Jerzy W a w r z y n e k (Wrocław)

WYBRANE TESTY NIEPARAMETRYCZNE

Wstęp

Centralnym zagadnieniem współczesnej statystyki jest sprawdzanie hipotez statystycznych. Hipotezą statystyczną nazywamy każde przypuszczenie dotyczące rozkładu cechy elementów w populacjach generalnych. Metody służące weryfikacji hipotez statystycznych noszą nazwę testów statystycznych.

Rozważmy weryfikowanie hipotezy na podstawie próbki o ustalonej liczności, że mediana rozkładu cechy w populacji jest równa zero. Gdy wiadomo, że rozkład cechy jest normalny, mediana jest równa wartości oczekiwanej. Hipotezę tę weryfikuje się wówczas za pomocą testu Studenta. Test Studenta oparty jest na statystyce Studenta, która przy prawdziwości weryfikowanej hipotezy i normalności rozkładu cechy ma znany rozkład prawdopodobieństwa, a mianowicie rozkład Studenta. Z tablic rozkładu Studenta znajdujemy dla tej statystyki wartości krytyczne i w ten sposób otrzyma się test, który prowadzi do odrzucenia weryfikowanej hipotezy, jeśli ona jest prawdziwa, ze znanym prawdopodobieństwem. Można też weryfikować rozważaną hipotezę za pomocą testu opartego na innej statystyce, a mianowicie na liczbie elementów próbki większych od zera. Przy prawdziwości hipotezy liczba elementów próbki większych od zera ma również znany rozkład prawdopodobieństwa. Jest to rozkład binomialny, w którym prawdopodobieństwa sukcesu i porażki są jednakowe. Wartości krytyczne dla tej statystyki wyznacza się na podstawie tego rozkładu binomialnego i otrzymuje test, który prowadzi do odrzucenia weryfikowanej hipotezy, jeśli ona jest prawdziwa, z prawdopodobieństwem znanym. Ten ostatni test nazywa się testem znaków.

Pierwszą różnicą między testem znaków a testem Studenta jest to, że dużo łatwiej jest obliczyć wartość statystyki, na której oparty jest test znaków niż wartość statystyki, na której oparty jest test Studenta. Druga różnica: o statystyce Studenta wiadomo, że ma rozkład Studenta pod warunkiem, że rozkład cechy w populacji jest normalny. Nie wiadomo, jak zmieni się ten rozkład, a wraz z nim prawdopodobieństwo odrzucenia weryfikowanej hipotezy, gdy ona jest prawdziwa, jeśli rozkład cechy w populacji nie jest normalny. Natomiast liczba elementów próbki większych od zera będzie mieć ten sam rozkład binomialny, jeśli tylko prawdopodobieństwo zaobserwowania dodatniej wartości cechy jest takie samo, jak prawdopodobieństwo zaobserwowania ujemnej wartości cechy, czyli równe $\frac{1}{2}$. A więc rozkład cechy w populacji wcale nie musi być normalny. Znaczy to, że za pomocą testu znaków możemy weryfikować hipotezę, że mediana rozkładu cechy jest równa zero, bez żadnych dodatkowych informacji o rozkładzie cechy w populacji, podczas gdy dla stosowania testu Studenta trzeba wiedzieć, że rozkład cechy w populacji jest normalny. Test znaków nazywamy nieparametrycznym dlatego, że nadaje się on do weryfikowania sformułowanej hipotezy bez ograniczających dodatkowych założeń o rozkładzie cechy w populacji, w odróżnieniu od parametrycznego testu Studenta, który oparty jest na dodatkowych założeniach (normalność rozkładu). Testy nieparametryczne wyróżniają się tym, że statystyki, na których są oparte, mają rozkłady niezależne od dodatkowych ograniczających założeń o rozkładzie cechy w populacji i często są łatwe do obliczenia.

Bibliografia tematu jest obszerna. Z książek wymienimy tu monografię Frasera [4], poglądową pracę Siegela [19] i niedawno wydane dzieło Hájeka i Šidáka [7], zaś z autorów polskich - prace Elandt [2], Łukaszewicza i Sadowskiego [11] oraz Sadowskiego [17].

W tej pracy przedstawimy najpierw potrzebne dalej definicje z ogólnej teorii testów. Potem nastąpi opis kilkunastu testów nieparametrycznych. Przedstawimy kolejno test o losowości próbki, testy dla dwu próbek łączonych w pary, testy dla jednej próbki, testy dla dwu próbek niezależnych oraz testy dla większej liczby próbek, i wreszcie test istotności korelacji. Opis informuje o tym, jaką hipotezę można za pomocą danego testu zweryfikować, jak należy wtedy wybrać próbki, jakie założenia musi spełniać rozkład badanej cechy, jaka statystyka służy za podstawę testu i w jaki sposób z zaobserwowanej wartości tej statystyki wnioskować o hipotezie. Do każdego testu dołączony jest przykład liczbowy. Podane są też rozkłady asymptotyczne statystyk występujących w omawianych testach. Na końcu pracy umieszczono tablice rozkładów związanych z omawianymi testami, których brak w [18].

Weryfikowanie hipotez statystycznych

n-elementową próbką losową z populacji generalnej nazywamy ciąg wartości badanej cechy u elementów z tej populacji.

Hipotezę, którą mamy zweryfikować, oznaczamy przez H. Pozostałe możliwe hipotezy nazywamy hipotezą konkurencyjną i oznaczamy przez K.

Test jest to przyporządkowanie próbkom jednego z dwu orzeczeń: "hipoteza H jest prawdziwa", albo "hipoteza K jest prawdziwa". Test jest określony przez wskazanie tych wszystkich próbek, dla których orzekamy prawdziwość hipotezy K. Zbiór wszystkich takich próbek nazywamy przestrzenią odrzuceń hipotezy H.

Przestrzeń odrzuceń określa się z reguły na podstawie pewnej statystyki U, to jest liczbowej funkcji próbki. Gdy elementy próbki traktujemy jako zmienne losowe, sama statystyka U staje się zmienną losową. Statystyki, które będziemy rozważać, mają na ogół dyskretny rozkład prawdopodobieństwa.

Znając rozkład statystyki U przy prawdziwości hipotezy H możemy dla danej liczby α spełniającej warunki $0<\alpha<1$ znaleźć liczbę $u_{\frac{\alpha}{2}}$ taką, że nierówność $U\leqslant u_{\frac{\alpha}{2}}$ zachodzi z prawdopodobieństwem nie większym niż $\frac{\alpha}{2}$, ale możliwie bliskim $\frac{\alpha}{2}$ oraz liczbę $u_{1-\frac{\alpha}{2}}$ taką, że nierówność $u_{1-\frac{\alpha}{2}}\leqslant U$ zachodzi z prawdopodobieństwem nie większym niż $\frac{\alpha}{2}$, ale możliwie bliskim $\frac{\alpha}{2}$. Następnie orzekamy, że prawdziwa jest hipoteza K wtedy i tylko wtedy, gdy zaobserwowana wartość statystyki U nie jest większa od $u_{\frac{\alpha}{2}}$ lub nie jest mniejsza od $u_{1-\frac{\alpha}{2}}$. Zbudowany w ten sposób test nazywamy testem dwustronnym. Możemy również znaleźć taką liczbę u_α, że nierówność $U\leqslant u_\alpha$ zachodzi z prawdopodobieństwem możliwie bliskim, choć nie większym od α i orzekać, że prawdziwa jest hipoteza K, gdy zaobserwowana wartość statystyki U jest równa lub mniejsza od u_α. Wreszcie możemy wyznaczyć taką liczbę $u_{1-\alpha}$, żeby przy prawdziwości hipotezy H nierówność $u_{1-\alpha} \leqslant U$ zachodziła z prawdopodobieństwem możliwie bliskim, choć nie większym niż α i orzekać, że prawdziwa jest hipoteza K, gdy zaobserwowana wartość statystyki U jest równa lub większa od $u_{1-\alpha}$. Tak zbudowane testy nazywamy jednostronnymi. Wszystkie opisane wyżej testy mają tę własność, że prowadzą do odrzucenia hipotezy H, jeśli ona jest prawdziwa, z prawdopodobieństwem bliskim α, ale nie przekraczającym tej liczby. Liczbę α nazywa się poziomem istotności testu.

Test dwustronny buduje się często wtedy, gdy hipoteza K jest zaprzeczeniem hipotezy H. Np. gdy hipoteza H mówi, że mediana w populacji jest równa zero, a hipoteza K mówi, że mediana jest różna od zera, właściwe jest użycie testu dwustronnego, opartego na liczbie dodatnich elementów w próbce. W niektórych przypadkach właściwsze jest użycie testu jednostronnego. Np. jeśli hipoteza H pozostaje niezmieniona, a hipoteza K mówi, że mediana w populacji jest większa od zera, należy użyć testu jednostronnego.

Test o losowości próbki

Będziemy weryfikowali hipotezę H, że próbka została pobrana losowo, czyli że można ją uważać za realizację niezależnych zmiennych losowych o tym samym rozkładzie prawdopodobieństwa.

Oto przykład dwu hipotez konkurencyjnych: tendencja do wybierania coraz większych wartości do próbki, tendencja do wybierania na przemian wartości dużych i małych do próbki.

Zbiór wartości w próbce dzielimy na dwie rozłączne części. Elementy próbki należące do jednej części oznaczamy literami a, pozostałe - literami b.

Serią nazywamy ciąg identycznych liter (lub innego rodzaju symboli), po którym występują i który poprzedzają inne litery, albo żadna.

Niech w n-elementowej próbce będzie n_1 elementów a oraz $n_2 = n - n_1$ elementów b. Prawdopodobieństwo tego, że elementy tej próbki ułożą się w u serii wynosi (dowód - patrz Fisz § 11.3 [3])

$$(1) \qquad P(U = 2k | N_1 = n_1, N_2 = n_2) = 2\frac{\binom{n_1 - 1}{k - 1}\binom{n_2 - 1}{k - 1}}{\binom{n}{n_1}}$$

dla u parzystych

oraz

$$P(U = 2k+1 \mid N_1 = n_1,\ N_2 = n_2) = \frac{\binom{n_1 - 1}{k}\binom{n_2 - 1}{k - 1} + \binom{n_1 - 1}{k - 1}\binom{n_2 - 1}{k}}{\binom{n}{n_1}}$$

dla u nieparzystych, gdzie k = 0, 1, 2, ...

Hipotezę o losowości próbki odrzucamy na poziomie istotności α, gdy ilość serii w danej próbce nie przekracza ilości $u_{\frac{\alpha}{2}}$ serii lub przekracza ilość $u_{1-\frac{\alpha}{2}}$ serii, gdzie

$$P(u \leqslant u_{\frac{\alpha}{2}}) \leqslant \frac{\alpha}{2} \quad \text{zaś} \quad P(u \leqslant u_{\frac{\alpha}{2}} + 1) > \frac{\alpha}{2}$$

oraz

$$P(u \geqslant u_{1-\frac{\alpha}{2}}) \leqslant \frac{\alpha}{2} \quad \text{zaś} \quad P(u \geqslant u_{1-\frac{\alpha}{2}} - 1) > \frac{\alpha}{2}.$$

W tablicy 21 u Sadowskiego [18] podane są wartości $u_{\frac{\alpha}{2}}$ i $u_{1-\frac{\alpha}{2}}$ dla różnych n_1 i n_2 (= 1, 2, ..., 20) i α (= 0,05, 0,1).

<u>Przykład</u> (zaczerpnięty z książki Freunda [5]). Na punkcie kontrolnym dokonano pomiarów prędkości 25 samochodów osobowych. Uzyskano w kolejności pomiaru następujące dane (w milach/godz.): 55, 57, 52, 46, 50, 48, 45, 44, 50, 52, 55, 41, 42, 58, 60, 45, 53, 54, 48, 46, 51, 49, 44, 43, 56.

H: Przejeżdżające samochody były wybierane do kontroli prędkości w sposób losowy.

Zweryfikujemy tę hipotezę na poziomie $\alpha = 0,05$. Wyznaczamy medianę próbkową Me = 50. Odrzucamy elementy próbki równe medianie. Elementy próbki większe od mediany oznaczamy przez a, pozostałe przez b. Otrzymujemy ciąg:

aaabbbbaabbaabaabbabbba

Mamy: $n_1 = 11$, $n_2 = 12$, $u = 11$.

Z tablicy 21 [18] znajdujemy dla $n_1 = 11$, $n_2 = 12$ (test dwustronny) $P(u \leqslant 7) = 0,025$ i $P(u \geqslant 17) = 0,025$. Ponieważ $7 < 11 < 17$, więc nie ma powodu do odrzucenia hipotezy o losowości próbki.

<u>Przypadek dużej próbki</u>. Jeśli n_1 lub n_2 jest większe od 20, nie możemy użyć tablicy 21 [18]. Gdy $n_1 \to \infty$ i $n_2 = kn_1$, gdzie k jest stałą liczbą dodatnią, ilość serii ma rozkład asymptotycznie normalny $N\left(\frac{2n_1 n_2}{n} + 1,\ \sqrt{\frac{2n_1 n_2(2n_1 n_2 - n)}{n^2(n-1)}}\right)$ (Wald i Wolfowitz [21]).

Statystyka

$$Z = \frac{U - \left(\frac{2n_1n_2}{n} + 1\right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}}} \tag{2}$$

ma rozkład asymptotycznie normalny $N(0, 1)$. Odpowiednie prawdopodobieństwa podaje tablica 2 [18].

Test znaków dla par

Dla porównania dwu warunków, zabiegów czy doświadczeń używa się często próbek łączonych w pary. Z populacji wybiera się n par elementów tak, żeby elementy tworzące jedną parę były możliwie podobne. Następnie po jednym elemencie z każdej pary, które stanowić będą pierwszą próbkę n-elementową, poddajemy działaniu jednego warunku, pozostałe elementy, które tworzą drugą próbkę n-elementową, działaniu drugiego warunku.

Niech X będzie zmienną losową, która ma taki rozkład prawdopodobieństwa, jak rozkład badanej cechy u elementów poddanych działaniu pierwszego warunku, a Y niech będzie niezależną od X zmienną losową, która ma taki rozkład prawdopodobieństwa, jak rozkład badanej cechy u elementów poddanych działaniu drugiego warunku.

Zastosujemy test znaków do zweryfikowania hipotezy H, że dwa warunki (zabiegi, doświadczenia) nie różnią się w skutkach. Hipotezę H można wyrazić następująco:

$$P(X>Y) = P(X<Y) = \frac{1}{2}.$$

Test znaków stosujemy przy założeniu, że badane cechy mają rozkład ciągły.

Oznaczymy wartości cechy zaobserwowane u i-tej ($i = 1, 2, \ldots, \ldots, n$) pary przez x_i dla elementu tej pary poddanego działaniu pierwszego warunku oraz przez y_i dla elementu poddanego działaniu drugiego warunku.

Znajdujemy znaki różnic $x_i - y_i$ ($i = 1, 2, \ldots, n$). Przypuśćmy, że ilość różnic dodatnich jest m, wówczas różnic ujemnych jest $n - m$ (wykluczamy wypadek $x_i - y_i$). Przy prawdziwości H należy oczekiwać, że ilość różnic dodatnich będzie równa ilości różnic ujemnych, tzn. $m = n - m$. Prawdopodobieństwo pojawienia się wśród n różnic j dodatnich oraz $n - j$ ujemnych można określić za pomocą rozkładu binomialnego z parametrami $p = q = \frac{1}{2}$. Niech $u = \max(m, n-m)$, tzn. u jest równe większej z dwu liczb: m oraz $n - m$. Wówczas prawdopodobieństwo, że co najmniej u różnic będzie miało jeden znak, wyraża się wzorem

$$(3) \qquad P(u \leqslant j \leqslant n) = \sum_{j=u}^{n} \binom{n}{j} p^j q^{n-j} = \frac{\sum_{j=u}^{n} \binom{n}{j}}{2^n} .$$

Test znaków może być testem jednostronnym lub dwustronnym.

Omówimy najpierw test jednostronny. Hipotezę H odrzucamy na poziomie istotności α na korzyść hipotezy K: $P(X>Y)>\frac{1}{2}$ lub hipotezy K: $P(X>Y)<\frac{1}{2}$, jeśli (3):

$$P(u \leqslant j \leqslant n) \leqslant \alpha.$$

W przeciwnym wypadku przyjmujemy hipotezę H.

T a b l i c a 1

L.p.	Waga w funtach przed dietą	Waga w funtach po diecie	Sgn (X-Y)
1	186	175	+
2	147	144	+
3	128	125	+
4	167	167	0
5	183	182	+
6	176	177	-
7	159	154	+
8	212	203	+
9	192	187	+
10	177	169	+
11	158	158	0
12	204	197	+
13	188	182	+
14	157	160	-
15	189	181	+
16	149	151	-
17	172	169	+
18	185	184	+
19	191	187	+
20	200	195	+

Omówimy teraz test dwustronny. Hipotezę H odrzucamy na poziomie istotności α na korzyść hipotezy K: $P(X>Y) \neq \frac{1}{2}$, jeśli $P(u \leqslant j \leqslant n) \leqslant \frac{\alpha}{2}$. W przeciwnym wypadku przyjmujemy hipotezę H.

Prawdopodobieństwa (3) można znaleźć w tablicy 5 u Sadowskiego [18].

Przykład (zaczerpnięty z książki Freunda [5]). 20 Amerykanów poddano 2-tygodniowej diecie (wyniki obserwacji umieszczono w tablicy 1). Chcemy zweryfikować hipotezę H: $P(X>Y) = P(X<Y) = \frac{1}{2}$ (nieskuteczność diety), przeciw hipotezie K: $P(X>Y)>\frac{1}{2}$ (dieta wpłynęła na obniżenie wagi pacjentów; test jednostronny).

Zweryfikujemy tę hipotezę na poziomie istotności $\alpha = 0,01$.

Odrzucamy te pary elementów, dla których różnica wyniosła zero. Mamy więc $n = 18$ oraz $u = \max(15, 3) = 15$. Ze wzoru (3) znajdujemy w tablicy 5 [18].

$$P(15 \leqslant j \leqslant 18) = 0,0037 < 0,01.$$

Odrzucamy więc hipotezę H i stwierdzamy, że dieta wpłynęła na obniżenie wagi pacjentów.

Przypadek dużej próbki. Tablica 5 [18] staje się bezużyteczna, jeżeli $n > 20$. Korzystając z twierdzenia Moivre'a-Laplace'a można w wypadku dużej próbki skorzystać z przybliżenia rozkładu dwumianowego rozkładem normalnym N $(np, \sqrt{npq})$.

Statystyka

$$Z = \frac{U - np}{\sqrt{npq}} = \frac{U - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} \tag{4}$$

ma rozkład asymptotycznie normalny N(0, 1).

Odpowiednich prawdopodobieństw szukamy w tablicy 2 [18].

Test Wilcoxona dla par

Test Wilcoxona stosuje się dla próbek łączonych w pary przy założeniu ciągłości badanych cech. Hipotezę H formułuje się jak dla testu znaków

$$H: \; P(X > Y) = P(X < Y) = \frac{1}{2}.$$

W odróżnieniu od testu znaków uwzględnia on nie tylko znaki badanych różnic $d_i = x_i - y_i$ (wyłączamy przypadek $x_i = y_i$), ale i ich wartości liczbowe. W teście Wilcoxona wykorzystuje się więcej informacji zawartych w próbce.

Tworzymy niemalejący n-elementowy ciąg bezwzględnych wartości różnic d_i. Przyporządkowujemy k-temu ($k = 1, 2, \ldots, n$) elementowi tego ciągu numer tego elementu, a więc liczbę k. Sumę numerów elementów dodatnich ciągu oznaczamy przez T_+, a sumę numerów elementów ujemnych przez T_-. Oczywiście $T_+ + T_- = \frac{n(n+1)}{2}$. Niech $T = \min(T_+, T_-)$. Oznaczymy przez α poziom istotności testu.

Elementy utworzonego poprzednio ciągu bezwzględnych wartości różnic można poprzedzić znakami "+" i "-" na 2^n sposobów. Każdemu uporządkowaniu odpowiada jedna para wartości (T_+, T_-). Przy prawdziwości H każde z tych uporządkowań może się pojawić z tym samym prawdopodobieństwem.

Hipotezę H odrzucamy na korzyść hipotezy K: $P(X > Y) > \frac{1}{2}$, gdy zaobserwowana wartość sumy T_- (równa wtedy wartości statystyki T) odpowiada jednemu z $E[\alpha \cdot 2^n]$ ($E[k]$ oznącza tu część całkowitą liczby k) uporządkowań, dla których sumy T_- są najmniejsze.

Hipotezę H odrzucamy na korzyść hipotezy K: $P(X > Y) < \frac{1}{2}$, gdy zaobserwowana wartość sumy T_+ (równa wtedy wartości statystyki T) odpowiada jednemu z $E[\alpha \cdot 2^n]$ uporządkowań, dla których sumy T_+ są najmniejsze.

Hipotezę H odrzucamy na korzyść hipotezy K: $P(X>Y) \neq \frac{1}{2}$, gdy zaobserwowana wartość sumy T_+ odpowiada jednemu z $E[\alpha \cdot 2^{n-1}]$ uporządkowań, dla których sumy T_+ są najmniejsze lub gdy zaobserwowana wartość sumy T_- odpowiada jednemu z $E[\alpha \cdot 2^{n-1}]$ uporządkowań, dla których sumy T_- są najmniejsze.

Tablica I na końcu pracy podaje dla różnych n (= 6, 7, ..., ..., 25) i α (= 0,025, 0,01, 0,005 dla testu jednostronnego i 0,05, 0,02, 0,01 dla testu dwustronnego) takie wartości T_o, że jeżeli $T \leqslant T_o$, hipotezę H należy odrzucić na korzyść hipotezy K.

T a b l i c a 2

L.p. pary	Aktywność społeczna dziecka z domu małego dziecka	Aktywność społeczna dziecka w domu	D = X - Y
1	82	63	19
2	69	42	27
3	73	74	- 1
4	43	37	6
5	58	51	7
6	56	43	13
7	76	80	- 4
8	65	62	3

Przykład (zaczerpnięty z książki Siegela [19], rozdz. 5). Za pomocą pewnej skali zmierzono aktywność społeczną 8 par bliźniąt. W każdej z tych par jedno dziecko przebywa w domu małego dziecka, drugie - w domu. Wyniki obserwacji umieszczono w tablicy nr 2.

Na poziomie istotności $\alpha = 0{,}05$ weryfikujemy hipotezę H, że aktywność społeczna dzieci z domów małego dziecka nie różni się od stopnia aktywności społecznej dzieci pozostających w domach. Hipoteza K niech będzie zaprzeczeniem hipotezy H (test dwustronny).

1	2	3	4	5	6	7	8
-1	3	-4	6	7	13	19	27

Mamy: $n = 8$, $T_+ = 2 + 4 + 5 + 6 + 7 + 8 = 32$, $T_- = 1 + 3 = 4$,

$$T = \min(T_+, T_-) = \min(32, 4) = 4.$$

Z tablicy I znajdujemy $T_o = 4$. Ponieważ $T = 4 = T_o$, więc odrzucamy hipotezę H.

Przypadek dużej próbki. Można pokazać (Wilcoxon [23]), że gdy $n \longrightarrow \infty$, statystyka T ma rozkład asymptotycznie normalny

$$N\left(\frac{n(n+1)}{4}, \sqrt{\frac{n(n+1)(2n+1)}{24}}\right).$$

Statystyka

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \tag{5}$$

ma rozkład asymptotycznie normalny N(0, 1).

Test Walsha

Test Walsha stosuje się do próbek łączonych w pary, pod warunkiem, że zmienne losowe X i Y z danych warunków mają rozkład równocześnie ciągły i symetryczny.

Obliczamy różnice $x_i - y_i$ $(i = 1, 2, \ldots, n)$ dla każdej z n par. Niech d_1 oznacza najmniejszą spośród n różnic, d_2 - następną najmniejszą różnicę itd. Otrzymujemy ciąg $d_1 \leqslant d_2 \leqslant \ldots \leqslant d_n$.

Weryfikujemy hipotezę H, że zmienne losowe X i Y mają jednakową medianę i - co wynika z warunku symetrii rozkładów - jednakową wartość oczekiwaną. Inaczej mówiąc, zmienna losowa X - Y ma wartość oczekiwaną $\mu = 0$. Hipoteza K dla testu dwustronnego ma postać: $\mu \neq 0$, zaś dla testu jednostronnego: $\mu > 0$ albo $\mu < 0$.

Tablica II na końcu pracy informuje, kiedy należy odrzucić hipotezę H.

Przykład (zaczerpnięty z książki Siegela [19], rozdz. 5). 15 osób nauczono 10 nie mających sensu sylab. 5 sylab wybranych losowo spośród 10 uczono w chwili, gdy osoby poddawano wstrząsom elektrycznym. Po 48 godzinach osoby te miały powtórzyć zapamiętane sylaby. Wyniki obserwacji umieszczono w tablicy 3.

H: Obie grupy sylab zostały zapamiętane w równym stopniu ($\mu = 0$).

K: Zapamiętana liczba sylab "bezwstrząsowych" jest większa od zapamiętanej liczby sylab "wstrząsowych" ($\mu > 0$, test jednostronny).

T a b l i c a 3

L.p.	Liczba zapamiętanych sylab bezwstrząsowych	Liczba zapamiętanych sylab wstrząsowych	Różnica liczb zapamiętanych sylab
1	5	2	$3 = d_{12}$
2	4	2	$2 = d_9$
3	3	0	$3 = d_{13}$
4	5	3	$2 = d_{10}$
5	2	3	$-1 = d_1$
6	4	2	$2 = d_{11}$
7	2	3	$-1 = d_2$
8	2	1	$1 = d_6$
9	4	1	$3 = d_{14}$
10	4	3	$1 = d_7$
11	3	4	$-1 = d_3$
12	1	2	$-1 = d_4$
13	5	2	$3 = d_{15}$
14	3	4	$-1 = d_5$
15	1	0	$1 = d_8$

Test przeprowadzimy na poziomie $\alpha = 0{,}05$. Z tablicy II znajdujemy, że hipotezę H należy odrzucić, jeśli

$$\min\left[\tfrac{1}{2}(d_1 + d_{12}),\ \tfrac{1}{2}(d_2 + d_{11})\right] > 0 .$$

W naszym przykładzie $d_1 = -1$, $d_2 = -1$, $d_{11} = 2$, $d_{12} = 3$ i $\min\left[\frac{1}{2}(-1 + 3),\ \frac{1}{2}(-1 + 2)\right] = \min\left[1, \frac{1}{2}\right] = \frac{1}{2} > 0$, przyjmujemy więc hipotezę K.

Test zrandomizowany dla próbek łączonych w pary

Załóżmy, że zmienne losowe X i Y mają rozkład ciągły. Obliczmy różnice $d_i = x_i - y_i$ $(i = 1, 2, \ldots, n)$. Zweryfikujemy hipotezę H, że zmienne losowe X i Y mają jednakowe rozkłady, tzn. warunki, którym poddano elementy pierwszej i drugiej próbki, są równoważne. Elementy ciągu $\{|d_i|\}$ $(i = 1, 2, \ldots, n)$ mogą być poprzedzone znakami "+" i "-" na 2^n sposobów. Każdemu z tych uporządkowań odpowiada pewna wartość sumy $\sum_{i=1}^{n} d_i$. Przy prawdziwości H każde spośród 2^n uporządkowań ma jednakowe prawdopodobieństwo pojawienia się.

Mówimy, że zmienna losowa X jest stochastycznie większa niż zmienna losowa Y, jeżeli dla każdej liczby k prawdopodobieństwo zaobserwowania wartości zmiennej losowej X, które są mniejsze od k, nie przekracza prawdopodobieństwa zaobserwowania mniejszych od k wartości zmiennej losowej Y.

Niech α oznacza poziom istotności testu.

Hipotezę H odrzucamy na korzyść hipotezy K, że zmienna losowa X jest stochastycznie większa niż zmienna losowa Y, gdy zaobserwowana w próbce wartość sumy $\sum_{i=1}^{n} d_i$ jest jedną z sum, które odpowiadają ilości $E\left[\alpha \cdot 2^n\right]$ uporządkowań, dla których sumy $\sum_{i=1}^{n} d_i$ przyjmują największe wartości dodatnie (test jednostronny).

Hipotezę H odrzucamy na korzyść hipotezy K, że zmienna losowa X jest stochastycznie mniejsza niż zmienna losowa Y, gdy zaobserwowana w próbce wartość sumy $\sum_{i=1}^{n} d_i$ jest jedną z sum, które odpowiadają ilości $E\left[\alpha \cdot 2^n\right]$ uporządkowań, dla których sumy $\sum_{i=1}^{n} d_i$ przyjmują najmniejsze wartości ujemne (test jednostronny).

Hipotezę H odrzucamy na korzyść hipotezy K, że zmienne losowe X i Y mają różne rozkłady prawdopodobieństw, gdy zaobserwowana w próbce wartość sumy $\sum_{i=1}^{n} d_i$ jest jedną z sum, które odpowiadają ilości $E[\alpha \cdot 2^{n-1}]$ uporządkowań, dla których sumy $\sum_{i=1}^{n} d_i$ przyjmują największe wartości dodatnie lub gdy zaobserwowana wartość sumy $\sum_{i=1}^{n} d_i$ jest jedną z sum, które odpowiadają ilości $E[\alpha \cdot 2^{n-1}]$ uporządkowań, dla których sumy $\sum_{i=1}^{n} d_i$ przyjmują najmniejsze wartości ujemne (test dwustronny).

Przykład. Zweryfikujemy na poziomie $\alpha = 0,05$ hipotezę przedstawioną na str. 10.

Zaobserwowany ciąg $\{d_i\}$ wygląda następująco:

$$+19, \ +27, \ -1, \ +6, \ +7, \ +13, \ -4, \ +3, \ \text{więc} \ \sum_{i=1}^{8} d_i = +70.$$

Ilość możliwych permutacji znaków wynosi $2^8 = 256$. Test jest dwustronny. Hipotezę H odrzucimy, jeśli $\sum_{i=1}^{8} d_i = +70$ jest jedną z sum, które odpowiadają $E[\alpha \cdot 2^{n-1}] = E[0,05 \cdot 128] = E[6,4] = 6$ uporządkowaniom, dla których sumy $\sum_{i=1}^{8} d_i$ przyjmują największe wartości dodatnie lub jeśli zaobserwowana suma odpowiada jednemu z 6 uporządkowań, dla których sumy $\sum_{i=1}^{8} d_i$ przyjmują najmniejsze wartości ujemne.

W tablicy 4 znaleźć można te uporządkowania i odpowiadające im wartości sumy.

Ponieważ zaobserwowana wartość $\sum_{i=1}^{8} d_i = +70$ jest jedną z wartości w tablicy 4, przeto hipotezę H odrzucamy na korzyść hipotezy K.

Tablica 4

d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	$\sum_{i=1}^{8} d_i$
+19	+27	+1	+6	+7	+13	+4	+3	+80
+19	+27	-1	+6	+7	+13	+4	+3	+78
+19	+27	+1	+6	+7	+13	+4	-3	+74
+19	+27	+1	+6	+7	+13	-4	+3	+72
+19	+27	-1	+6	+7	+13	+4	-3	+72
+19	+27	-1	+6	+7	+13	-4	+3	+70
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	
-19	-27	-1	-6	-7	-13	-4	-3	-80
-19	-27	+1	-6	-7	-13	-4	-3	-78
-19	-27	-1	-6	-7	-13	-4	+3	-74
-19	-27	-1	-6	-7	-13	+4	-3	-72
-19	-27	+1	-6	-7	-13	-4	+3	-72
-19	-27	+1	-6	-7	-13	+4	-3	-70

Przypadek dużej próbki. Jeśli np. $n = 13$, to $2^{13} = 8192$, a np. $0{,}05 \cdot 8192 = 409{,}6$ i obliczenie 409 sum staje się uciążliwe.

Dla dużych liczebności n stosujemy test Wilcoxona.

Moses [13] udowodnił, że jeśli $n > 25$ oraz spełniony jest warunek

$$\frac{d_{max}^2}{\sum_{i=1}^{n} d_i^2} \leqslant \frac{5}{2n}, \quad \text{gdzie} \quad d_{max} = \max(d_1, d_2, \ldots, d_n), \tag{6}$$

to suma $\sum_{i=1}^{n} d_i$ ma rozkład asymptotycznie normalny $N\left(0, \sqrt{\sum_{i=1}^{n} d_i^2}\right)$.

Statystyka

$$Z = \frac{\sum_{i=1}^{n} d_i}{\sqrt{\sum_{i=1}^{n} d_i^2}} \tag{7}$$

ma rozkład asymptotycznie normalny N(0, 1), jeśli zachodzi (6).

Testy nieparametryczne dla hipotez o parametrach rozkładu

Hipotezy o parametrach rozkładu badanej cechy weryfikuje się na podstawie jednej próbki. Testy nieparametryczne takich hipotez można otrzymać przez zmodyfikowanie opisanych wyżej testów dla dwu próbek łączonych w pary. Gdy dane są dwie próbki łączone w pary, każda o liczebności n, to statystyki używane w tych testach są funkcjami n różnic wartości badanej cechy u odpowiednich par elementów. Różnice te można traktować jako wartości innej cechy w jednej próbce o liczebności n. Tak więc za pomocą tej samej statystyki można wtedy weryfikować inne hipotezy.

Za pomocą testu znaków można weryfikować hipotezę, że mediana badanej cechy jest równa pewnej liczbie k. Należy policzyć elementy próbki większe od tej liczby i z ich ilości wnioskować o tej hipotezie, tak jak z ilości różnic dodatnich w teście znaków dla próbek łączonych w pary.

Na podstawie jednej próbki za pomocą testu znaków można też zweryfikować hipotezę, że kwantyl rzędu p badanej cechy jest równy pewnej liczbie r. Należy policzyć elementy próbki większe od r. Ilość takich elementów ma rozkład binomialny z prawdopodobieństwem porażki równym p. Wartości krytyczne wyznacza się z tego rozkładu.

Hipotezę, że mediana badanej cechy wynosi k, można też weryfikować za pomocą testu Wilcoxona, testu Walsha i testu zrandomizowanego. Posługiwanie się tymi testami wymaga wtedy założenia, że rozkład badanej cechy jest symetryczny. Tym samym hipoteza ta staje się hipotezą o wartości oczekiwanej tej cechy. Od wartości zaobserwowanych w próbce należy odjąć liczbę k. Otrzymamy n różnic $d_i = x_i - k$ $(i = 1, 2, \ldots, n)$. Dalej postępujemy tak samo, jak z różnicami d_i dla próbek łączonych w pary.

Przykład. Badanie pewnego gatunku benzyny ze względu na ilość mil, które można przejechać spalając 1 galon tego paliwa, dało następującą próbkę:

17,0 17,8 15,3 16,8 18,4 16,3 18,3 18,1 17,3

Na poziomie istotności $\alpha = 0{,}01$ zweryfikujemy hipotezę H, że rozkład ilości mil ma medianę równą 16,5 przeciw hipotezie K, że mediana jest większa niż 16,5. Zastosujemy tu dla ilustracji kolejno test znaków, test Wilcoxona, test Walsha i test zrandomizowany.

Test znaków. Mamy $n = 9$. W próbce 7 elementów jest większych od 16,5. Stąd $u = \max(7, 2) = 7$. Ze wzoru (3) w tablicy 5 [18] znajdujemy

$$P(7 \leqslant j \leqslant 9) = 0{,}0898 > 0{,}01.$$

Nie ma więc powodów do odrzucenia hipotezy H.

Test Wilcoxona. Obliczmy różnice $d_i = x_i - k$. Mamy:

0,5 1,3 -1,2 0,3 1,9 -0,2 1,8 1,6 0,8

Tworzymy niemalejący ciąg bezwzględnych wartości tych różnic:

1	2	3	4	5	6	7	8	9
-0,2	0,3	0,5	0,8	-1,2	1,3	1,6	1,8	1,9

Mamy $T_+ = 2 + 3 + 4 + 6 + 7 + 8 + 9 = 39$, $T_- = 1 + 5 = 6$,
$T = \min(T_+, T_-) = \min(39, 6) = 6$.

Z tablicy I znajdujemy $T_0 = 3$. Ponieważ $T = 6 > T_0 = 3$, nie odrzucamy hipotezy H.

Test Walsha. W teście tym przez d_1 oznacza się najmniejszą różnicę, przez d_2 - następną najmniejszą różnicę, wreszcie przez d_n - największą spośród różnic. Dla danej próbki mamy więc

$d_1 = -1{,}2$ $d_4 = 0{,}5$ $d_7 = 1{,}6$
$d_2 = -0{,}2$ $d_5 = 0{,}8$ $d_8 = 1{,}8$
$d_3 = 0{,}3$ $d_6 = 1{,}3$ $d_9 = 1{,}9$

Z tablicy II znajdujemy, że hipotezę H należy odrzucić, jeśli

$$\min\left[d_2, \tfrac{1}{2}(d_1 + d_5)\right] > 0.$$

Mamy $d_1 = -1{,}2$, $d_2 = -0{,}2$, $d_5 = 0{,}8$ i $\min\left[-0{,}2,\ \tfrac{1}{2}(-1{,}2 + 0{,}8)\right] = \min\left[-0{,}2;\ -0{,}2\right] = -0{,}2 < 0$, nie odrzucamy więc hipotezy H.

Test zrandomizowany. Obliczamy najpierw $\sum_{i=1}^{9} d_i$. Suma ta wynosi +6,8. Hipotezę H odrzucimy na korzyść hipotezy K, gdy $\sum_{i=1}^{9} d_i = +6{,}8$ odpowiada jakiemuś spośród $E[\alpha \cdot 2^n] = E[0{,}01 \cdot 2^9] = E[0{,}01 \cdot 512] = 5$ uporządkowań, dla których sumy $\sum_{i=1}^{9} d_i$ przybierają największe wartości dodatnie. Takie uporządkowania i odpowiadające im sumy znajdują się w tablicy 5.

Zaobserwowana wartość $\sum_{i=1}^{9} d_i = +6,8$ nie jest żadną wartością z tablicy 5, dlatego przyjmujemy hipotezę H.

T a b l i c a 5

d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	$\sum_{i=1}^{9} d_i$
+0,5	+1,3	+1,2	+0,3	+1,9	+0,2	+1,8	+1,6	+0,8	+9,6
+0,5	+1,3	+1,2	+0,3	+1,9	-0,2	+1,8	+1,6	+0,8	+9,2
+0,5	+1,3	+1,2	-0,3	+1,9	+0,2	+1,8	+1,6	+0,8	+9,0
-0,5	+1,3	+1,2	+0,3	+1,9	+0,2	+1,8	+1,6	+0,8	+8,6
+0,5	+1,3	+1,2	-0,3	+1,9	-0,2	+1,8	+1,6	+0,8	+8,6

Test X

Przedstawimy teraz kilka testów stosowanych w wypadkach, gdy mamy dwie próbki niezależne. Zacznijmy od testu X Van der Waerdena.

Zakładamy, że badane zmienne losowe X i Y mają rozkład ciągły.

Niech pierwsza próbka $(x_1, x_2, \ldots, x_{n_1})$ liczy n_1 elementów, druga $(y_1, y_2, \ldots, y_{n_2})$ - n_2 elementów i niech $n = n_1 + n_2$.

Ustawmy obie próbki w n-elementowy ciąg rosnący i niech r_i $(i = 1, 2, \ldots, n_1)$ oznacza numer elementu x_i w tym ciągu, a s_j $(j = 1, 2, \ldots, n_2)$ - numer elementu y_j w tym ciągu.

Niech $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{z^2}{2}} dz$. Niech $\Psi(t)$ $(0<t<1)$ będzie funkcją odwrotną do funkcji $\Phi(u)$. Mamy więc

$$\Psi\big(\Phi(u)\big) = u.$$

Utwórzmy sumy

$$X = \sum_{i=1}^{n_1} \Psi\left(\frac{r_i}{n+1}\right), \tag{8}$$

$$(9) \qquad Y = \sum_{j=1}^{n_2} \Psi\left(\frac{s_j}{n+1}\right).$$

Wartości liczbowe sum (8) i (9) oblicza się za pomocą tablicy 2 [18].

Łatwo zauważyć, że $X + Y = 0$.

Hipotezę H, że obie próbki pochodzą z populacji o tym samym rozkładzie badanej cechy odrzucamy na danym poziomie istotności α na korzyść hipotezy K, że próbki pochodzą z populacji o różnych rozkładach, jeśli $|X|$ przekracza wartość podaną w tablicy III na końcu pracy (test dwustronny).

Hipotezę H odrzucamy na danym poziomie istotności α na korzyść hipotezy K, że pierwsza próbka pochodzi z populacji, w której cecha jest stochastycznie większa niż w drugiej próbce jeśli zaobserwowana wielkość X przekracza wartości z tablicy III (test jednostronny).

Gdy hipoteza K mówi, że druga próbka pochodzi z populacji, w której cecha jest stochastycznie większa, posługujemy się statystyką Y.

Przykład (zaczerpnięty z książki Van der Waerdena [20], § 65). W pewnej fabryce czas przestojów maszyn ilustruje próbka:

$$x_1 = 11, \quad x_2 = 34, \quad x_3 = 13, \quad x_4 = 18.$$

Po reorganizacji toku produkcji pobrano drugą próbkę:

$$y_1 = 8, \quad y_2 = 10, \quad y_3 = 7, \quad y_4 = 6.$$

Zastosujemy test dwustronny na poziomie $\alpha = 0{,}05$. Mamy:

$$n_1 = 4, \quad n_2 = 4, \quad n = 8.$$

$$X = \Psi\left(\frac{5}{9}\right) + \Psi\left(\frac{8}{9}\right) + \Psi\left(\frac{6}{9}\right) + \Psi\left(\frac{7}{9}\right) = 0{,}14 + 1{,}22 + 0{,}43 + 0{,}76 = 2{,}55;$$

$$Y = -X = -2{,}55.$$

W tablicy III znajdujemy wartość $X_{0,05} = 2{,}4$.

Ponieważ $2{,}55 > 2{,}40 = X_{0,05}$, odrzucamy hipotezę H i stwierdzamy, że rozkład czasu przestojów przed i po reorganizacji toku produkcji zmienił się.

Test U Manna-Whitheya-Wilcoxona

Zakładając ciągłość badanej cechy można za pomocą testu U zweryfikować hipotezę H, że dwie próbki niezależne pochodzą z populacji, w których badana cecha ma ten sam rozkład prawdopodobieństwa.

Niech n_1 będzie liczebnością mniejszej próbki, n_2 - liczebnością większej próbki ($n_1 \leqslant n_2$) i niech $n = n_1 + n_2$. Utwórzmy n-elementowy ciąg rosnący. Przyporządkowujemy k-temu ($k = 1, 2, \ldots, \ldots, n$) elementowi tego ciągu numer tego elementu, a więc liczbę k. Niech R_1 oznacza sumę numerów elementów pochodzących z próbki o liczebności n_1, a R_2 - sumę numerów elementów z próbki o liczebności n_2.

Określamy statystykę $U = \min(U_1, U_2)$, gdzie

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 , \tag{10}$$
$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 .$$

Łatwo pokazać, że

$$U_1 + U_2 = n_1 n_2. \tag{11}$$

Spośród $n_1 + n_2$ elementów można na $\binom{n_1+n_2}{n_1}$ sposobów wybrać n_1 elementów i traktować je jako pochodzące z pierwszej próbki, a pozostałe n_2 elementów zaliczyć do drugiej. Każdej kombinacji przyporządkowana jest para wartości (U_1, U_2). Przy prawdziwości H każda z tych kombinacji może pojawić się z tym samym prawdopodobieństwem.

Hipotezę H odrzucamy na poziomie istotności α na korzyść hipotezy K, że pierwsza próbka (n_1) pochodzi z populacji, w której badana cecha jest stochastycznie większa niż w drugiej, jeśli zaobserwowana wartość U_1 odpowiada jednej z $E\left[\alpha\binom{n_1+n_2}{n_1}\right]$ kombinacji, dla których U_1 przybiera najmniejsze wartości.

Hipotezę H odrzucamy na poziomie istotności α na korzyść hipotezy K, że pierwsza próbka pochodzi z populacji, w której badana cecha jest stochastycznie mniejsza niż w drugiej, jeśli zaobserwowana wartość U_2 odpowiada jednej z $E\left[\alpha\binom{n_1 + n_2}{n_1}\right]$ kombinacji, dla których U_2 przybiera najmniejsze wartości.

Hipotezę H odrzucamy na poziomie istotności α na korzyść hipotezy K, że próbki pochodzą z populacji o różnych rozkładach, jeśli zaobserwowana wartość U_1 odpowiada jednej z $E\left[\frac{\alpha}{2}\binom{n_1 + n_2}{n_1}\right]$ kombinacji, dla których U_1 przybiera najmniejsze wartości lub gdy zaobserwowana wartość U_2 odpowiada jednej z $E\left[\frac{\alpha}{2}\binom{n_1 + n_2}{n_1}\right]$ kombinacji, dla których U_2 przybiera najmniejsze wartości.

Tablica IV na końcu pracy podaje takie wartości U_α dla różnych n_1 (= 1, 2, ..., 20) i n_2 (= 9, 10, ..., 20), że jeżeli zaobserwowana wartość $U = \min(U_1, U_2)$ nie przekracza wartości z tablicy, to hipotezę H należy odrzucić na poziomie $\alpha = 0{,}025$ dla testu jednostronnego i $\alpha = 0{,}05$ dla testu dwustronnego.

Przykład (zaczerpnięty z książki Freunda [5]). Wydajność (w buszlach z akra) dwu odmian A i B kukurydzy amerykańskiej ilustrują próbki:

A:	83,4	91,5	86,4	86,9	90,3	88,6	92,7	84,8	92,8	
B:	84,1	97,3	95,2	88,9	103,2	89,0	99,7	101,4	87,2	91,0

H. Wydajność obu odmian jest jednakowa.

K: Odmiany A i B różnią się wydajnością (test dwustronny).

Mamy: $n_1 = 9$, $n_2 = 10$ i niech $\alpha = 0{,}05$.

1	2	3	4	5	6	7	8	9	10
83,4	84,1	84,8	86,4	86,9	87,2	88,6	88,9	89,o	90,3
A	B	A	A	A	B	A	B	B	A

11	12	13	14	15	16	17	18	19
91,0	91,5	92,7	92,8	95,2	97,3	99,7	101,4	103,2
B	A	A	A	B	B	B	B	B

$R_1 = 1 + 3 + 4 + 5 + 7 + 10 + 12 + 13 + 14 = 69$

$R_2 = 2 + 6 + 8 + 9 + 11 + 15 + 16 + 17 + 18 + 19 = 121,$

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 9 \cdot 10 + \frac{9 \cdot 10}{2} - 69 = 66,$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = 9 \cdot 10 + \frac{10 \cdot 11}{2} - 121 = 24,$$

$$U = \min(U_1, U_2) = \min(66, 24) = 24$$

W tablicy IV znajdujemy dla $n_1 = 9$, $n_2 = 10$ i $\alpha = 0{,}05$ wartość $U_\alpha = 20$, nie ma więc powodów dla odrzucenia hipotezy H.

Przypadek dużej próbki. Mann i Whitney [12] pokazali, że gdy $n_1 \to \infty$ i $n_2 \to \infty$, statystyka U ma rozkład asymptotycznie normalny $N\left(\frac{n_1 n_2}{2}, \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}\right)$.

Statystyka

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}} \tag{12}$$

ma rozkład asymptotycznie normalny $N(0, 1)$.

Test Walda - Wolfowitza

Test ten można stosować przy założeniu, że badana cecha jest zmienną losową o ciągłej dystrybuancie. Weryfikowana hipoteza H orzeka, że dwie próbki niezależne zostały wylosowane z populacji o jednakowym rozkładzie badanej cechy.

Niech $(x_1, x_2, \ldots, x_{n_1})$ i $(y_1, y_2, \ldots, y_{n_2})$ będą dwiema niezależnymi próbkami losowymi z dwóch populacji. Niech $n = n_1 + n_2$. Uporządkujmy wartości x_i $(i = 1, 2, \ldots, n_1)$ i y_j $(j = 1, 2, \ldots, n_2)$ według wielkości, tj. utwórzmy n-elementowy ciąg rosnący $(z_1, z_2, \ldots, z_n)$. Oznaczmy te z_k $(k = 1, 2, \ldots, n)$, które są elementami pierwszej próbki, przez a, natomiast pozostałe z_k przez b. Prawdopodobieństwo, że ciąg złożony z n_1 liter a oraz n_2 liter b ułoży się w u serii dane jest wzorami (1).

Hipotezę H odrzucamy, gdy ogólna ilość serii jest zbyt mała, albowiem wówczas występują długie serie, a więc elementy próbki tworzą duże skupienia, co świadczy o tym, że badana cecha nie ma jednakowego rozkładu w obydwu populacjach. Tak więc test polega na znalezieniu takiej wartości u_α (w tablicy 21 [18]), że $P(U \leqslant u_\alpha) \leqslant \alpha$, gdzie α jest poziomem istotności oraz na odrzuceniu weryfikowanej hipotezy, gdy zaobserwujemy $u \leqslant u_\alpha$.

Przykład (zaczerpnięty z podręcznika Perkala [16] § 9, 5). W poniższym zestawieniu podane są średnie plony żyta w województwach grupy I o liczebności $n_1 = 9$, których nazwa zaczyna się na litery A, B, ..., Ł oraz grupy II o liczebności $n_2 = 8$; na litery M, ..., Z.

I	10,2	10,5	10,6	13,2	13,5	14,2	14,3	14,5	15,4
II	11,4	11,9	13,0	13,1	13,7	13,8	14,1	15,9	

Stawiamy hipotezę H, że te dwie grupy województw nie różnią się pod względem wysokości plonu żyta ($\alpha = 0,05$).

Budujemy ciąg liter a i b

aaabbbbaabbbaaaab

Mamy u = 6 serii. Ponieważ w tablicy 21 [18] mamy $u_{0,05} = 5$, więc nie ma powodu do odrzucenia hipotezy H.

Przypadek dużej próbki. Patrz: Test o losowości próbki, str. 5 oraz wzór (2), str. 6.

Test zrandomizowany dla próbek niezależnych

Test ten służy do weryfikacji hipotezy H, że dwie próbki niezależne pochodzą z populacji, w których badana cecha ma jednakowy rozkład prawdopodobieństwa, przeciw hipotezie, że w populacjach tych rozkłady badanej cechy różnią się wartością oczekiwaną.

Niech będą dane próbki: $(x_1, x_2, \ldots, x_{n_1})$ i $(y_1, y_2, \ldots, y_{n_2})$. Mamy więc $n_1 + n_2$ obserwacji. Obliczmy sumę obserwacji w poszczególnych próbkach, czyli $\sum_{i=1}^{n_1} x_i$ oraz $\sum_{j=1}^{n_2} y_j$. Spośród $n_1 + n_2$ obserwacji można na $\binom{n_1 + n_2}{n_1}$ sposobów wybrać n_1 i traktować je jako pochodzące z pierwszej próbki, a pozostałe n_2 obserwacji zaliczyć do drugiej. Każdej kombinacji odpowiada pewna wartość różnicy $\sum_{i=1}^{n_1} x_i - \sum_{j=1}^{n_2} y_j$. Przy prawdziwości H każda z tych kombinacji ma to samo prawdopodobieństwo pojawienia się.

Jeżeli K: $\mu_1 > \mu_2$, to hipotezę H odrzucamy, gdy zaobserwowana wartość różnicy $\sum_{i=1}^{n_1} x_i - \sum_{j=1}^{n_2} y_i$ odpowiada jednej spośród $E\left[\alpha\binom{n_1+n_2}{n_1}\right]$ kombinacji, dla których różnice $\sum_{i=1}^{n_1} x_i - \sum_{j=1}^{n_2} y_j$ są największe.

Jeśli K: $\mu_1 < \mu_2$, zamieniamy kolejność próbek i postępujemy jak poprzednio.

Jeżeli K: $\mu_1 \neq \mu_2$, to hipotezę H odrzucamy, gdy zaobserwowana wartość różnicy $\sum_{i=1}^{n_1} x_i - \sum_{j=1}^{n_2} y_j$ odpowiada jednej spośród $E\left[\frac{\alpha}{2}\binom{n_1+n_2}{n_1}\right]$ kombinacji, dla których różnice $\sum_{i=1}^{n_1} x_i - \sum_{j=1}^{n_2} y_j$ są największe lub jednej spośród $E\left[\frac{\alpha}{2}\binom{n_1+n_2}{n_1}\right]$ kombinacji, dla których różnice $\sum_{i=1}^{n_1} x_i - \sum_{j=1}^{n_2} y_j$ są najmniejsze.

<u>Przykład</u> (dane fikcyjne). W dwóch wydziałach A i B pewnego przedsiębiorstwa badano ilość godzin nadliczbowych, którą każdy robotnik przepracował w danym miesiącu. W próbkach (uporządkowanych) zaobserwowano:

A: $x_1 = 16$, $x_2 = 19$, $x_3 = 22$, $x_4 = 24$, $x_5 = 29$,

B: $y_1 = 0$, $y_2 = 11$, $y_3 = 12$, $y_4 = 20$.

H: Średnia ilość godzin nadliczbowych na jednego robotnika jest jednakowa dla obydwu wydziałów.

K: Średnia ta jest większa dla oddziału A.

Niech $\alpha = 0{,}05$. Mamy: $n_1 = 5$, $n_2 = 4$, więc $\binom{n_1+n_2}{n_1} = \binom{9}{5} = 126$, i $E[0{,}05 \cdot 126] = E[6, 3] = 6$.

Obliczamy zaobserwowaną różnicę $\sum_{i=1}^{5} x_i - \sum_{j=1}^{4} y_j = 110 - 43 = 67$.

Kombinacje i odpowiadające im różnice, których zaobserwowanie powoduje odrzucenie hipotezy H, zamieszczono w tablicy 6. Ponieważ zaobserwowana wartość 67 jest jedną z wartości różnic z tej tablicy, przyjmujemy hipotezę K.

T a b l i c a 6

x_1	x_2	x_3	x_4	x_5	y_1	y_2	y_3	y_4	$\sum_{i=1}^{5} x_i - \sum_{j=1}^{4} y_j$
19	20	22	24	29	0	11	12	16	114 - 39 = 75
16	20	22	24	29	0	11	12	19	111 - 42 = 69
16	19	22	24	29	0	11	12	20	110 - 43 = 67
16	19	20	24	29	0	11	12	22	108 - 45 = 63
12	20	22	24	29	0	11	16	19	107 - 46 = 61
16	19	20	22	29	0	11	12	24	106 - 47 = 59

<u>Przypadek dużej próbki</u>. Gdy $n_1 \longrightarrow \infty$ i $n_2 \longrightarrow \infty$, statystyka

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\dfrac{\sum_{i=1}^{n_1}(x_i - \bar{x})^2 + \sum_{j=1}^{n_2}(y_j - \bar{y})^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \tag{13}$$

ma asymptotycznie rozkład normalny N (0, 1). Odpowiednie prawdopodobieństwa - patrz tablica 2 [18].

Test Friedmana dla $k>2$ zależnych próbek

Wybieramy n grup jednorodnych ze względu na pewne cechy, po k elementów w każdej grupie. Następnie w sposób losowy zaliczamy w i-tej grupie ($i = 1, 2, \ldots, n$) po jednym elemencie do pierwszej, drugiej, ..., k-tej próbki. Otrzymamy w ten sposób k próbek n-elementowych. Poddajemy pierwszą próbkę działaniu pierwszego warunku (zabiegu, doświadczenia), drugą próbkę - działaniu drugiego warunku, k-tą - działaniu k-tego warunku.

Weryfikujemy hipotezę H, że próbki po zastosowaniu tych zabiegów pochodzą z populacji o tym samym rozkładzie badanej cechy, tzn. wszystkie warunki są równoważne.

Niech x_{ij} oznacza zaobserwowaną wartość cechy i-tego elementu w j-tej ($j = 1, 2, \ldots, k$) próbce. Wyniki obserwacji przedstawmy w tablicy 7.

T a b l i c a 7

	1	2	...	...	...	k
1	x_{11}	x_{12}	...	...	...	x_{1k}
2	x_{21}	x_{22}	...	...	...	x_{2k}
...	...	...	...	...	...	...
...	...	...	...	x_{ij}	...	...
...	...	...	...	...	...	...
n	x_{n1}	x_{n2}	...	...	...	x_{nk}

Przekształcimy tablicę 7 w następujący sposób: W i-tym ($i = 1, 2, \ldots$ $\ldots, n$) wierszu w miejscu najmniejszego elementu tego wiersza wpisujemy liczbę 1, w miejsce kolejnego najmniejszego elementu tego wiersza wpisujemy liczbę 2, ..., w miejsce największego elementu i-tego wiersza - liczbę k. W tak przekształconej tablicy 7 obliczamy sumę R_j ($j = 1, 2, \ldots, k$) liczb znajdujących się w j-tej kolumnie.

Określamy statystykę

$$\chi_F^2 = \frac{12}{nk(k+1)} \cdot \sum_{j=1}^{k} R_j^2 - 3n(k+1). \tag{14}$$

Hipotezę H odrzucamy, gdy zaobserwowana wartość statystyki χ_F^2 przekracza wartość $\chi_{F\alpha}^2$, gdzie $P\left(\chi_F^2 \geqslant \chi_{F\alpha}^2\right) = \alpha$ (α - poziom istotności testu). Odpowiednie prawdopodobieństwa podane są w tablicy V na końcu pracy.

<u>Przykład</u>. (Zaczerpnięty z książki Siegela [19], rozdz. 7). Żeby zweryfikować hipotezę H, że cztery metody nauczania A, B, C, D dają te same wyniki, wybrano n = 3 grupy po k = 4 uczniów. Każda grupa składa się z dzieci w jednakowym wieku, jednakowo inteligentnych, o jednakowych dotychczasowych wynikach i warunkach nauczania, z tego samego środowiska, itp. Zaliczenie dziecka do nauczania go jedną z tych metod odbywa się w poprzednio opisany spo-

sób. Przy systemie ocen od 1 do 10 otrzymano wyniki nauczania przedstawione w tablicy 8.

T a b l i c a 8

	A	B	C	D
1	9	4	1	7
2	6	5	2	8
3	9	1	2	6

Jeśli przekształcimy tablicę 8 podobnie jak tablicę 7, otrzymamy tablicę 9.

T a b l i c a 9

	A	B	C	D
1	4	2	1	3
2	3	2	1	4
3	4	1	2	3
R_j	11	5	4	10

Mamy z (14):

$$\chi_F^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1) = \frac{12}{3\cdot 4\cdot 5}(11^2 + 5^2 + 4^2 + 10^2) - 3\cdot 3\cdot 5 = 7{,}4.$$

Na poziomie $\alpha = 0{,}05$ odrzucamy H, gdyż z tablicy V znajdujemy, że $P(\chi_F^2 \geqslant 7{,}4) < 0{,}05$.

<u>Przypadek dużej próbki</u>. Dokładnym rozkładem statystyki χ_F^2 posługujemy się tylko dla niewielkich n oraz k. Friedman [6] pokazał, że jeśli $n + k \longrightarrow \infty$, to statystyka χ_F^2 określona wzorem (14) ma asymptotycznie rozkład χ^2 z k-1 stopniami swobody (patrz: tablica 10 [18]).

Test mediany

Test mediany pozwala zweryfikować hipotezę H, że $k \geqslant 2$ niezależnych próbek (niekoniecznie o tych samych liczebnościach) pochodzi z populacji, w których mediana badanej cechy ma wspólną wartość.

Niech j-ta ($j = 1, 2, \ldots, k$) próbka liczy n_j elementów. Obliczamy wspólną dla k próbek wartość mediany.

Możemy teraz wypełnić tablicę 10.

T a b l i c a 10

Liczba elementów próbki	Numer próbki			
	1	2	...	k
O wartościach mniejszych od mediany				
O wartościach większych od mediany				

Niech l_{ij} oznacza zaobserwowaną liczbę elementów w i-tym ($i = 1, 2$) wierszu j-tej kolumny tablicy 10. Pod warunkiem H oczekiwana liczba elementów w pierwszym wierszu j-tej kolumny jest równa oczekiwanej liczbie elementów w drugim wierszu tej kolumny i wynosi $m_j = \frac{n_j}{2}$.

Statystyka

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(l_{ij} - m_j)^2}{m_j} \tag{15}$$

ma asymptotycznie rozkład χ^2 z k-1 stopniami swobody. Odpowiednie prawdopodobieństwa podaje tablica 10 [18].

Przykład (zaczerpnięty z książki Siegela [19], rozdz. 8).

Na podstawie danych z tablicy 11 zweryfikujemy na poziomie istotności $\alpha = 0{,}05$ hipotezę H, że częstość wizyt matki w szkole nie zależy od poziomu jej wykształcenia.

Obliczmy według (15)

$$\chi^2 = \sum_{k=1}^{2} \sum_{j=1}^{4} \frac{(l_{ij} - m_j)^2}{m_j} = \frac{(5-5)^2}{5} + \frac{(7-5{,}5)^2}{5{,}5} + \frac{(6-6{,}5)^2}{6{,}5} +$$

$$+ \frac{(4-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(4-5,5)^2}{5,5} + \frac{(7-6,5)^2}{5} + \frac{(6-5)^2}{5} =$$

$= 1,295.$

W tablicy 10 [18] znajdujemy dla $k-1 = 3$ stopni swobody, że $P(\chi^2 \geqslant 1,295) > 0,7 > 0,05$, nie ma więc powodu do odrzucenia hipotezy H.

Tablica 11

Liczba matek, które były w szkole	Wykształcenie matek: 7 klas	9 klas	11 klas	więcej
Rzadziej, niż wynosi wspólna mediana wizyt w szkole	5	7	6	4
Częściej, niż wynosi wspólna mediana wizyt w szkole	5	4	7	6
$m_j = \frac{n_j}{2}$	5	5,5	6,5	5

Test Kruskala-Wallisa

Za pomocą tego testu można zweryfikować hipotezę H, że $k > 2$ niezależnych próbek pochodzi z populacji, w których badana cecha jest ciągłą zmienną losową, o jednakowej wartości oczekiwanej.

Niech j-ta ($j = 1, 2, \ldots, k$) próbka liczy n_j elementów.

W sumie mamy $n = \sum_{j=1}^{k} n_j$ obserwacji. Ustawmy wszystkie obserwacje w n-elementowy ciąg rosnący. Pierwszemu elementowi tego ciągu przyporządkujmy numer tego elementu, a więc liczbę 1, drugiemu - liczbę 2, itd. Niech R_j oznacza sumę numerów tych elementów tego ciągu, które pochodzą z j-tej próbki.

W celu zweryfikowania hipotezy H, posługujemy się statystyką

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1). \tag{16}$$

Odpowiednie prawdopodobieństwa dla $k = 3$ oraz $\max_{j=1,2,3} n_j \leqslant 5$ podaje tablica VI na końcu pracy.

Przykład (opracowany na podstawie materiałów zamieszczonych w książce Perkala [16], cz. III). Zbadano plony 3 odmian pszenicy uprawianych w jednakowych warunkach. W sposób losowy wybrano 5 poletek doświadczalnych, na których rosła odmiana "rokicka", 5 poletek z odmianą "puławską" i 4 na których rosła "ostka". Plony (w gramach na m^2) podane są w tablicy 12.

T a b l i c a 12

Rokicka	Puławska twarda	Ostka chłopicka
208,5 (4)	202,0 (2)	215,5 (7)
220,1 (9)	216,9 (8)	229,4 (13)
207,0 (3)	222,1 (10)	236,3 (14)
194,2 (1)	222,5 (11)	228,9 (12)
209,7 (5)	211,0 (6)	
$R_1 = 22$	$R_2 = 37$	$R_3 = 46$

Liczby w nawiasach w tablicy 12 oznaczają numer elementu w ciągu rosnącym n = 14 obserwacji.

Zweryfikujemy na poziomie $\alpha = 0{,}05$ hipotezę H, że nie ma różnicy między średnimi plonami trzech odmian pszenicy przy danych warunkach uprawy.

Mamy z (16)

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1) =$$

$$= \frac{12}{14 \cdot 15}\left(\frac{22^2}{5} + \frac{37^2}{5} + \frac{46^2}{4}\right) - 3 \cdot 15 = 6{,}4$$

Z tablicy VI znajdujemy, że $P(H \geqslant 6{,}4) < 0{,}05$, odrzucamy więc hipotezę H.

Przypadek dużej próbki. Jeśli $n \to \infty$, statystyka H określona wzorem (16) ma asymptotycznie rozkład χ^2 z k-1 stopniami swobody (dowód patrz: Kruskall [9]). Rozkładem χ^2 posługujemy się w praktyce jako aproksymacją statystyki H, jeśli tylko $\min\limits_{j=1,2,\dots,k} n_j > 5$. Odpowiednich prawdopodobieństw należy szukać w tablicy 10 [18].

Test R_S istotności korelacji

Chcemy zbadać istotność korelacji dwóch cech X i Y o nieznanym rozkładzie. Wybieramy w tym celu z populacji n elementów w sposób losowy. Otrzymujemy 2 ciągi $(x_1, x_2, \ldots, x_n)$ i $(y_1, y_2, \ldots \ldots, y_n)$ wartości tych cech.

Niech f oznacza funkcję przekształcającą zbiór $(x_1, x_2, \ldots \ldots, x_n)$ na zbiór $(1, 2, \ldots, n)$ w ten sposób, że $f(x_i) < f(x_j)$, gdy $x_i < x_j$ $(i, j = 1, 2, \ldots, n)$. Niech g oznacza funkcję przekształcającą zbiór $(y_1, y_2, \ldots, y_n)$ na zbiór $(1, 2, \ldots, n)$ w ten sposób, że $g(y_i) < g(y_j)$, gdy $y_i < y_j$. Inaczej mówiąc, numerujemy każdy ciąg według wielkości. Niech $d_i = f(x_i) - g(y_i)$.

Wskaźnikiem korelacji uporządkowania Spearmana nazywamy statystykę

$$R_S = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n^3 - n}. \tag{17}$$

Weryfikujemy hipotezę H, że badane cechy są nieskorelowane. Niech α oznacza poziom istotności testu.

Jeśli ustalona jest numeracja według wielkości jednego ciągu, drugi ciąg można ponumerować na n! sposobów. Każdej z tych permutacji odpowiada pewna wartość wskaźnika R_S. Przy prawdziwości H wszystkie permutacje są jednakowo prawdopodobne.

Hipotezę H odrzucamy na korzyść hipotezy K, że korelacja między badanymi cechami jest istotna (test dwustronny), jeśli zaobserwowana wartość r_S odpowiada jednej z $E[\alpha \cdot n!]$ permutacji, dla których wartości wskaźniką R_S różnią się najbardziej od zera.

Hipotezę H odrzucamy na korzyść hipotezy K, że pomiędzy badanymi cechami istnieje korelacja dodatnia (ujemna)(test jednostronny), jeśli zaobserwowana wartość r_S odpowiada jednej z $E[\alpha \cdot n!]$ permutacji, dla których wartości wskaźnika R_S różnią się najmniej od wartości +1 (-1).

Tablica VII na końcu pracy podaje dla n = 4, 5, ..., 10, 12, 14, ..., 30 i α = 0,05 i 0,01 dla testu jednostronnego wartości $R_{S\alpha}$ takie, że jeśli zaobserwowana wartość r_S przekracza wartość z tablicy, hipotezę H należy odrzucić. Dla testu dwustronnego należy podwoić poziom istotności.

Przykład (zaczerpnięty z książki Perkala [16], cz. III, str. 117). Pytamy, czy ciężar bulw pod krzakiem ziemniaczanym jest skorelowany z ciężarem bulwy, z której wyrósł krzak. Stawiamy hipotezę H, że korelacja między tymi dwoma cechami krzaków jest nieistotna i zweryfikujemy tę hipotezę na poziomie $\alpha = 0,05$ przeciw hipotezie K, że pomiędzy tymi cechami istnieje korelacja dodatnia. W tablicy 13 podane są ciężary sadzeniaków i uzyskanych bulw 12 krzaków.

T a b l i c a 13

Ciężar w g		$f(x_i)$	$g(y_i)$	d_i	d_i^2
sadzeniaka x_i	bulw y_i				
47	880	7	11	-4	16
14	550	1	4	-3	9
22	540	2	3	-1	1
42	682	6	9	-3	9
38	665	4	8	-4	16
31	890	3	12	-9	81
59	480	10	2	8	64
52	655	9	7	2	4
49	550	8	5	3	9
40	475	5	1	4	16
66	700	11	10	1	1
68	570	12	6	6	36

$$\sum_{i=1}^{12} d_i^2 = 262$$

Według wzoru (17) mamy na podstawie tablicy 13

$$r_S = 1 - \frac{6 \cdot 262}{12^3 - 12} = 0,084 .$$

W tablicy VII znajdujemy, że $P(R_S \geqslant 0,084) > 0,05$, nie można więc twierdzić, że ciężar bulw pod krzakiem jest dodatnio skorelowany z ciężarem sadzeniaka.

<u>Przypadek dużej próbki</u>. Jeśli $n \geqslant 10$, hipotezę H, że korelacja między dwoma cechami jest nieistotna, można weryfikować za pomocą statystyki

$$T = R_S \sqrt{\frac{n - 2}{1 - R_S^2}}, \tag{18}$$

gdzie R_S wyraża się wzorem (17). Statystyka określona wzorem (18) ma w przybliżeniu rozkład t Studenta z n-2 stopniami swobody (patrz: Kendall [8]). Odpowiednie prawdopodobieństwa patrz: tablica 12 [18].

T a b l i c a I

Test Wilcoxona dla par

Maksymalne wartości T_o statystyki T na poziomie istotności $\alpha^{(*)}$

n	Poziom istotności testu jednostronnego		
	0,025	0,01	0,005
	Poziom istotności testu dwustronnego		
	0,05	0,02	0,01
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

(*) Sporządzona na podstawie [24].

T a b l i c a II

Test Walsha (*)

n	Poziom istotności testu		Test	
			dwustronny: przyjąć $\mu \neq 0$, jeśli max $[\dots] < 0$ lub min $[\dots] > 0$	
	jedno-stronnego	dwustronnego	jednostronny: przyjąć $\mu < 0$	jednostronny: przyjąć $\mu > 0$
8	0,043	0,086	$\max[d_6, \frac{1}{2}(d_4+d_8)] < 0$	$\min[d_3, \frac{1}{2}(d_1+d_5)] > 0$
	0,027	0,055	$\max[d_6, \frac{1}{2}(d_5+d_8)] < 0$	$\min[d_3, \frac{1}{2}(d_1+d_4)] > 0$
	0,012	0,023	$\max[d_7, \frac{1}{2}(d_6+d_8)] < 0$	$\min[d_2, \frac{1}{2}(d_1+d_3)] > 0$
	0,008	0,016	$\frac{1}{2}(d_7+d_8) < 0$	$\frac{1}{2}(d_1+d_2) > 0$
	0,004	0,008	$d_8 < 0$	$d_1 > 0$
9	0,051	0,102	$\max[d_6, \frac{1}{2}(d_4+d_9)] < 0$	$\min[d_4, \frac{1}{2}(d_1+d_6)] > 0$
	0,022	0,043	$\max[d_7, \frac{1}{2}(d_5+d_9)] < 0$	$\min[d_3, \frac{1}{2}(d_1+d_5)] > 0$
	0,010	0,020	$\max[d_8, \frac{1}{2}(d_5+d_9)] < 0$	$\min[d_2, \frac{1}{2}(d_1+d_5)] > 0$
	0,006	0,012	$\max[d_8, \frac{1}{2}(d_7+d_9)] < 0$	$\min[d_2, \frac{1}{2}(d_1+d_3)] > 0$
	0,004	0,008	$\frac{1}{2}(d_8+d_9) < 0$	$\frac{1}{2}(d_1+d_2) > 0$
10	0,056	0,111	$\max[d_6, \frac{1}{2}(d_4+d_{10})] < 0$	$\min[d_5, \frac{1}{2}(d_1+d_7)] > 0$
	0,025	0,051	$\max[d_7, \frac{1}{2}(d_5+d_{10})] < 0$	$\min[d_4, \frac{1}{2}(d_1+d_6)] > 0$
	0,011	0,021	$\max[d_8, \frac{1}{2}(d_6+d_{10})] < 0$	$\min[d_3, \frac{1}{2}(d_1+d_5)] > 0$
	0,005	0,010	$\max[d_9, \frac{1}{2}(d_6+d_{10})] < 0$	$\min[d_2, \frac{1}{2}(d_1+d_5)] > 0$
11	0,048	0,097	$\max[d_7, \frac{1}{2}(d_4+d_{11})] < 0$	$\min[d_5, \frac{1}{2}(d_1+d_8)] > 0$
	0,028	0,056	$\max[d_7, \frac{1}{2}(d_5+d_{11})] < 0$	$\min[d_5, \frac{1}{2}(d_1+d_7)] > 0$
	0,011	0,021	$\max[\frac{1}{2}(d_6+d_{11}), \frac{1}{2}(d_8+d_9)] < 0$	$\min[\frac{1}{2}(d_1+d_6), \frac{1}{2}(d_3+d_4)] > 0$
	0,005	0,011	$\max[d_9, \frac{1}{2}(d_7+d_{11})] < 0$	$\min[d_3\ \frac{1}{2}(d_1+d_5)] > 0$

Tablica II (c. d.)

n	Poziom istotności testu		Test	
			dwustronny: przyjąć $\mu \neq 0$, jeśli max $[\ldots] < 0$ lub min $[\ldots] > 0$	
	jedno-stronnego	dwustronnego	jednostronny: przyjąć $\mu < 0$	jednostronny: przyjąć $\mu < 0$
12	0,047	0,094	$\max\left[\frac{1}{2}(d_4+d_{12}),\ \frac{1}{2}(d_5+d_{11})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_9),\ \frac{1}{2}(d_2+d_8)\right] > 0$
	0,024	0,048	$\max\left[d_8,\ \frac{1}{2}(d_5+d_{12})\right] < 0$	$\min\left[d_5,\ \frac{1}{2}(d_1+d_8)\right] > 0$
	0,010	0,020	$\max\left[d_9,\ \frac{1}{2}(d_6+d_{12})\right] < 0$	$\min\left[d_4,\ \frac{1}{2}(d_1+d_7)\right] > 0$
	0,005	0,011	$\max\left[\frac{1}{2}(d_7+d_{12}),\ \frac{1}{2}(d_9+d_{10})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_6),\ \frac{1}{2}(d_3+d_4)\right] > 0$
13	0,047	0,094	$\max\left[\frac{1}{2}(d_4+d_{13}),\ \frac{1}{2}(d_5+d_{12})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_{10}),\ \frac{1}{2}(d_2+d_9)\right] > 0$
	0,023	0,047	$\max\left[\frac{1}{2}(d_5+d_{13}),\ \frac{1}{2}(d_6+d_{12})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_9),\ \frac{1}{2}(d_2+d_8)\right] > 0$
	0,010	0,020	$\max\left[\frac{1}{2}(d_6+d_{13}),\ \frac{1}{2}(d_9+d_{10})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_8),\ \frac{1}{2}(d_4+d_5)\right] > 0$
	0,005	0,010	$\max\left[d_{10},\ \frac{1}{2}(d_7+d_{13})\right] < 0$	$\min\left[d_4,\ \frac{1}{2}(d_1+d_7)\right] > 0$
14	0,047	0,094	$\max\left[\frac{1}{2}(d_4+d_{14}),\ \frac{1}{2}(d_5+d_{13})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_{11}),\ \frac{1}{2}(d_2+d_{10})\right] > 0$
	0,023	0,047	$\max\left[\frac{1}{2}(d_5+d_{14}),\ \frac{1}{2}(d_6+d_{13})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_{10}),\ \frac{1}{2}(d_2+d_9)\right] > 0$
	0,010	0,020	$\max\left[d_{10},\ \frac{1}{2}(d_6+d_{14})\right] < 0$	$\min\left[d_5,\ \frac{1}{2}(d_1+d_9)\right] > 0$
	0,005	0,010	$\max\left[\frac{1}{2}(d_7+d_{14}),\ \frac{1}{2}(d_{10}+d_{11})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_8),\ \frac{1}{2}(d_4+d_5)\right] > 0$
15	0,047	0,094	$\max\left[\frac{1}{2}(d_4+d_{15}),\ \frac{1}{2}(d_5+d_{14})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_{12}),\ \frac{1}{2}(d_2+d_{11})\right] > 0$
	0,023	0,047	$\max\left[\frac{1}{2}(d_5+d_{14}),\ \frac{1}{2}(d_6+d_{14})\right] < 0$	$\min\left[\frac{1}{2}(d_1+d_{11}),\ \frac{1}{2}(d_2+d_{10})\right] > 0$
	0,010	0,020	$\max\left[\frac{1}{2}(d_6+d_{15}),\ \frac{1}{2}(d_{10}+d_{11})\right] < 0$	$\min\left[\frac{1}{2}(d_5+d_{10}),\ \frac{1}{2}(d_5+d_6)\right] > 0$
	0,005	0,010	$\max\left[d_{11},\ \frac{1}{2}(d_7+d_{15})\right] < 0$	$\min\left[d_5,\ \frac{1}{2}(d_1+d_9)\right] > 0$

(*) Sporządzona na podstawie [22].

T a b l i c a III

Test X

Minimalne wartości X_0 statystyki X na poziomie istotności α (x)

n	Test dwustronny: $\alpha = 0,05$ Test jednostronny: $\alpha = 0,025$			n	Test dwustronny: $\alpha = 0,02$ Test jednostronny: $\alpha = 0,01$		
	$\|n_1-n_2\|=$ 0 lub 1	$\|n_1-n_2\|=$ 2 lub 3	$\|n_1-n_2\|=$ 4 lub 5		$\|n_1-n_2\|=$ 0 lub 1	$\|n_1-n_2\|=$ 2 lub 3	$\|n_1-n_2\|=$ 4 lub 5
8	2,40	2,30		8			
9	2,38	2,20		9	2,80		
10	2,60	2,43	2,30	10	3,00	2,90	2,80
11	2,72	2,58	2,40	11	3,20	3,00	2,90
12	2,86	2,79	2,68	12	3,29	3,30	3,20
13	2,96	2,91	2,78	13	3,50	3,36	3,18
14	3,11	3,06	3,00	14	3,62	3,55	3,46
15	3,24	3,19	3,06	15	2,74	3,68	3,57
16	3,39	3,36	3,28	16	3,92	3,90	3,80
17	3,49	3,44	3,36	17	4,06	4,01	3,90
18	3,63	3,60	3,53	18	4,23	4,21	4,14
19	3,73	3,69	3,61	19	4,37	4,32	4,23
20	3,86	3,84	3,78	20	4,52	4,50	4,44
21	3,96	3,92	3,85	21	4,66	4,62	4,53
22	4,07	4,06	4,01	22	4,80	4,78	4,72
23	4,18	4,15	4,08	23	4,92	4,89	4,81
24	4,29	4,27	4,23	24	5,06	5,04	4,99
25	4,39	4,36	4,30	25	5,18	5,14	5,08
26	4,50	4,48	4,44	26	5,30	5,29	5,24
27	4,59	4,56	4,51	27	5,42	5,39	5,33

Tablica III (c. d.)

	Test dwustronny : $\alpha = 0{,}05$ Test jednostronny: $\alpha = 0{,}025$				Test dwustronny : $\alpha = 0{,}02$ Test jednostronny: $\alpha = 0{,}01$		
n	$\lvert n_1-n_2\rvert =$ 0 lub 1	$\lvert n_1-n_2\rvert =$ 2 lub 3	$\lvert n_1-n_2\rvert =$ 4 lub 5	n	$\lvert n_1-n_2\rvert =$ 0 lub 1	$\lvert n_1-n_2\rvert =$ 2 lub 3	$\lvert n_1-n_2\rvert =$ 4 lub 5
28	4,69	4,68	4,64	28	5,54	5,52	5,48
29	4,78	4,76	4,72	29	5,65	5,62	5,57
30	4,88	4,87	4,84	30	5,77	5,75	5,72
31	4,97	4,95	4,91	31	5,87	5,85	5,80
32	5,07	5,06	5,03	32	5,99	5,97	5,94
33	5,15	5,13	5,10	33	6,09	6,07	6,02
34	5,25	5,24	5,21	34	6,20	6,19	6,16
35	5,33	5,31	5,28	35	6,30	6,28	6,24
36	5,42	5,41	5,38	36	6,40	6,39	6,37
37	5,50	5,48	5,45	37	6,50	6,48	6,45
38	5,59	5,58	5,55	38	6,60	6,59	6,57
39	5,67	5,65	5,62	39	6,70	6,68	6,65
40	5,75	5,74	5,72	40	6,80	6,79	6,77
41	5,83	5,81	5,79	41	6,89	6,88	6,85
42	5,91	5,90	5,88	42	6,99	6,98	6,96
43	5,99	5,97	5,95	43	7,08	7,07	7,04
44	6,06	6,06	6,04	44	7,17	7,27	7,14
45	6,14	6,12	6,10	45	7,26	7,25	7,22
46	6,21	6,21	6,19	46	7,35	7,35	7,32
47	6,29	6,27	6,25	47	7,44	7,43	7,40
48	6,36	6,35	6,34	48	7,53	7,52	7,50
49	6,43	6,42	6,39	49	7,61	7,60	7,57
50	6,50	6,50	6,48	50	7,70	7,69	7,68

(*) Przedrukowana z [20]

T a b l i c a IV

Test Manna-Whitneya-Wilcoxona

Maksymalne wartości U_0 statystyki U dla testu jednostronnego na poziomie istotności $\alpha = 0{,}025$ i dla testu dwustronnego na poziomie istotności $\alpha = 0{,}05$ (*)

n_1 \ n_2	9	10	11	12	13	14	15	16	17	18	19	20
1	0											
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	99	106	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

(*) Sporządzona na podstawie [1].

T a b l i c a V

Test Friedmana

Prawdopodobieństwa wartości niemniejszych od zaobserwowanej wartości statystyki χ_F^2 (*)

k = 3

n = 2		n = 3		n = 4		n = 5	
χ_F^2	α	χ_F^2	α	χ_F^2	α	χ_F^2	α
4	0,167	4,667	0,194	4,5	0,125	4,8	0,124
		6,0	0,028	6,0	0,069	5,2	0,093
				6,5	0,042	6,4	0,039
				8,0	0,0046	7,6	0,024
						8,4	0,0085

n = 6		n = 7		n = 8		n = 9	
χ_F^2	α	χ_F^2	α	χ_F^2	α	χ_F^2	α
4,33	0,142	4,571	0,112	4,75	0,120	4,667	0,107
5,33	0,072	5,429	0,085	5,25	0,079	6,0	0,057
6,33	0,052	6,0	0,052	6,25	0,047	6,222	0,048
7,00	0,029	7,714	0,021	7,00	0,030	6,889	0,031
8,33	0,012	8,0	0,016	7,75	0,018	8,0	0,019
9,00	0,0081	8,857	0,0084	9,00	0,0099	8,667	0,010

k = 4

n = 2		n = 3		n = 4	
χ_F^2	α	χ_F^2	α	χ_F^2	α
5,4	0,167	5,8	0,148	6,0	0,105
6,0	0,042	6,6	0,075	6,3	0,094
		7,0	0,054	7,5	0,052
		7,4	0,033	7,8	0,036
		8,2	0,017	8,1	0,033
		9,0	0,0017	8,4	0,019
				9,3	0,012
				9,6	0,0069

(*) Sporządzona na podstawie [6].

T a b l i c a VI

Test Kruskala-Wallisa

Prawdopodobieństwa wartości niemniejszych od zaobserwowanych wartości h statystyki H (*)

Liczebności próbek			H	α	Liczebności próbek			H	α
n_1	n_2	n_3			n_1	n_2	n_3		
2	2	2	4,5714	0,067	2	2	2	6,5333	0,008
								5,0400	0,056
3	2	2	4,7143	0,048					
					5	3	1	6,4000	0,012
3	3	1	5,1429	0,043				4,9600	0,048
3	3	2	6,2500	0,011	5	3	2	6,8218	0,010
			5,1389	0,061				5,2509	0,049
3	3	3	6,4889	0,011	5	3	3	7,0788	0,009
			5,6000	0,050				5,6485	0,049
4	2	1	4,8214	0,057	5	4	1	6,8400	0,011
								4,9855	0,044
4	2	2	6,0000	0,014					
			5,1250	0,052	5	4	2	7,1182	0,010
								5,2682	0,050
4	3	1	5,8333	0,021					
			5,2083	0,050	5	4	3	7,4449	0,010
								5,6308	0,050

Tablica VI (c. d.)

Liczebności próbek			H	α	Liczebności próbek			H	α
n_1	n_2	n_3			n_1	n_2	n_3		
4	3	2	6,3000	0,011					
			5,4000	0,051	5	4	4	7,7604	0,009
								5,6176	0,050
4	3	3	6,7455	0,010					
			5,7273	0,050	5	5	1	7,3091	0,009
								4,9091	0,053
4	4	1	6,6667	0,010					
			4,9667	0,048	5	5	2	7,2692	0,010
								5,2462	0,051
4	4	2	6,8727	0,011					
			5,2364	0,052	5	5	3	7,5429	0,010
								5,6264	0,051
4	4	3	7,1439	0,010					
			5,5985	0,049	5	5	4	7,7914	0,010
								5,6429	0,050
4	4	4	7,5385	0,011					
			5,6923	0,049	5	5	5	7,9800	0,010
								5,7800	0,049
5	2	1	5,0000	0,048					

(*) Sporządzona na podstawie [10].

T a b l i c a VII

Test korelacji uporządkowania

Minimalna wartość r_S wskaźnika R_S na poziomie istotności α (*)

n	Poziom istotności (test jednostronny)	
	0,05	0,01
4	1,000	
5	0,900	1,000
6	0,829	0,943
7	0,714	0,893
8	0,643	0,833
9	0,600	0,783
10	0,564	0,746
12	0,506	0,712
14	0,456	0,645
16	0,425	0,601
18	0,399	0,564
20	0,377	0,534
22	0,359	0,508
24	0,343	0,485
26	0,329	0,465
28	0,317	0,448
30	0,306	0,432

(*) Sporządzona na podstawie [14] i [15].

Literatura cytowana

[1] Auble, D., Extended tables for the Mann-Whitney statistic, Bulletin of the Institute of Educational Research of Indiana University 1 (1953), Nr 2.

[2] Elandt, R., A non-parametric test of tendency, Bull. Acad. Polon. Sci., cl. V, ser. sci. biol. 5 (1957), str. 187-190.

[3] Fisz, M., Rachunek prawdopodobieństwa i statystyka matematyczna, Warszawa 1958.

[4] Fraser, D., Non-parametric methods in statistics, New York 1957.

[5] Freund, J., Modern elementary statistics, New York 1961.

[6] Friedman, M., The use of ranks to avoid the assumption of normality implicit in the analysis of variance, Journal of the American Statistical Association 32 (1937), str. 675-701.

[7] Hájek, J., Šidák, Z., Theory of rank tests, Prague 1967.

[8] Kendall, M., Rank correlation methods, London 1948.

[9] Kruskal, W., A non-parametric test for the several sample problem, Annals of Mathematical Statistics 23 (1952), str. 525 -540.

[10] Kruskal, W., Wallis, W., Use of ranks in one-criterion variance analysis, J. Amer, Statist. Ass. 47 (1952), str. 583--621.

[11] Łukaszewicz, J., Sadowski, W., On comparing several populations with a control population, Zastosowania Matematyki 5 (1960/61), str. 309-320.

[12] Mann, H., Whitney, D., On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Statist. 18 (1947), str. 50-60.

[13] Moses, L., Non-parametric statistics for psychological research, Psychological Bulletin 49 (1952), str. 122-143.

[14] Olds, E., Distributions of sums of squares of rank differences for small numbers of individuals, Ann. Math. Statist. 9 (1938), str. 133-148.

[15] Olds, E., The 5% significance levels for sums of squares of rank differences and a correlation, Ann. Math. Statist. 20 (1949), str. 117-118.

[16] Perkal, J., Matematyka dla przyrodników i rolników, część III, Warszawa 1963.

[17] Sadowski, W., O nieparametrycznym teście na porównywanie rozsiewów, Zastos. Mat. 2 (1955), str. 161-171 (przekład angielski: A non-parametric test of comparing dispersions, tamże 5 (1960/61), str. 299-308).
[18] Sadowski, W. i inni, Tablice Statystyczne, Warszawa 1957.
[19] Siegel, S., Non-parametric statistics for the behavioral sciences, New York 1956.
[20] Van der Waerden, B., Mathematische Statistik, Berlin 1957 (przekład rosyjski: Matiematiczeskaja Statistika, Moskwa 1960).
[21] Wald, A., Wolfowitz, J., On a test whether two samples are from the same population, Ann. Math. Statist. 11 (1940), str. 147-162.
[22] Walsh, J., Applications of some significance tests for the median which are walid under very general conditions, J. Amer. Statist. Ass. 44 (1949), str. 342-355.
[23] Wilcoxon, F., Probability tables for individual comparisons by ranking methods, Biometrics 3 (1947), str. 119-122.
[24] Wilcoxon, F., Some rapid approximate statistical procedures, Stanford 1949.