

Podstawy statystyki praktycznej
Lista nr 7

1. Rozważmy problem testowania dla dwóch prób: $X_1, \dots, X_n \sim N(\mu_1, \sigma^2)$, $Y_1, \dots, Y_n \sim N(\mu_2, \sigma^2)$, σ znane, $H_0 : \mu_1 = \mu_2$ vs $H_A : \mu_1 \neq \mu_2$, za pomocą statystyki testowej $Z = \frac{|\bar{X} - \bar{Y}|}{\sigma} \sqrt{\frac{n}{2}}$, gdzie $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.
 - a) Wyznacz wartości krytyczne dla Z na poziomach istotności $\alpha = 0.1$ i $\alpha = 0.05$.
 - b) Zaprojektuj badania symulacyjne dla estymacji prawdopodobieństwa popełnienia błędu pierwszego rodzaju α .
 - i) Dla $\alpha = 0.05$ i 0.1 wyznacz liczbę powtórzeń eksperymentu, tak żeby z prawdopodobieństwem 0.95 błąd estymacji był nie większy niż $10\% \alpha$.
 - ii) Przeprowadź odpowiednie symulacje i zweryfikuj swoje obliczenia z punktu i).
 - e) Powtórz eksperyment w sytuacji gdy obie próby są pobrane z populacji o rozkładzie wykładniczym z parametrem $\lambda = 1$. Eksperyment przeprowadź dla $n=50$ i $n=100$. Podsumuj wyniki.
2. Rozważmy problem testowania dla dwóch prób: $X_1, \dots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$, $Y_1, \dots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$, gdzie σ_1 i σ_2 są nieznane, $H_0 : \mu_1 = \mu_2$ vs $H_A : \mu_1 \neq \mu_2$.

W tym zadaniu będziemy testować H_0 za pomocą statystyki ogólnego testu Studenta:

$$T = \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

gdzie $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$, $\bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$, $s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$, $s_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$. W oparciu o symulacje komputerowe skonstruuj 95% przedział ufności dla p-stwa błędu pierwszego rodzaju jeżeli wartość krytyczną weźmiemy z rozkładu

- i) Studenta z liczbą stopni swobody wyznaczoną w oparciu o poprawkę Welcha-Satterthwaite'a
- ii) Studenta z liczbą stopni swobody równą $n_1 + n_2 - 2$
- iii) normalnego $z_{kr} = \Phi^{-1}(1 - \alpha/2)$

Obliczenia wykonaj dla $\sigma_1 = 1$, $\sigma_2 = 5$, $\alpha \in \{0.05, 0.1\}$ i

- a) $n_1 = 5$, $n_2 = 10$
- b) $n_1 = 10$, $n_2 = 20$
- c) $n_1 = 20$, $n_2 = 40$
- d) $n_1 = 40$, $n_2 = 80$

3. Zbiór danych dane5.txt zawiera dane dotyczące poziomu cholesterolu u pacjentów kilka dni po zawale.
 - a) Narysuj histogramy i wykresy kwantylowe poziomu cholesterolu u osób z grupy kontrolnej oraz u pacjentów dwa, cztery i czternaście dni po zawale.
 - b) Narysuj wykresy ramkowe (pudełkowe) porównujące rozkłady poziomu cholesterolu u osób z grupy kontrolnej i pacjentów dwa dni po zawale.
 - c) Porównaj poziom cholesterolu u osób z grupy kontrolnej i pacjentów dwa dni po zawale za pomocą testu Studenta dla dwóch prób.
4. W tym zadaniu ponownie zajmiemy się danymi ze zbioru individuals.dat.
 - a) Narysuj wykresy kwantylowe zarobków osobno dla kobiet i mężczyzn. Podsumuj wyniki.
 - b) Utwórz nową zmienną $\text{transincome} = \text{income}^{0.25}$ i narysuj dla niej odpowiednie wykresy kwantylowe. Podsumuj wyniki.
 - c) Zastosuj test Studenta dla dwóch prób do sprawdzenia czy średnia wielkość zmiennej transincome jest istotnie większa u mężczyzn niż u kobiet. Sformułuj wniosek z testowania.
 - d) Utwórz jeden wykres porównujący wykresy ramkowe zmiennej transincome dla różnych poziomów wykształcenia. Krótko podsumuj uzyskane wyniki.
 - e) Zastosuj test Studenta do sprawdzenia czy średnia wielkość zmiennej transincome u osób z wykształceniem wyższym jest istotnie większa niż u osób z licencjatem. Sformułuj wniosek z testowania.
 - f) Utwórz jeden wykres porównujący wykresy ramkowe zmiennej transincome dla trzech sektorów zatrudnienia.
 - g) Zastosuj test Studenta do porównania wartości oczekiwanych zmiennej transincome dla osób pracujących w sektorze rządowym i sektorze prywatnym. Sformułuj wniosek z testowania.

Zadania do policzenia ręcznego

5. 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 $\bar{x} = 1.15$, $S_1^2 = 2.4$ oraz $\bar{y} = 1.05$, $S_2^2 = 2.3$. Czy można twierdzić, że średnie w tych populacjach są takie same? Przyjąć poziom istotności $\alpha = 0.05$.
6. Badając wymagania żywieniowe krów badacze zmierzyli przyrost wagi krów podczas 78 dniowego okresu testowego. W poniższej tabeli podsumowano wyniki dla dwóch ras krów, Hereford (HH) i Brown Swiss/Hereford (SH).

	HH	SH
N	33	51
$\bar{y}$	18.3	13.9
s	17.8	19.1

Na poziomie istotności $\alpha = 0.1$ porównaj średni przyrost wagi u obu ras krów.

Małgorzata Bogdan