

Egzamin: Wstęp do Statystycznej Analizy Danych

UMK, luty 2021, I termin. Zestaw A

1. Wiadomo, że $\mathbb{P}(A \cup B) = 0.8$, $\mathbb{P}(A \cap B) = 0.1$ i $\mathbb{P}(A \setminus B) = 0.4$ (prawdopodobieństwo różnicy zdarzeń). Oblicz

- (a) $\mathbb{P}(A)$.
- (b) $\mathbb{P}(B)$.
- (c) $\mathbb{P}(B \setminus A)$.
- (d) $\mathbb{P}(A|B)$.
- (e) Czy A i B są niezależne? (TAK/NIE)

2. Zmienna losowa X ma rozkład prawdopodobieństwa dany tabelką:

Wartość x	1	2	3
Prawdopodobieństwo $\mathbb{P}(X = x)$	0.1	0.8	0.1

- (a) Oblicz $\mathbb{P}(X \leq 2)$.
 - (b) Oblicz $\mathbb{E}(X)$
 - (c) Oblicz $\text{Var}(X)$.
 - (d) Oblicz $\mathbb{E}(X^2)$
 - (e) Oblicz $\mathbb{P}(X = 3|X > 1)$ (prawdopodobieństwo warunkowe).
3. Student wybiera losowo jedno zadanie z zestawu 20 zadań. Zestaw zawiera 12 zadań z algebry (A) i 8 zadań ze statystyki (S). Wiadomo, że student rozwiąże poprawnie zadanie z algebry z prawdopodobieństwem $5/6$, zaś zadanie ze statystyki z prawdopodobieństwem $1/2$.
- (a) Oblicz prawdopodobieństwo tego, że student rozwiąże zadanie (R).
 - (b) Oblicz prawdopodobieństwo tego, że student wylosował zadanie z algebry, jeśli wiadomo, że zadanie rozwiązał.
 - (c) Oblicz prawdopodobieństwo tego, że student wylosował zadanie z algebry, jeśli wiadomo, że zadania nie rozwiązał.
 - (d) Rozważmy zdarzenia „student wylosował zadanie z algebry” (A) i „student zadanie rozwiązał” (R). Czy te 2 zdarzenia są niezależne? (TAK/NIE)
 - (e) Rozważmy zdarzenia „student wylosował zadanie z algebry” (A) i „student wylosował zadanie ze statystyki” (S). Czy te 2 zdarzenia są niezależne? (TAK/NIE)
4. Zmienna losowa X ma rozkład prawdopodobieństwa o gęstości danej wzorem

$$f(x) = \begin{cases} \frac{1}{9}x^2 & \text{dla } 0 \leq x \leq 3; \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

- (a) Oblicz $\mathbb{E}(X)$.
- (b) Oblicz $\text{Var}(X)$.
- (c) Oblicz $\mathbb{P}(X \leq 0.5)$.
- (d) Podaj medianę rozkładu zmiennej losowej X .
- (e) Podaj medianę rozkładu zmiennej losowej $1/X$.

5. Wykonujemy 100 rzutów symetryczną monetą.

- (a) Oblicz prawdopodobieństwo otrzymania dokładnie 50 orłów. Możesz napisać wzór albo możesz podać wartość liczbową.
- (b) Oblicz prawdopodobieństwo otrzymania więcej niż 60 orłów. Możesz użyć przybliżenia rozkładem normalnym, albo możesz obliczyć dokładnie, używając R.
- (c) Podaj liczbę k taką, że prawdopodobieństwo otrzymania więcej niż k orłów jest równe 0.95. Możesz użyć przybliżenia rozkładem normalnym albo możesz obliczyć dokładnie, używając R.
- (d) Oblicz wartość oczekiwaną (średnią) liczby orłów.
- (e) Oblicz wariancję liczby orłów.

6. Zanotowano czasy życia lodówek pewnego producenta (ile miesięcy lodówka pracowała):

20, 15, 16, 34, 41, 13, 34, 15, 27, 52, 17, 19, 15, 27, 19, 26, 17, 31, 18, 25, 11, 30, 31, 23, 14, 27, 20, 21, 29, 24, 41, 14, 22, 28, 39, 48, 25, 20, 18, 20, 13, 27, 21, 20, 6, 21, 18, 11, 25, 26, 15

Na podstawie tych danych należy obliczyć następujące wielkości.

- (a) Oblicz wartość średnią czasu życia.
- (b) Oblicz medianę czasu życia.
- (c) Oblicz wariancję czasu życia.
- (d) Oblicz odchylenie standardowe czasu życia.
- (e) Oblicz rozstęp międzykwartyłowy.

7. Wysokości cen 15 mieszkań sprzedanych w pewnym mieście były następujące:

150, 245, 225, 195, 300, 170, 120, 350, 200, 190, 235, 220, 200, 230, 200

Traktujemy te dane jako losową próbkę z populacji sprzedawanych w tym mieście mieszkań i zakładamy rozkład normalny w tej populacji.

- (a) Oszacuj średnią cenę mieszkania (oblicz średnią z próbki).
- (b) Oblicz nieobciążony estymator wariancji.
- (c) Podaj przedział ufności dla średniej ceny mieszkania (średniej populacyjnej) . Przyjmij poziom ufności $1 - \alpha = 0.99$.
- (d) Przeprowadź test hipotezy zerowej, mówiącej, że średnia cena jest równa 200, przeciwko alternatywie, że jest większa. Podaj p-wartość.
- (e) Przeprowadź test hipotezy zerowej, mówiącej, że średnia cena jest równa 200, przeciwko alternatywie, że jest różna od 200. Podaj p-wartość.

8. Chcemy zbadać, czy istnieje zależność między zdrowotnością psa a posiadaniem przez niego rodowodu. Dane są przytoczone w tabelce:

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

Przeprowadź test hipotezy H_0 stwierdzającej niezależność zdrowotności od rodowodu.

- (a) Podaj p-wartość odpowiedniego testu.
- (b) Podaj wartość statystyki testowej.
- (c) Czy test na poziomie istotności 0.05 prowadzi do odrzucenia H_0 ? (TAK/NIE)
- (d) Podaj tabelkę wartości oczekiwanych.