma traci z powodu zużywanych gum                    | 455 |
| Przeprowadzamy testy na próbce, nie na całej populacji | 456 |
| Jak przebiega dobór próby                              | 457 |
| Kiedy próba nie jest reprezentatywna                   | 458 |
| Jak dobrać próbę                                       | 460 |
| Definiujemy operat losowania                           | 461 |
| Czasami dostajemy próby obciążone                      | 462 |
| Źródła obciążenia próby                                | 463 |
| Jak właściwie dobrać próbę                             | 468 |
| Losowanie próby prostej                                | 468 |
| Jak uzyskać próbę prostą                               | 469 |
| Istnieją także inne schematy losowania                 | 470 |
| Możemy przeprowadzić losowanie warstwowe...            | 470 |
| ...losowanie zespołowe...                              | 471 |
| ...a nawet losowanie systematyczne                     | 471 |
| Mighty Gumball dostał swoją próbę                      | 477 |

## Szacujemy parametry populacji na podstawie próby

## 11

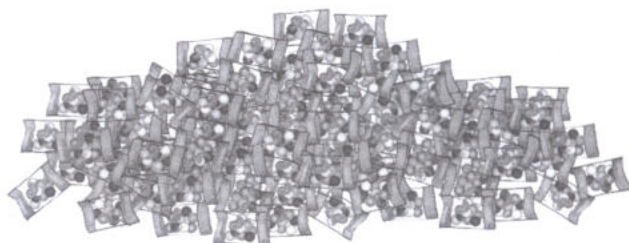
## Dokonujemy ocen

**Czy nie byłoby wspaniale, gdybyśmy potrafili scharakteryzować populację generalną na podstawie informacji zawartych w pojedynczej próbie?**

Zanim będziesz mógł świętować osiągnięcie **biegłości w prowadzeniu badań** statystycznych, musisz się nauczyć robić właściwy użytek z próbki pobieranej z populacji. Musisz więc posiadać umiejętność **przewidywania charakterystyk populacji** na podstawie informacji zawartych w próbie oraz nauczyć się, jak możesz ocenić **wiarygodność** swoich szacunków. W tym rozdziale pokażemy Ci, w jaki sposób posiadana próbka może być użyta jako użyteczne źródło informacji o badanej populacji i vice versa.

|                                                                                         |     |
|-----------------------------------------------------------------------------------------|-----|
| Jaka więc jest rzeczywista trwałość smaku tamtych gum?                                  | 480 |
| Zaczniemy od oszacowania średniej w populacji                                           | 481 |
| Estymatory punktowe pozwalają oszacować parametry populacji                             | 482 |
| Szacujemy wariancję populacji                                                           | 486 |
| Znajdujemy inny estymator niż wariancja z próby                                         | 487 |
| Która formuła co oznacza?                                                               | 489 |
| Wszystko jest kwestią proporcji                                                         | 492 |
| Jaki ma to związek z estymacją parametrów?                                              | 497 |
| Rozkład z próby estymatora $p$                                                          | 498 |
| Ile wynosi wartość oczekiwana $P_s$ ?                                                   | 500 |
| A ile wynosi wariancja $P_s$ ?                                                          | 501 |
| Ustalamy rozkład $P_s$                                                                  | 502 |
| $P_s$ ma rozkład normalny                                                               | 503 |
| Musimy znaleźć rozkład dla średniej z próby                                             | 509 |
| Rozkład z próby estymatora średniej                                                     | 510 |
| Znajdujemy wartość oczekiwaną $\bar{X}$                                                 | 512 |
| A co z wariancją zmiennej $\bar{X}$ ?                                                   | 514 |
| Jaki jest więc kształt rozkładu zmiennej $\bar{X}$ ?                                    | 518 |
| Jeśli $n$ jest odpowiednio duże, rozkład $\bar{X}$ jest zbliżony do rozkładu normalnego | 519 |
| Stosujemy centralne twierdzenie graniczne                                               | 520 |

Fantastycznie!  
Mamy komplet bardzo  
korzystnych dla nas  
statystyk, które  
możemy wykorzystać  
w naszych  
reklamach.





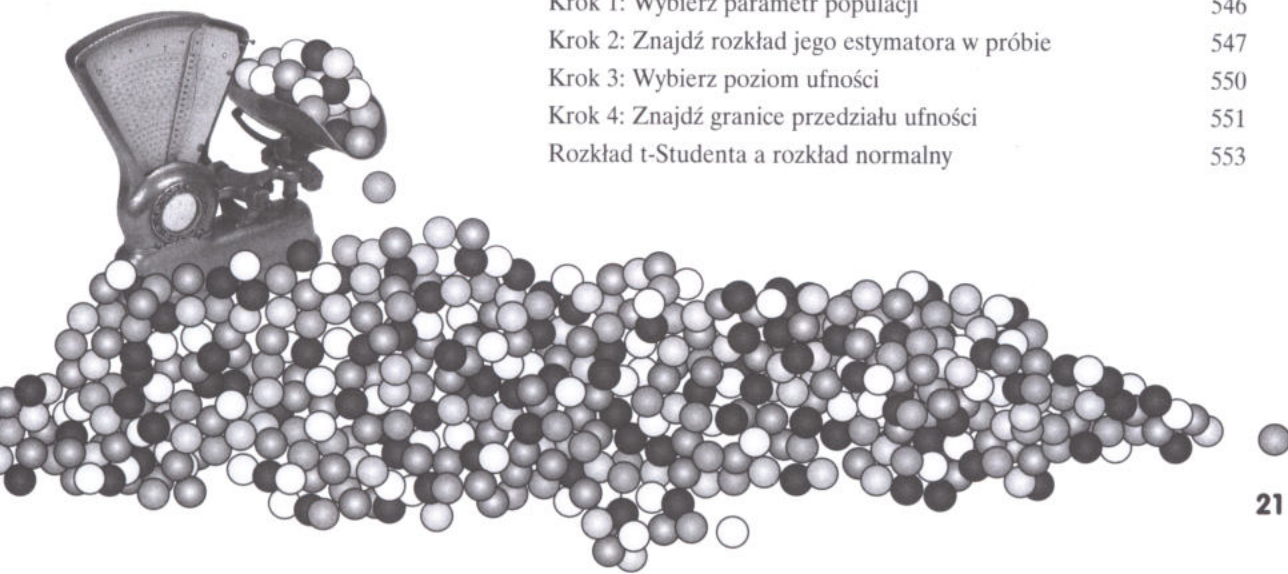
## Konstruujemy przedziały ufności

## 12

## Wyrażamy przekonania

**Czasami estymacja punktowa daje nie do końca trafione wyniki.** Wiesz już, jak za pomocą estymatorów punktowych uzyskać **dokładne oceny** parametrów populacji, takich jak wartość oczekiwana, wariancja czy wskaźnik struktury. Nie zawsze jednak ocena w postaci pojedynczej liczby zaspokoi w pełni Twoje oczekiwania. Bo jak ocenić, na ile jest ona dokładna? Bądź co bądź całe wnioskowanie o populacji generalnej opiera się na stosunkowo nielicznej próbie, która przecież nie zawsze musi w pełni odzwierciedlać charakterystyki populacji. W tym rozdziale poznasz **inną metodę szacowania nieznanymi wartościami parametrów populacji**, która uwzględnia pewien stopień **niepewności ocen** i — co więcej — pozwala ją zmierzyć. Czytaj dalej, a poznasz wszystkie tajemnice **przedziałów ufności**.

|                                                       |     |
|-------------------------------------------------------|-----|
| Mighty Gumball znów ma kłopot                         | 526 |
| Problemem pozostaje precyzja                          | 527 |
| Poznajemy przedziały ufności                          | 528 |
| Wyznaczamy przedział ufności w czterech krokach       | 529 |
| Krok 1: Wybierz parametr populacji                    | 530 |
| Krok 2: Znajdź rozkład jego estymatora w próbie       | 530 |
| Krok 3: Wybierz poziom ufności                        | 532 |
| Krok 4: Znajdź granice przedziału ufności             | 534 |
| Zacniemy od wyznaczenia $Z$                           | 535 |
| Zapisujemy prawdopodobieństwo z użyciem $\bar{X}$     | 536 |
| Znajdujemy ostatecznie wartość zmiennej $\bar{X}$     | 539 |
| Znaleźliśmy poszukiwany przedział ufności             | 540 |
| Podsumujmy wykonane kroki                             | 541 |
| Użyteczne skróty przy wyznaczaniu przedziałów ufności | 542 |
| Krok 1: Wybierz parametr populacji                    | 546 |
| Krok 2: Znajdź rozkład jego estymatora w próbie       | 547 |
| Krok 3: Wybierz poziom ufności                        | 550 |
| Krok 4: Znajdź granice przedziału ufności             | 551 |
| Rozkład t-Studenta a rozkład normalny                 | 553 |





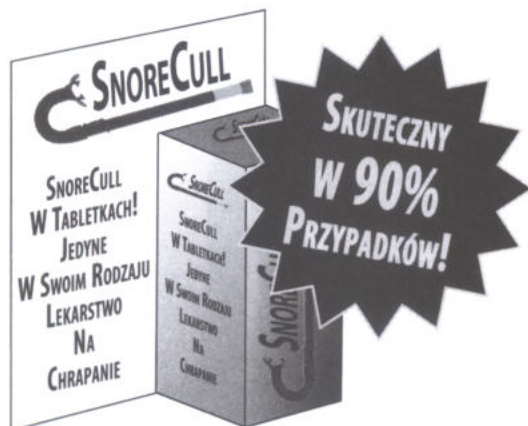
## Weryfikacja hipotez

## 13

## Oceniamy fakty

**Nie wszystko, co do Ciebie dociera, musi być prawdą.** Najgorsze jest jednak to, że trudno jest ocenić, kiedy ma się do czynienia z prawdą, a kiedy nie. **Weryfikacja hipotez**, drugi obok estymacji dział wnioskowania statystycznego, daje Ci narzędzie do oceny prawdziwości twierdzeń statystycznych. Za jej pomocą będziesz mógł ocenić, na ile takie, a nie inne charakterystyki próby mogą być efektem działalności określonych sił, stanowiących o kształcie całej populacji, a na ile są jedynie dziełem **czystego przypadku**. Z lektury tego rozdziału dowiesz się, w jaki sposób możesz potwierdzić lub obalić swoje przypuszczenia odnoszące się do otaczającej Cię rzeczywistości.

|                                                                       |     |
|-----------------------------------------------------------------------|-----|
| Cudowny lek na chrapanie                                              | 560 |
| Ogólne spojrzenie na problem                                          | 564 |
| Weryfikacja hipotez w sześciu krokach                                 | 565 |
| Krok 1: Sformułuj hipotezę, którą chcesz zweryfikować                 | 566 |
| Krok 2: Wybierz statystykę testową (sprawdzian testu)                 | 569 |
| Krok 3: Określ obszar odrzuceń testowanej hipotezy                    | 570 |
| Krok 4: Znajdź prawdopodobieństwo $p$ ( $p$ -wartość)                 | 573 |
| Krok 5: Sprawdź, czy sprawdzian testu<br>wpada do obszaru odrzuceń    | 575 |
| Krok 6: Podejmij decyzję                                              | 575 |
| Co by się stało, gdyby próba była większa?                            | 578 |
| Przeprowadzamy kolejny test                                           | 581 |
| Krok 1: Sformułuj hipotezę, którą chcesz zweryfikować                 | 581 |
| Krok 2: Wybierz statystykę testową (sprawdzian testu)                 | 582 |
| Przybliżamy rozkład statystyki testowej rozkładem normalnym           | 585 |
| Krok 3: Określ obszar odrzuceń testowanej hipotezy                    | 586 |
| Zaczniemy od błędu I rodzaju                                          | 594 |
| A co z błędem II rodzaju?                                             | 595 |
| Znajdujemy prawdopodobieństwa $\alpha$ i $\beta$ w naszym przykładzie | 596 |
| Znajdujemy zbiór wartości spoza obszaru krytycznego                   | 597 |
| Znajdujemy $P(\text{błąd II rodzaju})$                                | 598 |
| Moc przybywa                                                          | 599 |



Rozkład  $\chi^2$ 

## 14

## Gdy sprawy idą nie po naszej myśli

**Czasami sprawy toczą się zupełnie inaczej, niż się tego spodziewaliśmy.**

Kiedy decydujesz się na opis pewnego zjawiska za pomocą konkretnego rozkładu prawdopodobieństwa, zwykle masz jakieś wyobrażenia na temat tego, jak się ono rozwinie w dłuższym okresie. Czasem jednak te **wyobrażenia całkowicie rozmiijają się z rzeczywistością**. Co wtedy począć? Skąd masz wiedzieć, czy dostrzeżone różnice są jedynie dziełem przypadku, czy też może pierwszą oznaką błędnych założeń leżących u podstaw przyjętego przez Ciebie modelu? W tym rozdziale pokażemy Ci, jak możesz posłużyć się rozkładem  $\chi^2$  do **oceny rezultatów**, by móc wskazać wśród nich te najbardziej **podejrzane**.

|                                                                                |     |
|--------------------------------------------------------------------------------|-----|
| Przed kasynem Dana rysują się kłopoty                                          | 606 |
| Przyglądamy się automatom do gry                                               | 607 |
| Rozkład $\chi^2$ dobrze modeluje różnice                                       | 609 |
| O czym więc mówi ta statystyka?                                                | 610 |
| Główne zastosowania rozkładu $\chi^2$                                          | 611 |
| $\chi^2$ reprezentuje liczbę stopni swobody                                    | 612 |
| Czym jest istotność statystyczna?                                              | 613 |
| Testowanie hipotez z rozkładem $\chi^2$                                        | 614 |
| Rozwiązałeś tajemnicę wysokich wygranych w grach na automatach                 | 617 |
| Dan ma jeszcze jeden problem                                                   | 623 |
| Rozkład $\chi^2$ sprawdza się również w testach niezależności                  | 624 |
| Częstości teoretyczne możemy wyznaczyć w oparciu o rachunek prawdopodobieństwa | 625 |
| Ile więc wynoszą częstości teoretyczne?                                        | 626 |
| Musimy jeszcze poznać liczbę stopni swobody                                    | 629 |
| Ogólna metoda wyznaczania liczby stopni swobody                                | 634 |
| A zatem formuła ma postać...                                                   | 635 |
| Uratowałeś kasyno Dana od bankructwa                                           | 637 |



## Korelacja i regresja

## 15

## Co z moją linią?

**Czy zastanawiałeś się kiedyś, w jakim stopniu dwie rzeczy są ze sobą**

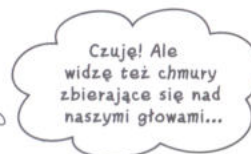
**powiążane?** W poprzednich rozdziałach przyglądaliśmy się użyciu statystyk, które opisywały zbiór danych z punktu widzenia wyłącznie jednej cechy — mówiliśmy na przykład o wzroście mężczyzn, punktach zdobytych przez zawodnika koszykówki czy też o trwałości smaku gum do żucia. Tymczasem istnieją statystyki, które pozwalają ocenić siłę **związku między większą liczbą zmiennych**. Ich znajomość dostarczy Ci znacznie bogatszych informacji na temat otaczającego Cię świata, które będziesz mógł wykorzystać we własnym interesie. W tym rozdziale pokażemy Ci, jak **wykrywać związki** między zmiennymi, korzystając z miar korelacji i regresji.



|                                                             |     |
|-------------------------------------------------------------|-----|
| Przyjrzyjmy się danym na temat frekwencji i nasłonecznienia | 645 |
| Rzut oka na wymiary                                         | 646 |
| Wykreślamy dane dwuwymiarowe                                | 647 |
| Wykresy rozrzutu pokazują trendy obecne w danych            | 650 |
| Korelacja a przyczynowość                                   | 652 |
| Wykorzystujemy do prognozowania                             |     |
| linię o najlepszym dopasowaniu                              | 656 |
| Najlepsze dopasowanie jest nadal tylko dopasowaniem         | 657 |
| Będziemy minimalizować odchylenia od wartości rzeczywistych | 658 |
| Wyznaczamy sumę kwadratów odchyłeń                          | 659 |
| Znajdujemy wartości nieznanymi parametrów                   | 660 |
| Obliczamy nachylenie linii najlepszego dopasowania          | 661 |
| Obliczamy nachylenie linii najlepszego dopasowania (cd.)    | 662 |
| Znaleźliśmy b, ale co z a?                                  | 663 |
| Znaleźliśmy związek między dwiema zmiennymi                 | 667 |
| Różne typy korelacji                                        | 668 |
| Współczynnik korelacji mierzy siłę związku między zmiennymi | 669 |
| Poznajemy wzór na wartość współczynnika r                   | 670 |
| Obliczamy wartość r dla naszego zbioru danych               | 671 |
| Obliczamy wartość r dla naszego zbioru danych (cd.)         | 672 |



Poczuj ten rytm,  
dziecino...



Czuję! Ale  
widzę też chmury  
zbierające się nad  
naszymi głowami...





## Dodatek uzupełniający

## A

**Dziesięć najważniejszych rzeczy  
(które pomineliśmy)**

**Choć powiedzieliśmy już wiele, coś musimy dopowiedzieć.** Jest jeszcze kilka rzeczy, o których, naszym zdaniem, powinienś wiedzieć. Byłoby trochę nie w porządku wobec Ciebie, gdybyśmy je całkowicie przemilczeli. Nie ma przy tym znaczenia, że są to naprawdę sprawy, które wymagają tylko krótkiej wzmianki.



|                                                                    |     |
|--------------------------------------------------------------------|-----|
| 1. Inne techniki wizualizacji danych                               | 682 |
| 2. Anatomia rozkładu prawdopodobieństwa                            | 683 |
| 3. Eksperyment statystyczny                                        | 684 |
| 4. Metoda najmniejszych kwadratów w notacji alternatywnej          | 686 |
| 5. Współczynnik determinacji                                       | 687 |
| 6. Zależności nieliniowe                                           | 688 |
| 7. Przedział ufności dla współczynnika nachylenia prostej regresji | 689 |
| 8. Rozkłady z próby — różnica między dwiema średnimi               | 690 |
| 9. Rozkłady z próby — różnica między wskaźnikami struktury         | 691 |
| 10. $E(X)$ i $Var(X)$ dla zmiennych ciągłych                       | 692 |

## Tablice statystyczne

## B

**Czasem trzeba coś sprawdzić**

**Co byśmy poczęli bez pocziwych tablic statystycznych?** Nie wystarczy znać zastosowanie poszczególnych rozkładów. Bardzo często trzeba coś policzyć. Dobrze jest mieć wtedy pod ręką **tablice statystyczne**, zawierające **standardowe prawdopodobieństwa** dla typowych rozkładów. Dlatego ten załącznik prezentuje tablice dla rozkładów: **normalnego**, **t-Studenta** i  $\chi^2$ .

|                                            |     |
|--------------------------------------------|-----|
| Standaryzowany rozkład normalny            | 696 |
| Wartości krytyczne dla rozkładu t-Studenta | 698 |
| Wartości krytyczne dla rozkładu $\chi^2$   | 699 |

## S

**Skorowidz**

701