

УДК 519.2:303.732.4

UDC 519.2:303.732.4

**ДВУХВЫБОРОЧНЫЙ КРИТЕРИЙ  
ВИЛКОКСОНА – АНАЛИЗ ДВУХ МИФОВ****TWO-SAMPLE WILCOXON TEST -  
ANALYSIS OF TWO MYTHS**

Орлов Александр Иванович  
д.э.н., д.т.н., к.ф.-м.н., профессор

Orlov Alexander Ivanovich  
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci.,  
professor  
Bauman Moscow State Technical University,  
Moscow, Russia

*Московский государственный технический  
университет им. Н.Э. Баумана, Россия, 105005,  
Москва, 2-я Бауманская ул., 5, [prof-orlov@mail.ru](mailto:prof-orlov@mail.ru)*

Установлено, что двухвыборочный критерий Вилкоксона (Манна-Уитни) предназначен для проверки гипотезы  $H_0: P(X < Y) = 1/2$ , где  $X$  - случайная величина, распределенная как элементы первой выборки, а  $Y$  - второй. Показано, что критерий Вилкоксона не позволяет проверять совпадение функций распределения двух независимых выборок, а также равенство медиан этих функций распределения

It was established that the two-sample Wilcoxon test (Mann-Whitney test) was designed to test the hypothesis  $H_0: P(X < Y) = 1/2$ , where  $X$  - a random variable distributed as the elements of the first sample, and  $Y$  - second. It is shown that the Wilcoxon test does not allow verifying the identity of the distribution functions of two independent samples, as well as equality of medians of these distribution functions

Ключевые слова: СТАТИСТИЧЕСКИЕ МЕТОДЫ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, ПРИКЛАДНАЯ СТАТИСТИКА, НЕЗАВИСИМЫЕ ВЫБОРКИ, ГИПОТЕЗА ОДНОРОДНОСТИ, ДВУХВЫБОРОЧНЫЙ КРИТЕРИЙ ВИЛКОКСОНА

Keywords: STATISTICAL METHODS, MATHEMATICAL STATISTICS, APPLIED STATISTICS, INDEPENDENT SAMPLES, THE HOMOGENEITY HYPOTHESIS, TWO-SAMPLE WILCOXON TEST

## 1. Введение

Критерий Вилкоксона (Манна – Уитни) – один из наиболее известных непараметрических статистических критериев. Он широко применяется при анализе конкретных данных.

Настоящая статья посвящена изучению асимптотических свойств двухвыборочного непараметрического рангового критерия Вилкоксона (Манна – Уитни). На основе асимптотической нормальности статистики этого критерия установлено, что критерий Вилкоксона (Манна – Уитни) предназначен для проверки гипотезы  $H_0: P(X < Y) = 1/2$ , где  $X$  - случайная величина, распределенная как элементы первой выборки, а  $Y$  - второй.

С критерием Вилкоксона связано два мифа:

- 1) он позволяет проверить совпадение (тождественное равенство) функций распределения двух независимых выборок;
- 2) с его помощью можно убедиться в равенстве (теоретических) медиан двух выборок.

Построены контрпримеры, позволяющие опровергнуть эти два мифа. Обсуждаются проблемы использования для проверки гипотез критериев, которые не являются состоятельными с точки зрения теории математической статистики. Рассмотрено теоретическое и прикладное значение гипотезы сдвига.

## **2. Критерии однородности Вилкоксона и Манна-Уитни**

В прикладных исследованиях часто возникает необходимость выяснить, различаются ли генеральные совокупности, из которых взяты две независимые выборки. Например, надо выяснить, влияет ли способ упаковки подшипников на их потребительские качества через год после хранения. Или: отличается ли потребительское поведение мужчин и женщин. Если отличается – рекламные ролики и плакаты надо делать отдельно для мужчин и отдельно для женщин. Если нет – рекламная кампания может быть единой.

В математико-статистических терминах постановка задачи такова: имеются две выборки  $x_1, x_2, \dots, x_m$  и  $y_1, y_2, \dots, y_n$  (т. е. наборы из  $m$  и  $n$  действительных чисел), требуется проверить их однородность. Термин «однородность» уточняется ниже.

Противоположным понятием является «различие» (другими словами, «наличие эффекта при переходе от одной выборки к другой»). Можно переформулировать задачу: требуется проверить, есть ли различие между выборками. Если различия нет, то для дальнейшего изучения две рассматриваемые выборки часто объединяют в одну.

Например, в маркетинге важно выделить сегменты потребительского рынка. Если установлена однородность двух выборок, то возможно объединение сегментов, из которых они взяты, в один. В дальнейшем это позволит осуществлять по отношению к ним одинаковую маркетинговую политику (проводить одни и те же рекламные мероприятия и т.п.). Если же

установлено различие, то поведение потребителей в двух сегментах различно, объединять эти сегменты нельзя, и могут понадобиться различные маркетинговые стратегии, своя для каждого из этих сегментов.

В прикладной математической статистике часто рассматривают вероятностную модель двух независимых выборок числовых результатов наблюдений. Первая выборка описывается набором  $m$  случайных величин  $X_1, X_2, \dots, X_m$ , имеющих одну и ту же функцию распределения  $F(x)$ , а вторая выборка - набором  $n$  случайных величин  $Y_1, Y_2, \dots, Y_n$ , имеющих одну и ту же функцию распределения  $G(x)$ , причем все эти  $m+n$  случайных величин  $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$  независимы в совокупности. Без ограничения общности можно считать, что  $m \leq n$ , в противном случае выборки можно поменять местами. Обычно предполагается, что функции  $F(x)$  и  $G(x)$  непрерывны и строго возрастают. Из непрерывности этих функций следует, что с вероятностью 1 все  $m+n$  результатов наблюдений различны. В реальных статистических данных иногда встречаются совпадения, но сам факт их наличия - свидетельство нарушений предпосылок только что описанной базовой математической модели.

Предполагается, что результаты наблюдений (измерений, испытаний, анализов, опытов) – числа  $x_1, x_2, \dots, x_m$  и  $y_1, y_2, \dots, y_n$  – являются реализациями случайных величин  $X_1, X_2, \dots, X_m$ , и  $Y_1, Y_2, \dots, Y_n$  соответственно, т.е. значениями случайных величин (как функций от элементарного события), принимаемыми для определенного случайного события.

Статистика  $S$  двухвыборочного критерия Вилкоксона определяется следующим образом. Выборки объединяют вместе, получают т.н. объединенную выборку. Все элементы объединенной выборки  $X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$  упорядочивают в порядке возрастания, строя общий (объединенный) вариационный ряд. Элементы первой выборки  $X_1, X_2, \dots, X_m$  занимают в общем вариационном ряду места с номерами  $R_1, R_2, \dots, R_m$ ,

другими словами, имеют ранги  $R_1, R_2, \dots, R_m$ . Тогда статистика Вилкоксона - это сумма рангов элементов первой выборки

$$S = R_1 + R_2 + \dots + R_m.$$

Статистика  $U$  Манна-Уитни определяется как число пар  $(X_i, Y_j)$  таких, что  $X_i < Y_j$ , среди всех  $mn$  пар, в которых первый элемент - из первой выборки, а второй - из второй. Как известно (см., например, [1, с.160]),

$$U = mn + m(m+1)/2 - S.$$

Правила принятия решения на основе статистик  $S$  и  $U$ , т.е. собственно критерии Вилкоксона и Манна-Уитни, рассмотрим ниже.

Поскольку  $S$  и  $U$  линейно связаны, то часто говорят не о двух критериях - Вилкоксона и Манна-Уитни, а об одном - критерии Вилкоксона (Манна-Уитни).

### 3. Асимптотическое распределение статистики критерия Вилкоксона

Критерий Вилкоксона - один из самых известных инструментов непараметрической статистики (наряду с критериями на основе статистик типа Колмогорова-Смирнова, омега-квадрат [2] и коэффициентами ранговой корреляции). Свойствам этого критерия и таблицам его критических значений уделяется место во многих монографиях по математической и прикладной статистике (см., например, [1, 3, 4]).

Однако в литературе имеются и неточные утверждения относительно возможностей критерия Вилкоксона. Так, одни полагают, что с его помощью можно обнаружить любое различие между функциями распределения  $F(x)$  и  $G(x)$ . По мнению других, этот критерий нацелен на проверку равенства медиан распределений, соответствующих выборкам. И то, и другое, строго говоря, неверно. Это будет ясно из дальнейшего изложения.

Введем некоторые обозначения. Пусть  $F^{-1}(t)$  - функция, обратная к функции распределения  $F(x)$ . Она определена на отрезке  $[0;1]$ . Положим

$L(t) = G(F^{-1}(t))$ . Поскольку  $F(x)$  непрерывна и строго возрастает, то  $F^{-1}(t)$  и  $L(t)$  обладают теми же свойствами. Важную роль в дальнейшем изложении будет играть величина  $a = P(X < Y)$ . Здесь  $X$  – случайная величина, распределение которой совпадает с распределением элементов первой выборки  $X_1, X_2, \dots, X_m$ , а  $Y$  – случайная величина, распределение которой совпадает с распределением элементов второй выборки  $Y_1, Y_2, \dots, Y_n$ . Как нетрудно показать,

$$a = P(X < Y) = \int_0^1 t dL(t).$$

Введем также параметры

$$b^2 = \int_0^1 L^2(t) dt - (1-a)^2, \quad g^2 = \int_0^1 t^2 dL(t) - a^2.$$

Тогда математические ожидания и дисперсии статистик Вилкоксона и Манна-Уитни согласно [1, с.160] выражаются через введенные величины:

$$\begin{aligned} M(U) = mna, \quad M(S) = mn + m(m+1)/2 - M(U) = mn(1-a) + m(m+1)/2, \\ D(S) = D(U) = mn [(n-1)b^2 + (m-1)g^2 + a(1-a)]. \end{aligned} \quad (1)$$

Когда объемы обеих выборок безгранично растут, распределения статистик Вилкоксона и Манна-Уитни являются асимптотически нормальными (см., например, [1, гл. 5 и 6]) с параметрами, задаваемыми формулами (1).

Если выборки полностью однородны, т.е. их функции распределения  $F(x)$  и  $G(x)$  тождественно совпадают, справедлива гипотеза

$$H_0: F(x) = G(x) \text{ при всех } x, \quad (2)$$

то  $L(t) = t$  для  $t$  из отрезка  $[0, 1]$ ,  $L(t) = 0$  для всех отрицательных  $t$  и  $L(t) = 1$  для  $t > 1$ , соответственно  $a = 1/2$ . Подставляя в формулы (1), получаем, что

$$M(S) = m(m+n+1)/2, \quad D(S) = mn(m+n+1)/12. \quad (3)$$

Следовательно, распределение нормированной и центрированной статистики Вилкоксона

$$T = (S - m(m+n+1)/2) (mn(m+n+1)/12)^{-1/2} \quad (4)$$

при росте объемов выборок приближается к стандартному нормальному распределению (с математическим ожиданием 0 и дисперсией 1).

Из асимптотической нормальности статистики  $T$  следует, что правило принятия решения для критерия Вилкоксона выглядит так:

– если  $|T| \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ , то гипотеза (2) однородности (тождества)

функций распределений принимается на уровне значимости  $\alpha$ ,

– если же  $|T| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ , то гипотеза (2) однородности (тождества)

функций распределений отклоняется на уровне значимости  $\alpha$ .

Здесь  $\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$  – квантиль порядка  $\left(1 - \frac{\alpha}{2}\right)$  стандартного нормального распределения (с математическим ожиданием 0 и дисперсией 1).

В прикладной статистике наиболее часто применяется уровень значимости  $\alpha = 0,05$ . Тогда значение модуля статистики  $T$  Вилкоксона надо сравнивать с граничным значением  $\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) = 1,96$ .

#### 4. Алгоритм проверки однородности с помощью критерия Вилкоксона

Представим алгоритм на примере.

*Пример 1.* Пусть даны две выборки. Первая содержит  $m = 12$  элементов 17; 22; 3; 5; 15; 2; 0; 7; 13; 97; 66; 14. Вторая содержит  $n = 14$  элементов 47; 30; 2; 15; 1; 21; 25; 7; 44; 29; 33; 11; 6; 15. Проведем проверку однородности функций распределения двух выборок с помощью критерия Вилкоксона.

Первым шагом является построение общего вариационного ряда для элементов двух выборок (табл.1).

**Таблица 1. Общий вариационный ряд для элементов двух выборок**

Ранги	1	2	3,5	3,5	5	6	7	8,5	8,5	10	11	12	14
Элементы выборок	0	1	2	2	3	5	6	7	7	11	13	14	15
Номера выборок	1	2	1	2	1	1	2	1	2	2	1	1	1
Ранги	14	14	16	17	18	19	20	21	22	23	24	25	26
Элементы выборок	15	15	17	21	22	25	29	30	33	44	47	66	97
Номера выборок	2	2	1	2	1	2	2	2	2	2	2	1	1

Хотя с точки зрения теории математической статистики вероятность совпадения двух элементов выборок равна 0, в реальных выборках статистических данных совпадения встречаются. Так, в рассматриваемых выборках, как видно из табл.1, два раза повторяется величина 2, два раза - величина 7 и три раза - величина 15. В таких случаях говорят о наличии «связанных рангов», а соответствующим совпадающим величинам приписывают среднее арифметическое тех рангов (т.е. мест в общем вариационном ряду), которые они занимают. Так, величины 2 и 2 занимают в объединенной выборке места 3 и 4, поэтому им приписывается связанный ранг  $(3+4)/2=3,5$ . Величины 7 и 7 занимают в объединенной выборке места 8 и 9, поэтому им приписывается связанный ранг  $(8+9)/2=8,5$ . Величины 15, 15 и 15 занимают в объединенной выборке места 13, 14 и 15, поэтому им приписывается связанный ранг  $(13+14+15)/3=14$ .

Следующий шаг - подсчет значения статистики Вилкоксона, т.е. суммы рангов элементов первой выборки

$$S = R_1 + R_2 + \dots + R_m = 1+3,5+5+6+8,5+11+12+14+16+18+25+26=146.$$

Подсчитаем также сумму рангов элементов второй выборки

$$S_1 = 2+3,5+7+8,5+10+14+14+17+19+20+21+22+23+24 = 205.$$

Величина  $S_1$  может быть использована для контроля вычислений. Дело в том, что суммы рангов элементов первой выборки  $S$  и второй выборки  $S_1$  вместе составляют сумму рангов объединенной выборки, т.е. сумму всех натуральных чисел от 1 до  $m+n$ . Следовательно,

$$S + S_1 = (m+n)(m+n+1)/2 = (12+14)(12+14+1)/2 = 351.$$

В соответствии с ранее проведенными расчетами  $S+S_1 = 146+205=351$ . Необходимое условие правильности расчетов выполнено. Ясно, что оно не является достаточным, т.е. справедливость этого условия не гарантирует правильности расчетов.

Перейдем к расчету статистики  $T$ . Согласно формуле (3)

$$M(S) = 12(12+14+1)/2 = 162, D(S) = 12 \cdot 14(12+14+1)/12 = 378.$$

Следовательно,

$$T = (S - 162) (378)^{-1/2} = (146 - 162) / 19,44 = -0,82.$$

Поскольку  $|T| \leq 1,96$ , то гипотеза однородности принимается на уровне значимости 0,05.

Что будет, если поменять выборки местами, вторую назвать первой?

Тогда вместо  $S$  надо рассматривать  $S_1$ . Имеем

$$M(S_1) = 14(12+14+1)/2 = 189, D(S) = D(S_1) = 378,$$

$$T_1 = (S_1 - 189) (378)^{-1/2} = (205 - 189) / 19,44 = 0,82.$$

Таким образом, значения статистики критерия отличаются только знаком (можно показать, что это утверждение верно всегда). Поскольку в правиле принятия решения используется только абсолютная величина статистики, то принимаемое решение не зависит от того, какую выборку считаем первой, а какую второй. Для уменьшения объема таблиц критических значений принято считать первой выборку меньшего объема.



## 5. Мощность критерия Вилкоксона

Продолжим обсуждение критерия Вилкоксона. Правила принятия решений и таблицы критических значений для критерия Вилкоксона строятся в предположении справедливости гипотезы полной однородности, описываемой формулой (2). А что будет, если эта гипотеза неверна? Другими словами, какова мощность критерия Вилкоксона?

Пусть объемы выборок достаточно велики, так что можно пользоваться асимптотической нормальностью статистики Вилкоксона. Тогда в соответствии с формулами (1) статистика  $T$  будет асимптотически нормальна с параметрами

$$M(T) = (12mn)^{1/2} (1/2 - a) (m+n+1)^{-1/2},$$

$$D(T) = 12 [(n - 1) b^2 + (m - 1) g^2 + a(1 - a)] (m+n+1)^{-1}. \quad (5)$$

Из формул (5) видно большое значение гипотезы

$$H_{01}: a = P(X < Y) = 1/2. \quad (6)$$

Её можно переформулировать так: медиана случайной величины  $X - Y$  равна 0. Если эта гипотеза неверна, то, поскольку  $m \leq n$ , справедлива оценка

$$|M(T)| \geq (12m n (2n+1)^{-1})^{1/2} |1/2 - a|,$$

а потому  $|M(T)|$  безгранично растет при росте объемов выборок. В то же время, поскольку

$$b^2 \leq \int_0^1 L^2(t) dt \leq 1, \quad g^2 \leq \int_0^1 t^2 dL(t) \leq 1, \quad \alpha(1 - \alpha) \leq 1/4,$$

то

$$D(T) \leq 12 [(n - 1) + (m - 1) + 1/4] (m+n+1)^{-1} \leq 12. \quad (7)$$

Следовательно, вероятность отклонения гипотезы  $H_{01}$ , когда она неверна, т.е. мощность критерия Вилкоксона как критерия проверки гипотезы (6), стремится к 1 при возрастании объемов выборок, т.е. критерий Вилкоксона является состоятельным для этой гипотезы при альтернативе

$$AH_{01}: a = P(X < Y) \neq 1/2. \quad (8).$$

Если же гипотеза (6) верна, то статистика  $T$  асимптотически нормальна с математическим ожиданием 0 и дисперсией, определяемой формулой

$$D(T) = 12[(n - 1)b^2 + (m - 1)g^2 + 1/4](m+n+1)^{-1}. \quad (9)$$

Гипотеза (6) является сложной, дисперсия (9), как показывают приводимые ниже примеры, в зависимости от значений  $b^2$  и  $g^2$  может быть как больше 1, так и меньше 1, но согласно неравенству (7) никогда не превосходит 12.

### **6. Критерий Вилкоксона не является состоятельным при проверке гипотезы однородности функций распределения**

Приведем пример двух функций распределения  $F(x)$  и  $G(x)$  таких, что гипотеза (6) выполнена, а гипотеза (2) – нет. Поскольку

$$a = P(X < Y) = \int_{-\infty}^{+\infty} F(x)dG(x), \quad 1 - a = P(Y < X) = \int_{-\infty}^{+\infty} G(x)dF(x) \quad (10)$$

и  $a = 1/2$  в случае справедливости гипотезы (2), то для выполнения условия (6) необходимо и достаточно, чтобы

$$\int_{-\infty}^{+\infty} (F(x) - G(x))dF(x) = 0 \quad (11),$$

а потому естественно в качестве  $F(x)$  рассмотреть функцию равномерного распределения на интервале  $(-1; 1)$ . Тогда формула (11) переходит в условие

$$\int_{-\infty}^{+\infty} (F(x) - G(x))dF(x) = -\frac{1}{2} \int_{-1}^{+1} \left( G(x) - \frac{(x+1)}{2} \right) dx = 0. \quad (11).$$

Это условие выполняется, если функция  $(G(x) - (x + 1)/2)$  является нечетной.

*Пример 2.* Пусть функции распределения  $F(x)$  и  $G(x)$  сосредоточены на интервале  $(-1; 1)$ , на котором

$$F(x) = (x + 1)/2, \quad G(x) = (x + 1 + 1/\pi \sin \pi x)/2.$$

Тогда

$$x=F^{-1}(t)=2t-1, L(t)=G(F^{-1}(t))=(2t+1/\pi \sin \pi (2t-1))/2=t+1/2 \pi \sin \pi (2t-1).$$

Условие (11) выполнено, поскольку функция  $(G(x) - (x + 1)/2)$  является нечетной. Следовательно,  $a = 1/2$ . Начнем с вычисления

$$g^2 = \int_0^1 t^2 dL(t) - 1/4 = \int_0^1 t^2 d\left(t + \frac{1}{2\pi} \sin \pi(2t-1)\right) - \frac{1}{4}.$$

Поскольку

$$d\left(t + \frac{1}{2\pi} \sin \pi(2t-1)\right) = (1 + \cos \pi(2t-1))dt,$$

то

$$g^2 = \int_0^1 t^2 (1 + \cos \pi(2t-1))dt - \frac{1}{4} = \frac{1}{12} + \int_0^1 t^2 \cos \pi(2t-1)dt.$$

С помощью замены переменных  $t = (x + 1) / 2$  получаем, что

$$\int_0^1 t^2 \cos \pi(2t-1)dt = \frac{1}{8} \left( \int_{-1}^1 x^2 \cos \pi x dx + 2 \int_{-1}^1 x \cos \pi x dx + \int_{-1}^1 \cos \pi x dx \right).$$

В правой части последнего равенства стоят табличные интегралы (см., например, справочник [5, с.71]). Проведя соответствующие вычисления, получаем, что в правой части стоит  $1/8 \times (-4/\pi^2) = -1/(2\pi^2)$ . Следовательно,

$$g^2 = 1/12 - 1/(2\pi^2) = 0,032672733...$$

Перейдем к вычислению  $b^2$ . Поскольку

$$b^2 = \int_0^1 L^2(t)dt - \frac{1}{4} = \int_0^1 \left( t + \frac{1}{2} \pi \sin \pi(2t-1) \right)^2 dt - \frac{1}{4},$$

то

$$b^2 = \frac{1}{12} + \frac{1}{\pi} \int_0^1 (t \sin \pi(2t-1))dt + \left( \frac{\pi}{2} \right)^2 \int_0^1 \sin^2 \pi(2t-1)dt.$$

С помощью замены переменных  $t = (x+1)/2$  переходим к табличным интегралам (см., например, справочник [5, с.65]):

$$b^2 = \frac{1}{12} + \frac{1}{4\pi} \int_{-1}^1 x \sin \pi x dx + \frac{1}{4\pi} \int_{-1}^1 \sin \pi x dx + \frac{1}{8\pi^2} \int_{-1}^1 \sin^2 \pi x dx.$$

Проведя необходимые вычисления, получим, что

$$b^2 = \frac{1}{12} + \frac{1}{4\pi} \left( -\frac{2}{\pi} \right) + 0 + \frac{1}{8\pi^2} = \frac{1}{12} - \frac{3}{8\pi^2} = 0,045337893\dots$$

Следовательно, для рассматриваемых функций распределения нормированная и центрированная статистика Вилкоксона (см. формулу (4)) асимптотически нормальна с математическим ожиданием 0 и дисперсией (см. формулу (9))

$$D(T) = (0,544 n + 0,392 m + 2,064) (m+n+1)^{-1}.$$

Как легко видеть, дисперсия всегда меньше 1. Это значит, что в рассматриваемом случае гипотеза полной однородности (2) при проверке с помощью критерия Вилкоксона будет приниматься чаще, чем если она на самом деле верна.

На наш взгляд, это означает, что критерий Вилкоксона нельзя считать критерием для проверки гипотезы (2) при альтернативе общего вида. Он не всегда позволяет проверить однородность – не при всех альтернативах. Точно так же критерии типа хи-квадрат нельзя считать критериями проверки гипотез согласия и однородности – они позволяют обнаружить не все различия, поскольку некоторые из различий «скрадывает» группировка.

## 7. Критерий Вилкоксона не позволяет проверять равенство медиан

Обсудим теперь, действительно ли критерий Вилкоксона нацелен на проверку равенства медиан распределений, соответствующих выборкам.

*Пример 3.* Построим семейство пар функций распределения  $F(x)$  и  $G(x)$  таких, что их медианы различны, но для  $F(x)$  и  $G(x)$  выполнена гипотеза (6). Пусть распределения сосредоточены на интервале  $(0; 1)$ , и на нем  $G(x) = x$ , а  $F(x)$  имеет кусочно-линейный график с вершинами в точках  $(0; 0)$ ,  $(\lambda, 1/2)$ ,  $(\delta, 3/4)$ ,  $(1; 1)$ . Следовательно,

$$F(x) = 0 \text{ при } x < 0;$$

$$F(x) = x/(2\lambda) \text{ на } [0; \lambda];$$

$$F(x) = 1/2 + (x - \lambda)/(4\delta - 4\lambda) \text{ на } [\lambda; \delta];$$

$$F(x) = 3/4 + (x - \delta)/(4 - 4\delta) \text{ на } [\delta; 1];$$

$$F(x) = 1 \text{ при } x > 1.$$

Очевидно, что медиана  $F(x)$  равна  $\lambda$ , а медиана  $G(x)$  равна  $1/2$ .

Согласно соотношению (9) для выполнения гипотезы (6) достаточно определить  $\delta$  как функцию  $\lambda$ , т.е.  $\delta = \delta(\lambda)$ , из условия

$$\int_0^1 F(x)dx = \frac{1}{2}.$$

Вычисления дают

$$\delta = \delta(\lambda) = 3(1 - \lambda)/2.$$

Учитывая, что  $\delta$  лежит между  $\lambda$  и 1, не совпадая ни с тем, ни с другим, получаем ограничения на  $\lambda$ , а именно,  $1/3 < \lambda < 3/5$ . Итак, построено искомое семейство пар функций распределения.

*Пример 4.* Пусть, как и в примере 3, распределения сосредоточены на интервале  $(0; 1)$ , и на нем  $F(x) = x$ . А  $G(x)$  - функция распределения, сосредоточенного в двух точках, а именно,  $\beta$  и 1. Т.е.  $G(x) = 0$  при  $x$ , не превосходящем  $\beta$ ;  $G(x) = h$  на  $(\beta; 1]$ ;  $G(x) = 1$  при  $x > 1$ . С такой функцией  $G(x)$  легко проводить расчеты. Однако она не удовлетворяет принятым выше условиям непрерывности и строгого возрастания. Вместе с тем легко видеть, что она является предельной (сходимость в каждой точке отрезка  $[0; 1]$ ) для последовательности функций распределения, удовлетворяющих этим условиям. А распределение статистики Вилкоксона для пары функций распределения примера 4 является предельным для последовательности соответствующих распределений статистики Вилкоксона, полученных в рассматриваемых условиях непрерывности и строгого возрастания.

Условие  $P(X < Y) = 1/2$  выполнено, если  $h = (1 - \beta)^{-1/2}$  (при  $\beta$  из отрезка  $[0; 1/2]$ ). Поскольку  $h > 1/2$  при положительном  $\beta$ , то очевидно, что медиана  $G(x)$  равна  $\beta$ , в то время как медиана  $F(x)$  равна  $1/2$ . Значит,

при  $\beta = 1/2$  медианы совпадают, при всех иных положительных  $\beta$  – различны. При  $\beta = 0$  медианой  $G(x)$  является любая точка из отрезка  $[0; 1]$ .

Легко подсчитать, что в условиях примера 4 параметры предельного распределения имеют вид

$$b^2 = \beta(1 - \beta)^{-1}/4, \quad g^2 = (1 - 2\beta)/4.$$

Следовательно, распределение нормированной и центрированной статистики Вилкоксона будет асимптотически нормальным с математическим ожиданием 0 и дисперсией

$$D(T) = 3 [(n - 1) \beta (1 - \beta)^{-1} + (m - 1) (1 - 2\beta) + 1] (m+n+1)^{-1}.$$

Проанализируем величину  $D(T)$  в зависимости от параметра  $\beta$  и объемов выборок  $m$  и  $n$ . При достаточно больших  $m$  и  $n$

$$D(T) = 3w \beta (1 - \beta)^{-1} + 3(1 - w)(1 - 2\beta)$$

с точностью до величин порядка  $(m+n)^{-1}$ , где  $w = n/(m+n)$ . Значит,  $D(T)$  - линейная функция от  $w$ , а потому достигает экстремальных значений на границах интервала изменения  $w$ , т.е. при  $w = 0$  и  $w = 1$ . Легко видеть, что при  $\beta(1 - \beta)^{-1} < 1 - 2\beta$  минимум равен  $3\beta(1 - \beta)^{-1}$  (при  $w = 1$ ), а максимум равен  $3(1 - 2\beta)$  (при  $w = 0$ ). В случае  $\beta(1 - \beta)^{-1} > 1 - 2\beta$  максимум равен  $3\beta(1 - \beta)^{-1}$  (при  $w = 1$ ), а минимум равен  $3(1 - 2\beta)$  (при  $w = 0$ ). Если же  $\beta(1 - \beta)^{-1} = 1 - 2\beta$  (это равенство справедливо при  $\beta = \beta_0 = 1 - 2^{-1/2} = 0,293$ ), то  $D(T) = 3(2^{1/2} - 1) = 1,2426...$  при всех  $w$  из отрезка  $[0; 1]$ .

Первый из описанных выше случаев имеет быть при  $\beta < \beta_0$ . При этом минимум  $D(T)$  возрастает от 0 (при  $\beta = 0, w = 1$  - предельный случай) до  $3(2^{1/2} - 1)$  (при  $\beta = \beta_0, w$  - любом), а максимум уменьшается от 3 (при  $\beta = 0, w = 0$  - предельный случай) до  $3(2^{1/2} - 1)$  (при  $\beta = \beta_0, w$  - любом). Второй случай относится к  $\beta$  из интервала  $(\beta_0; 1/2]$ . При этом минимум убывает от приведенного выше значения для  $\beta = \beta_0$  до 0 (при  $\beta = 1/2$ ,

$w=0$  – предельный случай), а максимум возрастает от того же значения при  $\beta = \beta_0$  до 3 (при  $\beta = 1/2, w = 0$ ).

Таким образом,  $D(T)$  может принимать все значения из интервала (0; 3) в зависимости от значений  $\beta$  и  $w$ . Если  $D(T) < 1$ , то при применении критерия Вилкоксона к выборкам с рассматриваемыми функциями распределения гипотеза однородности (2) будет приниматься чаще (при соответствующих значениях  $\beta$  и  $w$  – с вероятностью, сколь угодно близкой к 1), чем если бы она самом деле была верна. Если  $1 < D(T) < 3$ , то гипотеза (2) также принимается достаточно часто. Так, если уровень значимости критерия Вилкоксона равен 0,05, то (асимптотическая) критическая область этого критерия, как показано выше, имеет вид  $\{T: |T| \geq 1,96\}$ . Если – самый плохой случай –  $D(T) = 3$ , то гипотеза (2) принимается с вероятностью 0,7422.

## 8. Гипотеза сдвига

При проверке гипотезы однородности мы рассмотрели различные виды нулевых и альтернативных гипотез – гипотезу (2) и ее отрицание в качестве альтернативы, гипотезу (6) и ее отрицание, гипотезы о равенстве или различии медиан. В теоретических работах по математической статистике часто рассматривают гипотезу сдвига, в которой альтернативой гипотезе (2) является гипотеза

$$H_1: F(x) = G(x + r) \quad (12)$$

при всех  $x$  и некотором сдвиге  $r$ , отличным от 0. Если верна альтернативная гипотеза  $H_1$ , то вероятность  $P(X < Y)$  отлична от 1/2, а потому при альтернативе (12) критерий Вилкоксона является состоятельным.

В некоторых прикладных постановках гипотеза (12) представляется естественной. Например, если одним и тем же прибором проводятся две серии измерений двух значений некоторой величины (физической,

химической и т.п.). При этом функция распределения  $G(x)$  описывает результаты измерений (вместе с погрешностями) одного значения, а  $G(x+r)$  - другого. Вопреки распространенному заблуждению, хорошо известно, что распределение погрешностей измерений, как правило, не является нормальным [6].

Однако при анализе конкретных статистических данных, как правило, нет никаких оснований считать, что отсутствие однородности всегда выражается столь однозначным образом, как следует из формулы (12). Поэтому для проверки однородности (тождественного совпадения функций распределения двух выборок) необходимо использовать статистические критерии, состоятельные против любого отклонения от гипотезы однородности (2).

Почему же математики так любят гипотезу сдвига (12)? Наше мнение таково: потому, что она дает возможность доказывать глубокие математические результаты, например, об асимптотической оптимальности критериев [7]. К сожалению, с точки зрения прикладной статистики (как науки об обработке реальных данных в конкретных предметных областях) это напоминает поиск ключей под фонарем, где светло, а не в кустах, где они потеряны.

Отметим еще одно обстоятельство. Часто говорят (в соответствии с классическим подходом математической статистики), что нельзя проверять нулевые гипотезы без рассмотрения альтернативных. Однако при анализе данных, полученных в ходе технических, экономических, социологических, биологических, медицинских или иных исследований, зачастую полностью ясна формулировка той гипотезы, которую желательно проверить (например, гипотезы полной однородности - см. формулу (2)), в то время как формулировка альтернативной гипотезы не очевидна. То ли это гипотеза о неверности равенства (2) хотя бы для одного значения  $x$ , то ли это альтернатива (8), то ли - альтернатива сдвига



(12), и т. д. В таких случаях целесообразно «обернуть» задачу - исходя из статистического критерия найти альтернативы, относительно которых он состоятелен. Именно это и проделано в настоящей статье для критерия Вилкоксона.

## 9. Заключение

Подведем итоги рассмотрения критерия Вилкоксона.

1. Критерий Вилкоксона (Манна-Уитни) является одним из самых распространенных непараметрических ранговых критериев, используемых для проверки однородности двух выборок. Значение статистики критерия Вилкоксона не меняется при любом монотонном преобразовании шкалы измерения (т.е. он пригоден для статистического анализа данных, измеренных в порядковой шкале).

2. Распределение статистики критерия Вилкоксона определяется функциями распределения  $F(x)$  и  $G(x)$  и объемами  $m$  и  $n$  двух выборок. При больших объемах выборок распределение статистики Вилкоксона является асимптотически нормальным с параметрами, выписанными выше (см. формулы (1), (3) и (5)).

3. При альтернативной гипотезе, когда функции распределения выборок  $F(x)$  и  $G(x)$  не совпадают, распределение статистики Вилкоксона зависит от величины  $a = P(X < Y)$ . Если  $a$  отличается от  $1/2$ , то мощность критерия Вилкоксона стремится к 1, и он отличает нулевую гипотезу  $F = G$  от альтернативной. Если же  $a = 1/2$ , то это не всегда имеет место. В примере 2 приведены две *различные* функции распределения выборок  $F(x)$  и  $G(x)$  такие, что гипотеза однородности  $F = G$  при проверке с помощью критерия Вилкоксона будет приниматься *чаще*, чем если бы она на самом деле была верна.

4. Следовательно, в случае общей альтернативы критерий Вилкоксона не является состоятельным, как это иногда утверждают [8],

т.е. не всегда позволяет обнаружить различие функций распределения. Однако это не лишает его практической ценности, точно так же, как несостоятельность критериев типа хи-квадрат при проверке согласия, независимости или однородности не мешает отклонять нулевую гипотезу во многих практически важных случаях. Однако принятие нулевой гипотезы с помощью критерия Вилкоксона может означать не совпадение  $F$  и  $G$ , а лишь выполнение равенства  $a = 1/2$ .

5. Иногда утверждают [8], что с помощью критерия Вилкоксона можно проверять равенство медиан функций распределения  $F$  и  $G$ . Это не так. В примерах 3 и 4 указаны функции распределения  $F$  и  $G$  с  $a = 1/2$ , но с различными медианами. Во многих случаях это различие нельзя обнаружить с помощью критерия Вилкоксона, как это показано при численном анализе асимптотической дисперсии в примере 4.

6. Указанные выше недостатки критерия Вилкоксона исчезают для специального вида альтернативы - т.н. «альтернативы сдвига»  $H_1: F(x) = G(x + r)$ . В этом частном случае при справедливости альтернативной гипотезы мощность стремится к 1, различие медиан также всегда обнаруживается. Однако альтернатива сдвига не всегда естественна. Ее целесообразно принять, если одним и тем же прибором проводятся две серии измерений двух значений некоторой величины (физической, химической и т.п.). При этом функция распределения  $G(x)$  описывает результаты измерений (с погрешностями) одного значения, а  $F(x) = G(x+r)$  - другого. Другими словами, меняется лишь измеряемое значение, а собственно распределение погрешностей - одно и то же, присущее используемому средству измерения (и обычно описанное в его техническом паспорте). Однако в большинстве прикладных статистических исследований нет никаких оснований считать, что при альтернативе функция распределения второй выборки лишь сдвигается, но не меняется каким-либо иным образом.

7. При всех своих недостатках критерий Вилкоксона прост в применении и часто позволяет обнаруживать различие групп (поскольку оно часто сводится к отличию  $a = P(X < Y)$  от  $1/2$ , т.е. медианы разности  $X - Y$  от 0). Приведенные здесь критические замечания не следует понимать как призыв к полному отказу от использования критерия Вилкоксона [9]. Однако для проверки гипотезы однородности в случае альтернативы общего вида можно порекомендовать состоятельные критерии, в частности, критерии Смирнова и типа омега-квадрат (Лемана – Розенблатта) [10].

8. В литературе по прикладным статистическим методам соседствуют два стиля изложения. Один из них в соответствии с математико-статистической теорией исходит из формулировок нулевой и альтернативных гипотез (или описания набора гипотез, из которого надо выбрать наиболее адекватную), для проверки которых строятся те или иные критерии. При другом стиле изложения упор делается на алгоритмическое описание критериев для проверки тех или иных гипотез, а об альтернативах даже не упоминается.

Например, в литературе по математической статистике часто говорится, что для проверки нормальности используются критерии асимметрии и эксцесса (они описаны, например, в главе 2.3 и в лучшем справочнике 1960 – 1980-х годов [3, табл. 4.7]). Однако эти критерии позволяют проверять некоторые соотношения между моментами распределения, но отнюдь не являются состоятельными критериями нормальности (не все отклонения от нормальности обнаруживают). Впрочем, для прикладной статистики эти критерии большого практического значения в настоящее время не имеют, поскольку заранее известно, что распределения конкретных технических, экономических, медицинских и иных статистических данных скорее всего отличны от нормальных [6].

Констатируем, что недостатки критерия Вилкоксона не являются исключением. Мощность ряда иных популярных в прикладной математической статистике критериев заслуживает тщательного изучения, при этом заранее можно сказать, что зачастую они не позволяют проверять те гипотезы, с которыми традиционно связаны. При применении подобных критериев к анализу реальных данных необходимо тщательно взвешивать их достоинства и недостатки.

### Литература

1. Гаек Я., Шидак З. Теория ранговых критериев / Пер. с англ. - М.: Наука, 1971. – 376 с.
2. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, Омега-квадрат и ошибки при их применении / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 647 – 675. – IDA [article ID]: 0971403047. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>
3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
4. Холлендер М., Вульф Д. Непараметрические методы статистики. – М.: Финансы и статистика, 1983. – 518 с.
5. Смолянский М.Л. Таблицы неопределенных интегралов. – М.: ГИФМЛ, 1961. – 108 с.
6. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. Диагностика материалов. 1991. Т.57. №7. С. 64 – 66.
7. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.
8. Закс Л. Статистическое оценивание. – М.: Статистика, 1976. – 598 с.
9. Орлов А.И. Какие гипотезы можно проверять с помощью двухвыборочного критерия Вилкоксона? // Заводская лаборатория. Диагностика материалов. 1999. Т.65. №1. С. 51-55.
10. Орлов А.И. Состоятельные критерии проверки абсолютной однородности независимых выборок // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №11. С. 66 – 70.

### References

1. Gaek Ja., Shidak Z. Teorija rangovyh kriteriev / Per. s angl. - M.: Nauka, 1971. – 376 s.
2. Orlov A.I. Neparametricheskie kriterii soglasija Kolmogorova, Smirnova, Omega-kvadrat i oshibki pri ih primenenii / A.I. Orlov // Politematicheskij setevoy jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. –

№03(097). S. 647 – 675. – IDA [article ID]: 0971403047. – Rezhim dostupa: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>

3. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoy statistiki. – M.: Nauka, 1983. – 416 s.
4. Hollender M., Vul'f D. Neparametricheskie metody statistiki. – M.: Finansy i statistika, 1983. – 518 s.
5. Smoljanskij M.L. Tablicy neopredelennyh integralov. – M.: GIFML, 1961. – 108 s.
6. Orlov A.I. Chasto li raspredelenie rezul'tatov nabljudenij javljaetsja normal'nym? // Zavodskaja laboratorija. Diagnostika materialov. 1991. T.57. №7. S. 64 – 66.
7. Nikitin Ja.Ju. Asimptoticheskaja jeffektivnost' neparametricheskikh kriteriev. - M.: Nauka, 1995. - 240 s.
8. Zaks L. Statisticheskoe ocenivanie. – M.: Statistika, 1976. – 598 s.
9. Orlov A.I. Kakie gipotezy mozno proverjat' s pomoshh'ju dvuhvyborochnogo kriterija Vilkoksona? // Zavodskaja laboratorija. Diagnostika materialov. 1999. T.65. №1. S. 51-55.
10. Orlov A.I. Sostojatel'nye kriterii proverki absoljutnoj odnorodnosti nezavisimyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. 2012. T.78. №11. S. 66 – 70.