Dynamika zjawisk (indeksy)

Numer artykułu	Ilość Rok 0	Ilość Rok 1	Cena jednostkowa Rok 0	Cena jednostkowa Rok 1
1	q_{10}	q_{11}	p_{10}	p_{11}
2	q_{20}	q_{21}	p_{20}	p_{21}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	q_{k0}	q_{1k}	p_{k0}	p_{1k}

Numer	Wartość	Wartość	Wartość	Wartość
1	$w_{1,00} = q_{10}p_{10}$	$w_{1,11} = q_{11}p_{11}$	$w_{1,01} = q_{10}p_{11}$	$w_{1,10} = q_{11}p_{10}$
2	$w_{2,00} = q_{20}p_{20}$	$w_{2,11} = q_{21}p_{21}$	$w_{2,01} = q_{20}p_{21}$	$w_{2,10} = q_{21}p_{20}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$w_{k,00} = q_{k0}p_{k0}$	$w_{k,11} = q_{k1}p_{k1}$	$w_{k,01} = q_{k0}p_{k1}$	$w_{k,10} = q_{k1}p_{k0}$
Razem	w_{00}	w_{11}	w_{01}	w_{10}

Indeks zmian wartości $I_w = w_{11}/w_{00}$
Indeks Laspayresa zmian ilości ${}_L I_{qp} = w_{10}/w_{00}$
Indeks Laspayresa zmian cen ${}_L I_{pq} = w_{01}/w_{00}$
Indeks Paaschego zmian ilości ${}_P I_{qp} = w_{11}/w_{01}$
Indeks Paaschego zmian cen ${}_P I_{pq} = w_{11}/w_{10}$
Indeks Fishera zmian ilości ${}_F I_q = \sqrt{{}_L I_{qp} \cdot \ {}_P I_{qp}}$
Indeks Fishera zmian cen ${}_F I_p = \sqrt{{}_L I_{pq} \cdot \ {}_P I_{pq}}$

Czas	Obserwacja	Indeksy jednopodstawowe absolutne	Indeksy jednopodstawowe względne	$i_{t\|c}$	Indeksy łańcuchowe absolutne	Indeksy łańcuchowe względne	$i_{t\|t-1}$
t_0	y_0	0	$\cdot$	1	0	$\cdot$	$\cdot$
t_1	y_1	$y_1 - y_0$	$(y_1 - y_0)/y_0$	y_1/y_0	$y_1 - y_0$	$(y_1 - y_0)/y_0$	y_1/y_0
t_2	y_2	$y_2 - y_0$	$(y_2 - y_0)/y_0$	y_2/y_0	$y_2 - y_1$	$(y_2 - y_1)/y_1$	y_2/y_1
t_3	y_3	$y_3 - y_0$	$(y_3 - y_0)/y_0$	y_3/y_0	$y_3 - y_2$	$(y_3 - y_2)/y_2$	y_3/y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
t_k	y_k	$y_k - y_0$	$(y_k - y_0)/y_0$	y_k/y_0	$y_k - y_{k-1}$	$(y_k - y_{k-1})/y_{k-1}$	y_k/y_{k-1}

Średnie tempo zmian $\bar{i}_{t|t-1} = \sqrt[k-1]{i_{2|1} i_{3|2} \dots i_{k|k-1}}$

1. W celu zbadania dynamiki kosztów utrzymania w zakresie dwóch grup towarowych: nabiał oraz owoce ustalono, że towarami–reprezentantami będą: ser biały oraz jabłka. W badanym okresie cena sera nie uległa zmianie, natomiast cena jabłek wzrosła o 20%. Jak zmieniły się w rozpatrywanym okresie koszty utrzymania w zakresie dwóch grup towarowych, jeśli wiadomo, że udział wartości grupy towarowej „nabiał" w łącznej wartości sprzedaży obu grup zmalał z 48% w okresie podstawowym do 40% w okresie badanym? Zastosować w analizie odpowiednie formuły indeksowe.

2. Przedsiębiorstwo branży dziewiarskiej produkuje w trzech zakładach odpowiednio: w zakładzie A dzianinę (w metrach bieżących), w zakładzie B dresy (w sztukach) i w zakładzie C włóczkę bawełnianą (w kilogramach).

Zakład	Wielkość eksportu		Cena jednostkowa	
	1991	1992	1991	1992
dzianina	50	50	30	45
dresy	10	20	60	80
włóczka	20	15	20	40

Przeprowadzić wszechstronną analizę dynamiki eksportu obu produktów.

3. Zbadać dynamikę cen w latach 1972–1976 dla dwóch arytułów łącznie, jeżeli cena pierwszego nie zmieniła się, zaś drugiego zmalała o 10%. Wiadomo ponadto, że sprzedaż piewszego artykułu w ujęciu wartościowym była w 1976 roku trzykrotnie większa niż drugiego.

4. Zbadać dynamikę zbytu dwóch artykułów A i B jeżeli wiadomo, że cena artykułu A zmalała w okresie badanym w porównaniu z okresem podstawowym o 8%, natomiast artykułu B o 10% oraz, że udział wartości sprzedaży artykułu A wynosił 40% obrotów łącznych w okresie badanym wynoszących 2 mld zł.

5. Wartość skupu czterech podstawowych zbóż i dynamika wielkości dostaw w latach 1971 i 1973 podane są w poniższej tabeli:

Zboża	Wartość skupu w cenach bieżących (w mln zł)		Dynamika wielkości dostaw (1971 = 100)
	1971	1973	
Pszenica	8424	9124	102
Żyto	4557	6273	117
Jęczmień	3072	2988	95
Owies	951	520	48

Zbadać dynamikę skupu zbóż.

6. Zbadać dynamikę zmian zbiorów winogron w latach 1983 i 1986.

Odmiana	Ilość		Cena jednostkowa	
	1983	1986	1983	1986
A	1280	1360	108	111
B	830	890	93	101
C	1640	1660	97	107

7. Na pewnym targowisku ustalono, że wartość sprzedaży jaj wzrosła z 250 tys zł w 1984 roku do 500 tys zł w roku 1986, sera białego z 80 tys zł do 120 tys zł, zaś wartość sprzedaży śmietany zmalała z 60 tys zł do 30 tys zł. Wiadomo, że ilościowo sprzedaż jaj wzrosła o 30%, sera o 10%, natomiast śmietany zmalała dwukrotnie. Opisać dynamikę sprzedaży nabiału na tym targowisku.

Rozkład dwumianowy (Bernoulliego) $B(n, p)$. Zmienna losowa X ma rozkład $B(n, p)$, jeżeli

$$P_{n,p}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

$$P_{n,p}\{X = k\} = P_{n,1-p}\{X = n - k\}$$
$$EX = np,\ D^2X = np(1 - p)$$

Schemat Bernoulliego. Wykonujemy dwuwynikowe doświadczenie D. Wyniki nazywane są umownie *sukces* oraz *porażka*. Prawdopodobieństwo sukcesu wynosi p (porażki — $1 - p$). Doświadczenie wykonujemy w sposób niezależny n krotnie. Niech zmienną losową X będzie ilość sukcesów. Zmienna losowa X ma rozkład $B(n, p)$.

Rozkład dwumianowy jest stablicowany:

$$Q(k; n, p) = \sum_{i=k}^{n} P_{n,p}\{X = i\} = \sum_{i=k}^{n} \binom{n}{i} p^i (1 - p)^{n-i}$$

---

8. Partia dostarczanych detali ma wadliwość 5%. Niech zmienną losową będzie liczba dobrych detali spośród czterech dostarczonych. Obliczyć prawdopodobieństwo, że dostarczono co najmniej dwa detale dobre.

9. Z partii nasion o sile kiełkowania 75% losujemy osiem nasion. Obliczyć prawdopodobieństwo, że nie wykiełkuje co najmniej pięć nasion.

10. Obliczyć prawdopodobieństwo, że na siedem rzutów kostką co najwyżej trzy razy wypadnie liczba oczek nie mniejsza niż 4.

11. Załóżmy, że prawdziwa jest hipoteza Mendla, iż dla krzyżówki grochu w drugim pokoleniu stosunek nasion żółtych do zielonych jest jak 3 : 1. Wylosowano dziesięć nasion. Obliczyć prawdopodobieństwo, że będą co najwyżej cztery nasiona żółte.

12. Siła kiełkowania nasion wynosi 0.75. Obliczyć prawdopodobieństwo, że na dziesięć wysianych nasion wzejdzie co najmniej osiem nasion.

13. Środek owadobójczy zabija przeciętnie 90% owadów. Środek ten zastosowano na dziesięciu owadach. Obliczyć prawdopodobieństwo, że co najwyżej dwa osobniki przeżyją.

14. Wadliwość procesu produkcyjnego wynosi 10%. Obliczyć prawdopodobieństwo, że na osiem wylosowanych produktów będą co najwyżej dwa złe.

15. W pewnym gatunku zwierząt prawdopodobieństwo urodzenia osobnika płci męskiej wynosi 0.6. Obliczyć prawdopodobieństwo, że w miocie, w którym urodziło się pięcioro młodych będą co najmniej cztery osobniki męskie.

16. W stawie hodowlanym są dwa gatunki ryb w proporcji 8 : 2. Obliczyć prawdopodobieństwo, że wśród dziesięciu złowionych ryb będzie co najmniej siedem ryb liczniejszego gatunku.

17. W jeziorze jest 1000 ryb, w tym 100 ryb zaobrączkowanych. Obliczyć prawdopodobieństwo, że wśród 10 złowionych ryb będzie co najmniej siedem ryb zaobrączkowanych.

18. Przyjmując, że co czwarte wezwanie pogotowia jest nieuzasadnione określić prawdopodobieństwo, że na osiem wyjazdów co najmniej połowa z nich będzie uzasadniona.

19. Załóżmy, że co czwarty zawał kończy się zejściem oraz, że prawdopodobieństwo przeżycia zawału nie zależy od tego, który to zawał z kolei. Jakie jest prawdopodobieństwo przeżycia trzech zawałów?

20. Właściciel kurzej fermy stwierdził, że kogutków wykluwa się trzy razy więcej niż kurek. Obliczyć prawdopodobieństwo, że z pięciu losowo wybranych jajek wykluje się co najmniej jeden kogutek, ale nie mniej niż dwie kurki.

21. Producent podaje, że w co czwartym jajku niespodziance znajduje się zajączek Ribbon. Jakie jest prawdopodobieństwo, że wśród dwudziestu kupionych jajek jest a) przynajmniej pięć jajek z zajączkiem Ribbon; b) nie więcej niż piętnaście jajek bez zajączka. Jaka jest najbardziej prawdopodobna ilość jajek z zajączkami?

Rozkład normalny $N(\mu, \sigma^2)$. Zmienna losowa X ma rozkład normalny o wartości średniej μ i wariancji σ^2, jeżeli jej funkcja gęstości wyraża się wzorem

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \ -\infty < x < \infty.$$

$$EX = \mu, D^2X = \sigma^2.$$

Standardowy rozkład normalny: $N(0,1)$
Dystrybuanta $F(x)$ standardowego rozkładu normalnego jest stablicowana
Dla $x \leq 0$ zachodzi $F(x) = 1 - F(-x)$

Standaryzacja. Jeżeli $X \sim N(\mu, \sigma^2)$, to $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

$$P\{X \in (a,b)\} = P\left\{Z \in \left(\frac{a-\mu}{\sigma}, \frac{b-\mu}{\sigma}\right)\right\} = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

---

22. Niech X będzie zmienną losową o rozkładzie $N(5,4)$. Obliczyć prawdopodobieństwa

$$P\{X \leq 0\}, \ P\{X > 6\}, \ P\{X \in (2,7)\}.$$

23. Niech X będzie zmienną losową o rozkładzie $N(-3/2, 4)$. Obliczyć prawdopodobieństwa

$$P\{X < 2.5\}, \ P\{0.5 < X < 2\}, \ P\{|2X-1| < 1\}.$$

24. Niech X będzie zmienną losową o rozkładzie $N(-5, 100)$. Obliczyć prawdopodobieństwa

$$P\{X \in (-1,5)\}, \ P\{X \leq -9\}, \ P\{X \in (-7,1)\}, \ P\{X \geq -7\}, \ P\{|X+5| \leq 10\}.$$

25. Niech X będzie zmienną losową o rozkładzie $N(10, 25)$. Obliczyć prawdopodobieństwa

$$P\{X \in (12,15)\}, \ P\{X \leq 8\}, \ P\{X \in (9,13)\}, \ P\{X \geq 9\}, \ P\{|X-10| \leq 5\}.$$

26. Niech X będzie zmienną losową o rozkładzie $N(-10, 25)$. Obliczyć prawdopodobieństwa

$$P\{X \in (-8,-5)\}, \ P\{X \leq -12\}, \ P\{X \in (-11,-7)\}, \ P\{X \geq -11\}, \ P\{|X+10| \leq 5\}.$$

27. Niech X będzie zmienną losową o rozkładzie $N(5, 100)$. Obliczyć prawdopodobieństwa

$$P\{X \in (9,15)\}, \ P\{X \leq 1\}, \ P\{X \in (3,11)\}, \ P\{X \geq 3\}, \ P\{|X-5| \leq 10\}.$$

28. Wzrost kobiety jest zmienną losową o rozkładzie $N(160, 100)$. Obliczyć jaki jest procent kobiet wyższych niż 170 lub niższych niż 150.

29. Wzrost kobiety jest zmienną losową o rozkładzie $N(158, 100)$. Obliczyć jaki jest procent kobiet o wzroście pomiędzy 148 a 168.

30. Dzienna mleczność krowy jest zmienną losową o rozkładzie $N(10, 4)$. Jaki jest procent krów o mleczności między 7 a 12.

31. Dzienna mleczność krowy jest zmienną losową o rozkładzie $N(10, 4)$. Jaki jest procent krów o mleczności mniejszej niż 7, a jaki jest procent krów o mleczności większej niż 12.

32. Plon ziemniaka z poletka jest zmienną losową o rozkładzie $N(20, 25)$. Jaki procent poletek da plon większy niż 23, a jaki procent poletek da plon mniejszy niż 18.

33. Plon ziemniaka z poletka jest zmienną losową o rozkładzie $N(20, 25)$. Jaki procent poletek da plon między 18 a 23.

34. Automat tokarski produkuje śrubki, których średnica ma rozkład normalny ze średnią 3 mm i odchyleniem standardowym 0.05 mm. Śrubkę uważa się za dobrą, jeżeli jej średnica mieści się w przedziale $(2.9, 3.1)$. Jakie jest prawdopodobieństwo wyprodukowania dobrej śrubki?

35. Automat tokarski produkuje nity, których średnica ma rozkład normalny z odchyleniem standardowym 0.04 mm. Wartość średnia tej zmiennej losowej może być dowolnie regulowana przez odpowiednie ustawienie automatu. Nit uważa się za dobry, jeżeli jego średnica mieści się w przedziale $(2.9, 3.1)$. Jakie jest prawdopodobieństwo wyprodukowania braku, gdy automat tokarski ustawiony jest tak, że średnia średnica jest równa 3.05 mm? Jak powinien być ustawiony automat, by wadliwość procesu produkcyjnego była najmniejsza?

36. Automat produkuje nity. Średnica nitu jest zmienną losową o rozkładzie $N(2, 0.01)$. Dla jakiego ε średnica nitu z prawdopodobieństwem 0.95 znajduje się w przedziale $(2 - \varepsilon, 2 + \varepsilon)$?

37. Obliczyć odchylenie standardowe przyrządu pomiarowego o którym wiadomo, że z prawdopodobieństwem 0.95 daje błąd nie przekraczający trzech jednostek. Zakładamy, że rozkład błędu jest normalny z wartością średnią zero.

38. Cecha X o rozkładzie normalnym przyjmuje wartości z przedziału $\langle 28, 46 \rangle$ (przedział ten ustalony został mocy prawa trzech sigm). Wartości $40 - 46$ zaliczamy do pierwszej klasy, wartości $28 - 34$ do trzeciej klasy. Jakie jest prawdopodobieństwo tego, że losowo wybrana wartość cechy X należy do pierwszej lub trzeciej klasy?

39. Ciężar jabłek dostarczanych do skupu ma rozkład normalny ze średnią 8 dag i wariancją 9. Jaki procent jabłek dostarczanych do skupu nadaje się na eksport, jeżeli za jabłka eksportowe uważa się tylko te, które ważą więcej niż 11 dag.

40. Ciężar jajek dostarczanych do skupu ma rozkład normalny ze średnią 2 dag i wariancją 0.01. Jajko kwalifikuje się do pierwszego gatunku, jeżeli jego waga wynosi co najmniej 2.096 g. Jaki procent jajek dostarczanych do skupu można uznać za jajka pierwszego gatunku?

41. W pewnej okolicy wartość pH gleby jest zmienną losową o rozkładzie normalnym ze średnią 6.9 i wariancją 0.25. Czy w tej okolicy pszenica ozima ma większe szanse dawać bardziej optymalne plony niż burak cukrowy, jeżeli wiadomo, że graniczne wartości pH, przy których rośliny rozwijają się dobrze są następujące:
dla pszenicy ozimej: $6.3 - 7.6$,
dla buraka cukrowego: $7 - 7.5$?

42. Błąd pomiaru pewnym urządzeniem ma rozkład normalny o wartości średniej zero (pomiar nieobciążony) i odchyleniu standardowym 20 mm. Obliczyć prawdopodobieństwo, że błąd przynajmniej jednego z trzech niezależnych pomiarów nie przekroczy 4 mm.

43. Zawartość tłuszczu w mleku pewnej rasy krów ma rozkład normalny o wartości średniej 5% i wariancji 4. Mleko uważa się za bardzo tłuste, jeżeli zawartość tłuszczu przekracza 7%. Obliczyć prawdopododobieństwo, że przynajmniej jedna z trzech niezależnych próbek mleka będzie uznana za bardzo tłustą

44. Wzrost dzieci jest zmienną losową o rozkładzie normalnym o wartości średniej 110 cm i wariancji 400. Obliczyć prawdopodobieństwo, że przynajmniej jedno z trójki losowo wybranych dzieci będzie miało wzrost większy od przeciętnej.

45. Średnica nitu ma rozkład normalny o wartości oczekiwanej 2 mm i wariancji 0.01. Obliczyć prawdopodobieństwo, że przynajmniej wśród trzech losowo pobranych nitów wszystkie okażą się brakami, jeżeli za brak uważany jest nit o średnica mniejszej niż 1.8 mm lub większej niż 2.2 mm.

46. Ciężar jajka kurzego jest zmienną losową o rozkładzie normalnym o wartości średniej 10 g i wariancji 4. Obliczyć prawdopodobieństwo, że z trzech losowo wybranych jajek wszystkie będą miały ciężar poniżej przeciętnej.

47. Średni czas żarzenia się liści tytoniu wynosi 17 sekund, liście tlące się krócej niż 12 sekund są dyskwalifikowane. Jaki jest procent liści przydatnych do produkcji, jeżeli wariancja wspomnianej cechy wynosi 6.25?

48. Przyjmując, że przeciętna waga (w kilogramach) noworodka jest zmienną losową o rozkładzie $N(3, 0.25)$ określić jaki jest procent noworodków o wadze z przedziału $(3, 3.5)$.

49. Stwierdzono, że 80% ludzi o IQ powyżej 90 jest w stanie nauczyć się posługiwać pewnym urządzeniem. Jaki jest to procent całej populacji jeśli wiadomo, że iloraz inteligencji jest cechą o rozkładzie $N(100, 100)$?

50. Aby zdać egzamin ze statystyki należy prawidłowo rozwiązać co najmniej 70% zadań z testu egzaminacyjnego. Przyjmując, że wyniki testu dla studentów zdających w pierwszym terminie mają rozkład normalny ze średnią 76% i odchyleniem standardowym 8.0%, obliczyć jaki procent studentów zda egzamin w pierwszym terminie.

Estymacja parametrów cechy X o rozkładzie normalnym $N(\mu, \sigma^2)$

Próba (prosta): $X_1, X_2, \ldots, X_n$

Średnia arytmetyczna $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n}$

Suma kwadratów odchyleń od średniej $\mathrm{var}X = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n(\bar{X})^2 = \sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2$

Wariancja próbkowa $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{\mathrm{var}X}{n-1}$

Próbkowe odchylenie standardowe $S = \sqrt{S^2}$

Przedział ufności dla średniej w rozkładzie normalnym (wariancja σ^2 jest nieznana)

$$\left(\bar{X} - t(\alpha; n-1)\frac{S}{\sqrt{n}}, \quad \bar{X} + t(\alpha; n-1)\frac{S}{\sqrt{n}}\right)$$

---

51. Przeprowadzono anonimową ankietę wśród 26 uczniów najstarszych klas szkoły podstawowej. Celem ankiety było określenie ile uczniowie wydają tygodniowo na różnego rodzaju używki. Wyniki przedstawiają się następująco: $\sum x_i = 650$, $\sum x_i^2 = 16650$. Na podstawie wyników ankiety oszacować przeciętne tygodniowe wydatki uczniów na używki.

52. Wśród 17 losowo wybranych osób przeprowadzono ankietę na temat czasu dojazdu do pracy. Wyniki ankiety przedstawiały się następująco: $\sum x_i = 425$, $\sum x_i^2 = 10881$. Na podstawie wyników ankiety oszacować średni czas dojazdu do pracy.

53. Badano zawartość mięsa wołowego w hamburgerach. W dziewięcioelementowej próbie stwierdzono następujące wyniki w procentach wagowych: $\sum x_i = 180$, $\sum x_i^2 = 3618$. Oszacować przeciętną zawartość mięsa w hamburgerach.

54. Przy wycenie użytkowności mlecznej buhaja wzięto pod uwagę mleczność piętnastu jego córek w okresie pierwszych stu dni laktacji. Uzyskano następujące wyniki oznaczania zawartości białka (w procentach): $\sum x_i = 45.46$, $\sum(x_i - \bar{x})^2 = 0.240093$. Oszacować na tej podstawie średnią zawartość białka w mleku.

55. Przez dwieście dni obserwowano ilość znoszonych jaj przez szesnaście kur i uzyskano następujące wyniki: $n = 16$, $\sum x_i = 3408$, $\sum(x_i - \bar{x})^2 = 23718$. Na podstawie zebranych danych oszacować przeciętną dzienną liczbę jaj znoszonych przez kury.

56. Oszacować przeciętną procentową zawartość tłuszczu w mleku krów na podstawie danych o procentowej zawartości tłuszczu w mleku 50 krów: $\sum x_i = 191.91$, $\sum(x_i - \bar{x})^2 = 4.534138$.

57. Na podstawie zaobserwowanych mleczności dziesięciu krów oszacować średnią mleczność krowy: $\sum x_i = 100$, $\sum(x_i - \bar{x})^2 = 0.36$.

58. Notowano wielkość suchej masy dziesięciu roślin pewnej odmiany jęczmienia i uzyskano wyniki: $\sum x_i = 10$, $\sum(x_i - \bar{x})^2 = 0.36$. Oszacować przeciętną suchą masę rośliny.

59. Z dziesięciu poletek zebrano plon pewnego zboża i uzyskano wyniki: $\sum x_i = 200$, $\sum(x_i - \bar{x})^2 = 36$. Oszacować przeciętny plon z poletka.

60. Należy oszacować średnią żywotność (w godzinach świecenia) wyprodukowanej partii świetlówek. Wiadomo, że czas świecenia świetlówek ma rozkład normalny. Wylosowana z tej partii próba o liczności $n = 25$ świetlówek dała następujące rezultaty: $\sum x_i = 69990$, $\sum(x_i - \bar{x})^2 = 345600$.

61. W pewnym doświadczeniu medycznym bada się czas snu (w minutach) pacjentów chorych na pewną chorobę. Można przyjąć, że czas snu ma rozkład normalny. Należy oszacować średni czas snu pacjentów, jeżeli pomiary w grupie szesnastu pacjentów dały następujące rezultaty: $\sum x_i = 7172$, $\sum(x_i - \bar{x})^2 = 73500$.

62. Ciężar jajka kurzego jest zmienną losową o rozkładzie normalnym. Na podstawie następujących pomiarów oszacować średni ciężar jaja kurzego: $\sum_{i=1}^{10} x_i = 65$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 1.5$.

63. Wzrost dzieci jest zmienną losową o rozkładzie normalnym. Na podstawie następujących pomiarów oszacować średni wzrost dzieci: $\sum_{i=1}^{10} x_i = 1350$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 150$.

64. Średnica nitu ma rozkład normalny. Na podstawie następujących pomiarów oszacować średnią średnicę nitu: $\sum_{i=1}^{10} x_i = 15$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 1.5$.

65. Zawartość tłuszczu w mleku pewnej rasy krów ma rozkład normalny. Na podstawie następujących pomiarów oszacować średnią zawartość tłuszczu w mleku: $\sum_{i=1}^{10} x_i = 35$, $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 1.5$.

66. Oszacować przeciętną zawartość procentową skrobii w bulwach ziemniaka odmiany *Irys*, jeżeli w losowo wybranej dziewięcioelementowej próbie uzyskano następujące wyniki: $\sum x_i = 135$, $S^2 = 9$.

67. Badano wydatki studentów warszawskich na rozrywkę. W tym celu wylosowano 200 studentów i zanotowano ich wydatki. Okazało się, że średnie wydatki w tej grupie wyniosły 120 zł, zaś odchylenie standardowe 84 zł. Oszacować średnie miesięczne wydatki studentów warszawskich na rozrywkę.

Szacowanie prawdopodobieństwa sukcesu p

Niech X oznacza liczbę sukcesów w próbie n elementowej.
Dokładny przedział ufności ma postać

$$p \in (p_1(1-\alpha; X, n-X),\ p_2(1-\alpha; X, n-X)),$$

gdzie $p_2(1-\alpha; X, n-X) = 1 - p_1(1-\alpha; n-X, X)$, zaś liczby $p_1 = p_1(1-\alpha; k, n-k)$ są stablicowane. Końce przedziału ufności można również wyznaczać ze wzorów

$$\begin{aligned} p_1(1-\alpha; k, n-k) =& \frac{k}{k + (n-k+1)F(\frac{\alpha}{2}, 2(n-k+1), 2k)}\ , \\ p_2(1-\alpha; k, n-k) =& \frac{(k+1)F(\frac{\alpha}{2}, 2(k+1), 2(n-k))}{n-k+(k+1)F(\frac{\alpha}{2}, 2(k+1), 2(n-k))}\ . \end{aligned}$$

gdzie $F(\alpha; r_1, r_2)$ jest wartością krytyczną rozkładu F o (r_1, r_2) stopniach swobody

Przybliżony przedział ufności ma postać

$$p \in \left(\hat{p} - u_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + u_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right),$$

gdzie u_α jest kwantylem rzędu α rozkładu $N(0,1)$.

---

68. Wyprodukowano pewien nowy środek owadobójczy. Środek ten zastosowano na tysiącu owadach, z których 852 padły. Oszacować skuteczność tego środka owadobójczego.

69. Oszacować wadliwość procesu produkcyjnego wiedząc, że na 50 przebadanych wyrobów stwierdzono dwa braki.

70. Z pewnej partii owoców pobrano do badania 200 sztuk. Stwierdzono, że 60 jest zepsutych. Ocenić na tej podstawie procent owoców zepsutych w całej partii.

71. Na 800 zbadanych pacjentów pewnego szpitala 320 miało grupę krwi „O”. Oszacować odsetek pacjentów z tą grupą krwi zgłaszających się do szpitala.

72. Na 150 wylosowanych studentów pewnej Akademii Medycznej 114 stwierdziło, że systematycznie pali papierosy. Ocenić ogólny odsetek palaczy wśród studentów Akademii Medycznej.

73. Przeprowadzono ankietę dotyczącą poparcia dla pewnego ruchu społecznego. Wśród tysiąca ankietowanych 850 wyraziło poparcie. Ocenić odsetek ludności popierającej wspomniany ruch społeczny.

74. W pewnym zakładzie wybrano losowo dwadzieścia osób okazało się, że cztery z nich nigdy nie były na zwolnieniu chorobowym. Oszacować jaki odsetek pracowników tego zakładu nie korzystał ze zwolnienia lekarskiego?

75. Na sześćset przypadków wezwań pogotowia czterysta było uzasadnionych. Ocenić jaki procent wezwań pogotowia jest nieuzasadniony.

Testowanie hipotezy $H_0 : \mu = \mu_0$ w rozkładzie normalnym

Cecha X ma rozkład normalny $N(\mu, \sigma^2)$.

Test Studenta

$$t_{\text{emp}} = \frac{\bar{X} - \mu_0}{S}\sqrt{n}$$

Jeżeli $|t_{\text{emp}}| > t(\alpha; n-1)$, to hipotezę $H_0 : \mu = \mu_0$ odrzucamy.

---

76. Badano zawartość alkoholu w piwie pewnej marki. W siedemnastu zbadanych butelkach zawartość alkoholu w procentach objętościowych wyniosła: $\sum x_i = 76.5$, $\sum x_i^2 = 360.25$. Producent piwa twierdzi, że przeciętna zawartość alkoholu w piwie tej marki wynosi 5%. Czy wyniki badania dowodzą, że twierdzenie producenta nie jest pozbawione podstaw?

77. Przypuśćmy, że producent gwarantuje uzyskanie średnio 220 jaj od kury i przypuśćmy, że hodowca zakupił partię piskląt, z których uzyskał 25 kur z ukończoną niośnością. Średnia arytmetyczna niośności wyniosła 213.8 jaj, a odchylenie standardowe było równe 16 jaj. Zbadać, czy w warunkach fermy hodowcy możliwa jest do uzyskania gwarantowana przez producenta niośność.

78. Według norm, przebieg opon samochodowych powinien wynosić 35 tys. km. W celu sprawdzenia, czy nowy rodzaj opon spełnia wymagania normy, zbadano 200 opon i zanotowano ich przebiegi. Uzyskano wyniki: $\sum x_i = 7450$, $\sum(x_i - \bar{x})^2 = 6307.6831$. Czy można uznać, że nowy rodzaj opon spełnia wymagania normy?

79. W poszukiwaniu tzw. markerów genetycznych bada się zależność między takimi czynnikami jak grupa krwi czy typ białka krwi (*transferyny*) a poziomem cech produkcyjnych. W pewnej populacji krów o średniej wydajności mleka $\mu_0 = 3500$ kg zidentyfikowano 400 krów z transferyną EE. Średnia wydajność tych 400 krów wyniosła 3560 kg, a odchylenie standardowe 620 kg. Czy można uznać, że krowy z tą transferyną różnią się wydajnością od innych krów?

80. Dzienne zużycie wody w pewnej fabryce podlega wahaniom losowym. Przez 315 dni notowano zużycie wody i otrzymano wyniki: $\bar{x}_i = 1029$, $\sum(x_i - \bar{x})^2 = 59974$. Czy można przyjąć, że średnie dzienne zużycie wody wynosi 1000 m^3?

81. Czy można uznać, że przeciętny czas pracy bateryjki radiowej jest zgodny z normą wynoszącą 35 godzin, jeżeli dokonano pomiaru czasu pracy (w godzinach) piętnastu bateryjek radiowych i otrzymano wyniki: $\sum x_i = 484$, $\sum(x_i - \bar{x})^2 = 514.1304$?

82. Maszyna jest ustawiona w taki sposób, by produkowała kulki łożyskowe o średnicy 1 cm. Próba dziesięciu wylosowanych z produkcji kulek dała wyniki: $\sum x_i = 10.04$, $\sum(x_i - \bar{x})^2 = 0.000081$. Czy można uznać, że maszyna nie rozregulowała się w trakcie pracy?

83. W pewnym rejonie Polski uzyskiwano przeciętny plon pszenicy równy 22.6 q/ha. Chcemy sprawdzić, czy po zmianie sposobu uprawy przeciętny plon pszenicy ulegnie zmianie. W tym celu wylosowano z tego rejonu dwanaście gospodarstw i zbadano, jaki uzyskały one plon przy nowym sposobie uprawy. Uzyskano wyniki: $\sum x_i = 267.12$, $\sum(x_i - \bar{x})^2 = 0.9724$. Czy uzyskane wyniki świadczą o zmianie wielkości plonu przy nowym sposobie uprawy?

84. Automat produkuje blaszki o nominalnej grubości 0.04 mm. Z produkcji automatu wylosowano 25 blaszek i suma grubości tych blaszek wyniosła 0.925 mm, zaś odchylenie standardowe 0.04 mm. Czy można na podstawie tych wyników twierdzić, że grubość blaszek produkowanych przez automat jest równa grubości nominalnej.

85. Paczka powinna ważyć 1 kg. Kontroli poddano 64 paczki i okazało się, że średnia waga paczki w tej próbie wynosi 0.96 kg, zaś odchylenie standardowe jest równe 0.4 kg. Na poziomie istotności 0.05 sprawdzić, czy badana partia spełnia wymagania odnośnie do ciężaru paczki.

86. Badano średnią głębokość morza w pewnym rejonie. Otrzymano wyniki: suma czterech pomiarów wyniosła 3412 m, natomiast odchylenie standardowe było równe 8 m. Czy można twierdzić, że głębokość morza w tym rejonie wynosi 870 m.

Testowanie hipotezy $H_0 : p = p_0$

Cecha X ma rozkład dwupunktowy $D(p)$

Test przybliżony

$$u_{\text{emp}} = \frac{Y - np_0}{\sqrt{np_0(1-p_0)}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

gdzie Y oznacza liczbę sukcesów w próbie n–elementowej, natomiast $\hat{p} = Y/n$.
Jeżeli $|u_{\text{emp}}| > u_{1-\alpha/2}$, to hipotezę $H_0 : p = p_0$ odrzucamy.

---

87. Lider pewnej partii politycznej powiedział w wywiadzie, że jego partia ma poparcie 25% społeczeństwa. W odpowiedzi przytoczono wyniki ankiety przeprowadzonej wśród tysiąca osób. Spośród ankietowanych tylko 240 osób popierało wspomnianą partię. Czy wyniki ankiety dowodzą, że lider nie jest zorientowany w rzeczywistym poparciu dla swojej partii?

88. Wyprodukowano pewien nowy środek owadobójczy. Producent gwarantuje 90% skuteczności. Środek ten zastosowano na tysiącu owadach, z których 852 padły. Czy środek ma taką skuteczność jaką gwarantuje producent?

89. Czy można twierdzić, że wadliwość procesu produkcyjnego wynosi 2%, jeżeli na 50 przebadanych wyrobów stwierdzono dwa braki.

90. Czy można stwierdzić, że w transporcie psuje się 25% owoców, jeżeli na 200 przebadanych owoców było 60 zepsutych.

91. Na 800 zbadanych pacjentów pewnego szpitala 320 miało grupę krwi „O". Zweryfikować hiptezę, że procent pacjentów z tą grupą wynosi 35.

92. Na 150 wylosowanych studentów pewnej Akademii Medycznej 114 stwierdziło, że systematycznie pali papierosy. Zweryfikować hipotezę, że palących studentów jest 60%.

Testowanie hipotezy $H_0 : \mu_1 = \mu_2$ w dwóch rozkładach normalnych

Cecha X_1 ma rozkład normalny $N(\mu_1, \sigma_1^2)$.
Cecha X_2 ma rozkład normalny $N(\mu_2, \sigma_2^2)$.
Wariancje σ_1^2 oraz σ_2^2 są sobie równe: $\sigma_1^2 = \sigma_2^2$.
Test Studenta

$$t_{\text{emp}} = \frac{\bar{X}_1 - \bar{X}_2}{S_r}; \qquad S_r = \sqrt{S_e^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad S_e^2 = \frac{\text{var}X_1 + \text{var}X_2}{n_1 + n_2 - 2}\ .$$

Jeżeli $|t_{\text{emp}}| > t(\alpha, n_1 + n_2 - 2)$, to hipotezę $H_0 : \mu_1 = \mu_2$ odrzucamy.

---

93. W pewnym sklepie zważono jaja dostarczane przez dwóch różnych dostawców. Pobrano po dziesięć jaj od każdego dostawcy i otrzymano wyniki:
dostawca I: $\sum x_{1i} = 645$, $\sum x_{1i}^2 = 41615$; dostawca II: $\sum x_{2i} = 680$, $\sum (x_{2i} - \bar{x}_2)^2 = 10$.
Na podstawie uzyskanych wyników stwierdzić, czy średnie ciężary jaj dostarczane przez obu dostawców są takie same.

94. Badano zawartość tłuszczu w serach żółtych produkowanych zimą i latem. W każdym z dwóch okresów zbadano zawartość tłuszczu w dziesięciu serach i otrzymano wyniki:
zima: $\sum x_{1i} = 265$, $\sum x_{1i}^2 = 7034$; lato: $\sum x_{2i} = 276$, $\sum (x_{2i} - \bar{x}_2)^2 = 11$.
Na podstawie uzyskanych wyników stwierdzić, czy zawartość tłuszczu w serze żółtym zależy od pory roku.

95. Badano ciężar owoców jabłoni dwóch odmian. Z każdej odmiany pobrano po dziesięć owoców i otrzymano wyniki:
odmiana I: $\sum x_{1i} = 129$, $\sum x_{1i}^2 = 1664.6$; odmiana II: $\sum x_{2i} = 136$, $\sum (x_{2i} - \bar{x}_2)^2 = 0.4$.
Na podstawie uzyskanych wyników stwierdzić, czy odmiany różnią się pod względem średniego ciężaru owocu.

96. Spośród uczniów pewnego liceum wylosowano piętnastu z klas pierwszych oraz dwunastu z klas drugich i obliczono średnią ocen uzyskanych w semestrze dla każdego z uczniów. Otrzymano rezultaty:
klasy I: $\bar{x}_I = 3.70$, $s_I^2 = 0.32$; klasy II: $\bar{x}_{II} = 3.79$, $s_{II}^2 = 0.32$.
Zbadać, czy osiągnięcia klas pierwszych i klas drugich można uznać za takie same.

97. Średnia prędkość tramwaju (w km/h) obliczona na podstawie zmierzonych w środę prędkości 200 tramwajów była równa 15.1, natomiast średnia prędkość obliczona dla 120 tramwajów w niedzielę wynosiła 16.4. Wariancje prędkości obliczone na podstawie tych wyników wynosiły odpowiednio $s_1^2 = 6.8$, $s_2^2 = 4.3$. Na podstawie uzyskanych wyników zbadać, czy tramwaje jeżdżą tak samo szybko w środę i w niedzielę.

98. Dwóm grupom robotników zlecono wykonanie tej samej pracy z tym jednak, że robotnicy grupy pierwszej przeszli wcześniej przeszkolenie. Zaobserwowana wydajność pracy w pierwszej grupie kształtowała się następująco (w szt/h): 18.6, 17.9, 18.1, 17.0, 18.7, 18.3, podczas gdy w grupie drugiej zaobserwowano następujące wydajności: 17.3, 17.6, 17.1, 16.0, 17.8. Na poziomie istotności 0.05 sprawdzić, czy przeszkolenie zmieniło wydajność pracy robotnika.

99. W celu sprawdzenia, czy średni wzrost dwulatków w Warszawie i Łodzi jest taki sam zmierzono 400 dzieci warszawskich oraz 400 dzieci łódzkich. Okazało się, że średni wzrost dzieci warszawskich wynosi 86.5 cm, a odchylenie standardowe jest równe 8.5 cm. Odpowiednie wielkości dla dzieci łódzkich wyniosły: 87.5 cm oraz 5.2 cm. Na poziomie istotności 0.05 zweryfikować przypuszczenie o równości średnich wzrostów.

100. Ocenić, czy jest istotna różnica między wynikami egzaminu ze statystyki na dwóch wydziałach SGGW, jeżeli dla dwudziestu losowo wybranych studentów z wydziału A średnia ocen wyniosła 3.8, a odchylenie standardowe 0.3, zaś dla dwunastu studentów wydziału B otrzymano średnią 3.4 i wariancję 0.04.

101. Pobrano dwie dziesięcioelementowe próby, jedną dla traw, drugą dla roślin motylkowych, i zbadano zawartość procentową bezazotowych substancji wyciągowych. Uzyskano następujące wyniki:
Trawy: $\sum x_i = 451$, $\sum x_i^2 = 20352.1$; Motylkowe: $\sum x_i = 409$, $\sum x_i^2 = 16738.1$.
Czy trawy i rośliny motylkowe różnią się pod względem średniej zawartości procentowej bezazotowych substancji wyciągowych?

102. Kosiarka bębnowa Z105 przeznaczona jest do koszenia wszelkich upraw zielonych. Producent podaje, że kosiarka kosi na wysokości niskiej 32 cm oraz wysokiej 42 cm. Maszynę wykorzystywano do koszenia, traw i lucerny i uzyskano następujące wysokości pokosów:
Niski pokos: 29, 30.5, 28.5, 31, 31.2, 32.3;
Wysoki pokos: 43.2, 39.1, 32.8, 35.5, 41.8, 40.5.
Czy zapewnienia producenta można uznać za słuszne?

Testowanie hipotezy $H_0 : p_1 = p_2$

test przybliżony

Cecha X ma w pierwszej populacji rozkład dwupunktowy $D(p_1)$.

Cecha X ma w drugiej populacji rozkład dwupunktowy $D(p_2)$.

Niech $\hat{p}_1 = k_1/n_1,\ \hat{p}_2 = k_2/n_2$ oraz $\hat{p} = (k_1 + k_2)/(n_1 + n_2)$. Statystyka testowa ma postać:

$$u_{\text{emp}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}\ .$$

Jeżeli $|u_{\text{emp}}| \geq u_{1-\alpha/2}$, to hipotezę odrzucamy.

---

103. Wysunięto przypuszczenie, że jakość produkcji pewnego wyrobu po wprowadzeniu nowej, tańszej technologii nie uległa zmianie. Wylosowano próbę 120 sztuk tego wyrobu spośród wyprodukowanych starą technologią i otrzymano 12 sztuk złych. Wśród 160 wylosowanych sztuk wyprodukowanych nową technologią było 20 sztuk wadliwych. Czy wysunięte przypuszczenie można w świetle uzyskanych wyników uznać za uzasadnione?

104. W pewnej szkole rozeszła się plotka, że uczniowie chcą ogolić dyrektora. Nauczyciel matematyki zapytał o to 150 dziewcząt i 200 chłopców. Dziewięćdziesięciu chłopców i 70% dziewcząt odpowiedziało twierdząco. Czy można uznać, że chęć ogolenia dyrektora zależy od płci?

105. Na 200 przebadanych szczurów u 60 stwierdzono objawy obniżonego refleksu. Wśród chorych szczurów tylko 20 dostawało pewien preparat P, a wszystkich szczurów karmionych tym preparatem było 80. Czy można uznać, że karmienie preparatem P wpływa na obniżenie refleksu u szczurów?

106. Badano, czy młodzież męska nosząca modną fryzurę ma inne wyniki niż pozostali młodzieńcy. W tym celu zbadano 492 uczniów i okazało się, że wśród 94 modnych młodzieńców aż 51 miało złe wyniki w nauce. Wszystkich źle uczących się było 245. Czy na tej podstawie można sądzić, że fryzura ma wpływ na wyniki nauczania?

107. Na 180 przebadanych studentów i studentek stwierdzono, że 100 zdało egzamin ze statystyki. Wśród 130 studentek egazminu nie zaliczyło 55. Czy można na tej podstawie uznać, że wynik egzaminu zależy od płci zdającego?

108. Wysunięto przypuszczenie, że pacjenci z objawem klinicznej niewydolności oddechowej charakteryzują się zawyżonym poziomem aktywności pewnego enzymu. Wśród 357 pacjentów zanotowano 49 z niewydolnością oddechową, natomiast wśród 43 pacjentów o podwyższonej aktywności enzymu niewydolnością oddechową charakteryzowało się 18. Czy na tej podstawie wysunięte przypuszczenie można uznać za uzasadnione?

Porównanie wartości średnich
Rozpatrujemy k populacji oraz cechę X.
X_i — cecha X w i–tej populacji
Cecha X_i ma rozkład normalny $N(\mu_i, \sigma_i^2)$
Średnie μ_i oraz wariancje σ_i^2 są nieznane
$\sigma_1^2 = \cdots = \sigma_k^2$
Próba z cechy X_i: $X_{i1}, \ldots, X_{in_i}$

$$H_0 : \mu_1 = \cdots = \mu_k$$

Test F (poziom istotności α)
Statystyka testowa:

$$F_{\text{emp}} = \frac{S_a^2}{S_e^2}$$

gdzie

$$S_a^2 = \frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2,\ S_e^2 = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2,$$

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij},\ \bar{X} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij},\ N = \sum_{i=1}^{k} n_i$$

Jeżeli $F_{\text{emp}} > F(\alpha; k-1, N-k)$, to hipotezę $H_0 : \mu_1 = \cdots = \mu_k$ odrzucamy.
Wniosek praktyczny: przynajmniej jedna ze średnich $\mu_1, \ldots, \mu_k$ jest inna od pozostałych
Tabela analizy wariancji

Źródło zmienności	Stopnie swobody	Sumy kwadratów	Średnie kwadraty	F_{emp}
Czynnik	$k-1$	$\text{var}A$	$S_a^2 = \frac{\text{var}A}{k-1}$	S_a^2/S_e^2
Błąd losowy	$N-k$	$\text{var}E$	$S_e^2 = \frac{\text{var}E}{N-k}$	
Ogółem	$N-1$	$\text{var}T$		

$$\text{var}A = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2,\ \text{var}E = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2,\ \text{var}T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^2 - N(\bar{X})^2,$$

$$\text{var}A + \text{var}E = \text{var}T$$

Grupy jednorodne — podzbiory średnich, które można uznać za takie same
Procedury porównań wielokrotnych — postępowanie statystyczne zmierzające do podzielenia zbioru średnich na grupy jednorodne

109. Badano średnią zawartość (w ppm na kg s.m.) żelaza w roślinach łąkowych. Uzyskano następujące wyniki

dla traw	174.0	172.0	175.0	173.0	177.0	175.5
dla turzycowatych	134.0	135.0	137.0	138.0	135.0	135.2
dla motylkowych	117.0	118.0	116.0	119.0	116.3	116.5
dla ziół (chwastów)	111.0	112.0	113.0	111.5	112.7	114.2

Sformułować i zweryfikować odpowiednią hipotezę.

110. Sprawdzano ceny jednego kwiatu róży ogrodowej w trzech różnych miastach: M, W, P. Czy poniższe dane udowadniają zależność ceny róży od miasta?

miasto	średnia cena (z dziesięciu powtórzeń)
M	10.5
W	9.1
P	8.3

$$s_e^2 = 0.1$$

111. W metodzie Kjeldahla oznaczania procentowej zawartości białka stosowany jest katalizator miedziowy. Podejrzewa się, że wynik oznaczenia zależy od ilości dodanego katalizatora. W tym celu przeprowadzono doświadczenie, w którym badano trzy różne ilości katalizatora. Czy poniższe dane udowadaniają zależność wyniku oznaczenia od ilości katalizatora miedziowego?

ilość katalizatora	średni wynik (z pięciu pomiarów)
1.0	18.48
2.0	18.44
5.0	18.02

$$s_e^2 = 6.408$$

112. Liczba błędów popełnionych przez tresowane szczury w toku przejścia przez labirynt ma rozkład normalny. Do pewnych doświadczeń wylosowano po pięć szczurów do czterech grup, które powinny być jednorodne pod względem stopnia wytresowania. Otrzymano dla tych szczurów następujące liczby popełnianych błędów:

I	II	III	IV
10	7	8	16
8	10	13	10
7	6	15	8
6	14	6	10
11	5	3	4

Czy można na podstawie uzyskanych wyników sądzić, że badane grupy są jednakowo wytresowane?

113. *Hipoteza merytoryczna.* Zastąpienie sacharozy innymi środkami słodzącymi w odmienny sposób wpływa na występowanie próchnicy zębów.
Przebieg doświadczenia. Samice szczurów w wieku sześciu tygodni podzielono losowo na cztery grupy po dwanaście osobników, którym podawano tą samą dietę hodowlaną. Poszczególnym grupom do paszy dodawano codziennie w ilości $50ml/100g\ paszy$ wodne roztwory różnych środków słodzących o równoważnym natężeniu smaku słodkiego, odpowiednio 30% roztwór sacharozy, 40% roztwór glukozy, 60% roztwór glicyny i 30% roztwór sorbitolu. Po trzech miesiącach żywienia zwierzęta uśmiercono i w wypreparowanych im szczękach oraz żuchwach oceniano odsetek zębów objętych próchnicą. Wyniki doświadczenia były następujące:

sacharoza: 40.3, 37.8, 32.1, 48.2, 39.1, 36.7, 44.3, 38.3, 40.1, 37.2, 30.1, 40.2
glukoza: 17.1, 10.2, 14.4, 16.0, 14.1, 9.7, 12.3, 13.1, 7.6, 11.5, 15.8, 14.3
glicyna: 21.7, 18.8, 18.7, 14.5, 22.8, 19.3, 24.8, 23.3, 19.7, 19.5, 20.8, 18.3
sorbitol: 16.2, 14.1, 16.0, 10.5, 14.0, 8.5, 7.8, 12.5, 12.8, 15.8, 14.2, 13.7

Sprawdzić postawioną hipotezę merytoryczną. W przypadku jej odrzucenia przeprowadzić analizę szczegółową.

114. Do pewnych doświadczeń farmakologicznych hodowane są cztery grupy królików. Czy poniższe dane udowadniają zróżnicowanie pomiędzy wymienionymi grupami hodowlanymi królików pod względem średniego ciężaru?

grupa	średni ciężar (z czterech królików)
I	37
II	30
III	35
IV	34

$$s_e^2 = 8$$

115. W czterech ulach zmierzono średnice plastrów zbudowanych przez pszczoły. W każdym ulu wykonano dziesięć pomiarów. Otrzymano następujące wyniki:

$$\bar{x}_1 = 5.6, \quad \bar{x}_2 = 5.4, \quad \bar{x}_3 = 5.1, \quad \bar{x}_4 = 5.5, \quad \sum x_{ij}^2 = 1170.$$

Czy można uznać,że przeciętne średnice we wszystkich czterech ulach są jednakowe?

116. Trzech nauczycieli A, B, C statystyki oceniało w skali punktowej $1 - 20$ prace czterech wylosowanych uczniów. Wyniki były następujące:

A	B	C
19	17	20
20	20	19
10	11	9
14	15	12

Czy można uznać, że wszyscy trzej nauczyciele są jednakowi w swoich ocenach?

117. W doświadczeniu wazonowym badano wpływ wzrastających dawek nawożenia azotowego na plon pszenicy pewnej odmiany. Uzyskano wyniki

Dawka N	Plony z poszczególnych wazonów					Sumy
0	1.01	1.21	1.11	1.15	1.02	5.50
3	2.03	2.23	2.41	2.22	2.11	11.00
6	2.64	2.51	2.32	2.72	2.31	12.50
12	2.81	2.71	2.94	2.60	2.44	13.50

Sformułować i zweryfikować odpowiednią hipotezę.

118. W pewnym doświadczeniu badano plon pewnej odmiany pszenicy dla trzech różnych kombinacji nawozowych: A, B, C. Czy poniższe dane udowadniają różny wpływ kombinacji nawozowych na plon pszenicy?

kombinacja	średni plon (z pięciu poletek)
A	58
B	60
C	62

$$s_e^2 = 4$$

119. Porównywano rezultaty w rzucie oszczepem dla czterech zawodników. Zmierzono po dziesięć wyników dla każdego zawodnika. Otrzymano następujące wyniki:

$$\bar{x}_1 = 88,\ \bar{x}_2 = 94,\ \bar{x}_3 = 91,\ \bar{x}_4 = 89,\ \sum x_{ij}^2 = 328135.$$

Czy można uznać,że przeciętne wyniki wszystkich zawodników są jednakowe?

120. W pewnym doświadczeniu fizycznym przeprowadzonym trzema metodami A, B, C badano czas pewnego efektu świetlnego występującego w tym doświadczeniu. Czy można uznać, że średni czas występowania tego efektu jest dla wszystkich metod przeprowadzania doświadczenia taki sam? Wyniki eksperymentu:

metoda	średni czas (z pięciu powtórzeń)
A	3.22
B	3.24
C	3.14

$$s_e^2 = 0.004$$

121. Mierzono czas świecenia trzech typów żarówek. Na podstawie poniższych danych stwierdzić, czy można uznać, że średni czas świecenia tych trzech typów żarówek jest taki sam?

typ żarówki	średni czas świecenia (z pięciu żarówek)
I	1864.8
II	1776.0
III	1827.8

$$s_e^2 = 8405.5$$

122. Porównywano działanie trzech leków podawanych świniom chorym na różycę (skuteczność działania danego leku mierzono czasem trwania kuracji). Każde lekarstwo zostało zaaplikowane pięciu chorym zwierzętom. Uzyskano następujące wyniki:

Lekarstwo	Średni czas trwania kuracji
L_1	16
L_2	20
L_3	21

$$s_e^2 = 5$$

Czy na podstawie powyższych danych można przyjąć, że te trzy lekarstwa dają jednakowy efekt?

123. Wysunięto przypuszczenie, że czas przeznaczony na kolokwium ma wpływ na jego wynik. Chcąc stwierdzić czy tak jest pobrano następującą próbę:

Czas w minutach	Wyniki kolokwium w punktach
Od 15 do 20	$11; 12; 13; 12; 10$
Od 20 do 30	$15; 18; 21; 23; 24$
Od 30 do 60	$30; 25; 32; 28; 20$

Czy powyższe dane pozwalają odpowiedzieć na pytanie, czy czas przeznaczony na kolokwium ma wpływ na jego wyniki?

124. Chcąc sprawdzić, czy stabilność pracy maszyny zależy od czasu jej eksploatacji wykonano pewne pomiary i uzyskano następujące wyniki:

Czas eksploatacji w dniach	Stabilność pracy (w g)
Od 15 do 20	$250; 251; 253; 249; 251$
Od 30 do 60	$260; 263; 259; 273; 264$
Od 80 do 160	$285; 290; 288; 283; 284$

Czy powyższe dane pozwalają odpowiedzieć na pytanie, że awaryjność maszyny zależy od czasu jej eksploatacji?

125. Chcąc stwierdzić czy czas świecenia pewnego typu żarówek zależy od wielkości bańki szklanej przeprowadzono pewne pomiary:

Wielkość bańki	Czas świecenia (wminutach)
Mniej niż 5 cm średnicy	$10; 15; 18; 18; 12; 15; 13; 15; 12; 18; 17$
Więcej niż 5, ale mniej niż 10 cm średnicy	$18; 19; 20; 19; 22; 25; 19; 22; 25; 25; 24$
Więcej niż 10 cm średnicy	$38; 40; 45; 44; 42; 41; 43; 40; 35; 30; 29$

Czy powyższe dane pozwalają stwierdzić, czy wielkość bańki szklanej ma wpływ na czas świecenia żarówki?

126. W pewnej miejscowości położonej blisko trasy szybkiego ruchu, kierownik mleczarni stwierdził, że rolnicy pasą krowy w przydrożnych rowach. Jak wiadomo zawartość metali ciężkich jest większa w roślinach rosnących przy drodze. Zbadano po dziesięć próbek mleka od dostawców A, B i C. Otrzymano następujące średnie zawartości metali ciężkich: $A = 4.41,\ B = 3.56,\ C = 4.35$. Ponadto obliczono $\mathrm{var}E = 2.13$. Czy można na tej podstawie stwierdzić, który z dostawców pasie krowy przy szosie?

127. Właściciel palarni kawy twierdzi, że wszystkie gatunki kawy które produkuje mają podobną zawartość kofeiny. W celu udowodnienia tej hipotezy wybrano trzy mieszanki kawy i po poddaniu ich procesowi palenia, uzyskano następujące zawartości kofeiny (w mg w jednej filiżance kawy) :

gatunek 1	gatunek 2a	gatunek 3
122.24	117.47	129.96
122.49	123.56	114.32
123.22	112.53	118.30
123.09	114.54	115.65
120.43	118.71	127.74
121.26	124.17	128.02

W oparciu o powyższe dane odpowiedzieć na pytanie czy właściciel palarni ma rację?

Badanie zależności między cechami (współczynnik korelacji i regresja)

Obserwujemy parę cech ilościowych (X, Y)
Niech $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ będzie próbą
Suma iloczynów odchyleń

$$\mathrm{cov}(X, Y) = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} = \sum_{i=1}^{n} X_i Y_i - \frac{1}{n} \left(\sum_{i=1}^{n} X_i \right) \left(\sum_{i=1}^{n} Y_i \right)$$

Kowariancja z próby

$$\mathrm{C}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{\mathrm{cov}(X, Y)}{n-1}$$

Współczynnik korelacji z próby (próbkowy)

$$r = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}X}\sqrt{\mathrm{var}Y}}$$

(X, Y) ma dwuwymiarowy rozkład normalny

$$H_0 : \text{Cechy } X \text{ oraz } Y \text{ są niezależne} \Leftrightarrow H_0 : \varrho = 0$$

test współczynnika korelacji Pearsona
Statystyka $r_{\mathrm{emp}} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}X}\sqrt{\mathrm{var}Y}}$ Wartość krytyczna $r(\alpha; n)$ (dwustronna)
Jeżeli $|r_{\mathrm{emp}}| > r(\alpha; n)$, to hipotezę H_0 odrzucamy.

Ilościowy opis zależności
Obserwujemy parę cech ilościowych (X, Y)
(X, Y) ma dwuwymiarowy rozkład normalny
Zakładamy niezerowość współczynnika korelacji: $\varrho \neq 0$
Ilościowy opis zależności Y od X (**funkcja regresji**): $E(Y|X = x) = ax + b$
Ocena parametrów funkcji regresji
Próba $(X_1, Y_1), \ldots, (X_n, Y_n)$

$$\hat{a} = \frac{\mathrm{cov}(X, Y)}{\mathrm{var}X} \qquad \hat{b} = \bar{Y} - \hat{a}\bar{X}$$

Resztowa suma kwadratów $\mathrm{var}R = \mathrm{var}Y(1 - r^2) = \mathrm{var}Y - \frac{(\mathrm{cov}(X,Y))^2}{\mathrm{var}X} = \mathrm{var}Y - \hat{a}\mathrm{cov}(X, Y)$
Wariancja resztowa $S_{y \cdot x}^2 = \frac{\mathrm{var}R}{n-2}$
Przedział ufności dla współczynnika regresji (poziom ufności $1 - \alpha$)

$$a \in (\hat{a} - t(\alpha; n-2)S_a; \hat{a} + t(\alpha; n-2)S_a)$$

gdzie $S_a^2 = \frac{S_{y \cdot x}^2}{\mathrm{var}X}$

128. Badano zależność między ilością godzin przebywania samolotu w powietrzu („nalot” lotniczy) a ilością wypadków. Na podstawie zebranych danych z pewnego okresu czasu zbadać czy taka zależność istnieje, a jeżeli istnieje, to opisać ją za pomocą liniowej funkcji regresji.

„nalot”	167	167	155	118	159	141	130	131	141	135	145	147	128	137	138
wypadki	135	129	118	134	120	109	92	129	138	132	138	135	111	123	99

129. Zbadano w dziesięciu wylosowanych zakładach przemysłowych wielkość zaplanowanego i wykonanego funduszu na akcję socjalną. Zabadać, czy istnieje zależność między badanymi cechami. Jeżeli taka zależność istnieje, to opisać ją za pomocą liniowej funkcji regresji.

planowany	3.60	4.65	5.20	1.86	3.06	1.36	2.46	3.93	5.80	6.35
wykonany	3.56	4.59	5.13	1.84	3.02	1.35	2.43	3.89	5.72	6.26

130. Istnieje podejrzenie że ludzie dzielą się na humanistów i matematyków tzn. jeśli ktoś jest dobry z przedmiotów humanistycznych to z matematyką może już mieć problemy. Wylosowano ośmiu uczniów z czwartej klasy liceum i obliczono dla nich średnie z ocen semestralnych z języka polskiego i z matematyki :

J. polski	3.4	2.9	4.3	3.8	3.3	4.7	3.6	3.9
Matematyka	4.6	4.4	3.6	3.2	3.8	3.5	4.6	3.1

Interesuje nas, czy oceny uczniów potwierdzają wspomniane twierdzenie. Jeśli tak to proszę wyznaczyć równanie regresji. Jaki jest przewidywana średnia ocena z matematyki dla ucznia którego średnia z języka polskiego wynosi 5.0?

131. Poniższe dane z dziesięciu poletek dotyczą efektywności nawożenia łąk azotem (w kg siana na 1 kg N) w zależności od poziomu nawożenia azotem:

$$\bar{x} = 40, \quad \bar{y} = 16, \quad \sum x_i^2 = 20200, \quad \sum y_i^2 = 2564.625, \quad \sum x_i y_i = 6295.$$

Zbadać, czy istnieje zależność między cechami. Wyznaczyć liniową funkcję zależności przeciętnej efektywności nawożenia łąk azotem od poziomu nawożenia.

132. Poniższe dane dotyczą ciężaru owoców (Y) pewnej rośliny oraz ilości zastosowanego pewnego preparatu (X) na dziesięciu poletkach:

$$\bar{x} = 3.9, \quad \bar{y} = 3.6, \quad \sum x_i^2 = 152.82, \quad \sum y_i^2 = 131.22, \quad \sum x_i y_i = 141.21.$$

Zbadać, czy istnieje zależność między cechami. Wyznaczyć liniową funkcję zależności przeciętnego ciężaru owoców od ilości zastosowanego preparatu.

133. Zbadać, czy istnieje korelacja między wielkością produkcji X pewnego artykułu (w mln metrów), a zużyciem Y pary technologicznej (w tys. ton). Dane pochodzą z dziesięciu wylosowanych zakładów.

x_i	1.5	3.0	2.0	3.5	1.5	4.5	2.5	4.0	4.5	3.0
y_i	4.5	7.0	7.5	6.5	6.5	7.5	5.5	4.5	5.5	5.0

Jeżeli zależność istnieje, to opisać ją za pomocą liniowej funkcji regresji.

134. W pewnym gospodarstwie wiejskim badano w ciągu dziesięciu kolejnych lat przeciętne dzienne spożycie zmieniaków w kg (X) i wielkość spożycia artykułów zbożowych w kg (Y) przypadającą na jednego członka rodziny. Zbadać, czy istnieje zależność między cechami X oraz Y. Jeżeli zależność istnieje, to opisać ją za pomocą liniowej funkcji regresji.

x_i	0.70	0.60	0.80	0.85	0.55	0.65	0.90	1.00	0.75	0.50
y_i	0.50	0.70	0.50	0.40	0.75	0.60	0.30	0.20	0.55	0.70

135. Na podstawie poniższych danych zbadać, czy istnieje zależność między zawartością tłuszczu (X) i białka (Y) w mleku krów.

$$\sum_{i=1}^{10} x_i = 38.6, \quad \sum_{i=1}^{10} x_i^2 = 150.16, \quad \sum_{i=1}^{10} y_i = 36.2, \quad \sum_{i=1}^{10} y_i^2 = 131.74, \quad \sum_{i=1}^{10} x_i y_i = 140.34$$

Jeżeli tak, to opisać ilościowo tę zależność.

136. Na podstawie poniższych danych zbadać, czy istnieje zależność między grubością włókna lnu (X) i grubością łodygi (Y).

$$\sum_{i=1}^{8} x_i = 83.5, \quad \sum_{i=1}^{8} x_i^2 = 1269.25, \quad \sum_{i=1}^{8} y_i = 8.74, \quad \sum_{i=1}^{8} y_i^2 = 18.5454, \quad \sum_{i=1}^{8} x_i y_i = 145.780$$

Jeżeli tak, to opisać ilościowo tę zależność.

137. Wyniki ogólnopolskiego konkursu maszynopisania dostarczyły informacji o liczbie błędów popełnionych przez maszynistki zależnie od czasu pisania tego samego tekstu. Podane zestawienie uwzględnia odpowiednie informacje dla dziesięciu wylosowanych maszynistek:

Czas	10	12	7	8	9	11	10	13	9	10
Liczba błędów	3	0	5	4	4	0	1	0	3	2

Zbadać istnienie zależności między liczbą popełnianych błędów a czasem pisania. Opisać tę zależność za pomocą liniowej funkcji regresji.

138. Spośród studentów pewnego wydziału wylosowano niezależnie dziesięciu studentów IV roku i otrzymano dla nich następujące średnie oceny uzyskane na I roku oraz na IV roku. Zabadać, czy istnieje zależność między wynikami studiów na I i na IV roku. Jeżeli taka zależność istnieje, to opisać ją za pomocą liniowej funkcji regresji.

I rok	3.5	4.0	3.8	4.6	3.9	3.0	3.5	3.9	4.5	4.1
IV rok	4.2	3.9	3.8	4.5	4.2	3.4	3.8	3.9	4.6	4.0

139. Badano zależność między miesięcznym dochodem na jednego członka rodziny (X) a wyrażoną w procentach częścią budżetu domowego przeznaczoną na zakup artykułów żywnościowych (Y). Na podstawie badania dziesięciu rodzin otrzymano następujące wyniki:

$$\bar{x} = 2.375, \sum x_i^2 = 60.9375, \bar{y} = 77, \sum y_i^2 = 66600, \sum x_i y_i = 1887.5$$

Zbadać, czy istnieje zależność między dochodem a wydatkami na żywność. Jeżeli tak, to dopasować liniową funkcję regresji.

140. Badano zależność między wysokością nad pewnym punktem terenowym (X), a ciśnieniem atmosferycznym (Y). Wykonano dziesięć pomiarów i otrzymano następujące wyniki:

$$\sum x_i = 45, \ \sum x_i^2 = 285, \ \sum y_i = 6060, \ \sum y_i^2 = 3672690, \ \sum x_i y_i = 27237.$$

Zbadać istnienie zależności między wysokością a ciśnieniem oraz wyznaczyć liniową funkcję regresji opisującą zależność ciśnienia od wysokości, na której dokonywany jest pomiar.

141. W celu zbadania zależności między kątami nachylenia terenu a wielkością błędów wysokościowych popełnianych w pewnej metodzie aerotriangulacji dokonano dziewięciu pomiarów. Zabadać, czy istnieje zależność między badanymi cechami. Jeżeli taka zależność istnieje, to opisać ją za pomocą liniowej funkcji regresji.

kąt	0.157	0.140	0.122	0.105	0.087	0.070	0.052	0.035	0.017
błąd	0.111	0.097	0.183	0.215	0.214	0.209	0.200	0.178	0.225

142. Do uprawy pewnej rośliny użyto nawozów sztucznych o różnej zawartości jednego składnika. Obserwowano procentową zawartość tego składnika (X) oraz przeciętny ciężar masy zielonej jednej rośliny w gramach (Y) uprawianej tym nawozem.

x_i	1.1	2.1	3.0	3.8	5.3	6.1	7.0	8.0
y_i	2.6	4.4	3.9	6.1	5.8	7.1	6.6	7.1

Zbadać istnienie zależności między cechami oraz wyznaczyć równanie regresji prostoliniowej.

143. W pewnej miejscowości dokonano sześciu pomiarów temperatury (Y) dla różnych głębokości pod powierzchnią ziemi (X). Otrzymano następujące wyniki:

$$\sum x_i = 4200, \quad \sum x_i^2 = 3640000, \quad \sum y_i = 144, \quad \sum y_i^2 = 3988, \quad \sum x_i y_i = 120000.$$

Zbadać istotność tej zależności. Wyznaczyć liniową funkcję regresji opisującą zależność temperatury od głębokości pod powierzchnią ziemi.

144. Poziom zatrudnienia X (w tys. osób) oraz wielkość produkcji Y (w tys. ton) w ośmiu wylosowanych przedsiębiorstwach przemysłowych kształtowały się następująco:

$$\sum x_i = 10.2, \quad \sum y_i = 22.1, \quad \sum x_i^2 = 13.42, \quad \sum y_i^2 = 62.71, \quad \sum x_i y_i = 29.$$

Zbadać istnienie zależności między poziomem zatrudnienia a wielkością produkcji. Opisać ilościowo tę zależność przy pomocy liniowej funkcji regresji. Zinterpretować oszacowany przedziałowo współczynnik kierunkowy regresji. Oszacować ile wynosi przeciętny poziom produkcji zakładów zatrudniających 1.2 tys. pracowników.

145. Zbiorowość pracowników pewnego dużego przędsiębiorstwa postanowiono przebadać ze względu na zależność wydajności pracy (Y) od stażu pracy (X). W tym celu wylosowano 20 pracowników i uzyskano następujące wyniki:

$$\sum x_i = 204, \quad \sum y_i = 442, \quad , \sum x_i^2 = 5000, \quad \sum y_i^2 = 25000, \quad \sum x_i y_i = 10000.$$

Zbadać istotność zależności między wydajnością a stażem pracy. Opisać ilościowo tę zależność przy pomocy liniowej funkcji regresji. Zinterpretować oszacowany przedziałowo współczynnik kierunkowy regresji. Oszacować ile wynosi przeciętna wydajność pracowników o stażu pracy $x = 10.2$. Zinterpretować uzyskany wynik.

Badanie zależności między cechami (testy niezależności)

Obserwujemy parę cech (X, Y)

$$H_0 : \text{cechy } X \text{ oraz } Y \text{ są niezależne}$$

Test chi–kwadrat niezależności

W celu zastosowania tego testu próba musi być zapisana w postaci szeregu rozdzielczego:

Klasy cechy Y	Klasy cechy X: 1	2	$\dots$	m
1	n_{11}	n_{12}	$\dots$	n_{1m}
2	n_{21}	n_{22}	$\dots$	n_{2m}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
k	n_{k1}	n_{k2}	$\dots$	n_{km}

Statystyka testowa

$$\chi^2_{\text{emp}} = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(n_{ij} - n^t_{ij})^2}{n^t_{ij}} ,$$

gdzie

$$n^t_{ij} = n_{i\cdot} n_{\cdot j}/N, \ n_{i\cdot} = \sum_{j=1}^{m} n_{ij}, \ n_{\cdot j} = \sum_{i=1}^{k} n_{ij}, \ N = \sum_{i=1}^{k} \sum_{j=1}^{m} n_{ij}$$

Jeżeli $\chi^2_{\text{emp}} > \chi^2_{\alpha,(k-1)(m-1)}$, to hipotezę H_0 : *Cechy X oraz Y są niezależne* odrzucamy.
Liczby $n_{i\cdot}$, $i = 1, \dots, k$ oraz $n_{\cdot j}$, $j = 1, \dots, m$ noszą nazwę *próbkowych rozkładów brzegowych* cech odpowiednio Y oraz X

Test współczynnika korelacji rangowej Spearmana

Obserwacjami są pary liczb $(X_i, Y_i), i = 1, \dots, n$. Każdej obserwacji X_i nadajemy rangę R_i, tzn. jej numer w niemalejąco uporządkowanym ciągu X–ów, oraz każdej obserwacji Y_i nadajemy rangę Q_i, tzn. jej numer w niemalejąco uporządkowanym ciągu Y–ów. Otrzymujemy w ten sposób pary liczb naturalnych (R_i, Q_i).
Statystyka testowa

$$r_{\text{emp}} = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - Q_i)^2$$

Jeżeli $|r_{\text{emp}}| > r_{\alpha/2,n}$, to hipotezę H_0 : *Cechy X oraz Y są niezależne* odrzucamy. Liczby $r_{\alpha,n}$ są stablicowanymi dwustronnymi wartościami krytycznymi rozkładu współczynnika korelacji Spearmana.

Test współczynnika korelacji rangowej Kendalla

Obserwacjami są pary liczb $(X_i, Y_i), i = 1, \dots, n$. Pary te porządkujemy według wzrastających wartości X-ów otrzymując następujący ciąg par:

$$(X_{(1)}, Y^*_1), \dots, (X_{(n)}, Y^*_n), \ X_{(1)} < \cdots X_{(n)}.$$

Niech s_i będzie liczbą tych par $(X_{(j)}, Y^*_i), j > i$, w których $Y^*_j > Y^*_i$.
Statystyka testowa

$$t_{\text{emp}} = \frac{4 \sum_{i=1}^{n} s_i}{n(n-1)} - 1$$

Jeżeli $|t_{\text{emp}}| > t_{\alpha/2,n}$, to hipotezę H_0 : *Cechy X oraz Y są niezależne* odrzucamy. Liczby $t_{\alpha,n}$ są stablicowanymi dwustronnymi wartościami krytycznymi rozkładu współczynnika korelacji Kendalla.

146. Właściciel palarni kawy twierdzi, że stopień palenia kawy niema wpływu na jej smak, a dokładnie na gorzkość. W celu udowodnienia tej hipotezy wybrano pewną mieszankę kawy i poddano ją procesowi palenia w różnym stopniu, uzyskano następujące wyniki:

	smak kawy		
	normalna	gorzka	bardzo gorzka
słabo palona	5	9	4
mocno palona	2	12	8
bardzo mocno palona	1	7	14

W oparciu o powyższe dane odpowiedzieć na pytanie czy właściciel palarni ma rację?

147. Poniższa tabela przedstawia liczbę psów zdrowych i chorych na nosówkę w zależności od tego, czy pies ma rodowód, czy go nie ma. Zbadać, czy istnieje zależność między zdrowotnością psa a posiadaniem przez niego rodowodu.

	Psy z rodowodem	Psy bez rodowodu
Psy zdrowe	300	200
Psy chore	40	20

148. Poniższa tabela przedstawia liczby prosiąt zdrowych i chorych na nosoryjówke w zależności od tego, czy matka była zdrowa, czy też chora. Zbadać, czy istnieje zależność między zdrowotnością matek i potomków.

	Matka zdrowa	Matka chora
Potomek zdrowy	300	50
Potomek chory	40	60

149. Przypuszczano, że w populacji małych i średnich firm gminy Centrum organizowanie przez firmę regularnego wyżywienia pracownikom zależy od liczby zatrudnionych w niej pracowników. Czy przypuszczenia można uznać za uzasadnione, jeżeli na 110 wylosowanych firm uzyskano następujące wyniki:

	nieregularne	regularne
Małe firmy	60	5
Średnie firmy	36	9

150. Przypuszczano, że sposób zapewniania sobie posiłków w pracy przez pracowników, którym firma nie zapewnia regularnego wyżywienia, zależy od płci. W tym celu wylosowano pewną grupę pracowników i uzyskano następujące wyniki:

Płeć	śniadanie z domu	obiad na mieście	zamówienie do pracy
Mężczyźni	68	36	23
Kobiety	36	50	18

151. Pewien produkt można wytwarzać trzema metodami produkcji. Wysunięto hipotezę, że wadliwość produkcji nie jest zależna od metody produkcji. Zweryfikować to przypuszczenie na podstawie poniższych danych.

Jakość	Metoda produkcji		
	I	II	III
dobra	40	80	60
zła	10	60	20

$(n_{ij} - n^t_{ij})^2/n^t_{ij}$		
1.33	1.90	0.83
		1.67

152. W celu stwierdzenia, czy podanie chorym na pewną chorobę nowego leku przynosi poprawę w ich stanie zdrowia wylosowano dwie grupy pacjentów w jednakowym stopniu chorych na tę chorobę. Jednej grupie podawano nowy lek, zaś drugiej podawano leki tradycyjne. Po pewnym czasie zanotowano następujące zmiany stanu zdrowia. Zbadać, czy nowy lek daje inne efekty leczenia niż lek tradycyjny.

	bez poprawy	wyraźna poprawa	całkowite wyleczenie	$(n_{ij} - n^t_{ij})^2/n^t_{ij}$		
nowy lek	20	40	60	0.90	0.13	0.94
lek tradycyjny	45	20	15			2.26

153. Pracownicy fabryk pewnego zjednoczenia charakteryzują się różną absencją. Wysunięto przypuszczenie, że absencja zależy do płci. Zweryfikować to przypuszczenie na podstawie poniższych danych.

Liczba dni nieobecności	Płeć Kobiety	Mężczyźni
0–5	300	500
5–20	80	70
20 i więcej	20	30

$(n_{ij} - n^t_{ij})^2/n^t_{ij}$	
	0.83
	4.44
0.00	0.00

154. W poniższej tablicy podany jest podział ze względu na dwie cechy jaj zniesionych przez 23 kury. Zbadać, czy płeć kurczaka i okres zniesienia jaja są niezależne.

	Jaja zniesione w okresie trzech pierwszych miesięcy nieśności	Jaja zniesione po trzech pierwszych miesiącach nieśności
Jaja, z których wykluły się kogutki	449	821
Jaja, z których wykluły się kurki	578	1048
Jaja niezapłodnione	269	1269

$(n_{ij} - n^t_{ij})^2/n^t_{ij}$	
16.3	6.73
	9.17
	29.94

155. W badaniach budżetów rodzinnych wylosowano 2000 gospodarstw domowych i zanotowano średni miesięczny dochód na głowę oraz fakt posiadania magnetowidu. Czy można na tej podstawie powiedzieć, że fakt posiadania magnetowidu jest wskaźnikiem zamożności rodziny?

Dochód na głowę	Magnetowid jest	nie ma
poniżej 200	404	231
200– 400	486	300
400– 600	242	137
600– 800	57	44
800–1000	29	28
1000 i więcej	24	18

$(n_{ij} - n^t_{ij})^2/n^t_{ij}$	
0.24	0.39
0.01	0.01
0.19	0.31
0.52	0.86
1.16	
0.17	

156. Zbadać, czy istnieje zależność między stopniem związania kiełbasy a jej smakowitością.

	słabo związana	związana	dobrze związana
dostateczna	9	5	3
dobra	4	12	6
b. dobra	1	6	14

$(n_{ij} - n^t_{ij})^2/n^t_{ij}$		
	0.35	1.90
0.25	1.51	0.70
3.10	0.52	

157. W pewnym doświadczeniu chemicznym bada się grubość powłoki niklowej uzyskiwanej dla trzech różnych rodzajów kąpieli galwanicznych. Uzyskano następujące wyniki. Czy na tej podstawie można powiedzieć, że grubość powłoki zależy od rodzaju kąpieli?

Grubość powłoki	Liczba pomiarów w kąpieli A	B	C
4– 8	32	51	68
8–12	123	108	80
12–16	10	26	26
16–20	41	34	28
20–24	18	20	24

$(n_{ij} - n^t_{ij})^2/n^t_{ij}$		
5.95	0.04	
4.74	0.00	
5.12	0.94	1.58
1.69	0.08	0.99
0.23	0.11	0.66

158. W ankiecie rozesłanej wśród pracowników pewnego konsorcjum pytano, czy chcieliby zmienić obecne miejsce pracy. Uzyskano następujące wyniki. Czy chęć zmiany pracy zależy od aktualnych zarobków?

Zarobek aktualny	Odpowiedź Tak	Nie	$(n_{ij} - n_{ij}^t)^2/n_{ij}^t$	
500– 700	46	62		1.78
700– 900	94	146		1.80
900–1100	249	501	0.33	0.15
1100–1300	126	326	2.42	1.14
1300–1500	43	135	3.43	1.62
1500–1700	26	70	0.73	0.34

159. Kierowców pewnego przedsiębiorstwa poddano ocenie za jazdę w mieście i w terenie oceniając ich umiejętności w skali dwudziestopunktowej. Na podstawie badania dziesięciu kierowców, zbadać czy badane umiejętności są od siebie niezależne.

kierowca	1	2	3	4	5	6	7	8	9	10
jazda w mieście	19	17	15	14	13	10	7	3	2	1
jazda w terenie	14	20	11	18	6	12	13	2	4	9

160. W pewnej grupie gospodarstw domowych badano zużycie dwóch artykułów spożywczych. Na podstawie poniższych danych sprawdzić, czy taka zależność istnieje.

artykuł A	10	20	25	30	40	45	60
artykuł B	20	25	35	30	45	50	60

161. W pewnej fabryce badano zależność między średnią wydajnością a czasem nieprzerwanej pracy. Na podstawie poniższych wyników zbadać, czy taka zależność istnieje.

czas pracy	1	2	3	4	5	6	7	8	9	10
wydajność	18	20	18	17	15	15	14	12	10	10

162. Dwóch kiperów poddano badaniu na zgodność ocen. Na podstawie ocen wystawionych przez kiperów dziesięciu próbkom wina zbadać, czy oceny kiperów są niezależne od siebie.

kiper 1	190	170	150	140	130	100	70	30	20	10
kiper 2	140	190	110	160	50	120	130	10	20	80

163. Wylosowano dziesięć par zawierających związak małżeński i zanotowano ich wiek. Czy istnieje zależność między wiekiem kobiety i mężczyzny wstępujących w związek małżeński?

wiek kobiety	23	24	29	27	33	29	19	22	21	23
wiek mężczyzny	27	28	30	30	35	41	22	25	26	26

164. Poddano badaniom dwa różne testy inteligencji. Testy takie powinny być oczywiście „zgodne” w tym sensie, że ich wyniki nie są niezależne. Jeżeli przez X oznaczymy wyniki uzyskiwane w pierwszym teście, zaś przez Y wyniki drugiego testu, to taka para zmiennych losowych nie powinna być niezależna. Zbadać, czy zmienne losowe X i Y są niezależne. W eksperymencie wzięło udział dwadzieścia osób. Każdą z nich zbadano obydwoma testami uzyskując następujące wyniki:

$$(502, 564), (678, 787), (727, 851), (724, 767), (930, 789),$$
$$(576, 722), (527, 585), (705, 739), (737, 865), (714, 768),$$
$$(999, 901), (955, 922), (529, 444), (603, 492), (858, 809),$$
$$(825, 951), (504, 616), (646, 635), (663, 574), (582, 573).$$

Pierwsza liczba jest wynikiem pierwszego testu (X), natomiast druga jest wynikiem drugiego testu (Y) uzyskanego przez tą samą osobę.

165. Dziesięciu uczniów rozwiązywało dwa testy ze statystyki. Wyniki testów (w punktach) podane są w poniższej tabeli. Czy można przyjąć, że występuje zależność między wynikami tych testów?

Test 1	20	19	11	18	17	16	15	14	13	12
Test 2	19	20	16	18	17	11	15	12	14	10

Testy zgodności: chi–kwadrat, Kołmogorowa, Shapiro–Wilka

Test chi–kwadrat zgodności

W celu zastosowania tego testu próba musi być zapisana w postaci szeregu rozdzielczego:

Klasa	1	2	$\cdots$	k
Liczebność	n_1	n_2	$\cdots$	n_k

Statystyka testowa

$$\chi^2_{\text{emp}} = \sum_{i=1}^{k} \frac{(n_i - n_i^t)^2}{n_i^t} \ ,$$

gdzie

$$n_i^t = Np_i^t,\ N = \sum_{i=1}^{k} n_i,\ p_i^t = P_F\{X \text{ przyjęła wartość z klasy } i\},$$

Jeżeli $\chi^2_{\text{emp}} > \chi^2_{\alpha,k-u-1}$, gdzie u jest ilością nieznanych parametrów hipotetycznego rozkładu F, to hipotezę odrzucamy.

Test Kołmogorowa zgodności

Próbę $X_1, \ldots, X_n$ porządkujemy niemalejąco: $X_{(1)} \leq \cdots \leq X_{(n)}$.
Statystyka testowa

$$D_n = \sup_{-\infty<x<\infty} \left|F(X_{(i)}) - F_n(x)\right| \ .$$

gdzie F_n jest dystrybuantą empiryczną

$$F_n(x) = \begin{cases} 0, & x \leq X_{(1)}, \\ \frac{k}{n}, & X_{(k-1)} < x \leq X_{(k)}, \\ 1, & X_{(n)} < x. \end{cases}$$

Jeżeli $D_n > D_{\alpha,n}$, to hipotezę odrzucamy. Liczby $D_{\alpha,n}$ są stablicowanymi wartościami krytycznymi testu Kołmogorowa.

Test Shapiro–Wilka

Próbę $X_1, \ldots, X_n$ porządkujemy: $X_{(1)} \leq \cdots \leq X_{(n)}$.
Statystyka testowa

$$W = \frac{\left(\sum_{i=1}^{[n/2]} a_{i:n}(X_{(n-i+1)} - X_{(i)})\right)^2}{\text{var}X} \ ,$$

gdzie $a_{i:n}$ są stablicowanymi współczynnikami oraz

$$[n/2] = \begin{cases} n/2, & \text{dla } n \text{ parzystych} \\ (n-1)/2, & \text{dla } n \text{ nieparzystych} \end{cases}$$

Jeżeli $W \leq W_n(\alpha)$, to hipotezę odrzucamy, gdzie $W_n(\alpha)$ jest stablicowaną wartością krytyczną.

166. W celu sprawdzenia, czy kostka sześcienna do gry jest rzetelna wykonano 120 rzutów i otrzymano następujące wyniki:

Liczba oczek	1	2	3	4	5	6
Liczba rzutów	11	30	14	10	33	22

Na podstawie powyższych wyników sprawdzić, czy badaną kostkę można uznać za rzetelną.

167. Ścianki kostki do gry pomalowane są na dwa kolory: czarny i biały. Czy prawdziwe jest stwierdzenie, że białym kolorem pomalowana jest dokładnie jedna ścianka, jeżeli na 30 rzutów kolor biały pojawił się dokładnie 4 razy.

168. Badano stopień nieopanowania materiału ze statystyki przez studentów pewnego wydziału zliczając ilość ocen niedostatecznych na zaliczeniu przedmiotu w każdej z siedmiu grup.

Grupa	1	2	3	4	5	6	7
Ilość dwój	7	9	14	6	4	2	7

Czy na podstawie powyższych danych można sądzić, że stopień nieopanowania wiedzy jest we wszystkich grupach jednakowy?

169. W klasycznych doświadczeniach dotyczących selekcji grochu Mendel obserwował liczności występowania różnych rodzajów nasion otrzymywanych przy krzyżowaniu roślin z okrągłymi i żółtymi nasionami oraz roślin z pomarszczonymi i zielonymi nasionami. Uzyskał on następujące wyniki obserwacji:

pomarszczone i zielone	32
okrągłe i zielone	108
pomarszczone i żółte	101
okrągłe i żółte	315

Według teoretycznych rozważań prawdopodobieństwa otrzymania wymienionych rodzajów grochu winny być w stosunku 1 : 3 : 3 : 9. Na podstawie powyższych danych zweryfikować teorię.

170. Wyznaczono liczbę błędów na stronie przy korekcie 500–stronnicowej książki i otrzymano:

Liczba błędów	0	1	2	3	4	5	6	7	8
Liczba stron	67	139	134	90	44	15	6	4	1

Zbadać, czy można uznać, że liczba błędów na stronie ma rozkład Poissona.

171. Czy można twierdzić, że posiadane przez nas monety są symetryczne, jeżeli na 100 rzutów obiema monetami naraz 23 razy uzyskaliśmy dwa orły, 27 razy dwie reszki, a w pozostałych przypadkach uzyskaliśmy orła i reszkę.

172. Wykonano 200 serii po cztery rzuty monetą i uzyskano następujące liczby orłów:

Liczba orłów	0	1	2	3
Liczba rzutów	20	80	70	30

Czy na podstawie uzyskanych wyników można monetę traktować jako symetryczną?

173. Właściciel sklepu podejrzewa, że czarne szale kupowane są dwa razy częściej niż brązowe, a te z kolei dwa razy częściej niż białe. Czy można uznać przypuszczenie sprzedawcy za uzasadnione, jeżeli na 350 sprzedanych szali 220 było czarnych, a 90 brązowych.

174. Przypuszcza się, że u ludzi zamieszkujących środkową Europę włosy naturalnie ciemne występują sześć razy częściej niż blond, a włosy blond dwa razy częściej niż rude. Czy można to przypuszczenie uznać za uzasadnione, jeżeli wśród 150 losowo wybranych osób stwierdzono 120 osób z włosami ciemnymi, a 15 z włosami blond.

175. W poniższej tablicy podano procentową zawartość skrobii w każdym z 80 ziemniaków wylosowanych z partii ziemniaków:

Zawartość skrobii	Liczba ziemniaków
9 – 11	1
11 – 13	2
13 – 15	7
15 – 17	20
17 – 19	30
19 – 21	16
21 – 23	3
23 – 25	1

Zbadać normalność rozkładu zawartości skrobi.

176. W celu zbadania zjawiska absencji chorobowej pracowników pewnego zakładu pracy wybrano losowo grupę 100 pracowników i zanotowano liczby dni opuszczonych z powodu choroby w ciągu ostatniego roku. Otrzymano wyniki:

Liczba dni	0	1–3	4–6	7-9	10–15	16–24	24–30
Liczba pracowników	13	37	22	17	8	2	1

Sprawdzić, czy rozkład absencji chorobowej pracowników jest rozkładem Poissona.

Dynamika zjawisk (indeksy)

Numer artykułu	Ilość Rok 0	Ilość Rok 1	Cena jednostkowa Rok 0	Cena jednostkowa Rok 1
1	q_{10}	q_{11}	p_{10}	p_{11}
2	q_{20}	q_{21}	p_{20}	p_{21}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	q_{k0}	q_{1k}	p_{k0}	p_{1k}

Numer	Wartość	Wartość	Wartość	Wartość
1	$w_{1,00} = q_{10}p_{10}$	$w_{1,11} = q_{11}p_{11}$	$w_{1,01} = q_{10}p_{11}$	$w_{1,10} = q_{11}p_{10}$
2	$w_{2,00} = q_{20}p_{20}$	$w_{2,11} = q_{21}p_{21}$	$w_{2,01} = q_{20}p_{21}$	$w_{2,10} = q_{21}p_{20}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$w_{k,00} = q_{k0}p_{k0}$	$w_{k,11} = q_{k1}p_{k1}$	$w_{k,01} = q_{k0}p_{k1}$	$w_{k,10} = q_{k1}p_{k0}$
Razem	w_{00}	w_{11}	w_{01}	w_{10}

Indeks zmian wartości $I_w = w_{11}/w_{00}$
Indeks Laspayresa zmian ilości ${}_LI_{qp} = w_{10}/w_{00}$
Indeks Laspayresa zmian cen ${}_LI_{pq} = w_{01}/w_{00}$
Indeks Paaschego zmian ilości ${}_PI_{qp} = w_{11}/w_{01}$
Indeks Paaschego zmian cen ${}_PI_{pq} = w_{11}/w_{10}$
Indeks Fishera zmian ilości ${}_FI_q = \sqrt{{}_LI_{qp} \cdot \ {}_PI_{qp}}$
Indeks Fishera zmian cen ${}_FI_p = \sqrt{{}_LI_{pq} \cdot \ {}_PI_{pq}}$

Czas	Obserwacja	Indeksy jednopodstawowe absolutne	Indeksy jednopodstawowe względne	$i_{t\|c}$	Indeksy łańcuchowe absolutne	Indeksy łańcuchowe względne	$i_{t\|t-1}$
t_0	y_0	0	$\cdot$	1	0	$\cdot$	$\cdot$
t_1	y_1	$y_1 - y_0$	$(y_1 - y_0)/y_0$	y_1/y_0	$y_1 - y_0$	$(y_1 - y_0)/y_0$	y_1/y_0
t_2	y_2	$y_2 - y_0$	$(y_2 - y_0)/y_0$	y_2/y_0	$y_2 - y_1$	$(y_2 - y_1)/y_1$	y_2/y_1
t_3	y_3	$y_3 - y_0$	$(y_3 - y_0)/y_0$	y_3/y_0	$y_3 - y_2$	$(y_3 - y_2)/y_2$	y_3/y_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
t_k	y_k	$y_k - y_0$	$(y_k - y_0)/y_0$	y_k/y_0	$y_k - y_{k-1}$	$(y_k - y_{k-1})/y_{k-1}$	y_k/y_{k-1}

Średnie tempo zmian $\bar{i}_{t|t-1} = \sqrt[k-1]{i_{2|1} i_{3|2} \dots i_{k|k-1}}$

177. W celu zbadania dynamiki kosztów utrzymania w zakresie dwóch grup towarowych: nabiał oraz owoce ustalono, że towarami–reprezentantami będą: ser biały oraz jabłka. W badanym okresie cena sera nie uległa zmianie, natomiast cena jabłek wzrosła o 20%. Jak zmieniły się w rozpatrywanym okresie koszty utrzymania w zakresie dwóch grup towarowych, jeśli wiadomo, że udział wartości grupy towarowej „nabiał" w łącznej wartości sprzedaży obu grup zmalał z 48% w okresie podstawowym do 40% w okresie badanym? Zastosować w analizie odpowiednie formuły indeksowe.

178. Przedsiębiorstwo branży dziewiarskiej produkuje w trzech zakładach odpowiednio: w zakładzie A dzianinę (w metrach bieżących), w zakładzie B dresy (w sztukach) i w zakładzie C włóczkę bawełnianą (w kilogramach).

Zakład	Wielkość eksportu		Cena jednostkowa	
	1991	1992	1991	1992
dzianina	50	50	30	45
dresy	10	20	60	80
włóczka	20	15	20	40

Przeprowadzić wszechstronną analizę dynamiki eksportu obu produktów.

179. Zbadać dynamikę cen w latach 1972–1976 dla dwóch arytułów łącznie, jeżeli cena pierwszego nie zmieniła się, zaś drugiego zmalała o 10%. Wiadomo ponadto, że sprzedaż piewszego artykułu w ujęciu wartościowym była w 1976 roku trzykrotnie większa niż drugiego.

180. Zbadać dynamikę zbytu dwóch artykułów A i B jeżeli wiadomo, że cena artykułu A zmalała w okresie badanym w porównaniu z okresem podstawowym o 8%, natomiast artykułu B o 10% oraz, że udział wartości sprzedaży artykułu A wynosił 40% obrotów łącznych w okresie badanym wynoszących 2 mld zł.

181. Wartość skupu czterech podstawowych zbóż i dynamika wielkości dostaw w latach 1971 i 1973 podane są w poniższej tabeli:

Zboża	Wartość skupu w cenach bieżących (w mln zł)		Dynamika wielkości dostaw (1971 = 100)
	1971	1973	
Pszenica	8424	9124	102
Żyto	4557	6273	117
Jęczmień	3072	2988	95
Owies	951	520	48

Zbadać dynamikę skupu zbóż.

182. Zbadać dynamikę zmian zbiorów winogron w latach 1983 i 1986.

Odmiana	Ilość		Cena jednostkowa	
	1983	1986	1983	1986
A	1280	1360	108	111
B	830	890	93	101
C	1640	1660	97	107

183. Na pewnym targowisku ustalono, że wartość sprzedaży jaj wzrosła z 250 tys zł w 1984 roku do 500 tys zł w roku 1986, sera białego z 80 tys zł do 120 tys zł, zaś wartość sprzedaży śmietany zmalała z 60 tys zł do 30 tys zł. Wiadomo, że ilościowo sprzedaż jaj wzrosła o 30%, sera o 10%, natomiast śmietany zmalała dwukrotnie. Opisać dynamikę sprzedaży nabiału na tym targowisku.

184. Przedsiębiorstwo Handlu Materiałami Budowalnymi miało następujące dane o obrotach w 1976 roku:

Materiał	Wartość obrotów (w mln zł)	Zmiany cen w 1976 roku w stosunku do 1974 roku
I	400	spadek o 5%
II	800	wzrost o 10%
III	200	bez zmian

Łączne obroty w 1974 roku wyniosły 1 mld zł. Ocenić jaki wpływ na dynamikę wartości sprzedaży tych materiałów wywarły zmiany cen, a jaki wpływ zmiany ilości zakupów.

185. Obroty (w tys. zł.) sklepów „Społem" w pewnej miejscowości przedstawiały się następująco:

Towary	Wartość obrotów		Obniżka cen
	1972	1974	(%)
A	400	400	10
B	500	600	7
C	500	200	5

Zbadać dynamikę wartości obrotów trzema artykułami łącznie.

186. W 1974 roku sprzedano w mieście M jabłek za 120 mln zł i gruszek za 60 mln zł, a w 1976 roku sprzedano jabłek za 160 mln zł oraz gruszek za 70 mln zł. Wiadomo, że ceny jabłek wzrosły o 30%, natomiast cecny gruszek spadły o 10% w 1976 roku w stosunku do roku 1974. Scharakteryzować dynamikę sprzedaży owoców w mieście M.

187. Wielkość eksportu i ceny jednostkowe dwóch produktów kształtowały się następująco:

Produkt	Wielkość eksportu		Cena jednostkowa	
	1972	1976	1972	1976
A	25	30	10	9
B	40	80	12	12

Przeprowadzić wszechstronną analizę dynamiki eksportu obu produktów.

188. Przedsiębiorstwo L wytwarza trzy rodzaje obrabiarek A, B i C (o różnym przeznaczeniu, wielkości i wartości). W 1972 roku przedsiębiorstwo wyprodukowało 400 obrabiarek typu A, 100 obrabiarek typu B oraz 600 obrabiarek typu C. W roku 1975 produkcja obrabiarek wynosiła odpowiednio: 600, 50 oraz 1000. Cena obrabiarki typu A wynosiła w 1972 roku 40 tys. zł., a w 1975 roku była o 20% wyższa. Cena obrabiarki typu B nie zmieniła się w badanym okresie i wynosiła 15 tys. zł. Cena obrabiarki typu C wynosiła w 1972 roku 80 tys. zł., a w 1975 — 120 tys. zł. Przeprowadzić wszechstronną analizę dynamiki produkcji obrabiarek.

189. Badano wielkość spożycia herbaty w Polsce:

Rok	1978	1979	1980	1981	1982	1983	1984	1985	1986
Spożycie	175	190	185	195	180	200	185	190	205

Czy średnie tempo zmian spożycia herbaty w latach 1978–1982 było takie same jak w latach 1982–1986?

190. Obserwowano wielkość zbiorów pszenicy latach 1980–1988:

Rok	1980	1981	1982	1983	1984	1985	1986	1987	1988
Zbiory	7.5	7.8	8.2	8.2	8.4	8.5	8.7	9.1	9.2

Oszacować średnie tempo zmian zbiorów pszenicy.

191. W latach 1970–1978 notowano ilość zgonów niemowląt w Polsce.

Rok	1970	1971	1972	1973	1974	1975	1976	1977	1978
Liczba zgonów	9468	8644	7965	8599	8460	7857	8088	7979	7217

Oszacować średnie tempo zmian zgonów niemowląt.

192. Badano ilość ofiar śmiertelnych w wypadkach samochodowych spowodowanych przez pijanych kierowców:

Rok	1978	1979	1980	1981	1982	1983	1984	1985	1986
Ofiary	175	190	185	195	180	200	185	190	205

Czy średnie tempo wzrostu ilości ofiar w latach 1978–1982 było takie same jak w latach 1982–1986?

Analiza szeregów czasowych
y_t — szereg czasowy ($t = 0, 1, 2, \ldots, n-1$); r — okres szeregu czasowego
Średnie ruchome (długość cyklu wynosi r)

$$\bar{y}_m = \begin{cases} \frac{1}{r}(y_{m-\frac{r-1}{2}} + \cdots + y_m + \cdots + y_{m+\frac{r-1}{2}}), \ m = \frac{r-1}{2}, \ldots, k - \frac{r-1}{2}, \ (r \text{ nieparzyste}), \\ \frac{1}{r}(\frac{1}{2}y_{m-\frac{r}{2}} + \cdots + y_m + \cdots + \frac{1}{2}y_{m+\frac{r}{2}}), \ m = \frac{r}{2}, \ldots, k - \frac{r}{2}, \ (r \text{ parzyste}), \end{cases}$$

$f(t)$ — funkcja trendu; $\hat{y}_t = \hat{f}(t)$ — dopasowana funkcja trendu
Surowy wskaźnik O_{si}

$$O_{si} = \frac{Y_i}{\hat{Y}_i}, \ Y_i = y_i + y_{r+i} + y_{2r+i} + \cdots, \ \hat{Y}_i = \hat{y}_i + \hat{y}_{r+i} + \hat{y}_{2r+i} + \cdots$$

Wskaźnik korygujący

$$k = \frac{r}{O_{s1} + O_{s2} + \cdots + O_{sr}}$$

Poprawiony wskaźnik O_i

$$O_i = k \cdot O_{si}, \quad i = 0, 1, \ldots, r-1$$

Absolutny wskaźnik okresowości g_i

$$g_i = (O_i - 1) \cdot \bar{y}, \quad i = 0, 1, \ldots, r-1, \ \bar{y} = (\sum_{j=0}^{n-1} y_j)/n$$

Oszacowanie odchylenia standardowego

$$S = \sqrt{\frac{1}{n-2} \sum_{j=0}^{n-1} (y_j - \hat{y}_j - g_t)^2}, \ \ (t \text{ jest resztą z dzielenia } j/r)$$

Prognoza w chwili $m > n$

$$y_m = \hat{y}_m + g_t \pm S, \ \ (\text{gdzie } t \text{ jest resztą z dzielenia } m/r)$$

193. Zanotowano ilość zleceń wykonanych przez pewną firmę kolejnych kwartałach lat 1989–1993.

Dane oryginalne

Lata	Kwartały			
	I	II	III	IV
1989	4	8	12	15
1990	6	14	18	20
1991	12	16	22	28
1992	20	25	30	35
1993	30	36	39	43

Wartości trendu $y = 1.7962t + 4.5857$

Lata	Kwartały			
	I	II	III	IV
1989	4.586	6.382	8.178	9.974
1990	11.771	13.567	15.363	17.159
1991	18.956	20.752	22.548	24.344
1992	26.141	27.937	29.733	31.529
1993	33.326	35.122	36.918	38.714

Wiedząc, że $s^2 = 4.123$ zrobić prognozę ilości zleceń w trzecim kwartale 1994 roku.

194. Wielkość produkcji jaj (w mln. szt.) w poszczególnych kwartałach lat 1970–1974 w gospodarstwach rolnych (państwowych i indywidualnych) była następująca

Dane oryginalne

Lata	Kwartały			
	I	II	III	IV
1970	1450	1600	2100	1770
1971	1520	1710	2350	1830
1972	1570	1750	2410	1870
1973	1650	1850	2690	1910
1974	1750	1910	2750	1970

Wartości trendu $y = 27.51t + 1659.14$

Lata	Kwartały			
	I	II	III	IV
1970	1659.14	1686.65	1714.17	1741.68
1971	1769.19	1796.70	1824.21	1851.72
1972	1879.23	1906.74	1934.26	1961.77
1973	1989.28	2016.79	2044.30	2071.81
1974	2099.32	2126.83	2154.35	2181.86

Wiedząc, że $s^2 = 5959.155$ zrobić prognozę produkcji jaj na trzeci kwartał 1975.

195. W tablicy podano szereg dynamiczny, który przedstawia kwartalne dane dotyczące ilości zajętych miejsc hotelowych w poszczególnych kwartałach lat 1982–1986

Dane oryginalne

Lata	Kwartały			
	I	II	III	IV
1982	1861	2203	2415	1908
1983	1921	2343	2514	1986
1984	1834	2154	2098	1799
1985	1837	2025	2304	1965
1986	2073	2414	2339	1967

Wartości trendu $y = 1.699t + 2081.857$

Lata	Kwartały			
	I	II	III	IV
1982	2081.86	2083.56	2085.26	2086.95
1983	2088.65	2090.35	2092.05	2093.75
1984	2095.45	2097.15	2098.85	2100.55
1985	2102.25	2103.95	2105.65	2107.35
1986	2109.05	2110.74	2112.44	2114.14

Wiedząc, że $s^2 = 14081.04$ zrobić prognozę zajętych miejsc hotelowych w pierwszym kwartale 1987.

196. Badano ilość zgonów niemowląt w kolejnych kwartałach lat 1970–1974.

Dane oryginalne

Lata	Kwartały			
	I	II	III	IV
1970	5664	4811	3804	3833
1971	4347	4412	3618	4187
1972	4912	4131	3548	3726
1973	3979	4109	3509	3870
1974	3995	3726	3222	3767

Wartości trendu $y = -58.07t + 4610.20$

Lata	Kwartały			
	I	II	III	IV
1970	4610.20	4552.13	4494.05	4435.98
1971	4377.91	4319.83	4261.76	4203.68
1972	4145.61	4087.54	4029.46	3971.39
1973	3913.32	3855.24	3797.17	3739.09
1974	3681.02	3622.95	3564.87	3506.80

Wiedząc, że $s^2 = 82812.94$ zrobić prognozę ilości zgonów w drugim kwartale 1975 roku.

197. Badano ilość nieobecności w kolejnych kwartałach lat 1980–1984.

Dane oryginalne

Lata	Kwartały I	II	III	IV
1980	19	17	16	21
1981	20	18	16	23
1982	21	19	17	25
1983	22	19	19	26
1984	24	21	20	27

Wartości trendu $y = 0.314t + 17.514$

Lata	Kwartały I	II	III	IV
1980	17.51	17.83	18.14	18.46
1981	18.77	19.09	19.40	19.71
1982	20.03	20.34	20.66	20.97
1983	21.29	21.60	21.91	22.23
1984	22.54	22.86	23.17	23.49

Wiedząc, że $s^2 = 0.28$ zrobić prognozę ilości nieobecności w drugim kwartale 1985 roku.

198. Poniżej podana jest sprzedaż pewnego artykułu w kolejnych kwartałach lat 1976–1980

Dane oryginalne

Lata	Kwartały I	II	III	IV
1976	3995	699	746	6421
1977	4225	840	914	6801
1978	5572	847	959	7629
1979	5824	1060	1137	8668
1980	6274	1169	1218	8916

Wartości trendu $y = 120.84t + 2547.69$

Lata	Kwartały I	II	III	IV
1976	2547.69	2668.53	2789.37	2910.22
1977	3031.06	3151.90	3272.75	3393.59
1978	3514.43	3635.28	3756.12	3876.97
1979	3997.81	4118.65	4239.50	4360.34
1980	4481.18	4602.03	4722.87	4843.71

Wiedząc, że $s^2 = 217888.07$ zrobić prognozę wielkości sprzedaży w drugim kwartale 1981 roku.

199. Poniżej podana jest sprzedaż pewnego artykułu w kolejnych kwartałach lat 1981–1985

Dane oryginalne

Lata	Kwartały I	II	III	IV
1981	7268	1233	1347	9560
1982	7487	1430	1463	11954
1983	8770	1495	1469	11823
1984	9311	1615	1543	13512
1985	9264	1738	1653	13941

Wartości trendu $y = 150.54t + 4463.67$

Lata	Kwartały I	II	III	IV
1981	4463.67	4614.21	4764.75	4915.29
1982	5065.83	5216.37	5366.91	5517.45
1983	5667.99	5818.53	5969.07	6119.61
1984	6270.15	6420.69	6571.23	6721.77
1985	6872.31	7022.85	7173.39	7323.93

Wiedząc, że $s^2 = 501291.14$ zrobić prognozę wielkości sprzedaży w pierwszym kwartale 1986 roku.

200. Poniżej podane są kwartalne ilości (w tysiącach) przewozów lotniczych w latach 1949–1953

Dane oryginalne

Lata	Kwartały I	II	III	IV
1949	362	385	432	341
1950	382	409	498	387
1951	473	513	582	474
1952	544	582	681	557
1953	628	707	773	592

Wartości trendu $y = 18.53t + 339.10$

Lata	Kwartały I	II	III	IV
1949	339.10	357.63	376.15	394.68
1950	413.21	431.73	450.26	468.78
1951	487.31	505.84	524.36	542.89
1952	561.42	579.94	598.47	616.99
1953	635.52	654.05	672.57	691.10

Wiedząc, że $s^2 = 453.39$ zrobić prognozę wielkości przewozów lotniczych w pierwszym kwartale 1954 roku.

201. Poniżej podane są kwartalne ilości (w tysiącach) przewozów lotniczych w latach 1956–1960

Dane oryginalne

Lata	Kwartały I	II	III	IV
1956	878	1005	1173	883
1957	972	1125	1336	988
1958	1020	1146	1400	1006
1959	1108	1288	1570	1174
1960	1227	1468	1736	1283

Wartości trendu $y = 26.86t + 934.14$

Lata	Kwartały I	II	III	IV
1956	934.14	961.00	987.86	1014.72
1957	1041.58	1068.44	1095.29	1122.15
1958	1149.01	1175.87	1202.73	1229.59
1959	1256.45	1283.31	1310.16	1337.02
1960	1363.88	1390.74	1417.60	1444.46

Wiedząc, że $s^2 = 1752.40$ zrobić prognozę wielkości przewozów lotniczych w czwartym kwartale 1961 roku.

202. Poniżej podane są miesięczne wielkości sprzedaży pewnego towaru w kolejnych pięciu latach

Dane oryginalne

Lata	Kwartały I	II	III	IV
rok_1	8.242	8.703	7.416	17.377
rok_2	8.047	10.272	8.382	19.669
rok_3	10.166	11.446	8.361	22.079
rok_4	12.181	13.573	10.045	24.393
rok_5	12.079	13.521	10.736	28.111

Wartości trendu $y = 0.48t + 8.64$

Lata	Kwartały I	II	III	IV
rok_1	8.639	9.124	9.608	10.092
rok_2	10.576	11.061	11.545	12.029
rok_3	12.514	12.998	13.482	13.966
rok_4	14.451	14.935	15.419	15.903
rok_5	16.388	16.872	17.356	17.841

Wiedząc, że $s^2 = 1.745$ zrobić prognozę wielkości sprzedaży w drugim kwartale szóstego roku.

203. Poniżej podana jest sprzedaż w kolejnych miesiącach lat 1976–1985

	I	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	*X*	*XI*	*XII*
1976	1490	1597	908	437	150	112	87	161	498	1433	1793	3195
1977	1543	1529	1153	503	205	132	89	189	636	1185	2358	3258
1978	2184	2221	1167	522	184	141	106	201	652	1543	2434	3652
1979	2279	2256	1289	649	241	170	115	207	815	1520	2825	4323
1980	2583	2303	1388	772	228	169	131	234	853	1653	2841	4422
1981	2911	2786	1571	770	275	188	152	257	938	1694	3269	4597
1982	2972	2841	1674	947	296	187	160	288	1015	2348	3790	5816
1983	3579	3213	1978	951	332	212	162	314	993	2255	3659	5909
1984	3681	3365	2265	1043	336	236	177	322	1044	2670	4405	6437
1985	3899	3217	2148	1134	348	256	180	321	1152	2681	4611	6649

Wyznaczyć równanie trendu oraz wskaźniki okresowości. Zrobić prognozę ceny jabłek na marzec 1984.

204. Poniżej podane są miesięczne ilości (w tysiącach) przewozów lotniczych w latach 1949–1960

	I	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	*X*	*XI*	*XII*
1949	112	118	132	129	121	135	148	148	136	119	104	118
1950	115	126	141	135	125	149	170	170	158	133	114	140
1951	145	150	178	163	172	178	199	199	184	162	146	166
1952	171	180	193	181	183	218	230	242	209	191	172	194
1953	196	196	236	235	229	243	264	272	237	211	180	201
1954	204	188	235	227	234	264	302	293	259	229	203	229
1955	242	233	267	269	270	315	364	347	312	274	237	278
1956	284	277	317	313	318	374	413	405	355	306	271	306
1957	315	301	356	348	355	422	465	467	404	347	305	336
1958	340	318	362	348	363	435	491	505	404	359	310	337
1959	360	342	406	396	420	472	548	559	463	407	362	405
1960	417	391	419	461	472	535	622	606	508	461	390	432

Wyznaczyć równanie trendu oraz wskaźniki okresowości. Zrobić prognozę wielkości przewozów w kwietniu 1961 roku.

205. Poniżej podane są miesięczne wielkości sprzedaży pewnego towaru w kolejnych siedmiu latach

	I	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	*X*	*XI*	*XII*
rok_1	2.815	2.672	2.755	2.721	2.946	3.036	2.282	2.212	2.922	4.301	5.764	7.312
rok_2	2.541	2.475	3.031	3.266	3.776	3.230	3.028	1.759	3.595	4.474	6.838	8.357
rok_3	3.113	3.006	4.047	3.523	3.937	3.986	3.260	1.573	3.528	5.211	7.614	9.254
rok_4	5.375	3.088	3.718	4.514	4.520	4.539	3.663	1.643	4.739	5.428	8.314	10.651
rok_5	3.633	4.292	4.154	4.121	4.647	4.753	3.965	1.723	5.048	6.922	9.858	11.331
rok_6	4.016	3.957	4.510	4.276	4.968	4.677	3.523	1.821	5.222	6.872	10.803	13.916
rok_7	2.639	2.899	3.370	3.740	2.927	3.986	4.217	1.738	5.221	6.424	9.842	13.076

Wyznaczyć równanie trendu oraz wskaźniki okresowości. Zrobić prognozę wielkości sprzedaży w maju ósmego roku.

206. Poniżej podane są miesięczne wielkości sprzedaży pewnego towaru w kolejnych czternastu latach

	I	II	III	IV	V	VI	VII	$VIII$	IX	X	XI	XII
rok_1	73.637	77.136	81.481	84.127	84.562	91.959	94.174	96.087	88.952	83.479	80.814	77.466
rok_2	75.225	79.418	84.813	85.691	87.490	92.995	95.375	98.396	92.791	88.018	86.899	83.636
rok_3	79.245	85.536	89.313	88.785	91.307	96.394	99.864	100.744	96.009	89.428	85.518	84.603
rok_4	85.169	87.973	88.696	92.686	91.807	98.593	100.677	103.084	97.644	92.648	89.486	88.593
rok_5	86.141	89.156	93.795	93.422	94.277	99.927	102.451	105.003	98.477	93.219	90.410	87.763
rok_6	84.883	89.476	93.191	96.919	95.869	100.272	102.103	104.381	97.779	94.369	88.432	84.314
rok_7	81.714	81.222	74.290	87.121	89.099	93.743	96.741	99.366	95.184	90.407	88.968	86.765
rok_8	83.674	84.265	88.551	89.998	93.421	98.109	100.235	102.866	96.903	92.426	91.024	91.070
rok_9	87.650	87.534	92.763	97.420	99.845	103.349	102.438	104.467	99.740	96.101	93.858	93.265
rok_{10}	89.609	92.934	95.272	96.791	96.478	101.938	105.094	108.158	101.625	97.255	95.378	92.997
rok_{11}	88.047	92.172	98.050	97.874	101.700	105.794	108.016	111.475	104.945	100.496	97.555	95.556
rok_{12}	92.926	96.196	99.117	97.932	89.270	98.836	102.335	106.912	102.047	97.428	95.025	91.530
rok_{13}	87.958	89.159	94.853	95.799	98.193	103.468	104.989	108.179	103.084	99.781	96.403	95.072
rok_{14}	90.707	94.949	94.970	100.286	101.497	106.352	107.415	109.385	103.266	99.432	93.965	94.385

Wyznaczyć równanie trendu oraz wskaźniki okresowości. Zrobić prognozę wielkości sprzedaży w kwietniu piętnastego roku.

207. Ceny targowiskowe jabłek deserowych za 1 kg w poszczególnych miesiącach lat 1970–1973 kształtowały się następująco

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1980	11.6	12.4	12.8	13.7	15.0	24.7	14.0	10.4	9.0	8.5	9.0	11.3
1981	12.7	12.5	12.8	15.5	20.0	21.5	15.0	12.0	10.5	11.0	11.0	12.0
1982	12.5	12.3	12.7	14.8	19.0	31.0	15.4	10.4	10.8	12.0	13.5	15.0
1983	15.5	15.8	18.0	23.0	24.0	21.0	13.0	11.2	10.7	12.5	12.8	16.5

Wyznaczyć równanie trendu oraz wskaźniki okresowości. Zrobić prognozę ceny jabłek na marzec 1984.

Analiza danych

Postać danych (próba prosta, dane indywidualne): $X_1, X_2, \ldots, X_n$. Niech $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ będzie uporządkowanym ciągiem danych.
Postać danych (szereg rozdzielczy, dane skumulowane):

Przedział klasowy	Liczebność	Liczebność skumulowana
$x_0 - x_1$	n_1	$n_{(1)} = n_1$
$x_1 - x_2$	n_2	$n_{(2)} = n_1 + n_2$
$\vdots$	$\vdots$	$\vdots$
$x_{k-1} - x_k$	n_k	$n_{(k)} = n_1 + n_2 + \cdots + n_k (= n)$

Dla liczby p takiej, że $0 \leq p \leq 1$ niech x_p, n_p, h_p oznaczają początek, liczebność i długość przedziału zawierającego obserwację o numerze $p \cdot n$ oraz niech $n_{(p)}$ oznacza liczebność skumulowaną przedziału poprzedzającego przedział o początku x_p.

Mierniki położenia	Próba prosta	Szereg rozdzielczy
średnia $\bar{x}$	$\frac{1}{n}\sum_{i=1}^{k} X_i$	$\frac{1}{n}\sum_{i=1}^{k} \dot{x}_i n_i$
mediana Me	$X_{(n+1)/2:n}$(n nieparzyste) $(X_{n/2:n} + X_{n/2+1:n})/2$($n$ parzyste)	$x_{0.5} + \frac{h_{0.5}}{n_{0.5}}\left(\frac{n}{2} - n_{(0.5)}\right)$
dolny kwartyl Q_1	$X_{[n/4]:n}$	$x_{0.25} + \frac{h_{0.25}}{n_{0.25}}\left(\frac{n}{4} - n_{(0.25)}\right)$
górny kwartyl Q_3	$X_{[3n/4]:n}$	$x_{0.75} + \frac{h_{0.75}}{n_{0.75}}\left(\frac{3n}{4} - n_{(0.75)}\right)$
dominanta (moda) D	najczęściej występująca wartość	$x_D + h_D \frac{n_D - n_{D-1}}{2n_D - n_{D+1} - n_{D-1}}$
minimum Min	$X_{1:n}$	x_0
maksimum Max	$X_{n:n}$	x_k

Mierniki rozproszenia	Próba prosta	Szereg rozdzielczy
rozstęp R	$Max - Min$	$Max - Min$
wariancja $\mathcal{S}^2$	$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})^2$	$\frac{1}{n}\sum_{i=1}^{k} n_i (X_i - \bar{x})^2$
odchylenie standardowe $\mathcal{S}$	$\sqrt{\mathcal{S}^2}$	$\sqrt{\mathcal{S}^2}$
odchylenie przeciętne d	$\frac{1}{n}\sum_{i=1}^{n}\lvert X_i - \bar{x}\rvert$	$\frac{1}{n}\sum_{i=1}^{k} n_i \lvert X_i - \bar{x}\rvert$
odchylenie ćwiartkowe Q	$\frac{Q_3 - Q_1}{2}$	$\frac{Q_3 - Q_1}{2}$
współczynnik zmienności V	$\frac{\mathcal{S}}{\bar{x}}100\%$	$\frac{\mathcal{S}}{\bar{x}}100\%$

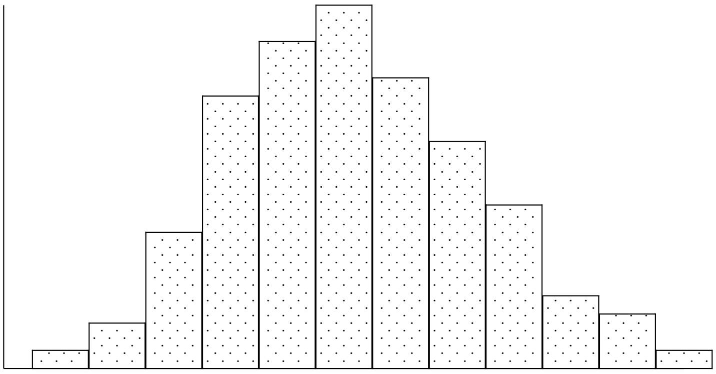
Koncentracja Lorentza

Przedział	Liczebność	Częstość	Środek	t_i	z_i	$z_{(i)}$
$x_0 - x_1$	n_1	$w_1 = n_1/n$	$\dot{x}_1$	$t_1 = n_1\dot{x}_1$	$z_1 = t_1/t$	$z_{(1)} = z_1$
$x_1 - x_2$	n_2	$w_2 = n_2/n$	$\dot{x}_2$	$t_2 = n_2\dot{x}_2$	$z_2 = t_2/t$	$z_{(2)} = z_1 + z_2$
$\vdots$	$\vdots$	$\vdots$				
$x_{k-1} - x_k$	n_k	$w_k = n_k/n$	$\dot{x}_k$	$t_k = n_k\dot{x}_k$	$z_k = t_k/t$	$z_{(k)} = z_1 + z_2 + \cdots + z_k$
Razem	n	1		t	1	

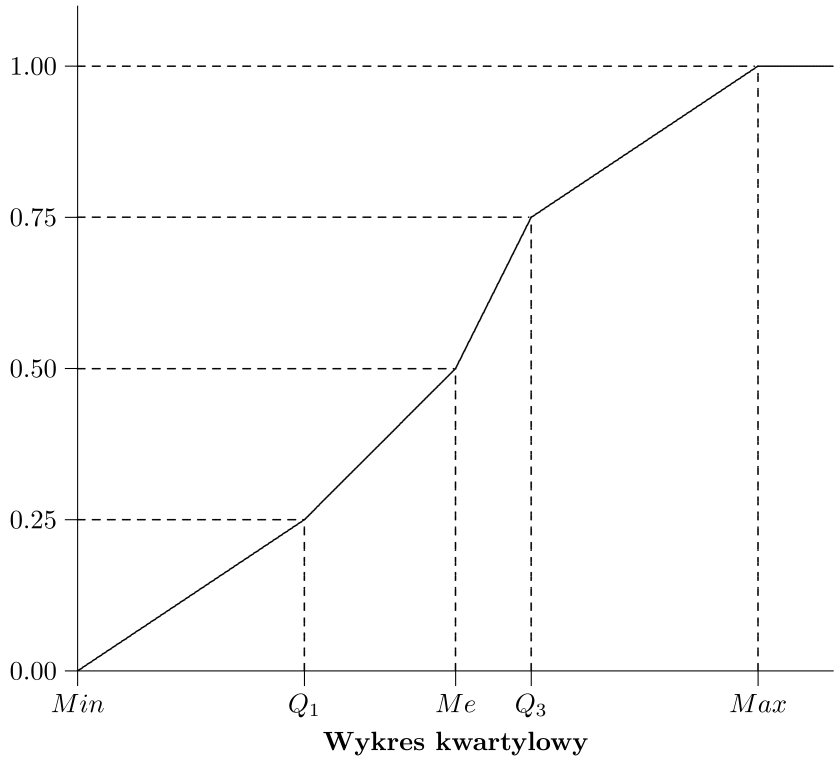
Współczynnik koncentracji Lorentza

$$K = 1 - \sum_{i=1}^{k} [z_{(i)} + z_{(i-1)}] w_i$$

Analiza danych — prezentacja graficzna



Histogram



Wykres kwartylowy

208. Oszacować przeciętną procentową zawartość tłuszczu w mleku krów na podstawie danych o procentowej zawartości tłuszczu w mleku 50 krów.

Nr.	%	Nr.	%	Nr.	%	Nr.	%	Nr.	%
1	3.35	11	4.16	21	3.24	31	4.23	41	3.42
2	3.73	12	3.56	22	3.98	32	3.70	42	4.47
3	3.94	13	3.92	23	3.62	33	3.53	43	3.93
4	4.16	14	3.22	24	4.10	34	3.72	44	4.26
5	3.92	15	3.66	25	3.78	35	3.96	45	3.81
6	4.28	16	3.50	26	3.39	36	3.83	46	4.27
7	4.26	17	3.71	27	3.93	37	4.27	47	4.06
8	3.78	18	3.96	28	3.89	38	3.93	48	4.06
9	3.99	19	3.77	29	4.22	39	3.78	49	3.66
10	3.41	20	3.53	30	3.54	40	4.08	50	3.44

Skonstruować szereg rozdzielczy (*od* 3.2 *co* 0.2) i na podstawie tego szeregu również oszacować przeciętną procentową zawartość tłuszczu w mleku krów. Porównać uzyskane wyniki.

209. Spożycie papieru (w kg) w Polsce w latach 1960 — 1971 na jednego mieszkańca wynosiło: 2.2, 2.2, 2.1, 2.3, 2.3, 2.3, 2.3, 2.4, 2.4, 2.5, 2.4, 2.5, 2.5. Przyjmując upraszczające założenie, że liczba mieszkańców w Polsce w tym okresie była stała, obliczyć średnie spożycie papieru na jednego mieszkańca w tym okresie, odchylenie standardowe, modę, medianę oraz współczynnik zmienności. Podać interpretację obliczonych wielkości.

210. Liczba koni (w mln. szt.) w Polsce w latach 1947 — 1974 wynosiła: 2.0, 2.3, 2.7, 2.8, 2.9, 2.7, 2.7, 2.6, 2.6, 2.5, 2.6, 2.7, 2.8, 2.8, 2.7, 2.7, 2.6, 2.6, 2.6, 2.6, 2.6, 2.7, 2.6, 2.6, 2.5, 2.4, 2.4, 2.3. Obliczyć średnią liczbę koni w Polsce w tym okresie, medianę, modę, odchylenie standardowe, współczynnik zmienności oraz współczynnik asymetrii.

211. W poniższej tablicy podano procentową zawartość skrobii w każdym z 80 ziemniaków wylosowanych z partii ziemniaków:

Zawartość skrobii	9–11	11–13	13–15	15–17	17–19	19–21	21–23	23–25
Liczba ziemniaków	1	2	7	20	30	16	3	1

Obliczyć średnią arytmetyczną, medianę, odchylenie standardowe, współczynnik zmienności, kwartyle. Zinterpretować wyznaczone wielkości.

212. W celu zbadania zjawiska absencji chorobowej pracowników pewnego zakładu pracy wybrano losowo grupę 100 pracowników i zanotowano liczby dni opuszczonych z powodu choroby w ciągu ostatniego roku. Otrzymano wyniki:

Liczba dni	0	1–3	4–6	7-9	10–15	16–24	24–30
Liczba pracowników	13	37	22	17	8	2	1

Przeprowadzić analizę nieobecności pracowników z powodu chorby.

213. Przeprowadzić analizę czasu dojazdu pracowników z miejsca zamieszkania do pracy na podstawie danych zawartych w poniższej tabeli.

Czas dojazdu	5–15	15–25	25–35	35-45	45–55	55–65
Liczba pracowników	3	5	25	15	5	2

214. Przeprowadzić wszechstronną analizę porównawczą wydajności pracy dwóch zakładów pracy na podstawie poniższych danych (dzienna wydajność pracowników).

Wydajność	poniżej 50	50–60	60–70	70–80	80–90	90 i więcej
Zakład A	65	50	43	11	8	3
Zakład B	10	35	92	50	17	4

215. Przeprowadzić analizę porównawczą powierzchni użytkowej mieszkań na wsiach w latach 1978 i 1988.

Powierzchnia	1978	1988
20– 40	50	60
40– 60	300	300
60– 80	400	350
80–100	150	150
100–120	80	100
120–140	20	40

Mierniki	1978	1988
średnia	69.40	71.00
dominanta	65.71	
dolny kwartyl	53.33	52.67
mediana	67.50	68.00
górny kwartyl		85.33

216. Porównać wyniki testów na inteligencję przeprowadzone wśród uczniów klas siódmych.

Wyniki	dziewczęta	chłopcy
20– 40	1	1
40– 60	4	4
60– 80	17	11
80–100	25	19
100–120	8	13
120–140	1	8

Mierniki	dziewczęta	chłopcy
średnia	83.6	92.5
dominanta	86.4	91.4
dolny kwartyl	70.6	
mediana		92.6
górny kwartyl	96.0	110.8

217. Porównać strukturę bezrobocia wśród mężczyzn i kobiet

Miesiące bez pracy	mężczyźni	kobiety
0– 3	214	153
3– 6	161	139
6– 9	121	116
9–12	108	96
12–15	396	496

Mierniki	meżczyźni	kobiety
średnia	8.43	9.43
dolny kwartyl	3.67	5.09
mediana		11.88
górny kwartyl	13.11	

218. Przeprowadzić analizę porównawczą powierzchni użytkowej mieszkań w miastach w latach 1978 i 1988.

Powierzchnia	1978	1988
20– 40	80	50
40– 60	350	250
60– 80	400	350
80–100	100	200
100–120	50	120
120–140	20	30

Mierniki	1978	1988
średnia	65.00	73.60
dominanta	60.00	68.00
dolny kwartyl	49.71	
mediana		71.43
górny kwartyl	76.00	90.00

219. Przeprowadzić analizę porównawczą powierzchni użytkowej mieszkań na wsiach i w miastach.

Powierzchnia	miasto	wieś
20– 40	80	50
40– 60	350	300
60– 80	400	400
80–100	100	150
100–120	50	80
120–140	20	20

Mierniki	miasto	wieś
średnia	65.00	69.40
dominanta	60.00	
dolny kwartyl	49.71	53.33
mediana		67.50
górny kwartyl	76.00	80.00

220. Porównać strukturę wynagrodzeń miesięcznych w przemyśle i budownictwie.

Płaca	przemysł	budownictwo
0– 200	250	100
200– 400	450	350
400– 600	150	400
600– 800	100	100
800–1000	50	50

Mierniki	przemysł	budownictwo
średnia	350	430
dominanta	280	
dolny kwartyl	200	286
mediana		425
górny kwartyl	467	550

221. Przeprowadzić analizę porównawczą „zaludnienia” mieszkań (ilość osób przypadająca na jeden pokój) na wsiach i w miastach.

„Zaludnienie”	miasto	wieś
0.0–0.5	2	3
0.5–1.0	13	11
1.0–1.5	41	31
1.5–2.0	20	24
2.0–2.5	17	18
2.5–3.0	7	13

Mierniki	miasto	wieś
średnia	1.54	1.66
dominanta		1.37
dolny kwartyl	1.12	1.18
mediana	1.43	1.60
górny kwartyl	1.98	

222. Przeprowadzić wszechstronną analizę porównawczą struktury wieku nowożeńców obu płci mając następujące dane (w latach):

Parametr	Kobiety	Mężczyźni
Średnia arytmetyczna	23.0	26.0
Mediana	21.4	24.7
Dominanta	19.0	22.0
Wariancja	16.0	36.0

223. Porównać opinie konsumentów dotyczące dwóch gatunków kawy na podstawie badań sondażowych. Każdy z sześćdziesięciu konsumentów oceniał każdą z dwóch kaw w skali punktowej.

Ocena	kawa Szatanex	kawa Lureksja
3	4	6
4	14	16
5	17	21
6	19	13
7	6	4

Mierniki	kawa Szatanex	kawa Lureksja
średnia	5.15	4.88
dominanta	5.87	
dolny kwartyl	4.00	4.00
mediana		5.00
górny kwartyl	6.00	6.00

224. Przeprowadzić wszechstronną analizę porównawczą dochodów w gospodarstwach domowych dwóch grup pracowniczych na podstawie niżej podanych danych (uzyskanych z dwóch stuelementowych prób)

	Grupa A	Grupa B
najniższy dochód	2.2	2.2
najwyższy dochód	4.0	4.8
średnia arytmetyczna	3.1	3.4
dolny kwartyl	3.0	2.7
górny kwartyl	3.6	3.5
odchylenie standardowe	0.8	0.6
mediana	3.4	3.0
dominanta	3.5	2.8

225. Na podstawie poniższych danych prezentujących dane dotyczące wagi urodzeniowej chłopców i dziewczynek dokonać wszechstronnej analizy porównawczej. Wyznaczyć brakujące parametry.

	Dziewczynki	Chłopcy
wielkość próby	60	40
średnia	339.3	334.2
mediana	345.5	349.5
dominanta	380	385
wariancja	3140	4741
minimum	178	132
maksimum	439	454
dolny kwartyl	313	315
górny kwartyl	375	381