PRZYKŁADOWE PYTANIA NA SPRAWDZIAN KOŃCOWY

1. Na czym polega różnica pomiędzy analizą danych a wnioskowaniem statystycznym?

2. Podać miary

   (a) tendencji centralnej,

   (b) pozycji,

   (c) rozproszenia

   dla danych ilościowych.

3. Drugi decyl to

   (a) który percentyl,

   (b) kwantyl którego rzędu?

4. Wymienić i opisać (bez podawania wzorów) miary kształtu zbioru danych ilościowych.

5. Dla następujących danych: $7, 5, 1, 9, 5, 3$ sporządzić i opisać wykres skrzynkowy.

6. Co to znaczy, że $X_1, X_2, \ldots, X_n$ jest prostą próbą losową z X?

7. Niech $X_1, X_2, \ldots, X_n$ będzie próbą losową z rozkładu Poissona z parametrem $\theta > 0$, tzn. z rozkładu dyskretnego o funkcji masy prawdopodobieństwa $P(X = x) = \mathrm{e}^{-\theta}\frac{\theta^x}{x!}$, $x = 0, 1, 2, \ldots$. Wyznaczyć estymator parametru θ, stosując metodę największej wiarygodności.
   Następnie wyznaczyć estymator największej wiarygodności odchylenia standardowego X_1. Przytoczyć, wykorzystywane do jego wyznaczenia, twierdzenie. Warto pamiętać, że dla zmiennej losowej X o rozkładzie Poissona z parametrem θ mamy $Var(X) = \theta$.

8. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu Pareto o gęstości danej wzorem $f_a(x) = \begin{cases} \frac{a}{x^{a+1}} & \text{dla } x \geq 1 \\ 0 & \text{dla } x < 1 \end{cases}$. Stosując metodę największej wiarygodności, wyznaczyć

   (a) estymator parametru a,

   (b) estymator prawdopodobieństwa $P(X_1 < 2)$.

9. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu gamma $Gamma(a, 1/s)$, tzn. z rozkładu o gęstości

$$f_{a,s}(x) = \begin{cases} \frac{1}{s^a \Gamma(a)} x^{a-1} \exp\left(-x/s\right) & \text{dla } x > 0 \\ 0 & \text{dla } x \leq 0 \end{cases}, \; a > 0, s > 0,$$

   gdzie $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$. Załóżmy, że $a > 0$ jest znane. Wyznaczyć estymator największej wiarygodności

   (a) parametru s,

   (b) wartości oczekiwanej X_1,

   (c) wariancji X_1.

   Wiadomo, że gdy zmienna losowa X ma rozkład $Gamma(a, \beta)$, to $E(X) = a\beta$, $Var(X) = a\beta^2$.

10. Zdefiniować błąd I-go rodzaju, błąd II-go rodzaju i poziom istotności oraz moc testu.

11. Zdefiniować współczynnik zwany *p-value*. Omówić zasadę posługiwania się tą wielkością.

12. Zakłada się, że napięcie wyjściowe na zasilaczu ma rozkład normalny. Przeprowadzono szesnaście niezależnych pomiarów napięcia (w woltach) i otrzymano średnie napięcie z próby równe $12,5$ wolta oraz wariancję z próby równą $2,25$ wolta^2.

(a) Chcemy sprawdzić czy średnie napięcie przekracza 12 woltów. Jakie hipotezy postawimy? Ile wyniesie statystyka testowa odpowiedniego testu? Podać wzór na zbiór krytyczny tego testu.

(b) Zakładając, że rzeczywiste średnie napięcie to 12,5 wolta, wyznaczyć prawdopodobieństwo, że test używany w pkt. (a)

   i. przyjmie hipotezę, że średnie napięcie przekracza 12 woltów;

   ii. da odpowiedź, że średnie napięcie nie przekracza 12 woltów.

   Odpowiedzi zapisać za pomocą funkcji $m(\cdot)$, gdzie $m(\cdot)$ to funkcja mocy testu używanego w pkt. (a).

(c) Chcemy sprawdzić czy wariancja napięcia wynosi 3 wolt^2. Jakie hipotezy postawimy? Ile wyniesie statystyka testowa odpowiedniego testu? Podać wzór na zbiór krytyczny tego testu.

13. Wiadomo, że 50% mężczyzn, którzy przeżyli zawał serca, umiera na skutek drugiego zawału. Na 100 losowo wybranych pacjentów, poddanych po pierwszym zawale terapii nowym lekiem, 40 zmarło w wyniku drugiego zawału. Chcemy sprawdzić czy przeprowadzone badanie potwierdza, że nowy lek zwiększa szansę przeżycia drugiego zawału. Jakie hipotezy postawimy? Jakiego użyjemy testu? Dlaczego można użyć właśnie tego testu? Ile wyniesie statystyka testowa tego testu? Podać wzór na zbiór krytyczny tego testu.

14. Niech $moc(\cdot)$ oznacza funkcję mocy testu używanego w zadaniu 13. Za pomocą tej funkcji zapisać prawdopodobieństwo, że test, używany w zadaniu 13, da odpowiedź, że nowy lek nie zwiększa szansy przeżycia drugiego zawału w sytuacji, gdy rzeczywiste prawdopodobieństwo przeżycia drugiego zawału dla mężczyzny leczonego nowym lekiem to 0,38.

15. Pewna firma szyje prześcieradła o nominalnej długości 200 cm. Jest jednak podejrzenie, że średnia długość tych prześcieradeł jest mniejsza niż 200 cm. Aby sprawdzić czy owo podejrzenie jest słuszne, zmierzono 100 losowo wybranych prześcieradeł pochodzących od tej firmy i otrzymano średnią długość 190 cm z odchyleniem standardowym długości wynoszącym 10 cm. Jakie hipotezy należy postawić? Ile wyniesie statystyka testowa odpowiedniego testu? Podać wzór na zbiór krytyczny tego testu.

16. 55 spośród 100 absolwentów techników oraz 45 spośród 100 absolwentów liceów nie rozwiązało poprawnie zadań z geometrii podczas egzaminu maturalnego. Chcemy sprawdzić czy na podstawie tych danych można stwierdzić, że absolwenci techników są słabiej przygotowani do egzaminu z matematyki niż absolwenci liceów. Jakie hipotezy postawimy? Jakiego użyjemy testu? Dlaczego można użyć właśnie tego testu? Ile wyniesie statystyka testowa tego testu? Podać wzór na zbiór krytyczny tego testu.

17. Hotel zaopatruje się w pościel u dwóch firm: firmy A i firmy B. Zbadano wytrzymałość 100 losowo wybranych prześcieradeł pochodzących od firmy A i 100 losowo wybranych prześcieradeł pochodzących od firmy B. Dla firmy A otrzymano średnią długość przydatności prześcieradła 1000 dni z odchyleniem standardowym 300 dni, zaś dla firmy B - średnią długość przydatności przećieradła 1200 dni z odchyleniem standardowym 400 dni. Na poziomie istotności 0,05 chcemy sprawdzić czy na podstawie tych danych można uznać, że prześcieradła produkowane przez firmę A są mniej wytrzymałe niż prześcieradła pochodzące od firmy B. Jakie hipotezy postawimy? Ile wyniesie statystyka testowa odpowiedniego testu? Podać wzór na zbiór krytyczny tego testu.

18. Niech $power(\cdot)$ oznacza funkcję mocy testu używanego w zadaniu 17 (argumentem tej funkcji jest $\mu_A - \mu_B$). Otrzymaliśmy, że $power(-500) = 0,89$. Podać interpretację tego wyniku.

19. Niech $\theta \in \Theta$ i $\Theta_0 \subset \Theta$. Jakie warunki musi spełniać test by był to jednostajnie najmocniejszy test na poziomie istotności α dla $H_0 : \theta \in \Theta_0$ przeciwko $H_1 : \theta \in \Theta \setminus \Theta_0$?

20. Przytoczyć lemat Neymana-Pearsona.

21. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu $Gamma(2, s)$, czyli rozkładu o gęstości $f_s(x) = \begin{cases} s^2 x e^{-sx}, & x > 0 \\ 0, & x \leq 0 \end{cases}$, gdzie $s > 0$.

    (a) Wyznaczyć najmocniejszy test na poziomie istotności α służący do weryfikacji hipotezy $H_0 : s = s_0$ przeciwko $H_1 : s = s_1$, gdzie $0 < s_0 < s_1$.

    (b) Wyznaczyć jednostajnie najmocniejszy test na poziomie istotności α do weryfikacji hipotezy $H_0 : s = s_0$ przeciwko $H_1 : s > s_0$.

    Podać statystyki testowe i wzory na zbiory krytyczne szukanych testów.

22. Jakich metod graficznych można użyć by sprawdzić czy analizowane dane pochodzą z rozkładu normalnego?

23. Co to jest wykres normalności i do czego służy? Omówić budowę tego wykresu i sposób korzystania z niego.

24. Wymienić uniwersalne testy normalności. Sformuować hipotezy, które w tych testach weryfikujemy. Który z tych testów uznawany jest za najlepszy?

25. Chcemy sprawdzić czy kostka do gry jest symetryczna. W tym celu wykonaliśmy 120 rzutów i wyniki zebraliśmy w poniższej tabeli

Liczba oczek	1	2	3	4	5	6
ile razy wypadła ta liczba oczek	10	20	30	20	20	20

    Jakie postawimy hipotezy? Jakiego użyjemy testu w celu ich weryfikacji? Ile wynosi statystyka testowa tego testu?

ODPOWIEDZI:

7. $\hat{\theta}_{NW} = \bar{X}$, $\hat{\sigma}_{NW} = \sqrt{\bar{X}}$

8. (a) $\hat{a}_{NW} = \frac{n}{\sum_{i=1}^{n} \ln X_i}$, (b) $1 - 2^{\frac{-n}{\sum_{i=1}^{n} \ln X_i}}$

9. (a) $\hat{s}_{NW} = a/\bar{X}$, (b) $E(\hat{X}_1)_{NW} = \bar{X}$, (c) $Var\hat{(}X_1)_{NW} = \frac{(\bar{X})^2}{a}$

12. Niech X to zmienna losowa opisująca napięcie wyjściowe na zasilaczu (w woltach).

(a) $H_0 : \mu = 12$, $H_1 : \mu > 12$; $T = \frac{4}{3} \approx 1,33$; $W = \langle t_{1-\alpha,n-1}; +\infty)$, gdzie $t_{1-\alpha,n-1}$ to kwantyl rzędu $1 - \alpha$ rozkładu t-Sudenta o $n - 1 = 15$ stponich swobody zaś α to poziom istotności testu

(b) i. $m(12, 5)$ ii. $1 - m(12, 5)$

(c) $H_0 : \sigma^2 = 3$, $H_1 : \sigma^2 \neq 3$; $\chi^2 = 11\frac{1}{4} = 11,25$; $W = \left(0, \chi^2_{\alpha/2;n-1}\right\rangle \cup \left\langle \chi^2_{1-\alpha/2;n-1}; +\infty\right)$, gdzie $\chi^2_{\alpha/2;n-1}$ oraz to $\chi^2_{1-\alpha/2;n-1}$ to kwantyle rozkładu chi-kwadrat o $n - 1 = 15$ stponich swobody rzędu odpowiednio $\alpha/2$ i $1 - \alpha/2$, zaś α to poziom istotności testu

13. Niech $X = \begin{cases} 1 & \text{jeśli mężczyzna, co przeżył zawał serca, umiera w wyniku drugiego zawału} \\ 0 & \text{jeśli mężczyzna, co przeżył zawał serca, nie umiera w wyniku drugiego zawału} \end{cases}$

$H_0 : p = 0,5$, $H_1 : p < 0,5$; $U = -2$; $W = (-\infty; -u_{1-\alpha}\rangle$, gdzie $u_{1-\alpha}$ to kwantyl rzędu $1 - \alpha$ standardowego rozkładu normalnego, zaś α to poziom istotności testu

UWAGA: Równie dobrze można przyjąć:

$X = \begin{cases} 0 & \text{jeśli mężczyzna, co przeżył zawał serca, umiera w wyniku drugiego zawału} \\ 1 & \text{jeśli mężczyzna, co przeżył zawał serca, nie umiera w wyniku drugiego zawału} \end{cases}$

Wówczas:

$H_0 : p = 0,5$, $H_1 : p > 0,5$; $U = 2$; $W = \langle u_{1-\alpha}; +\infty)$, gdzie $u_{1-\alpha}$ to kwantyl rzędu $1-\alpha$ standardowego rozkładu normalnego, zaś α to poziom istotności testu

14. Jeśli przyjmiemy konwencję

$X = \begin{cases} 1 & \text{jeśli mężczyzna, co przeżył zawał serca, umiera w wyniku drugiego zawału} \\ 0 & \text{jeśli mężczyzna, co przeżył zawał serca, nie umiera w wyniku drugiego zawału} \end{cases}$,

to odpowiedzią będzie $1 - moc(0,62)$.

Natomiast jeśli przyjmiemy

$X = \begin{cases} 0 & \text{jeśli mężczyzna, co przeżył zawał serca, umiera w wyniku drugiego zawału} \\ 1 & \text{jeśli mężczyzna, co przeżył zawał serca, nie umiera w wyniku drugiego zawału} \end{cases}$,

to odpowiedzią będzie $1 - moc(0,38)$.

15. Niech X to zmienna losowa opisująca długość prześcieradła (w cm).

$H_0 : \mu = 200$, $H_1 : \mu < 200$; $U = -10$; $W = (-\infty; -u_{1-\alpha}\rangle$, gdzie $u_{1-\alpha}$ to kwantyl rzędu $1-\alpha$ standardowego rozkładu normalnego, zaś α to poziom istotności testu

16. Niech

$X = \begin{cases} 1 & \text{jeśli absolwent technikum na maturze nie rozwiązał poprawnie zadań z geometrii} \\ 0 & \text{jeśli absolwent technikum na maturze rozwiązał poprawnie zadania z geometrii} \end{cases}$

$Y = \begin{cases} 1 & \text{jeśli absolwent liceum na maturze nie rozwiązał poprawnie zadań z geometrii} \\ 0 & \text{jeśli absolwent liceum na maturze rozwiązał poprawnie zadania z geometrii} \end{cases}$

$H_0 : p_{tech} = p_{lic}$, $H_1 : p_{tech} > p_{lic}$; $U = \frac{\sqrt{50}}{5} \approx 1,414$; $W = \langle u_{1-\alpha}; +\infty)$, gdzie $u_{1-\alpha}$ to kwantyl rzędu $1-\alpha$ standardowego rozkładu normalnego, zaś α to poziom istotności testu

UWAGA: Równie dobrze można przyjąć:

$X = \begin{cases} 0 & \text{jeśli absolwent technikum na maturze nie rozwiązał poprawnie zadań z geometrii} \\ 1 & \text{jeśli absolwent technikum na maturze rozwiązał poprawnie zadania z geometrii} \end{cases}$

$Y = \begin{cases} 0 & \text{jeśli absolwent liceum na maturze nie rozwiązał poprawnie zadań z geometrii} \\ 1 & \text{jeśli absolwent liceum na maturze rozwiązał poprawnie zadania z geometrii} \end{cases}$

Wówczas:

$H_0 : p_{tech} = p_{lic}$, $H_1 : p_{tech} < p_{lic}$; $U = -\frac{\sqrt{50}}{5} \approx -1,414$; $W = (-\infty; -u_{1-\alpha}\rangle$, gdzie $u_{1-\alpha}$ to kwantyl rzędu $1-\alpha$ standardowego rozkładu normalnego, zaś α to poziom istotności testu

17. Niech X to zmienna losowa opisująca wytrzymałość (w dniach) prześcieradeł od firmy A, zaś Y to zmienna losowa opisująca wytrzymałość (w dniach) prześcieradeł od firmy B.

$H_0 : \mu_A = \mu_B$, $H_1 : \mu_A < \mu_B$; $U = -4$; $W = (-\infty; -u_{0,95}\rangle$, gdzie $u_{0,95}$ to kwantyl rzędu $0,95$ standardowego rozkładu normalnego.

21. (a) Statystyka testowa szukanego testu to $S = \sum_{i=1}^{n} X_i$, zaś jego zbiór krytyczny to $W = (0; \gamma_{\alpha,2n,s_0}\rangle$, gdzie $\gamma_{\alpha,2n,s_0}$ oznacza kwantyl rzędu α rozkładu $Gamma(2n, s_0)$.

WSKAZÓWKA: Przydaje się tu fakt (podawany na wykładzie) dotyczący dodawania niezależnych zmiennych losowych o rozkładach gamma.

(b) Dokładnie ten sam test co w pkt. (a) (ale trzeba wyjaśnić dlaczego).

25. H_0: badana kostka do gry jest symetryczna

H_1: badana kostka do gry nie jest symetryczna

Użyjemy testu zgodności χ^2-Pearsona, statystyka testowa wynosi $\chi^2 = 10$.