

Testowanie hipotez statystycznych.

Krzysztof Topolski

Bioinformatyka

Wrocław, 8 maja 2019

Temat.

Testy nieparametryczne.

Założmy, że mamy do czynienia z dwoma cechami X i Y , których rozkłady w populacji są opisane rozkładami o dystrybuantach odpowiednio F i G . Rozpatrzmy dwa niezależne ciągi niezależnych zmiennych losowych $X_1, \dots, X_{n_1}$ i $Y_1, \dots, Y_{n_2}$ o rozkładach odpowiednio z dystrybuantami F i G .

Wektory $(X_1, \dots, X_{n_1})$ i $(Y_1, \dots, Y_{n_2})$ stanowią one próby losowe na podstawie, których chcemy przetestować hipotezę zerową

$$H_0 : F(t) = G(t) \quad \text{dla wszystkich } t.$$

Jako hipotezę alternatywną możemy przyjąć

$$H_1 : G(t) \leq F(t) \quad \text{dla wszystkich } t \text{ i } G \neq F.$$

Lub gdy sądzimy, że kształt dystrybuanty jest taki sam w obu przypadkach

$$H_1 : G(t) = F(t - a) \quad \text{dla wszystkich } t \text{ gdzie } a > 0.$$

Wektory $(X_1, \dots, X_{n_1})$ i $(Y_1, \dots, Y_{n_2})$ stanowią one próby losowe na podstawie, których chcemy przetestować hipotezę zerową

$$H_0 : F(t) = G(t) \quad \text{dla wszystkich } t.$$

Jako hipotezę alternatywną możemy przyjąć

$$H_1 : G(t) \leq F(t) \quad \text{dla wszystkich } t \text{ i } G \neq F.$$

Lub gdy sądzimy, że kształt dystrybucyjny jest taki sam w obu przypadkach

$$H_1 : G(t) = F(t - a) \quad \text{dla wszystkich } t \text{ gdzie } a > 0.$$

Wektory $(X_1, \dots, X_{n_1})$ i $(Y_1, \dots, Y_{n_2})$ stanowią one próby losowe na podstawie, których chcemy przetestować hipotezę zerową

$$H_0 : F(t) = G(t) \quad \text{dla wszystkich } t.$$

Jako hipotezę alternatywną możemy przyjąć

$$H_1 : G(t) \leq F(t) \quad \text{dla wszystkich } t \text{ i } G \neq F.$$

Lub gdy sądzimy, że kształt dystrybucyjny jest taki sam w obu przypadkach

$$H_1 : G(t) = F(t - a) \quad \text{dla wszystkich } t \text{ gdzie } a > 0.$$

Przykład

	Liczba godzin pracy noża
Typ 1	28,5 20 26,5 24 30,5 27 19 19,5 31,5 20,5 17 24,5
Typ 2	34 22,5 25 29 38 25,5 28 23 33 26 21 30

Oznaczmy przez F dystrybuantę rozkładu czasu pracy noża typu 1 a przez G dystrybuantę czasu pracy noża typu 2.

Przykład

	Liczba godzin pracy noża
Typ 1	28,5 20 26,5 24 30,5 27 19 19,5 31,5 20,5 17 24,5
Typ 2	34 22,5 25 29 38 25,5 28 23 33 26 21 30

Oznaczmy przez F dystrybuantę rozkładu czasu pracy noża typu 1 a przez G dystrybuantę czasu pracy noża typu 2.

Przykład

	Liczba godzin pracy noża
Typ 1	28,5 20 26,5 24 30,5 27 19 19,5 31,5 20,5 17 24,5
Typ 2	34 22,5 25 29 38 25,5 28 23 33 26 21 30

Oznaczmy przez F dystrybuantę rozkładu czasu pracy noża typu 1 a przez G dystrybuantę czasu pracy noża typu 2.

Chcemy przetestować hipotezę

$$H_0 : F(t) = G(t) \quad \text{dla wszystkich } t.$$

Jako hipotezę alternatywną przyjmujemy hipotezę

$$H_1 : G(t) \leq F(t) \quad \text{dla wszystkich } t \text{ i } G \neq F.$$

Chcemy przetestować hipotezę

$$H_0 : F(t) = G(t) \quad \text{dla wszystkich } t.$$

Jako hipotezę alternatywną przyjmujemy hipotezę

$$H_1 : G(t) \leq F(t) \quad \text{dla wszystkich } t \text{ i } G \neq F.$$

Chcemy przetestować hipotezę

$$H_0 : F(t) = G(t) \quad \text{dla wszystkich } t.$$

Jako hipotezę alternatywną przyjmujemy hipotezę

$$H_1 : G(t) \leq F(t) \quad \text{dla wszystkich } t \text{ i } G \neq F.$$

W pierwszym kroku łączymy dane w jeden zbiór, a następnie porządkujemy go w porządku rosnącym i otrzymujemy

17 19 19,5 20 20,5 **21 22,5 23** 24 24,5 **25 25,5**

26 26,5 27 **28** 28,5 **29 30** 30,5 31,5 **33 34 38**

Obserwacje pogrubione to czasy pracy noży typu 2.

Widać, że mają one tendencję do zajmowania dalszych miejsc w tym ciągu. Może to świadczyć o tym, że czas pracy noży typu 2 jest statystycznie większy niż czas pracy noży typu 1. Aby to przypuszczenie zweryfikować wykonajmy drugi krok.

W kroku drugim przypiszmy każdej obserwacji rangę jaką zajmuje w połączonym ciągu obserwacji ciągu. Obserwacje odpowiadające czasom pracy noży typu 2 mają rangi

6 7 8 11 12 13 16 18 19 22 23 24.

W przypadku gdy prawdziwa jest hipoteza alternatywna czasy pracy noży typu 2 będą miały tendencje do zajmowania dalszych miejsc w połączonym ciągu.

Tak więc przypisane tym obserwacjom rangi będą miały tendencje do bycia wyższymi niż rangi przypisane obserwacjom czasów pracy noży typu 1.

W konsekwencji prowadzi to do wyższej wartości sumy rang obserwacji czasów pracy noży typu 2.

W przypadku gdy prawdziwa jest hipoteza alternatywna czasy pracy noży typu 2 będą miały tendencje do zajmowania dalszych miejsc w połączonym ciągu.

Tak więc przypisane tym obserwacjom rangi będą miały tendencje do bycia wyższymi niż rangi przypisane obserwacjom czasów pracy noży typu 1.

W konsekwencji prowadzi to do wyższej wartości sumy rang obserwacji czasów pracy noży typu 2.

W przypadku gdy prawdziwa jest hipoteza alternatywna czasy pracy noży typu 2 będą miały tendencje do zajmowania dalszych miejsc w połączonym ciągu.

Tak więc przypisane tym obserwacjom rangi będą miały tendencje do bycia wyższymi niż rangi przypisane obserwacjom czasów pracy noży typu 1.

W konsekwencji prowadzi to do wyższej wartości sumy rang obserwacji czasów pracy noży typu 2.

Dla danych z przykładu suma tych rang jest równa

$$w = 6 + 7 + 8 + 11 + 12 + 13 + 16 + 18 + 19 + 22 + 23 + 24 = 179.$$

Wartość ta jest większa niż 150, oczekiwana wartość tej sumy w przypadku zachodzenia hipotezy zerowej.

Definicja.

Statystyką Wilcoxona dla testowania hipotezy

$$H_0 : F(t) = G(t) \quad \text{dla wszystkich } t.$$

przeciwko hipotezie alternatywnej

$$H_1 : G(t) \leq F(t) \quad \text{dla wszystkich } t \text{ i } G \neq F,$$

nazywamy sumę rang obserwacji $Y_1, Y_2, \dots, Y_{n_2}$ w połączonej próbie $X_1, X_2, \dots, X_{n_1}, Y_1, Y_2, \dots, Y_{n_2}$.

Twierdzenie

Jeśli $F = G$ i dystrybuanta F jest ciągła to

- rozkład statystyki Wilcoxona W nie zależy od postaci dystrybuanty F
- $EW = \frac{n_2(n_1+n_2+1)}{2}$,
- $Var(W) = \frac{n_1 n_2(n_1+n_2+1)}{12}$,
- dla dowolnej liczby t

$$P\left(\frac{W - EW}{\sqrt{Var(W)}} \leq t\right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Twierdzenie

Jeśli $F = G$ i dystrybuanta F jest ciągła to

- rozkład statystyki Wilcoxona W nie zależy od postaci dystrybuanty F
- $EW = \frac{n_2(n_1+n_2+1)}{2}$,
- $Var(W) = \frac{n_1 n_2(n_1+n_2+1)}{12}$,
- dla dowolnej liczby t

$$P\left(\frac{W - EW}{\sqrt{Var(W)}} \leq t\right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Twierdzenie

Jeśli $F = G$ i dystrybuanta F jest ciągła to

- rozkład statystyki Wilcoxon W nie zależy od postaci dystrybuanty F
- $EW = \frac{n_2(n_1+n_2+1)}{2}$,
- $Var(W) = \frac{n_1 n_2 (n_1+n_2+1)}{12}$,
- dla dowolnej liczby t

$$P\left(\frac{W - EW}{\sqrt{Var(W)}} \leq t\right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Twierdzenie

Jeśli $F = G$ i dystrybuanta F jest ciągła to

- rozkład statystyki Wilcoxon W nie zależy od postaci dystrybuanty F
- $EW = \frac{n_2(n_1+n_2+1)}{2}$,
- $Var(W) = \frac{n_1 n_2 (n_1+n_2+1)}{12}$,
- dla dowolnej liczby t

$$P\left(\frac{W - EW}{\sqrt{Var(W)}} \leq t\right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Twierdzenie

Jeśli $F = G$ i dystrybuanta F jest ciągła to

- rozkład statystyki Wilcoxona W nie zależy od postaci dystrybuanty F
- $EW = \frac{n_2(n_1+n_2+1)}{2}$,
- $Var(W) = \frac{n_1 n_2 (n_1+n_2+1)}{12}$,
- dla dowolnej liczby t

$$P\left(\frac{W - EW}{\sqrt{Var(W)}} \leq t\right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Dla danych z przykładu wartość statystyki Wilcoxsona wynosi $w = 179$. Przy spełnieniu hipotezy zerowej, na mocy twierdzenia,

- $EW = \frac{n_2(n_1+n_2+1)}{2} = \frac{12 \times (12+12+1)}{2} = 150,$
- $Var(W) = \frac{n_1 n_2 (n_1+n_2+1)}{12} = \frac{12 \times 12 \times (12+12+1)}{12} = 300,$
-

$$P(W \geq 179) = P\left(\frac{W - 150}{\sqrt{300}} \geq \frac{179 - 150}{\sqrt{300}}\right) \approx 1 - \Phi(1,6743) \\ = 0,0505.$$

Stosując test t –Studenta otrzymamy p –wartość równą 0,034.

Dla danych z przykładu wartość statystyki Wilcoxsona wynosi $w = 179$. Przy spełnieniu hipotezy zerowej, na mocy twierdzenia,

- $EW = \frac{n_2(n_1+n_2+1)}{2} = \frac{12 \times (12+12+1)}{2} = 150,$
- $Var(W) = \frac{n_1 n_2 (n_1+n_2+1)}{12} = \frac{12 \times 12 \times (12+12+1)}{12} = 300,$



$$P(W \geq 179) = P\left(\frac{W - 150}{\sqrt{300}} \geq \frac{179 - 150}{\sqrt{300}}\right) \approx 1 - \Phi(1,6743) \\ = 0,0505.$$

Stosując test t –Studenta otrzymamy p –wartość równą 0,034.

Dla danych z przykładu wartość statystyki Wilcoxsona wynosi $w = 179$. Przy spełnieniu hipotezy zerowej, na mocy twierdzenia,

- $EW = \frac{n_2(n_1+n_2+1)}{2} = \frac{12 \times (12+12+1)}{2} = 150,$
- $Var(W) = \frac{n_1 n_2 (n_1+n_2+1)}{12} = \frac{12 \times 12 \times (12+12+1)}{12} = 300,$
-

$$\begin{aligned} P(W \geq 179) &= P\left(\frac{W - 150}{\sqrt{300}} \geq \frac{179 - 150}{\sqrt{300}}\right) \approx 1 - \Phi(1,6743) \\ &= 0,0505. \end{aligned}$$

Stosując test t –Studenta otrzymamy p –wartość równą 0,034.

Dla danych z przykładu wartość statystyki Wilcoxsona wynosi $w = 179$. Przy spełnieniu hipotezy zerowej, na mocy twierdzenia,

- $EW = \frac{n_2(n_1+n_2+1)}{2} = \frac{12 \times (12+12+1)}{2} = 150,$
- $Var(W) = \frac{n_1 n_2 (n_1+n_2+1)}{12} = \frac{12 \times 12 \times (12+12+1)}{12} = 300,$
-

$$\begin{aligned} P(W \geq 179) &= P\left(\frac{W - 150}{\sqrt{300}} \geq \frac{179 - 150}{\sqrt{300}}\right) \approx 1 - \Phi(1,6743) \\ &= 0,0505. \end{aligned}$$

Stosując test t –Studenta otrzymamy p –wartość równą 0,034.

Dla danych z przykładu wartość statystyki Wilcoxsona wynosi $w = 179$. Przy spełnieniu hipotezy zerowej, na mocy twierdzenia,

- $EW = \frac{n_2(n_1+n_2+1)}{2} = \frac{12 \times (12+12+1)}{2} = 150,$
- $Var(W) = \frac{n_1 n_2 (n_1+n_2+1)}{12} = \frac{12 \times 12 \times (12+12+1)}{12} = 300,$
-

$$\begin{aligned} P(W \geq 179) &= P\left(\frac{W - 150}{\sqrt{300}} \geq \frac{179 - 150}{\sqrt{300}}\right) \approx 1 - \Phi(1,6743) \\ &= 0,0505. \end{aligned}$$

Stosując test t –Studenta otrzymamy p –wartość równą 0,034.

Test Wilcoxona dla par obserwacji

Test Wilcoxona możemy również użyć do testowania hipotez dotyczących prób zależnych. Rozpatrzmy sytuację, w której mamy do czynienia z parami obserwacji $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, gdzie pary są wzajemnie niezależne ale obserwacje w parach są zależne. Zakładamy, że pary pochodzą z obserwacji tego samego rozkładu dwuwymiarowego.

Test Wilcoxona dla par obserwacji

Chcemy przetestować hipotezę mówiącą o tym, że rozkłady brzegowe są takie same. Czyli, że zmienna losowa X ma taki sam rozkład jak zmienna losowa Y . Hipotezę o równości rozkładów możemy zastąpić ogólniejszą hipotezą mówiącą, że rozkłady zmiennych losowych $D_i = X_i - Y_i$ oraz $-D_i = Y_i - X_i$ są takie same.

$$H_0 : F(t) = 1 - F(-t) \quad \text{dla każdego } t.$$

Test Wilcoxona dla par obserwacji

Jako hipotezę alternatywną możemy w zależności od naszych potrzeb wybrać

$$H_0 : F(t) \leq 1 - F(-t) \quad \text{dla każdego } t$$

gdy przypuszczamy, że dystrybuanta zmiennej Y przewyższa dystrybuantę X ,

lub

$$H_0 : F(t) \geq 1 - F(-t) \quad \text{dla każdego } t$$

gdy przypuszczamy, że dystrybuanta zmiennej X przewyższa dystrybuantę Y .

Test Wilcoxona dla par obserwacji

Jako hipotezę alternatywną możemy w zależności od naszych potrzeb wybrać

$$H_0 : F(t) \leq 1 - F(-t) \quad \text{dla każdego } t$$

gdy przypuszczamy, że dystrybuanta zmiennej Y przewyższa dystrybuantę X ,
lub

$$H_0 : F(t) \geq 1 - F(-t) \quad \text{dla każdego } t$$

gdy przypuszczamy, że dystrybuanta zmiennej X przewyższa dystrybuantę Y .

Przykład

	Czas pokonywania labiryntu
Pierwszy dzień eksperymentu	x_i : 85 122 162 206 121 250
Dziesiąty dzień eksperymentu	y_i : 87 82 158 96 131 121

Ciąg różnic ma postać

$$d_i : -2, 40, 4, 110, -10, 129.$$

Przykład

	Czas pokonywania labiryntu
Pierwszy dzień eksperymentu	x_i : 85 122 162 206 121 250
Dziesiąty dzień eksperymentu	y_i : 87 82 158 96 131 121

Ciąg różnic ma postać

$$d_i : -2, 40, 4, 110, -10, 129.$$

Przykład

	Czas pokonywania labiryntu
Pierwszy dzień eksperymentu	x_i : 85 122 162 206 121 250
Dziesiąty dzień eksperymentu	y_i : 87 82 158 96 131 121

Ciąg różnic ma postać

$$d_i : -2, 40, 4 \ 110 \ -10 \ 129.$$

Jeśli czasy przebycia labiryntu ulegają systematycznemu skróceniu, to dodatnie różnice powinny bardziej odstawać od zera niż różnice ujemne.

Porządkując wartości bezwzględne otrzymujemy

$$|d_i| : 2 \ 4 \ 10 \ 40 \ 110 \ 129,$$

gdzie wyróżnione są wartości dodatnie.

Jeśli czasy przebycia labiryntu ulegają systematycznemu skróceniu, to dodatnie różnice powinny bardziej odstawać od zera niż różnice ujemne.

Porządkując wartości bezwzględne otrzymujemy

$$|d_i| : 2 \ 4 \ 10 \ 40 \ 110 \ 129,$$

gdzie wyróżnione są wartości dodatnie.

Jeśli czasy przebycia labiryntu ulegają systematycznemu skróceniu, to dodatnie różnice powinny bardziej odstawać od zera niż różnice ujemne.

Porządkując wartości bezwzględne otrzymujemy

$$|d_i| : 2 \ 4 \ 10 \ \mathbf{40} \ \mathbf{110} \ \mathbf{129},$$

gdzie wyróżnione są wartości dodatnie.

Definicja.

Statystyką Wilcoxona W^+ dla par obserwacji jest zdefiniowana jako suma rang wartości bezwzględnych różnic odpowiadających różnicom dodatnim.

W przykładzie

$$w^+ = 2 + 4 + 5 + 6 = 17.$$

Definicja.

Statystyką Wilcoxona W^+ dla par obserwacji jest zdefiniowana jako suma rang wartości bezwzględnych różnic odpowiadających różnicom dodatnim.

W przykładzie

$$w^+ = 2 + 4 + 5 + 6 = 17.$$

Definicja.

Statystyką Wilcoxona W^+ dla par obserwacji jest zdefiniowana jako suma rang wartości bezwzględnych różnic odpowiadających różnicom dodatnim.

W przykładzie

$$w^+ = 2 + 4 + 5 + 6 = 17.$$

Twierdzenie

Jeśli F jest ciągła oraz prawdziwa jest hipoteza

$$H_0 : F(t) = 1 - F(-t) \quad \text{dla każdego } t,$$

to

- rozkład statystyki Wilcoxa W^+ nie zależy od postaci dystrybuanty F
- $EW^+ = \frac{n(n+1)}{4},$
- $Var(W^+) = \frac{n(n+1)(2n+1)}{24},$
- dla dowolnej liczby t

$$P\left(\frac{W^+ - EW^+}{\sqrt{Var(W^+)}} \leq t\right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Twierdzenie

Jeśli F jest ciągła oraz prawdziwa jest hipoteza

$$H_0 : F(t) = 1 - F(-t) \quad \text{dla każdego } t,$$

to

- rozkład statystyki Wilcoxona W^+ nie zależy od postaci dystrybuanty F
- $EW^+ = \frac{n(n+1)}{4}$,
- $Var(W^+) = \frac{n(n+1)(2n+1)}{24}$,
- dla dowolnej liczby t

$$P\left(\frac{W^+ - EW^+}{\sqrt{Var(W^+)}} \leq t\right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Twierdzenie

Jeśli F jest ciągła oraz prawdziwa jest hipoteza

$$H_0 : F(t) = 1 - F(-t) \quad \text{dla każdego } t,$$

to

- rozkład statystyki Wilcoxona W^+ nie zależy od postaci dystrybuanty F
- $EW^+ = \frac{n(n+1)}{4}$,
- $Var(W^+) = \frac{n(n+1)(2n+1)}{24}$,
- dla dowolnej liczby t

$$P\left(\frac{W^+ - EW^+}{\sqrt{Var(W^+)}} \leq t\right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Twierdzenie

Jeśli F jest ciągła oraz prawdziwa jest hipoteza

$$H_0 : F(t) = 1 - F(-t) \quad \text{dla każdego } t,$$

to

- rozkład statystyki Wilcoxona W^+ nie zależy od postaci dystrybuanty F
- $EW^+ = \frac{n(n+1)}{4},$
- $Var(W^+) = \frac{n(n+1)(2n+1)}{24},$
- dla dowolnej liczby t

$$P \left(\frac{W^+ - EW^+}{\sqrt{Var(W^+)}} \leq t \right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Twierdzenie

Jeśli F jest ciągła oraz prawdziwa jest hipoteza

$$H_0 : F(t) = 1 - F(-t) \quad \text{dla każdego } t,$$

to

- rozkład statystyki Wilcoxona W^+ nie zależy od postaci dystrybuanty F
- $EW^+ = \frac{n(n+1)}{4}$,
- $Var(W^+) = \frac{n(n+1)(2n+1)}{24}$,
- dla dowolnej liczby t

$$P \left(\frac{W^+ - EW^+}{\sqrt{Var(W^+)}} \leq t \right) \rightarrow \Phi(t),$$

gdy $\min(n_1, n_2) \rightarrow \infty$ a Φ oznacza dystrybuantę standardowego rozkładu normalnego.

Dla danych z przykładu wartość statystyki Wilcoxsona W^+ wynosi $w^+ = 17$. Przy spełnieniu hipotezy zerowej, na mocy twierdzenia,

- $EW^+ = \frac{n(n+1)}{4} = \frac{6 \times 7}{4} = 10,5,$

- $\sqrt{\text{Var}(W^+)} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{6 \times 7 \times 13}{24}} = 4,77.$

-

$$P(W^+ \geq 17) = P\left(\frac{W^+ - 10,5}{4,77} \geq \frac{16,5 - 10,5}{4,77}\right) \approx 1 - \Phi(1,05) \\ = 0,147$$

Dla danych z przykładu wartość statystyki Wilcoxsona W^+ wynosi $w^+ = 17$. Przy spełnieniu hipotezy zerowej, na mocy twierdzenia,

- $EW^+ = \frac{n(n+1)}{4} = \frac{6 \times 7}{4} = 10,5,$
- $\sqrt{\text{Var}(W^+)} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{6 \times 7 \times 13}{24}} = 4,77.$

•

$$P(W^+ \geq 17) = P\left(\frac{W^+ - 10,5}{4,77} \geq \frac{16,5 - 10,5}{4,77}\right) \approx 1 - \Phi(1,05) \\ = 0,147$$

Dla danych z przykładu wartość statystyki Wilcoxsona W^+ wynosi $w^+ = 17$. Przy spełnieniu hipotezy zerowej, na mocy twierdzenia,

- $EW^+ = \frac{n(n+1)}{4} = \frac{6 \times 7}{4} = 10,5,$
- $\sqrt{\text{Var}(W^+)} = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{6 \times 7 \times 13}{24}} = 4,77.$
-

$$P(W^+ \geq 17) = P\left(\frac{W^+ - 10,5}{4,77} \geq \frac{16,5 - 10,5}{4,77}\right) \approx 1 - \Phi(1,05) \\ = 0,147$$

Uwaga

Dla małych liczebności próby, pakiety statystyczne wyliczają rozkład dokładny statystyki testowej a następnie na jej podstawie wyliczają p -wartość. Dla rozpatrywanego przykładu p -wartość wyliczona na podstawie rozkładu dokładnego statystyki W^+ wynosi 0,147.