*) Так как ... — нет оснований отвергнуть нулевую гипотезу; ранговая корреляционная связь между признаками незначимая.

§ 27. Критерий Вилкоксона и проверка гипотезы об однородности двух выборок

Критерий Вилкоксона[*] служит для проверки однородности двух независимых выборок: $x_1, x_2, \ldots, x_{n_1}$ и $y_1, y_2, \ldots, y_{n_2}$. Достоинство этого критерия состоит в том, что он применим к случайным величинам, распределения которых неизвестны; требуется лишь, чтобы величины были непрерывными.

Если выборки однородны, то считают, что они извлечены из одной и той же генеральной совокупности и, следовательно, имеют одинаковые, причем неизвестные, непрерывные функции распределения $F_1(x)$ и $F_2(x)$.

Таким образом, нулевая гипотеза состоит в том, что при всех значениях аргумента (обозначим его через x) функции распределения равны между собой: $F_1(x) = F_2(x)$.

Конкурирующими являются следующие гипотезы: $F_1(x) \neq F_2(x)$, $F_1(x) < F_2(x)$ и $F_1(x) > F_2(x)$.

Заметим, что принятие конкурирующей гипотезы H_1: $F_1(x) < F_2(x)$ означает, что $X > Y$. Действительно, неравенство $F_1(x) < F_2(x)$ равносильно неравенству $P(X < x) < P(Y < x)$. Отсюда легко получить, что $P(X > x) > P(Y > x)$. Другими словами, вероятность того, что случайная величина X превзойдет фиксированное действительное число x, больше, чем вероятность случайной величине Y оказаться большей, чем x; в этом смысле $X > Y$.

Аналогично, если справедлива конкурирующая гипотеза $H_1: F_1(x) > F_2(y)$, то $X < Y$.

---

[*] В 1945 г. Вилкоксон опубликовал критерий сравнения двух выборок одинакового объема; в 1947 г. Манн и Уитни обобщили критерий на выборки различного объема.

Далее предполагается, что объем первой выборки меньше (не больше) объема второй: $n_1 \leqslant n_2$; если это не так, то выборки можно перенумеровать (поменять местами).

А. Проверка нулевой гипотезы в случае, если объем обеих выборок не превосходит 25. Правило 1. Для того чтобы при заданном уровне значимости $\alpha = 2Q$ проверить нулевую гипотезу $H_0: F_1(x) = F_2(x)$ об однородности двух независимых выборок объемов n_1 и n_2 ($n_1 \leqslant n_2$) при конкурирующей гипотезе $H_1: F_1(x) \neq F_2(x)$, надо:

1) расположить варианты обеих выборок в возрастающем порядке, т. е. в виде одного вариационного ряда, и найти в этом ряду наблюдаемое значение критерия $W_{набл}$ — сумму порядковых номеров вариант первой выборки;

2) найти по таблице приложения 10 нижнюю критическую точку $w_{нижн.\,кр}(Q;\ n_1,\ n_2)$, где $Q = \alpha/2$;

3) найти верхнюю критическую точку по формуле

$$w_{верхн.\,кр} = (n_1 + n_2 + 1)\,n_1 - w_{нижн.\,кр}.$$

Если $W_{набл} < w_{нижн.\,кр}$ или $W_{набл} > w_{верхн.\,кр}$ — нулевую гипотезу отвергают.

Если $w_{нижн.\,кр} < W_{набл} < w_{верхн.\,кр}$ — нет оснований отвергнуть нулевую гипотезу.

Пример 1. При уровне значимости 0,05 проверить нулевую гипотезу об однородности двух выборок объемов $n_1 = 6$ и $n_2 = 8$:

x_i	15	23	25	26	28	29		
y_i	12	14	18	20	22	24	27	30

при конкурирующей гипотезе $H_1: F_1(x) \neq F_2(x)$.

Решение. Расположим варианты обеих выборок в виде одного вариационного ряда и перенумеруем их:

порядковые номера . . .	1	2	*3*	4	5	6	*7*	8	*9*	*10*	11	*12*	*13*	14
варианты . . .	12	14	15	18	20	22	23	24	25	26	27	28	29	30

Найдем наблюдаемое значение критерия Вилкоксона — сумму порядковых номеров (они набраны курсивом) вариант первой выборки:

$$W_{набл} = 3 + 7 + 9 + 10 + 12 + 13 = 54.$$

Найдем по таблице приложения 10 нижнюю критическую точку, учитывая, что $Q = \alpha/2 = 0{,}05/2 = 0{,}025$, $n_1 = 6$, $n_2 = 8$:

$$w_{нижн.\,кр}(0{,}025;\ 6,\ 8) = 29.$$

Найдем верхнюю критическую точку:

$$w_{верхн.\,кр} = (n_1 + n_2 + 1)n_1 - w_{нижн.\,кр} = (6 + 8 + 1)\cdot 6 - 29 = 61.$$

Так как $29 < 54 < 61$, т. е. $w_{нижн.\,кр} < W_{набл} < w_{верхн.\,кр}$, — нет оснований отвергнуть нулевую гипотезу об однородности выборок.

Правило 2. При конкурирующей гипотезе $F_1(x) > F_2(x)$ надо найти по таблице нижнюю критическую точку $w_{\text{нижн. кр}}(Q;\ n_1;\ n_2)$, где $Q=\alpha$.

Если $W_{\text{набл}} > w_{\text{нижн. кр}}$ — нет оснований отвергнуть нулевую гипотезу.

Если $W_{\text{набл}} < w_{\text{нижн. кр}}$ — нулевую гипотезу отвергают.

Правило 3. При конкурирующей гипотезе $H_1: F_1(x) < F_2(x)$ надо найти верхнюю критическую точку: $w_{\text{верхн. кр}}(Q;\ n_1,\ n_2) = (n_1+n_2+1)\,n_1 - w_{\text{нижн. кр}}(Q;\ n_1,\ n_2)$, где $Q=\alpha$.

Если $W_{\text{набл}} < w_{\text{верхн. кр}}$ — нет оснований отвергнуть нулевую гипотезу.

Если $W_{\text{набл}} > w_{\text{верхн. кр}}$ — нулевую гипотезу отвергают.

Замечание. Если несколько вариант только одной выборки одинаковы, то в общем вариационном ряду им приписывают обычные порядковые номера (совпавшие варианты нумеруют так, как если бы они были различными числами); если же совпадают варианты разных выборок, то всем им присваивают один и тот же порядковый номер, равный среднему арифметическому порядковых номеров, которые имели бы эти варианты до совпадения.

Б. Проверка нулевой гипотезы в случае, если объем хотя бы одной из выборок превосходит 25. 1. При конкурирующей гипотезе $F_1(x) \neq F_2(x)$ нижняя критическая точка

$$w_{\text{нижн. кр}}(Q;\ n_1,\ n_2) = \left[\frac{(n_1+n_2+1)\,n_1-1}{2} - z_{\text{кр}}\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}\right], \qquad (*)$$

где $Q=\alpha/2$; $z_{\text{кр}}$ находят по таблице функции Лапласа по равенству $\Phi(z_{\text{кр}})=(1-\alpha)/2$; знак $[a]$ означает целую часть числа a.

В остальном правило 1, приведенное в п. А, сохраняется.

2. При конкурирующих гипотезах $F_1(x) > F_2(x)$ и $F_1(x) < F_2(x)$ нижнюю критическую точку находят по формуле (*), положив $Q=\alpha$; соответственно $z_{\text{кр}}$ находят по таблице функции Лапласа по равенству $\Phi(z_{\text{кр}})=(1-2\alpha)/2$. В остальном правила 2—3, приведенные в п. А, сохраняются.

Правило 2. При конкурирующей гипотезе $F_1(x) > F_2(x)$ надо найти по таблице нижнюю критическую точку $w_{\text{нижн. кр}}(Q;\ n_1;\ n_2)$, где $Q=\alpha$.

Если $W_{\text{набл}} > w_{\text{нижн. кр}}$ — нет оснований отвергнуть нулевую гипотезу.

Если $W_{\text{набл}} < w_{\text{нижн. кр}}$ — нулевую гипотезу отвергают.

Правило 3. При конкурирующей гипотезе $H_1: F_1(x) < F_2(x)$ надо найти верхнюю критическую точку: $w_{\text{верхн. кр}}(Q;\ n_1,\ n_2) = (n_1+n_2+1)\,n_1 - w_{\text{нижн. кр}}(Q;\ n_1,\ n_2)$, где $Q=\alpha$.

Если $W_{\text{набл}} < w_{\text{верхн. кр}}$ — нет оснований отвергнуть нулевую гипотезу.

Если $W_{\text{набл}} > w_{\text{верхн. кр}}$ — нулевую гипотезу отвергают.

Замечание. Если несколько вариант только одной выборки одинаковы, то в общем вариационном ряду им приписывают обычные порядковые номера (совпавшие варианты нумеруют так, как если бы они были различными числами); если же совпадают варианты разных выборок, то всем им присваивают один и тот же порядковый номер, равный среднему арифметическому порядковых номеров, которые имели бы эти варианты до совпадения.

Б. Проверка нулевой гипотезы в случае, если объем хотя бы одной из выборок превосходит 25. 1. При конкурирующей гипотезе $F_1(x) \neq F_2(x)$ нижняя критическая точка

$$w_{\text{нижн. кр}}(Q;\ n_1,\ n_2) = \left[\frac{(n_1+n_2+1)\,n_1-1}{2} - z_{\text{кр}}\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}\right], \qquad (*)$$

где $Q=\alpha/2$; $z_{\text{кр}}$ находят по таблице функции Лапласа по равенству $\Phi(z_{\text{кр}}) = (1-\alpha)/2$; знак $[a]$ означает целую часть числа a.

В остальном правило 1, приведенное в п. А, сохраняется.

2. При конкурирующих гипотезах $F_1(x) > F_2(x)$ и $F_1(x) < F_2(x)$ нижнюю критическую точку находят по формуле (*), положив $Q=\alpha$; соответственно $z_{\text{кр}}$ находят по таблице функции Лапласа по равенству $\Phi(z_{\text{кр}}) = (1-2\alpha)/2$. В остальном правила 2—3, приведенные в п. А, сохраняются.

Пример 2. При уровне значимости 0,01 проверить нулевую гипотезу об однородности двух выборок объемов $n_1=30$ и $n_2=50$ при конкурирующей гипотезе $H_1: F_1(x) \neq F_2(x)$, если известно, что в общем вариационном ряду, составленном из вариант обеих выборок, сумма порядковых номеров вариант первой выборки $W_{\text{набл}}=1600$.

Решение. По условию, конкурирующая гипотеза имеет вид $F_1(x) \neq F_2(x)$, поэтому критическая область — двусторонняя.

Найдем $z_{кр}$ по равенству

$$\Phi(z_{кр}) = (1-\alpha)/2 = (1-0{,}01)/2 = 0{,}495.$$

По таблице функции Лапласа (см. приложение 2) находим $z_{кр} = 2{,}58$.

Подставив $n_1 = 30$, $n_2 = 50$, $z_{кр} = 2{,}58$ в формулу (*), получим $w_{нижн.\,кр} = 954$.

Найдем верхнюю критическую точку:

$$w_{верхн.\,кр} = (n_1 + n_2 + 1)\, n_1 - w_{нижн.\,кр} = 2430 - 954 = 1476.$$

Так как $1600 > 1476$, т. е. $W_{набл} > w_{верх.\,кр}$ — нулевая гипотеза отвергается.

Задачи

1. По двум независимым выборкам, объемы которых соответственно равны n_1 и n_2, извлеченным из нормальных генеральных совокупностей X и Y, найдены исправленные выборочные дисперсии s_X^2 и s_Y^2. При уровне значимости α проверить нулевую гипотезу H_0: $D(X) = D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе H_1: $D(X) > D(Y)$, если:

а) $n_1 = 10$, $n_2 = 16$, $s_X^2 = 3{,}6$, $s_Y^2 = 2{,}4$, $\alpha = 0{,}05$;

б) $n_1 = 13$, $n_2 = 18$, $s_X^2 = 0{,}72$, $s_Y^2 = 0{,}20$, $\alpha = 0{,}01$.

Отв. а) $F_{набл} = 1{,}5$; $F_{кр}(0{,}05;\, 9;\, 15) = 2{,}59$. Нет оснований отвергнуть нулевую гипотезу; б) $F_{набл} = 3{,}6$; $F_{кр}(0{,}01;\, 12;\, 17) = 3{,}45$. Нулевая гипотеза отвергается.

2. По двум независимым выборкам, объемы которых соответственно равны n и m, извлеченным из нормальных генеральных совокупностей X и Y, найдены выборочные средние $\overline{x}$ и $\overline{y}$. Генеральные дисперсии $D(X)$ и $D(Y)$ известны. При уровне значимости α проверить нулевую гипотезу H_0: $M(X) = M(Y)$ о равенстве математических ожиданий при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$, если:

а) $n = 30$, $m = 20$, $D(X) = 120$, $D(Y) = 100$, $\alpha = 0{,}05$;

б) $n = 50$, $m = 40$, $D(X) = 50$, $D(Y) = 120$, $\alpha = 0{,}01$.

Отв. а) $Z_{набл} = 1$, $z_{кр} = 1{,}96$. Нет оснований отвергнуть нулевую гипотезу; б) $Z_{набл} = 10$, $z_{кр} = 2{,}58$. Нулевая гипотеза отвергается.

3. По двум независимым выборкам, объемы которых соответственно равны $n = 5$ и $m = 6$, извлеченным из нормальных генеральных совокупностей X и Y, найдены выборочные средние $\overline{x} = 15{,}9$, $\overline{y} = 14{,}1$ и исправленные выборочные дисперсии $s_X^2 = 14{,}76$, $s_Y^2 = 4{,}92$. При уровне значимости 0,05 проверить нулевую гипотезу $H_0: M(X) = M(Y)$ о равенстве математических ожиданий при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$.

Указание. Предварительно сравни[illegible]