Obserwacje powiązane w pary

Rozważmy sytuację, w której mamy do czynienia z parami obserwacji

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n),$$

gdzie **pary są wzajemnie niezależne, ALE zmienne w parze już NIE**.

Przykład

- Populacja X opisuje umiejętności pracowników danej firmy przed szkoleniem.
- Populacja Y opisuje umiejętności pracowników danej firmy po szkoleniu.

Testy różnicy średnich dla obserwacji powiązanych w pary

Sparowany test t

Sparowany test t – założenia

Przykład

- Populacja X opisuje umiejętności pracowników danej firmy przed szkoleniem.
- Populacja Y opisuje umiejętności pracowników danej firmy po szkoleniu.

Sparowany test t – założenia

Przykład

- Populacja X opisuje umiejętności pracowników danej firmy przed szkoleniem.
- Populacja Y opisuje umiejętności pracowników danej firmy po szkoleniu.

- Załóżmy dodatkowo, że obie populacje mają **rozkłady normalne**.

Sparowany test t – założenia

Przykład

- Populacja X opisuje umiejętności pracowników danej firmy przed szkoleniem.
- Populacja Y opisuje umiejętności pracowników danej firmy po szkoleniu.

- Załóżmy dodatkowo, że obie populacje mają **rozkłady normalne**.
- Przypuśćmy, że interesuje nas **porównanie wartości średnich** m_X **i** m_Y **tych populacji**.

Sparowany test t – założenia

Przykład

- Populacja X opisuje umiejętności pracowników danej firmy przed szkoleniem.
- Populacja Y opisuje umiejętności pracowników danej firmy po szkoleniu.

- Załóżmy dodatkowo, że obie populacje mają **rozkłady normalne**.
- Przypuśćmy, że interesuje nas **porównanie wartości średnich** m_X **i** m_Y **tych populacji**.

Uwaga: Ze względu na zależność populacji X i Y ich wartości średnie również są **zależne**.

Sparowany test t – hipotezy zerowa i alternatywna

- Zdefiniujmy różnice

$$Z_i = X_i - Y_i,$$

które tworzą już próbę **niezależnych zmiennych losowych**.

Sparowany test t – hipotezy zerowa i alternatywna

- Zdefiniujmy różnice
$$Z_i = X_i - Y_i,$$
  które tworzą już próbę **niezależnych zmiennych losowych**.
- Sprawdzaną **hipotezą zerową** jest
$$H_0 : \ m_Z = 0,$$
  gdzie m_Z jest wartością średnią obliczaną z przyrostów par.

Sparowany test t – hipotezy zerowa i alternatywna

- Zdefiniujmy różnice

$$Z_i = X_i - Y_i,$$

które tworzą już próbę **niezależnych zmiennych losowych**.

- Sprawdzaną **hipotezą zerową** jest

$$H_0 : \ m_Z = 0,$$

gdzie m_Z jest wartością średnią obliczaną z przyrostów par.

- Jako **hipotezę alternatywną** możemy wybrać

$$H_1 : \ m_Z > 0 \quad \text{lub} \quad H_1 : \ m_Z < 0 \quad \text{lub} \quad H_1 : \ m_Z \neq 0.$$

Sparowany test t – statystyka testowa

- Do oceny istotności różnic pomiędzy średnimi zastosujemy **statystykę testową**:

$$t = \frac{\bar{Z} - m_Z}{\sqrt{\frac{S_Z^2}{n}}} = \frac{\bar{Z}}{S_Z}\sqrt{n},$$

Sparowany test t – statystyka testowa

- Do oceny istotności różnic pomiędzy średnimi zastosujemy **statystykę testową**:

$$t = \frac{\bar{Z} - m_Z}{\sqrt{\frac{S_Z^2}{n}}} = \frac{\bar{Z}}{S_Z}\sqrt{n},$$

gdzie

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \frac{1}{n}\sum_{i=1}^{n} Z_i, \qquad S_Z^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2\right).$$

Sparowany test t – statystyka testowa

- Do oceny istotności różnic pomiędzy średnimi zastosujemy **statystykę testową**:

$$t = \frac{\bar{Z} - m_Z}{\sqrt{\frac{S_Z^2}{n}}} = \frac{\bar{Z}}{S_Z}\sqrt{n},$$

gdzie

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \frac{1}{n}\sum_{i=1}^{n} Z_i, \qquad S_Z^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} Z_i^2 - n\bar{Z}^2\right).$$

- Statystyka t ma – przy założeniu prawdziwości H_0 – **rozkład t Studenta z $n-1$ stopniami swobody**.

Sparowany test t – przykład

Przykład

- W badaniu analizowano skuteczność nowego leku na nadciśnienie tętnicze.

Sparowany test t – przykład

Przykład

- W badaniu analizowano skuteczność nowego leku na nadciśnienie tętnicze.
- Zastosowano go na próbie 22 pacjentów, u których ciśnienie skurczowe przed rozpoczęciem terapii było bliskie wartości 144 mmHg.

Sparowany test t – przykład

Przykład

- W badaniu analizowano skuteczność nowego leku na nadciśnienie tętnicze.
- Zastosowano go na próbie 22 pacjentów, u których ciśnienie skurczowe przed rozpoczęciem terapii było bliskie wartości 144 mmHg.
  (Górna granica normy ciśnienia skurczowego wynosi 140 mmHg.)

Sparowany test t – przykład

Przykład

- W badaniu analizowano skuteczność nowego leku na nadciśnienie tętnicze.
- Zastosowano go na próbie 22 pacjentów, u których ciśnienie skurczowe przed rozpoczęciem terapii było bliskie wartości 144 mmHg.
  (Górna granica normy ciśnienia skurczowego wynosi 140 mmHg.)
- Celem badania było sprawdzenie, czy zastosowanie terapii z nowym lekiem obniża ciśnienie o 5 mmHg.

Sparowany test t – przykład

Przykład

- W badaniu analizowano skuteczność nowego leku na nadciśnienie tętnicze.
- Zastosowano go na próbie 22 pacjentów, u których ciśnienie skurczowe przed rozpoczęciem terapii było bliskie wartości 144 mmHg.
  (Górna granica normy ciśnienia skurczowego wynosi 140 mmHg.)
- Celem badania było sprawdzenie, czy zastosowanie terapii z nowym lekiem obniża ciśnienie o 5 mmHg.
- Każdemu pacjentowi zmierzono ciśnienie skurczowe przed rozpoczęciem terapii oraz po jej zakończeniu, tworząc pary wyników (X_i, Y_i) dla $i \in \{1, \ldots, 22\}$.

Sparowany test t – przykład

- X_i – opisuje ciśnienie i-tego pacjenta przed terapią nowym lekiem,
  Y_i – opisuje ciśnienie i-tego pacjenta po terapii nowym lekiem,
- $Z_i = X_i - Y_i$ dla $i \in \{1, \ldots, 22\}$,
- **Hipoteza zerowa**:

  $H_0 : \; m_Z = 5$ (*średnie obniżenie ciśnienia wynosi dokładnie* 5 *mmHg*).
- **Hipoteza alternatywna**:

  $H_1 : \; m_Z \neq 5$ (*średnie obniżenie ciśnienia jest inne niż założona wartość* 5 *mmHg*).
- Do testowania tej hipotezy używamy **statystyki testowej**

$$t = \frac{\bar{Z} - 5}{S_Z} \sqrt{n}.$$

Sparowany test t – przykład

- Przyjmijmy, że dla próby $n = 22$ pacjentów otrzymano następujące wartości:
  - $\bar{Z} = 4,691$ (średnia różnica ciśnień przed i po terapii nowym lekiem),
  - $S_Z = 0,404$ (odchylenie standardowe różnic).

Sparowany test t – przykład

- Przyjmijmy, że dla próby $n = 22$ pacjentów otrzymano następujące wartości:
  - $\bar{Z} = 4,691$ (średnia różnica ciśnień przed i po terapii nowym lekiem),
  - $S_Z = 0,404$ (odchylenie standardowe różnic).
- Podstawiając te wartości do wzoru na statystykę testową, mamy:

$$t = \frac{4,691 - 5}{0,404}\sqrt{22} \approx -3,59.$$

Sparowany test t – przykład

k \ α	0,2	0,1	0,05	0,04	0,02	0,01	0,002	0,001	...
1	3,078	6,314	12,708	15,895	31,821	63,657	318,306	636,627	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	1,323	1,721	2,080	2,189	2,518	2,831	3,527	3,819	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- Dla $t = -3,59$ oraz $k = 21$ stopni swobody, p-wartość dla testu dwustronnego wynosi $\alpha = 0,001$ (mniej niż typowy poziom istotności, np. $0,05$), zatem **odrzucamy** H_0.
- Różnica między średnią różnicą ciśnień po terapii a założoną wartością 5 mmHg jest **statystycznie istotna**.

Sparowany test t w *Pythonie* – przykład

- Różnice pomiędzy pomiarami dla 22 pacjentów:

3,9	4	4,2	4,2	4,3	4,4	4,4	4,4	4,5	4,6	4,8
4,8	4,9	4,9	5	5	5	5,1	5,1	5,2	5,2	5,3

- $n = 22, \quad \bar{Z} = 4,691, \quad S_Z = 0,404$
- $H_0 : \ m_Z = 5$

Sparowany test t w *Pythonie* – przykład

```
import numpy as np
from scipy import stats

u = np.array([3.9, 4, 4.2, 4.2, 4.3, 4.4, 4.4, 4.4, 4.5, 4.6, 4.8,
              4.8, 4.9, 4.9, 5, 5, 5, 5.1, 5.1, 5.2, 5.2, 5.3])

print(stats.ttest_1samp(u,popmean=5))
```

```
TtestResult(statistic=-3.473558170209485, pvalue=0.0022694734402210063, df=21)
```