

Metody nieparametryczne

Wojciech Kotłowski

Statystyka i analiza danych 2019/2020

26.05.2020

Metody nieparametryczne

- Dotychczasowe testy działają poprawnie tylko przy pewnych **założeniach o populacji**.
 - Przykład: test T działa poprawnie, jeśli X ma rozkład normalny, itp.
- Złamanie tych założeń powoduje, że prawdopodobieństwo błędu I rodzaju może nie być na poziomie istotności α .
- **Metody nieparametryczne** czynią znacznie mniej założeń o populacji.
 - Ogólniejsze – stosowalne w szerszym zakresie.
 - Słabsze – moc testu jest niższa.

Test znaków

- Brak założeń o rozkładach X i Y .
- Odpowiednik testu **sparowanego** T dla dwóch populacji.

Test znaków

- Brak założeń o rozkładach X i Y .
- Odpowiednik testu **sparowanego** T dla dwóch populacji.

Zasada działania: zamiast różnicy $Y - X$, testujemy jej **znak**:

$$S = \text{sgn}(Y - X),$$

przy czym przypadki $Y = X$ pomijamy.

X	Y	S
4	6	+
7	7	pomijamy
3	2	-
1	1.5	+
2	9	+
6	2	-

Test znaków

- Brak założeń o rozkładach X i Y .
- Odpowiednik testu **sparowanego** T dla dwóch populacji.

Zasada działania: zamiast różnicy $Y - X$, testujemy jej **znak**:

$$S = \text{sgn}(Y - X),$$

przy czym przypadki $Y = X$ pomijamy.

Wniosek: Niezależnie od oryginalnego rozkładu $Y - X$, zmienna S ma rozkład **dwupunktowy** z parametrem sukcesu:

$$p = P(S = +)$$

Test znaków

- Brak założeń o rozkładach X i Y .
- Odpowiednik testu **sparowanego** T dla dwóch populacji.

Zasada działania: zamiast różnicy $Y - X$, testujemy jej **znak**:

$$S = \text{sgn}(Y - X),$$

przy czym przypadki $Y = X$ pomijamy.

Wniosek: Niezależnie od oryginalnego rozkładu $Y - X$, zmienna S ma rozkład **dwupunktowy** z parametrem sukcesu:

$$p = P(S = +) = P(Y > X | Y \neq X)$$

Zredukowaliśmy problem do testu Z dla rozkładu dwupunktowego!

Przykład

W ramach zajęć 14 studentów napisało wejściówkę i wyjściówkę. Sprawdź na poziomie $\alpha = 0.05$, czy studenci poszerzyli wiedzę w trakcie zajęć.

ID studenta	ocena z wejściówki	ocena z wyjściówki
1	4.5	3.5
2	3.5	4.5
3	2.0	4.5
4	5.0	5.0
5	4.0	5.0
6	3.5	4.5
7	3.0	3.5
8	2.0	3.5
9	2.0	2.0
10	4.5	4.0
11	3.0	5.0
12	4.0	3.5
13	5.0	4.5
14	2.0	3.5

Przykład

W ramach zajęć 14 studentów napisało wejściówkę i wyjściówkę. Sprawdź na poziomie $\alpha = 0.05$, czy studenci poszerzyli wiedzę w trakcie zajęć.

ID studenta	ocena z wejściówki	ocena z wyjściówki	S
1	4.5	3.5	—
2	3.5	4.5	+
3	2.0	4.5	+
4	5.0	5.0	pomiń
5	4.0	5.0	+
6	3.5	4.5	+
7	3.0	3.5	+
8	2.0	3.5	+
9	2.0	2.0	pomiń
10	4.5	4.0	—
11	3.0	5.0	+
12	4.0	3.5	—
13	5.0	4.5	—
14	2.0	3.5	+

Przykład

W ramach zajęć 14 studentów napisało wejściówkę i wyjściówkę. Sprawdź na poziomie $\alpha = 0.05$, czy studenci poszerzyli wiedzę w trakcie zajęć.

ID studenta	ocena z wejściówki	ocena z wyjściówki	S
1	4.5	3.5	—
2	3.5	4.5	+
3	2.0	4.5	+
4	5.0	5.0	pomiń
5	4.0	5.0	+
6	3.5	4.5	+
7	3.0	3.5	+
8	2.0	3.5	+
9	2.0	2.0	pomiń
10	4.5	4.0	—
11	3.0	5.0	+
12	4.0	3.5	—
13	5.0	4.5	—
14	2.0	3.5	+

Liczba **par znaczących** $n = 12$.

Liczba **sukcesów** $S_n = 8$.

Test znaków – przykład

X – ocena z wejściówki, Y – ocena z wyjściówki

- **Układ hipotez:**

$$H_0 : p = P(Y > X | Y \neq X) = \frac{1}{2}$$

$$H_1 : p > \frac{1}{2}$$

- **Dane:**

- Liczba par znaczących $n = 12$
- Liczba sukcesów $S_n = 8$

- **Statystyka testowa.** Ponieważ $np_0 = n(1 - p_0) = 12 \cdot \frac{1}{2} > 5$, korzystamy z Centralnego Twierdzenia Granicznego:

$$Z = \frac{S_n - np_0}{\sqrt{np_0(1 - p_0)}} \sim N(0, 1).$$

$$Z = \frac{8 - 12 \cdot 0.5}{\sqrt{12 \cdot 0.5 \cdot 0.5}} = \frac{2}{\sqrt{3}} \simeq 1.15.$$

- **Zbiór krytyczny** dla $\alpha = 0.05$: $C_{kr} = (1.64, \infty)$.

Wniosek: Nie ma podstaw do odrzucenia H_0 .



Test Wilcoxona

- Odpowiednik testu **sparowanego** T dla dwóch populacji.
- Silniejszy od testu znaków, ale słabszy od testu T .
- Zakłada jedynie, że różnice $Y - X$ można ze sobą porównywać (większa/mniejsza/równa).

Układ hipotez:

$$H_0 : \text{median}(Y - X) = 0$$

$$H_1 : \text{median}(Y - X) \neq 0 \text{ (lub odpowiednia wersja jednostronna)}$$

Wymaga skomplikowanych obliczeń na **rangach modułów różnic**, które najlepiej przeprowadzić na przykładzie.

Używamy następującej obserwacji: gdy H_0 jest prawdziwe, znaki różnic są równo prawdopodobne.

Test Wilcoxona – przykład

wejściówka	wyjściówka	różnica	moduł różnicy	ranga
4.5	3.5			
3.5	4.5			
2.0	4.5			
5.0	5.0			
4.0	5.0			
3.5	4.5			
3.0	3.5			
2.0	3.5			
2.0	2.0			
4.5	4.0			
3.0	5.0			
4.0	3.5			
5.0	4.5			
2.0	3.5			

Test Wilcoxona – przykład

wejściówka	wyjściówka	różnica	moduł różnicy	ranga
4.5	3.5	-1		
3.5	4.5	1		
2.0	4.5	2.5		
5.0	5.0	0		
4.0	5.0	1		
3.5	4.5	1		
3.0	3.5	0.5		
2.0	3.5	1.5		
2.0	2.0	0		
4.5	4.0	-0.5		
3.0	5.0	2		
4.0	3.5	-0.5		
5.0	4.5	-0.5		
2.0	3.5	1.5		

Test Wilcoxona – przykład

wejściówka	wyjściówka	różnica	moduł różnicy	ranga
4.5	3.5	-1	1	
3.5	4.5	1	1	
2.0	4.5	2.5	2.5	
5.0	5.0	0	0	
4.0	5.0	1	1	
3.5	4.5	1	1	
3.0	3.5	0.5	0.5	
2.0	3.5	1.5	1.5	
2.0	2.0	0	0	
4.5	4.0	-0.5	0.5	
3.0	5.0	2	2	
4.0	3.5	-0.5	0.5	
5.0	4.5	-0.5	0.5	
2.0	3.5	1.5	1.5	

Test Wilcoxona – przykład

wejściówka	wyjściówka	różnica	moduł różnicy	ranga
4.5	3.5	-1	1	5 – 8
3.5	4.5	1	1	5 – 8
2.0	4.5	2.5	2.5	12
5.0	5.0	0	0	–
4.0	5.0	1	1	5 – 8
3.5	4.5	1	1	5 – 8
3.0	3.5	0.5	0.5	1 – 4
2.0	3.5	1.5	1.5	9 – 10
2.0	2.0	0	0	–
4.5	4.0	-0.5	0.5	1 – 4
3.0	5.0	2	2	11
4.0	3.5	-0.5	0.5	1 – 4
5.0	4.5	-0.5	0.5	1 – 4
2.0	3.5	1.5	1.5	9 – 10

Rangujemy **niezerowe** moduły od najmniejszego (ranga 1) do największego (ranga 12).

Test Wilcoxona – przykład

wejściówka	wyjściówka	różnica	moduł różnicy	ranga
4.5	3.5	-1	1	6.5
3.5	4.5	1	1	6.5
2.0	4.5	2.5	2.5	12
5.0	5.0	0	0	—
4.0	5.0	1	1	6.5
3.5	4.5	1	1	6.5
3.0	3.5	0.5	0.5	2.5
2.0	3.5	1.5	1.5	9.5
2.0	2.0	0	0	—
4.5	4.0	-0.5	0.5	2.5
3.0	5.0	2	2	11
4.0	3.5	-0.5	0.5	2.5
5.0	4.5	-0.5	0.5	2.5
2.0	3.5	1.5	1.5	9.5

Rangujemy **niezerowe** moduły od najmniejszego (ranga 1) do największego (ranga 12).

Równe moduły dzielą się rangami (przydzielamy **średnią** rangę).

Test Wilcoxona – przykład

wejściówka	wyjściówka	różnica	moduł różnicy	ranga
4.5	3.5	-1	1	6.5
3.5	4.5	1	1	6.5
2.0	4.5	2.5	2.5	12
5.0	5.0	0	0	—
4.0	5.0	1	1	6.5
3.5	4.5	1	1	6.5
3.0	3.5	0.5	0.5	2.5
2.0	3.5	1.5	1.5	9.5
2.0	2.0	0	0	—
4.5	4.0	-0.5	0.5	2.5
3.0	5.0	2	2	11
4.0	3.5	-0.5	0.5	2.5
5.0	4.5	-0.5	0.5	2.5
2.0	3.5	1.5	1.5	9.5

Sumujemy osobno rangi dodatnich różnic i rangi ujemnych różnic:

$$\Sigma_+ = 6.5 + 12 + 6.5 + 6.5 + 2.5 + 9.5 + 11 + 9.5 = 64$$

$$\Sigma_- = 6.5 + 2.5 + 2.5 + 2.5 = 14$$

Test Wilcoxona – przykład

X – ocena z wejściówki, Y – ocena z wyjściówki

- **Układ hipotez:**

$$H_0 : \text{median}(Y - X) = 0$$

$$H_1 : \text{median}(Y - X) > 0$$

- **Dane:** $n = 12$ (liczba par znaczących), $\Sigma_+ = 64$, $\Sigma_- = 14$.
- **Statystyka testowa** $T = \min\{\Sigma_+, \Sigma_-\}$.
Otrzymujemy $T = 14$.
- Wartość krytyczną T_{kr} otrzymujemy z tablic, np. [tutaj](#).
Dla $n = 12$ i $\alpha = 0.05$, $T_{\text{kr}} = 17$.

Zbiór krytyczny: $C_{\text{kr}} = [0, T_{\text{kr}}]$ (zawsze $\leq T_{\text{kr}}$)

$T \leq T_{\text{kr}} \implies$ **Odrzucamy** H_0 , studenci się czegoś nauczyli.

Współczynnik korelacji rangowej Spearmana

- Nieparametryczny współczynnik korelacji.
- Zasada obliczania: zamień X i Y na **rangi** i policz na tym zwykły współczynnik korelacji Pearsona.

Przykład:

X	Y
0.3	2.5
1.7	1.2
2.2	1.2
-1.1	-0.9
0.2	5
0.3	-3
-0.5	-3
-4.1	-1

Współczynnik korelacji rangowej Spearmana

- Nieparametryczny współczynnik korelacji.
- Zasada obliczania: zamień X i Y na **rangi** i policz na tym zwykły współczynnik korelacji Pearsona.

Przykład:

X	rangi X	Y	rangi Y
0.3	5.5	2.5	7
1.7	7	1.2	5.5
2.2	8	1.2	5.5
-1.1	2	-0.9	4
0.2	4	5	8
0.3	5.5	-3	1.5
-0.5	3	-3	1.5
-4.1	1	-1	3

Współczynnik korelacji rangowej Spearmana

- Nieparametryczny współczynnik korelacji.
- Zasada obliczania: zamień X i Y na **rangi** i policz na tym zwykły współczynnik korelacji Pearsona.

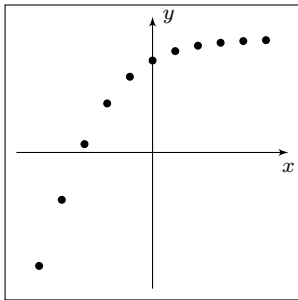
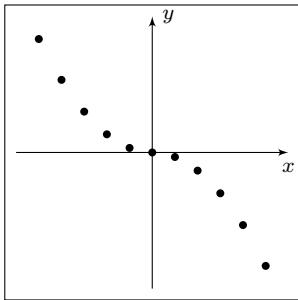
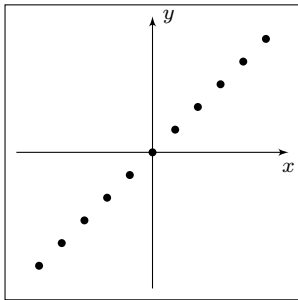
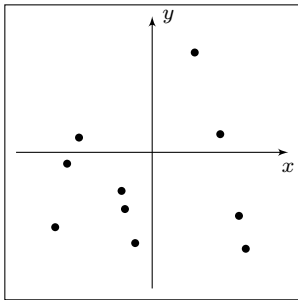
Przykład:

X	rangi X	Y	rangi Y
0.3	5.5	2.5	7
1.7	7	1.2	5.5
2.2	8	1.2	5.5
-1.1	2	-0.9	4
0.2	4	5	8
0.3	5.5	-3	1.5
-0.5	3	-3	1.5
-4.1	1	-1	3

Współczynnik korelacji
między rangami:

$$r_s = 0.358$$

Wsp. korelacji Pearsona vs. Spearmana



Wsp. korelacji Pearsona vs. Spearmana

