LISTA ZADAŃ NR 10: Zaawansowane metody statystyczne

Zadanie 1

Zmierzono czasy wykonania pewnego zadania. Uporządkowano wyniki w kolejności ich otrzymywania (w czasie). Następnie dla każdego wyniku określono, czy jest on powyżej (a) czy poniżej (b) mediany.

Otrzymano ciąg symboli: $a, a, b, b, a, a, a, b, b, b, b, a, a, b, ...$

Zweryfikować hipotezę, że próba jest losowa (tzn. wyniki nie zależą od czasu/kolejności), stosując **test serii**.

Zadanie 2

Mamy dwa algorytmy sortowania (A i B). Wykonano po 5 niezależnych pomiarów czasu dla każdego z nich. Wyniki nie mają rozkładu normalnego (występują elementy odstające).

- Algorytm A: $12, 18, 14, 15, 13$
- Algorytm B: $19, 21, 23, 20, 22$

Zweryfikować hipotezę, że algorytm A jest szybszy od B, stosując test sumy rang (Manna-Whitneya-Wilcoxona).

Zadanie 3

W tabeli (tablicy kontyngencji) zebrano dane o awariach w zależności od producenta sprzętu:

Producent Typ awarii	Przegrzanie	Błąd dysku	Błąd pamięci
Producent X	20	10	15
Producent Y	30	50	25

Sprawdzić na poziomie istotności $\alpha = 0.05$, czy rodzaj awarii zależy od producenta.

Zadanie 4

Testujemy wydajność 3 różnych frameworków (X, Y, Z). Ponieważ dane są mocno asymetryczne, zamiast klasycznej analizy wariancji (ANOVA), stosujemy test nieparametryczny Kruskala-Wallisa.

Dla danych rankingowych z tabeli zweryfikować hipotezę, że wszystkie frameworki mają taką samą medianę wydajności.

Zadanie 5

Mamy dwa zbiory danych o ruchu sieciowym (przed i po wdrożeniu firewalla). Chcemy sprawdzić, czy **cały rozkład** (nie tylko średnia) uległ zmianie.

Na podstawie dystrybuant empirycznych obu prób obliczyć statystykę $D_{n,m}$ i zweryfikować hipotezę o identyczności rozkładów (test Kołmogorowa-Smirnowa).

Zadanie 6

Badamy czas kompilacji kodu (Y) w zależności od liczby plików (X_1) i liczby linii kodu w pliku (X_2).

Dla podanych danych wyznaczyć równanie płaszczyzny regresji:

$$y = ax_1 + bx_2 + c$$

Zadanie 7

Liczba tranzystorów w procesorach rośnie wykładniczo: $y = a \cdot e^{bx}$. Mając dane historyczne, sprowadzić to zagadnienie do regresji liniowej poprzez logarytmowanie ($\ln y = \ln a + bx$) i wyznaczyć parametry wzrostu.

Zadanie 8

Produkcja procesorów generuje pewien procent braków. Zamiast pobierać stałą próbkę 100 sztuk, pobieramy sztuki jedna po drugiej. Po każdym pobraniu decydujemy: „partia dobra", „partia zła" lub „pobieramy dalej".

Skonstruować test sekwencyjny (test Walda) dla weryfikacji hipotezy $p = 0.01$ przeciw $p = 0.10$.

Zadanie 9

Dla próby prostej $x_1, ..., x_n$ z rozkładu wykładniczego (czas bezawaryjnej pracy) o gęstości $f(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right)$, gdzie λ jest wartością oczekiwaną, wyznaczyć estymator parametru λ metodą największej wiarygodności (MNW).

Zadanie 10

Mamy 3 serwery. Chcemy sprawdzić, czy działają tak samo stabilnie (czy mają taką samą wariancję czasów odpowiedzi), zanim porównamy ich średnie czasy. Wariancje z prób wynoszą: $s_1^2 = 1.4,\ s_2^2 = 1.8,\ s_3^2 = 1.2.$

Zweryfikować hipotezę $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ (np. testem Bartletta).