

Małgorzata Rabej

Analizy statystyczne z programami Statistica i Excel



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: Małgorzata Kulik

Projekt okładki: Studio Gravite / Olsztyn
Obarek, Pokoński, Pazdrijowski, Zaprucki

Grafika na okładce została wykorzystana za zgodą Shutterstock.com

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/warsta>
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Kody źródłowe wybranych przykładów dostępne są pod adresem:
<ftp://ftp.helion.pl/przyklady/warsta.zip>

ISBN: 978-83-283-3922-4

Copyright © Helion 2018

Printed in Poland.

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

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

Spis treści

Wprowadzenie	7
---------------------------	----------

Rozdział 1. Metody gromadzenia i prezentacji danych	11
--	-----------

Wprowadzenie	11
1.1. Program Statistica	11
Ćwiczenie 1.1.1. Wprowadzanie i zapisywanie danych	13
Ćwiczenie 1.1.2. Tworzenie wyrażeń matematycznych	17
Ćwiczenie 1.1.3. Sortowanie danych	17
Ćwiczenie 1.1.4. Wzory matematyczne. Przeglądarka funkcji	18
Ćwiczenie 1.1.5. Zarządzanie wynikami	19
Ćwiczenie 1.1.6. Wykresy funkcji	20
Ćwiczenie 1.1.7. Wykresy słupkowe/kolumnowe	22
Ćwiczenie 1.1.8. Wykresy liniowe	22
Ćwiczenie 1.1.9. Wykres słupkowy wielokrotny	23
Ćwiczenie 1.1.10. Wykres 3W sekwencyjny	25
Ćwiczenie 1.1.11. Wykresy obrazkowe	27
Ćwiczenie 1.1.12. Wykres liniowy	29
1.2. Program Microsoft Excel	30
Ćwiczenie 1.2.1. Instalacja dodatków do analizy statystycznej	32
Ćwiczenie 1.2.2. Wprowadzanie wyrażeń arytmetycznych i funkcji	33
Ćwiczenie 1.2.3. Adresy względne i bezwzględne	33
Ćwiczenie 1.2.4. Funkcje daty i czasu	34
Ćwiczenie 1.2.5. Funkcje statystyczne	35
Ćwiczenie 1.2.6. Wykres funkcji	36
Ćwiczenie 1.2.7. Wykres słupkowy	37
Ćwiczenie 1.2.8. Wykres kolumnowy	37
Ćwiczenie 1.2.9. Sortowanie i filtrowanie tabel	38
Ćwiczenie 1.2.10. Tabele przestawne	40

Rozdział 2. Statystyka opisowa	43
Ćwiczenie 2.1. Analiza wyników pomiarów dla cechy dyskretnej	47
Ćwiczenie 2.2. Analiza wyników pomiarów dla cechy ciągłej	54
Ćwiczenie 2.3. Wykres ramka-wąsy w Excelu	60
Ćwiczenie 2.4. Analiza danych dla dwóch grup (populacji)	63
Rozdział 3. Rozkłady zmiennych losowych	67
Ćwiczenie 3.1. Rozkład dwumianowy	76
Ćwiczenie 3.2. Standardowy rozkład normalny	80
Ćwiczenie 3.3. Rozkład t-Studenta	85
Ćwiczenie 3.4. Rozkład chi-kwadrat	91
Ćwiczenie 3.5. Rozkład średniej z próby	93
Ćwiczenie 3.6. Rozkład sumy zmiennych losowych	94
Rozdział 4. Wyznaczanie przedziałów ufności dla średniej i odchylenia standardowego	97
Ćwiczenie 4.1. Przedział ufności dla wartości średniej	99
Ćwiczenie 4.2. Przedział ufności dla odchylenia standardowego	105
Rozdział 5. Testy statystyczne	109
5.1. Wprowadzenie	109
5.2. Testy jednorodności wariancji	112
Ćwiczenie 5.2.1. Test F	113
Ćwiczenie 5.2.2. Test Levene'a	115
5.3. Badanie normalności rozkładu zmiennych	117
Ćwiczenie 5.3.1. Wykres normalności	117
Ćwiczenie 5.3.2. Test zgodności χ^2	120
Ćwiczenie 5.3.3. Test Kołmogorowa-Smirnowa i test Shapiro-Wilka	124
5.4. Testy t-Studenta	126
Ćwiczenie 5.4.1. Test t dla pojedynczej próby	130
Ćwiczenie 5.4.2. Test t dla dwóch prób niezależnych	135
Ćwiczenie 5.4.3. Test t dla dwóch prób zakładający nierówne wariancje	139
Ćwiczenie 5.4.4. Test t dla dwóch prób zależnych	148
5.5. Testy nieparametryczne dla dwóch prób niezależnych	150
Ćwiczenie 5.5.1. Test U Manna-Whitneya	151
Ćwiczenie 5.5.2. Test serii Walda-Wolfowitza	153
5.6. Testy nieparametryczne dla prób zależnych	154
Ćwiczenie 5.6.1. Test znaków	155
Ćwiczenie 5.6.2. Test kolejności par Wilcoxon'a	156

Rozdział 6. Porównywanie wielu średnich	157
Ćwiczenie 6.1. Badanie jednorodności wariancji dla wielu grup	158
Ćwiczenie 6.2. Test Bartletta	159
Ćwiczenie 6.3. Test Browna-Forsythe'a	161
Ćwiczenie 6.4. Testy Levene'a i Browna-Forsythe'a	162
Ćwiczenie 6.5. ANOVA jednoczynnikowa	163
Ćwiczenie 6.6. Porównania wielokrotne	176
Ćwiczenie 6.7. MANOVA — analiza wariancji dwuczynnikowa	179
Ćwiczenie 6.8. ANOVA dla układów czynnikowych	183
Ćwiczenie 6.9. Porównania zaplanowane. Analiza kontrastów	185
Ćwiczenie 6.10. Test Kruskala-Wallisa i test mediany	194
Ćwiczenie 6.11. Układy z powtarzаныmi pomiarami	196
Rozdział 7. Analiza zmiennych jakościowych	203
Ćwiczenie 7.1. Tabele wielozdzielcze. Test niezależności χ^2	206
Ćwiczenie 7.2. Testy McNemary i Fishera	211
Ćwiczenie 7.3. Test Q Cochrańa	214
Rozdział 8. Analiza współzależności między zmiennymi	219
8.1. Regresja liniowa	223
Ćwiczenie 8.1.1. Badanie korelacji	224
Ćwiczenie 8.1.2. Regresja liniowa	227
8.2. Regresja wieloraka	234
Ćwiczenie 8.2.1. Liniowy model regresji wielorakiej	236
Ćwiczenie 8.2.2. Regresja krokowa	244
8.3. Linearyzowana regresja nieliniowa	250
Ćwiczenie 8.3.1. Logarytmiczna funkcja regresji	250
Ćwiczenie 8.3.2. Wykładnicza funkcja regresji	258
Ćwiczenie 8.3.3. Hiperboliczna funkcja regresji	262
Ćwiczenie 8.3.4. Aproksymacja wielomianem drugiego stopnia	266
8.4. Estymacja nieliniowa	271
Ćwiczenie 8.4.1. Regresja użytkownika. Solver	271
Ćwiczenie 8.4.2. Regresja logistyczna	278
Rozdział 9. Szeregi czasowe. Metody prognozowania	285
Ćwiczenie 9.1. Prognozowanie metodą średniej ruchomej	288
Ćwiczenie 9.2. Wygładzanie wykładnicze	297
Ćwiczenie 9.3. Model Holta	302
Ćwiczenie 9.4. Model trendu liniowego	309

Ćwiczenie 9.5.	Metoda wskaźników. Dekompozycja sezonowa (Census 1) ...	313
Ćwiczenie 9.6.	Model ARIMA dla pojedynczego szeregu	325

Rozdział 10. Wielowymiarowe techniki eksploracyjne 333

10.1. Analiza skupień	334
Ćwiczenie 10.1.1. Aglomeracja	334
Ćwiczenie 10.1.2. Analiza skupień. Grupowanie metodą k-średnich	338
10.2. Analiza czynnikowa	341
Ćwiczenie 10.2.1. Zastosowanie analizy czynnikowej do redukcji zmiennych	342
10.3. Analiza składowych głównych	349
Ćwiczenie 10.3.1. Zastosowanie analizy składowych głównych do klasyfikacji	350

Rozdział 11. Sieci neuronowe 355

Ćwiczenie 11.1.	Zastosowanie automatycznych sieci neuronowych do klasyfikacji danych doświadczalnych	360
Ćwiczenie 11.2.	Zastosowanie SSN do prognozowania na podstawie szeregów czasowych	366

Bibliografia 373

Skorowidz 376

Rozdział 4.

Wyznaczanie przedziałów ufności dla średniej i odchylenia standardowego

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 wartości 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 dystrybucją standardowego rozkładu normalnego.

Długość przedziału ufności zależy od 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 jest z relacji: $P(-t_{\alpha} < t < t_{\alpha}) = 1 - \alpha$.

Im większa jest wartość współczynnika ufności, tym szerszy jest przedział ufności, a więc mniejsza dokładność estymacji parametru. Przy zmniejszaniu wartości współczynnika ufności maleje długość przedziału ufności. Długość przedziału ufności jest miarą **precyzji estymacji przedziałowej**. Szeroki przedział ufności oznacza możliwość dużych odchyłeń 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. Przedział ufności dla wartości średniej

Przykład 35. Przedział ufności dla średniej 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, i przyjmując współczynnik ufności 0,95, wyznacz przedział ufności dla wartości średniej.

Dane

Współczynnik ufności oznaczany jako $1 - \alpha$ wynosi 0,95. Zmienną losową jest zanieczyszczenie gleby ołowiem.

- Ponieważ próba jest duża, można przyjąć, że $\sigma = s$.

Sposób wykonania w programie Microsoft Excel

- Otwórz plik *ólów.xlsx* lub wybierz z menu *Plik/Nowy* i wprowadź dane.
- Wybierz z menu *Dane/Analiza danych/Statystyka opisowa*.
- W oknie wprowadź zakres danych, zakres wyjściowy i poziom ufności dla średniej, jak pokazano na rysunku 4.1.

RYSUNEK 4.1.
Okno statystyk
opisowych

- Program wyświetla wszystkie parametry i poziom ufności (połowa przedziału ufności):

	A	B
1	Kolumna1	
2		
3	Średnia	64,62
4	Błąd standardowy	0,43
5	Mediana	65,00
6	Tryb	64,00
7	Odchylenie standardowe	3,13
8	Wariancja próbek	9,81
9	Kurtoza	-0,67
10	Skośność	-0,09
11	Zakres	13
12	Minimum	58
13	Maksimum	71
14	Suma	3360
15	Licznik	52
16	Poziom ufności(95,0%)	0,87

- Dolną granicę przedziału ufności można wyznaczyć, odejmując od średniej obliczony poziom ufności, czyli 0,87. Otrzymamy 63,7. Górną granicę można obliczyć, dodając do średniej poziom ufności: 65,5.

II sposób

Przedział ufności wyznacza się według wzoru. Do wzoru potrzebne są wartość średnia, odchylenie standardowe oraz u_α

Końce przedziału obliczane są według wzoru:

$$\left(\bar{X} - u_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X} + u_\alpha \frac{\sigma}{\sqrt{n}} \right).$$

Współczynnik ufności $1 - \alpha = 0,95$, czyli $\alpha = 0,05$.

u_α należy wyznaczyć z relacji $\Phi(u_\alpha) = 1 - (\alpha/2)$, gdzie Φ jest dystrybuantą standardowego rozkładu normalnego.

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

Do wyznaczenia u_α w Excelu służy funkcja ROZKŁAD.NORMALNY.S.ODW.

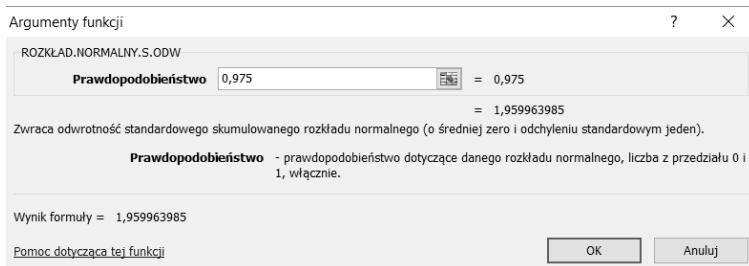
W oknie funkcji należy podać wartość prawdopodobieństwa = 0,975 (rysunek 4.2).

RYСУNEK 4.2.

Okno funkcji

Rozkład.

Normalny.S.Odw



Funkcja zwraca wartość $u_\alpha = 1,96$.

Po podstawieniu do wzoru otrzymamy przedział (63,76, 65,5).

Odpowiedź

Otrzymany przedział ufności (63,7, 65,5) jest jednym z możliwych do otrzymania przedziałów, które z ufnością równą 95% pokrywają średnie zanieczyszczenie gleby.

Rozwiązanie z użyciem programu Statistica

- 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 *ołów.sta*.

I sposób

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

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

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

Odpowiedź

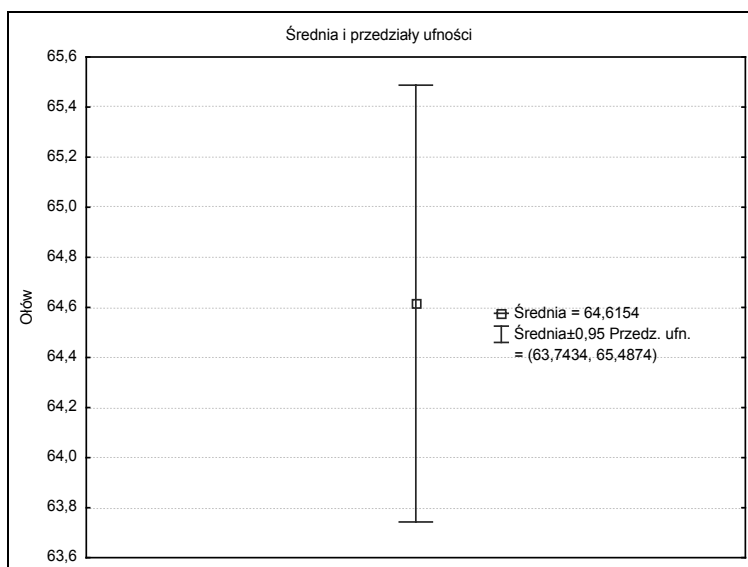
Przedział ufności (w programie Statistica: (*Ufność -95%*) = 63,7; (*Ufność +95%*) = 65,5) ma postać (63,7, 65,5).

II sposób

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

- Wybierz z menu: *Wykresy/Wykresy średnia i błędy*.
- Zdefiniuj zmienną Ołów.
- Wprowadź współczynnik ufności. Współczynnik ufności jest wyświetlany w polu *Prawdopodob.* (domyślna wartość tego współczynnika wynosi 0,95). Kliknij OK.
- Program tworzy wykres średniej i przedziałów ufności dla tej średniej (rysunek 4.4). Przedział ufności jest wyświetlany na wykresie.

RYСУNEK 4.4.
Wykres średniej i przedziałów ufności dla tej średniej



Odpowiedź

Z prawdopodobieństwem równym 0,95 można stwierdzić, że średnie zanieczyszczenie gleby ołowiem zawiera się w przedziale (63,7, 65,5) mg/kg suchej masy.

Przykład 36. 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 równy 0,95, wyznacz przedział ufności dla średniej wartości zanieczyszczenia gleby ołowiem.

Dane

$1 - \alpha = 0,95$, $n = 12$, $\alpha = 0,05$, liczba stopni swobody jest równa $n - 1$, czyli 11.

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 $P(-t_{\alpha} < t < t_{\alpha}) = 1 - \alpha$.

Sposób wykonania w programie Microsoft Excel

- Program Excel nie rozpoznaje, czy próba jest mała, czy duża. Dlatego należy znaleźć wartość t_{α} i wyznaczyć przedział ufności według podanego wyżej wzoru.
- Do wzoru trzeba obliczyć wartość średnią, odchylenie standardowe oraz wartość statystyki t-Studenta t_{α} dla $\alpha = 0,05$.
- Do wyznaczania wartości t_{α} w Excelu służy funkcja Rozkład.T.Odw. Funkcja ta ma 2 argumenty: prawdopodobieństwo α oraz liczbę stopni swobody.
- Kliknij przycisk *Wstaw funkcję* i wybierz Rozkład.T.odw. Wprowadź: Liczba stopni swobody = liczbie pomiarów-1 = 11, prawdopodobieństwo = 0,05.
- Funkcja zwraca wartość $t_{\alpha} = 2,2$.
- Aby wyznaczyć wartość średnią i odchylenie standardowe, wstaw funkcje: Średnia oraz Odchylenie.standardowe. W obu funkcjach wystarczy wskazać zakres danych. Wartość średnia = 62,83, s = 5,17.
- Gdy dane te zostaną podstawione do wzoru, przedział ufności otrzymuje postać: (59,4, 66,3).
- Można sprawdzić, że przedział ufności liczony podobnie jak w poprzednim przykładzie, z wykorzystaniem modułu *Statystyka opisowa*, ma postać: (59,55, 66,12).

Sposób wykonania w programie Statistica

- 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 4.5).

RYСУNEK 4.5.

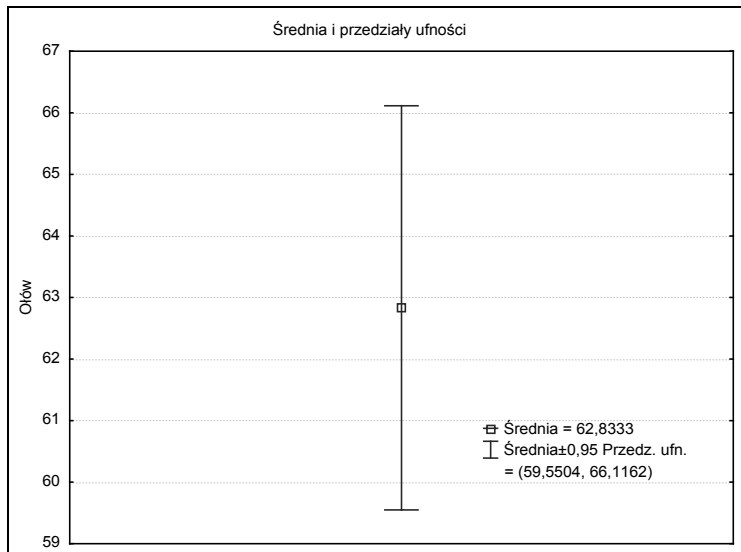
Arkusz
z wynikami
obliczeń

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

- 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 4.6).

RYСУNEK 4.6.

Wykres średniej
i przedziałów
ufności



Odpowiedź

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

Ćwiczenie 4.2. Przedział ufności dla odchylenia standardowego**Przykład 37. Przedział ufności dla odchylenia standardowego**

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

Dane

$1 - \alpha = 0,98$, $n = 12$, $\alpha = 0,02$, liczba stopni swobody jest równa $n - 1$, czyli 11.

Sposób wykonania w programie Microsoft Excel

Przedział ufności dla wariancji dla małych prób ma postać:

$$P\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right) = 1 - \alpha$$

Statystyki $\chi^2_{\frac{\alpha}{2}, n-1}$ oraz $\chi^2_{1-\frac{\alpha}{2}, n-1}$ wyznacza się z rozkładu chi-kwadrat. W Excelu służy do tego funkcja ROZKŁAD.CHI.ODW.

Wariancja jest kwadratem odchylenia standardowego. Aby wyznaczyć przedział ufności dla odchylenia standardowego, należy obliczyć pierwiastek z wyznaczonych wartości przedziału ufności dla wariancji.

Jeśli $\alpha = 0,02$, to $\alpha / 2 = 0,01$, $1 - \alpha / 2 = 0,99$.

- Otwórz plik *zanieczyszczenie łożem.xlsx*.
- Oblicz wariancję s^2 . W tym celu wybierz przycisk *Wstaw funkcję*. Wybierz funkcję statystyczną *Wariancja*. Jako argument funkcji należy podać zakres danych. Otrzymamy $s^2 = 26,7$.
- Kliknij w dowolnej komórce arkusza i wstaw funkcję statystyczną: *Rozkład.Chi.ODW*.
- W oknie wprowadź: prawdopodobieństwo: 0,01, stopnie swobody: 11. Funkcja zwraca wartość $\chi^2_{\frac{\alpha}{2}, n-1} = 24,72$.
- Powtórz wywołanie funkcji i wprowadź: prawdopodobieństwo = 0,99, stopnie swobody = 11. Funkcja zwraca wartość $\chi^2_{1-\frac{\alpha}{2}, n-1} = 3,05$.
- Po podstawieniu do wzoru otrzymamy przedział ufności dla wariancji (11,89. 96,29).

Aby obliczyć przedział ufności dla odchylenia standardowego, należy obliczyć pierwiastek z 11,89 oraz pierwiastek z 96,29.

- Przedział ufności dla odchylenia standardowego ma postać (3,45, 9,81).

Przykład 38. Przedział ufności dla odchylenia standardowego dla dużych prób

Dla danych z przykładu 35. wyznacz przedział ufności dla średniej, dla współczynnika ufności równego 0,98.

Jeśli $\alpha = 0,02$, to $\alpha / 2 = 0,01$, $1 - \alpha / 2 = 0,99$.

- Otwórz plik *ołów.xlsx*.
- Dla dużych prób przedział ufności wyznaczany jest ze wzoru:

$$P \left\{ \frac{s}{1 + \frac{u_\alpha}{\sqrt{2n}}} < \sigma < \frac{s}{1 - \frac{u_\alpha}{\sqrt{2n}}} \right\} \approx 1 - \alpha$$

Wartość u_α można wyznaczyć z zależności: $\Phi(u_\alpha) = 1 - \alpha / 2$, $\Phi(u_\alpha) = 0,99$.

Do wyznaczenia u_α służy w Excelu funkcja ROZKŁAD.NORMALNY.S.ODW.

W oknie funkcji należy podać wartość prawdopodobieństwa = 0,99.

Otrzymamy: $u_\alpha = 2,33$.

Wartość średnia = 64,62.

Odchylenie standardowe = 3,13.

Po podstawieniu do powyższego wzoru przedział ufności dla odchylenia standardowego ma postać: (2,55, 4,06).

Odpowiedź

Otrzymany przedział (2,55, 4,06) z prawdopodobieństwem równym 0,98 pokrywa odchylenie standardowe zanieczyszczenia gleby ołowiem.

Końce przedziału mogą się nieznacznie różnić — zależy to od tego, z jaką dokładnością było obliczane odchylenie standardowe oraz u_α .

Sposób wykonania w programie Statistica

- 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.*
- Pole edycji *Przedział* zawiera współczynnik ufności podawany w procentach. Wprowadź 98.
- Program tworzy arkusz z wynikami (rysunek 4.7).

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

Rysunek 4.7. Arkusz z wynikami statystyk opisowych

Odpowiedź

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

Skorowidz

A

adresowanie mieszane, 34
adresy

- bezwzględne, 33
- względne, 33

aglomeracja, 334
aksjomatyczna definicja
prawdopodobieństwa, 68
algorytm

- aglomeracji, 333
- RBFT, 364

algorytmy metody redukcji, 358
analiza

- czynnikowa, 333, 341
- danych dla dwóch grup, 63
- danych jakościowych, 207
- danych na wykresach, 29
- danych skategoryzowanych, 63
- kontrastów, 186–190, 194
- korelacji, 203
- log-liniowa, 205
- reszt, 240
- składowych głównych, 349
- skupień, 205, 333, 334, 338
- wariancji, 164, 166, 172, 175
- ANOVA, 157
 - dwuczynnikowa MANOVA, 180
 - jednoczynnikowa, 167, 190
- wartości średnich, 191
- wrażliwości, 358
- współzależności między zmiennymi, 219
- wyników pomiarów
 - dla cechy ciągłej, 54
 - dla cechy dyskretnej, 47
- zależności dwóch zmiennych, 227
- zmiennych ilościowych, 279
- zmiennych jakościowych, 203

Analysis ToolPak, 32
ANOVA

- dla układów czynnikowych, 183
- jednoczynnikowa, 157, 163

aproksymacja wielomianem drugiego stopnia, 266
ARIMA, 325, 330
arkusz, 12, 13

- analiza wrażliwości, 366
- predykcji, 366
- z danymi, 16

B

badanie

- interakcji między czynnikami, 183
- jednorodności wariancji
 - dla wielu grup, 158
 - test Bartletta, 160
 - test Browna-Forsythe'a, 161
 - test F, 113
 - test Levene'a, 115
- korelacji, 224
- normalności rozkładu zmiennych, 117

błąd

- bezwzględny procentowy, 287
- drugiego rodzaju, 109
- pierwszego rodzaju, 109
- standardowy, 75
- średni prognozy, 287
- względny, 306, 308
- względny prognozy, 311

błędy

- ex post, 286
- prognozy, 299, 311

brak autokorelacji reszt, 239, 249, 254

C

cechy

- ciągłe, 54
- dyskretne, 47
- ilościowe, 43
- statystyczne, 43

Census 1, 313

centralne twierdzenie graniczne, 74

cząstkowe współczynniki regresji, 234

czynnik

- główny, 341
- pomiarów powtarzanych, 196

D

definicja prawdopodobieństwa

- aksjomatyczna, 68
- klasyczna, 68

definiowanie funkcji matematycznych, 18

dekompozycja

- sezonowa, 313, 319
- szeregu czasowego, 313

dendrogram, 334

dodatek Analysis ToolPak, 32

dodawanie przypadków, 16

dopasowanie funkcji wykładniczej, 258

dystybuanta, 70

- zmiennej losowej ciągłej, 71
- zmiennej losowej skokowej, 70

E

ekstrapolacja funkcji trendu, 310

entropia wzajemna, 358, 364

estymacja, 97

- nieliniowa, 220, 271
- nieparametryczna, 97
- parametryczna, 97
- przedziałowa, 98, 205
- punktowa, 97

estymator, 97

eta-kwadrat cząstkowe, 173

Excel, 30

- adresowanie, 34
- analiza

wariancji dwuczynnikowa, 182

wariancji jednoczynnikowa, 167

wyników pomiarów, 58

aproksymacja wielomianem drugiego stopnia, 269

badanie korelacji, 226

dodatek Analysis ToolPak, 32

estymacja nieliniowa, 275

filtrowanie tabel, 38

funkcje

- daty i czasu, 34
- statystyczne, 35, 77
- wbudowane, 33

hiperboliczna funkcja regresji, 264

instalacja dodatków, 32

liniowy model regresji, 241

porównanie

- dwóch średnich, 149
- kilku prób zależnych, 216
- kilku średnich, 173

poziom ufności, 100

prognozowanie metodą

- Holta, 303
- średniej ruchomej, 289
- wygładzania wykładniczego, 297

przedział ufności dla odchylenia standardowego, 105

regresja liniowa, 230

rozkład

- chi-kwadrat, 91
- dwumianowy, 77
- t-Studenta, 86, 129

sortowanie tabel, 38

standardowy rozkład normalny, 80

tabele przestawne, 40

test

- normalności chi-kwadrat, 123
- t dla dwóch prób niezależnych, 138
- t dla pojedynczej próby, 133
- t z dwiema próbami zakładający nierówne wariancje, 140, 145

Test.F, 114

tworzenie

- histogramu, 51
- szeregu rozdzielczego, 51

wprowadzanie

- funkcji, 33
- wyrażeń, 33

Excel

- wykładnicza funkcja regresji, 260
- wykres
 - funkcji, 36
 - kolumnowy, 37
 - normalności, 120
 - ramka-wąsy, 60
 - słupkowy, 37

F

filtrowanie

- danych, 39
- tabel, 38

funkcja

- 3W, 283
- aktywacji, 356
- autokorelacji, 326
- błędu, 355
- błędu SOS, 358, 364
- celu, 298, 304
- gęstości rozkładu prawdopodobieństwa,
 - 70, 86
- Kwartyl, 53, 60, 173
- LN, 160
- Mediana, 173
- Odch.standardowe, 119
- prawdopodobieństwa, 70
- regresji
 - hiperboliczna, 262
 - logarytmiczna, 250
 - wykładnicza, 258
- Rozkład.Chi, 91, 160, 216
- Rozkład.Dwum, 77
- Rozkład.Normalny, 81, 94, 95, 124
- Rozkład.Normalny.S.Odw, 101, 232
- Rozkład.T, 86
- Rozkład.T.Odw, 103, 134
- Skośność, 138, 167
- Softmax, 364
- Średnia, 119
- Test.F, 112, 114, 147
- Ufność, 176
- Usuń duplikaty, 119
- Wariancja, 160
- Wsp.Korelacji, 226, 277

funkcje

- aktywacji, 369
- autokorelacji, 326
- celu, 308
- daty i czasu, 34
- statystyczne, 35, 77
- trendu, 310, 317
- użytkownika, 272

G

- globalna analiza wrażliwości, 358
- graficzna prezentacja statystyk opisowych, 47
- granica przedziału ufności, 98
- gromadzenie danych, 11
- grupowanie
 - danych, 334
 - metodą k-średnich, 333, 338

H

- hiperboliczna funkcja regresji, 262
- hipoteza
 - alternatywna, 110
 - statystyczna, 109
 - zerowa, 109
- histogram, 44, 47
 - reszt, 282
- horyzont prognozy, 370

I

- identyfikacja poliamidu, 360
- iloczyn zdarzeń, 67
- interakcja między czynnikami, 183
- interpretacja wyników, 346, 354
- istotność
 - cząstkowych współczynników regresji, 238, 249
 - regresji liniowej, 238, 243, 249
 - współczynników regresji, 229, 243, 254

J

- jednorodność wariancji, 191
- jednostki statystyczne, 43

K

Kalkulator prawdopodobieństwa, 77, 93
 dla rozkładu normalnego, 83
 karta
 Przegląd, 294
 Przesunięcie, 293
 Przewidywania, 365
 wyboru podzbiorów, 362
 wyboru zmiennych, 362
 Wygładzanie, 292
 klasyczna definicja prawdopodobieństwa, 68
 klasyfikacja, 350
 danych doświadczalnych, 360
 komputerowe prawdopodobieństwo, 110
 konstrukcja prognozy, 305
 korelacja, 224
 cząstkowa, 235
 semicząstkowa, 235
 korelacje między zmiennymi, 344
 kryterium Kaisera, 342
 kwantyl, 72
 rzędu p , 72
 kwartył, 53, 72

L

liczebność, 43
 linearyzowana regresja nieliniowa, 250
 liniowy model regresji, 236
 logarytmiczna funkcja regresji, 250
 losowość odchylenia, 221

Ł

ładunki czynnikowe, 342, 346

M

MANOVA, 179
 metoda
 dekompozycji Census 1, 325
 średniej ruchomej, 288
 wskaźników, 313
 wstecznej propagacji błędów, 358
 metody prognozowania, 285
 miary
 asymetrii, 45, 46
 koncentracji, 45, 46

położenia, 45
 rozproszenia, 45
 moc testu, 111
 model
 addytywny, 288
 ARIMA, 325
 Holta, 303
 multiplikatywny, 288
 regresji linearyzowanej, 252
 trendu liniowego, 309

N

neutron, 355
 niecentralność, 173
 nieobciążoność reszt, 222
 normalność
 rozkładu, 191
 reszt, 221, 229, 239, 249, 254
 zmiennych losowych, 117

O

obracanie wykresu 3W, 26
 obserwacje odstające, 74
 obszar krytyczny testu, 130–132
 ocena
 dopuszczalności prognozy, 296, 305, 307
 modelu, 229, 233, 241, 250, 255
 odchylenie standardowe, 97
 odległość wiązania, 334
 okno
 analizy szeregów, 301
 ARIMA, 330
 Automatyczny projekt sieci, 363
 Definicja modelu, 245
 dekompozycji sezonowej, 320
 funkcji
 Rozkład normalny, 81
 Rozkład T, 87
 Test F, 114
 Kalkulator prawdopodobieństwa, 77, 83, 93
 Regresja, 231
 Solver, 298, 304
 statystyk opisowych, 100
 ustawień tabeli licznosci, 55
 wyboru metody estymacji, 273
 wyboru zmiennych, 253, 322, 323

P

perceptron, 356
 wielowarstwowy, 356
 pionowy wykres soplekowy, 337
 pliki
 *.spf, 13
 *.sta, 13
 pobudzenie neuronu, 356
 pomiary powtarzane, 198
 populacja generalna, 43
 porównania
 wielokrotne, 176, 177
 zaplanowane, 185
 dwóch
 grup, 154
 prób, 212
 prób zależnych, 155, 156
 średnich z prób niezależnych, 151, 153
 średnich z prób zależnych, 148
 kilku prób zależnych, 215
 prognoz, 332
 średnich, 157, 166, 172
 z prób dla zmiennych porządkowych, 194
 z prób niezależnych, 169
 z prób zależnych, 148
 poszukiwanie skupień
 hierarchiczne, 333
 niehierarchiczne, 333
 poziom istotności, 109
 poziomy wykres drzewa, 337
 prawdopodobieństwo, 68
 całkowite, 69
 warunkowe, 69
 prawo wielkich liczb Bernoulliego, 69
 precyzja estymacji przedziałowej, 99
 predyktor jakościowy, 157
 prezentacja
 danych, 11, 27
 danych na wykresach, 29
 prognozowanie
 metodą
 Holta, 303
 średniej ruchomej, 288
 wygładzania wykładniczego, 297
 na podstawie szeregów czasowych, 366

z trendem liniowym, 309
 z trendem i sezonowością, 326
 program
 Analysis ToolPak, 8
 Microsoft Excel, 30
 Solver, 8
 Statistica, 8, 11
 projekcja zmiennych na płaszczyznę
 czynników, 353
 próba, 43
 losowa, 43
 losowa prosta, 70
 próbka, 44
 przedział ufności, 97, 98
 dla dużych prób, 99, 106
 dla małych prób, 103
 dla odchylenia standardowego, 105
 dla wartości średniej, 99
 przeglądarka funkcji, 18, 19
 przestrzeń zdarzeń elementarnych, 67
 przewidywanie, 365

R

raport, 20
 redukcja zmiennych, 342
 regresja
 krokowa, 244, 246
 krokowa postępująca, 245
 linearyzowana, 219
 liniowa, 219, 223, 227
 logistyczna, 205, 278
 użytkownika, 271
 wieloraka, 220, 234
 reguła trzech sigm, 74
 reszty, 220
 odstające, 240
 rotacja ładunków, 342
 rozkład
 chi-kwadrat, 76, 91
 dwumianowy, 72, 76
 empiryczny, 44
 F Fishera, 76
 jednomodalny, 44
 normalny, 73
 Poissona, 73
 sumy zmiennych losowych, 94

średniej arytmetycznej, 75
 średniej z próby, 93
 t-Studenta, 75, 85
 rozkłady zmiennych losowych, 67, 69
 równanie
 regresji, 231
 regresji wielorakiej, 236
 różnica zdarzeń, 67

S

seria danych, 19
 sezonowość, 313
 sferyczność, 197
 sieci
 MLP, 356
 neuronowe, 355
 identyfikacja poliamidu, 360
 prognozowanie na podstawie szeregu czasowego, 367
 uczenie, 363
 RBF, 357
 siła skupienia, 334
 skategoryzowane wykres normalności, 165, 171
 składowa główna, 341
 skoroszyt, 13
 skupienie, 340
 Solver, 271, 276, 298, 304
 sortowanie
 danych, 17
 tabel, 38, 39
 specyfikacja zmiennych, 17
 sprawdzanie
 normalności, 121
 założenia jednorodności wariancji, 162, 166, 167
 założeń normalności, 167
 SSCP
 błędów, 198
 efektów, 197
 standardowy
 błąd bezwzględny, 306, 308
 rozkład normalny, 73, 80
 standaryzacja zmiennych, 23, 73
 Statistica
 analiza
 danych skategoryzowanych, 63
 kontrastów, 187, 190

wariancji dwuczynnikowa, 180
 wariancji jednoczynnikowa, 190
 wyników pomiarów, 48, 54
 aproksymacja wielomianem drugiego stopnia, 266
 automatyczne sieci neuronowe, 360
 badanie
 interakcji między czynnikami, 183
 jednorodności wariancji, 161
 korelacji, 224
 dekompozycja szeregów czasowych, 319
 estymacja nieliniowa, 271
 grupowania danych, 334
 hiperboliczna funkcja regresji, 262
 identyfikacja poliamidu, 360
 liniowy model regresji, 237
 logarytmiczna funkcja regresji, 250
 model ARIMA, 326
 porównanie
 dwóch prób, 212
 dwóch średnich, 148
 dwóch średnich z prób niezależnych, 152
 kilku prób zależnych, 215
 prognozowanie metodą
 Holta, 306
 średniej ruchomej, 291
 wygładzania wykładniczego, 300
 przedział ufności, 101
 dla odchylenia standardowego, 106
 regresja
 krokowa, 247
 liniowa, 227
 logistyczna, 278
 rozkład
 chi-kwadrat, 92
 dwumianowy, 78
 sferyczność, 197
 sprawdzanie
 jednorodności wariancji, 162
 normalności, 121
 standardowy rozkład normalny, 82
 test
 Kruskala-Wallisa, 158
 t dla dwóch prób niezależnych, 137, 140, 145
 t dla pojedynczej próby, 134
 t-Studenta, 128
 wykładnicza funkcja regresji, 258

Statistica
 wykres
 normalności, 119
 skategoryzowany ramka-wąsy, 64
 statystyka, 7
 matematyczna, 67
 opisowa, 43, 56, 59
 statystyki Durбина-Watsona, 236
 suma
 kwadratów różnic, 287
 zdarzeń, 67
 zmiennych losowych, 94
 symetria
 połączona, 196
 reszt, 222
 szereg
 czasowy, 285, 302, 308, 319, 324
 czasowy zróżnicowany, 328
 prognozowany, 308
 rozdzielczy, 44, 47, 52–55
 rozdzielczy dla cechy ciągłej, 58
 skorygowany, 322
 wygładzony, 302
 sztuczne sieci neuronowe, 355

Ś

średni
 bezwzględny błąd procentowy, 287
 bezwzględny błąd prognozy, 287
 błąd procentowy, 287
 kwadrat, 287
 kwadratowy błąd prognozy, 287
 średnia, 97
 arytmetyczna, 94
 z próby, 93
 średnie skupień, 339

T

tabele
 licznosci, 44, 48
 przestawne, 40
 wielodzielcze, 206
 tablice korelacyjne, 203
 test
 Bartletta, 159, 174
 Bonferroniego, 187

Browna-Forsythe'a, 161, 162
 chi-kwadrat, 117, 120
 Duncana, 177
 Dunnetta, 178
 F, 112, 113
 Fishera, 205, 211
 G-G, 201
 H-F, 201
 HSD Tukeya, 187
 istotności, 205
 kolejności par Wilcoxon, 156
 Kołmogorowa-Smirnowa, 117
 Kruskala-Wallis, 158, 194
 K-S, 125
 Levene'a, 115, 162
 Lillieforsa, 125
 Mauchleya, 197
 McNemary, 205, 211
 mediany, 194
 Newman i Keulsa, 177
 niezależności chi-kwadrat, 205, 206
 NIR, 177
 normalności chi-kwadrat, 123
 osypiska, 342
 Q Cochra, 205, 214
 Scheffego, 177, 187
 serii Walda-Wolfowitz, 153
 Shapiro-Wilka, 117, 171
 t dla dwóch prób niezależnych, 135, 137, 145
 t dla dwóch prób zakładający nierówne wariancje, 139
 t dla dwóch prób zależnych, 148
 t dla pojedynczej próby, 129–133
 t dla prób niezależnych, 128
 t dla prób zależnych, 128
 t-Studenta, 126
 Tukeya, 177
 U Manna-Whitneya, 151
 W Shapiro-Wilka, 125
 zgodności chi-kwadrat, 120
 znaków, 155
 testy
 ANOVA, 157
 istotności, 110
 jednorodności wariancji, 112
 nieparametryczne, 109

dla dwóch prób niezależnych, 150
 dla prób zależnych, 154
 parametryczne, 109
 post-hoc, 177, 193
 statystyczne, 109
 tolerancja, 235
 trend, 313, 316
 twierdzenie
 Lindeberga-Lévy'ego, 95
 o prawdopodobieństwie całkowitym, 69
 tworzenie
 arkusza z danymi, 14
 histogramu, 51
 raportów, 20
 szeregu rozdzielczego, 51, 58
 wykresów, 22
 liniowych, 22
 sekwencyjnych, 25
 typu ramka-wąsy, 56, 57, 64
 wielokrotnych, 23
 wyrażeń matematycznych, 17

U

uczenie, 369
 układy z powtarzaniem pomiarów, 196

V

varimax, 341

W

wahania sezonowe, 319
 wariancja składnika resztowego, 287
 wartości własne, 342, 351
 wartość oczekiwana, 71
 weryfikacja modelu, 274
 wielomian drugiego stopnia, 266
 wielowymiarowe techniki eksploracyjne, 333
 wkłady przypadków, 354
 wnioskowanie statystyczne, 97, 205
 wprowadzanie danych, 13
 wskaźnik WKP, 199
 współczynnik
 determinacji R^2 , 220
 f_i , 208

kontyngencji, 208
 korelacji liniowej, 223
 korelacji wielorakiej R , 234
 regresji, 223
 ufności, 98
 V Craméra, 204, 208
 Yule'a, 204
 zbieżności Czuprowa, 204
 korelacji, 204
 współrzędne czynnikowe zmiennych, 353
 wygładzanie wykładnicze, 297
 wykładnicza funkcja regresji, 258
 wykres
 3W sekwencyjny, 25
 absolutnych wahań sezonowych, 316
 autokorelacji, 328
 autokorelacji zróżnicowany, 329
 drzewa pionowego, 336
 drzewkowy, 334
 funkcji, 20, 21, 36
 funkcji dopasowanej, 274
 interakcji, 167, 185, 202
 kolumnowy, 22, 37
 liniowy, 22, 29, 30
 normalności, 117, 118, 233
 normalności reszt, 230, 244, 275
 normalności rozkładu reszt, 232
 obrazkowy, 27–31
 ospiska, 345, 352
 prawdopodobieństwa normalnego, 233
 przewidywanych względem wartości
 resztowych, 239
 ramka-wąsy, 56, 57
 rozrzutu, 225, 232, 252
 skategoryzowany ramka-wąsy, 64
 słupkowy, 37
 słupkowy wielokrotny, 23
 szeregu czasowego, 371
 średnich, 170, 189
 Twarze Chernoffa, 27
 wahań sezonowych, 315, 322
 wyniki analizy wariancji, 192
 wyrażenia matematyczne, 17
 wyznaczanie
 parametrów równania regresji wielorakiej,
 236
 prognozy, 293

wyznaczanie

przedziałów ufności, 97

trendu, 316

wahań przypadkowych, 317

wzory matematyczne, 18

Z

zakładka Funkcje aktywacji, 369

założenie homoscedastyczności, 238, 243

zapisywanie

danych, 13, 16

raportów, 20

zarządzanie wynikami, 19

zastosowanie

aglomeracji, 334

analizy czynnikowej, 342

analizy składowych głównych, 350

automatycznych sieci neuronowych, 360

modelu ARIMA, 326

regresji krokowej, 246

zbiorowość, 43

zdarzenia

losowe, 67

niemożliwe, 67

niezależne, 69

pewne, 67

przeciwne, 67

sprzyjające, 68

zmiana ustawień wykresów, 21, 22, 25

zmiennie

ciągłe, 69

dychotomiczne, 212, 215

dyskretne, 69

grupujące, 157

ilościowe, 279

jakościowe, 203

losowe, 69

o rozkładzie chi-kwadrat, 91

o rozkładzie dwumianowym, 76

o rozkładzie normalnym, 80

o rozkładzie t-Studenta, 86, 88

skokowe, 69

zmiennosć

międzygrupowa, 164

wewnątrzgrupowa, 164

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>

Analizy statystyczne z programami Statistica i Excel

W naszych czasach statystyka jest obecna dosłownie wszędzie.

Opisuje procesy społeczne, wskazuje kierunki rozwoju, dyktuje strategię działania rządów i międzynarodowych korporacji. Pomaga ekonomistom, naukowcom i inżynierom, umożliwia tworzenie prognoz gospodarczych, pozwala opracowywać nowe technologie i rozwiązania techniczne, wspiera walkę z epidemiami i odkrywanie nowych terapii. Jest po prostu nieodzowna, aby radzić sobie z wyzwaniami współczesnego świata i skalą zjawisk, z którymi mamy do czynienia.

Podstawą jest zgromadzenie oraz analiza danych w celu pozyskania z nich jak największej wiedzy o badanym zjawisku. Nie da się analizować ogromnych zbiorów danych bez pomocy komputerów i właściwego oprogramowania. Wśród najlepszych programów statystycznych jest Statistica firmy Statsoft, a do najpopularniejszych aplikacji biurowych należy Excel firmy Microsoft. Obydwa rozwiązania oferują szereg specjalistycznych narzędzi wspomagających obróbkę danych statystycznych, przeprowadzanie analiz i prezentowanie wyników. Obydwa warto poznać, a najlepiej zrobić to w praktyczny sposób — za pomocą ćwiczeń. Właśnie takich, jakie zostały przedstawione w tej książce!

- Metody graficznej prezentacji danych
- Charakterystyki liczbowe i rozkłady zmiennych losowych
- Estymacja parametrów badanej zbiorowości
- Testy statystyczne
- Analiza wskaźników i zależności
- Liniowe i nieliniowe modele regresji
- Analiza szeregów czasowych i metody prognozowania
- Techniki eksploracji danych
- Zastosowanie sieci neuronowych

Naucz się analizować i prezentować dane statystyczne!

Helion 		<i>Sprawdź nasze szkolenia!</i>		KOD KORZYŚCI <i>Sięgnij po więcej!</i> ▶			
	helion.pl	 AKADEMIA IT & BUSINESS WWW.SZKOLENIA.HELION.PL		ISBN 978-83-283-3922-4			
	0 801 339900						
	0 601 339900			9 788328 339224			
INFORMATYKA W NAJLEPSZYM WYDANIU				Cena: 59,00 zł			