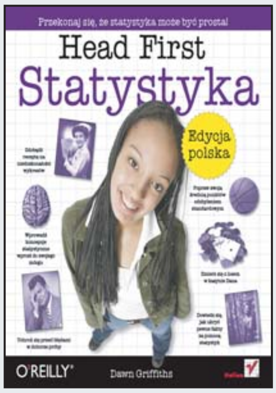
Head First. Statystyka. Edycja polska



Autor: Dawn Griffiths
Tłumaczenie: Przemysław Janicki
ISBN: 978-83-246-2065-4
Tytuł oryginału: Head First Statistics
Format: 200×230, stron: 452

Przekonaj się, że statystyka może być prosta!

Jeśli statystyka wydaje Ci się nudna i skomplikowana, to tylko dlatego, że nie korzystałeś dotąd z odpowiedniego podręcznika! Ta innowacyjna, przyjazna dla czytelnika książka z pewnością zmieni Twoją opinię. Wiedzę z zakresu statystyki podano tu w sposób uwzględniający najnowsze badania w zakresie efektywnego nauczania. Dzięki przyjęciu takiej formuły tekst dostosowano do sposobu funkcjonowania Twojego mózgu, aby w pełni wykorzystać jego możliwości i bezboleśnie wprowadzić Cię w świat skomplikowanych obliczeń.

Najważniejsze zagadnienia zostały tu zilustrowane za pomocą – nierzadko zabawnych – przykładów z życia codziennego, takich jak analiza statystyk sportowych, wyników gier hazardowych czy testów nowych leków. Dzięki tej książce dowiesz się m.in., jak wybrać optymalny wykres do wizualizacji określonych danych, szybko wskazać wartości reprezentatywne dla danego zbioru danych i za pomocą rachunku prawdopodobieństwa przewidywać skutki powtarzalnych zdarzeń w długich seriach. Z łatwością nie tylko przyswoisz zawartą tu wiedzę, ale i wykorzystasz ją w codziennym życiu!

- Wizualizacja danych
- Wykresy
- Praca z danymi zgrupowanymi
- Rozkład geometryczny, dwumianowy i Poissona
- Miary zróżnicowania
- Szacowanie parametrów populacji na podstawie próby
- Kwartyle i wariancje
- Prawdopodobieństwo zdarzeń
- Konstruowanie przedziału ufności
- Podstawy kombinatoryki
- Weryfikacja hipotez
- Korelacja i regresja

Z dobrym podręcznikiem nawet statystyka może być łatwa i ciekawa!

Spis treści (skrócony)

Spis treści (z prawdziwego zdarzenia)

W

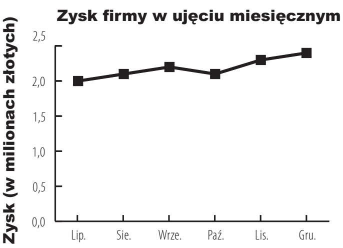
Wprowadzenie

Twój mózg a statystyka. Czytasz tę książkę, ponieważ chcesz się czegoś nauczyć. W tym czasie Twój mózg będzie Ci wyświadczał przysługę, dbając o to, byś się za bardzo nie przemęczał. Będzie Ci podpowiadał: „Zajmij się lepiej ważniejszymi sprawami, na przykład tym, jakich dzikich zwierząt trzeba się wystrzegać albo czy jeżdżenie na snowboardzie nago to aby na pewno dobry pomysł". Jak w tej sytuacji mógłbyś przekonać swój mózg, że Twoje życie zależy od znajomości statystyki?

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.



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


20 lat


21 lat



20 lat


19 lat

Miary zróżnicowania

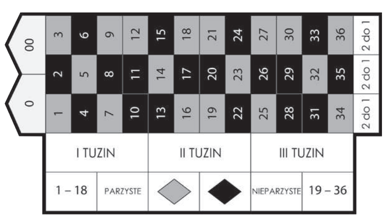
3 Potęga zmienności

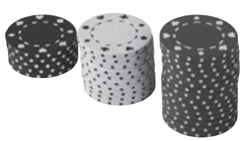
Nie wszystkiemu można wierzyć, ale jak się o tym przekonać?
Średnie pozwalają nam poznać typową wartość dla naszych danych, ale **nie mówią nam wszystkiego**. Umiemy już znajdować wartości centralne zbioru danych, ale średnia arytmetyczna, mediana czy dominanta nie zawsze wystarczają do wyciągnięcia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

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!

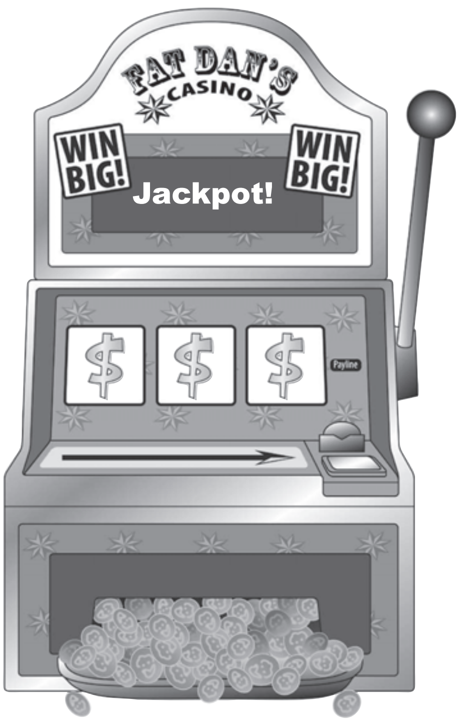




Dyskretne rozkłady prawdopodobieństwa

5 Zarządzamy oczekiwaniami

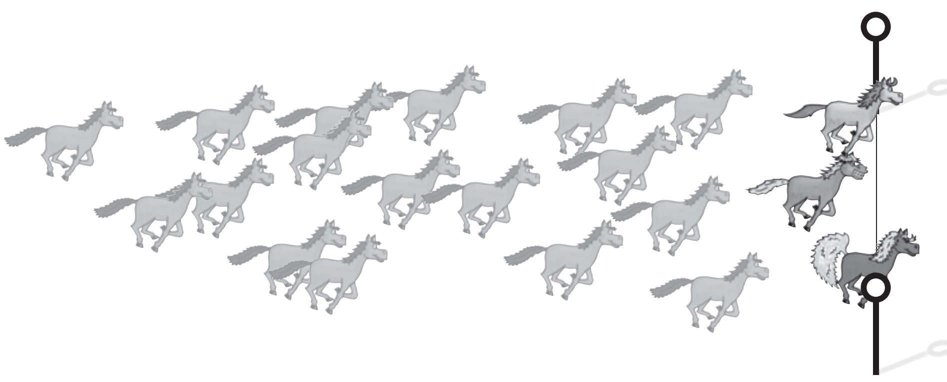
Zdarzenia mało prawdopodobne czasem się zdarzają, ale jakie są ich konsekwencje? W poprzednim rozdziale przekonaliśmy się, jak rachunek prawdopodobieństwa może pomóc nam ocenić szanse zajścia pewnych zdarzeń. Jednak nie powie nam on nic na temat **wpływu** tych zdarzeń na nasze życie. Choć przy stole 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.



Podstawy kombinatoryki

6 Porządkujemy obiekty

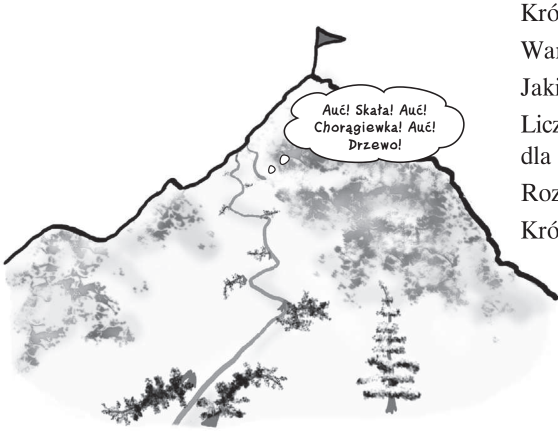
Czasami kolejność ma znaczenie. Policzenie **wszystkich możliwych sposobów** grupowania czy porządkowania pewnego zbioru obiektów może być niezwykle pracochłonne. Często jednak nie mamy wyjścia, bo takie informacje są **kluczowe** dla rachunku prawdopodobieństwa. W tym rozdziale poznamy **szybki i efektywny** 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**.



7 Poznajemy rozkłady: geometryczny, dwumianowy i Poissona

Stajemy się dyskretni

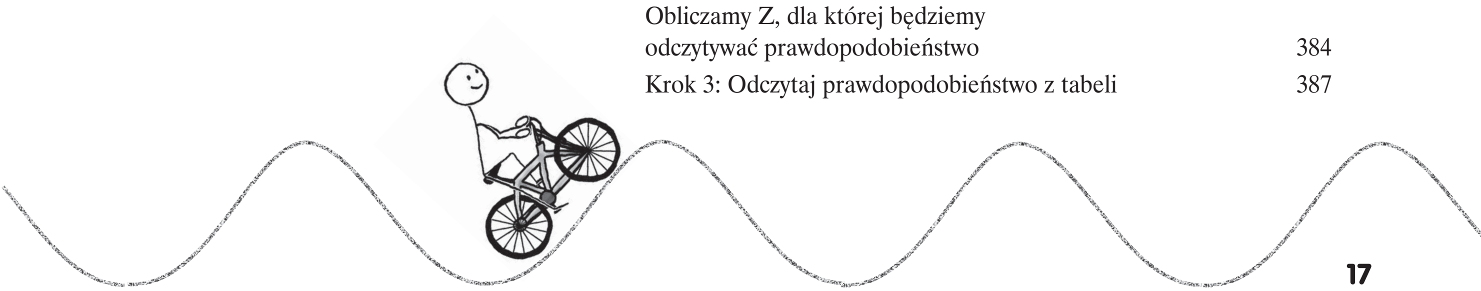
Wyznaczanie rozkładów prawdopodobieństwa zabiera sporo czasu. Wiemy już, jak wyznaczać rozkłady prawdopodobieństwa, choć nie zawsze jest to łatwe. Pewnie więc zgodzisz się z nami, że byłoby dobrze, gdyby istniały ogólne, bardziej uniwersalne rozkłady prawdopodobieństwa, którymi **łatwo** można by się 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.



Poznajemy rozkład normalny

8 Wybieramy normalność

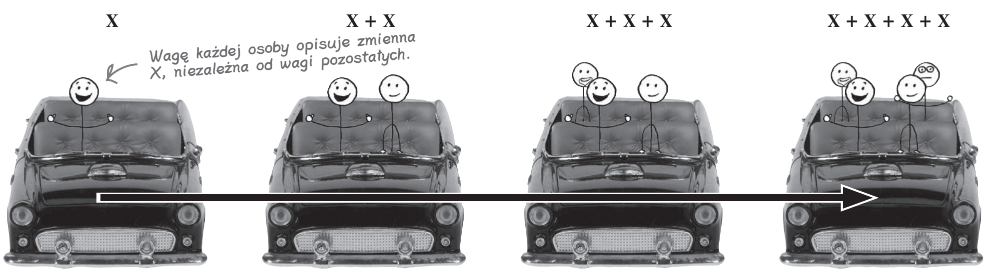
Dyskretne rozkłady prawdopodobieństwa nie w każdej sytuacji się sprawdzają. W poprzednim rozdziale poznaliśmy trzy rozkłady prawdopodobieństwa, dla których byliśmy w stanie wymienić wszystkie wartości, jakie może przyjąć zmienna losowa. Nie zawsze jednak jest to możliwe. Niekiedy posiadane przez nas 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**.



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.



Przeprowadzamy losowanie

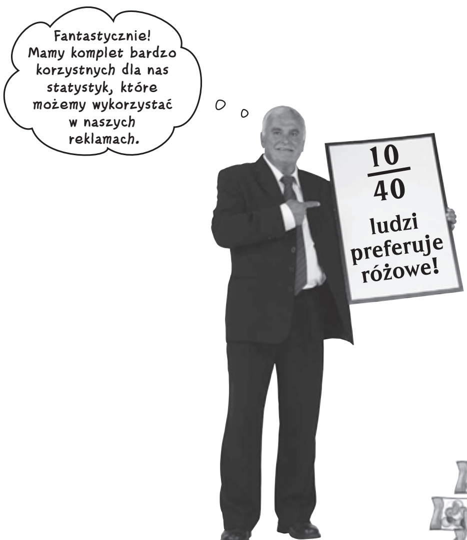
10 Pobieramy próbkę

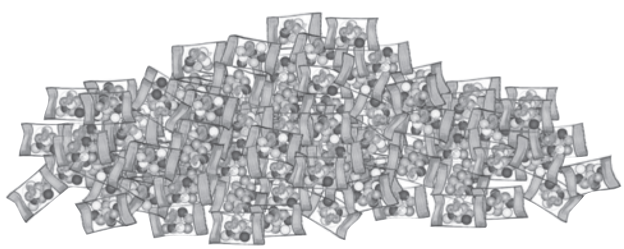
Statystyka zajmuje się analizą danych, ale skąd właściwie bierze dane? Czasami zebranie danych potrzebnych do analiz nie jest wcale trudne. Tak było wtedy, gdy potrzebowaliśmy informacji o przeciętnym wieku klientów klubu odnowy czy też danych o wielkości sprzedaży gier komputerowych. Ale co w sytuacji, 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!

11 Szacujemy parametry populacji na podstawie próby

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 posiąść 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.

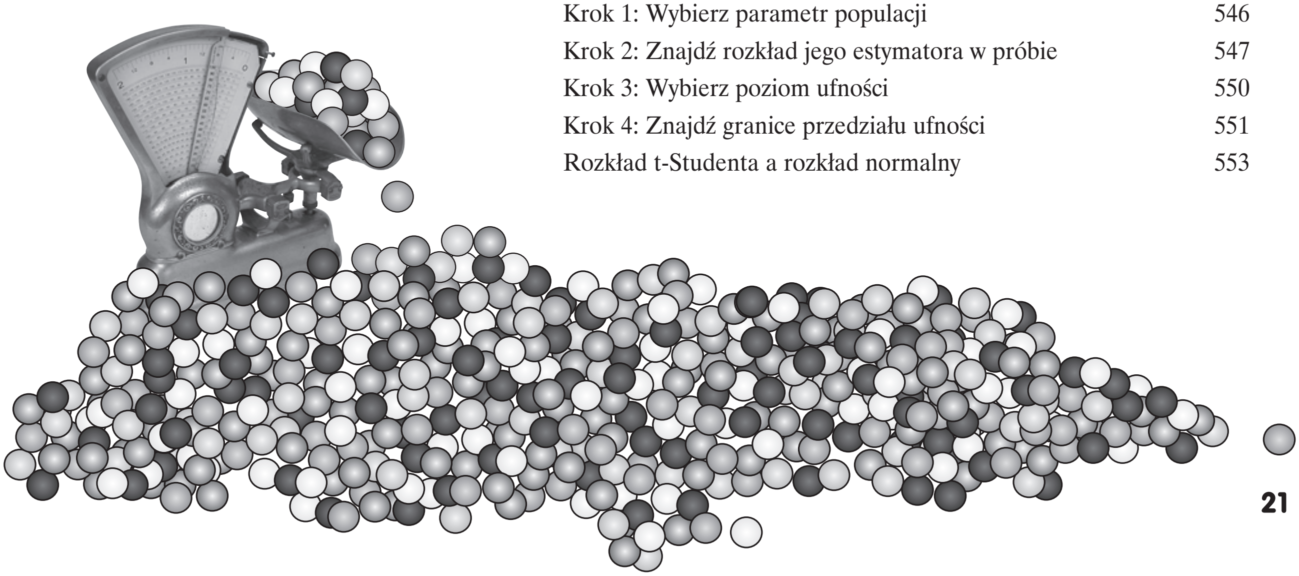




Konstruujemy przedziały ufności

12 Wyrażamy przekonania

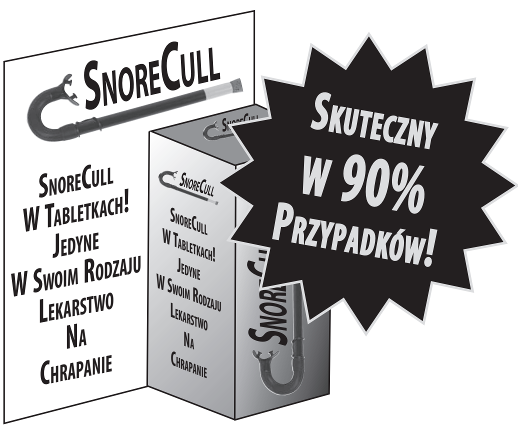
Czasami estymacja punktowa daje nie do końca trafione wyniki. Wiesz już, jak za pomocą estymatorów punktowych uzyskać **dokładne oceny** parametrów populacji, takich jak wartość oczekiwana, wariancja czy wskaźnik struktury. Nie zawsze jednak ocena w postaci pojedynczej liczby zaspokoi w pełni Twoje oczekiwania. 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 nieznanych wartości parametrów populacji**, która uwzględnia pewien stopień **niepewności ocen** i — co więcej — pozwala ją zmierzyć. Czytaj dalej, a poznasz wszystkie tajemnice **przedziałów ufności**.



13 Weryfikacja hipotez

Oceniamy fakty

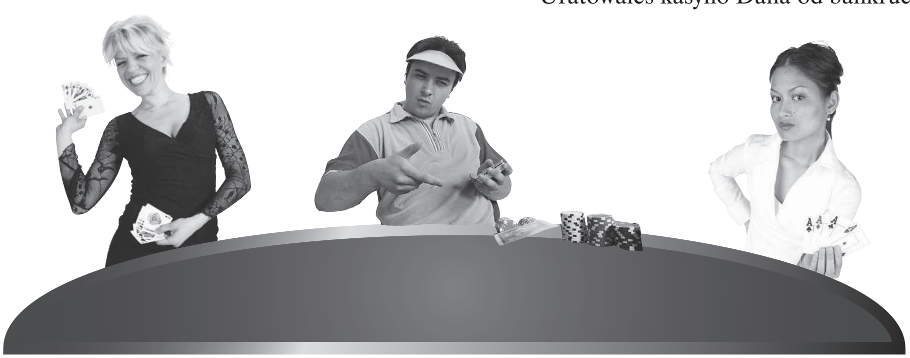
Nie wszystko, co do Ciebie dociera, musi być prawdą. Najgorsze jest jednak to, że trudno jest ocenić, kiedy ma się do czynienia z prawdą, a kiedy nie. ***Weryfikacja hipotez***, drugi obok estymacji dział wnioskowania statystycznego, daje Ci narzędzie do oceny prawdziwości twierdzeń statystycznych. Za 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.



14 Rozkład χ^2

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 rozmijają 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 χ^2 do **oceny rezultatów**, by móc wskazać wśród nich te najbardziej **podejrzane**.



Korelacja i regresja

15 Co z moją linią?

Czy zastanawiałeś się kiedyś, w jakim stopniu dwie rzeczy są ze sobą powiązane? 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.

Dodatek uzupełniający

A

Dziesięć najważniejszych rzeczy (które pominęliśmy)

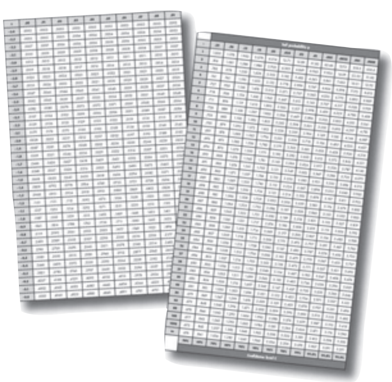
Choć powiedzieliśmy już wiele, coś musimy dopowiedzieć. Jest jeszcze kilka rzeczy, o których, naszym zdaniem, powinieneś wiedzieć. Byłoby trochę nie w porządku wobec Ciebie, gdybyśmy je całkowicie przemilczeli. Nie ma przy tym znaczenia, że są to naprawdę **sprawy, które wymagają tylko krótkiej wzmianki**.

Tablice statystyczne

B

Czasem trzeba coś sprawdzić

Co byśmy poczęli bez poczciwych 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 χ^2**.



S

5. Dyskretne rozkłady prawdopodobieństwa

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

Czy i Ty poddajesz się hipnotyzującemu wpływowi migających światełek automatu do gry zwanego „jednorękim bandytą”? A zatem jesteś prawdziwym szczęściarzem. W Fat Dan’s Casino znajduje się cały rząd automatów do gry, które tylko czekają na to, by ktoś na nich zagrał. Spróbujmy więc i my (pociągnięcie dźwigni kosztuje 1 dolara). Kto wie, może rozbijemy bank!

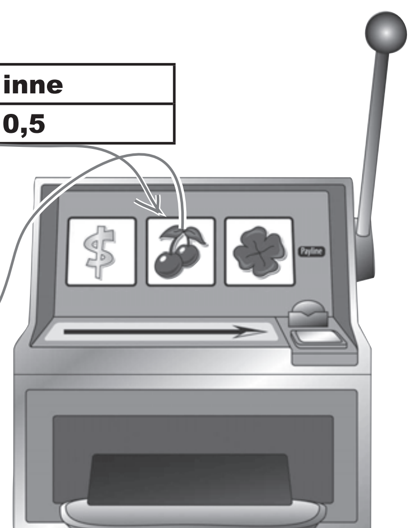
Jednoręki bandyta posiada trzy okienka, w których pojawiają się różne symbole. Gdy pojawią się one w odpowiedniej kombinacji, z umieszczonego poniżej otworu posypią się monety.

Kwota, jaką można wygrać, wygląda zachęcająco, ale najpierw chciałbym się dowiedzieć, jakie mam szanse na wygraną.

Cóż, wydaje się, że szanse te można by łatwo policzyć. Poniżej zamieszczono prawdopodobieństwo pojawienia się danego symbolu w okienku:

$	wisienka	cytrynka	inne
0,1	0,2	0,2	0,5

Wszystkie trzy okienka są od siebie niezależne, zatem pojawienie się danego symbolu w jednym z nich nie ma wpływu na to, co pojawi się w pozostałych okienkach.



Prawdopodobieństwo pojawienia się wisienki w tym okienku wynosi 0,2.

Wczuj się w rolę gracza

Spójrz na planszę z wygranymi zamieszczoną na poprzedniej stronie. Wyobraź sobie, że jesteś graczem, który chce poznać prawdopodobieństwo pojawienia się każdej z kombinacji gwarantujących wygraną. Uzupełnij tabelkę. Jakie jest prawdopodobieństwo tego, że nie wygrasz nic?

prawdopodobieństwo 

prawdopodobieństwo (kolejność dowolna) 

prawdopodobieństwo 

prawdopodobieństwo 

prawdopodobieństwo braku wygranej

Wczuj się w rolę gracza: Rozwiązanie

Spójrz na planszę z wygranymi zamieszczoną na poprzedniej stronie. Wyobraź sobie, że jesteś graczem, który chce poznać prawdopodobieństwo pojawienia się każdej z kombinacji gwarantujących wygraną. Uzupełnij tabelkę. Jakie jest prawdopodobieństwo tego, że nie wygrasz nic?

prawdopodobieństwo

$P(\$, \$, \$) = P(\$) \times P(\$) \times P(\$) =$

$= 0{,}1 \times 0{,}1 \times 0{,}1 =$

$= 0{,}001$

Prawdopodobieństwo pojawienia się symbolu dolara wynosi 0,1.

prawdopodobieństwo (kolejność dowolna)

Są trzy możliwości uzyskania takiego wyniku:

$P(\$, \$, \text{wisienka}) + P(\$, \text{wisienka}, \$) + P(\text{wisienka}, \$, \$) =$

$= (0{,}1^2 \times 0{,}2) + (0{,}1^2 \times 0{,}2) + (0{,}1^2 \times 0{,}2) =$

$= 0{,}006$

prawdopodobieństwo

$P(\text{cytrynka, cytrynka, cytrynka}) =$

$= P(\text{cytrynka}) \times P(\text{cytrynka}) \times P(\text{cytrynka}) =$

$= 0{,}2 \times 0{,}2 \times 0{,}2 =$

$= 0{,}008$

Pojawienie się cytrynki w jednym z okienek nie wpływa na jej pojawienie się w dwóch pozostałych. Dlatego mnożymy przez siebie te prawdopodobieństwa.

prawdopodobieństwo

$P(\text{wisienka, wisienka, wisienka}) =$

$= P(\text{wisienka}) \times P(\text{wisienka}) \times P(\text{wisienka}) =$

$= 0{,}2 \times 0{,}2 \times 0{,}2 =$

$= 0{,}008$

prawdopodobieństwo braku wygranej

Zamiast wypisywać wszystkie kombinacje symboli prowadzące do przegranej, zauważamy, że: P(przegrana) = 1 – P(wygrana).

To oznacza, że nie uzyskamy żadnej z wygrywających kombinacji.

$P(\text{przegrana}) = 1 - P(\$, \$, \$) - P(\$, \$, \text{wisienka (kolejność dowolna)}) - P(\text{wisienka, wisienka, wisienka}) +$
$- P(\text{cytrynka, cytrynka, cytrynka})$

$= 1 - 0{,}001 - 0{,}006 - 0{,}008 - 0{,}008 =$

$= 0{,}977$

To są prawdopodobieństwa wygranej obliczone wyżej.

Tworzymy rozkład prawdopodobieństwa wygranej na automacie

Oto prawdopodobieństwa pojawienia się wszystkich kombinacji symboli gwarantujących wygraną:

To jest zestawienie prawdopodobieństw, które przed chwilą policzyliśmy.

Kombinacja	brak wygranej	cytrynki	wisienki	dolary/ wisienki	dolary
Prawdopodobieństwo	0,977	0,008	0,008	0,006	0,001

Pewnie takie zestawienie może się do czegoś przydać. Czy jednak nie moglibyśmy pójść o krok dalej? Dobrze, że znamy te prawdopodobieństwa, ale nadal nie wiemy, ile możemy wygrać.

Chcielibyśmy znać nie tylko prawdopodobieństwo wygranej, ale również to, jak *wiele* moglibyśmy wygrać.

Na razie znamy jedynie prawdopodobieństwa pojawienia się na automacie poszczególnych kombinacji symboli, ale nie wiemy nic o wysokości ewentualnej wygranej.

Dlatego nieco zmodyfikujemy naszą tabelkę, przyporządkowując kolejne prawdopodobieństwa do odpowiadającej im wygranej. Wysokość wygranej odczytaną z informacji umieszczonej na automacie pomniejszymy o kwotę 1 dolara, którą musimy zapłacić, by wziąć udział w grze.

Kombinacja	brak wygranej	cytrynki	wisienki	dolary/ dolary	dolary
Wypłata	**-1 $**	**4 $**	**9 $**	**14 $**	**19 $**
Prawdopodobieństwo	0,977	0,008	0,008	0,006	0,001

Tracimy 1 dolara, jeśli nie pojawi się żadna z kombinacji gwarantujących wygraną.

To te same prawdopodobieństwa co powyżej, ale przyporządkowane wygranej, jaka się z nimi wiąże.

Wypłata przy każdej kombinacji: kwota wygranej pomniejszona o 1 dolara opłaty za udział w grze.

W tabeli tej zawarto tak zwany ***rozkład prawdopodobieństwa*** wypłat z gry na automacie, bowiem każdej wygranej pomniejszonej o koszt wzięcia udziału w grze przyporządkowano prawdopodobieństwo, z jakim może ona paść.

Rozkłady prawdopodobieństwa z bliska

Aby stworzyć rozkład prawdopodobieństwa wypłat z gry na automacie, dla każdej możliwej kwoty wypłaty obliczyliśmy prawdopodobieństwo, z jakim może ona paść. Przy opisie tego rodzaju zdarzeń statystycy posługują się tak zwaną ***zmienną losową***, czyli taką zmienną, która może przyjmować różne wartości z pewnego zbioru z określonym prawdopodobieństwem. W naszym przykładzie zmienna losowa reprezentuje wypłatę z gry, a więc wielkość, jaką mamy szansę zarobić, pociągając raz za dźwignię automatu.

Zmienne losowe oznaczamy zwykle wielkimi literami alfabetu łacińskiego, na przykład X czy Y. Poszczególne wielkości, jakie może przyjąć zmienna (tak zwane ***realizacje***), oznaczamy odpowiadającymi im małymi literami, na przykład x czy y. Tak więc zapis P(X = x) odczytujemy: „Prawdopodobieństwo, że zmienna X przyjmie wartość x”.

Oto nasz rozkład prawdopodobieństwa zapisany przy wykorzystaniu tej notacji:

Tutaj x wynosi 19 dolarów.

Wartość wypłaty dla każdej kombinacji symboli jest reprezentowana przez x.

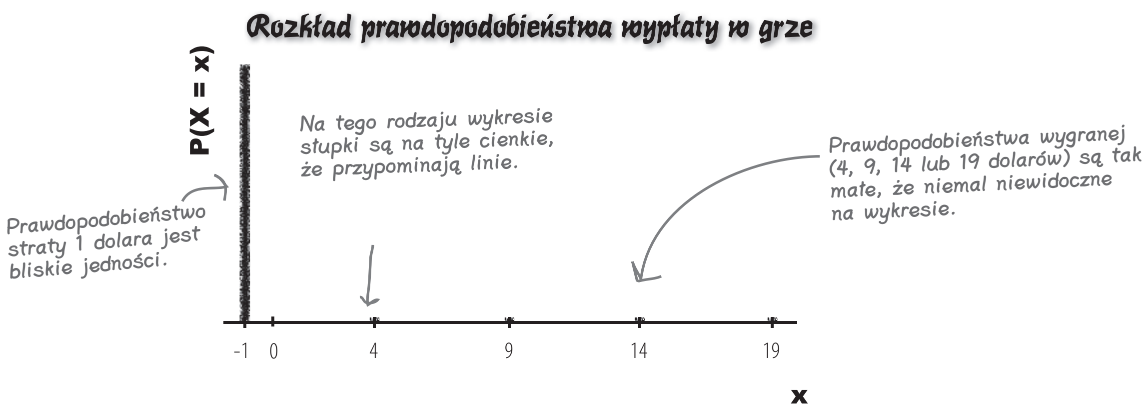
Kombinacja	brak wygranej	cytrynki	wisienki	dolary/wisienki	dolary
x	-1 $	4 $	9 $	14 $	19 $
P(X = x)	0,977	0,008	0,008	0,006	0,001

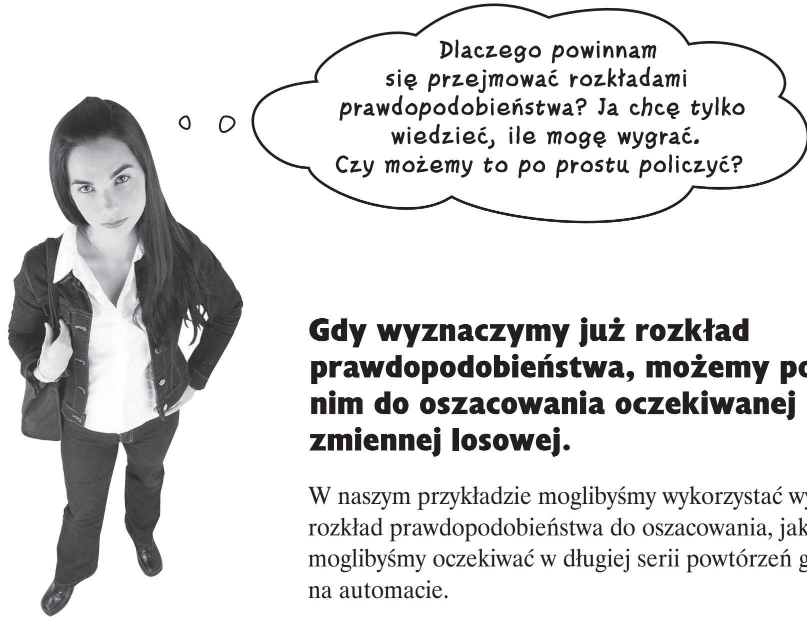
X to nasza zmienna losowa.

To jest P(X = 9), czyli prawdopodobieństwo tego, że na grze zyskamy 9 dolarów.

Ta zmienna losowa jest zmienną ***dyskretną***. Oznacza to, że przyjmuje ona tylko wybrane, dokładne wartości. Rozkład Poissona też jest dyskretny, ale nie jest skończony.

Oprócz przedstawienia rozkładu prawdopodobieństwa w formie tabelarycznej możemy przedstawić go na wykresie, co znacząco ułatwia dalszą analizę rozkładu. Poniżej zamieszczono przykładowy wykres (słupkowy) dla rozkładu naszej zmiennej.





Gdy wyznaczymy już rozkład prawdopodobieństwa, możemy posłużyć się nim do oszacowania oczekiwanej realizacji zmiennej losowej.

W naszym przykładzie moglibyśmy wykorzystać wyznaczony rozkład prawdopodobieństwa do oszacowania, jakiej wypłaty moglibyśmy oczekiwać w długiej serii powtórzeń gry na automacie.

Nie istnieją głupie pytania

P: Dlaczego zamiast symboli do budowy rozkładu wykorzystujemy liczby? Czy naprawdę tak wiele na tym zyskujemy?

O: Oczywiście moglibyśmy posługiwać się różnymi symbolami, jednak mają one podstawową wadę — nie nadają się do obliczeń. Już niedługo przekonasz się, że na bazie rozkładu opisanego za pomocą liczb można wyznaczyć różne parametry zmiennej losowej. Nie byłoby to możliwe, gdybyśmy posługiwali się wyłącznie symbolami.

P: Czy mogę przedstawić rozkład prawdopodobieństwa na diagramie Venna?

O: Diagramy Venna raczej nie nadają się do tego celu. Zarówno one, jak i drzewa stochastyczne są pomocne przy obliczaniu prawdopodobieństw, niekoniecznie przy ich wizualizacji.

P: Czy mogę wykorzystać dowolną literę do oznaczenia zmiennej losowej?

O: Oczywiście że tak, pod warunkiem że nie będzie to prowadziło do nieporozumień. Najczęściej wykorzystuje się litery z końca alfabetu: X, Y czy Z.

P: Czy zmienną i jej realizacje powinienem oznaczać tą samą literą?

O: W zasadzie mógłbyś użyć różnych liter, ale jak się jeszcze nieraz przekonasz, mogłoby to nadmiernie utrudnić rozeznanie w danych. Lepiej więc pozostać przy tej samej literze.

P: Stwierdziliście, że zmienna dyskretna może przyjmować tylko wybrane, dokładne wartości. Czy nie jest to regułą?

O: Nie jest. W przykładzie z „jednorękim bandytą” wiemy z góry, jaką wypłatę gwarantują poszczególne kombinacje symboli. Nigdy nie uda nam się uzyskać kombinacji innej niż przewidziana przez konstruktora automatu.

Czasami mamy do czynienia ze zmiennymi losowymi, które mogą przyjmować dowolną wartość z pewnego przedziału.* Średnica jabłka może przyjmować dowolną wartość na przykład z przedziału 5 – 15 cm.

Na razie nie musisz się jednak za bardzo tym przejmować. Wrócimy do tego zagadnienia w dalszej części książki. Każda zmienna, którą będziemy się zajmować w tym rozdziale, będzie zmienną dyskretną.

* Liczb naturalnych też jest nieskończenie wiele, ale zmienna losowa przyjmująca wartości naturalne jest nadal dyskretna.

Wartość oczekiwana pozwala przewidzieć wynik…

Znamy już rozkład prawdopodobieństwa wypłaty w grze na automacie, ale nie wiemy, jakiej tak naprawdę wygranej moglibyśmy oczekiwać. Odpowiedź na to pytanie da nam ***wartość oczekiwana***. Jest to jeden z parametrów rozkładu prawdopodobieństwa, mówiący o tym, jakiej przeciętnej wartości zmiennej losowej należy się spodziewać w długiej serii jej realizacji.

Jestem wartością oczekiwaną. Możesz mnie traktować jak średnią arytmetyczną.

Wartość oczekiwana zmiennej X przypomina nieco średnią arytmetyczną, tyle że jest wyznaczana na podstawie rozkładu prawdopodobieństwa. Liczymy ją w bardzo podobny sposób: wartość każdej realizacji x zmiennej losowej X mnożymy przez jej prawdopodobieństwo, a następnie sumujemy otrzymane iloczyny.

Wartość oczekiwaną zmiennej losowej X oznaczamy zwykle symbolem E(X), jednak równie często wykorzystuje się symbol μ, którym wcześniej oznaczaliśmy średnią arytmetyczną.

$$E(X) = \mu$$

Oto wzór na policzenie wartości oczekiwanej E(X):

Pomnóż każdą wartość x przez jej prawdopodobieństwo.

$$E(X) = \sum xP(X = x)$$

E(X) oznacza wartość oczekiwaną X.

Gdy wykonasz mnożenie, dodaj do siebie otrzymane iloczyny.

Posłużymy się tym wzorem do policzenia wartości oczekiwanej naszej zmiennej. Oto jej rozkład (pominęliśmy jednostki):

x	**–1**	**4**	**9**	**14**	**19**
P(X = x)	0,977	0,008	0,008	0,006	0,001

$$E(X) = (-1 \times 0{,}977) + (4 \times 0{,}008) + (9 \times 0{,}008) + (14 \times 0{,}006) + (19 \times 0{,}001) =$$

$$= -0{,}977 + 0{,}032 + 0{,}072 + 0{,}084 + 0{,}019 =$$

$$= -0{,}77$$

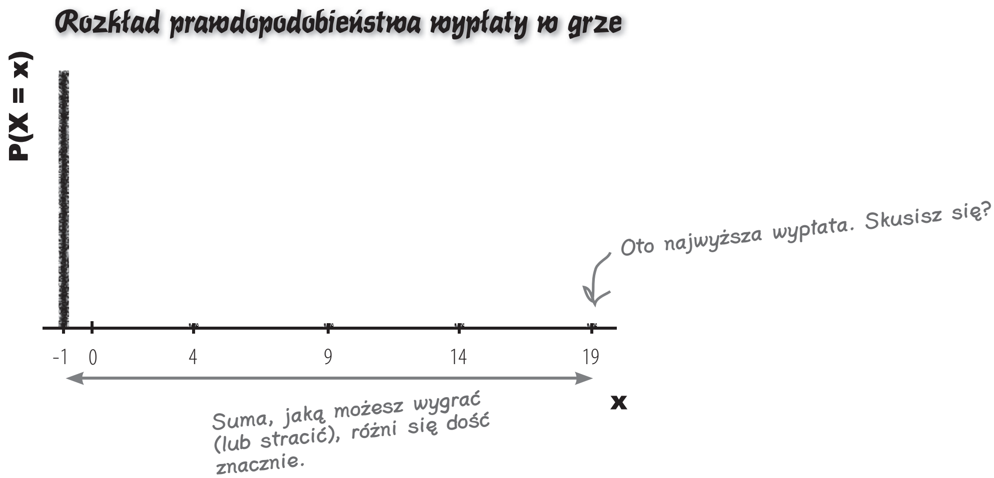
Oto wypłata (w dolarach), jakiej możesz oczekiwać, grając na naszym automacie — jak widzisz, jest ujemna!

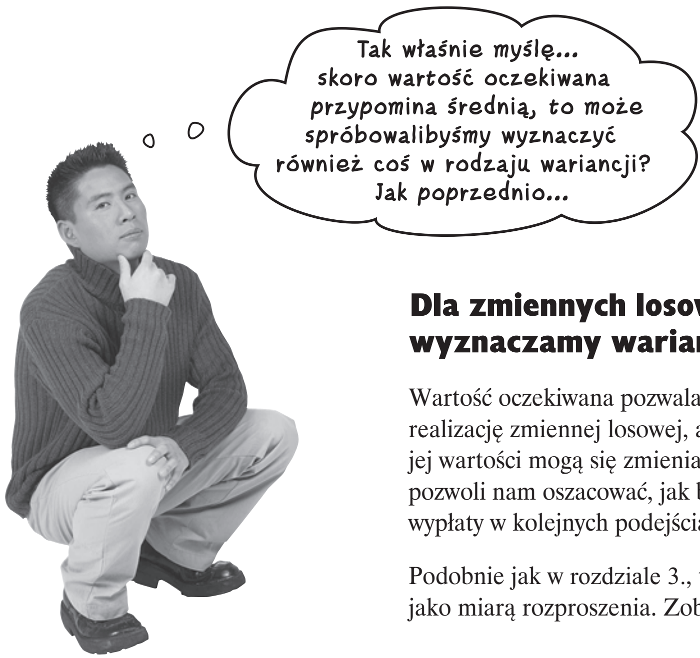
Innymi słowy, przy dużej liczbie powtórzeń gry powinieneś oczekiwać, że w każdej z nich stracisz przeciętnie 0,77 dolara. Czyli grając 100 razy, możesz oczekiwać straty na poziomie 77 dolarów.

...a wariancja mówi o tym, jak bardzo jest on zmienny

Wartość oczekiwana mówi o tym, jakiej przeciętnie wypłaty powinniśmy oczekiwać w pojedynczej grze. Ponieważ w naszym przykładzie jest ona ujemna, zastanawiasz się zapewne, dlaczego ludzie decydują się na grę.

Odpowiedź jest prosta: to, że powinieneś oczekiwać raczej strat niż zysków z gry na automacie, nie oznacza jeszcze, że nie masz w ogóle szans na wygraną. Tak jak średnia arytmetyczna, wartość oczekiwana nie mówi nam wszystkiego o danym rozkładzie. Wyniki kolejnych podejść do gry mogą się dość znacząco różnić. Jak sądzisz, czy możemy zmierzyć tę zmienność?





Dla zmiennych losowych również wyznaczamy wariancję.

Wartość oczekiwana pozwala oszacować typową, przeciętną realizację zmiennej losowej, ale nie mówi nic o tym, jak bardzo jej wartości mogą się zmieniać. W naszym przykładzie wariancja pozwoli nam oszacować, jak bardzo mogą różnić się od siebie wypłaty w kolejnych podejściach do gry.

Podobnie jak w rozdziale 3., tak i tutaj posłużymy się **wariancją** jako miarą rozproszenia. Zobaczmy, jak to wygląda w praktyce.

Wariancja a rozkład prawdopodobieństwa

W rozdziale 3. wyznaczaliśmy wariancję dla pewnego zbioru liczb. W pierwszym kroku obliczaliśmy wartość wyrażenia $(x-\mu)^2$ dla każdej liczby ze zbioru danych, a następnie wyznaczaliśmy ich średnią arytmetyczną.

Podobnie możemy postąpić w przypadku zmiennej losowej. Nie będziemy jednak liczyć średniej wyrażenia $(X-\mu)^2$, ale jego wartość oczekiwaną, tak jak w poniższej formule:

Oto wariancja. Wariancję zmiennej X oznaczamy w skrócie przez Var(X).

$$\mathrm{Var}(X) = E(X - \mu)^2$$

μ jest innym oznaczeniem E(X).

Musimy znaleźć wartość oczekiwaną wyrażenia (X–μ)². Tylko jak?

Powstał jednak problem: jak policzyć wartość oczekiwaną $(X-\mu)^2$?

Jak więc wyznaczyć $E(X-\mu)^2$?

Obliczanie $E(X-\mu)^2$ przypomina bardzo liczenie zwykłej E(X).

Aby policzyć E(X), musimy przemnożyć każdą wartość zmiennej losowej przez odpowiadające jej prawdopodobieństwo, a następnie dodać do siebie otrzymane iloczyny. Innymi słowy, postępujemy według wzoru:

$$E(X) = \Sigma x P(X = x)$$

Aby policzyć wariancję zmiennej losowej X, wyznaczamy wartość wyrażenia $(x-\mu)^2$ dla każdej wartości x zmiennej X, mnożymy tę wartość przez odpowiadające jej prawdopodobieństwo, a następnie dodajemy otrzymane iloczyny:

Dla każdej wartości x wyznacz (x–μ)². Następnie pomnóż to przez prawdopodobieństwo przypisane x...

$$E(X - \mu)^2 = \sum (x - \mu)^2 P(X = x)$$

...i dodaj do siebie wyznaczone iloczyny.

Innymi słowy, zamiast mnożyć x przez odpowiadające mu prawdopodobieństwo, mnożymy przez nie wyrażenie o postaci $(x-\mu)^2$.

Obliczamy wariancję dla naszego przykładu

Oto rozkład prawdopodobieństwa dla naszego przykładu.

x	–1	4	9	14	19
P(X = x)	0,977	0,008	0,008	0,006	0,001

Zobaczmy, jak możemy wykorzystać tę formułę do wyznaczenia wariancji naszej zmiennej losowej. Najpierw odejmiemy wartość oczekiwaną tej zmiennej od każdej z jej realizacji, podniesiemy tę różnicę do kwadratu, a następnie pomnożymy ją przez prawdopodobieństwo przypisane każdej realizacji. Dla przypomnienia: E(X), czyli μ, wynosi –0,77 dolara.

Na stronie 242 obliczyliśmy E(X) = –0,77.

$$\begin{aligned}
\mathrm{Var}(X) &= E(X-\mu)^2 = \\
&= (-1+0{,}77)^2 \times 0{,}977 + (4+0{,}77)^2 \times 0{,}008 + (9+0{,}77)^2 \times 0{,}008 + (14+0{,}77)^2 \times 0{,}006 + (19+0{,}77)^2 \times 0{,}001 = \\
&= (-0{,}23)^2 \times 0{,}977 + 4{,}77^2 \times 0{,}008 + 9{,}77^2 \times 0{,}008 + 14{,}77^2 \times 0{,}006 + 19{,}77^2 \times 0{,}001 = \\
&= 0{,}0516833 + 0{,}1820232 + 0{,}7636232 + 1{,}3089174 + 0{,}3908529 = \\
&= 2{,}6971
\end{aligned}$$

$(X-\mu)^2 \times P(X = x)$

Oznacza to, że wariancja zmiennej losowej X (wypłaty w grze) wynosi 2,6971.

A co z odchyleniem standardowym?
Czy także możemy je policzyć?

Podobnie jak wariancję, dla zmiennych losowych możemy także wyznaczyć odchylenie standardowe.

Pełni ono podobną rolę jak odchylenie standardowe, które wyznaczyliśmy w rozdziale 3. dla pewnego zbioru liczb. Określa ono, jak daleko od centrum rozkładu znajdują się przeciętnie wartości zmiennej.

Tak jak poprzednio, wyznaczamy je jako pierwiastek kwadratowy z wariancji:

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

Odchylenie standardowe możemy oznaczyć tym samym symbolem co poprzednio.

Oznacza to, że odchylenie standardowe wypłaty w grze na naszym automacie wynosi $\sqrt{2{,}6971}$, czyli 1,642 dolara. Można powiedzieć, że wypłata w grze różni się przeciętnie o 1,642 dolara od wartości oczekiwanej, wynoszącej –0,77 dolara.

WYSIL SZARE KOMÓRKI

Czy wolałbyś grać na automacie, który charakteryzuje duża, czy mała zmienność wypłat? Dlaczego?

Nie istnieją głupie pytania

P: Czyli wartość oczekiwana jest czymś w rodzaju średniej. Czy dla rozkładów prawdopodobieństwa istnieją również odpowiedniki mediany i dominanty?

O: Można wyznaczyć wartość zmiennej, która ma najwyższe prawdopodobieństwo realizacji, a więc coś w rodzaju dominanty. Zwykle jednak nie robi się tego. W przypadku rozkładów statystycy najczęściej posługują się wartością oczekiwaną jako miarą tendencji centralnej.

P: Czy wartość oczekiwana nie powinna znajdować się w zbiorze możliwych realizacji zmiennej losowej?

O: Niekoniecznie. Tak jak średnia nie musi znajdować się w zbiorze danych, dla którego została wyznaczona, tak i wartość oczekiwana nie musi znajdować się w zbiorze realizacji zmiennej losowej.

P: Czy wariancja i odchylenie standardowe to te same miary, które wyznaczaliśmy dla zbiorów liczb w rozdziale 3.?

O: W zasadzie tak, z wyjątkiem tego, że tym razem operujemy na rozkładach prawdopodobieństwa. I wariancja, i odchylenie standardowe pozostają jednak nadal miarami rozproszenia.

P: Nie do końca rozumiem, dlaczego liczymy $E(x-\mu)^2$. Czy to nie to samo, jak byśmy policzyli $E(x-\mu)$ i podnieśli wynik do kwadratu?

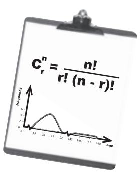
O: Nie, to dwie różne wartości. W przypadku $E(x-\mu)^2$ podnosimy do potęgi drugiej wyrażenie $(x-\mu)$ dla każdego x, a następnie wyznaczamy wartość oczekiwaną takich potęg. Gdy najpierw wyznaczymy $E(x-\mu)$, a następnie podniesiemy ją do potęgi drugiej, otrzymamy zupełnie inny wynik.

Tak naprawdę w pierwszym przypadku liczymy więc $E((x-\mu)^2)$, ale upraszczamy zapis.

P: Jaka jest więc różnica między automatem o małej wariancji a tym o dużej wariancji wypłat?

O: Automat o dużej wariancji wypłaca generalnie bardziej zróżnicowane kwoty. Sumę, jaką w ogólnym przypadku otrzymasz, trudniej jest przewidzieć.

Zapamiętaj, że im mniejsza wariancja, tym bliżej oczekiwanej wypłaty będą się znajdować kwoty rzeczywiście wypłacone. Decydując się na grę na automacie o większej wariancji, będziesz musiał się liczyć z wynikiem, który trudniej przewidzieć.

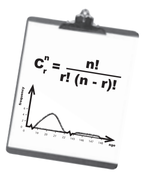


Podstawowe terminy

Wartość oczekiwana

Poniższą formułę wykorzystujemy do wyznaczania wartości oczekiwanej zmiennej losowej X:

$$E(X) = \sum xP(X = x)$$



Podstawowe terminy

Wariancja

Według tej formuły wyznaczamy wariancję zmiennej X:

$$Var(X) = E(X - \mu)^2$$

Ćwiczenie

Oto rozkład prawdopodobieństwa zmiennej losowej X:

x	**1**	**2**	**3**	**4**	**5**
P(X = x)	0,1	0,25	0,35	0,2	0,1

1. Ile wynosi E(X)?

2. Ile wynosi Var(X)?

Ćwiczenie
Rozwiązanie

Oto rozkład prawdopodobieństwa zmiennej losowej X:

x	**1**	**2**	**3**	**4**	**5**
P(X = x)	0,1	0,25	0,35	0,2	0,1

1. Ile wynosi E(X)?

Pomnóż każdą wartość zmiennej przez prawdopodobieństwo jej realizacji i dodaj do siebie tak uzyskane iloczyny.

$E(X) = \Sigma x P(X = x) =$

$= 1 \times 0{,}1 + 2 \times 0{,}25 + 3 \times 0{,}35 + 4 \times 0{,}2 + 5 \times 0{,}1 =$

$= 0{,}1 + 0{,}5 + 1{,}05 + 0{,}8 + 0{,}5 =$

$= 2{,}95$

2. Ile wynosi Var(X)?

Wyznacz $(x-\mu)^2$ dla każdej wartości x. Następnie pomnóż tę wartość przez prawdopodobieństwo x i dodaj do siebie tak uzyskane iloczyny.

$Var(X)^2 = E(X-\mu)^2 =$

$= \Sigma(x-\mu)^2 P(X = x) =$

$= (1-2{,}95)^2 \times 0{,}1 + (2-2{,}95)^2 \times 0{,}25 + (3-2{,}95)^2 \times 0{,}35 + (4-2{,}95)^2 \times 0{,}2 + (5-2{,}95)^2 \times 0{,}1 =$

$= (-1{,}95)^2 \times 0{,}1 + (-0{,}95)^2 \times 0{,}25 + 0{,}05^2 \times 0{,}35 + 1{,}05^2 \times 0{,}2 + 2{,}05^2 \times 0{,}1 =$

$= 3{,}8025 \times 0{,}1 + 0{,}9025 \times 0{,}25 + 0{,}0025 \times 0{,}35 + 1{,}1025 \times 0{,}2 + 4{,}2025 \times 0{,}1 =$

$= 0{,}38025 + 0{,}225625 + 0{,}000875 + 0{,}2205 + 0{,}42025 =$

$= 1{,}2475$

Przypadek średnich ruchomych

Lokalna stacja telewizyjna w Statsville nadaje kilka popularnych teleturniejów, z których największą widownią cieszy się „Idź na całość”. Każdemu uczestnikowi zabawy prezentowane są skrzynki z umieszczonymi wewnątrz pieniędzmi. Jego zadaniem jest wybór jednej z nich bez zaglądania do środka. Następnie skrzynki, które nie zostały wybrane przez zawodnika, są stopniowo otwierane. Po ujawnieniu zawartości każdej kolejno otwieranej skrzynki zawodnik musi zdecydować, czy chce pozostać przy swoim pierwotnym wyborze (dokonanym w ciemno), czy też wybrać inną ofertę, którą w międzyczasie złożył mu gospodarz programu. Każda wygrana zawodnika wiąże się z datkiem na miejscowe towarzystwo opieki nad fokami.

Na tropie wielkiej tajemnicy

Uczestnik dzisiejszego teleturnieju jest statystykiem-amatorem. Zdaje sobie sprawę z tego, że znalazłby się w dużo lepszym położeniu, gdyby poznał wartość oczekiwaną kwot zamieszczonych w poszczególnych skrzynkach. Właśnie skończył ją obliczać, gdy podszedł do niego producent teleturnieju.

„Wchodzi pan na antenę za 3 minuty” — powiedział producent. „Zmieniliśmy zawartość poszczególnych skrzynek. Zawierają one teraz niemal dwa razy tyle pieniędzy co do tej pory. Bez 10 dolarów”.

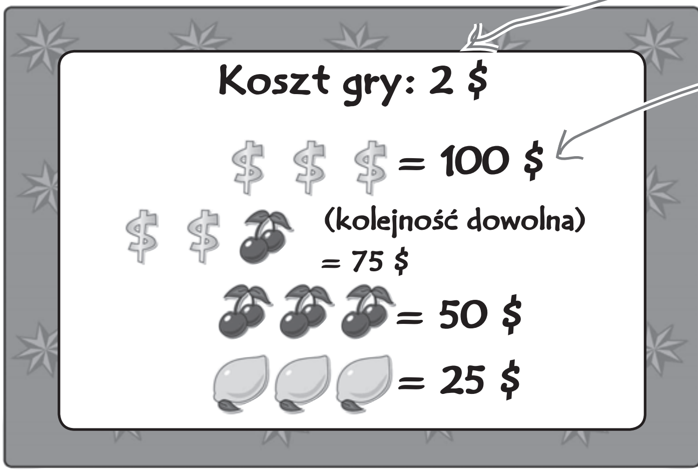
Zawodnik spojrzał na producenta z wyrzutem. Czy to oznacza, że jego obliczenia poszły na marne? Wątpił w to, by w ciągu trzech minut udało mu się przeprowadzić wszystkie obliczenia od początku. Co więc powinien zrobić?

Czy nasz zawodnik mógłby w prostszy sposób wyznaczyć wartość oczekiwaną, tak by nie zajęło mu to więcej niż 3 minuty?

Gdy ceny idą w górę

W ciągu kilku ostatnich minut właściciel kasyna podniósł cenę zakładu w grze na automatach, zwiększając jednocześnie wygrane. Oto plansza z nowymi stawkami:

Udział w grze kosztuje 2 dolary, a nie 1 dolar, jak dotąd.



Stawki wygranych wzrosły pięciokrotnie.

Koszt udziału w grze (pociągnięcia za dźwignię automatu) wzrósł do 2 dolarów, ale w ślad za tym poszły w górę stawki wygranych — są teraz pięć razy wyższe. Jeśli więc szczęście będzie nam sprzyjać, zarobimy dużo więcej niż poprzednio.

Oto nowy rozkład prawdopodobieństwa wypłat w tej grze:

y	-2	23	48	73	98
P(Y = y)	0,977	0,008	0,008	0,006	0,001

Tym razem zmienną losową oznaczyliśmy jako Y, nie X.

Gdybyśmy wiedzieli, jaka jest wartość oczekiwana i wariancja tej zmiennej, moglibyśmy oszacować wypłatę w długiej serii *powtórzeń*.

Zaostrz ołówek

Jaka jest wartość oczekiwana i wariancja zmiennej losowej Y?
Jak się mają te wartości do wyznaczonych poprzednio: wartości oczekiwanej wynoszącej –0,77 dolara oraz wariancji równej 2,6971?

y	-2	23	48	73	98
P(Y = y)	0,977	0,008	0,008	0,006	0,001

Zaostrz ołówek
Rozwiązanie

Jaka jest wartość oczekiwana i wariancja zmiennej losowej Y? Jak się mają te wartości do wyznaczonych poprzednio: wartości oczekiwanej wynoszącej –0,77 dolara oraz wariancji równej 2,6971?

y	**-2**	**23**	**48**	**73**	**98**
P(Y = y)	0,977	0,008	0,008	0,006	0,001

$E(Y) = (-2) \times 0{,}977 + 23 \times 0{,}008 + 48 \times 0{,}008 + 73 \times 0{,}006 + 98 \times 0{,}001 =$

$= -1{,}954 + 0{,}184 + 0{,}384 + 0{,}438 + 0{,}098 =$

$= -0{,}85$

$Var(Y) = E(Y-\mu)^2 =$

$= \Sigma(y-\mu)^2 P(Y = y) =$

$= (-2+0{,}85)^2 \times 0{,}977 + (23+0{,}85)^2 \times 0{,}008 + (48+0{,}85)^2 \times 0{,}008 + (73+0{,}85)^2 \times 0{,}006 + (98+0{,}85)^2 \times 0{,}001 =$

$= (-1{,}15)^2 \times 0{,}977 + 23{,}85^2 \times 0{,}008 + 48{,}85^2 \times 0{,}008 + 73{,}85^2 \times 0{,}006 + 98{,}85^2 \times 0{,}001 =$

$= 1{,}3225 \times 0{,}977 + 568{,}8225 \times 0{,}008 + 2386{,}3225 \times 0{,}008 + 5453{,}8225 \times 0{,}006 + 9771{,}3225 \times 0{,}001 =$

$= 1{,}2920825 + 4{,}55058 + 19{,}09058 + 32{,}722935 + 9{,}7713225 =$

$= 67{,}4275$

Wartość oczekiwana jest nieco niższa, a zatem w długiej serii powtórzeń oczekujemy przeciętnej straty na poziomie 85 centów w każdej grze. Wariancja jest dużo większa niż poprzednio. Przeciętnie spodziewamy się więc większej straty w każdej grze, choć mamy co do tego mniejszą pewność.

Stare i nowe stawki wygranych są ze sobą powiązane.

Koszt udziału w grze zwiększył się do 2 dolarów, zaś stawki wygranych poszły w górę pięciokrotnie. Skoro jest między nimi tak prosta i bezpośrednia zależność, to warto się przekonać, czy podobne zależności istnieją również między parametrami rozkładów zmiennych X i Y.

Przekonajmy się zatem.

Łamigłówka

Czas na odrobinę algebry. Uzupełnij luki w poniższych obliczeniach odpowiednimi wartościami znajdującymi się w basenie. Każdą wartość możesz wykorzystać tylko **jeden** raz, choć niektóre nie będą Ci potrzebne w ogóle. Jeśli to zadanie wykonasz bezbłędnie, otrzymasz formułę pokazującą relację między starymi (X) i nowymi (Y) stawkami wypłat w grze na automacie w kasynie Dana.

X = (stara stawka wygranej) – (stary koszt gry) =

= (stara stawka wygranej) – ....

(stara stawka wygranej) = ... + ...

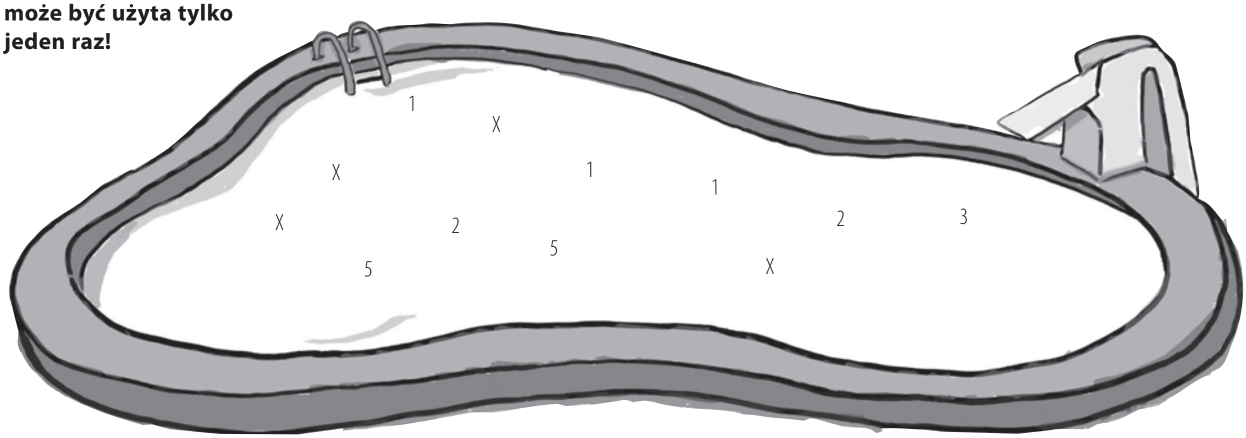
Y = 5 (stara stawka wygranej) – (nowy koszt gry) =

= 5 (............. + ...) – .... =

= 5 .... + ... – .... =

= .... + ...

Uwaga: każda wartość może być użyta tylko jeden raz!



Łamigłówka: Rozwiązanie

Czas na odrobinę algebry. Uzupełnij luki w poniższych obliczeniach odpowiednimi wartościami znajdującymi się w basenie. Każdą wartość możesz wykorzystać tylko **jeden** raz, choć niektóre nie będą Ci potrzebne w ogóle. Jeśli to zadanie wykonasz bezbłędnie, otrzymasz formułę pokazującą relację między starymi (X) i nowymi (Y) stawkami wypłat w grze na automacie w kasynie Dana.

X = (stara stawka wygranej) – (stary koszt gry) =

Koszt gry wynosił przedtem 1 dolar.

= (stara stawka wygranej) – ....1.......

(stara stawka wygranej) = ...X....... + ...1.......

W ten sposób przedstawiliśmy starą stawkę wypłaty jako funkcję X.

Podstawiamy tutaj wartości starych stawek wygranej.

Y = 5 (stara stawka wygranej) – (nowy koszt gry) =

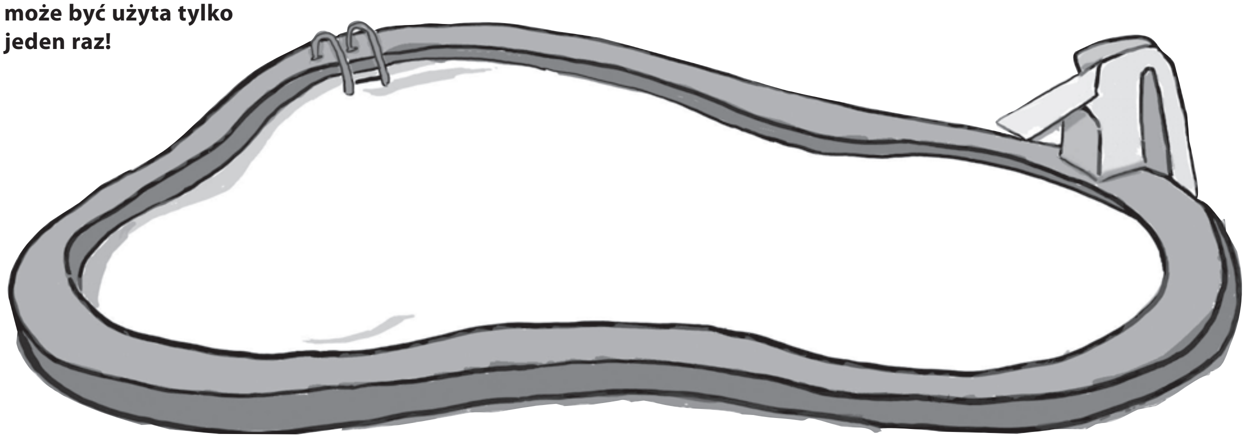
= 5 (.......X....... + ...1.......) – ....2....... =

= 5 ....X....... + ...5....... – ....2....... =

= ....5....... ....X....... + ...3.......

A zatem Y = 5X + 3. Między X i Y istnieje ściśle określony związek.

Uwaga: każda wartość może być użyta tylko jeden raz!



Między E(X) i E(Y) istnieje związek liniowy

Wiemy już, że między zmiennymi X (stara wypłata) i Y (nowa wypłata) istnieje zależność liniowa postaci $Y = 5X + 3$. Chcielibyśmy się teraz dowiedzieć, czy podobna zależność istnieje między E(X) i E(Y) oraz Var(X) i Var(Y).

Gdyby udało nam się potwierdzić istnienie takiej zależności, kolejne zmiany stawek w kasynie Dana nie byłyby dla nas problemem. Potrafilibyśmy bowiem łatwo wyznaczyć nową wartość oczekiwaną i wariancję na podstawie ich wartości sprzed zmiany.

Zaostrz ołówek

Spróbujmy sprawdzić, czy istnieje zależność między E(X) i E(Y) oraz Var(X) i Var(Y).

1. $E(X) = -0{,}77$, zaś $E(Y) = -0{,}85$. Ile wynosi $5 \times E(X)$? A ile $5 \times E(X) + 3$? Jak się to ma do E(Y)?

2. $Var(X) = 2{,}6971$, zaś $Var(Y) = 67{,}4275$. Ile wynosi $5 \times Var(X)$? A ile $5^2 \times Var(X)$? Jak się to ma do Var(Y)?

3. Czy dałoby się uogólnić zaobserwowane zależności dla dowolnych zmiennych losowych pozostających w relacji: $Y = aX + b$?

Zaostrz ołówek
Rozwiązanie

Spróbujmy sprawdzić, czy istnieje zależność między E(X) i E(Y) oraz Var(X) i Var(Y).

1. E(X) = –0,77, zaś E(Y) = –0,85. Ile wynosi 5 × E(X)? A ile 5 × E(X) + 3? Jak się to ma do E(Y)?

   $5 \times E(X) = -3{,}85$

   $5 \times E(X) + 3 = -0{,}85$

   $E(Y) = 5 \times E(X) + 3$

2. Var(X) = 2,6971, zaś Var(Y) = 67,4275. Ile wynosi 5 × Var(X)? A ile 5^2 × Var(X)? Jak się to ma do Var(Y)?

   $5 \times Var(X) = 13{,}4855$

   $5^2 \times Var(X) = 67{,}4275$

   $Var(Y) = 5^2 \times Var(X)$

3. Czy dałoby się uogólnić zaobserwowane zależności dla dowolnych zmiennych losowych pozostających w relacji: Y = aX + b?

   $E(aX + b) = a\,E(X) + b$

   $Var(aX + b) = a^2\,Var(X)$

Podsumujmy nasze rozważania

Podsumujmy to, czego dowiedziałeś się na kilku ostatnich stronach.

Najpierw obliczyłeś wartość oczekiwaną i wariancję zmiennej losowej X, która reprezentuje wypłatę w pojedynczej grze na automacie.

Następnie chciałeś się przekonać, jaki wpływ na parametry rozkładu zmiennej X wywrze zmiana stawek zarządzona przez Dana, właściciela kasyna. Nie chciałeś jednak liczyć wszystkiego od początku, dlatego spróbowałeś znaleźć relację między parametrami starego i nowego rozkładu prawdopodobieństwa. Przekonałeś się, że:

$$E(5X + 3) = 5E(X) + 3$$

$$Var(5X + 3) = 5^2Var(X)$$

Ogólne wzory na przekształcenia liniowe

Nasze rezultaty możemy uogólnić na dowolną zmienną losową. Dla każdej zmiennej losowej X zachodzą poniższe równości:

$$E(aX + b) = aE(X) + b$$

Pomnóż wartość oczekiwaną przez a, a następnie dodaj b.

$$Var(aX + b) = a^2Var(X)$$

Pomnóż wariancję przez kwadrat a (pomijając b).

Tego rodzaju transformacje nazywamy przekształceniami liniowymi, ponieważ zmienna X występuje w pierwszej potędze, a o zmiennej Y mówimy, że jest funkcją X. Zmieniamy tu jedynie potencjalne wartości X, przekształcając je do postaci aX + b, lecz nie zmieniając prawdopodobieństwa ich realizacji.

Nie istnieją głupie pytania

P: Czy a i b muszą być stałymi liczbami?

O: Muszą. Jeśli a i b nie są stałe, powyższe wzory nie są prawdziwe.

P: Co się stało z b we wzorze na wariancję?

O: Dodanie ustalonej wartości do zmiennej losowej nie zmienia jej wariancji, a jedynie wartość oczekiwaną.

Spowoduje ono przesunięcie środka rozkładu w pewnym kierunku, ale nie zmieni ogólnego kształtu jego wykresu. Oznacza to, że zmieni się wartość oczekiwana, ale rozrzut danych, a więc i wariancja, pozostanie niezmieniony.

P: Zaskoczyło mnie to, że wariancję trzeba pomnożyć przez a^2. Dlaczego nie przez a?

O: Mnożąc zmienną losową przez stałą, każdą jej wartość mnożymy przez tą stałą.

We wzorze na wariancję wartości zmiennej losowej podnoszone są do potęgi drugiej. Dlatego gdy pomnożymy je przez a, wariancja zwiększy się a^2 razy.

P: Czy naprawdę muszę pamiętać wzory na przekształcenia liniowe parametrów? Czy są one aż tak ważne?

O: Tak, są ważne. Dzięki nim możesz zaoszczędzić mnóstwo czasu, który musiałbyś przeznaczyć na wyznaczanie wartości parametrów rozkładu za każdym razem, gdy zmieniają się wartości zmiennej losowej. Zamiast więc wyznaczać rozkład nowej zmiennej, obliczać od nowa jej wartość oczekiwaną i wariancję, musisz podstawić jedynie do odpowiedniego wzoru stare wartości parametrów.

Znajomość tych wzorów może się też przydać w czasie egzaminów ze statystyki. Znajomość tego rodzaju skrótów pozwoli Ci oszczędzić sporo cennego czasu. Musisz też pamiętać, że na egzaminach nie zawsze podawane są pełne rozkłady prawdopodobieństwa, a jedynie pewne ich charakterystyki. Niekiedy więc możesz nie mieć wyjścia.

P: Próbowałem wyznaczyć oba parametry na piechotę, ale otrzymałem inne wartości. Dlaczego?

O: A zatem przekonałeś się, że liczenie na piechotę może być ryzykowne. Po drodze masz bowiem wiele okazji do popełnienia błędów. Zawsze, kiedy masz taką możliwość, powinieneś korzystać ze znanych Ci skrótów.

Przypadek średnich ruchomych: Rozwiązanie

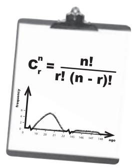
Czy nasz zawodnik mógłby w prostszy sposób wyznaczyć wartość oczekiwaną, tak by nie zajęło mu to więcej niż 3 minuty?

Przez krótką chwilę nasz zawodnik rozglądał się nerwowo dokoła, jednak szybko się uspokoił. Zmiana stawek wygranej to w końcu nie taki duży problem.

Już raz policzył wartość oczekiwaną kwot znajdujących się w kolejnych skrzynkach, co dało mu jakiś ogląd tego, ile może wygrać w tym teleturnieju.

Producent programu powiedział, że nowe kwoty pieniędzy ukryte w skrzynkach są niemal dwa razy wyższe od dotychczasowych, z dokładnością do 10 dolarów. Oznacza to, że między nowymi (Y) i starymi (X) kwotami istnieje linowy związek, który można by zapisać jako $Y = 2X - 10$.

To pozwala szybko obliczyć $E(Y)$ przy wykorzystaniu faktu, że $E(2X - 10) = 2E(X) - 10$. Zatem jedyne, co powinien zrobić nasz zawodnik, to pomnożyć obliczoną wcześniej wartość oczekiwaną wygranej przez 2 i od tego wyniku odjąć 10 dolarów.



Podstawowe terminy

Przekształcenia liniowe

Dla każdej zmiennej losowej X i stałych wartości a i b prawdziwe są równości:

$$E(aX + b) = aE(X) + b$$

$$Var(aX + b) = a^2Var(X)$$

CELNE SPOSTRZEŻENIA

- **Rozkład prawdopodobieństwa** przypisuje każdej wartości zmiennej losowej prawdopodobieństwo jej realizacji.
- **Wartość oczekiwana** informuje o przeciętnej realizacji zmiennej losowej w długiej serii powtórzeń. Oznacza się ją jako $E(X)$ lub μ i oblicza ze wzoru: $E(X) = \Sigma xP(X = x)$.
- **Wartość oczekiwana funkcji** zmiennej losowej X dana jest wzorem: $E(f(X)) = \Sigma f(x)P(X = x)$.
- **Wariancję** zmiennej losowej X wyznacza się według wzoru: $Var(X) = E(X - \mu)^2$.
- **Odchylenie standardowe** zmiennej losowej dane jest wzorem: $\sigma = \sqrt{Var(X)}$
- Z **przekształceniem liniowym** zmiennej losowej X mamy do czynienia wtedy, gdy jej wartości przekształcamy do ogólnej postaci: $aX + b$, gdzie a i b są wartościami ustalonymi. Wartość oczekiwana i wariancja przekształconej zmiennej są równe:
  $E(aX + b) = aE(X) + b$
  $Var(aX + b) = a^2Var(X)$

A zatem przedstawiając jedną zmienną losową jako funkcję innej zmiennej, będę mogła łatwo wyznaczyć parametry rozkładu także wtedy, gdy zagram więcej niż raz?

Przekształcenia liniowe nie sprawdzą się wtedy, gdy zmienia się prawdopodobieństwo realizacji zmiennej.

W wyniku liniowego przekształcenia zmiennej losowej zmieniają się jedynie wartości, jakie może ona przyjąć. Prawdopodobieństwa ich realizacji nie ulegają zmianie. Nie zmienia się również liczba potencjalnych wartości.

Gdy myślimy o rozegraniu kilku gier z rzędu, rozkład prawdopodobieństwa wypłaty będzie całkowicie inny: zmienią się zarówno wartości samej zmiennej, jak i prawdopodobieństwa ich realizacji. Nie wystarczy więc policzyć wyłącznie nowe wartości rozkładu, ale i ich prawdopodobieństwa, co nie musi być zadaniem łatwym.

Spójrzmy na przykład. Załóżmy, że gra toczy się na bardzo prostym automacie, który oferuje wypłaty według poniższego rozkładu:

x	–1	5
P(X = x)	0,9	0,1

Wartości zmiennej zostały pomnożone przez 2. Prawdopodobieństwa nie zmieniły się.

Aby znaleźć rozkład prawdopodobieństwa zmiennej 2X, wystarczy pomnożyć przez 2 wszystkie wartości zmiennej X:

2x	-2	10
P(2X = 2x)	0,9	0,1

Jak wyglądałby rozkład prawdopodobieństwa wypłaty z gry przy założeniu, że zagralibyśmy dwukrotnie? Tym razem wszystkie prawdopodobieństwa musimy obliczyć od nowa, rozpatrując wszystkie możliwe kombinacje wygranej (przegranej) w obu grach:

y = –2, gdy przegrasz w obu grach.

Tym razem ciągniemy dźwignię dwukrotnie. Potencjalne zyski i ich prawdopodobieństwa różnią się.

w	-2	4	10
P(Y = y)	0,81	0,18	0,01

y = 10, gdy wygrasz w obu grach.

Są to łączne wyniki uzyskane w obu grach

y = 4, gdy przegrasz w jednej grze, ale wygrasz w drugiej.

Tym razem zarówno wartości zmiennej, jak i ich prawdopodobieństwa są nieco inne. Czy istnieje jakiś sposób na skrócenie obliczeń w podobnych przypadkach?

Każde pociągnięcie dźwigni jest niezależnym zdarzeniem

Każdą grę możemy traktować jako niezależne zdarzenie, opisywane za pomocą oddzielnej zmiennej losowej, a jej wynik jako oddzielną obserwację. Wszystkie obserwacje będące realizacjami zmiennych losowych o tym samym rozkładzie, mają tę samą wartość oczekiwaną i wariancję, choć same w sobie mogą się od siebie różnić — za każdym razem możemy uzyskać nieco inną kwotę.

Przydałby nam się jakiś sposób na to, by łatwo odróżnić od siebie poszczególne zdarzenia czy obserwacje. Jeśli wypłatę w pojedynczej grze oznaczyliśmy przez X, to zmienne opisujące wygrane w poszczególnych kolejkach moglibyśmy oznaczyć na przykład poprzez X_1, X_2 itd.

Każda gra to jedno zdarzenie. Wynik każdej gry to jedna obserwacja.

Mamy tę samą wartość oczekiwaną i wariancję, ale jesteśmy innymi zmiennymi.

Zmienna opisująca wynik pierwszej gry.

X_1

X_2

Zmienna opisująca wynik drugiej gry.

Obie zmienne: X_1 i X_2 mają te same prawdopodobieństwa realizacji oraz tę samą wartość oczekiwaną i wariancję co zmienna X. Innymi słowy, mają one ten sam rozkład, choć w rzeczywistości są to oddzielne zmienne, których zaobserwowane wartości mogą się różnić.

X_1 i X_2 opierają się na mnie jako zmiennej wyjściowej.

x	-1	5
P(X = x)	0,9	0,1

Rozkład prawdopodobieństwa wyjściowej zmiennej X.

Właśnie stąd wziął się nasz rozkład prawdopodobieństwa.

x_1	-1	5
$P(X_1 = x_1)$	0,9	0,1

x_2	-1	5
$P(X_2 = x_2)$	0,9	0,1

Jeśli więc chcemy znaleźć wartość oczekiwaną i wariancję łącznej wypłaty w obu grach, tak naprawdę musimy policzyć je dla sumy zmiennych $X_1 + X_2$. Spójrzmy, jak możemy tego dokonać jak najmniejszym kosztem.

Przydatne skróty

Spróbujmy znaleźć wartość oczekiwaną i wariancję dla sumy zmiennych $X_1 + X_2$.

$X_1 + X_2$ to nie to samo co 2X.

Sumując zmienne X_1 i X_2, musisz się oprzeć na realizacjach każdej z nich. 2X oznacza tylko jedną realizację, tyle że podwojoną w stosunku do X.

Wartość oczekiwana

Policzmy najpierw $E(X_1 + X_2)$:

$$E(X_1 + X_2) = E(X_1) + E(X_2) =$$

$$= E(X) + E(X) =$$

Zarówno $E(X_1)$, jak i $E(X_2)$ są równe E(X), ponieważ obie zmienne mają rozkład taki jak X.

$$= 2E(X)$$

Innymi słowy, aby wyznaczyć $E(X_1 + X_2)$, wystarczy pomnożyć E(X) przez 2. Gdybyśmy więc dwukrotnie zagrali na automacie, dla którego E(X) = –0,77 dolara, moglibyśmy oczekiwać straty na poziomie –0,77×2, czyli –1,54 dolara.

Wynik ten możemy uogólnić na dowolną liczbę zmiennych. Gdybyśmy chcieli wyznaczyć wartość oczekiwaną *n* zmiennych tego rodzaju, moglibyśmy wykorzystać wzór:

Gdy mamy n zmiennych, mnożymy przez n wartość E(X).

$$\mathbf{E(X_1 + X_2 + \ldots X_n) = nE(X)}$$

Wariancja

A jak policzyć $Var(X_1 + X_2)$? Odpowiedź poniżej:

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) =$$

$$= Var(X) + Var(X) =$$

$Var(X_1)$ i $Var(X_2)$ są takie same jak Var(X), ponieważ obie zmienne mają rozkład taki jak X.

$$= 2Var(X)$$

Dlatego wariancja wypłaty w przypadku dwukrotnej gry na automacie o wariancji równej 2,6971 wyniosłaby 2×2,6971, czyli 5,3942.

Wynik ten możemy uogólnić na dowolną liczbę niezależnych obserwacji. Gdybyśmy chcieli wyznaczyć wariancję *n* obserwacji tego rodzaju, moglibyśmy wykorzystać wzór:

Pomnóż Var(X) przez liczbę obserwacji n.

$$\mathbf{Var(X_1 + X_2 + \ldots X_n) = nVar(X)}$$

Podsumowując: aby znaleźć wartość oczekiwaną i wariancję w przypadku wielokrotnej gry, pomnóż E(X) i Var(X) przez liczbę obserwacji (kolejek gry).

Nie istnieją głupie pytania

P: Czy $E(X_1 + X_2)$ to nie to samo, co $E(2X)$?

O: Nie, choć rzeczywiście wyglądają podobnie.

$E(2X)$ jest wartością oczekiwaną zmiennej losowej, której wartości zostały podwojone względem zmiennej X. Mamy więc do czynienia z wartościami jednej zmiennej.

$E(X_1 + X_2)$ to wartość oczekiwana sumy dwóch różnych zmiennych losowych: X_1 i X_2. Jeśli X_1 oznacza wygraną w jednej grze, a X_2 w drugiej, $X_1 + X_2$ oznaczać będzie łączną wygraną w obu grach.

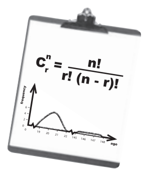
P: Rozumiem więc, że X_1 i X_2 są tym samym?

O: Niezupełnie. Obie zmienne mają ten sam rozkład prawdopodobieństwa, choć są to dwie różne zmienne. X_1 może oznaczać wygraną w jednej kolejce, a X_2 w drugiej. Choć mają ten sam rozkład prawdopodobieństwa, to wygrane w każdej kolejce mogą się od siebie znacząco różnić.

P: Dlaczego wariancja w przypadku n zmiennych wynosi nVar(X), a nie n^2Var(X), jak dla zmiennych przekształconych liniowo?

O: Tym razem mamy do czynienia z szeregiem niezależnych zmiennych losowych o tym samym rozkładzie prawdopodobieństwa. Całkowitą wariancję możemy wyznaczyć jako sumę wariancji poszczególnych zmiennych. Ponieważ mamy n zmiennych niezależnych, wynosi ona nVar(X).

W przypadku zmiennej nX każdą z wartości zmiennej X mnożymy przez n. Ponieważ w formule na wariancję występują kwadraty tych wartości, stąd w wyniku n^2Var(X), a nie nVar(X).



Podstawowe terminy

Zmienne niezależne

Dla niezależnych zmiennych losowych prawdziwe są wzory:

$$E(X_1 + X_2 + \ldots + X_n) = nE(X)$$
$$Var(X_1 + X_2 + \ldots + X_n) = nVar(X)$$

CELNE SPOSTRZEŻENIA

- Rozkład prawdopodobieństwa opisuje prawdopodobieństwo realizacji każdej możliwej wartości zmiennej losowej.
- Wartość oczekiwana zmiennej losowej X opisuje przeciętny wynik możliwy do uzyskania w długiej serii powtórzeń. Oznaczamy ją jako E(X) lub μ i obliczamy ze wzoru:

  $$E(X) = \Sigma x P(X=x)$$
- Wariancja zmiennej losowej X dana jest wzorem:

  $$Var(X) = E(X - \mu)^2$$
- Odchylenie standardowe σ jest pierwiastkiem kwadratowym z wariancji.
- Przekształcenie liniowe polega na przekształceniu zmiennej X w zmienną aX + b, gdzie a i b są stałymi. Wartość oczekiwaną i wariancję nowej zmiennej można obliczyć ze wzorów:

  $$E(aX + b) = aE(X) + b$$
  $$Var(aX + b) = a^2Var(X)$$

Zmienne przekształcone czy niezależne?

Poniżej zamieszczono opisy kilku sytuacji z życia codziennego. Zakładając, że znasz rozkład zmiennej X, zdecyduj, czy występujące w opisach zdarzenia można opisać za pomocą przekształcenia liniowego (funkcji) zmiennej X, czy też występują w nich zmienne niezależne.

	Przekształcenie liniowe	Zmienne niezależne
Ilość kawy, która składa się na dużą porcję; X — ilość kawy zawarta w standardowej porcji.	☐	☐
Spożywanie dodatkowego kubka kawy każdego dnia; X — ilość kawy w jednym kubku.	☐	☐
Wypłata możliwa do uzyskania po nabyciu 10 losów na loterii; X — wypłata możliwa do uzyskania po nabyciu jednego losu.	☐	☐
Wypłata możliwa do uzyskania po nabyciu pojedynczego losu, którego cena poszła w górę; X — wypłata możliwa do uzyskania po nabyciu pojedynczego losu w starej cenie.	☐	☐
Kupno dodatkowych kur znoszących jajka; X — liczba znoszonych jaj w zależności od gatunku kury.	☐	☐

Zmienne przekształcone czy niezależne?

Poniżej zamieszczono opisy kilku sytuacji z życia codziennego. Zakładając, że znasz rozkład zmiennej X, zdecyduj, czy występujące w opisach zdarzenia można opisać za pomocą przekształcenia liniowego (funkcji) zmiennej X, czy też występują w nich zmienne niezależne.

	Przekształcenie liniowe	Zmienne niezależne
Ilość kawy, która składa się na dużą porcję; X — ilość kawy zawarta w standardowej porcji.	☑	☐
Spożywanie dodatkowego kubka kawy każdego dnia; X — ilość kawy w jednym kubku.	☐	☑
Wypłata możliwa do uzyskania po nabyciu 10 losów na loterii; X — wypłata możliwa do uzyskania po nabyciu jednego losu.	☐	☑
Wypłata możliwa do uzyskania po nabyciu pojedynczego losu, którego cena poszła w górę; X — wypłata możliwa do uzyskania po nabyciu pojedynczego losu w starej cenie.	☑	☐
Kupno dodatkowych kur znoszących jajka; X — liczba znoszonych jaj w zależności od gatunku kury.	☐	☑

Wygrane z każdego losu są niezależne od siebie.

Zmiana ceny losu wpływa na wartość wypłaty, ale nie na prawdopodobieństwo wygranej, dlatego mamy do czynienia z przekształceniem liniowym.

Ćwiczenie

Lokalna cukiernia włączyła do swojej oferty ciasteczka z niespodzianką. Ich cena to 0,5 dolara za sztukę. Większość ciasteczek zawiera pomyślną wróżbę na przyszłość, ale w niektórych kryją się pieniądze. Prawdopodobieństwo wygrania 2 dolarów wynosi 0,1, 5 dolarów — 0,07, zaś 10 dolarów — 0,03.

Niech X oznacza wypłatę (wygrana pomniejszona o koszt ciasteczka) z tej „gry". Sporządź rozkład prawdopodobieństwa zmiennej X. Ile wynosi E(X) i Var(X)?

Cukiernia zdecydowała się podnieść cenę ciasteczek do 1 dolara za sztukę. Ile teraz wynosi wartość oczekiwana i wariancja wypłaty?

Ćwiczenie
Rozwiązanie

Lokalna cukiernia włączyła do swojej oferty ciasteczka z niespodzianką. Ich cena to 0,5 dolara za sztukę. Większość ciasteczek zawiera pomyślną wróżbę na przyszłość, ale w niektórych kryją się pieniądze. Prawdopodobieństwo wygrania 2 dolarów wynosi 0,1, 5 dolarów — 0,07, zaś 10 dolarów — 0,03.

Niech X oznacza wypłatę (wygrana pomniejszona o koszt ciasteczka) z tej „gry". Sporządź rozkład prawdopodobieństwa zmiennej X. Ile wynosi E(X) i Var(X)?

Oto rozkład prawdopodobieństwa zmiennej X:

x	$-0,5$	$1,5$	$4,5$	$9,5$
$P(X = x)$	$0,8$	$0,1$	$0,07$	$0,03$

$$E(X) = (-0,5)\times 0,8 + 1,5\times 0,1 + 4,5\times 0,07 + 9,5\times 0,03 =$$

$$= -0,4 + 0,15 + 0,315 + 0,285 =$$

$$= 0,35$$

$$Var(X) = E(X-\mu)^2 =$$

$$= \sum(x-\mu)^2 P(X = x) =$$

$$= (-0,5-0,35)^2\times 0,8 + (1,5-0,35)^2\times 0,1 + (4,5-0,35)^2\times 0,07 + (9,5-0,35)^2\times 0,03 =$$

$$= (-0,85)^2\times 0,8 + 1,15^2\times 0,1 + 4,15^2\times 0,07 + 9,15^2\times 0,03 =$$

$$= 0,7225\times 0,8 + 1,3225\times 0,1 + 17,2225\times 0,07 + 83,7225\times 0,03 =$$

$$= 0,578 + 0,13225 + 1,205575 + 2,511675 =$$

$$= 4,4275$$

Cukiernia zdecydowała się podnieść cenę ciasteczek do 1 dolara za sztukę. Ile teraz wynosi wartość oczekiwana i wariancja wypłaty?

Cukiernia podniosła cenę ciasteczek o 0,5 dolara, a zatem wypłatę po podwyżce można przedstawić jako X–0,5.

$$E(X-0,5) = E(X) - 0,5 =$$

$$= 0,35 - 0,5 =$$

$$= -0,15$$

$$Var(X-0,5) = Var(X) =$$

$$= 4,4275$$

Nowe automaty wchodzą do gry!

Właściciel kasyna sprowadził całkiem nowy model automatów do gry. Każda gra kosztuje teraz więcej, ale jeśli szczęście będzie Ci sprzyjać, będziesz mógł liczyć na naprawdę spore pieniądze. Oto nowy rozkład prawdopodobieństwa:

Gra kosztuje więcej niż na innych automatach, ale spójrz na te wygrane!

x	-5	395
P(X = x)	0,99	0,01

Wiemy już, jak wyznaczyć wartość oczekiwaną i wariancję zarówno w pojedynczej grze, jak i w przypadku kilku gier rozegranych na tym samym automacie. Co by się jednak stało, gdybyśmy chcieli rozegrać kilka kolejek i na starych, i na nowych urządzeniach?

W tej sytuacji mamy do czynienia z dwoma niezależnymi rozkładami prawdopodobieństwa — dla maszyn starego i nowego typu:

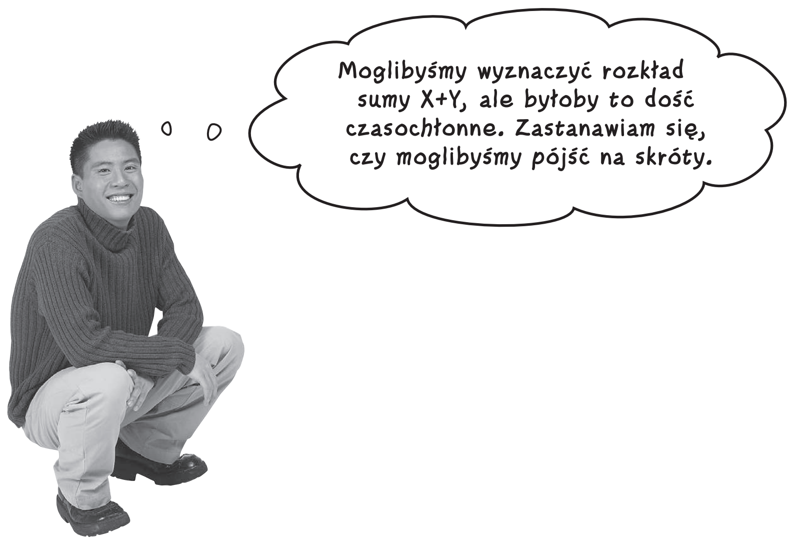
x	-5	395
P(X = x)	0,99	0,01

Oto wygrane możliwe do uzyskania na automatach nowego typu.

y	-2	23	48	73	98
P(Y = y)	0,977	0,008	0,008	0,006	0,001

Oto wygrane w automatach starego typu.

Czy możemy łatwo policzyć wartość oczekiwaną i wariancję wypłaty w przypadku wielokrotnej gry na obu rodzajach automatów?



Dodaj E(X) do E(Y), by uzyskać E(X+Y)...

Chcemy znaleźć wartość oczekiwaną i wariancję wypłaty przy założeniu, że gramy raz na automacie starego typu i raz na automacie nowego typu. Innymi słowy, interesuje nas wartość E(X+Y) i Var(X+Y), gdzie X i Y są zmiennymi losowymi opisującymi grę na starych i nowych urządzeniach. Są to zmienne niezależne.

W tym celu moglibyśmy wyznaczyć rozkład prawdopodobieństwa sumy X+Y, a następnie znaleźć jego wartość oczekiwaną i wariancję.

x		

y		

Nie martw się, nie będziesz musiał tego liczyć.

x + y		

Wyobrażasz sobie, ile czasu zajęłoby nam policzenie tego wszystkiego i ile błędów moglibyśmy przy tym popełnić?

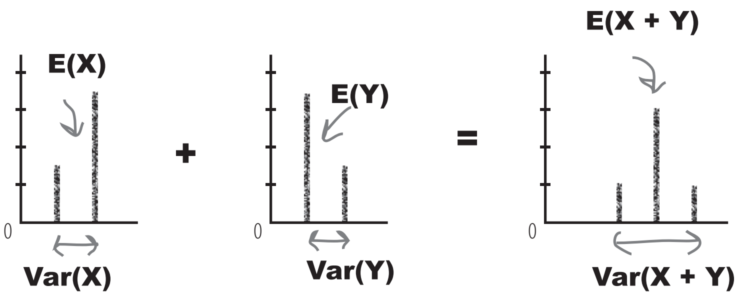
Na szczęście nie musimy tego robić. Aby znaleźć E(X+Y), możemy po prostu dodać do siebie E(X) i (Y).

E(X + Y) = E(X) + E(Y)

Przyznasz sam, że ma to sens. Gdybyś grał w dwie gry, przy czym w pierwszej oczekiwałbyś wygranej w wysokości 5 dolarów, a w drugiej 10 dolarów, łącznie spodziewałbyś się wygrać 5+10 = 15 dolarów.

Var(X + Y) = Var(X) + Var(Y)

Podobnie możemy postąpić w przypadku wariancji. Aby znaleźć Var(X+Y), możemy dodać do siebie wariancje obu zmiennych. Jednak tylko wtedy, gdy X i Y są zmiennymi niezależnymi.



Wariancja się zwiększa — rozkład prawdopodobieństwa jest bardziej rozproszony.

Swobodnie można dodawać tylko wariancje zmiennych niezależnych.

Jeśli X i Y nie są zmiennymi niezależnymi, wtedy Var(X+Y) nie jest równa Var(X) + Var(Y).

...lub odejmij E(Y) od E(X), by uzyskać E(X–Y)

W podobny sposób możemy również wyznaczyć parametry zmiennej będącej różnicą dwóch zmiennych losowych, czyli X–Y.

Równie łatwo jak w przypadku sumy zmiennych można wyznaczyć E(X–Y). Wystarczy odjąć E(Y) od E(X).

Formuła na wariancję jest nieco mniej intuicyjna. Aby znaleźć Var(X–Y), musimy… dodać do siebie Var(X) i Var(Y).

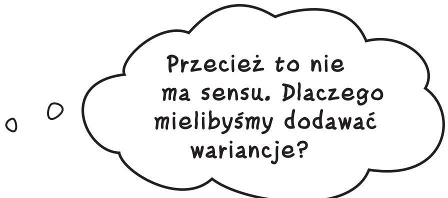
$$E(X - Y) = E(X) - E(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

Tutaj wariancje sumujemy, więc bądź ostrożny!

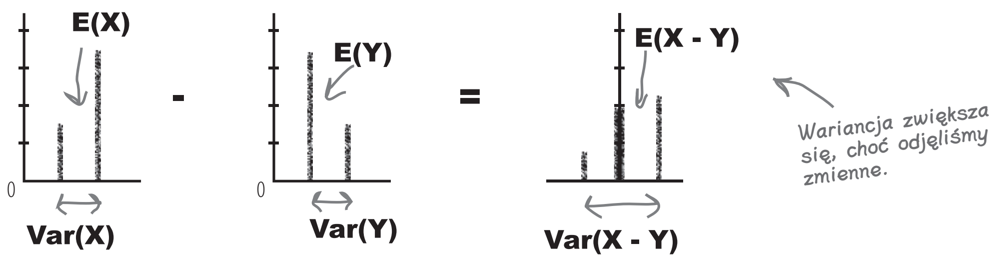
Wyznaczając wariancję różnicy zmiennych losowych, zsumuj ich wariancję.

Łatwo tu o pomyłkę, ponieważ w pierwszej chwili wydaje się to sprzeczne z intuicją. Zapamiętaj jednak, że jeśli X i Y są niezależnymi zmiennymi losowymi, to Var(X–Y) = Var(X) + Var(Y).



Ponieważ w tym przypadku zmienność również się zwiększa.

Chociaż odejmujemy od siebie dwie zmienne losowe, to wariancja rozkładu wynikowego nie zmniejsza się, ale zwiększa.



Wariancja zmiennej będącej różnicą dwóch niezależnych zmiennych losowych jest dokładnie taka sama, jak wariancja zmiennej będącej ich sumą. W obu przypadkach zmienność może jedynie wzrosnąć.

Odjęcie od siebie dwóch zmiennych losowych zwiększa wariancję.

Podobne operacje możesz wykonywać na zmiennych przekształcanych liniowo

To jeszcze nie wszystko. Oprócz dodawania i odejmowania samych zmiennych losowych te same operacje można wykonywać na ich przekształceniach liniowych.

Zastanów się, co by się stało, gdyby właściciel kasyna zmienił stawki opłat i wygranych w maszynach starego i nowego typu. Ostatnią rzeczą, jaką chcielibyśmy robić, to tworzyć od początku jeszcze jeden rozkład prawdopodobieństwa i wyznaczać jego parametry.

Na szczęście istnieje wygodna droga na skróty.

Załóżmy, że Dan zmienił stawki wygranych na wszystkich maszynach dostępnych w kasynie, tak że wypłata z maszyn starego typu zmieniła się z X na aX, zaś z Y na bY, gdzie a i b są ustalonymi wartościami.

$$X \longrightarrow aX$$

$$Y \longrightarrow bY$$

a i b mogą być dowolnymi liczbami.

Aby znaleźć wartość oczekiwaną i wariancję liniowej kombinacji aX i bY, możemy postąpić w opisany niżej sposób.

Sumujemy aX i bY

Wartość oczekiwaną i wariancję zmiennej aX + bY znajdziemy, korzystając ze wzorów:

$$E(aX + bY) = aE(X) + bE(Y)$$

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y)$$

Jest to rodzaj przekształcenia liniowego, stąd te kwadraty.

W formule na wariancję występują kwadraty, ponieważ mamy tutaj do czynienia z przekształceniem liniowym.

Odejmujemy aX i bY

Jeśli chcemy znaleźć E(X–Y) lub Var(X–Y), możemy skorzystać z poniższych formuł:

$$E(aX - bY) = aE(X) - bE(Y)$$

$$Var(aX - bY) = a^2Var(X) + b^2Var(Y)$$

Pamiętaj, że wariancje <u>dodajemy</u>.

Tak jak poprzednio, dodajemy wariancje obu zmiennych, choć szukamy wariancji ich różnicy.

Nie istnieją głupie pytania

P: Skoro X i Y oznaczają wypłaty w dwóch grach, to czy aX + bY oznacza łączną wygraną w a grach X i b grach Y?

O: Zapis aX + bY przedstawia właściwie sumę dwóch przekształceń liniowych zmiennych X i Y. Innymi słowy, zmianie ulegają zarówno wartości zmiennej X, jak i Y.

P: Nie wiem, czy kiedykolwiek przyda mi się znajomość rozkładu X − Y. Czy ma on w ogóle jakiś sens?

O: X–Y przydaje się wówczas, gdy interesują nas różnice między dwiema wielkościami. E(X−Y) daje odpowiedź na pytanie: „Jakiej różnicy między X i Y oczekujesz?”. Var(X−Y) informuje o wariancji takiej różnicy.

P: Dlaczego przy liczeniu Var(X − Y) musimy dodać wariancje zmiennych X i Y? Czy nie powinniśmy ich odjąć?

O: Choć z pozoru może się to wydawać sprzeczne z intuicją, to jednak odejmując od siebie dwie zmienne losowe, zwiększamy zmienność wartości zmiennej wynikowej, dlatego jej wariancja rośnie. Wariancja sumy dwóch zmiennych niezależnych jest taka sama jak wariancja ich różnicy.

Można to wyjaśnić jeszcze inaczej: przy liczeniu wariancji podnosimy do kwadratu wartości zmiennych losowych. Var(X + bY) jest równa Var(X) + b^2Var(Y). Podstawiając b = –1, otrzymamy szukaną Var(X−Y). Ponieważ $(-1)^2 = 1$, zatem Var(X−Y) = Var(X) + Var(Y).

P: Czy te same obliczenia możemy wykonywać dla zmiennych, które nie są niezależne?

O: Nie, formuły dla wariancji zachowują ważność tylko w przypadku zmiennych niezależnych. Jeśli chcesz wyznaczyć Var(X+Y) dla zmiennych, które są zależne, musisz najpierw wyznaczyć rozkład prawdopodobieństwa zmiennej X+Y.

P: Wygląda na to, że te same zasady, które stosują się do X+Y, są również prawdziwe w przypadku $X_1 + X_2$. Mam rację?

O: Oczywiście. Możesz je stosować do wszystkich zmiennych losowych, pod warunkiem że są one niezależne.

CELNE SPOSTRZEŻENIA

- Przez **niezależne obserwacje zmiennej X** rozumieć należy kolejne realizacje zmiennej X. Można je opisać oddzielnymi zmiennymi losowymi, z których każda ma ten sam rozkład prawdopodobieństwa co X.
- Jeśli $X_1, X_2, X_3, \ldots, X_n$ są zmiennymi niezależnymi o rozkładzie takim jak X, wówczas:

$$E(X_1 + X_2 + \ldots + X_n) = nE(X)$$
$$Var(X_1 + X_2 + \ldots X_n) = nVar(X)$$

- Jeśli X i Y są niezależnymi zmiennymi losowymi, prawdziwe są wzory:

$$E(X + Y) = E(X) + E(Y)$$
$$E(X - Y) = E(X) - E(Y)$$
$$Var(X + Y) = Var(X) + Var(Y)$$
$$Var(X - Y) = Var(X) + Var(Y)$$

- Wartość oczekiwaną i wariancję liniowych funkcji zmiennych losowych X i Y można wyznaczyć ze wzorów:

$$E(aX + bY) = aE(X) + bE(Y)$$
$$E(aX - bY) = aE(X) - bE(Y)$$
$$Var(aX + bY) = a^2Var(X) + b^2Var(Y)$$
$$Var(aX - bY) = a^2Var(X) + b^2Var(Y)$$

Ćwiczenie

W poniższej tabeli zamieszczono wartości oczekiwane i wariancje różnych zmiennych losowych. Spróbuj podać najprostszy sposób na wyznaczenie każdej z tych wartości. W razie potrzeby przyjmij założenie o niezależności zmiennych losowych.

Parametr	Sposób obliczenia
$E(aX + b)$	
$Var(aX + b)$	
$E(X)$	
$E(f(X))$	
$Var(aX - bY)$	
$Var(X)$	
$E(aX - bY)$	
$E(X_1 + X_2 + X_3)$	
$Var(X_1 + X_2 + X_3)$	
$E(X^2)$	
$Var(aX - b)$	

Ćwiczenie

Pewna restauracja oferuje dwa menu: jedno przeznaczone na dni robocze, drugie ważne w weekendy. Każde z nich zawiera potrawy w czterech różnych kategoriach cenowych. Rozkłady prawdopodobieństw wydatków klientów zamieszczono w poniższych tabelkach:

Dni robocze:

x	**10**	**15**	**20**	**25**
P(X = x)	0,2	0,5	0,2	0,1

Weekend:

y	**15**	**20**	**25**	**30**
P(Y = y)	0,15	0,6	0,2	0,05

Kto, Twoim zdaniem, wyda więcej pieniędzy na posiłek w tej restauracji: grupa 20 klientów weekendowych czy 25 klientów odwiedzających restaurację w pozostałe dni?

Ćwiczenie
Rozwiązanie

W poniższej tabeli zamieszczono wartości oczekiwane i wariancje różnych zmiennych losowych. Spróbuj podać najprostszy sposób na wyznaczenie każdej z tych wartości. W razie potrzeby przyjmij założenie o niezależności zmiennych losowych.

Parametr	Sposób obliczenia
$E(aX + b)$	$aE(X) + b$
$Var(aX + b)$	$a^2Var(X)$
$E(X)$	$\sum xP(X = x)$
$E(f(X))$	$\sum f(x)P(X = x)$
$Var(aX - bY)$	$a^2Var(X) + b^2Var(Y)$
$Var(X)$	$E(X - \mu)^2 = E(X^2) - \mu^2$
$E(aX - bY)$	$aE(X) - bE(Y)$
$E(X_1 + X_2 + X_3)$	$3E(X)$
$Var(X_1 + X_2 + X_3)$	$3Var(X)$
$E(X^2)$	$\sum x^2P(X = x)$
$Var(aX - b)$	$a^2Var(X)$

Ćwiczenie
Rozwiązanie

Pewna restauracja oferuje dwa menu: jedno przeznaczone na dni robocze, drugie ważne w weekendy. Każde z nich zawiera potrawy w czterech różnych kategoriach cenowych. Rozkłady prawdopodobieństw wydatków klientów zamieszczono w poniższych tabelkach:

Dni robocze:

x	10	15	20	25
P(X = x)	0,2	0,5	0,2	0,1

Weekend:

y	15	20	25	30
P(Y = y)	0,15	0,6	0,2	0,05

Kto, Twoim zdaniem, wyda więcej pieniędzy na posiłek w tej restauracji: grupa 20 klientów weekendowych czy 25 klientów odwiedzających restaurację w pozostałe dni?

Zacznijmy od wyznaczenia wartości oczekiwanej kwoty, jaką wyda przeciętny klient odwiedzający restaurację w weekend (Y) oraz w pozostałe dni (X).

$E(X) = 10 \times 0,2 + 15 \times 0,5 + 20 \times 0,2 + 25 \times 0,1 =$

$= 2 + 7,5 + 4 + 2,5 =$

$= 16$

$E(Y) = 15 \times 0,15 + 20 \times 0,6 + 25 \times 0,2 + 30 \times 0,05 =$

$= 2,25 + 12 + 5 + 1,5 =$

$= 20,75$

Wydatek każdego klienta można opisać za pomocą niezależnej obserwacji. Aby wyznaczyć łączne wydatki klientów zaliczonych do obu grup, pomnożymy E(X) i E(Y) przez liczebność każdej grupy.

25 klientów w dni robocze wyda razem: $25 \times E(X) = 25 \times 16 = 400$

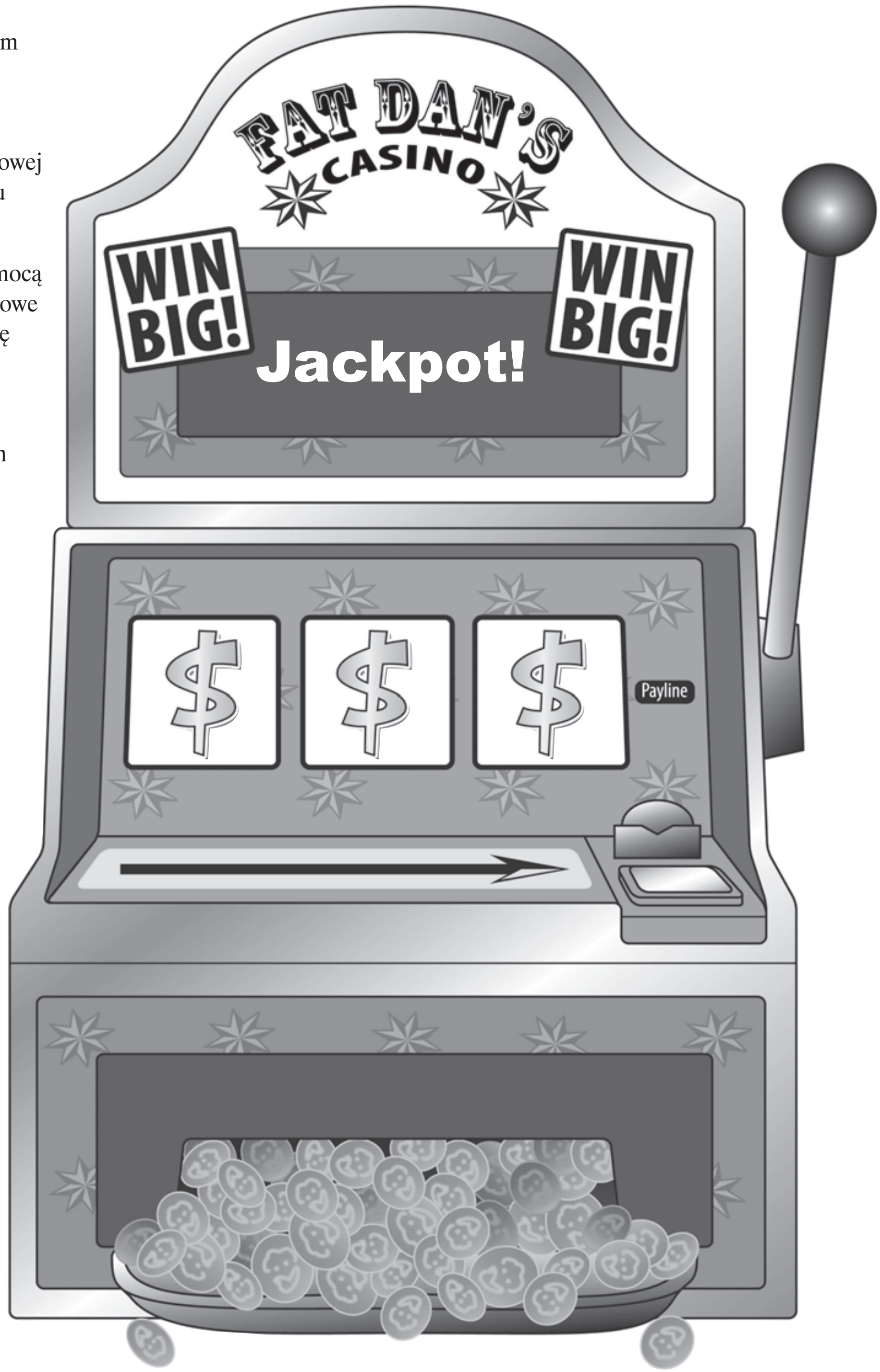
20 klientów weekendowych wyda zaś: $20 \times E(Y) = 20 \times 20,75 = 415$

Możemy więc oczekiwać, że 20 klientów odwiedzających restaurację w ciągu weekendu wyda więcej niż 25 klientów odwiedzających ją w tygodniu.

Rozbiłeś bank!

W tym rozdziale poruszyliśmy całkiem sporo zagadnień. Dowiedziałeś się, jak można posłużyć się rozkładem prawdopodobieństwa, wartością oczekiwaną i wariancją zmiennej losowej do przewidywania wygranej w starciu z „jednorękim bandytą”.

Dowiedziałeś się również, jak za pomocą przekształceń liniowych wyznaczyć nowe wartości parametrów, gdy zmienia się profil wypłaty w grze. Dzięki wiedzy o tym, czym są zmienne niezależne, umiesz już wyznaczać parametry rozkładu w przypadku gier złożonych z wielu kolejek.



Ćwiczenie

Sam zwykł jadać w dwóch restauracjach. Restauracja A jest droższa niż B, ale serwuje znacznie lepsze jakościowo dania.

Poniższe tabelki prezentują rozkłady prawdopodobieństw wydatków Sama w obu restauracjach. Jak mógłbyś scharakteryzować różnicę w poziomie cen między restauracją A i B? Ile wynosi jej wariancja?

Restauracja A:

x	**20**	**30**	**40**	**45**
P(X = x)	0,3	0,4	0,2	0,1

Restauracja B:

y	**10**	**15**	**18**
P(Y = y)	0,2	0,6	0,2

Ćwiczenie
Rozwiązanie

Sam zwykł jadać w dwóch restauracjach. Restauracja A jest droższa niż B, ale serwuje znacznie lepsze jakościowo dania.

Poniższe tabelki prezentują rozkłady prawdopodobieństw wydatków Sama w obu restauracjach. Jak mógłbyś scharakteryzować różnicę w poziomie cen między restauracją A i B? Ile wynosi jej wariancja?

Restauracja A:

x	**20**	**30**	**40**	**45**
P(X = x)	0,3	0,4	0,2	0,1

Restauracja B:

y	**10**	**15**	**18**
P(Y = y)	0,2	0,6	0,2

Wyznaczmy najpierw parametry rozkładów zmiennych X i Y:

$$E(X) = 20\times0{,}3 + 30\times0{,}4 + 40\times0{,}2 + 45\times0{,}1 =$$
$$= 6 + 12 + 8 + 4{,}5 =$$
$$= 30{,}5$$

$$Var(X) = (20-30{,}5)^2\times0{,}3 + (30-30{,}5)^2\times0{,}4 + (40-30{,}5)^2\times0{,}2 + (45-30{,}5)^2\times0{,}1 =$$
$$= (-10{,}5)^2\times0{,}3 + (-0{,}5)^2\times0{,}4 + 9{,}5^2\times0{,}2 + 14{,}5^2\times0{,}1 =$$
$$= 110{,}25\times0{,}3 + 0{,}25\times0{,}4 + 90{,}25\times0{,}2 + 210{,}25\times0{,}1 =$$
$$= 33{,}075 + 0{,}1 + 18{,}05 + 21{,}025 =$$
$$= 72{,}25$$

$$E(Y) = 10\times0{,}2 + 15\times0{,}6 + 18\times0{,}2 =$$
$$= 2 + 9 + 3{,}6 =$$
$$= 14{,}6$$

$$Var(Y) = (10-14{,}6)^2\times0{,}2 + (15-14{,}6)^2\times0{,}6 + (18-14{,}6)^2\times0{,}2 =$$
$$= (-4{,}6)^2\times0{,}2 + 0{,}4^2\times0{,}6 + 3{,}4^2\times0{,}2 =$$
$$= 21{,}16\times0{,}2 + 0{,}16\times0{,}6 + 11{,}56\times0{,}2 =$$
$$= 4{,}232 + 0{,}096 + 2{,}312 =$$
$$= 6{,}64$$

Różnicę między poziomami cen w obu restauracjach można przedstawić jako X–Y:

$$E(X-Y) = E(X) - E(Y) =$$
$$= 30{,}5 - 14{,}6 =$$
$$= 15{,}9$$

$$Var(X-Y) = Var(X) + Var(Y) =$$
$$= 72{,}25 + 6{,}64 =$$
$$= 78{,}89$$