

Wydaje Ci się, że **statystyka** jest trudna?
Z programem **Statistica** na pewno nie!

Statystyka

z programem

Statistica

Małgorzata Rabiej



Helion



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

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi bądź towarowymi ich właścicieli.

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

Redaktor prowadzący: Tomasz Waryszak

Projekt okładki: Maciej Pasek

Materiały graficzne na okładce zostały wykorzystane za zgodą Shutterstock.

Wydawnictwo HELION
ul. Kościuszki 1c, 44-100 GLIWICE
tel. 32 231 22 19, 32 230 98 63
e-mail: helion@helion.pl
WWW: <http://helion.pl> (księgarnia internetowa, katalog książek)

Drogi Czytelniku!
Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
<http://helion.pl/user/opinie?statis>
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

ISBN: 978-83-246-4110-9

Copyright © Helion 2012

Printed in Poland.

- [Kup książkę](#)
- [Poleć książkę](#)
- [Oceń książkę](#)

- [Księgarnia internetowa](#)
- [Lubię to! » Nasza społeczność](#)

Spis treści

Wprowadzenie	7
Rozdział 1. Organizacja pracy w programie Statistica	11
1.1. Interfejs programu. Zmienne i przypadki	11
Ćwiczenie 1.1.1. Wprowadzanie i zapisywanie danych	14
Ćwiczenie 1.1.2. Edycja danych w arkuszu	18
Ćwiczenie 1.1.3. Tworzenie wyrażeń matematycznych	24
Ćwiczenie 1.1.4. Sortowanie danych	25
Ćwiczenie 1.1.5. Filtrowanie danych	27
Ćwiczenie 1.1.6. Wzory matematyczne. Przeglądarka funkcji	28
Ćwiczenie 1.1.7. Zarządzanie wynikami	29
1.2. Graficzna prezentacja danych w programie Statistica	31
Ćwiczenie 1.2.1. Wykresy rozrzutu	31
Ćwiczenie 1.2.2. Wykresy słupkowe/kolumnowe	35
Ćwiczenie 1.2.3. Wykresy liniowe	38
Ćwiczenie 1.2.4. Wykres słupkowy wielokrotny	39
Ćwiczenie 1.2.5. Wykres 3W sekwencyjny	39
Ćwiczenie 1.2.6. Wykresy obrazkowe	42
Ćwiczenie 1.2.7. Wykres liniowy	45
Ćwiczenie 1.2.8. Wykres powierzchniowy i warstwicowy	47
Rozdział 2. Statystyka opisowa	51
Ćwiczenie 2.1.1. Szereg rozdzielczy. Histogramy	52
Ćwiczenie 2.1.2. Opisowe charakterystyki rozkładów	57
Ćwiczenie 2.1.3. Wykresy ramka-wąsy	60
Ćwiczenie 2.1.4. Analiza wielu zmiennych niezależnych	64
Ćwiczenie 2.1.5. Dane skategoryzowane	66
Zadania	70
Testy wielokrotnego wyboru	71
Rozdział 3. Zmienne losowe. Kalkulator prawdopodobieństwa	73
Ćwiczenie 3.1.1. Rozkład dwumianowy	83
Ćwiczenie 3.1.2. Standardowy rozkład normalny	85
Ćwiczenie 3.1.3. Rozkład t-Studenta	89
Ćwiczenie 3.1.4. Rozkład chi-kwadrat	93
Ćwiczenie 3.1.5. Rozkład średniej z próby	96
Ćwiczenie 3.1.6. Rozkład sumy zmiennych losowych	97
Zadania	97
Testy wielokrotnego wyboru	99

Rozdział 4. Estymacja parametryczna	103
Ćwiczenie 4.1.1. Przedział ufności dla wartości oczekiwanej	105
Ćwiczenie 4.1.2. Przedział ufności dla odchylenia standardowego	110
Zadania	111
Testy wielokrotnego wyboru	112
Rozdział 5. Testy statystyczne	115
5.1. Badanie normalności rozkładu zmiennych	118
Ćwiczenie 5.1.1. Testy normalności rozkładu	120
Ćwiczenie 5.1.2. Wykresy normalności	125
5.2. Testy jednorodności wariancji	126
Ćwiczenie 5.2.1. Test F	126
Ćwiczenie 5.2.2. Test Levene'a	128
Ćwiczenie 5.2.3. Test Browna-Forsythe'a	130
5.3. Testy t-Studenta	131
Ćwiczenie 5.3.1. Test t dla pojedynczej próby	139
Ćwiczenie 5.3.2. Test t dla dwóch prób niezależnych	141
Ćwiczenie 5.3.3. Testy istotności różnic dla dwóch prób zależnych	148
Ćwiczenie 5.3.4. Inne testy istotności	149
5.4. Testy nieparametryczne dla prób niezależnych	151
Ćwiczenie 5.4.1. Test U Manna-Whitneya	152
Ćwiczenie 5.4.2. Test serii Walda-Wolfowitza	154
5.5. Testy nieparametryczne dla prób zależnych	156
Ćwiczenie 5.5.1. Test znaków	156
Ćwiczenie 5.5.2. Test kolejności par Wilcoxon'a	157
Zadania	158
Testy wielokrotnego wyboru	159
Rozdział 6. Porównanie wielu średnich	165
Ćwiczenie 6.1.1. ANOVA jednoczynnikowa. Przekroje, prosta ANOVA	166
Ćwiczenie 6.1.2. ANOVA jednoczynnikowa. Moduł ANOVA	173
Ćwiczenie 6.1.3. Wielokrotne porównania	177
Ćwiczenie 6.1.4. ANOVA efektów głównych	181
Ćwiczenie 6.1.5. ANOVA dla układów czynnikowych	183
Ćwiczenie 6.1.6. Porównania zaplanowane	186
Ćwiczenie 6.1.7. Test Kruskala-Wallisa i test mediany	195
Ćwiczenie 6.1.8. Układy z powtarzanymi pomiarami	197
Zadania	205
Testy wielokrotnego wyboru	206
Rozdział 7. Analiza zmiennych jakościowych	209
Ćwiczenie 7.1.1. Test McNemary	209
Ćwiczenie 7.1.2. Test Q Cochraha	211
Ćwiczenie 7.1.3. Tabele wielodzidelcze. Test niezależności χ^2	213
Rozdział 8. Analiza współzależności między zmiennymi	217
8.1. Regresja liniowa	217
Ćwiczenie 8.1.1. Badanie korelacji	219
Ćwiczenie 8.1.2. Regresja liniowa	221
8.2. Regresja wieloraka	229
Ćwiczenie 8.2.1. Liniowy model regresji wielorakiej	231
Ćwiczenie 8.2.2. Predykcja zmiennej zależnej	243
Ćwiczenie 8.2.3. Regresja krokowa	244

8.3.	Linearyzowana regresja nieliniowa	250
	Ćwiczenie 8.3.1. Logarytmiczna funkcja regresji	251
	Ćwiczenie 8.3.2. Wykładnicza funkcja regresji	257
	Ćwiczenie 8.3.3. Hiperboliczna funkcja regresji	259
	Ćwiczenie 8.3.4. Aproksymacja wielomianem drugiego stopnia	263
8.4.	Estymacja nieliniowa	266
	Ćwiczenie 8.4.1. Funkcja użytkownika	266
	Ćwiczenie 8.4.2. Regresja logistyczna	271
	Zadania	277
	Testy wielokrotnego wyboru	278
Rozdział 9.	Szeregi czasowe. Metody prognozowania	281
	Ćwiczenie 9.1.1. Prognozowanie metodą średniej ruchomej	284
	Ćwiczenie 9.1.2. Wygładzanie wykładnicze	295
	Ćwiczenie 9.1.3. Model Holta	301
	Ćwiczenie 9.1.4. Model trendu liniowego	308
	Ćwiczenie 9.1.5. Metoda wskaźników. Dekompozycja sezonowa (Census 1)	312
	Ćwiczenie 9.1.6. Model ARIMA dla pojedynczego szeregu	326
	Zadania	333
	Testy wielokrotnego wyboru	334
	Odpowiedzi do testów	337
	Bibliografia	339
	Skorowidz	341

Rozdział 4.

Estymacja parametryczna

Głównym zadaniem badań statystycznych jest wnioskowanie o całej populacji generalnej na podstawie wyników uzyskanych w próbie losowej. Dział statystyki zajmujący się tym zagadnieniem jest nazywany **wnioskowaniem statystycznym**.

Estymacja to dział wnioskowania statystycznego, który zajmuje się szacowaniem wartości parametrów oraz postaci rozkładu w populacji generalnej na podstawie obserwacji uzyskanych w próbie losowej. Metody znajdowania nieznanymi wartościami parametrów rozkładu określa **estymacja parametryczna**. Wnioskowaniem o postaci rozkładu w populacji generalnej zajmuje się **estymacja nieparametryczna**. Punktem wyjściowym w estymacji jest wylosowanie z populacji n -elementowej próby i wyznaczenie na jej podstawie wartości estymatora nieznanego parametru. **Estymatorem** parametru θ rozkładu populacji generalnej jest funkcja wyznaczona na podstawie próby losowej, służąca do oceny wartości tego parametru. Teoria estymacji zajmuje się konstruowaniem estymatorów mających określone właściwości, takie jak nieobciążoność, zgodność, efektywność i dostateczność. Więcej o metodach wyznaczania takich estymatorów można znaleźć w pozycjach [4, 21, 27, 30]. Zgodnym, nieobciążonym i najefektywniejszym estymatorem wartości oczekiwanej populacji jest wartość średnia $\bar{x}$ z próby losowej wyrażona wzorem $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Zgodnym i nieobciążonym estymatorem wariancji populacji σ^2 jest wariancja z próby prostej wyrażona wzorem

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Estymacja parametryczna może być punktowa lub przedziałowa. **W estymacji punktowej** za parametr populacji przyjmuje się wartość estymatora otrzymaną z danej, n -elementowej próby losowej. Estymacja punktowa nie daje oszacowania nieznanego parametru θ rozkładu populacji. Prawdopodobieństwo, że estymator przyjmie wartość równą wartości szacowanego parametru, jest równe 0. Z tego wynika, że przy stosowaniu estymacji punktowej prawdopodobieństwo popełnienia błędu w ocenie parametru populacji jest równe 1.

Błąd oceny parametru populacji θ za pomocą jego estymatora Q nie powinien przekraczać odpowiednio małej wartości ε z przyjętym dużym prawdopodobieństwem $1-\alpha$, czyli musi być spełnione równanie: $P(|\theta - Q| < \varepsilon) = 1 - \alpha$. Przedział liczbowy $(Q-\varepsilon, Q+\varepsilon)$, który z określonym z góry, dużym (bliskim jedności) prawdopodobieństwem będzie zawierał nieznaną wartość parametru zbiorowości generalnej, jest nazywany **przedziałem ufności**, a prawdopodobieństwo $1-\alpha$ — **współczynnikiem ufności**. Do wyznaczenia wartości ε potrzebna jest znajomość rozkładu estymatora Q . Procedura wyznaczania przedziału ufności jest nazywana **estymacją przedziałową**.

Przedział ufności to losowy przedział wyznaczony za pomocą rozkładu estymatora, mający tę własność, że z dużym, z góry zadanyim prawdopodobieństwem pokrywa wartość szacowanego parametru. Zapisujemy go zwykle w postaci: $P(a < \theta < b) = 1 - \alpha$. Liczby a i b są nazywane **dolną** i **górną granicą** przedziału ufności. Współczynnik ufności $1-\alpha$ jest miarą zaufania do prawidłowego szacunku. Najczęściej ma on wartość 0,99, 0,95 lub 0,90.

Przedział ufności dla wartości oczekiwanej $E(X)$ populacji o rozkładzie normalnym $N(m, \sigma)$ jest wyznaczany według wzoru:

$$P\left(\bar{x} - u_{\alpha} \frac{\sigma}{\sqrt{n}} < m < \bar{x} + u_{\alpha} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha,$$

gdzie u_{α} jest taką wartością w standardowym rozkładzie, że pole pod krzywą gęstości w przedziale $(-u_{\alpha}, u_{\alpha})$ wynosi $1-\alpha$, a pole pod krzywą gęstości na prawo od u_{α} i na lewo od $-u_{\alpha}$ wynosi po $\alpha/2$. Z tego wynika, że u_{α} można wyznaczyć z relacji:

$\Phi(u_{\alpha}) = 1 - (\alpha/2)$, gdzie Φ jest dystrybuantą standardowego rozkładu normalnego.

Długość przedziału ufności zależy od wartości średniej, obliczonej na podstawie próby, przyjętego współczynnika ufności $1-\alpha$, liczebności próby oraz wariancji σ . Aby zatem oszacować przedział ufności z jak najmniejszym błędem, należy dokładnie określić wartość średnią.

Przedział ufności dla wartości oczekiwanej dla małych prób oblicza się według wzoru:

$$P\left(\bar{x} - t_{\alpha} \frac{s}{\sqrt{n-1}} < m < \bar{x} + t_{\alpha} \frac{s}{\sqrt{n-1}}\right) = 1 - \alpha,$$

gdzie t_{α} wartość zmiennej losowej t-Studenta dla $n-1$ stopni swobody wyznaczana z relacji: $P(-t_{\alpha} < t < t_{\alpha}) = 1 - \alpha$.

Im wartość współczynnika ufności jest większa, tym szerszy jest przedział ufności, a więc mniejsza dokładność estymacji parametru. Długość przedziału ufności jest miarą **precyzji estymacji przedziałowej**. Szeroki przedział ufności oznacza możliwość dużych odchyień wartości z próby od wartości rzeczywistych, czyli wartości oczekiwanych z populacji. Im krótszy jest przedział ufności, tym dokładniej obliczony przez nas estymator przybliży wartość oczekiwaną populacji, czyli tym precyzyjniejsza jest estymacja

przedziałowa. Błędy przybliżeń popełniane przy szacowaniu średniej maleją wraz ze zwiększaniem liczebności próby. Jednym z zadań estymacji jest wyznaczenie minimalnej liczebności próby tak, by oszacować przedział ufności z jak najmniejszym błędem. Zbyt mała próba może prowadzić do fałszywych wniosków o populacji generalnej. Aby zwiększyć dokładność estymacji, należy także poprawić dokładność pomiarów.

Przedziały ufności są wyznaczane dla wartości oczekiwanej, wariancji, odchylenia standardowego i wskaźnika struktury. Wyznacza się je z rozkładów odpowiednich statystyk będących estymatorami tych parametrów.

Ćwiczenie 4.1.1. Przedział ufności dla wartości oczekiwanej

Przykład 34. Przedział ufności dla średniej (wartości oczekiwanej $m = E(X)$) dla dużych prób

Dokonano 52 pomiarów zanieczyszczenia gleby ołowiem (w mg/kg suchej masy gleby), otrzymane wyniki zapisano w tabeli.

59	60	62	58	59	61	65	67	65	65	62	62	65	67	69	64	65	66	67	64	64	66	68	64	63	64
67	69	62	64	67	68	69	61	62	69	66	69	63	65	60	60	65	63	70	68	67	71	61	64	63	66

Zakładając, że rozkład wyników pomiarów jest rozkładem normalnym, wyznacz przedział ufności ze współczynnikiem ufności 0,95 dla wartości średniej.

Dane

Współczynnik ufności $1-\alpha = 0,95$. Zmienną jest zanieczyszczenie gleby ołowiem.

Rozwiązanie

- ♦ Wybierz z menu *Plik/Nowy*. W oknie *Utwórz nowy dokument* wprowadź:
Liczba zmiennych: 1, *Liczba przypadków*: 52.
- ♦ Wprowadź dane z tabeli.
- ♦ Zapisz arkusz w pliku *olów.sta*.
- ♦ Ponieważ próba jest duża, można przyjąć, że $\sigma = s$. Przedział ufności dla dużych prób obliczany jest według wzoru:

$$P\left(\bar{X} - u_{\alpha} \frac{\sigma}{\sqrt{n}} < m < \bar{X} + u_{\alpha} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

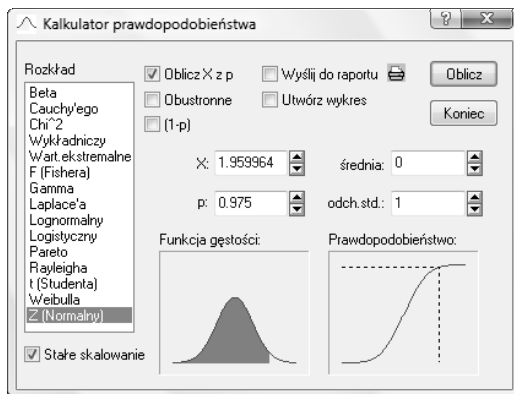
Dane jest $1-\alpha = 0,95$, czyli $\alpha = 0,05$. u_{α} należy wyznaczyć z relacji $\Phi(u_{\alpha}) = 1-(\alpha/2)$, gdzie Φ jest dystrybucją standardowego rozkładu normalnego.

Po podstawieniu $\Phi(u_{\alpha}) = 1-(\alpha/2) = 0,975$.

- ◆ Uruchom kalkulator prawdopodobieństwa. Wybierz *Rozkład Z (Normalny)*.
- ◆ Wprowadź $p = 0,975$ (rysunek 4.1). Kliknij przycisk *Oblicz*. Program oblicza u_α wyświetlane w polu X , czyli $u_\alpha = 1,96$.

Rysunek 4.1.

Okno Kalkulator
prawdopodobieństwa



- ◆ Aby wyznaczyć wartość średnią i odchylenie standardowe, kliknij lewym przyciskiem nazwę zmiennej, wybierz *Statystyki bloku danych/Kolumny/Średnia*, a następnie *Odchylenie standardowe*.
- ◆ Program wyświetla wyniki $\bar{x} = 64,615$, $\sigma = 3,13$.

Po podstawieniu do wzoru otrzymujemy przedział ufności (63,7, 65,5).

Rozwiązanie z programem Statistica

I sposób

- ◆ Wybierz z menu: *Statystyka/Statystyki podstawowe i tabele/Statystyki opisowe*.
- ◆ Kliknij przycisk *Zmienne* i jako zmienną wprowadź 01ów.
- ◆ Aby wyświetlić przedziały ufności, kliknij zakładkę *Więcej* i zaznacz parametry: *Średnia, Przedz. ufn. średniej*. W polu *Przedział [%]* podany jest współczynnik ufności równy 95% (rysunek 4.2).
- ◆ Kliknij przycisk *Statystyki* lub *Podsumowanie*. Program wyświetla arkusz wynikowy w postaci tabeli (rysunek 4.3).

Odpowiedź

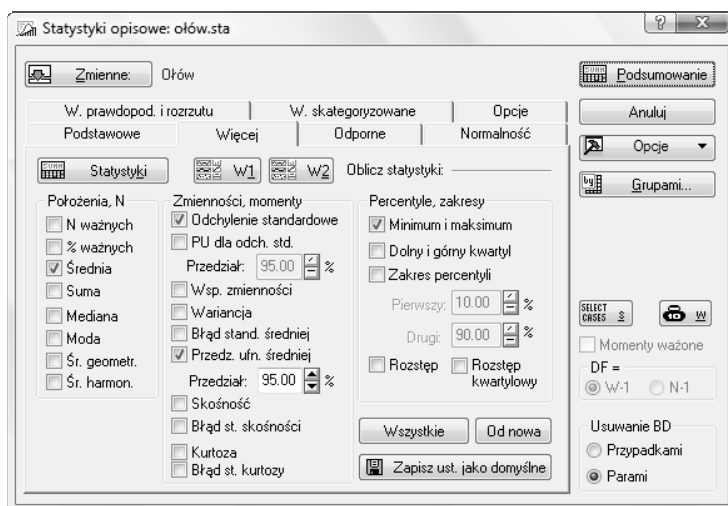
Przedział ufności (w programie Statistica: $(Ufność-95\%) = 63,7$; $(Ufność+95\%) = 65,5$) ma postać (63,7, 65,5).

II sposób

Przedział ufności jest wyświetlany na wykresie *Średnia i błędy*.

- ◆ Wybierz z menu: *Wykresy/Wykresy średnia i błędy*.
- ◆ Zdefiniuj zmienną 01ów.

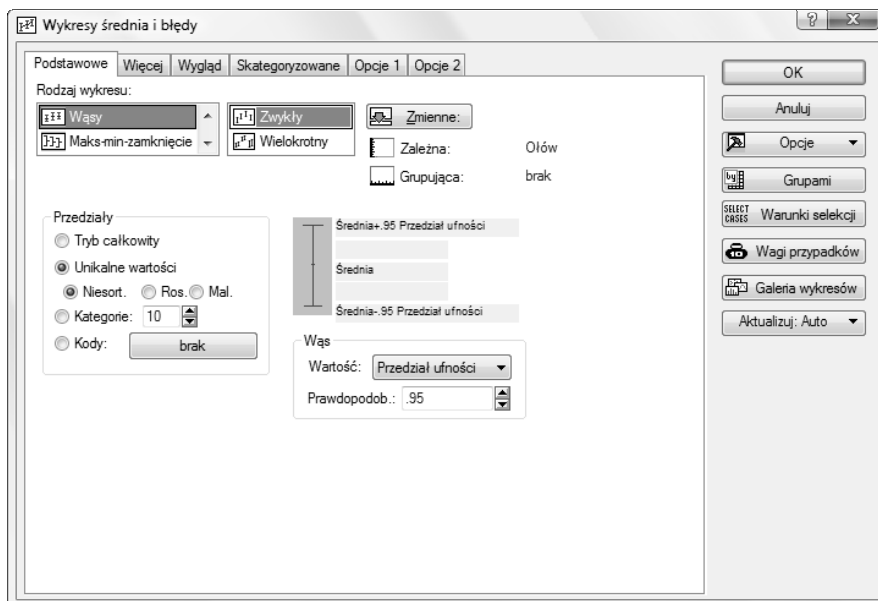
Rysunek 4.2.
Okno wyboru
statystyk opisowych



Zmienna	Statystyki opisowe (ołów.sta)				
	Średnia	Ufność -95,000%	Ufność 95,000%	Minimum	Maksimum
Ołów	64,6	63,7	65,5	58,0	71,0

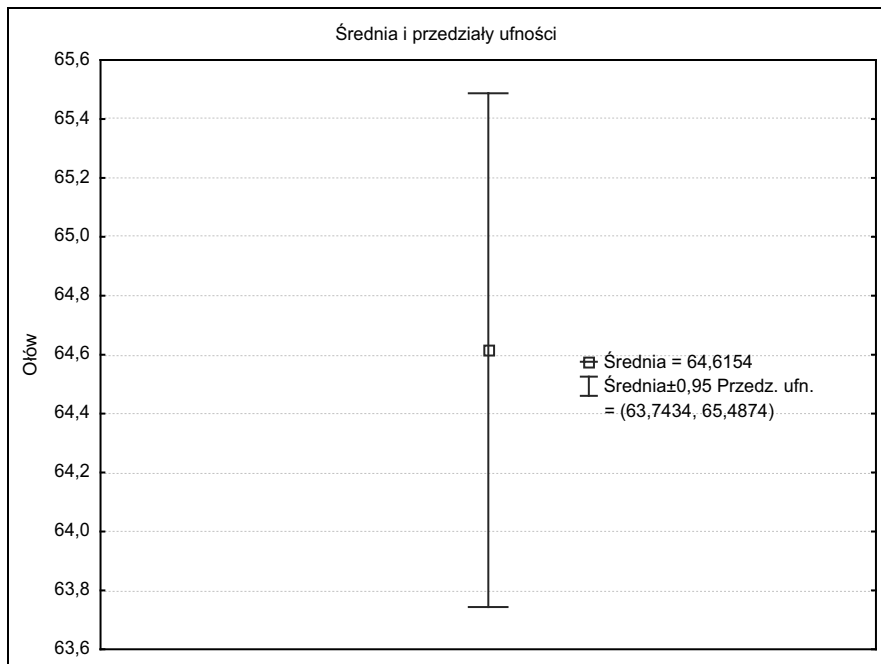
Rysunek 4.3. Arkusz z wynikami obliczeń (Ufność-95%, Ufność+95% to granice przedziału ufności dla współczynnika $1-\alpha = 0,95$)

- ♦ Współczynnik ufności jest wyświetlany w polu *Prawdopodob.* (domyślna wartość tego współczynnika wynosi 0,95) (rysunek 4.4). Kliknij OK.



Rysunek 4.4. Okno tworzenia wykresów średniej i przedziałów ufności

- ♦ Program tworzy wykres średniej i przedziałów ufności dla tej średniej (rysunek 4.5).



Rysunek 4.5. Wykres średniej i przedziałów ufności dla tej średniej

Odpowiedź

Przedział ufności jest wyświetlany na wykresie. Z prawdopodobieństwem 0,95 można twierdzić, że średnie zanieczyszczenie gleby ołowiem zawiera się w przedziale (63,7, 65,5) mg/kg suchej masy.

Przykład 35. Przedział ufności dla średniej dla małych prób

Dokonano 12 pomiarów zanieczyszczenia gleby ołowiem (w mg/kg suchej masy gleby), otrzymane wyniki zapisano w tabeli.

54	60	65	55	70	68	67	59	61	64	63	68
----	----	----	----	----	----	----	----	----	----	----	----

Zakładając, że rozkład zmiennej, czyli zanieczyszczenia gleby ołowiem, jest rozkładem normalnym, i przyjmując współczynnik ufności 0,95, wyznacz przedział ufności dla średniej wartości zanieczyszczenia gleby ołowiem.

Dane

$$1-\alpha = 0,95, n = 12$$

Rozwiązanie

Przedział ufności dla małych prób oblicza się według wzoru:

$$P\left(\bar{X} - t_{\alpha} \frac{s}{\sqrt{n-1}} < m < \bar{X} + t_{\alpha} \frac{s}{\sqrt{n-1}}\right) = 1 - \alpha,$$

gdzie t_{α} wartość zmiennej losowej t-Studenta dla $n-1$ stopni swobody jest wyznaczana tak, że spełniona jest relacja $P(-t_{\alpha} < t < t_{\alpha}) = 1 - \alpha$.

- ♦ Wybierz z menu *Plik/Nowy*. W oknie *Utwórz nowy dokument* wprowadź: *Liczba zmiennych*: 1, *Liczba przypadków*: 12.
- ♦ Wprowadź dane z tabeli i zachowaj w pliku *zanieczyszczenie ołowiem.sta*.
- ♦ Wybierz z menu: *Statystyka/Statystyki podstawowe i tabele/Statystyki opisowe*.
- ♦ Kliknij przycisk *Zmienne* i jako zmienną wprowadź *Ołów*.
- ♦ Kliknij zakładkę *Więcej* i zaznacz pola wyboru: *średnia*, *Przedz. ufn. średniej*.
- ♦ Pole edycji *Przedział* zawiera domyślny współczynnik ufności (0,95) podawany w procentach.

Program wyświetla tabelę z przedziałami ufności (rysunek 4.6).

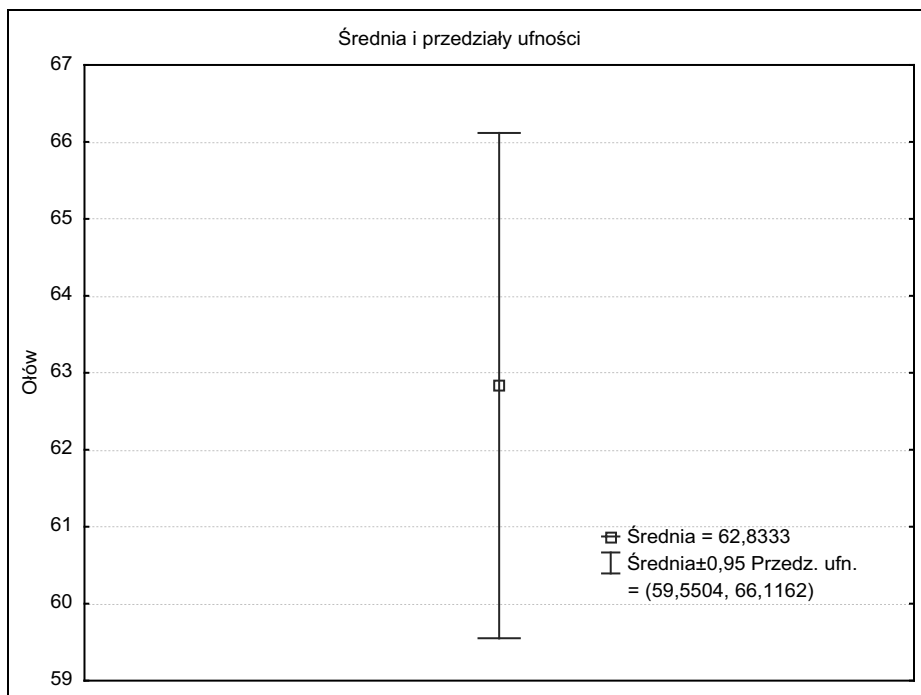
Zmienna	Statystyki opisowe (zanieczyszczenie ołowiem.sta)		
	Średnia	Ufność -95,000%	Ufność 95,000
Ołów	62,83	59,55	66,12

Rysunek 4.6. Arkusz z wynikami obliczeń

- ♦ Wybierz z menu: *Wykresy/Wykresy średnia i błędy*.
- ♦ Zdefiniuj zmienną *Ołów*.
- ♦ Współczynnik ufności jest wyświetlany w polu *Prawdopodob.* (domyślna wartość tego współczynnika wynosi 0,95). Kliknij *OK*.
- ♦ Program tworzy wykres średniej i wyświetla na wykresie przedziały ufności dla tej średniej (rysunek 4.7).

Odpowiedź

Z prawdopodobieństwem 0,95 można twierdzić, że zanieczyszczenie ołowiem zawiera się w przedziale (59,6 mg/kg, 66,1 mg/kg).



Rysunek 4.7. Wykres średniej i przedziałów ufności

Ćwiczenie 4.1.2. Przedział ufności dla odchylenia standardowego

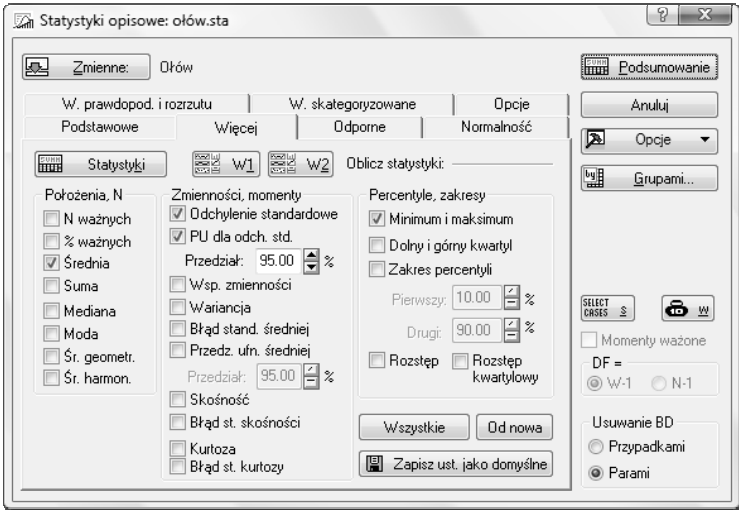
Przykład 36. Przedział ufności dla odchylenia standardowego

Przyjmując współczynnik ufności 0,98, wyznacz przedziały ufności dla odchylenia standardowego dla danych z poprzedniego przykładu (plik *zanieczyszczenie ołowiem.sta*).

Sposób wykonania

- ◆ Otwórz plik *zanieczyszczenie ołowiem.sta*.
- ◆ Wybierz z menu: *Statystyka/Statystyki podstawowe i tabele/Statystyki opisowe*.
- ◆ Kliknij przycisk *Zmienne* i jako zmienną wprowadź *Ołów*.
- ◆ Kliknij zakładkę *Więcej* i zaznacz pola wyboru: *Odchylenie standardowe, PU dla odch. std.* (rysunek 4.8).
- ◆ Pole edycji *Przedział* zawiera współczynnik ufności podawany w procentach. Wprowadź 98.
- ◆ Program tworzy arkusz z wynikami (rysunek 4.9).

Rysunek 4.8.
*Okno wyboru
statystyk opisowych*



Statystyki opisowe (zanieczyszczenie ołowiem.sta)					
	N ważnych	Średnia	Odch.std	P. ufności odch. std. -98,000%	P. ufności odch. std. +98,000%
Zmienna					
Ołów	12	62,8	5,2	3,4	9,8

Rysunek 4.9. *Arkusz z wynikami statystyk opisowych*

Odpowiedź

Otrzymany przedział (3,4, 9,8) z prawdopodobieństwem 0,98 pokrywa odchylenie standardowe zanieczyszczenia gleby ołowiem.

Przy zmniejszaniu wartości współczynnika ufności maleje długość przedziału ufności.

Można sprawdzić, że przedział (3,7, 8,8) z prawdopodobieństwem 0,95 pokrywa wartość odchylenia standardowego.

Zadania

Zadanie 1.

Przeprowadzono badanie stężenia azotynów NNO_2 (mg/l) w wodzie na dwóch odcinkach rzeki. Wyniki pomiarów zawiera tabela.

Odcinek 1	12	16	14	15	13	17	13	13	15	14	16	12
Odcinek 2	11	9	13	15	12	13	11	10	15	13	14	12

Wyznacz przedziały ufności dla średniej dla każdej grupy. Porównaj wyniki w grupach.

Zadanie 2.

W celu ustalenia stopnia krystaliczności pewnego polimeru przeprowadzono pomiary tego parametru dla 10 próbek i otrzymano wyniki (w %): 61, 57, 63, 62, 59, 60, 58, 62, 59, 61. Zbuduj przedziały ufności, które z prawdopodobieństwem 98% pokryją średnią wartość tego parametru.

Zadanie 3.

W celu ustalenia temperatury topnienia pewnego polimeru przeprowadzono pomiary tego parametru dla ośmiu próbek. Otrzymano (w °C): 220, 225, 223, 226, 224, 225, 223, 221. Przy zadanym współczynniku ufności 0,99 wyznacz przedział ufności dla temperatury topnienia.

Zadanie 4.

W celu ustalenia stężenia ozonu występującego przy powierzchni Ziemi przeprowadzono pomiary i otrzymano wyniki (w ppb): 61, 57, 63, 62, 59. Przyjmując współczynnik ufności 0,98:

- ♦ Zbuduj przedział ufności, który z prawdopodobieństwem 98% pokryje średnią wartość tego parametru.
- ♦ Wyznacz przedział ufności dla odchylenia standardowego stężenia ozonu.

Zadanie 5.

Dokładny pomiar odczynu pH dla oznaczenia stanu surowych ścieków jest ważny przy sterowaniu dozowaniem chemikaliów w procesie neutralizacji. Wykonano 60 pomiarów tego parametru i otrzymano: $\bar{x} = 7,9$ (°pH) oraz odchylenie standardowe 0,8.

Zbuduj przedział, który z prawdopodobieństwem 0,99 pokryje pH ścieków.

Zadanie 6.

Biologiczne zapotrzebowanie tlenu Q (mg O_2/l) określa wskaźnik BZT_5 . Wykonano 12 pomiarów tego wskaźnika i otrzymano wyniki widoczne w tabeli.

BZT_5	41	39	42	40	38	43	39	44	37	40	39	38
---------	----	----	----	----	----	----	----	----	----	----	----	----

Przyjmując współczynnik ufności 0,95, zbuduj przedział ufności dla nieznannej średniej wskaźnika BZT_5 .

Testy wielokrotnego wyboru

1. Estymacja parametryczna

- a) dotyczy szacowania wartości parametrów rozkładu populacji generalnej.
 - b) polega na oszacowaniu nieznanego rozkładu zmiennej losowej.
 - c) polega na szacowaniu wartości lub przedziału pokrywającego z pewnym prawdopodobieństwem parametr populacji generalnej.
2. Zmienna losowa X populacji generalnej ma rozkład normalny o nieznannej wartości średniej oraz nieznanym odchyleniu standardowym. Z populacji pobrano małą próbę.
- a) Przedział ufności dla wariancji wyznaczonej na podstawie tej próby jest oparty na rozkładzie chi-kwadrat.
 - b) Przedział ufności dla odchylenia standardowego zmiennej losowej X jest oparty na rozkładzie F Snedecora.
 - c) Przedział ufności dla wartości średniej wyznaczonej na podstawie tej próby jest oparty na rozkładzie t-Studenta o $n-1$ stopniach swobody.

3. Losowy przedział wyznaczony za pomocą rozkładu estymatora, mający tę własność, że z dużym, z góry zadany prawdopodobieństwem pokrywa wartość szacowanego parametru, jest nazywany
 - a) przedziałem dopuszczalnym.
 - b) przedziałem ufności.
 - c) przedziałem krytycznym.
4. Współczynnik ufności to
 - a) prawdopodobieństwo odrzucenia prawdziwej hipotezy zerowej.
 - b) z góry zadane duże prawdopodobieństwo równe $1-\alpha$.
 - c) z góry zadane małe prawdopodobieństwo równe α .
5. Wyznaczono przedział ufności dla wartości średniej, przyjmując wartość współczynnika ufności 0,95. Oznacza to, że
 - a) $\alpha = 0,95$.
 - b) $1-\alpha = 0,95$.
 - c) wyznaczony przedział z ufnością 0,05 pokrywa nieznaną wartość średniej.
6. Aby zwiększyć precyzję estymacji przedziałowej, należy
 - a) zwiększyć liczebność próby.
 - b) zwiększyć współczynnik ufności.
 - c) zmniejszyć współczynnik ufności.
7. Wybierz poprawne stwierdzenia.
 - a) Im wyższa jest wartość współczynnika ufności, tym szerszy jest przedział ufności.
 - b) Im niższa jest wartość współczynnika ufności, tym szerszy jest przedział ufności.
 - c) Długość przedziału ufności nie zależy od współczynnika ufności.

Skorowidz

.spf, 14
.sta, 14
.stw, 14

A

addytywny, model, 284
analiza
 kontrastów, 186, 189, 190
 log-liniowa, 213
 zmiennych jakościowych, 209
analiza wariancji, 165
 dwuczynnikowa, 181
 jednoczynnikowa, 165, 166, 190
ANOVA, 165, 169, 174
 dla układów czynnikowych, 165, 183
 efektów głównych, 165, 181
 jednoczynnikowa, 165, 166, 167, 168, 173
 moduł, 173
 założenia testów, 167
ARIMA, model, 326, 327
arkusz, 13, 14
Arkusz, zakładka, 14

B

badanie statystyczne, 51
błąd drugiego rodzaju, 115
błąd pierwszego rodzaju, 115
błąd standardowy, 81

C

cechy statystyczne, 51
Census 1, 312, 319
centralne twierdzenie graniczne, 80
czynnik pomiarów powtarzanych, 197

D

dane
 edycja, 18
 filtrowanie, 27
 graficzna prezentacja, 31
 skategoryzowane, 66
 sortowanie, 25, 26
 wprowadzanie, 14, 15
 zapisywanie, 14, 17
decyle, 78
dekompozycja sezonowa, 312, 319
dokument, tworzenie, 14, 15
dominanta, 58
Durbina-Watsona, statystyka, 231
dystrybuanta, 77
 zmiennej losowej ciągłej, 77
 zmiennej losowej skokowej, 77

E

estymacja, 103
 nieliniowa, 266
 nieparametryczna, 103
 parametryczna, 103
 przedziałowa, 104
 punktowa, 103
estymator, 103
Eta-kwadrat cząstkowe, 176
ex ante, 282
ex post, 282

F

formuły matematyczne, tworzenie, 24
funkcja
 gęstości rozkładu prawdopodobieństwa, 76
 prawdopodobieństwa, 76
 rozkładu prawdopodobieństwa, 76
funkcje matematyczne, definiowanie, 28

H

heteroscedantyczność, 235
 hipoteza, 115
 alternatywna, 115
 zerowa, 115
 histogram, 52, 54
 Holta, model, 282, 301
 homoscedantyczność, 235

I

iloczyn zdarzeń, 74
 interfejs graficzny, 11

J

jednostki statystyczne, 51

K

Kalkulator prawdopodobieństwa, 82, 86, 89
 klucz sortowania, 25
 kod braku danych, 24
 kolumna
 zaznaczanie, 18
 zmiana szerokości, 18
 komórki
 formatowanie, 20, 21
 kopiowanie zawartości, 20
 przenoszenie zawartości, 19
 wypełnianie serią danych, 23, 24
 zaznaczanie, 18
 kontrast, 186
 korelacja, 217
 badanie, 219
 cząstkowa, 230
 semicząstkowa, 230
 korelogram, 326
 kurtoza, 59
 kwantyl, 78
 rzędu p, 78
 kwartył, 78
 dolny, 58
 górny, 58
 pierwszy, 58
 trzeci, 58

L

logit, 271, 272

M

matematyczne formuły, tworzenie, 24
 mediana, 58, 78
 metoda
 średniej ruchomej, 284
 wskaźników, 312
 miary
 asymetrii, 58, 59
 koncentracji, 58, 59
 położenia, 57, 58
 rozproszenia, 58
 moc testu, 117
 mocne prawo wielkich liczb, 78
 moda, 58
 model
 addytywny, 284
 ARIMA, 326, 327
 Holta, 282, 301
 multiplikatywny, 284
 trendu liniowego, 308
 Wintersa, 282
 multiplikatywny, model, 284

N

niecentralność, 176

O

obserwacja statystyczna, 51
 obserwacje odstające, 80
 odchylenie ćwiartkowe, 58
 odchylenie standardowe, 58
 okno
 ANOVA, 165, 168, 178, 197
 edytora wykresu, 34
 Inne testy istotności, 149
 Kalkulator prawdopodobieństwa, 82, 84, 86, 90, 91, 151
 klasycznej dekompozycji sezonowej, 320
 Kryteria autofiltera, 27
 podstawowe, 11, 12
 Przeglądarka funkcji, 29
 raportu, 30
 Solver, 296
 sortowania, 26
 testu t dla pojedynczej próby, 140
 tworzenia wykresów, 33
 tworzenie nowego dokumentu, 15
 wyboru statystyk nieparametrycznych, 152
 wyboru statystyk opisowych, 60
 wyboru testu, 137
 wyboru typu rozkładu, 119
 wykresy sekwencyjne 3W, 41

P

populacja generalna, 51
porównywanie zaplanowane, 186
poziom istotności, 115
prawdopodobieństwo
całkowite, 75
definicja, 74
warunkowe, 75
prawo wielkich liczb Bernoulliego, 75
precyzja estymacji przedziałowej, 104
predyktory jakościowe, 165
prezentowanie danych, 31
prognoza dopuszczalna, 283
prognoza wygasła, 282
prognozowanie, 281, 282
błędy, 283
metodą średniej ruchomej, 284, 285
model Holta, 282, 301
model trendu liniowego, 308
model Wintersa, 282
wygładzanie wykładnicze, 295, 301
próba, 51
liczebność, 51
losowa, 51
losowa prosta, 75
reprezentatywna, 51
próbka, 51
przedziały ufności, 104, 105
dla odchylenia standardowego, 110
dla wartości oczekiwanej, 105
Przeglądarka funkcji, 28, 29
przestrzeń zdarzeń elementarnych, 73
przypadki
dodawanie, 18, 19
formatowanie nazw, 21, 23
selekcja, 24
punkty
ekstremalne, 62
odstające, 62, 125

R

raport, 30
tworzenie, 30
zapisywanie, 30
regresja, 217
częstkowe współczynniki, 229
hiperboliczna, 259
krokowa, 244, 245
liniowa, 217, 221
logarytmiczna, 251, 252
logistyczna, 271
nieliniowa, 250
wielokrotna, 229

wieloraka, 229, 231, 238
wykładnicza, 257
reguła trzech sigm, 80
reszty, 218
analiza, 237
warunki, 218
rozkład
asymetryczny, 59
bimodalny, 52
chi-kwadrat, 82, 93
dwumianowy, 78, 83
empiryczny, 52
F Fishera, 82
jednomodalny, 52
normalny, 79, 85
opisowe charakterystyki, 57
Poissona, 79
siodłowy, 52
sumy zmiennych losowych, 97
symetryczny, 59
średniej z próby, 96
t-Studenta, 81, 89
zmiennej losowej, 75
rozstęp, 58
kwartylowy, 58
różnica zdarzeń, 74

S

sferyczność, 198
skoroszyty, 14
słabe prawo wielkich liczb, 78
standardowy rozkład normalny, 79, 85
standaryzacja zmiennej, 39, 79
Statistica, 7, 9
statystyka matematyczna, 73
statystyka opisowa, 51, 52, 57, 59, 73
STR, format, 30
suma zdarzeń, 73
symetria połączona, 198
szereg czasowy, 281, 282
szereg rozdzielczy, 52
dla cechy ciągłej, 55
dla cechy dyskretnej, 52

Ś

średnie, porównywanie, 165, 177

T

tabele
kontyngencji, 213
liczności, 52
wielodzielcze, 213

tekst, formatowanie, 21
 testy istotności, 115
 testy statystyczne, 115
 χ^2 , 118, 120
 Brown-Forsythe'a, 130, 169
 C Cochran-Coxa, 132
 Duncana, 177, 178
 Dunnetta, 180
 F, 126, 127, 128
 jednorodności wariancji, 126
 kolejności par Wilcoxa, 157
 Kołmogorowa-Smirnowa, 118, 119, 120, 122, 123, 151
 Kruskala-Wallisa, 195
 Levene'a, 128, 129, 169
 Lillieforsa, 119, 122
 Mauchleya, 198
 McNemary, 209, 210, 211
 Newman i Keuls, 177, 178
 nieparametryczne, 115, 117, 118, 151, 156
 niezależności χ^2 , 213
 NIR, 178
 normalności rozkładu, 120
 parametryczne, 115, 117
 Q Cochran, 211, 212, 213
 Scheffégo, 177
 schemat weryfikacji, 116
 test mediany, 195
 test znaków, 156
 t-Studenta, 118, 131, 132, 137
 Tukeya, 177
 U Manna-Whitneya, 132, 151, 152
 W Shapiro-Wilka, 119, 122
 Walda-Wolfowitza, 132, 151, 154
 zasady, 116

W

wariancja, 58
 wartość
 modalna, 58
 oczekiwana, 77
 średnia, 58
 widok klasyczny, 11, 12
 wielokrotne porównywania, 177
 wiersz, zaznaczanie, 18
 Wintersa, model, 282
 wnioskowanie statystyczne, 103
 współczynnik
 determinacji, 218
 Fi, 215
 kontyngencji C-Pearsona, 215
 korelacji liniowej, 217
 skośności, 59

ufności, 104
 V Craméra, 215
 zmienności, 58
 wstążka, 11, 12
 wyglądanie wykładnicze, 295, 301
 wykresy, 31
 3W sekwencyjne, 39, 40, 41
 interakcji, 172
 liniowe, 38, 45, 46
 normalności, 125
 obracanie, 41
 obrazkowe, 42, 43, 47
 powierzchniowe, 47, 48, 49, 50
 ramka-wąsy, 60, 61, 62
 rozrzutu, 31, 32
 skategoryzowane, 66
 słupkowe wielokrotne, 39
 słupkowe/kolumnowe, 35, 37, 40
 Twarze Chernoffa, 42, 43, 44, 47
 warstwiczne, 47, 48, 49
 zmiana ustawień, 34
 wyniki obliczeń, zarządzanie, 29, 30
 wyrażenia matematyczne, tworzenie, 24
 wzory matematyczne, 28

Z

zakresy, zaznaczanie, 20
 zbiorowość, 51
 zdarzenia elementarne, 73
 zdarzenia niemożliwe, 73
 zdarzenia niezależne, 75
 zdarzenia pewne, 73
 zdarzenia przeciwne, 73
 zdarzenia sprzyjające, 74
 zdarzenie losowe, 73
 zmienna losowa, 75
 badanie normalności rozkładu, 118
 ciągła, 75
 rozkład, 75
 skokowa (dyskretna), 75
 zmienne, 51
 dodawanie, 18, 19
 formatowanie nazw, 23
 grupujące, 165
 niezależne, analiza, 64
 specyfikacja, 16
 specyfikacja za pomocą wyrażeń
 matematycznych, 25
 standaryzacja, 39, 79
 zależne, 165, 243

PROGRAM PARTNERSKI

GRUPY WYDAWNICZEJ HELION



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

Zmień swoją stronę WWW
w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

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

GRUPA WYDAWNICZA

 **Helion SA**

Statystyka z programem Statistica

Statistica to pakiet oprogramowania statystycznego i analitycznego, uważany za jedno z najlepszych rozwiązań w swojej klasie. Oferuje szeroki wybór narzędzi statystycznych i procedur analizy danych oraz umożliwia zarządzanie nimi i wizualizację otrzymanych wyników. Zapewnia też dostęp do zaawansowanych algorytmów modelowania i prognozowania. Właśnie dlatego stanowi obecnie standardowe narzędzie pracy w wielu firmach analitycznych, na uczelniach oraz w różnego rodzaju instytucjach badawczych. Ponadto dzięki prostemu interfejsowi i dobrej dokumentacji stale zyskuje nowych użytkowników.

Jeśli jesteś zainteresowany poznaniem możliwości środowiska oraz praktyczną nauką sposobów prowadzenia analiz statystycznych, koniecznie sięgnij po książkę *Statystyka z programem Statistica*. Z nią poznasz podstawowe pojęcia i metody statystyczne oraz nauczysz się tworzyć i stosować odpowiednie modele. Dowiesz się, jak posługiwać się programem do szybkiego osiągnięcia wyznaczonych celów. Przedstawione tu zagadnienia przygotują Cię do samodzielnego opracowania wyników badań, ich prawidłowej analizy oraz formułowania wniosków.

- Interfejs i podstawowe możliwości środowiska Statistica
- Pojęcia, metody i narzędzia wykorzystywane w analizach statystycznych
- Badanie wpływu różnych czynników na wybrane cechy populacji
- Ocena błędów wnioskowania o parametrach populacji na podstawie wyników próby losowej
- Analiza współzależności między zmiennymi losowymi oraz ocena jakości zbudowanego modelu
- Konstruowanie prognoz na podstawie szeregów czasowych
- Sposoby wizualizacji wyników analiz statystycznych

Przekonaj się, co można osiągnąć
za pomocą programu Statistica!

helion.pl
księgarnia
internetowa



Helion

Nr katalogowy: 8846



Księgarnia internetowa:
<http://helion.pl>



Zamówienia telefoniczne:

0 801 339900



0 601 339900

Sprawdź najnowsze promocje:

• <http://helion.pl/promocje>

Książki najchętniej czytane:

• <http://helion.pl/bestsellery>

Zamów informacje o nowościach:

• <http://helion.pl/novosci>

Helion SA

ul. Kościuszki 1c, 44-100 Gliwice

tel.: 32 230 98 63

e-mail: helion@helion.pl

<http://helion.pl>

sięgnij po **WIĘCEJ**



400 KORZYŚCI

ISBN 978-83-246-4110-9



Cena 54,90 zł

Informatyka w najlepszym wydaniu