Politechnika Wrocławska

Lab 4

Statystyka w zastosowaniach

23.05.2025

Mateusz Nasewicz

Spis treści

1 Zadanie 1: Test t-Studenta dla dwóch niezależnych prób

Celem analizy było sprawdzenie, czy średni wzrost kobiet i mężczyzn różni się istotnie statystycznie. W tym celu przeprowadzono test t-Studenta dla dwóch niezależnych prób. Dane zostały wygenerowane losowo na podstawie rozkładu normalnego:

- Wzrost kobiet: średnia = 165 cm, odchylenie standardowe = 7 cm, liczebność = 30,
- Wzrost mężczyzn: średnia = 175 cm, odchylenie standardowe = 7 cm, liczebność = 30.

Sformułowano następujące hipotezy:

- H_0: Średni wzrost kobiet i mężczyzn jest taki sam.
- H_1: Średni wzrost kobiet i mężczyzn różni się.

Wynik testu t-Studenta:

- Statystyka $t = -6.3256$
- Wartość $p = 0.0000$

Ponieważ wartość p jest mniejsza niż przyjęty poziom istotności $\alpha = 0{,}05$, istnieją podstawy do odrzucenia hipotezy zerowej. Oznacza to, że średni wzrost kobiet i mężczyzn różni się istotnie statystycznie.

2 Zadanie 2: Test t dla jednej próby

Celem analizy było sprawdzenie, czy średnia wzrostu w losowo wygenerowanej próbce różni się istotnie od wartości referencyjnej 170 cm. W tym celu przeprowadzono test t-Studenta dla jednej próby. Założono poziom istotności $\alpha = 0{,}05$.

Sformułowano następujące hipotezy:

- H_0: Średni wzrost w populacji wynosi 170 cm,
- H_1: Średni wzrost w populacji różni się od 170 cm.

Przeanalizowano trzy przypadki z różnymi wartościami średnich w generowanych próbkach:
Przypadek 1: Średnia próbki = 170

- Średnia próbki: 172,25 cm
- Statystyka $t = 2{,}0453$
- Wartość $p = 0{,}0500$
- Wniosek: Odrzucamy hipotezę zerową — średnia różni się istotnie od 170 cm.

Przypadek 2: Średnia próbki = 172

- Średnia próbki: 172,44 cm
- Statystyka $t = 1{,}8444$
- Wartość $p = 0{,}0754$
- Wniosek: Brak podstaw do odrzucenia hipotezy zerowej — średnia nie różni się istotnie od 170 cm.

Przypadek 3: Średnia próbki = 175

- Średnia próbki: 176,16 cm
- Statystyka $t = 5{,}0832$
- Wartość $p = 0{,}0000$
- Wniosek: Odrzucamy hipotezę zerową — średnia różni się istotnie od 170 cm.

Analiza pokazuje, że im większa różnica między średnią próbki a wartością testowaną (170 cm), tym mniejsza wartość p i większa siła dowodów przeciwko hipotezie zerowej.

3 Zadanie 3: Test chi-kwadrat dla danych jakościowych

Celem analizy było sprawdzenie, czy istnieje zależność między płcią a preferowanym gatunkiem muzyki. W tym celu przeprowadzono test chi-kwadrat niezależności dla danych jakościowych. Dane zebrano w postaci tabeli kontyngencji 2×3:

Płeć	**Pop**	**Rock**	**Klasyczna**
Kobieta	20	15	5
Mężczyzna	10	25	5

Sformułowano hipotezy:

- H_0: Płeć i preferowany gatunek muzyki są niezależne,
- H_1: Płeć i preferowany gatunek muzyki są zależne.

Wyniki testu chi-kwadrat:

- Statystyka $\chi^2 = 5{,}8333$
- Wartość $p = 0{,}0541$
- Liczba stopni swobody: 2

Oczekiwane liczności w przypadku braku zależności:

Płeć	**Pop**	**Rock**	**Klasyczna**
Kobieta	15.0	20.0	5.0
Mężczyzna	15.0	20.0	5.0

Ponieważ wartość $p = 0{,}0541$ jest większa od przyjętego poziomu istotności $\alpha = 0{,}05$, nie ma podstaw do odrzucenia hipotezy zerowej. Oznacza to, że brak jest istotnej statystycznie zależności między płcią a preferowanym gatunkiem muzyki.

4 Zadanie 4: Porównanie testu parametrycznego i nieparametrycznego

Celem analizy było porównanie wyników testu parametrycznego (t-Studenta) oraz nieparametrycznego (Manna-Whitneya) w sytuacji, gdy dane pochodzą z rozkładu niespełniającego założenia normalności — rozkładu wykładniczego.

Wygenerowano dwie grupy po 30 obserwacji:

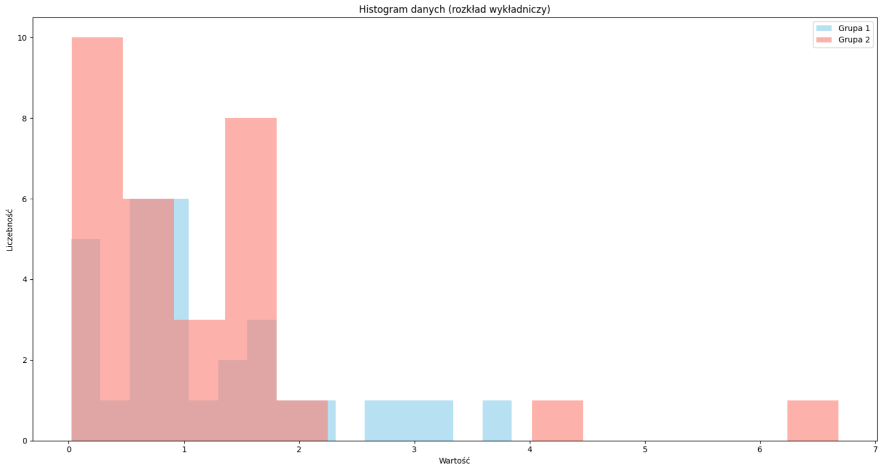

Fig... 1: Histogram danych z rozkładu wykładniczego dla dwóch grup.

- Grupa 1: dane z rozkładu wykładniczego o skali 1.0
- Grupa 2: dane z rozkładu wykładniczego o skali 1.5

Histogram rozkładu danych dla obu grup przedstawiono na Rysunku 1.
Przeprowadzono dwa testy statystyczne:

- **Test t-Studenta (parametryczny)**: zakłada normalność rozkładu,
- **Test Manna-Whitneya (nieparametryczny)**: nie wymaga założenia normalności.

Wyniki testów:

- Test t-Studenta: $t = -0{,}0514$, $p = 0{,}9592$
- Test Manna-Whitneya: $U = 471{,}0000$, $p = 0{,}7618$

W obu przypadkach wartość p jest znacznie większa niż przyjęty poziom istotności $\alpha = 0{,}05$, co oznacza brak podstaw do odrzucenia hipotezy zerowej — nie stwierdzono istotnych różnic pomiędzy grupami.

Wniosek: Mimo że dane nie mają rozkładu normalnego, oba testy dały zgodne wyniki. Jednakże w praktyce, przy nienormalnym rozkładzie danych (co można zauważyć na histogramie), zaleca się stosowanie testów nieparametrycznych (np. testu Manna-Whitneya), które są bardziej odporne na naruszenie założeń dotyczących rozkładu danych.

5 Zadanie 5: Wpływ liczności próby na wartość p

Celem tego zadania było zbadanie wpływu liczności próby na wartość p w teście t-Studenta dla dwóch grup o bardzo zbliżonych średnich. W tym celu wygenerowano dwie grupy danych: pierwsza z rozkładem normalnym o średniej 0 i odchyleniu standardowym 1, druga o średniej 0,1 i takim samym odchyleniu standardowym. Dla rosnącej liczności próby od 10 do 1000 (krok co 10) przeprowadzono test t-Studenta i zanotowano odpowiadające wartości p.

Wyniki przedstawiono na Rysunku 2, który pokazuje zależność wartości p od liczności próby.

Z wykresu wynika, że dla małych prób (np. $n < 200$) wartość p często przekracza poziom istotności 0,05, co sugeruje brak istotnej statystycznie różnicy między grupami. Jednak wraz ze wzrostem liczności

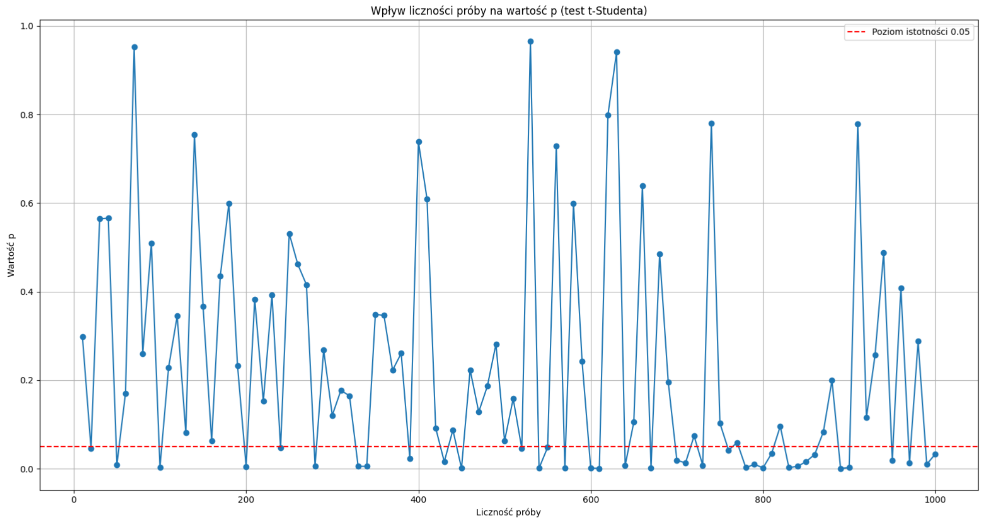

Fig... 2: Wpływ liczności próby na wartość p (test t-Studenta). Czerwona linia przerywana oznacza poziom istotności 0,05.

próby wartości p maleją i coraz częściej spadają poniżej poziomu 0,05, mimo że różnica między średnimi nadal wynosi jedynie 0,1.

Zjawisko to wynika z faktu, że wraz ze wzrostem liczby obserwacji rośnie moc testu statystycznego. Oznacza to, że test staje się bardziej czuły na nawet bardzo małe różnice między grupami, które przy małej próbie mogłyby zostać uznane za nieistotne statystycznie. Dlatego w analizie danych zawsze należy interpretować wartość p w kontekście wielkości próby oraz efektu praktycznego (np. poprzez analizę wielkości efektu).