Test Wilcoxona

Mateusz Lewko

6 czerwca 2018

1 Wstęp

Test sumy rang Wilcoxona służy do sprawdzenia, czy wartości próbek pobranych z dwóch niezależnych populacji, są jednakowo duże. Innymi słowy, pozwala sprawdzić, czy dwie, losowo wybrane próbki, zostały wybrane z populacji o takim samym rozkładzie.

1.1 Koniecznie założenia

1. Wszystkie obserwacje z obydwu grup są od siebie niezależne.
2. Obserwacje możemy porównać i uporządkować.
3. Hipoteza zerowa: obserwacje pochodzą z populacji o takim samym rozkładzie.
4. Hipoteza alternatywna: rozkłady populacji są różne.

2 Metoda obliczania

2.1 Opis testu

Po przeprowadzeniu eksperymentu dla dwóch grup osobników, otrzymaliśmy odpowiednio n_1 i n_2 pomiarów. Chcemy sprawdzić hipotezę, że rozkłady wyników dla obu populacji są takie same ($H_0 : G_1 = G_2$). Test Wilcoxona wykryje przesunięcie G_1 względem G_2. Gdy podejrzewamy, że grupa pierwsza jest przesunięta w prawo względem grupy pierwszej, to hipoteza alternatywna przyjmuje postać $H_a : G_1 < G_2$. W ogólnym przypadku, gdy nie podejrzewamy żadnego konkretnego kierunku przesunięcia, to $H_a : G_1 \neq G_2$.

Test Wilcoxona bazuje na uporządkowaniu wszystkich $n = n_1 + n_2$ obserwacji. Każda obserwacja otrzymuje rangę: najmniejsza ma rangę równą 1, a największa rangę n. Statystyką testową W będzie suma rang obserwacji dla jednej z grup. Niech

$$w_1 = \text{suma rang obserwacji z grupy pierwszej.} \tag{1}$$

2.2 Przykład obliczania rang

Po uporządkowaniu obserwacji z dwóch grup otrzymaliśmy następującą kolejność:

Number grupy	1	1	2	2	1	2
Ranga	1	2	3	4	5	6

Wtedy $w_1 = 1 + 2 + 5 = 8$ oraz $w_2 = 3 + 4 + 6 = 13$.

2.3 Rozkład testu rang

W celu wyznaczenia *wartości_p*, należy znać rozkład prawdopodobieństwa możliwych sum rang. Będziemy wtedy w stanie określić, jak prawdopodobne jest otrzymanie sumy rang jednej z grup, przy założeniu, że grupy pochodzą z tego samego rozkładu. Jeżeli obserwacje pochodzą z tych samych rozkładów, to każda ich kolejność (permutacja) jest tak samo prawdopodobna. To rozumowanie nasuwa następujący sposób na wyznaczenie gęstości sum rang:

1. Wyznaczyć wszystkie permutacje elementów z jednej grupy.
2. Dla każdej permutacji obliczyć sumę jej rang.
3. Częstość występowania danej sumy to prawdopodobieństwo jej otrzymania. Prawdopodobieństwo rang r i $n_1 + n_2 - r$ są takie same – gęstość jest symetryczna.

2.4 Obliczanie *wartości_p*

Załóżmy, że $H_a : G_1 > G_2$, wtedy oczekiwalibyśmy, że suma rang w grupie drugiej będzie nadzwyczaj duża, zgodnie z rozkładem sum rang. Stąd *wartość_p* to

$$wartości\_p = P(W \geq w_1) \tag{2}$$

W przypadku, gdy spodziewamy się, że $H_a : G_1 < G_2$, to *wartość_p* obliczamy analogicznie.

Dal dwustronnego testu z $H_a : G_1 \neq G_2$, *wartość_p* przyjmuje postać

$$wartości\_p = 2 * P(W \geq w_1) \tag{3}$$
$$\text{lub} \tag{4}$$
$$wartości\_p = 2 * P(W \leq w_1) \tag{5}$$

w zależności od tego, w którym kwantylu się znajdziemy.

3 Przybliżanie rozkładem normalnym

Jeżeli n_1 i n_2 są większe od 10, to statystyka W ma w przybliżaniu rozkład $N(\mu_1, \sigma_1)$, gdzie

$$\mu_1 = \frac{n_1(n_1 + n_2 + 1)}{2} \tag{6}$$
$$\sigma_1 = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \tag{7}$$

3.1 Przykład

Dane są grupy o licznościach $n_1 = 10$, $n_2 = 12$. Chcemy policzyć $P(W_1 \geq 145)$. Wiemy że

$$P(W_1 \geq w_a) \approx P(Z \geq z) \tag{8}$$
$$z = \frac{w_1 - \mu_1}{\sigma_1} \tag{9}$$
$$Z \sim N(0, 1) \tag{10}$$

Wtedy, $\mu_1 = 10 \times (10 + 12 + 1)/2 = 115$ oraz

$$\sigma_1 = \sqrt{\frac{10 \times 12 \times (10 + 12 + 1)}{12}} = 15.16575 \tag{11}$$

$$P(W_1 \geq 145) \approx P\Big(Z \geq \frac{145 - 115}{15.16575}\Big) = P(Z \geq 1.978) = 0.024 \tag{12}$$

3.2 Uporządkowanie zawierające remisy

W przypadku występowania remisów po uporządkowaniu obserwacji, wzór na σ_1 jest następujący

$$\sigma_1 = \sqrt{\frac{n_1 n_2}{12}\left((n+1) - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{n(n-1)}\right)} \tag{13}$$

gdzie $n = n_1 + n_2$, t_i to liczba różnych obserwacji z rangą i, a k to liczba różnych rang.

Literatura

[1] https://www.stat.auckland.ac.nz/~wild/ChanceEnc/Ch10.wilcoxon.pdf

[2] https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test?oldformat=true