METODY STATYSTYCZNE W BIOLOGII

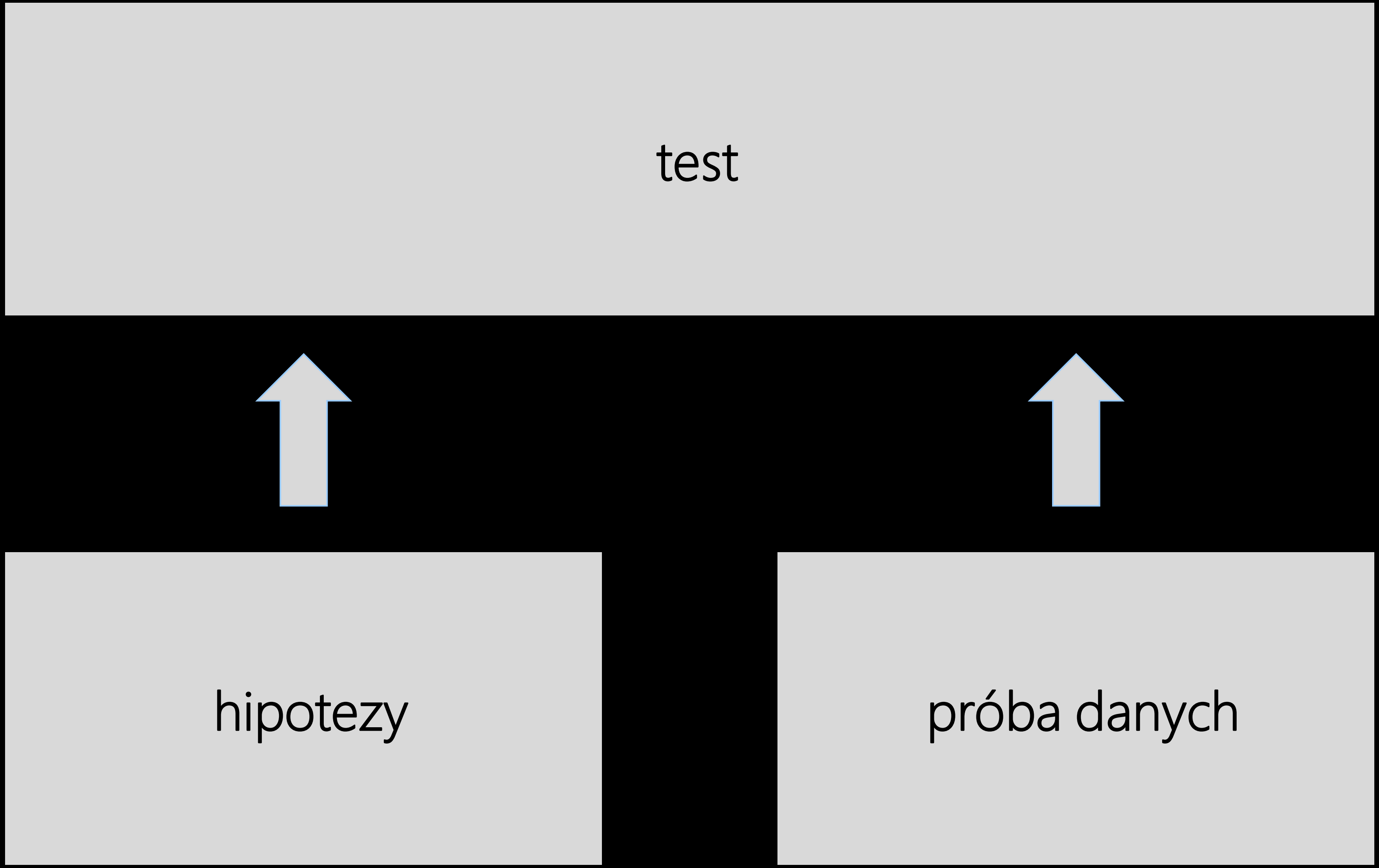
1. Wykład wstępny
2. Populacje i próby danych
3. Testowanie hipotez i estymacja parametrów
4. Planowanie eksperymentów biologicznych
5. Najczęściej wykorzystywane testy statystyczne I
6. Najczęściej wykorzystywane testy statystyczne II
7. Regresja liniowa
8. Regresja nieliniowa
9. Określenie jakości dopasowania równania regresji liniowej i nieliniowej
10. Korelacja
11. Elementy statystycznego modelowania danych
12. Porównywanie modeli
13. Analiza wariancji
14. Analiza kowariancji
15. Podsumowanie materiału, wspólna analiza przykładów, dyskusja

Wstęp

1. Test t
2. Test Manna-Whitneya
3. Test Wilcoxona

- Zakres stosowania
- Definicja
- Przykład

Wstęp



test t

Test t - zakres stosowania

1. Porównanie średnich arytmetycznych

2. Dane ciągłe

3. Rozkład normalny

4. Podobne wariancje

5. Warianty testu:

   - Pojedyncza próba $H_0: \mu = 1.0$

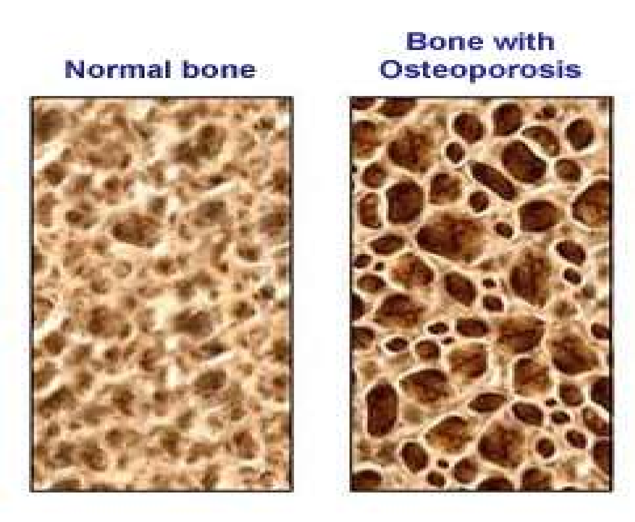
   - Dwie niezależne próby $H_0: \mu_A = \mu_B$
   - Dwie sparowane próby $H_0: \mu_A = \mu_B$

Test t – pojedyncza próba

próba danych

BMD	SEX
0.97	1
0.73	1
0.87	1
0.94	1
1.02	1
0.76	1
0.78	1
1.01	1
0.82	1
0.76	1
0.87	1
0.72	1
…	
0.91	2
1.02	2
0.87	2

1. Badanie osteoporozy
2. Medical Research Council, Cambridge
3. Gęstość kości [g/cm^2] 40 zdrowych osób dorosłych



Test t – pojedyncza próba

1. Określenie hipotez H_0 i H_1

   H_0: średnia gęstość kości w populacji wynosi 1.0 g/cm^2
   H_1: średnia gęstość kości w populacji różni się od 1.0 g/cm^2
   H_0: $\mu = 1.0$ H_1: $\mu \neq 1.0$

2. Ustalenie poziomu istotności

   $\alpha_{MAX} = 0.05$

3. Wybór i obliczenie wartości testu statystycznego

$$t = \frac{\bar{x} - 1.0}{\sqrt{\hat{\sigma}^2_{\bar{x}}}} = \frac{\bar{x} - 1.0}{\frac{\hat{\sigma}_x}{\sqrt{N}}} = \frac{0.85 - 1.0}{\frac{0.1040}{\sqrt{40}}} = -9.2434$$

Test t – pojedyncza próba

4. Określenie rozkładu testu

$$t = \frac{\bar{x} - 1.0}{\sqrt{\hat{\sigma}^2_{\bar{x}}}} = \frac{\bar{x} - 1.0}{\frac{\hat{\sigma}_x}{\sqrt{N}}} \quad \sim \quad t_{N-1}$$

5. Obliczenie wartości α_t

$$\alpha_t = 0.0000000000$$

Excel

R: `t_onesample <- t.test(bmd$BMD, mu = 1)`

6. Decyzja

$\alpha_t < \alpha_{max}$ $\cancel{H_0}$ H_1

średnia gęstość kości w populacji różni się od 1.0 g/cm^2

Test t - zakres stosowania

1. Porównanie średnich arytmetycznych
2. Dane ciągłe
3. Rozkład normalny
4. Podobne wariancje
5. Warianty testu:
   - Pojedyncza próba $H_0: \mu = 1.0$
   - Dwie niezależne próby $H_0: \mu_A = \mu_B$
   - Dwie sparowane próby $H_0: \mu_A = \mu_B$

This is probably the most widely used statistical test of all time.
Richard Lowry

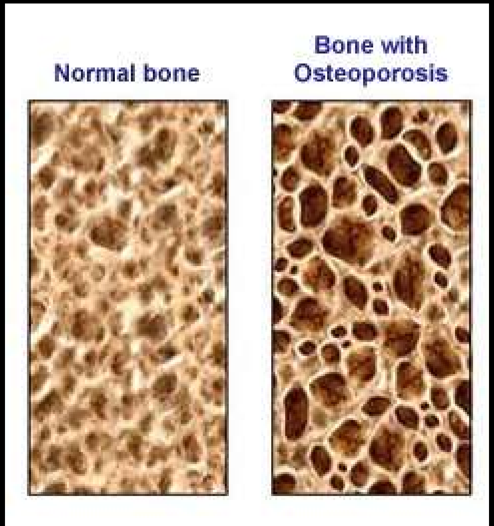
Test t – dwie niezależne próby

próba danych

BMD	SEX
0.97	1
0.73	1
0.87	1
0.94	1
1.02	1
0.76	1
0.78	1
1.01	1
0.82	1
0.76	1
0.87	1
0.72	1
…	
0.91	2
1.02	2
0.87	2

1. Badanie osteoporozy
2. Medical Research Council, Cambridge

1. Gęstość kości [g/cm2] 40 zdrowych osób dorosłych
2. Wartości znane dla mężczyzn i kobiet



Test t – dwie niezależne próby

1. Określenie hipotez H_0 i H_1

   H_0: gęstość kości kobiet jest taka sama jak mężczyzn
   H_1: gęstość kości kobiet jest różna niż mężczyzn
   H_0: $\mu_K = \mu_M$ H_1: $\mu_K \neq \mu_M$

2. Ustalenie poziomu istotności

   $\alpha_{MAX} = 0.05$

3. Wybór i obliczenie wartości testu statystycznego

$$t = \frac{\bar{x}_M - \bar{x}_K}{\sqrt{\frac{\hat{\sigma}_M^2}{N_M} + \frac{\hat{\sigma}_K^2}{N_K}}}$$

Test t – dwie niezależne próby

4. Określenie rozkładu testu

$$t = \frac{\bar{x}_M - \bar{x}_K}{\sqrt{\frac{\hat{\sigma}_M^2}{N_M} + \frac{\hat{\sigma}_K^2}{N_K}}} \sim t_{(N_M - 1)+(N_K - 1)}$$

5. Obliczenie wartości α_t

$$\alpha_t = 0.0466$$

Excel:

R: `t_twosamples <- t.test(genders$`1`,genders$`2`)`

6. Decyzja

$\alpha_t < \alpha_{max}$ $\cancel{H_0}$ H_1

gęstość kości u mężczyzn jest inna niż u kobiet, ..., jest wyższa

Test t - zakres stosowania

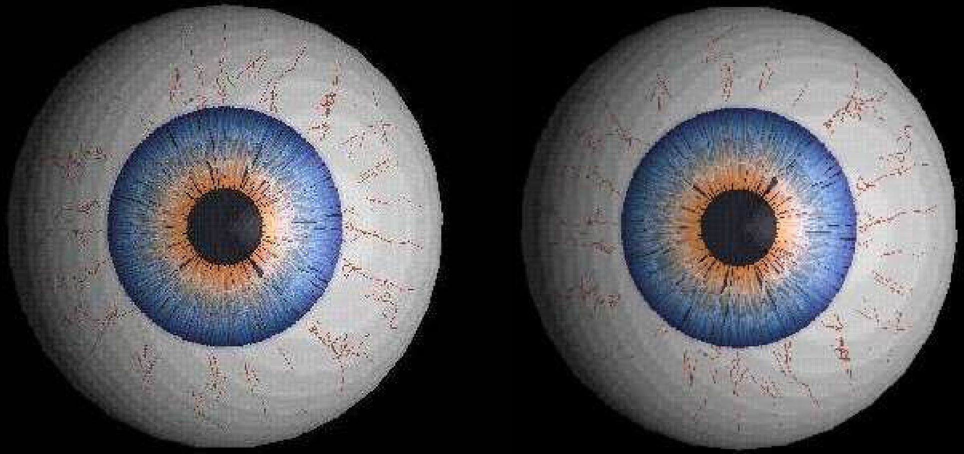
1. Porównanie średnich arytmetycznych
2. Dane ciągłe
3. Rozkład normalny
4. Podobne wariancje
5. Warianty testu:
   - Pojedyncza próba H_0: $\mu = 1.0$
   - Dwie niezależne próby H_0: $\mu_A = \mu_B$
   - Dwie sparowane próby H_0: $\mu_A = \mu_B$

Test t – dwie sparowane próby

PRÓBA DANYCH

Low CCT	High CCT
20.0	14.3
13.9	13.8
18.3	15.8
21.1	33.4
20.1	20.3
24.4	19.9
20.2	14.3
11.6	11.4
28.8	25.1
18.5	24.1

1. Badanie ciśnienia w gałce ocznej
2. Ciśnienie w 2 gałkach tej samej osoby
3. Podział oczu pod względem grubości rogówki (Low CCT i high CCT)



Test t – dwie sparowane próby

1. Określenie hipotez H_0 i H_1

   H_0: ciśnienie w gałce ocznej nie zależy od grubości rogówki
   H_1: ciśnienie w gałce ocznej zależy od grubości rogówki
   H_0: $\mu_L = \mu_H$ H_1: $\mu_L \neq \mu_H$

2. Ustalenie poziomu istotności

   $\alpha_{MAX} = 0.05$

3. Wybór i obliczenie wartości testu statystycznego

$$t = \frac{\overline{D}\sqrt{N}}{\sqrt{\frac{\sum_{i=1}^{N}\left(D_i - \overline{D}\right)^2}{N-1}}}, \qquad \overline{D} = \frac{\sum_{i=1}^{N}\left(x_{1i} - x_{2i}\right)}{N}$$

Test t – dwie sparowane próby

3. Wybór i obliczenie wartości testu statystycznego

$$\overline{D} = \frac{\sum_{i=1}^{N}\left(x_{1i} - x_{2i}\right)}{N} = \frac{4.5}{10} = 0.45$$

$$t = \frac{\overline{D}\sqrt{N}}{\sqrt{\frac{\sum_{i=1}^{N}\left(D_i - \overline{D}\right)^2}{N-1}}} = \frac{0.45\sqrt{10}}{\sqrt{\frac{288.21}{10-1}}} = 0.25$$

Test t – dwie sparowane próby

4. Określenie rozkładu testu

$$t = \frac{\overline{D}\sqrt{N}}{\sqrt{\frac{\sum_{i=1}^{N}\left(D_i - \overline{D}\right)^2}{N-1}}} \quad \sim \quad t_{N-1}$$

5. Obliczenie wartości α_t

$$\alpha_t = 0.8071$$

Excel: przykład

6. Decyzja

$\alpha_t > \alpha_{max}$ H_0 $\cancel{H_1}$

ciśnienie w gałce ocznej nie zależy od grubości rogówki

test Manna-Whitneya

test Manna-Whitneya - zakres stosowania

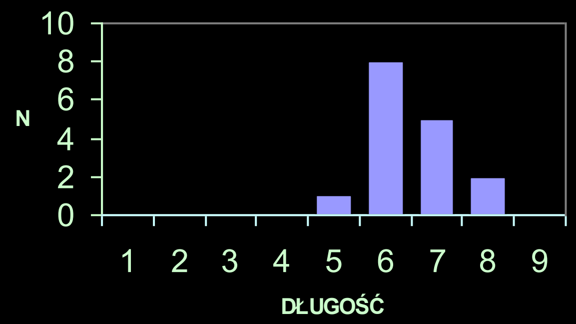
1. Porównanie średnich arytmetycznych
2. Dane ciągłe lub porządkowe (rangi)
3. Dane nie mają rozkładu normalnego
4. Dwie niezależne próby

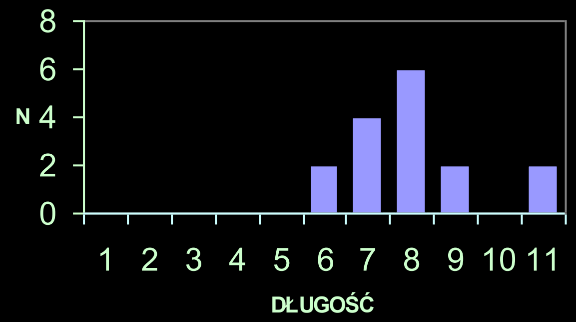
test Manna-Whitneya

próba danych

ŚREDNIE	WYSOKIE
5.5	6.0
6.0	7.0
5.0	7.5
7.0	6.0
5.5	7.5
6.0	8.0
7.0	11.0
6.0	9.0
8.0	8.0
7.0	11.0
6.0	8.0
7.0	8.0
6.0	7.0
8.0	7.0
6.0	7.0
7.0	9.0

- Długość krewetki w zależności od zasolenia wody
- Długość krewetki [mm] w wieku 4 tygodni





test Manna-Whitneya

1. Określenie hipotez H_0 i H_1

   H_0: długość krewetki nie zależy do zasolenia wody
   H_1: długość krewetki zależy od zasolenia wody
   H_0: $\mu_H = \mu_M$ H_1: $\mu_H \neq \mu_M$

2. Ustalenie poziomu istotności

   $\alpha_{MAX} = 0.05$

3. Wybór i obliczenie wartości testu statystycznego

$$U = \min\left(n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=1}^{n_2} r_{2i}, \quad n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum_{i=1}^{n_1} r_{1i} \right)$$

3. Wybór i obliczenie wartości testu statystycznego

$$U = min\left(n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=1}^{n_2} r_{2i}, n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum_{i=1}^{n_1} r_{1i}\right)$$

$$U = min\left(16 \cdot 16 + \frac{16(16+1)}{2} - 182, 16 \cdot 16 + \frac{16(16+1)}{2} - 342\right) = min(46{,}210)$$

$$U = 46$$

4. Określenie rozkładu testu

- Test nieparametryczny – brak rozkładu
- Dla $n_1 n_2 > 20$ – aproksymowany przez rozkład normalny:

$$U \sim N\left(\mu_U, \sigma_U^2\right) \quad \rightarrow \text{brak tablic}$$

$$z = \frac{U - \mu_U}{\sqrt{\sigma_U^2}} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0,1) \quad \rightarrow \text{tablice}$$

test Manna-Whitneya

4. Określenie rozkładu testu

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = \frac{46 - \frac{16*16}{2}}{\sqrt{\frac{16*16(33)}{12}}} = 3.09 \quad \sim \quad N(0,1)$$

5. Obliczenie wartości α_t:

$$\alpha_t = 0.002$$

Excel:

R: `wilcox.test(cm ~ zasolenie, data=shrimp)`

lub porównanie z wartością krytyczną:

$$U_{\alpha=0.05, n_1=16, n_2=16} = 181 \qquad U_t = 46$$

test Manna-Whitneya

6. Decyzja

$\alpha_t < \alpha_{max}$ $U_t < U$ $\cancel{H_0}$ H_1

UWAGA !!!

długość krewetki zależy od zasolenia wody

test Wilcoxona

zakres stosowalności testu Wilcoxona

1. Test nieparametryczny
2. Dane ciągłe lub porządkowe (rangi)
3. Dane nie mają rozkładu normalnego
4. Porównywane dwu zależnych = sparowanych prób danych

test Wilcoxona

próba danych

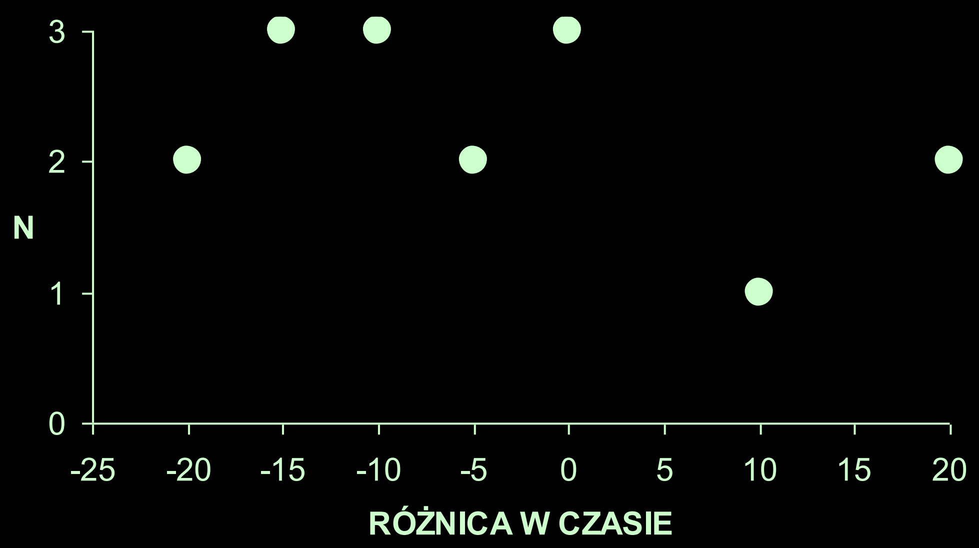
NR OWCY	BEZ JAGNIĘCIA	Z JAGNIĘCIEM
1	72.00	55.50
2	62.35	43.80
3	55.77	66.80
4	59.98	68.00
5	51.60	57.88
6	61.48	61.90
7	52.57	45.40
8	52.50	56.67
9	56.43	73.30
10	60.13	77.50
11	48.60	63.53
12	42.90	54.50
13	53.50	55.58
14	70.43	91.10
15	47.10	64.05
16	50.08	71.40

1. Próba danych – zachowanie się pokarmowe owiec
2. Dane zebrano w latach 1994-1996 w stadzie owiec utrzymywanym u podnóża Rocky Mountains w Kanadzie
3. Różnice w czasie pasienia się owcy z jagnięciem i bez
4. % czasu spędzanego na pasieniu się

test Wilcoxona

próba danych

NR OWCY	BEZ JAGNIĘCIA	Z JAGNIĘCIEM
1	72.00	55.50
2	62.35	43.80
3	55.77	66.80
4	59.98	68.00
5	51.60	57.88
6	61.48	61.90
7	52.57	45.40
8	52.50	56.67
9	56.43	73.30
10	60.13	77.50
11	48.60	63.53
12	42.90	54.50
13	53.50	55.58
14	70.43	91.10
15	47.10	64.05
16	50.08	71.40



test Wilcoxona

1. Określenie hipotez H_0 i H_1

   H_0: czas pasienia się owcy nie zależy od obecności jagnięcia
   H_1: czas pasienie się owcy zależy od obecności jagnięcia
   H_0: $\mu_J = \mu_B$ H_1: $\mu_J \neq \mu_B$

2. Ustalenie poziomu istotności

   $\alpha_{MAX} = 0.05$

3. Wybór i obliczenie wartości testu statystycznego

$$W = \min\left(\sum_{i=1}^{n_1} r_i^-, \sum_{i=n_1}^{N-n_1} r_i^+\right)$$

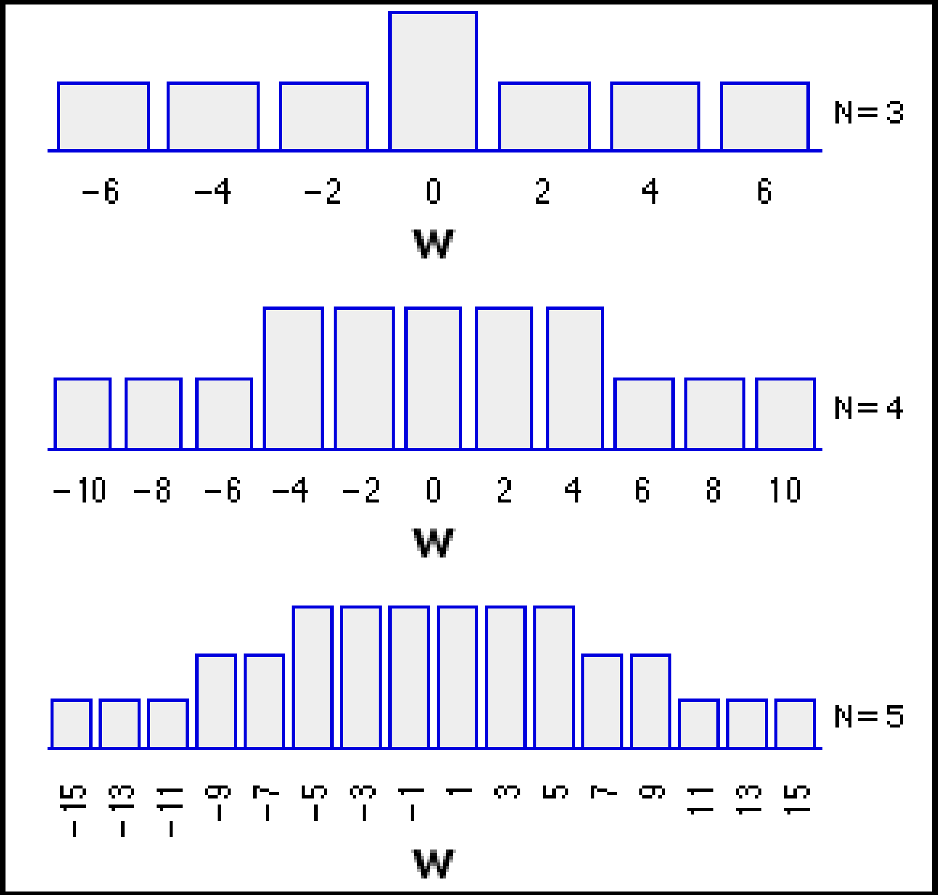
3. Wybór i obliczenie wartości testu statystycznego

$$W = \min\left(\sum_{i=1}^{n_1} r_i^-, \sum_{i=n_1}^{N-n_1} r_i^+\right) = \min(107,\ 29) = 29$$

test Wilcoxona

4. Określenie rozkładu testu

- Test nieparametryczny – brak rozkładu
- Dla N > 15 – aproksymowany przez rozkład normalny:



$$W = \quad \sim \quad N\left(\mu_W, \sigma_W^2\right)$$

$$z = \frac{W - \mu_W}{\sqrt{\sigma_W^2}} \quad \sim \quad N(0,1)$$

$$z = \frac{W - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \quad \sim \quad N(0,1)$$

$$z = \frac{29 - \frac{16*17}{4}}{\sqrt{\frac{16*17*33}{24}}} = -2.02 \quad \sim \quad N(0,1)$$

test Wilcoxona

5. Obliczenie wartości α_t:

$$\alpha_t = 0.0437$$

Excel: przykład

lub porównanie z wartością krytyczną:

$$W_{\alpha=0.05, N=16} = 29 \qquad W_t = 29$$

6. Decyzja

$\alpha_t < \alpha_{max}$ $\quad$ $W_t = W$ $\quad$ H_0 $\quad$ H_1 $\quad$?

czas pasienie się owcy nie/zależy od obecności jagnięcia ?

porównanie dwu prób

↓

próby zależne

↓

rozkład inny niż normalny

test Wilcoxona

porównanie dwu prób

próby niezależne

rozkład inny niż normalny

test Manna-Whitneya

porównanie dwu prób

próby zależne

rozkład normalny

test t

porównanie dwu prób

↓

próby niezależne

↓

rozkład normalny

test t

1. Test t
2. Test Manna-Whitneya
3. Test Wilcoxona

- Zakres stosowania
- Definicja
- Przykład