Statystyka stosowana MAP 1079

Lista 1a
(powtórka z rachunku prawdopodobieństwa)

1. Zmienna losowa X przyjmuje wartości $2, 3, 5, 8$ z prawdopodobieństwami odpowiednio równymi $2/10, 4/10, 3/10, 1/10$. Wyznaczyć dystrybuantę tej zmiennej i obliczyć

   (a) $P(X \leq 3)$, (b) $P(X \geq 2.5)$, (c) $P(2.7 \leq X < 5.1)$, (d) $\mathrm{E}(X)$, (e) $\mathrm{Var}(X)$.

2. Zmienna losowa X ma rozkład Bernoulliego $B(n, p)$ z $n = 10$ i $p = 1/2$. Obliczyć (a) $P(X = 5)$, (b) $P(X \geq 9)$, (c) $P(3 \leq X < 6)$, (d) $\mathrm{E}(X)$, (e) $\mathrm{Var}(X)$.

3. Test składa się z 25 pytań. Odpowiadając na każde z nich można wybrać jedną z 4 możliwych odpowiedzi, przy czym trzy z nich są błędne. Zakładając, że student zgaduje odpowiedzi obliczyć prawdopodobieństwo, że odpowie on poprawnie na:

   (a) co najmniej 20 pytań,
   (b) mniej niż 5 pytań.

4. Ziarna groszku ogrodowego są żółte lub zielone. W pewnej krzyżówce odmian groszku stosunek liczby roślin z żółtymi ziarnami do liczby roślin z zielonymi ziarnami jest jak 3 : 1. Losujemy cztery rośliny z tej populacji. Jakie jest prawdopodobieństwo, że

   (a) trzy rośliny będą miały żółte ziarna a jedna zielone?
   (b) wszystkie cztery będą miały ziarna tego samego koloru?

5. Pewne lekarstwo leczy 90% przypadków pewnej choroby. Poddajemy kuracji 20 losowo wybranych chorych. Znajdź prawdopodobieństwo tego, że wyleczymy

   (a) wszystkich chorych w naszej próbie,
   (b) wszystkich oprócz jednego,
   (c) dokładnie 18 chorych,
   (d) dokładnie 90% chorych w naszej próbie.

6. Pewne lekarstwo uszkadza wątrobę u 1% pacjentów. Testujemy lekarstwo na 50 pacjentach. Oblicz prawdopodobieństwo, że

   (a) żaden pacjent nie dozna uszkodzenia choroby,
   (b) co najmniej jeden pacjent dozna uszkodzenia wątroby.

7. Z talii kart wyciągnięto cztery karty. Znaleźć prawdopodobieństwo, tego że będą wśród nich dokładnie dwa asy.

8. W skrzynce znajduje się 47 żarówek dobrych i 3 przepalone. Wyciągamy losowo pięć żarówek. Jakie jest prawdopodobieństwo, że będą wśród nich najwyżej dwie przepalone?

9. Zmienna losowa X ma rozkład jednostajny na przedziale $(-1, 1)$.

   (a) Obliczyć $P(-0,5 < X <= 0,75)$.
   (b) Wyznaczyć liczbę x, dla której $P(-x < X < x) = 0.9$.

10. Czas potrzebny do przeprowadzenia pewnego testu krwi ma rozkład jednostajny na przedziale $(50, 75)$ s. Jaki procent testów

    (a) trwa dłużej niż 70 s.?

    (b) kończy się przed upływem minuty?

11. Zmienne losowe X i Y mają skończone wariancje σ_X^2, σ_Y^2 i wartości oczekiwane m_X, m_Y. Obliczyć

    (a) $\mathrm{E}(-2X+3)$, (b) $\mathrm{E}(2X+3Y)$ (c) $\mathrm{Var}(-X)$ (d) $\mathrm{Var}(-2X+3)$.

    Wyznaczyć także $\mathrm{Var}(2X-3Y+2001)$, zakładając dodatkowo, że X i Y są niezależne.

12. Obliczyć wartość oczekiwaną i wariancję zmiennej losowej X o rozkładzie

    (a) jednostajnym na przedziale (a, b),

    (b) wykładniczym z parametrem λ,

    (c) normalnym $N(m, \sigma^2)$,

    (d) ciągłym z gęstością $f_X(x) = \begin{cases} 1/4 & \text{dla } 0 < x < 1, \\ 3/8 & \text{dla } 3 < x \leq 5, \\ 0 & \text{w pozostałych przypadkach.} \end{cases}$.

13. Zmienna losowa X ma rozkład normalny $N(10, 2^2)$. Wyznaczyć prawdopodobieństwa $P(X < 13)$, $P(X > 9)$, $P(6 < X < 14)$, $P(2 < X < 4)$.

14. Niech Z będzie zmienną losową o rozkładzie normalnym $N(0, 1)$. Korzystając z tablic tego rozkładu znaleźć z_α takie, że

    (a) $P(Z > z_\alpha) = \alpha$ (*ogon jednostronny*),

    (b) $P(|Z| > z_\alpha) = \alpha$ (*ogon dwustronny*).

    Przyjąć $\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$.

15. Niech t_n będzie zmienną losową o rozkładzie t-Studenta z n stopniami swobody. Korzystając z tablic tego rozkładu znaleźć t_α takie, że

    (a) $P(t_n > t_\alpha) = \alpha$ (*ogon jednostronny*),

    (b) $P(|t_n| > t_\alpha) = \alpha$ (*ogon dwustronny*).

    Przyjąć $n = 2$, $n = 10$, $n = 50$ i $\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$.

16. Niech χ_n^2 będzie zmienną losową o rozkładzie chi-kwadrat Pearsona z n stopniami swobody. Korzystając z tablic tego rozkładu znaleźć χ_α^2, takie że $P(\chi_n^2 > \chi_\alpha^2) = \alpha$. Przyjąć $n = 2$, $n = 10$, $n = 50$ i $\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$.

Statystyka stosowana MAP 1079

Lista 1b
(powtórka z rachunku prawdopodobieństwa c.d.)

1. Samolot zabiera 80 osób. Zakładając, że waga pasażerów ma rozkład o wartości oczekiwanej 80 kg i wariancji 10 kg^2 oszacować, za pomocą nierówności Czebyszewa, prawdopodobieństwo tego, że łączna waga pasażerów przekroczy 7000 kg.

2. Prawdopodobieństwo sukcesu w jednej próbie równa się $\frac{1}{4}$. Oszacować ile prób należy wykonać aby prawdopodobieństwo tego, że liczba sukcesów odchyla się od wartości oczekiwanej liczby sukcesów o mniej niż 20% wszystkich prób było większe niż 0.8.

3. Niech X będzie zmienną losową o wartości oczekiwanej m i wariancji σ^2. Oszacować $\mathbb{P}(\ |X - m| < k\sigma)$ dla $k = 1, 2, 3$. Następnie podać wartości tych prawdopodobieństw, gdy X ma rozkład normalny (skorzystać z tablic rozkładu $N(0, 1)$).

4. Dla schematu Bernoulliego z $p = 0,4$ wyliczyć prawdopodobieństwo, że liczba sukcesów w $n = 100$ próbach przekroczy 2. Następnie oszacować to prawdopodobieństwo na podstawie tw. de Moivre'a-Laplace'a. Oszacować błąd przybliżenia. Porównać wyniki.

5. Z partii towaru o wadliwości 3% pobrano próbę 500 elementową. Korzystając z twierdzenie de Moivre'a-Laplace'a oszacować prawdopodobieństwo tego, że liczba wadliwych elementów w próbie

   a) nie będzie większa niż 3, b) nie przekroczy 4%,
   c) przekroczy 9%, d) znajdzie się w przedziale $[10, 20]$.

   Jaki jest błąd otrzymanych przybliżeń?

6. W pewnym towarzystwie ubezpieczeniowym jest ubezpieczonych 100000 samochodów. Każdy z właścicieli płaci roczną składkę 50 zł za samochód. Średnio 9 na 1000 samochodów ulega uszkodzeniu w ciągu roku. Właścicielowi uszkodzonego pojazdu towarzystwo wypłaca 5000 zł. Na podstawie tw. Moivre'a–Laplace'a oszacować, jakie jest prawdopodobienstwo, że w ciągu roku towarzystwo nie poniesie strat. Oszacować błąd przybliżenia.

7. Niech $X_1, \ldots, X_{100}$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie

   (a) Poissona z parametrem $\lambda = 2$,
   (b) wykładniczym z parametrem $\lambda = 0.5$.

   Za pomocą CTG Lindeberga–Lévy'ego oszacować wartość prawdopodobieństwa
   $$\mathbb{P}\left(180 < \sum_{i=1}^{100} X_i < 230\right).$$

8. W grupie studenckiej przeprowadza się test, w którym można uzyskać do 100 punktów. Średni wynik uzyskiwany przez studenta wynosi 40 pkt, a wariancja 202. Wyniki studentów są niezależne i o takim samym rozkładzie. Oszacować na podstawie CTG Lindeberga–Lévy'ego prawdopodobieństwo tego, ze przeciętna liczba punktów przypadająca na jednego studenta w grupie 150 osób zawiera się w przedziale od 35 do 45 pkt.

9. Rzucamy 1000 razy kostką do gry. Wykorzystując centralne twierdzenie graniczne oszacować prawdopodobieństwa, że suma wyrzuconych oczek znajdzie się między 3410 a 3590. Następnie wyznaczyć przedział, do którego ta suma będzie należeć z prawdopodobieństwem co najmniej 0.99.

Statystyka stosowana MAP 1079

Lista 2
(wstępna analiza danych)

1. Oblicz odchylenie standardowe, wariancję i średnią z próby dla każdej z poniższych fikcyjnych próbek. Korzystaj z definicji a nie gotowych funkcji na kalkulatorze.

   a) $16, 13, 18, 13$; b) $38, 30, 34, 38, 35$;
   c) $1, -1, 5, -1$; d) $4, 6, -1, 4, 2$.

   Jak zmienią się te parametry, gdy każdą z powyższych wartości wyrazimy w nowych jednostkach, przyjmując że dla pewnych ustalonych liczb A i B zachodzi:

   nowa wartość = A*stara wartość+B.

2. Dopamina jest związkiem chemicznym, który bierze udział w przekazywaniu sygnałów do mózgu. Farmakolog zmierzył ilość dopaminy w mózgu u siedmiu szczurów. Poziomy dopaminy były następujące (w molach/g): $6.8, 5.3, 6.0, 5.9, 6.8, 7.4, 6.2$. Oblicz

   (a) średnią i standardowe odchylenie z próby,

   (b) medianę oraz pierwszy i trzeci kwartyl z próby.

   Jak zmienią się te parametry, gdy zamiast wartości 6.2 pojawi się wartość 100?

3. Całkowitą ilość protein produkowanych przez krowy mleczne możżna ocenić okresowo badając ich mleko. W tabeli zawarto wartośći całkowitej rocznej produkcji protein (lb.) dla 28 krów rasy Holstein. Dieta i inne warunki były takie same dla wszystkich krów.

   $$425, 481, 477, 434, 410, 397, 438, 545, 528, 496, 502, 529, 500, 465,$$
   $$539, 408, 513, 496, 477, 445, 546, 471, 495, 445, 565, 499, 508, 426.$$

   Wyznacz rozkład częstości i przedstaw go w postaci tabeli, histogramu i wykresu łodyga–liście. Za przedziały klasowe przyjmij:

   $$[380, 410), [410, 440), [440, 470), [470, 500), [500, 530), [530, 560), [560, 590).$$

4. Obserwowano przyrost wagi u byków podczas 140 dniowego okresu testowego. Przeciętny dzienny przyrost wagi (lb./dzień) 13 byków na tej samej diecie jest zawarty w poniższej tabeli:

   $$3.89, 3.51, 3.97, 3.31, 3.21, 3.36, 3.67, 3.24, 3.27, 3.48, 3.52, 3.77, 3.90.$$

   Oblicz średnią i medianę tej próby. Ustal kwartyle i wykonaj wykres pudełkowy.

5. Odnotowano następujące wyniki testów badajacych graniczne wartości naprężeń powodujacych pękanie asfaltu (w megapascalach):
   30 75 79 80 80 105 126 138 149 179 179 191 223 232 232 236 240 242 245 247 254 274 384 470
   a) Ustalić kwartyle i rozstęp miedzykwartylowy.
   b) Narysować wykres pudełkowy dla tych danych.
   c) Opisać kształt rozkładu.

6. Badanie długości czasu T bezawaryjnej pracy 200 elementów danego typu pewnego urządzenia dało następujące wyniki:

przedział	liczba obserwacji	przedział	liczba obserwacji
[0,300)	53	[1800,2100)	9
$[300, 600)$	41	[2100,2400)	7
$[600, 900)$	30	[2400,2700)	5
$[900, 1200)$	22	[2700,3000)	3
$[1200, 1500)$	16	[3000,3300)	2
$[1500, 1800)$	12	$[3300, \infty)$	0

(a) za pomocą średniej i wariancji z próby oszacuj wartość oczekiwaną i wariancją zmiennej T, tzn. parametry $\mathbb{E}(T)$ i $\mathrm{Var}(T)$,

(b) oszacuj $\mathbb{P}(\ T \in [600, 1200)\ )$, tzn. prawdopodobieństwo tego, że zmienna T przyjmie wartość z przedziału $[600, 1200)$.

(c) naszkicuj histogram i porównaj go z wykresem wykresem funkcji

$$f(t) = \begin{cases} \lambda \exp(-\lambda t), & t \geq 0, \\ 0, & t < 0 \end{cases}$$

przyjmując, że nieznany parametr λ ma wartość $\dfrac{1}{\overline{\overline{X}}}$.

Uwaga. W obliczeniach przyjmij, że wszystkie obserwacje z ustalonego przedziału leżą w środku tego przedziału. Rysunek wykonaj za pomocą pakietu Excel.

7. Celem zbadania głębokości utleniania płytek półprzewodnikowych dokonano pomiarów tej wielkości dla 24 losowo wybranych płytek i otrzymano następujące wyniki (w angstremach):

$$425, 431, 416, 419, 421, 436, 418, 410, 431, 433, 423, 426,$$
$$410, 435, 436, 428, 411, 426, 409, 437, 422, 428, 413, 416.$$

Wyznacz wartości estymatorów następujących parametrów charakteryzujących populację płytek półprzewodnikowych:

(a) średniej głębokości utleniania,

(b) odchylenia standardowego głębokości utleniania,

(c) mediany głębokości utleniania,

(d) frakcji elementów, dla których głębokość utleniania jest większa niż 430 angstremów.

Statystyka stosowana MAP 1079

Lista 3
(estymatory i ich własności)

1. Niech $X_1, \ldots, X_7$ będzie próbą prostą z populacji o wartości oczekiwanej μ i wariancji σ^2 i niech

$$\hat{\mu}_1 = \frac{X_1 + \ldots + X_7}{7} \quad \hat{\mu}_2 = \frac{2X_1 - X_6 + X_4}{2}.$$

   będą dwoma estymatorami nieznanej średniej μ.

   (a) Który z tych estymatorów jest nieobciążony?

   (b) Który z tych estymatorów ma mniejszy błąd średniokwadratowy?

2. Niech $X_1, \ldots, X_n$ będzie próbą prostą z populacji o rozkładzie jednostajnym na $(0, \theta)$. Za pomocą metody momentów wyznacz estymator $\hat{\theta}_n$ parametru θ.

   (a) Wykorzystując ten estymator oszacuj θ dla próby $0.1, 0.3, 0.7, 0.8, 0.2$.

   (b) Oblicz obciążenie i błąd średniokwadratowy $\hat{\theta}_n$.

   (c) Zbadaj zgodność $\hat{\theta}_n$.

   Wyznacz estymator największej wiarogodności parametru θ i oblicz jego wartość dla próby z punktu (a).

3. Niech $X_1, \ldots, X_n$ będzie próbą prostą z populacji o rozkładzie ciągłym o gęstości

$$f(x) = \begin{cases} (\theta+1)x^{\theta}, & \text{gdy } 0 \le x \le 1, \\ 0, & \text{w przeciwnym razie.} \end{cases}$$

   (a) Znajdź estymator największej wiarogodności $\hat{\theta}$ parametru θ.

   (b) Wykorzystując $\widehat{\theta}$ oszacuj θ na podstawie próby $0.1, 0.3, 0.7, 0.8, 0.2$.

4. Niech $X_1, \ldots, X_n$ będzie próbą z populacji o rozkładzie trzypunktowym

$$P(X = -1) = p, \quad P(X = 0) = 0.4 - p, \quad P(X = 1) = 0.6.$$

   Za pomocą metody momentów wyznacz estymator $\widehat{p}_n$ parametru p. Wykorzystując ten estymator oszacuj p dla próby $-1, 0, 0, 1, 1, 1, -1, 0, -1, 1$. Czy to oszacowanie jest sensowne?

5. Na podstawie próby prostej $X_1, \ldots, X_n$ wyznacz estymatory

   (a) parametru p w rozkładzie geometrycznym;

   (b) parametrów m i σ^2 w rozkładzie normalnym $N(m; \sigma^2)$;

   (c) parametru θ w rozkładzie jednostajnym $U(0, \theta)$,

   (d) parametru c w rozkładzie o gęstości $f(x) = \begin{cases} 0 & \text{dla} \quad x \le 1, \\ \dfrac{c}{x^{c+1}} & \text{dla} \quad x > 1. \end{cases}$

   Wykorzystaj metodę momentów i metodę największej wiarygodności.

6. Niech $X_1, \ldots, X_n$ będzie próbą prostą z populacji o rozkładzie ciągłym o gęstości

$$f(x) = \begin{cases} c(1+\theta x), & \text{gdy } -1 \leq x \leq 1, \\ 0, & \text{w przeciwnym razie.} \end{cases}$$

   (a) znajdź c,

   (b) za pomocą metody momentów wyznacz estymator parametru θ,

   (c) czy można jawnie wyznaczyć estymator największej wiarogodności dla θ?

Statystyka stosowana MAP 1079

Lista 4
(przedziały ufności dla średniej i proporcji)

Oznaczenia: Średnia, wariancja i odchylenie standardowe w próbie $x_1, x_2, \ldots, x_n$ to

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2, \quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

1. Dla danych $-0.1, 0.15, 0.1, -0.05$, oszacować na poziomie ufności $1-\alpha = 0.9$ wartość oczekiwaną przyjmując, że rozkład jest normalny oraz $\sigma = 0.1$.

2. Z populacji o rozkładzie normalnym $N(m, \sigma^2)$ pobrano próbę pięcioelementową: $2.15, 2.08, 2.17, 1.95, 2.15$. Znaleźć przedział ufności dla wartości oczekiwanej na poziomie ufności $1 - \alpha = 0.9$, wiedząc że $\sigma = 1/20$.

3. Średni czas świecenia lampy, obliczony na podstawie próby losowej rozmiaru $n = 100$, wynosi 1000 godzin. Na poziomie ufności $1 - \alpha = 0.95$ wyznaczyć przedział ufności dla średniego czasu świecenia lampy z całej partii, jeśli wiadomo, że odchylenie standardowe długości świecenia lampy wynosi $\sigma = 40$ godzin.

4. Wytrzymałość pewnego materiału budowlanego ma rozkład normalny $N(m, \sigma^2)$. Próba $n = 5$ elementowa wylosowanych sztuk tego materiału dała wyniki: $\bar{x} = 20.8$ N/cm^2, $s = 2.8$N/cm^2.

   (a) Na poziomie ufności 0.99 zbudować przedział ufności dla średniej m.

   (b) Na poziomie ufności 0.95 zbudować przedział ufności dla wariancji σ^2.

5. Struktura wieku osób zatrudnionych w firmie komputerowej jest następująca:

Wiek w latach	18–20	20–22	22–24	24–26	26–28	28–30
Liczba osób	3	7	10	9	4	2

   Zakładając, że wiek osób ma rozkład normalny $N(m, \sigma^2)$, wyznaczyć przedział ufności dla wariancji wieku na poziomie ufności $0,98$.

6. Dokonano pomiarów zawartości pewnego enzymu w tkance 9 grzybów w pewnych ustalonych warunkach eksperymentalnych. Średnia z tych pomiarów wyniosła 5111 jednostek a odchylenie standardowe 818 jednostek.

   (a) Załóżmy, że zawartość badanego enzymu w populacji grzybów ma rozkład normalny. Skonstruuj 95% przedział ufności dla średniej zawartości tego enzymu w tkance grzybów.

   (b) Podaj interpretację skonstruowanego przedziału ufności.

   (c) W jaki sposób można zweryfikować założenie o normalności rozkładu.

7. Oblicz średnią, odchylenie standardowe i błąd standardowy średniej w pięcioelementowej próbie: $10.0, 8.9, 9.1, 11.7, 7.9$. Skonstruuj 90% przedział ufności dla μ, przy założeniu, że obserwacje pochodzą z rozkładu normalnego.

8. Zoolog zmierzył długość ogona u 86 myszy leśnych. Średnia długość ogona wyniosła 60.43 mm a odchylenie standardowe z próby 3.06 mm. 95% przedział ufności dla średniej długości ogona w tej populacji myszy wynosi $[59.77, 61.09]$.

   (a) Prawda czy fałsz (uzasadnij): Mamy 95% pewności, że średnia długość ogona w naszej próbie zawiera się między 59.77 mm a 61.09 mm.

   (b) Prawda czy fałsz (uzasadnij): Mamy 95% pewności, że średnia długość ogona w populacji myszy zawiera się w przedziale między 59.77 mm a 61.09 mm.

9. W celu sprawdzenia czy pewien lek obniża ciśnienie krwi u chorych na nadciśnienie, wylosowano $n = 20$ pacjentów i zmierzono im ciśnienie przed podaniem tego leku i po pewnym czasie po podaniu. Otrzymano następujące wyniki:

Przed podaniem	320	200	340	240	200	300	240	290	180	210
Po podaniu	270	180	260	250	150	260	200	310	150	220

Przed podaniem	250	300	180	270	290	200	280	190	220	290
Po podaniu	270	260	200	240	250	160	210	160	170	220

   Zakładając normalność odpowiedniego rozkładu skonstruuj 95% przedział ufności dla różnicy między średnim ciśnieniem przed i po podaniu leku.

10. Na podstawie danych z dwóch niezależnych próbek o liczności $n_1 = 10$ i $n_2 = 20$, wylosowanych z populacji o rozkładach normalnych, otrzymano następujące wartości średnich z prób badanej cechy: $\bar{x} = 14.3$ i $\bar{y} = 12.2$. Wariancje cech w obu populacjach są znane i wynoszą $\sigma_1^2 = 22$, $\sigma_2^2 = 18$. Skonstruuj 99% przedział ufności dla różnicy między średnim wartościami tych cech.

11. Cechy X i Y w dwóch populacjach mają rozkłady normalne o tej samej wariancji. Z dwóch niezależnych prób prostych o liczebnościach odpowiednio 100 i 120 obliczono $\bar{x} = 1.15$, $s_x^2 = 2.4$ oraz $\bar{y} = 1.05$, $s_y^2 = 2.3$. Skonstruuj 95% przedział ufności dla różnicy między średnim wartościami tych cech. Czy można twierdzić, że średnie w tych populacjach są takie same?

12. Przeprowadzono badanie krwi u 70 orangutanów i stwierdzono, że 14 z nich ma grupę krwi B. Skonstruuj 95% przedział ufności dla proporcji p występowania grupy krwi B u orangutanów.

13. Spośród 500 losowo wybranych mieszkańców Kalifornii 302 opowiedziało się za dopuszczalnością kary śmierci. Skonstruuj 99% przedział ufności dla oszacowania proporcji p wszystkich mieszkańców Kalifornii, którzy są za dopuszczalnością kary śmierci.

14. Spośród 296 losowo wybranych kobiet 63 stwierdziło, że przy kupowaniu koszul zwracają uwagę ma markę towaru. Wśród 251 mężczyzn 27 przyznało się do analogicznego zachowania. Skonstruuj 90% przedział ufności dla oszacowania różnicy $p_K - p_M$ między proporcjami kobiet i mężczyzn zwracających uwagę na markę kupowanych koszul.

Statystyka stosowana MAP 1079

Lista 5

(testy dla średniej i proprorcji)

1. Pewien automat w fabryce czekolady wytwarza tabliczki czekolady o nominalnej wadze 250 g. Wiadomo, że rozkład wagi produkowanych tabliczek jest normalny $N(m, 5^2)$. Kontrola techniczna pobrała w pewnym dniu próbę losową 16 tabliczek czekolady i otrzymała średnią wagę 244 g. Czy można stwierdzić, że automat rozregulował się i produkuje tabliczki czekolady o mniejszej niż przewiduje norma wadze? Na poziomie istotności $\alpha = 0.05$ zweryfikuj odpowiednią hipotezę.

2. Spośród pacjentów szpitala, leczonych na pewną chorobę, wylosowano próbę 26 chorych, którym następnie zmierzono ciśnienie tętnicze. Okazało się, że średnia i odchylenie standardowe z próby były równe $\overline{x} = 135$ i $s = 40$. Zakładając, że ciśnienie ma rozkład normalny $N(m, \sigma^2)$ zweryfikować hipotezę, że pacjenci pochodzą z populacji o średnim ciśnieniu tętniczym 120. Przyjąć poziom istotności $\alpha = 0.05$.

3. Firma doradztwa inwestycyjnego zapewnia, że przeciętny przychód z akcji w pewnej gałęzi przemysłu wynosi $11,5$ %. Inwestor chce sprawdzić tę opinię, więc pobiera próbę złożoną z akcji 100 spółek należących do tej gałęzi i stwierdza, że średni przychód z akcji w próbie wynosi $\bar{x} = 10,8$ % przy odchyleniu standardowym z próby $s = 3,4$ %. Czy inwestor ma dostateczne powody do odrzucenia zapewnienia firmy doradczej na poziomie istotności $\alpha = 0,05$?

4. W celu sprawdzenia czy pewien lek obniża ciśnienie krwi u chorych na nadciśnienie, wylosowano $n = 20$ pacjentów i zmierzono im ciśnienie przed podaniem tego leku i po pewnym czasie po podaniu. Otrzymano następujące wyniki:

Przed podaniem	320	200	340	240	200	300	240	290	180	210
Po podaniu	270	180	260	250	150	260	200	310	150	220

Przed podaniem	250	300	180	270	290	200	280	190	220	290
Po podaniu	270	260	200	240	250	160	210	160	170	220

   Zakładając normalność odpowiedniego rozkładu zweryfikuj hipotezę, że lek obniża ciśnienie krwi (odpowiednio dobierając hipotezę zerową i alternatywną). Przyjmij poziom istotności $\alpha = 0.05$

5. Na podstawie danych z dwóch niezależnych próbek o liczności $n_1 = 10$ i $n_2 = 20$, wylosowanych z populacji o rozkładach normalnych, otrzymano następujące wartości średnich z prób badanej cechy: $\bar{x} = 14.3$ i $\bar{y} = 12.2$. Wariancje cech w obu populacjach są znane i wynoszą $\sigma_1^2 = 22$, $\sigma_2^2 = 18$. Na poziomie istotności 0.05 zweryfikować hipotezę o równości średnich, tzn. $H : \mu_1 = \mu_2$, wobec hipotezy alternatywnej $K : \mu_1 \neq \mu_2$.

6. Cechy X i Y w dwóch populacjach mają rozkłady normalne o tej samej wariancji. Z dwóch niezależnych prób prostych o liczebnościach odpowiednio 100 i 120 obliczono $\bar{x} = 1.15$, $s_x^2 = 2.4$ oraz $\bar{y} = 1.05$, $s_y^2 = 2.3$. Na poziomie istotności $\alpha = 0.05$ zweryfikuj hipotezę o równości tych średnich, przyjmując alernatywę jednostronną.

7. Z dwóch dużych partii słupków betonowych wybrano próbki o liczebnościach $n_1 = 90$ oraz $n_2 = 110$. Średnie wytrzymałości na ściskanie osiowe obliczone z tych

próbek wynosiły: $\bar{x} = 248.31\,\mathrm{kG/cm^2}$, $\bar{y} = 240.2\,\mathrm{kG/cm^2}$, a odchylenia standardowe odpowiednio $s_x = 2\,\mathrm{kG/cm^2}$ i $s_y = 1.7\,\mathrm{kG/cm^2}$. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę o jednakowej wytrzymałości słupków w obu partiach.

8. Pobrano dwie losowe próby ziaren fasoli dwóch gatunków i zmierzono długości tychże ziaren. Dla gatunku A otrzymano $\overline{x} = 12.3$ mm, $s_x = 1.8$ mm, natomiast dla gatunku B otrzymano $\overline{y} = 11.9$ mm, $s_y = 2.1$ mm. Wiedząc, że liczebności tych prób wynosiły odpowiednio $n = 450$ i $m = 500$, zweryfikować hipotezę o równości średnich długości ziaren obu gatunków fasoli. Przyjąć poziom istotności $\alpha = 0.05$.

9. Przeprowadzono badanie krwi u 70 orangutanów i stwierdzono, że 14 z nich ma grupę krwi B. Niech p oznacza nieznaną proporcję występowania grupy krwi B u orangutanów. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę że $p = 0.25$ przy hipotezie alternatywnej $p < 0.25$.

10. Spośród 296 losowo wybranych kobiet 63 stwierdziło, że przy kupowaniu koszul zwracają uwagę ma markę towaru. Wśród 251 mężczyzn 27 przyznało się do analogicznego zachowania. Na poziomie istotności $\alpha = 0.01$ zweryfikuj hipotezę zerową o równości proporcji p_K i p_M kobiet i mężczyzn zwracających uwagę na markę kupowanych koszul.

Statystyka stosowana MAP 1079

Lista 6
(testy zgodności i niezależności chi-kwadrat)

1. W celu sprawdzenia czy kostka sześcienna do gry jest rzetelna (symetryczna), wykonano 120 rzutów tą kostką i otrzymano następujące wyniki:

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

   Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że kostka jest symetryczna.

2. W pewnej fabryce zaobserwowano następujący rozkład absencji w tygodniu, zbadany w wylosowanej grupie 900 pracowników z absencją:

Dzień tygodnia	PN	WT	ŚR	CZ	PT	SOB
Liczba nieobecnych	200	160	140	140	100	160

   Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że absencja w tej fabryce jest jednakowa w każdym dniu tygodnia.

3. Zbadano 300 losowo wybranych 5–sekundowych odcinków czasowych pracy pewnej centrali telefonicznej i otrzymano empiryczny rozkład liczby zgłoszeń :

Liczba zgłoszeń	0	1	2	3	4	5
Liczba odcinków	50	100	80	40	20	10

   Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że liczba zgłosień w tej centrali jest zmienną losową o rozkładzie Poissona.

4. Dla 200 próbek betonu przeprowadzono badanie wytrzymałości na ściskanie i uzyskano następujące wyniki (w [N/cm^2]):

wytrzymałość	liczba próbek
1900-2000	10
2000-2100	26
2100-2200	56
2200-2300	64
2300-2400	30
2400-2500	14

   Na poziomie istotności $\alpha = 0,05$ zweryfikować hipotezę, że rozkład wytrzymałości jest rozkładem normalnym.

5. W celu zweryfikowania hipotezy, że studentki pewnej uczelni lepiej zdają egzaminy niż studenci, wylosowano próbę 200 studentek i studentów i otrzymano następujące wyniki zaliczenia letniej sesji egzaminacyjnej:

	Zdany	Oblany
Studenci	55	45
Studentki	75	25

Na poziomie istotności $\alpha = 0.05$ zweryfikować, za pomocą testu chi-kwadrat, hipotezę o niezależności wyników egzaminacyjnych od płci.

6. Pewien produkt można wytwarzać trzema metodami. Wysunięto hipotezę, że wadliwość produkcji nie zależy od metody wytwarzania. Wylosowano niezależnie próbę 270 sztuk wyrobu i otrzymano następujące wyniki badania jakości dla poszczególnych metod:

Jakość	Metoda I	Metoda II	Metoda III
Dobra	40	80	60
Zła	10	60	20

Na poziomie istotności $\alpha = 0.05$ zweryfikuj hipotezę o niezależności jakości produkcji od metod produkcji.

Statystyka stosowana MAP 1079

Lista 7
(regresja liniowa)

1. Rodzicom pięcioletniej Kasi wydaje się, że ich córka rośnie zbyt wolno. W poniższej tabeli podano informacje dotyczące wzrostu dziecka:

Wiek (w miesiącach)	36	48	51	54	57	60
Wzrost (w cm)	86	90	91	93	94	95

   (a) Sporządź wykres rozproszenia, odpowiadający tym danym. Czy na podstawie tego rysunku można stwierdzić, że istnieje liniowa zależność między wzrostem a wiekiem Kasi?

   (b) Za pomocą metody najmniejszych kwadratów wyznacz równanie prostej regresji, opisującej zależność wzrostu Kasi od jej wieku.

   (c) Wykorzystując otrzymany model podaj prognozę wzrostu Kasi w wieku 40 i 60 miesięcy.

   (d) O ile średnio rośnie Kasia w ciągu miesiąca? Prawidłowo rozwijająca się dziewczynka przyrasta o 6 centymetrów między czwartym a piątym rokiem życia. Czy Kasia rośnie wolniej niż powinna?

2. Zbadano średnią grubość pni pewnego gatunku drzew w zależności od średniej temperatury otoczenia w ciągu dnia. Uzyskano następujące wyniki:

Temperatura (w °C)	15	20	25	30
Grubość pni (w cm)	25	26	28	29

   Za pomocą metody najmniejszych kwadratów skonstruować prostą regresji zależności grubości pni od temperatury. otoczenia. Wykorzystując otrzymany model znaleźć temperaturę, dla której prognozowana grubość pni osiągnie 30cm.

3. Naukowcy mają nadzieję, że pewien kwas pomoże ograniczyć rozwój chorób roślin uprawnych wywołanych przez grzyby. W poniższej tabeli podsumowano wyniki dotyczące rozwoju grzyba Pythium ultimum w naczyniach wypełnionych roztworem tego kwasu o różnych stężeniach. Każda wielkość opisująca wzrost jest średnią z czterech radialnych pomiarów kolonii tego grzyba po 24 godzinach przebywania w badanym środowisku. Każdemu stężeniu odpowiadają dwa naczynia.

stężenie kwasu	wzrost grzyba	stężenie kwasu	wzrost grzyba
X(g/ml)	Y(mm)	X(g/ml)	Y(mm)
0	33.3	10	25.5
0	31	10	23.8
3	29.8	20	18.3
3	27.8	20	15.5
6	28	30	11.7
6	29	30	10

   a) Wyznacz prostą regresji Y na X.

b) Wyznacz estymator $\widehat{\sigma}^2$ wariancji σ^2.

c) Znajdź błąd standardowy estymatora współczynnika nachylenia prostej regresji β_1.

d) Czy nasze dane potwierdzają oczekiwania, że badany kwas ogranicza wzrost grzyba? Zastosuj odpowiedni test statystyczny.

4. Za pomocą modelu regresji liniowej zbadano zależność między temperaturą powierzchni chodnika (x) a jego wygięciem (y). Na podstawie $n = 20$ pomiarów wyznaczono następujące wielkości:

$$\sum_{i=1}^{n} y_i = 12.75 \ \sum_{i=1}^{n} y_i^2 = 8.86, \ \sum_{i=1}^{n} x_i = 1478, \ \sum_{i=1}^{n} x_i^2 = 143,215.8, \ \sum_{i=1}^{n} x_i y_i = 1083.67.$$

(a) Wyznacz estymatory parametrów β_0 i β_1.

(b) Narysuj prostą regresji.

(c) Wykorzystaj otrzymany model do prognozy wygięcia chodnika przy temperaturze 85°F.

(d) Jakie jest średnie wygięcie chodnika mającego temperaturę 90°F?

(e) Jaka zmiana średnigo wygięcia chodnika odpowiada zmianie temperatury jego powierzchni o 1°F?