

# Egzamin: Wstęp do Statystycznej Analizy Danych

UMK, styczeń 2020.

Wypełnij miejsca wykropkowane [ "....." ]. 1 pytanie = 1 punkt; Oceniane będą wyłącznie odpowiedzi wpisane w wykropkowane miejsca na tej kartce, bez uzasadnień i rachunków. Wpisz odpowiedzi dopiero po ostatecznym ich sprawdzeniu; odpowiedzi pokreślone lub nieczytelne będą traktowane jako błędne! Możesz korzystać z kalkulatora, tablic statystycznych, notatek, książek. Nie wolno korzystać z komputerów, telefonów ani z tabletów. Nie wolno się porozumiewać.

Imię i NAZWISKO: .....

1. Rzucamy 3 razy kostką. Zmienna losowa  $X$  oznacza liczbę rzutów, w których otrzymaliśmy „szóstkę”.

(a) Podać rozkład prawdopodobieństwa zmiennej losowej  $X$  w postaci tabelki.

wartość $x$	.....	.....	.....	.....
prawdopodobieństwo $P(X = x)$	.....	.....	.....	.....

(b) Oblicz  $P(X \leq 1)$ . Odpowiedź: .....

(c) Oblicz wartość oczekiwaną i wariancję zmiennej losowej  $X$ . Odpowiedź: .....

(d) Oblicz wariancję zmiennej losowej  $X$ . Odpowiedź: .....

2. Zmienna losowa  $X$  ma rozkład prawdopodobieństwa o gęstości danej wzorem

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{dla } 0 \leq x \leq 2; \\ 0 & \text{w pozostałych przypadkach.} \end{cases}$$

(a) Oblicz  $E(X)$ . Odpowiedź: .....

(b) Oblicz  $\text{Var}(X)$ . Odpowiedź: .....

(c) Oblicz  $P(X \leq 1)$ . Odpowiedź: .....

(d) Podaj wzór na dystrybuantę:

$$F(x) = P(X \leq x) = \begin{cases} \dots\dots\dots & \text{dla } x < 0; \\ \dots\dots\dots & \text{dla } 0 \leq x \leq 2; \\ \dots\dots\dots & \text{dla } x > 2. \end{cases}$$

3. Zakładamy, że  $X_1, \dots, X_{100}$  jest próbą z rozkładu normalnego  $N(\mu, \sigma^2)$  z nieznaną wartością oczekiwaną  $\mu$  i wariancją  $\sigma^2$ . Obliczono średnią z próbki i nieobciążony szacownik wariancji:  $\bar{X} = 51.75$ ,  $S^2 = 20^2$ .

- Przeprowadź test hipotezy  $H_0: \mu = 50$  przeciw alternatywie  $H_1: \mu > 50$  na poziomie istotności  $\alpha = 0.05$ .
- (a) Oblicz wartość statystyki  $T$  (t-Studenta):  $T = \dots$
- (b) Oblicz p-wartość testu:  $P = \dots$
- (c) Podejmij decyzję: odrzucamy  $H_0$ ? (napisz TAK lub NIE)  $\dots$
- Przeprowadź test hipotezy  $H_0: \mu = 50$  przeciw alternatywie  $H_1: \mu \neq 50$  na poziomie istotności  $\alpha = 0.05$ .
- (d) Oblicz p-wartość testu:  $P = \dots$
- (e) Podejmij decyzję: odrzucamy  $H_0$ ? (napisz TAK lub NIE)  $\dots$
- (f) Oblicz przedział ufności dla  $\mu$  na poziomie  $1 - \alpha = 0.95$ .

Odpowiedź:  $\dots$

Uwaga: Rozkład t-Studenta z 399 stopniami swobody jest przybliżeniu równy standardowemu rozkładowi normalnemu  $N(0, 1)$ . W szczególności można wykorzystać następujące przybliżone wartości kwantyli:  $t_{0.95}(399) \approx z_{0.95} = 1.65$  i  $t_{0.975}(399) \approx z_{0.975} = 1.96 \approx 2$ . Do obliczania p-wartości można użyć tablic rozkładu normalnego.

4. Obecnie na rynku są 3 marki smartfonów: A, B i C. W celu zbadania, czy są one wybierane przez mężczyzn i przez kobiety z takim samym prawdopodobieństwem, zanotowano jaką markę wybrało 60 klientów i 60 klientek. Wyniki przedstawia następująca tabela:

marka	A	B	C
liczba klientów (M)	15	20	25
liczba klientek (K)	25	20	15

Czy mamy podstawę do odrzucenia hipotezy  $H_0$  mówiącej, że wybór marki jest niezależny od płci? Przeprowadź odpowiedni test niezależności  $\chi^2$  na poziomie istotności  $\alpha = 0.05$ .

- (a) Oblicz tabelkę wartości oczekiwanych:

marka	A	B	C
liczba klientów (M)	.....	.....	.....
liczba klientek (K)	.....	.....	.....

- (b) Oblicz statystykę testową  $\chi^2 = \dots$
- (c) Porównujemy tę statystykę z kwantylem rzędu  $\dots$  rozkładu  $\chi^2$  z  $\dots$  stopniami swobody.
- (d) Podaj p-wartość testu:  $P = \dots$
- (e) Podejmij decyzję: czy odrzucamy  $H_0$  na poziomie istotności  $\alpha = 0.05$  (tak/nie)?  $\dots$

Wskazówka: Wiadomo, że jeśli  $H_0$  jest prawdziwa to rozkład statystyki testowej  $Y = \chi^2$  ma w przybliżeniu rozkład o dystrybucancie dana wzorem  $F(y) = P(Y \leq y) = 1 - e^{-y/2}$  dla  $y > 0$ .

Odpowiedź na pytanie (d) będzie specjalnie premiowana.

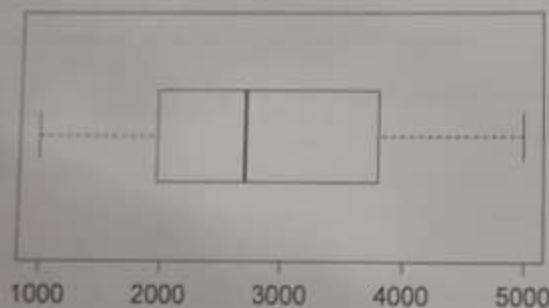
5. Wyniki cen 9 mieszkań sprzedanych przez pewnego pośrednika były następujące:

280, 150, 250, 220, 200, 300, 180, 110, 220

Na podstawie tych danych należy obliczyć następujące wielkości.

- (a) Oblicz wartość średnią ceny mieszkania. Odpowiedź: \_\_\_\_\_
- (b) Oblicz medianę ceny mieszkania. Odpowiedź: \_\_\_\_\_
- (c) Oblicz wariancję ceny mieszkania. Odpowiedź: \_\_\_\_\_
- (d) Oblicz odchylenie standardowe ceny mieszkania. Odpowiedź: \_\_\_\_\_

6. Na podstawie pomiarów wagi (w gramach) 120 noworodków, sporządzono następujący wykres pudełkowy (Box-and-Whiskers). Odczytaj z rysunku przybliżone wartości pewnych charakterystyk



liczbowych.

- (a) Podaj medianę wagi noworodka. Odpowiedź: \_\_\_\_\_
- (b) Podaj kwartyły  $Q_1$  i  $Q_3$  (kwartyły reguła  $1/4$  i  $3/4$ ). Odpowiedź: \_\_\_\_\_
- (c) Podaj rozstęp międzykwartyłowy. Odpowiedź: \_\_\_\_\_
- (d) Ile w przybliżeniu było noworodków o wadze większej, niż 3700 g? Odpowiedź: \_\_\_\_\_

**Wykazuje:** Dokładne wartości są niemożliwe do obliczenia bez znajomości danych. Niemniej, na podstawie rysunku można podać bardzo rozsądne przybliżenia.

7. W losowo wybranej próbie  $n = 400$  studentów, znalazło się  $x = 80$  palących. Interesuje nas frakcja  $p$  palących w populacji studentów.

(a) Zbuduj przedział ufności dla  $p$  na poziomie ufności 0,95.

Odpowiedź: .....

(b) Zbuduj przedział ufności dla  $p$  na poziomie ufności 0,99.

Odpowiedź: .....

Uwaga: Możesz przyjąć przybliżoną wartość kwantyla rozkładu normalnego:  $z_{0,975} \approx 2$ .

8. Windą jedzie 4 pasażerów. Każdy z nich, niezależnie od pozostałych, wysiada na jednym z 10 pięter (z jednakowym prawdopodobieństwem  $1/10$ ). Oblicz

(a) Prawdopodobieństwo, że każdy wysiadzie na innym piętrze. Odpowiedź: .....

(b) Prawdopodobieństwo, że wszyscy wysiądą na tym samym piętrze. Odpowiedź: .....

(c) Prawdopodobieństwo, że wszyscy wysiądą na pierwszym piętrze. Odpowiedź: .....

9. Zmienna losowa  $X$  ma rozkład normalny  $N(20, 5^2)$ . Oblicz

(a)  $P(X \leq 25)$ . Odpowiedź: .....

(b)  $P(X > 25)$ . Odpowiedź: .....

(c)  $P(15 \leq X \leq 25)$ . Odpowiedź: .....

(d) Podaj kwantyl rzędu 0.75 rozkładu zmiennej losowej  $X$ , tzn taką liczbę  $\xi$ , że  $P(X \leq \xi) = 0.75$ .

$\xi =$  .....

10. Wiadomo, że zmienne losowe  $X_1, \dots, X_{20}$  są niezależne, mają ten sam rozkład prawdopodobieństwa,  $EX_i = 7$  i  $\text{Var}X_i = 4$ . Niech  $\bar{X} = \frac{1}{20} \sum_{i=1}^{20} X_i$ . Oblicz

(a) Wartość oczekiwaną  $E(\bar{X}) =$  .....

(b) Wariancję  $\text{Var}(\bar{X}) =$  .....

(c) Odchylenie standardowe pojedynczej zmiennej:  $D(X_i) =$  .....

(d) Odchylenie standardowe średniej:  $D(\bar{X}) =$  .....