

# Статистическая обработка результатов фармакологических экспериментов, измеренных в порядковых и количественных шкалах, при невозможности анализа с помощью параметрических методов

Цорин И. Б.

ФГБНУ «НИИ фармакологии имени В.В. Закусова», Москва, Россия

**Аннотация.** В статье рассматриваются описательная (дескриптивная) статистика данных, измеренных в порядковых (ординальных) и количественных шкалах, и критерии для определения статистической значимости различий между выборками при невозможности анализа с помощью параметрических методов. Особое внимание уделяется проблеме множественных сравнений такого рода данных. Для каждого метода приводятся примеры обработки данных, получаемых в фармакологических исследованиях.

**Ключевые слова:** порядковые шкалы; количественные шкалы; фармакологические исследования; множественные сравнения; дескриптивная статистика

## Для цитирования:

Цорин И.Б. Статистическая обработка результатов фармакологических экспериментов, измеренных в порядковых и количественных шкалах, при невозможности анализа с помощью параметрических методов // *Фармакокинетика и фармакодинамика*. – 2020. – № 3. – С. 3–24. DOI: 10.37489/2587-7836-2020-3-3-24

## Statistical processing of pharmacological experiments results measured in ordinal and quantitative scales, if it is impossible to analyze using parametric methods

Tsorin IB

FSBI "Zakusov Institute of Pharmacology", Moscow, Russia

**Abstract.** The article discusses descriptive statistics of data measured in ordinal and quantitative scales, and criteria for determining the statistical significance of differences between samples when it is impossible to analyze using parametric methods. Special attention is paid to the problem of multiple comparisons of this type of data. For each method, examples of processing data obtained in pharmacological studies are given.

**Keywords:** ordinal scales; quantitative scales; pharmacological studies; multiple comparisons; descriptive statistics

## For citations:

Tsorin IB. Statistical processing of pharmacological experiments results measured in ordinal and quantitative scales, if it is impossible to analyze using parametric methods. *Farmakokinetika i farmakodinamika*. 2020;(3):3–24. DOI: 10.37489/2587-7836-2020-3-3-24

При проведении фармакологических исследований достаточно часто данные измеряют с помощью порядковых (ординальных) шкал, которые представляют собой разновидность качественных шкал. В порядковых шкалах измерения производятся в баллах, количестве плюсов (интенсивность окраски в гистологии) или в уровнях интенсивности процесса (лёгкая форма заболевания, средней тяжести, тяжёлая) [1, 2]. Иногда такие шкалы называют полуколичественными, однако такой термин является неприемлемым, так как это название не отражает их свойств. Исследователи, проводя статистическую обработку порядковых данных, часто не учитывают ограничения, которые имеет тот или иной статистический критерий. При этом ошибки допускаются уже в числовом описании полученных результатов. Точно такие же ошибки допускают исследователи, проводя обработку и описание количественных данных, в том случае, когда нельзя применить параметрические методы (t-критерий Стьюдента, дисперсионный анализ). Именно опи-

санию и обработке такого рода данных и посвящена настоящая работа.

## Свойства порядковых шкал и дескриптивная статистика

Порядковые (ординальные) шкалы, помимо распределения объектов по классам, позволяют упорядочить наблюдения, сравнивая их между собой в каком-то соотношении (лучше—хуже, больше—меньше). Например, степень тяжести патологического процесса, оцениваемая в баллах, интенсивность гистохимических реакций на срезах тканей, оцениваемая количеством плюсов, и т. д. Однако порядковые шкалы оценивают не дистанцию между классами, а только наличие разницы в выраженности процесса в наблюдениях. Поэтому над данными, описываемыми такого рода шкалами, даже если они имеют цифровое выражение, нельзя выполнять арифметические операции, так как они не являются числами в полном смысле этого

слова. Нельзя суммировать баллы, а следовательно, находить средний балл. Это связано с тем, что баллы — это всего лишь представленные в цифровой форме степени выраженности процесса у одного объекта по отношению к другим, при этом порой достаточно субъективные. Таким образом, в порядковой шкале, помимо частоты объектов в различных классах, можно определить ранг объекта [1, 2].

*Пример 1.* В экспериментах по изучению антиаритмического действия веществ во время окклюзии и реперфузии коронарной артерии у крыс животным присваивается определённое количество баллов в зависимости от тяжести аритмий. При ранжировании выборок (распределении объектов по нарастанию баллов) выявляют медиану баллов в каждой выборке и определяют в какой совокупности аритмии тяжелее (табл. 1).

Таблица 1

**Балльное шкалирование тяжести аритмий у крыс в условиях окклюзии и реперфузии коронарной артерии**

Table 1

**Ball scaling of the severity of arrhythmias in rats under conditions of occlusion and reperfusion of the coronary artery**

Виды аритмий	Количество баллов
Аритмии отсутствуют	0
Желудочковая экстрасистолия	1
Желудочковая тахикардия	2
Фибрилляция желудочков во время реперфузии коронарной артерии	3
Фибрилляция желудочков в течение окклюзии коронарной артерии	4

Как мы уже отметили выше, с данными, измеренными в порядковой шкале, никакие арифметические действия проводить нельзя, однако оценки, получаемые объектами изучения, можно ранжировать в порядке возрастания или убывания. В связи с вышесказанным для оценки математического ожидания (среднего) генеральной совокупности такого рода следует использовать выборочную медиану. Однако, если количество градаций порядковой шкалы менее 5, иногда следует перейти к категориальной (номинальной) шкале и указывать частоту градаций [2].

При проведении обработки количественных данных, имеющих распределение отличное от нормального и логнормального, исследователи часто описывают полученные результаты с помощью среднего арифметического, его стандартной ошибки и стандартного отклонения. Такой подход является ошибочным, так как эти показатели достаточно адекватно отражают математическое ожидание генеральной совокупности и вариацию показателя только в случае нормального распределения (случае логнормального распределения математическое ожидание генеральной совокупности отражает средняя геометрическая). В связи с изложенным, при описании такого рода данных следует

использовать те же статистики, что и в случае порядковых шкал.

*Медианой (Me) выборки называется значение варианты, делящей ранжированный несгруппированный ряд наблюдений на две равные части (приходящееся на середину ряда)* [3–5].

Если количество наблюдений в ряду нечётное, то медиана равна варианту, имеющей номер  $(n + 1)/2$ . Если  $n$  чётное, то медиана равна полусумме вариантов с номерами  $n/2$  и  $n/2 + 1$  [3–5].

*Пример 2.* На 7 гистологических срезах печени крыс с помощью порядковой шкалы оценивали выраженность жировой инфильтрации: 0 баллов — отсутствие жировых капель в поле зрения; 1 — одиночные мелкие капли жира; 2 — большое количество мелких капель жира; 3 — наличие одиночных крупных капель жира; 4 — большое количество крупных капель жира; 5 — разлитая жировая инфильтрация. При оценке срезов были получены следующие результаты: 0; 1; 3; 3; 4; 4; 4. Так как выборка имеет нечётное количество вариантов, то Me при ранжировании по возрастанию равна варианту с номером  $(n + 1)/2 = 4$ , следовательно  $Me = 3$ .

*Пример 3.* У 8 крыс измерили массу тела, которая оказалась равна 225; 230; 300; 325; 400; 420; 475 и 500 г. Необходимо вычислить медиану. Так как количество вариантов чётное ( $n = 8$ ), то медиана будет равна полусумме 4-й и 5-й вариант, т. е.  $Me = (x_4 + x_5)/2 = (325 + 400)/2 \approx 363$  г.

Для оценки вариации показателей, измеренных в порядковых шкалах, используют квантили (наиболее часто квартили).

*Квантилями называются статистики, отсекающие в пределах ряда определённую часть его членов* [4].

Иначе. Квантиль — значение, которое заданная случайная величина не превышает с фиксированной вероятностью.

К ним относятся: квартили — три значения признака, которые делят ранжированный ряд на четыре равные части (по 25 % значений в каждой), децили — 9 значений, делящих ряд на 10 частей, перцентили (процентили) — 99 значений, делящие ряд на 100 частей.

На практике чаще всего используют квантили  $P_3$ ;  $P_5$ ;  $P_{10}$ ;  $P_{25}$ ;  $P_{50}$ ;  $P_{75}$ ;  $P_{90}$ ;  $P_{95}$ ;  $P_{97}$ .  $P_{50}$  является вторым квартилем и равен медиане.  $P_{25}$  и  $P_{75}$  соответствуют первому (нижнему) и третьему (верхнему) квартилям, между которыми находится 50 % членов выборки. Именно нижний и верхний квартили и межквартильный размах используют в качестве показателей вариации порядковых признаков. Эти же показатели используют при оценке вариации количественных признаков, распределение которых значительно отличается от нормального и логнормального [4, 5].

Как и большинство разбираемых статистик, квантили в настоящее время рассчитывают с помощью пакетов статистических программ. Для вычисления вручную квантилей малых выборок, не сгруппиро-

ванных в вариационный ряд, используют следующий подход. Варианты сортируют в порядке возрастания. Весь объём выборки принимают за 100 % и определяют сколькоим процентам соответствует одна варианта. Затем строят таблицу кумуляции частот (%). После этого выбирают варианту, которая соответствует большей накопленной частоте, чем выбранный порядок квантиля, так, чтобы предыдущая была меньше. Именно большая из 2 вариант соответствует этому квантилю. Если накопленная частота меньшей из этих 2 вариант соответствует выбранному порядку квантиля, то этот показатель равен среднему арифметическому 2 вариант [4].

**Пример 4.** У 2 выборок крыс с 10-минутной окклюзией и последующей реперфузией коронарной артерии с помощью порядковой шкалы оценивали тяжесть аритмий (см. пример 1). В первой выборке было 11 крыс, во второй 8. Результаты измерений оказались следующими. I выборка: 0; 1; 2; 2; 3; 3; 3; 3; 4; 4; 4. II выборка: 0; 0; 2; 3; 3; 3; 3; 4. Необходимо рассчитать медианы, нижний и верхний квантили.

I выборка:  
I selection:

Тяжесть аритмий в баллах	0	1	2	2	3	3	3	3	4	4	4
%	9,09	18,18	27,27	36,36	45,45	54,54	63,63	72,72	81,81	90,90	100

Me = 3; P<sub>25</sub> = 2; P<sub>75</sub> = 4

II выборка:  
II selection:

Тяжесть аритмий в баллах	0	0	2	3	3	3	3	4
%	12,5	25,0	37,5	50,0	62,5	75,0	87,5	100

Me = 3; P<sub>25</sub> = 1; P<sub>75</sub> = 3

Рассчитав точечные характеристики изучаемой выборки, можно построить доверительный интервал, в котором с той или иной вероятностью находится генеральный параметр. Наиболее часто в качестве доверительных используют вероятности P<sub>1</sub> = 0,95, P<sub>2</sub> = 0,99. Это означает, что при оценке параметров генеральной совокупности по выборочным показателям вероятность ошибки составляет 5 % в первом случае и 1 % во втором. Выбор порога доверительной вероятности исследователь осуществляет, исходя из той ответственности, с какой делаются выводы о генеральных параметрах.

С доверительной вероятностью связан уровень значимости α, который равен 1-P. Геометрически эту величину можно представить как площадь под кривой распределения некоторой статистики, выходящую за пределы, которые ограничивают P % площади под этой кривой [4–6].

Для оценки доверительного интервала медианы выборка ранжируется в порядке возрастания. Если

объём выборки не превышает 50 вариант, то по таблице [2, 7, 8] в соответствии с n и уровнем доверительной вероятности (P = 1 – α) определяют номера вариант, которые являются границами доверительного интервала медианы. Например, для выборки, состоящей из 10 вариант, границами 95 % доверительного интервала служат 2-я и 9-я варианты. Если n > 50, то доверительный интервал медианы определяется следующим образом:

$$x_k \leq Me \leq x_{n-k+1},$$

где: k = (n – z<sub>α/2</sub> \* √(n – 1))/2, z<sub>α/2</sub> – (1 – α/2) квантиль верхнего хвоста стандартного нормального распределения [2].

Например, при выборке из 100 вариант и 95 % доверительном интервале (α = 0,05) k = (100 – 1,96 \* √(100 – 1))/2 = 40,2, тогда x<sub>40</sub> ≤ Me ≤ x<sub>61</sub>.

### Проверка гипотез об однородности двух выборок, распределённых по закону, отличающемуся от нормального (логнормального), или измеренных в порядковых шкалах

В фармакологических исследованиях очень часто возникает необходимость оценить различия между выборками, в которых оцениваемые показатели измерены в порядковой шкале или полученные количественные результаты имеют распределение, значительно отличающееся от нормального или логнормального. В частности, такие выборки встречаются при изучении влияния фармакологических веществ на поведение животных.

Для обработки такого рода данных используют различные непараметрические методы, которые независимы от распределения. Блок-схема алгоритма выбора метода приведена на рисунке 1, объяснения к нему см. далее в тексте.

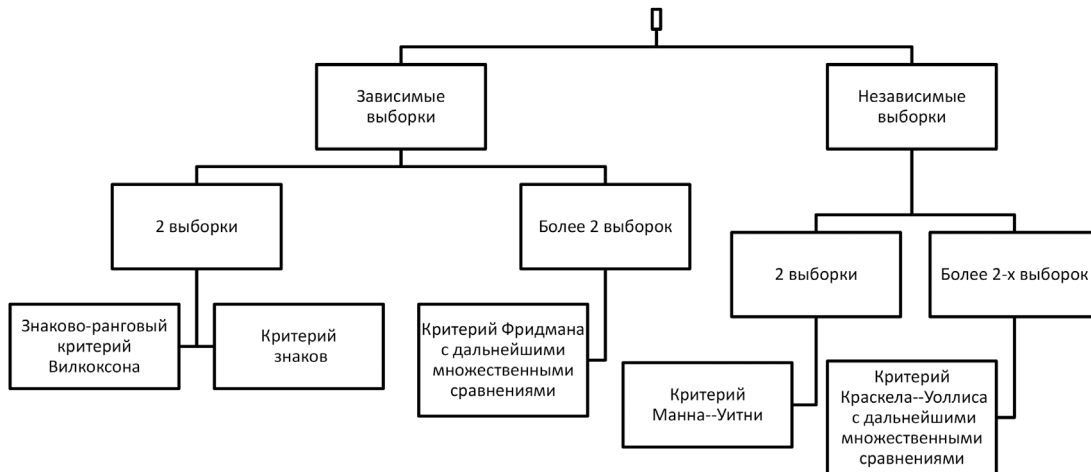
#### Сравнение двух независимых выборок

Одним из наиболее часто применяемых для сравнения двух независимых выборок методов является критерий Вилкоксона–Манна–Уитни. Критерий, как и многие непараметрические методы, является ранговым. Каждой варианте выборок присваивается определённый ранг, который соответствует её положению в общем ряду вариант обеих выборок после сортировки по возрастанию. Равным вариантам присваиваются средние значение тех рангов, которые были бы присвоены каждому из членов выборки при отсутствии совпадений. Сумма рангов в ряду равна N × (N + 1)/2.

#### Пример 5.

Вариант	0	1	2	2	4	5	5
Ранг	1	2	3,5	3,5	5	6,5	6,5

Проверим правильность ранжирования: сумма рангов R = 1 + 2 + 3,5 + 3,5 + 5 + 6,5 + 6,5 = 28; R = N × (N + 1)/2 = 7(7 + 1)/2 = 28, где N – количество вариант в выборке.



**Рис. 1.** Блок-схема выбора метода обработки данных, измеренных в порядковой шкале.  
**Figure 1.** Flowchart for selecting a method for processing data measured on an ordinal scale.

Критерий **Вилкоксона–Манна–Уитни** используется для оценки значимости различий между медианами совокупностей, варианты которых измерены в порядковых или количественных шкалах. Нулевая гипотеза  $H_0$  предполагает, что выборки извлечены из одной генеральной совокупности и различий между медианами не существует, т. е. изучаемое воздействие не вызывает сдвига ( $d = \mu_1 - \mu_2 = 0$ , где  $\mu_1$  и  $\mu_2$  медианы совокупностей). Альтернативная гипотеза  $H_1$  — выборки извлечены из разных генеральных совокупностей, медианы не равны.

Анализ данных с помощью критерия **Вилкоксона–Манна–Уитни** [6] проводится следующим образом: выборки  $X$  и  $Y$  объёмом  $n$  и  $m$  объединяют и сортируют по возрастанию, сохраняя при этом соответствие между отдельными вариантами и связанными с ними экспериментальными условиями; присваивают членам объединённой выборки соответствующие ранги; отделяют одну выборку от другой. Находят для каждой

из них сумму рангов  $\left( R_n = \sum_{i=1}^n r_i \text{ и } R_m = \sum_{j=1}^m r_j \right)$ . Затем

рассчитывают тестовые статистики (можно рассчитывать какую-либо одну, например, для выборки  $Y$ )

$$U_n = n \times m + n \times (n + 1)/2 - R_n, \quad (1)$$

$$U_m = n \times m + m \times (m + 1)/2 - R_m, \quad (2)$$

Эти два показателя связаны между собой следующим соотношением:

$$U_n = n \times m - U_m, \quad (3)$$

Рассчитанные статистики сравнивают с критическими значениями  $U_{кр.min}$  и  $U_{кр.max}$  при соответствующих  $n$ ,  $m$  и уровне значимости  $\alpha$ , которые приведены в соответствующих таблицах [6, 7].  $H_0$  отвергают в пользу односторонней  $H_1$   $d > 0$ , если  $U_m \leq U_{кр,\alpha,n,m,min}$ ; односторонней  $H_1$   $d < 0$ , если  $U_m \geq U_{кр,\alpha,n,m,max}$ ; двухсторонней  $H_1$   $d \neq 0$ , если  $U_m \geq U_{кр, \alpha/2, n, m, max}$  или  $U_m \leq U_{кр, \alpha/2, n, m, min}$ , в

противных случаях принимают  $H_0$ .  $U_n$  будет зеркально отражать положение  $U_m$ .

Этот критерий можно применять при  $n, m \geq 3$  или  $n = 2, m \geq 5$ . В том случае, если численность одной из выборок превышает максимальное  $n, m$  используемой таблицы, следует применить приближение для больших выборок.

Наиболее точной является аппроксимация, предложенная *Jman RL* [9]. Согласно этому приближению  $H_0$  отклоняется на уровне значимости  $\alpha$ , если  $|J| \geq J(\alpha)$  или  $|J| \geq J(\alpha/2)$  при односторонней и двухсторонней гипотезах, соответственно, где:

$$J = \frac{W}{2} \left[ 1 + \left( \frac{n+m-2}{n+m-1-W^2} \right)^{1/2} \right]; \quad J(\alpha) = \frac{z_\alpha}{2} + \frac{t_\alpha(v)}{2}, \quad (4)$$

где:  $z_\alpha$  —  $(1 - \alpha)$ -квантиль верхнего хвоста стандартного нормального распределения;  $t_\alpha(v)$  —  $\alpha$ -квантиль распределения Стьюдента с  $v = n + m - 2$  степенями свободы (значения квантилей для стандартных  $\alpha$  даны в соответствующих таблицах)

$$W = \frac{R_m - \frac{m(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \text{ или } W = \frac{R_n - \frac{n(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}}, \quad (5)$$

Если в двух анализируемых выборках есть совпадающие значения, то знаменатель статистики  $W$  умножают на следующую величину:

$$\sqrt{1 - \frac{\sum_{i=1}^k t_i(t_i^2 - 1)}{(m+n)(m+n-1)(m+n+1)}}, \quad (6)$$

где:  $k$  — общее число групп совпадающих вариантов;  $t_i$  — число совпавших величин в  $i$ -й группе (необходимо подчеркнуть, что поправка вносится только тогда, когда совпавшие величины принадлежат к разным



выборкам, т. е. группы совпадающих вариантов, целиком состоящие из элементов одной и той же выборки, на величину  $W$  не влияют).

**Пример 6.** На лабораторных мышах изучали влияние нового препарата на неспособность удерживать задние лапы под туловищем, вызываемую этиловым спиртом. Показатель оценивали в баллах табл. 2 [10]. Полученные данные представлены в табл. 3.

$H_0$ : выборки извлечены из одной генеральной совокупности, препарат не влияет на нарушения координации движений, вызываемые этанолом ( $d = 0$ ).

$H_1$ : выборки извлечены из разных генеральных совокупностей, препарат уменьшает тяжесть нарушений координации движений ( $d < 0$ ).

Уровень значимости  $\alpha = 0,05$ .

Таблица 2

Порядковая шкала неспособности удерживать задние лапы под туловищем

Table 2

Ordinal scale of inability to hold the hind legs under the body

Баллы	Характеристика
0	«Раскидывание» лап не наблюдается
1	Кратковременное эпизодическое «раскидывание» лап во время движения
2	Повторяющиеся «раскидывания» лап, способность двигаться по прямой линии
3	Постоянное «раскидывание» лап, ограниченная координация (часто движение по кругу)
4	Постоянное «раскидывание» лап, безуспешные попытки к движению
5	Постоянное «раскидывание» лап, отсутствие попыток к движению

Таблица 3

Table 3

Выборка	Неспособность удерживать задние лапы под туловищем, баллы
Контроль, $n = 6$	2; 3; 3; 4; 5; 5;
Опыт, $m = 7$	0; 1; 1; 2; 2; 3; 4

Для статистической обработки с помощью критерия Вилкоксона–Манна–Уитни объединяем выборки, сортируем варианты по возрастанию и присваиваем им ранги (табл. 4). Суммируем ранги по

Таблица 4

Table 4

Выборка	Варианта	Ранг	Выборка	Варианта	Ранг
Опыт	0	1	Контроль	3	8
Опыт	1	2,5	Опыт	3	8
Опыт	1	2,5	Опыт	4	10,5
Опыт	2	5	Контроль	4	10,5
Контроль	2	5	Контроль	5	12,5
Опыт	2	5	Контроль	5	12,5
Контроль	3	8			

выборкам:  $R_n = \sum_{i=1}^n r_i = 5 + 8 + 8 + 10,5 + 12,5 + 12,5 = 56,5$ ;

$R_m = \sum_{j=1}^m r_j = 1 + 2,5 + 2,5 + 5 + 5 + 8 + 10,5 = 34,5$ .

Откуда в соответствии с (1) и (2)  $U_n = 6 \cdot 7 + 6 \cdot (6 + 1) / 2 - 56,5 = 6,5$ ;  $U_m = 6 \cdot 7 + 7 \cdot (7 + 1) / 2 - 34,5 = 35,5$ . Следовательно,  $U_m = 35,5 > U_{кр. max, 0,05; 6; 7} = 34$ . Таким образом,  $H_0 d = 0$  можно отклонить в пользу альтернативной гипотезы  $d < 0$ .

**Пример 7.** На лабораторных крысах изучали влияние нового препарата на детоксикационную функцию печени, о которой судили по длительности гексеналового сна. Полученные данные представлены в табл. 5.

Таблица 5

Table 5

Выборка	Продолжительность гексеналового сна, мин
Контроль, $n = 6$	38,5; 39,3; 40,3; 42,1; 44,8; 45,0
Опыт, $m = 7$	34,1; 37,5; 38,0; 38,4; 39,0; 40,3; 41,0

$H_0$ : выборки извлечены из одной генеральной совокупности, препарат не влияет на продолжительность гексеналового сна ( $d = 0$ ).

$H_1$ : выборки извлечены из разных генеральных совокупностей, препарат уменьшает продолжительность гексеналового сна ( $d < 0$ ).

Уровень значимости  $\alpha = 0,05$ .

Для статистической обработки с помощью критерия Вилкоксона–Манна–Уитни объединяем выборки, сортируем варианты по возрастанию и присваиваем им ранги (табл. 6).

Таблица 6

Table 6

Выборка	Варианта	Ранг
Опыт	34,1	1
Опыт	37,5	2
Опыт	38,0	3
Опыт	38,4	4
Контроль	38,5	5
Опыт	39,0	6
Контроль	39,3	7
Контроль	40,3	8,5
Опыт	40,3	8,5
Опыт	41,0	10
Контроль	42,1	11
Контроль	44,8	12
Контроль	45,0	13

Суммируем ранги по выборкам:  $R_n = \sum_{i=1}^n r_i = 5 + 7 + 8,5 + 11 + 12 + 13 = 56,5$ ;  $R_m = \sum_{j=1}^m r_j = 1 + 2 +$

+ 3 + 4 + 6 + 8,5 + 10 = 34,5. Откуда в соответствии с (1) и (2)  $U_n = 6 \cdot 7 + 6 \cdot (6 + 1) / 2 = 56,5$ ;  $U_m = 6 \cdot 7 + 7 \cdot (7 + 1) / 2 = 34,5$ . Следовательно,  $U_m = 35,5 > U_{\text{кр. max}, 0,05; 6; 7} = 34$ . Таким образом,  $H_0: d = 0$  можно отклонить в пользу альтернативной гипотезы  $d < 0$ .

Так как критерий Вилкоксона–Манна–Уитни, как и большинство непараметрических критериев, основан на сравнении медиан, сравниваемые выборки описывают с помощью медиан, 25 и 75 % перцентилей (верхнего и нижнего квартилей). В то же время, если признак измерен в количественной шкале, то можно рассчитать сам сдвиг, вызванный изучаемым воздействием [3, 11].

Для выполнения этого расчёта необходимо построить  $n \times m$  разностей  $Y_j - X_i$ , где  $i = 1, \dots, n; j = 1, \dots, m$ . Полученный ряд сортируют по возрастанию и вычисляют оценку  $d$ , связанную со статистикой ранговой суммы Вилкоксона, которая равна медиане этого ряда:

$$d = \text{Me}\{(Y_j - X_i), i = 1, \dots, n; j = 1, \dots, m\}, \quad (7)$$

Если члены упорядоченного ряда представить как  $V^{(1)} \leq \dots \leq V^{(n \times m)}$ , то при  $m \times n$  нечётном ( $m \times n = 2k + 1$ )  $d = V^{(k+1)}$ . Если  $m \times n$  чётное ( $m \times n = 2k$ ), то  $\hat{d} = (V^{(k)} + V^{(k+1)}) / 2$ .

*Пример 8.* Используем данные примера 7, получим 42 упорядоченные разности  $(Y_j - X_i)$ , где  $j = 1, i = 1, \dots, 6$ , т. е.  $V^{(1)} \leq \dots \leq V^{(42)}$ : -10,9; -10,7; -8,1; -7,5; -7,3; -7,0; -6,8; -6,6; -6,4; -6,2; -6,0; -5,8; -5,2; -4,7; -4,5; -4,6; -4,4; -4,1; -4,0; -3,8; -3,7; -3,1; -2,8; -2,3; -1,9; -1,8; -1,8; -1,3; -1,3; -1,1; -1,0; -0,9; -0,5; -0,3; -0,1; 0,0; 0,5; 0,7; 1,0; 1,7; 1,8; 2,5.

Из этого ряда мы получаем:  $d = (V^{(21)} + V^{(22)}) / 2 = (-3,7 - 3,1) / 2 = -3,4$  мин составляет уменьшение гексеналового сна под влиянием препарата.

Для сдвига можно определить его свободный от распределения доверительный интервал [3, 12].

Для построения симметричного доверительного интервала для  $d$  с коэффициентом доверия  $1 - \alpha$  надо выполнить следующие действия.

1. Найти целое число

$$C_\alpha = U_{\text{кр. } \alpha/2, n, m, \min} + 1, \quad (8)$$

Тогда

$$P_0 = \{[n(n+1)/2 + C_\alpha] \leq R_n \leq [(n(2m+n+1)/2 - C_\alpha)]\} = 1 - \alpha \quad (m \geq n), \quad (9)$$

2. Необходимо провести сортировку разностей  $Y_j - X_i$ , где  $i = 1, \dots, n; j = 1, \dots, m$ , так, чтобы  $V^{(1)} \leq \dots \leq V^{(n \times m)}$ .

3. Границы доверительного интервала  $1 - \alpha$  будут равны:

$$d_L = V^{(C_\alpha)}; d_U = V^{(nm+1-C_\alpha)}, \quad (10)$$

Тогда для всех  $d$  получаем

$$P_d\{d_L < d < d_U\} = 1 - \alpha, \quad (11)$$

Для больших  $n$  и  $m$  целое число  $C_\alpha$  можно рассчитать с помощью аппроксимации стандартным нормальным распределением:

$$C_\alpha \approx \frac{mn}{2} - z_{(\alpha/2)} \left( \frac{mn(m+n+1)}{12} \right)^{1/2}, \quad (12)$$

где:  $z(\alpha/2) - (1 - \alpha/2)$  квантиль верхнего хвоста стандартного нормального распределения.

Если при расчёте получается не целое число, то его необходимо округлить до целого.

*Пример 9.* Возьмём данные о продолжительности гексеналового сна из примеров 8 и 7, рассчитаем 90 % доверительный интервал. В этом случае  $U_{\text{кр. } \alpha/2, n, m, \min} = U_{\text{кр. } 0,05, 6, 7, \min} = 8$ , тогда по формуле (8) рассчитаем  $C_\alpha = U_{\text{кр. } 0,05, 6, 7, \min} + 1 = 9$ . Следовательно,  $d_L = V^{(C_\alpha)} = V^{(9)}$ ;  $d_U = V^{(nm+1-C_\alpha)} = V^{(34)}$ . Из ряда 42 упорядоченных разностей  $Y_j - X_i$ , где  $j = 1, \dots, i = 1, \dots, 6$  ( $V^{(1)} \leq \dots \leq V^{(42)}$ : -10,9; -10,7; -8,1; -7,5; -7,3; -7,0; -6,8; -6,6; -6,4; -6,2; -6,0; -5,8; -5,2; -4,7; -4,5; -4,6; -4,4; -4,1; -4,0; -3,8; -3,7; -3,1; -2,8; -2,3; -1,9; -1,8; -1,8; -1,3; -1,3; -1,1; -1,0; -0,9; -0,5; -0,3; -0,1; 0,0; 0,5; 0,7; 1,0; 1,7; 1,8; 2,5), в соответствии с (10) находим, что  $d_L = -6,4$ ,  $d_U = -0,3$ . Таким образом, уменьшение изучаемого признака под влиянием препарата с вероятностью 0,90 находится в интервале от -6,4 мин до -0,3 мин.

Значительно реже по сравнению с критерием Вилкоксона–Манна–Уитни, используют **медианный критерий**, который обладает значительно меньшей мощностью [6].

С помощью медианного критерия проверяется нулевая гипотеза о том, отобраны ли две выборки из генеральных совокупностей с одинаковыми медианами. Этот критерий пригоден для обработки данных, измеренных в порядковой или количественных шкалах. Данный критерий достаточно прост, его можно описать следующим образом: все результаты измерений объединяются в один ряд, который сортируется по возрастанию или убыванию; находится общая медиана; для каждого условия определяется число отметок, находящихся выше или ниже общей медианы; эти данные записываются в таблицу сопряжённости признаков  $2 \times 2$ , и далее оценивается статистическая значимость различий с помощью метода точной вероятности Фишера или критерия  $\chi^2$  для 2 выборок (эти методы были разобраны в предыдущей статье [13]). Выборки следует описывать с помощью медиан и нижнего, и верхнего квартилей.

Для сравнения двух независимых выборок измеренных в любой количественной шкале может быть использован **критерий рандомизации** [6]. Этот критерий является одним из наиболее мощных непараметрических методов для сравнения двух выборок, измеренных в истинно количественных шкалах. К сожалению, он очень трудоёмок, это ограничивает его использование (в распространённых программных пакетах этот метод не разработан). Из-за трудоёмкости этот критерий может быть использован только для сравнения маленьких выборок (при количестве вариантов в выборках не более 10). В качестве  $H_0$  принимается

предположение о том, что анализируемые выборки извлечены из генеральных совокупностей с равными математическими ожиданиями и все видимые различия случайны;  $H_1$ -математические ожидания этих совокупностей действительно различны.

Суть метода заключается в следующем:

1. Вычисляют общее количество различных комбинаций, полученных данных в выборках. Число возможных исходов рассчитывают по формуле:

$$J = \frac{(n+m)!}{n!m!},$$

где:  $n$  и  $m$  — количество вариант в выборках;  $!$  — (факториал) указывает на необходимость перемножить все целые числа от 1 до  $n$  (по соглашению  $0! = 1$ ).

2. Путём перебора определяют количество ( $G$ ) столь же и ещё менее вероятных исходов распределения вариант по выборкам, чем наблюдаемый.

3. Рассчитывают вероятность получения исхода, столь же или ещё менее вероятного, чем наблюдаемый  $\left(P = \frac{G}{J}\right)$ .

Определение количества различных исходов может оказаться довольно утомительным занятием, однако при систематическом подходе эту процедуру можно минимизировать.

Например, можно использовать такой подход.

*Пример 10.* Пусть имеются 2 выборки А и Б, содержащие 3 и 4 варианты, соответственно:

А	Б
5	6
8	7
9	10
	12

1. Найдем количество возможных исходов распределения данных чисел по выборкам  $J = \frac{(n+m)!}{n!m!} = \frac{(3+4)!}{3!4!} = 35$ .

2. Находим сумму для каждой выборки:  $\Sigma A = 22$  и  $\Sigma B = 35$ .

3. Обратимся к выборке с меньшей суммой. Перечислим все комбинации по 3 из имеющихся чисел, сумма которых не будет превышать этой меньшей суммы:  $5 + 8 + 9 = 22$ ;  $5 + 6 + 9 = 20$ ;  $5 + 8 + 6 = 19$ ;  $5 + 6 + 10 = 21$ ;  $6 + 7 + 9 = 22$ ;  $6 + 8 + 7 = 21$ ;  $5 + 6 + 7 = 18$ ;  $5 + 7 + 9 = 21$ ;  $5 + 7 + 10 = 22$ ;  $5 + 8 + 7 = 20$ . Итак, получилось 10 комбинаций, сумма которых не превышает 22.

4. Находим вероятность получения суммы, меньшей или равной минимальной из наблюдаемых сумм:

$$P_{\Sigma A \leq 22} = \frac{G}{J} = \frac{10}{35} \approx 0,286.$$

Это — односторонняя вероятность наступления события столь же или более редкого, чем получение суммы 22 в комбинации 3 чисел из приведённых 7. Для определения двухсторонней вероятности наступления подобного события (при котором Б может иметь большую сумму) нужно увеличить одностороннее значение вероятности в 2 раза. Таким образом, вероятность наступления события столь же редкого,

Таблица 7

**Влияние вещества X на минутный объём сердца (мл/мин) в условиях экспериментального инфаркта миокарда**

Table 7

**Effect of substance X on the minute volume of the heart (ml/min) in experimental myocardial infarction**

Контроль $n = 3$	Опыт $m = 5$
35	39
38	44
43	45
	48
	50

как наблюдаемый результат, равна 0,572, т. е.  $H_0$  о том, что выборки взяты из одной совокупности, отвергнуть нельзя.

*Пример 11.* У крыс с экспериментальным инфарктом миокарда изучали влияние вещества на минутный объём сердца (мл/мин) в условиях 10-дневного введения. Данные представлены в таблице 7.

$H_0$ : Средние значения минутного объёма сердца в генеральных совокупностях, из которых извлечены выборки, совпадают ( $\mu_1 = \mu_2$ ).

$H_1$ : Средние значения минутного объёма сердца в изучаемых совокупностях различны ( $\mu_1 \neq \mu_2$ ).

Поскольку принять решение о нормальности распределения исходных совокупностей данных затруднительно, и показатель измерен в количественной шкале, подходящим критерием для решения вопроса о значимости различий является двухсторонний критерий рандомизации с уровнем значимости  $\alpha = 0,05$ .

1. Общее число возможных исходов при 2 группах, для которых  $n = 3$  и  $m = 5$ , равно:  $J = \frac{(3+5)!}{3!5!} = 56$ .

2. Находим сумму для каждой выборки:  $\Sigma K = 116$  и  $\Sigma O = 226$ .

3. Перечисляем все комбинации наблюдаемых чисел, которые приводят к получению суммы, не превышающей  $\Sigma K = 116$ . Эти исходы выглядят следующим образом:  $35 + 38 + 43 = 116$ ;  $35 + 38 + 39 = 112$ .

Таких комбинаций получилось 2.

4. Находим вероятность получения суммы, не превосходящую меньшую из наблюдаемых:

$$P_{\Sigma K \leq 116} = \frac{2}{56} \approx 0,0357.$$

5. Двухстороннее значение вероятности:

$$P_{\Sigma K \leq 116 \text{ или } \Sigma O \geq 226} = 4/56 \approx 0,0714.$$

Таким образом, мы не можем отвергнуть  $H_0$  при  $\alpha = 0,05$  и делаем вывод о том, что эффект вещества не доказан.

*Сравнение двух зависимых выборок*

Одним из наиболее часто применяемых для анализа зависимых выборок (с попарно связанными вариантами) методов является **независимый от распределения критерий знаковых рангов Вилкоксона** [3, 6].

Предположим, что у нас есть 2 равновеликие выборки случайных величин  $X$  и  $Y$ , являющихся одной и той же характеристикой изучаемых объектов до и после «обработки», измеренных в порядковой или количественных шкалах.

Тогда  $Z_i = Y_i - X_i = d + e_i, i = 1, \dots, n$ , где:  $e_i$  — ненаблюдаемые взаимно независимые случайные величины, принадлежащие непрерывной совокупности, симметричной относительно 0;  $d$  — неизвестный сдвиг, вызываемый «обработкой».

$H_0$  гласит:  $d = 0$ .

Для того, чтобы проверить эту гипотезу, необходимо провести сортировку по возрастанию абсолютных величин разностей  $|Z_i|$ . Если среди этих разностей есть  $Z_i = 0$ , их следует отбросить, уменьшив соответственно  $n$  до числа ненулевых значений. Оставшимся разностям следует присвоить ранги (в случае равенства значений используют средние ранги). Ранги положительных разностей суммируют:  $T^+ = \sum r_i^+$ .

$T^+$  сравнивают с критическим значениям  $t_{\alpha, n}$  при соответствующем  $n$  и уровне значимости  $\alpha$ , которые приведены в соответствующей таблице [3, 7, 14].  $H_0$  отвергают в пользу односторонней  $H_1 d > 0$ , если  $T^+ \geq t_{\alpha, n}$ ; односторонней  $H_1 d < 0$ , если  $T^+ \leq n(n+1)/2 - t_{\alpha, n}$ ; двухсторонней  $H_1 d \neq 0$ , если  $T^+ \geq t_{\alpha_1, n}$  или  $T^+ \leq n(n+1)/2 - t_{\alpha_2, n}$   $\alpha = \alpha_1 + \alpha_2$ , в противных случаях принимают  $H_0$ .

Этот критерий можно применять при  $n \geq 4$  (после исключения нулевых значений). В том случае, если  $n > 20$ , можно применить приближение для больших выборок.

Наиболее точной является аппроксимация, предложенная [15]. Согласно этому приближению,  $H_0$  отклоняется на уровне значимости  $\alpha$ , если  $|J| \geq J(\alpha)$  или  $|J| \geq J(\alpha/2)$  при односторонней и двухсторонней гипотезах, соответственно, где:

$$J = \frac{T^*}{2} \left[ 1 + \left( \frac{n-1}{n-(T^*)^2} \right)^{1/2} \right]; \quad J(\alpha) = \frac{z_\alpha}{2} + \frac{t_\alpha(v)}{2}, \quad (13)$$

где:  $z_\alpha$  —  $(1 - \alpha)$ -квантиль верхнего хвоста стандартного нормального распределения;  $t_\alpha(v)$  —  $\alpha$ -квантиль распределения Стьюдента с  $v = n - 1$  степенями свободы.

В случае наличия совпадающих вариантов вычисляют

$$T^* = \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1) - 0,5 \sum_{i=1}^g t_i(t_i-1)(t_i+1)}{24}}}, \quad (14)$$

где:  $g$  — общее число групп совпадающих вариантов;  $t_i$  — число совпавших величин в  $i$ -й группе. Если совпадающих вариант нет, выражение под знаком  $\Sigma$  равно 0.

**Пример 12.** На 8 крысах изучали влияние препарата А на атаксию, вызываемую этанолом. Атаксию оценивали по неустойчивости походки до и после введения препарата А. Атаксию оценивали с помощью

порядковой шкалы (табл. 8) [10]. Полученные данные представлены в табл. 9.

Таблица 8

Порядковая шкала оценки неустойчивости походки

Table 8

Ordinal scale for assessing gait instability

Баллы	Характеристика
0	Отсутствие покачивания
1	Покачивание при движении и на месте
2	Покачивания и падения
3	Падение из стороны в сторону, отсутствие движения
4	Падение, неспособность подняться

Таблица 9

Table 9

№	Оценка атаксии, баллы			Ранг
	До введения п-та	После введения п-та	Изменение (сдвиг), $Z_i$	
1	3	1	-2	4,5
2	4	3	-1	2
3	3	0	-3	6,5
4	4	4	0	-
5	3	2	-1	2
6	4	1	-3	6,5
7	3	4	+1	2
8	2	0	-2	4,5

$H_0$ : Сдвиг  $d = 0$ .

$H_1$ : Сдвиг  $d < 0$ .

Критический уровень значимости  $\alpha = 0,05$ .

Так как выборки измерены в порядковой шкале и являются зависимыми, то подходящим критерием для решения вопроса о статистической значимости сдвига является знаково-ранговый односторонний критерий Вилкоксона с уровнем значимости 0,05. В одном из оцениваемых случаев выраженность атаксии не меняется ( $Z_4 = 0$ ), в связи с этим  $n = 7$ .

$$T^+ = 2 = n(n+1)/2 - t_{0,023;7} = 28 - 26 = 2.$$

Таким образом, на уровне значимости 0,023 следует отвергнуть  $H_0$  и полагать, что препарат А статистически значимо уменьшает явления атаксии.

**Пример 13.** На 7 кошках изучали влияние пропранолола на подъём сегмента ST на эпикардиальной электрограмме, вызванный 3-минутной окклюзией передней нисходящей ветви левой коронарной артерии. Результаты представлены в табл. 10.

Так как выборки отличаются от нормального и логнормального распределений и являются зависимыми, то подходящим критерием для решения вопроса о статистической значимости сдвига является знаково-ранговый односторонний критерий Вилкоксона с уровнем значимости 0,05.

$$T^+ = 1,5 < n(n+1)/2 - t_{0,023;7} = 28 - 26 = 2.$$



Таблица 10

**Влияние пропранолола (0,5 мг/кг, в/в) на подъём сегмента ST на эпикардиальной электрограмме кошек, вызванный 3-минутной окклюзией коронарной артерии**

Table 10

**Effect of propranolol (0.5 mg / kg, IV) on ST segment elevation on the cat epicardial electrogram caused by 3-minute coronary artery occlusion**

№	Подъём сегмента ST, мв		Изменение (сдвиг) Z	Ранг
	До введения вещества	После введения вещества		
1	8	6	-2	3,5
2	10	9	-1	1,5
3	12	9	-3	5
4	13	8	-5	7
5	10	11	1	1,5
6	7	5	-2	3,5
7	14	10	-4	6

$H_0$ : Сдвиг  $d = 0$ .

$H_1$ : Сдвиг  $d < 0$ .

Критический уровень значимости  $\alpha = 0,05$ .

Таким образом, на уровне значимости 0,023 следует отвергнуть  $H_0$  и полагать, что препарат уменьшает подъём сегмента ST, вызванный окклюзией коронарной артерии.

Как и в случае анализа с помощью критерия Вилкоксона–Манна–Уитни при использовании критерия знаковых рангов Вилкоксона выборки описывают с помощью медиан и нижнего, и верхнего квартилей. Вместе с тем, при измерении показателя в количественных шкалах можно оценить непосредственно величину сдвига, вызываемого изучаемым воздействием [3, 11].

С этой целью строят ряд из  $M = n(n + 1)/2$  средних  $(Z_i + Z_j)/2$ ,  $i \leq j = 1, \dots, n$ . Полученный ряд сортируют по возрастанию. Вычисляют оценку сдвига, связанную со статистикой знаковых рангов Вилкоксона, которая равна медиане этого ряда:

$$\hat{d} = Me \left\{ \frac{Z_i + Z_j}{2}, i \leq j \right\}, \quad (15)$$

Медиану рассчитывают так, как было описано выше.

*Пример 14.* Возьмем данные из примера 11. Для них имеем  $M = n(n + 1)/2 = 7(7 + 1)/2 = 28$  упорядоченных значений  $(Z_i + Z_j)/2$ , т. е.  $W^{(1)} \leq \dots \leq W^{(28)}$ , представляют следующий ряд:  $-5,0; -4,5; -4,5; -4,0; -4,0; -3,5; -3,5; -3,0; -3,0; -3,0; -3,0; -2,5; -2,5; -2,5; -2,0; -2,0; -2,0; -2,0; -2,0; -1,5; -1,5; -1,5; -1,0; -1,0; -0,5; -0,5; 0,0; 1,0$ . Так как в ряду чётное количество членов  $M = 2k = 28$ , то:

$$\hat{d} = \frac{W^{(k)} + W^{(k+1)}}{2} = \frac{W^{(14)} + W^{(15)}}{2} =$$

$$= \frac{-2,5 - 2,0}{2} \approx -2,3 \text{ мв.}$$

Для сдвига можно определить его свободный от распределения доверительный интервал [3].

Для построения симметричного доверительного интервала для  $d$  с коэффициентом доверия  $1 - \alpha$  надо выполнить следующие действия:

1. Найти целое число

$$C_\alpha = \frac{n(n-1)}{2} + 1 - t_{\alpha/2, n}, \quad (16)$$

где:  $t_{\alpha/2, n}$  — критическое значение критерия Вилкоксона, тогда

$$P_0 \left\{ C_\alpha \leq T^+ \leq \left[ \frac{n(n-1)}{2} - C_\alpha \right] \right\} = 1 - \alpha. \quad (17)$$

2. Провести сортировку по возрастанию величин  $(Z_i + Z_j)/2$  при  $i \leq j = 1, \dots, n$ , получив, таким образом, ряд  $W^{(1)} \leq \dots \leq W^{(M)}$ , где  $M = n(n + 1)/2$ .

3. Границы доверительного интервала  $1 - \alpha$  будут равны:

$$d_L = W^{(C_\alpha)}; \quad d_U = W^{(M+1-C_\alpha)}. \quad (18)$$

Тогда для всех  $d$  получаем:

$$P_d \{ d_L < d < d_U \} = 1 - \alpha. \quad (19)$$

Для больших выборок  $C_\alpha$  можно рассчитать с помощью аппроксимации стандартным нормальным распределением:

$$C_\alpha \approx \frac{n(n+1)}{4} - z_{\alpha/2} \left[ \frac{n(n+1)(2n+1)}{24} \right]^{1/2}. \quad (20)$$

Это значение следует округлить до целого.

*Пример 15.* Возьмем данные примеров 11 и 12 и определим 95 % доверительный интервал. Рассчитаем  $C_\alpha = n(n + 1)/2 + 1 - t_{\alpha/2, n} = 7(7 + 1)/2 + 1 - t_{0,023,7} = 29 - 26 = 3$ .

Из примера 8 имеем упорядоченный ряд  $W^{(1)} \leq \dots \leq W^{(28)}$ :  $-5,0; -4,5; -4,5; -4,0; -4,0; -3,5; -3,5; -3,0; -3,0; -3,0; -3,0; -2,5; -2,5; -2,5; -2,0; -2,0; -2,0; -2,0; -2,0; -1,5; -1,5; -1,5; -1,0; -1,0; -0,5; -0,5; 0,0; 1,0$ .

Следовательно, из уравнений (18) имеем  $d_L = W^{(3)} = -4,5; d_U = W^{(26)} = -0,5$ .

Таким образом, уменьшение изучаемого признака под влиянием препарата с вероятностью 0,954 находится в интервале от  $-4,5$  мв до  $-0,5$  мв.

Для сравнения зависимых выборок признака, измеренного в порядковой или количественной шкале, может быть также использован критерий знаков. Однако этот метод, по сравнению с критерием знаковых рангов Вилкоксона, обладает значительно меньшей мощностью [3, 6].

Предположим, что у нас есть 2 равновеликие выборки случайных величин  $X$  и  $Y$ , являющихся одной

и той же характеристикой изучаемых объектов до и после «обработки». Признак измерен в порядковой или количественной шкале. В последнем случае выборки имеют распределение отличное от нормального и логнормального.

Тогда  $Z_i = Y_i - X_i = d + e_i$ ,  $i = 1, \dots, n$ , где:  $d$  — неизвестный сдвиг, вызываемый «обработкой»;  $e_i$  — ненаблюдаемые взаимно независимые случайные величины, принадлежащие непрерывной совокупности, имеющей медиану 0, так что:

$$P(e_i < 0) = P(e_i > 0) = 0,5, i = 1, \dots, n$$

$$H_0: d = 0.$$

Чтобы проверить гипотезу, подсчитаем статистику  $M$ , которая равна количеству положительных  $Z_i$ ,  $i = 1, \dots, n$ . Если среди  $Z_i$ ,  $i = 1, \dots, n$  есть нули, их необходимо отбросить, соответственно, уменьшив  $n$  до числа не нулевых значений.

$H_0$  против односторонней  $H_1 d > 0$  на уровне значимости  $\alpha$  отклоняют, если  $M \geq m_{\alpha, n, 1/2}$ , где константа  $m_{\alpha, n, 1/2}$  — верхняя  $\alpha$ -процентная точка биномиального распределения при объёме выборки  $n$  и  $p = 0,5$ , в противном случае нулевую гипотезу принимают.

$H_0$  против односторонней  $H_1 d < 0$  на уровне значимости  $\alpha$  отклоняют, если  $M \leq (n - m_{\alpha, n, 1/2})$ , в противном случае нулевую гипотезу принимают.

$H_0$  против двухсторонней  $H_1 d \neq 0$  на уровне значимости  $\alpha$  отклоняют, если  $M \geq m_{\alpha, n, 1/2}$  или  $M \leq (n - m_{\alpha, n, 1/2})$   $\alpha = \alpha_1 + \alpha_2$ ,  $m_{\alpha, n, 1/2}$  находим из таблицы биномиального распределения [3, 7].

*Пример 16.* Возьмем данные примера 12.  $H_0 d = 0$ ;  $H_1 d < 0$ , уровень значимости  $\alpha = 0,05$ , критерий односторонний. Так как  $Z_4 = 0$ , то  $n = 7$ . Только  $Z_7 > 0$ , следовательно,  $M = 1$ . Из таблицы получаем, что  $m_{0,0625,7,1/2} = 6$ , тогда  $M = 1 = n - m_{0,0625,7,1/2} = 7 - 6$ . Таким образом,  $H_0$  можно отвергнуть только на уровне значимости  $\alpha = 0,0625$ . Так как исходно уровень значимости был принят 0,05,  $H_0$  следует принять.

*Пример 17.* Возьмем данные из примера 13.  $H_0 d = 0$ ;  $H_1 d < 0$ , уровень значимости  $\alpha = 0,05$ , критерий односторонний. Видно, что только  $Z_5 > 0$ , следовательно,  $M = 1$ . Из таблицы получаем, что  $m_{0,0625,7,1/2} = 6$ , тогда  $M = 1 = n - m_{0,0625,7,1/2} = 7 - 6 = 1$ . Таким образом,  $H_0$  можно отвергнуть только на уровне значимости  $\alpha = 0,0625$ . Так как исходно уровень значимости был принят 0,05,  $H_0$  следует принять. Из обоих примеров хорошо видно, что критерий знаков обладает значительно меньшей мощностью, чем критерий знаковых рангов Вилкоксона.

Также как для предыдущего метода, для критерия знаков в случае регистрации признака в количественных шкалах существует связанная с ним оценка  $\hat{d}$ . Вычислить её достаточно просто. Необходимо упорядочить разности  $Z_i$  ( $i = 1, \dots, n$ ) по возрастанию и обозначить их так:  $Z^{(1)} \leq \dots \leq Z^{(n)}$ . Оценка  $\hat{d}$ , связанная со статистикой знаков, равна:

$$\hat{d} = Me\{Z_i, 1 \leq i \leq n\}. \quad (21)$$

Медиану определяют, как описано выше.

Для сдвига  $d$ , вызванного обработкой, можно вычислить доверительный интервал, основанный на критерии знаков [3].

Для построения симметричного двухстороннего интервала для  $d$ , с коэффициентом доверия  $1 - \alpha$ , надо выполнить следующие действия:

1. Найти целое число

$$C_\alpha = n + 1 - m_{\alpha/2, n, 1/2}, \quad (22)$$

тогда

$$P_0\{C_\alpha \leq B \leq (n - C_\alpha)\} = 1 - \alpha. \quad (23)$$

2. Используя упорядоченный по возрастанию ряд выборочных разностей  $Z^{(1)} \leq \dots \leq Z^{(n)}$  (см. предыдущий метод), найти

$$d_L = Z(C_\alpha) \text{ и } d_U = Z(n + 1 - C_\alpha), \quad (24)$$

Числа  $d_L$  и  $d_U$  таковы, что

$$P_0(d_L < d < d_U) = 1 - \alpha. \quad (25)$$

Следует отметить, что при использовании критерия знаков фактически происходит переход от порядковой или количественной шкалы к номинальной (дихотомической) шкале.

Одним из самых мощных непараметрических методов для сравнения двух зависимых выборок, измеренных в истинно количественных шкалах, является **критерий рандомизации** [6]. Однако он очень трудоёмок, в связи с этим его применяют только при анализе малых выборок ( $5 \leq n \leq 12$  при одностороннем уровне значимости  $\alpha = 0,05$ ).

Предположим, что у нас есть 2 равновеликие выборки случайных величин  $X$  и  $Y$ , являющихся одной и той же характеристикой изучаемых объектов до и после «обработки», имеющих распределение отличное от нормального, измеренных в истинно количественной шкале.

Нулевой гипотезой является предположение о том, что сдвиг  $d = 0$ , т. е. «обработка» не оказывает влияния на изучаемые объекты.

Процедура метода состоит в следующих действиях:

1. Вычитание результата первого измерения каждой однородной пары из соответствующего результата второго измерения ( $Z_i = Y_i - X_i$ ,  $i = 1, \dots, n$ );

2. Подсчёт алгебраической суммы найденных разностей;

3. Определение общего количества возможных исходов ( $J$ ), которое равно  $2n$ ;

4. Перечисление всех исходов ( $G$ ), которые приводят к получению результата, столь же или менее вероятного, чем наблюдаемый;

5. Расчёт суммарной односторонней вероятности столь же и менее вероятных исходов. Двухстороннюю вероятность можно получить путём удвоения односторонней.

*Пример 18.* На 6 крысах изучали влияние вещества А на частоту сокращений сердца. Данные представлены в таблице 11.

Таблица 11

Влияние вещества А на частоту сокращений сердца крыс

Table 11

The effect of compound A on heart rate of rats

№	Частота сокращений сердца, уд./мин		Изменение (сдвиг) $Z_i, i = 1, \dots, 6$
	До введения веществ	После введения вещества	
1	320	350	30
2	360	370	10
3	355	325	-30
4	310	350	40
5	365	400	35
6	400	445	45

$H_0$ :  $d = 0$ , вещество не влияет на частоту сокращений сердца.

$H_1$ :  $d > 0$ , вещество увеличивает частоту сокращений сердца.

Критерий односторонний, уровень значимости  $\alpha = 0,05$ .

Так как распределение выборок отличается от нормального, а признак измерен в количественной шкале, то может быть применён критерий рандомизации.

Определим общее количество возможных исходов:  $J = 2^6 = 64$ .

Находим  $\sum Z_i = 30 + 10 - 30 + 40 + 35 + 45 = 130$ .

Перечислим все столь же или менее вероятные исходы (приводящие к сумме равной наблюдаемой или большей её):

$$30 + 10 - 30 + 40 + 35 + 45 = 130;$$

$$-30 + 10 + 30 + 40 + 35 + 45 = 130;$$

$$30 - 10 + 30 + 40 + 35 + 45 = 170;$$

$$30 + 10 + 30 + 40 + 35 + 45 = 190.$$

Таких исходов оказалось:  $G = 4$ .

Рассчитаем одностороннюю вероятность:

$$P_{\sum Z_i \geq 130} = \frac{G}{J} = \frac{4}{64} = 0,0625.$$

Таким образом, на принятом уровне значимости  $H_0$  отвергнуть нельзя.

Перебор различных исходов может оказаться достаточно трудоёмким процессом, особенно когда  $H_0$  верна. Для экономии времени и сил можно определить пороговое значение, после достижения которого перебор вариантов можно прекратить и принять  $H_0$ . Пороговое значение рассчитывается следующим образом: общее количество возможных исходов  $J = 2n$  умножается на уровень значимости  $\alpha$  и, в случае односторонней гипотезы, округляется до целого числа с избытком. В случае двухсторонней гипотезы, полученное в результате умножения ( $J \cdot \alpha$ ) число делится на 2 и затем округляется. Например, если  $n = 9$  и  $\alpha = 0,05$ , то  $J = 2^9 = 512$ , тогда при односторонней гипотезе пороговое значение равно  $512 \cdot 0,05 = 25,6 \approx 26$ .

Сравнение одной выборки с известным математическим ожиданием генеральной совокупности

При проведении исследований иногда возникает необходимость сравнения, полученных результатов с референтными значениями. Сразу отметим, что референтных значений для показателей, измеряемых в порядковых шкалах, не существует, так как эти оценки крайне операторозависимы. Однако при измерении количественных показателей, распределение которых отличается от нормального и логнормального, сравнение с референтными значениями может быть проведено. В частности, такая задача возникает в токсикологии при оценке влияния веществ на биохимические и токсикологические показатели.

Одним из непараметрических критериев для решения вышеуказанной задачи является метод, основанный на статистике знаковых рангов [3].

Предположим, что у нас есть выборка  $n$  наблюдений  $Z_1, \dots, Z_n$ . Все  $Z$  взаимно независимы, измерены по количественной шкале, извлечены из симметричной относительно  $\theta$  совокупности, где  $\theta$  – неизвестное число.

Проверяем гипотезу  $H_0: \theta = \theta_0$  или  $\theta - \theta_0 = 0$ , где  $\theta_0$  некоторое референтное значение, являющееся оценкой генеральной совокупности.

Для проверки нулевой гипотезы модифицируем полученные наблюдения следующим образом:  $Z'_j = Z_j - \theta_0, j = 1, \dots, n$ . Если среди  $Z'_j$  встречаются нулевые значения, то их отбрасываем и объём выборки соответственно уменьшаем. Затем сортируем  $|Z'_j|$  по возрастанию и присваиваем им соответствующие ранги. Ранги положительных вариантов  $Z'_j$  суммируем:  $T^+ = \sum r_i^+$ .

$T^+$  сравниваем с критическим значениям  $t_{\alpha,n}$  при соответствующем  $n$  и уровне значимости  $\alpha$ , которые приведены в таблице распределения статистики знаковых рангов Вилкоксона [3, 7, 14].  $H_0$  отвергаем в пользу односторонней  $H_1 \theta - \theta_0 > 0$ , если  $T^+ \geq t_{\alpha,n}$ ; односторонней  $H_1 \theta - \theta_0 < 0$ , если  $T^+ \leq n(n+1)/2 - t_{\alpha,n}$ , двухсторонней  $H_1 \theta - \theta_0 \neq 0$ , если  $T^+ \geq t_{\alpha,n}$  или  $T^+ \leq n(n+1)/2 - t_{\alpha,n}$   $\alpha = \alpha_1 + \alpha_2$ , в противных случаях принимаем  $H_0$ .

Этот критерий можно применять при  $n \geq 4$  (после исключения нулевых значений). В том случае, если  $n \geq 20$  можно применить приближение для больших выборок.

Наиболее точной является аппроксимация предложенная [15]. Согласно этому приближению  $H_0$  отклоняется на уровне значимости  $\alpha$ , если  $|J| \geq J(\alpha)$  или  $|J| \geq J(\alpha/2)$  при односторонней и двухсторонней гипотезах, соответственно, где:

$$J = \frac{T^*}{2} \left[ 1 + \left( \frac{n-1}{n-(T^*)^2} \right)^{1/2} \right], \quad J(\alpha) = \frac{z_\alpha}{2} + \frac{t_\alpha(f)}{2}, \quad (26)$$

где:  $z_{\alpha}$  –  $(1-\alpha)$ -квантиль стандартного нормального распределения;  $t_{\alpha}(f)$  –  $\alpha$ -квантиль распределения Стьюдента с  $f = n - 1$  степенями свободы;

$$T^* = \frac{T^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1) - 0,5 \sum_{i=1}^g t_i(t_i-1)(t_i+1)}{24}}}, \quad (27)$$

где:  $g$  – общее число групп совпадающих вариантов;  $t_i$  – число совпавших величин в  $i$ -й группе. Если совпадающих вариант нет, выражение под знаком  $\Sigma$  равно 0.

Для того, чтобы рассчитать точечное значение  $\theta$  или его доверительный интервал, методы, предложенные при рассмотрении критерия знаковых рангов Вилкоксона (см. выше), следует применить к первоначальным вариантам  $Z_i, i = 1, \dots, n$ .

Таблица 12

Влияние вещества X на содержание гемоглобина в крови по отношению к референтному значению

Table 12

Effect of substance X on the content of hemoglobin in the blood in relation to the reference value

№	Содержание гемоглобина в крови женщин (Hb), г/л	Hb – Hb <sub>0</sub>	Ранг
1	113,0	-2	1,5
2	115,0	0	–
3	117,0	2	1,5
4	112,0	-3	3
5	110,0	-5	4,5
6	115,0	0	–
7	110,0	-5	4,5
8	109,0	-6	6
9	105,0	-10	7
10	100,0	-15	8

**Пример 19.** У 10 женщин, перенёвших сильное отравление неизвестным веществом, измеряли уровень гемоглобина в крови. Необходимо было выяснить, снижает ли это вещество уровень гемоглобина по сравнению с минимальным референтным значением ( $Hb_0 = \theta_0$ ), которое равно 115,0 г/л. Данные представлены в таблице 12.

$H_0: \theta = \theta_0 = 115,0$ .

$H_1: \theta < \theta_0$

Уровень значимости:  $\alpha = 0,05$ .

Так как в 2 случаях  $Hb - Hb_0 = 0$ , то  $n$  уменьшается до 8.  $T^+ = 1,5 < n(n+1)/2 - t_{0,012,8} = 36 - 34 = 2$ .

Таким образом,  $H_0$  может быть отвергнута на уровне значимости  $p = 0,012$  в пользу односторонней  $H_1: \theta < \theta_0 = 115,0$ . Следовательно, отравление данным веществом уменьшает содержание гемоглобина в крови ниже минимального референтного уровня.

Рассчитаем точечную оценку  $\theta$ . С этой целью создадим ряд величин  $(Hb_i + Hb_j)/2$ , где  $i \leq j, \dots, M, M = n(n+1)/2$ . Упорядочим этот ряд по возрастанию, т.е. получим  $W^{(1)} \leq \dots \leq W^{(55)}$ : 100,0; 102,5; 104,5; 105,0; 105,0; 105,0; 106,0; 106,5; 107,0; 107,5; 107,5; 107,5; 107,5; 108,5; 108,5; 109,0; 109,0; 109,5; 109,5; 110,0; 110,0; 110,0; 110,0; 110,5; 111,0; 111,0; 111,0; 111,0; 111,5; 111,5; 112,0; 112,0; 112,0; 112,5; 112,5; 112,5; 112,5; 113,0; 113,0; 113,5; 113,5; 113,5; 113,5; 114,0; 114,0; 114,5; 115,0; 115,0; 115,0; 115,0; 116,0; 116,0; 117,0.

$$\hat{\theta} = W^{(28)} = 111,0 \text{ г/л.}$$

Используя таблицу распределения статистики знаковых рангов Вилкоксона [3, 7, 14], определим доверительный интервал для  $\theta$ . Рассчитаем  $C_{\alpha} = C_{0,042} = n(n+1)/2 + 1 - t_{0,042,10} = 10 \cdot 11/2 + 1 - 45 = 11$ . Тогда  $\theta_L = W^{(11)} = 107,5$ ;  $\theta_U = W^{(M+1-C_{\alpha})} = W^{(45)} = 113,5$ .

Таким образом, 91,6 % доверительный интервал  $\theta_L \div \theta_U = 107,5 \div 113,5$ .

Одним из используемых в таких случаях методов является **критерий рандомизации** [6], который может быть применён только для анализа признаков, измеряемых в истинно количественных шкалах. При этом следует отметить, что объём выборки должен быть небольшим, в противном случае метод становится чрезвычайно трудоёмким. Суть метода заключается в следующем: определяют алгебраическую сумму разностей вариант выборки и референтного значения; с помощью перебора находят количество ( $G$ ) исходов столь же или менее вероятных, чем наблюдаемый исход. Определяют общее количество возможных исходов ( $J$ ), которое равно  $2N$ . Рассчитывают суммарную одностороннюю вероятность столь же или менее вероятных исходов  $\left( P = \frac{G}{J} \right)$ . Двухстороннюю вероятность получают удвоением односторонней.

**Пример 20.** У крыс линии SHR систолическое АД составляет в среднем 190 мм рт. ст. У 6 крыс были зарегистрированы следующие уровни систолического АД: 180, 181, 185, 188, 189, 193 мм рт. ст. Необходимо проверить нулевую гипотезу о том, что эта выборка извлечена из генеральной совокупности вышеупомянутой линии крыс.

$H_0: \mu = \mu_0$ .

$H_1: \mu \neq \mu_0$ .

Критерий двухсторонний. Уровень значимости:  $\alpha = 0,05$ .

Первым шагом анализа является нахождение разностей между измеренными уровнями АД и математическим ожиданием генеральной совокупности. Эти разностные отметки  $X_i (i = 1, \dots, 6)$  равны: -10, -9, -5, -2, -1, +3.

Затем находим сумму разностей:  $-10 -9 -5 -2 -1 + 3 = -24$ .



Находим общее количество возможных исходов, которое равно  $J = 2^N = 64$ .

Перечисляем всевозможные исходы, которые приводят к получению суммы, равной наблюдаемой или меньшей её (исходы получают за счёт изменения знаков у слагаемых):

$$\begin{aligned} -10 -9 -5 -2 -1 +3 &= -24 \\ -10 -9 -5 +2 -1 -3 &= -26 \\ -10 -9 -5 -2 +1 -3 &= -28 \\ -10 -9 -5 -2 -1 -3 &= -30 \\ -10 -9 -5 +2 +1 -3 &= -24. \end{aligned}$$

Таких исходов оказалось:  $G = 5$ . Рассчитываем вероятность получения исхода столь же или ещё более редкого, чем наблюдаемый. С этой целью количество найденных маловероятных исходов делим на количество возможных исходов:

$$P_{\Sigma X \leq -24} = G/J = 5/64 = 0,078.$$

Это одностороннее значение  $p$  – вероятность того, что сумма будет равна или меньше  $-24$ . Для получения двухстороннего значения  $p$  (вероятности того, что сумма будет меньше или равна  $-24$ , либо больше или равна  $24$ ) необходимо удвоить полученное значение:

$$P_{\Sigma X \leq -24 \text{ или } \geq 24} = 10/64 = 0,156.$$

Таким образом, мы принимаем нулевую гипотезу о том, что выборка извлечена из генеральной совокупности с  $\mu_0 = 190$ .

Рассмотренный критерий является одним из наиболее мощных, однако с увеличением размера выборки он становится чрезвычайно трудоёмким. Особенно утомителен этот критерий, когда  $H_0$  верна, так как в этом случае количество комбинаций, которые по крайней мере столь же маловероятны, как и наблюдаемый результат, велико. Для того, чтобы минимизировать процесс анализа, рассчитывают максимальное число исходов, которое позволяет отклонить  $H_0$ . Это пороговое значение для одностороннего критерия равно  $\alpha \cdot 2^N$ . Например, если критерий односторонний с уровнем значимости  $\alpha = 0,05$  и  $N = 10$ , то такое пороговое число исходов равно  $0,05 \cdot 2^{10} = 0,05 \cdot 1024 = 51,2$ . При двухстороннем критерии полученный результат необходимо разделить на 2. Как только пороговое значение будет достигнуто, можно прекратить перебор вариантов и принять нулевую гипотезу.

### **Методы сравнения нескольких выборок ( $k > 2$ ) порядковых данных**

Выше мы рассмотрели методы, позволяющие сравнить параметры двух статистических совокупностей. По своей сути в фармакологических исследованиях это выявление эффекта воздействия одного фактора, имеющего 2 уровня (градации). Однако в большинстве исследований действующих факторов и их градаций бывает, как правило, больше. Очень часто при такой ситуации исследователи проводят попарное сравнение большого количества групп с помощью методов, предназначенных для оценки различий 2 групп. Такой

подход является грубой ошибкой, так как в этом случае вступает в силу эффект множественных сравнений.

Рассмотрим пример. Исследовали влияние соединений А и В на тяжесть аритмий, вызываемых 7-минутной окклюзией и последующей реперфузией коронарной артерии. Тяжесть нарушений ритма оценивали в баллах. Эксперимент проводили на 3 группах крыс: I за 2 минуты до окклюзии коронарной артерии вводили вещество А; II — вещество Б; III — эквивалентный объём растворителя. Статистическая обработка с помощью критерия Вилкоксона–Манна–Уитни показала, что в каждом из трёх сравнений  $p < 0,05$ , т. е. вероятность ошибки I рода менее 5%. Однако это не совсем так. В действительности, так как мы допускаем 5% процентную ошибку в каждом из сравнений, вероятность ошибки будет значительно больше. В общем случае эта вероятность равна:  $P^* = 1 - (1 - p)k$ , где  $k$  — количество сравнений,  $p$  — принятый нами критический уровень значимости. При небольшом количестве сравнений  $P^* \approx pk$ . Следовательно, в нашем случае вероятность ошибиться хотя бы в 1 из сравнений составляет не многим менее 15% [5].

При проведении статистического анализа нескольких выборок, измеренных с помощью количественных шкал, обычно используют дисперсионный анализ с последующей обработкой каким-либо критерием множественных сравнений. Однако этот метод имеет определённые ограничения: данные должны иметь нормальное или близкое к нему распределение (некоторые отклонения от нормальности не являются критическими), дисперсии выборок должны быть гомогенны. Если указанные допущения не выполняются или показатель измерен с помощью порядковой шкалы, используют непараметрические аналоги дисперсионного анализа.

### *Однофакторный непараметрический анализ по Краскелу–Уоллису*

Аналогом однофакторного дисперсионного анализа является свободный от распределения критерий Краскела–Уоллиса, представляющий собой обобщение критерия Вилкоксона–Манна–Уитни [2, 3, 6]. При анализе полученных данных этим критерием проверяют нулевую гипотезу о том, что эффекты нескольких выборок, содержащих по  $n_1, n_2, \dots, n_m$  ( $m \geq 3$ ) вариант в каждой, равны против альтернативы о том, что не все обработки обладают одинаковым действием. При этом предполагается, что все случайные ошибки независимы и извлечены из одной непрерывной совокупности.

Алгоритм обработки данных с помощью критерия Краскела–Уоллиса выглядит следующим образом:

1. Объединить все выборки и проранжировать все

$N = \sum_{j=1}^m n_j$  наблюдений от меньшего к большему. Пусть

$r_{ij}$  обозначает ранг варианты  $X_{ij}$  в этой совместной ранжировке ( $i = 1, 2, \dots, n_j, j = 1, 2, \dots, m$ ).

$$2. \text{ Определить } R_j = \sum_{i=1}^{n_j} r_{ij}, R_{\bullet j} = \frac{R_j}{n_j}, R_{\bullet\bullet} = \frac{N+1}{2},$$

т. е.  $R_j$  — сумма рангов  $j$ -й обработки;  $R_{\bullet j}$  — средний ранг этой обработки;  $R_{\bullet\bullet}$  — средний ранг объединённой группы.

3. Вычислить показатель критерия Краскела–Уоллиса

$$H = \frac{12}{N(N+1)} \sum_{j=1}^m n_j (R_{\bullet j} - R_{\bullet\bullet})^2 = \left( \frac{12}{N(N+1)} \sum_{j=1}^m \frac{R_j^2}{n_j} \right) - 3(N+1). \quad (28)$$

4. На критическом уровне значимости  $\alpha$ :

отклонить  $H_0$ , если рассчитанный показатель

$$H \geq h_{(m, (n_1, n_2, \dots, n_m), \alpha)};$$

принять  $H_0$ , если рассчитанный показатель

$$H < h_{(m, (n_1, n_2, \dots, n_m), \alpha)};$$

где постоянная  $h_{(m, (n_1, n_2, \dots, n_m), \alpha)}$  удовлетворяет уравнению

$$P_0 \{ H \geq h_{(m, (n_1, n_2, \dots, n_m), \alpha)} \} = \alpha.$$

Однако достаточно полные таблицы критических значений критерия Краскела–Уоллиса очень громоздки и труднодоступны [3, 7], в связи с этим были разработаны различные аппроксимации этого критерия, которые применимы при  $n_i \geq 5$ . Мы приведём одну из них, разработанную [16]. В соответствии с этой аппроксимацией нулевая гипотеза равенства эффектов обработок отклоняется на уровне значимости  $\alpha$ , если  $J \geq J_\alpha$

$$J = \frac{H}{2} \left( 1 + \frac{N-m}{N-1-H} \right); \quad (29)$$

$$J_\alpha = \frac{1}{2} \left[ (m-1) F_{m-1; N-m; \alpha} + \chi_{m-1; \alpha}^2 \right],$$

где:  $F_{m-1; N-m; \alpha}$  и  $\chi_{m-1; \alpha}^2$  — соответственно, критические значения распределения Фишера–Снедекора и хи-квадрат с соответствующими степенями свободы.

При минимальном  $n_i \geq 20$  статистика  $H$  достаточно хорошо аппроксимируется распределением хи-квадрат с  $m-1$  степенями свободы.

Достаточно часто при проведении ранжирования величины вариант совпадают, в этом случае вариантам (наблюдениям) следует присваивать средние ранги. При этом статистику  $H$  следует заменить на  $H'$ , где:

$$H' = \frac{H}{1 - \left( \sum_{j=1}^g \frac{T_j}{[N^3 - N]} \right)}, \quad (30)$$

При этом  $g$  — число групп равных вариант,  $T_j = t_j^3 - t_j$ ;  $t_j$  — объём группы равных наблюдений

с номером  $j$ . Одиночные варианты рассматриваются как группы объёма 1. Таким образом, если равные наблюдения отсутствуют, то  $g = N$ ,  $t_j = 1$ ,  $T_j = 0$  при всех  $j = 1, \dots, N$ , тогда  $H' = H$  (Статистику  $H'$  рассчитывают в том случае, если совпадающие величины принадлежат разным выборкам).

Результаты представляют в виде медиан и нижнего, и верхнего квартилей. Указывают также показатель Краскела–Уоллиса ( $H$ ) и ошибку I рода ( $p$ ).

*Пример 21.* На беспородных мышах через 2 часа после отмены этанола при его положительном потреблении изучали влияние 2 новых веществ на судороги, вызываемые вертикальным подъёмом животных за хвост и вращением на  $180^\circ$  («хендлингом»). Выраженность судорог оценивали с помощью порядковой шкалы (табл. 13) [10]. Исследование проводили на 3 группах по 6 животных в каждой: 1-я — контроль; 2-я — вещество А; 3-я — вещество Б (табл. 14).

Таблица 13

Шкала оценки судорожных реакций, вызванных «хендлингом»

Table 13

The scale of evaluation of convulsive reactions caused by “handling”

Баллы	Характеристика поведения
0	Отсутствие судорог
1	Отсутствие судорог при подъёме за хвост
2	Отсутствие судорог при подъёме за хвост; тонические судороги после вращения на $180^\circ$
3	Судороги при подъёме за хвост и после вращения на $180^\circ$
4	Тонические судороги при подъёме за хвост
5	Тонико-клонические судороги при подъёме за хвост; запаздывание на 1-2 с
6	Сильные тонико-клонические судороги при подъёме за хвост с быстрым наступлением и длительной продолжительностью
7	Сильные тонико-клонические судороги до поднятия за хвост с быстрым наступлением и большой длительности

Таблица 14

Влияние веществ А и Б на судорожные реакции, вызываемые «хендлингом» в условиях абстиненции после отмены этанола

Table 14

Effect of substances A and B on convulsive reactions caused by “handling” in withdrawal conditions after ethanol withdrawal

Контроль		Вещество А		Вещество Б	
Баллы	Ранг	Баллы	Ранг	Баллы	Ранг
7	17,5	0	1,5	3	5,5
6	15	1	3	4	8,5
6	15	0	1,5	4	8,5
7	17,5	2	4	5	12
5	12	3	5,5	5	12
4	8,5	4	8,5	6	15

Таблица 15

## Влияние триметазида на количество заходов в светлый отсек приподнятого крестообразного лабиринта

Table 15

## Effect of trimetazidine on the number of visits to the light compartment of the raised cruciform labyrinth

Интактные крысы		Контрольные алкоголизованные крысы		Крысы, получавшие триметазидин	
Количество заходов	Ранг	Количество заходов	Ранг	Количество заходов	Ранг
1	8,5	0	3	2	13,5
1	8,5	0	3	3	17
0	3	2	13,5	2	13,5
1	8,5	0	3	5	20,5
5	20,5	1	8,5	4	19
2	13,5	1	8,5	7	22
1	8,5	0	3	3	17
3	17				

$H_0$ : Эффекты обработок равны (выборки принадлежат к одной генеральной совокупности).

$H_1$ : Эффекты обработок различаются (выборки принадлежат, по крайней мере, к 2 генеральным совокупностям).

Критический уровень значимости  $\alpha = 0,05$ .

Рассчитаем необходимые промежуточные показатели:  $R_1 = 3 + 4 \cdot 8,5 + 13,5 + 17 + 20,5 = 88$ ;  $R_{..1} = 11$ ;  $R_2 = 4 \cdot 3 + 8,5 \cdot 2 + 13,5 = 42,5$ ;  $R_{..2} = 6,07$ ;  $R_3 = 2 \cdot 13,5 + 2 \cdot 17 + 19 + 20,5 + 22 = 122,5$ ;  $R_{..3} = 17,5$ ;  $R_{..} = 11,5$ .

Рассчитаем статистику Краскела–Уоллиса (28):

$$H = \frac{12}{N(N+1)} \sum_{j=1}^m n_j (R_{.j} - R_{..})^2 =$$

$$= \frac{12}{22(22+1)} (8(11 - 11,5)^2 + 7(6,07 - 11,5)^2 +$$

$$+ 7(17,5 - 11,5)^2) = 10,92.$$

Так как в полученных результатах имеется несколько групп равных (связанных) наблюдений, то следует рассчитать показатель (30):

$$H' = \frac{H}{1 - \left( \sum_{j=1}^g \frac{T_j}{N^3 - N} \right)} =$$

$$= \frac{10,92}{1 - \frac{(6^3 - 6) + (5^3 - 5) + (4^3 - 4) + (3^3 - 3) + (2^3 - 2)}{22^3 - 22}} =$$

$$= \frac{10,92}{1 - 0,04} = 11,37.$$

$H_0$ : Эффекты обработок равны (выборки принадлежат к одной генеральной совокупности).

$H_1$ : Эффекты обработок различаются (выборки принадлежат, по крайней мере, к 2 генеральным совокупностям)

Критический уровень значимости  $\alpha = 0,05$ .

Объединим выборки, отсортируем наблюдения ( $N = 18$ ) по возрастанию и присвоим ранги каждой variante:

0; 0; 1; 2; 3; 3; 4; 4; 4; 4; 5; 5; 5; 6; 6; 6; 7; 7

Рассчитаем необходимые промежуточные показатели:

$R_1 = 8,5 + 12 + 15 + 15 + 17,5 + 17,5 = 85,5$ ;  $R_{..1} = 14,25$ ;  $R_2 = 1,5 + 1,5 + 3 + 4 + 5,5 + 8,5 = 24$ ;  $R_{..2} = 4$ ;  $R_3 = 5,5 + 8,5 + 8,5 + 12 + 12 + 15 = 61,5$ ;  $R_{..3} = 10,25$ ;  $R_{..} = 9,5$ .

Рассчитаем статистику Краскела–Уоллиса (28):

$$H = \frac{12}{N(N+1)} \sum_{j=1}^m n_j (R_{.j} - R_{..})^2 =$$

$$\frac{12}{18(18+1)} (6(14,25 - 9,5)^2 + 6(4 - 9,5)^2 +$$

$$+ 7(10,25 - 9,5)^2) = 11,86.$$

Так как в полученных результатах имеется несколько групп равных (связанных) наблюдений, то следует рассчитать показатель (30):

$$H' = \frac{H}{1 - \left( \sum_{j=1}^g \frac{T_j}{N^3 - N} \right)} =$$

$$= \frac{11,86}{1 - \frac{(4^3 - 4) + (3^3 - 3) + (3^3 - 3) + (2^3 - 2)}{18^3 - 18}} =$$

$$= \frac{11,86}{1 - 0,02} = 12,10.$$

Согласно таблице распределения статистики Краскела–Уоллиса, получаем, что  $H' = 12,10 > H_{3; (6,6,6); 0,001} = 10,889$ . Следовательно, мы отвергаем гипотезу  $H_0$  и принимаем альтернативную гипотезу  $H_1$  на уровне статистической значимости  $p < 0,001$ .

**Пример 22.** На крысах с развившейся алкогольной зависимостью в приподнятом крестообразном лабиринте изучали влияние триметазида на количество заходов в светлый отсек. Опыты были проведены на 3 группах животных:

1-я — интактные крысы ( $n = 8$ );

2-я — алкоголизованные контрольные животные ( $n = 7$ );

3-я — алкоголизованные крысы, получавшие триметазидин ( $n = 7$ ).

Данные представлены в таблице 15.

Поскольку разбираемый случай отсутствует в доступной таблице критических значений критерия Краскела–Уоллиса, то мы воспользуемся аппроксимацией Имана–Давенпорта (29):

$$J = \frac{H'}{2} \left( 1 + \frac{N - m}{N - 1 - H'} \right) = \frac{11,37}{2} \left( 1 + \frac{22 - 3}{22 - 1 - 11,37} \right) = 16,90;$$

$$J_{0,05} = \frac{1}{2} [(m - 1)F_{m-1; N-m; 0,05} + \chi_{m-1; 0,05}^2] = \frac{1}{2} [2F_{2; 19; 0,05} + \chi_{2; 0,05}^2] = \frac{1}{2} [2 \cdot 3,52 + 5,99] = 6,52.$$

Так как  $J > J_{0,05}$ , то мы отвергаем нулевую гипотезу на уровне  $p < 0,05$ .

В большинстве случаев исследователя-фармаколога интересует вопрос о том, эффекты каких обработок различаются между собой, а какие нет. Для решения этого вопроса используются различные непараметрические критерии множественных сравнений.

Так как таблицы для различных наборов данных  $(n_1, \dots, n_m)$  отсутствуют, да число таких возможных наборов чрезвычайно велико, то мы разберем один консервативный метод, связанный со статистикой Краскела–Уоллиса, и несколько приближений, использующих статистику, так или иначе, связанные с нормальным распределением.

После того как с помощью однофакторного анализа по Краскелу–Уоллису нулевая гипотеза была отвергнута, для попарного сравнения независимых выборок, для которых можно найти критические значения критерия Краскела–Уоллиса  $(h_{m, (n_1, n_2, \dots, n_m), \alpha})$ , может быть использован следующий критерий: при вероятности ошибочного решения, не большей, чем  $\alpha$ , принять решение о том, что нулевую гипотезу о равенстве эффектов двух обработок можно отвергнуть, если:

$$|R_{\bullet u} - R_{\bullet v}| \geq \{h_{m, (n_1, n_2, \dots, n_m), \alpha}\}^{0,5} \times \left[ \frac{N(N+1)}{12} \right]^{0,5} \left( \frac{1}{n_u} + \frac{1}{n_v} \right), \quad (31)$$

где  $R_{\bullet u}, R_{\bullet v}$  — средние ранги обработок,  $h_{m, (n_1, n_2, \dots, n_m), \alpha}$  критическое значение критерия Краскела–Уоллиса, которое можно найти в соответствующей таблице критических значений множественных сравнений, основанных на ранговых суммах Краскела–Уоллиса [3],  $N$  — общее количество вариантов в комплексе,  $n_u, n_v$  — объемы сравниваемых выборок,  $u = 1, \dots, m - 1, v = 2, \dots, m; u < v$ .

При выборках равных объемов может быть использован непараметрический вариант критерия мно-

жественных сравнений по Ньюмена–Кейлсу [5, 17]. Формула для этого критерия выглядит следующим образом:

$$q = \frac{R_v - R_u}{\sqrt{\frac{n^2 l (nl + 1)}{12}}}, \quad (32)$$

где  $R_u, R_v$  — суммы рангов двух сравниваемых выборок (выборки располагаются по нарастанию суммы рангов,  $u = 1, \dots, m - 1 < v = u + 1, \dots, m$ );  $n$  — объем каждой выборки;  $l$  — интервал сравнения (Интервал сравнения определяют следующим образом: при сравнении средних, стоящих на  $i$ -м и  $j$ -м месте в упорядоченном ряду,  $l = j - i + 1$ . Например, при сравнении 7-го и 2-го членов этого ряда  $l = 7 - 2 + 1 = 6$ );  $m$  — количество выборок. Вычисленное  $q$  сравнивают с критическим значением для бесконечного числа степеней свободы (таблица критических значений критерия Ньюмена–Кейлса [5]). Если рассчитанное  $q \geq q_{\infty, 1, \alpha}$ , то нулевую гипотезу о равенстве двух эффектов обработок отвергают на уровне значимости  $\alpha$ . Количество таких сравнений равно  $m(m - 1)/2$ .

Сравнение сумм рангов проводят в определенном порядке: сначала сравнивают наибольшую ( $m$ -ю) сумму рангов с наименьшей (1-й), затем  $m$ -ю со 2-й и т. д. до  $m - 1$ -й. Затем предпоследнее значение с 1-м, 2-м и т. д. Если какие-то выборки статистически значимо не различаются, то и все лежащие между ними также не различаются.

В том случае, когда необходимо сравнить контрольную выборку со всеми остальными, при равенстве их объемов можно использовать непараметрический вариант критерия Даннета [5, 17]. Формула его выглядит следующим образом:

$$q' = \frac{|R_{\text{contr}} - R_i|}{\sqrt{\frac{n^2 l (n_1 + 1)}{6}}}, \quad (33)$$

где  $R_{\text{contr}}$  — сумма рангов контрольной выборки;  $R_i$  — сумма рангов выборки, сравниваемой с контрольной;  $n$  — объем каждой выборки;  $l$  — число выборок. Вычисленное значение  $q'$  сравнивают с критическим значением для бесконечного числа степеней свободы (таблица критических значений критерия Даннета [5, 18]). Если рассчитанное  $q' \geq q'_{\infty, 1, \alpha}$ , то нулевую гипотезу о равенстве двух эффектов обработок отвергают на уровне значимости  $\alpha$ . Количество таких сравнений равно  $m - 1$ .

Для сравнения выборок как равного, так и разного объема может быть использован критерий Данна [3, 19]. Этот критерий выглядит следующим образом: принять альтернативную гипотезу (эффекты 2 обработок различаются) на критическом уровне значимости  $\alpha$ , следует в том случае, если:



$$|R_{\bullet u} - R_{\bullet v}| \geq z_{\alpha/[m(m-1)]} \sqrt{\frac{N(N+1)(n_u + n_v)}{12n_u n_v}}, \quad (34)$$

где  $R_{\bullet u}, R_{\bullet v}$  — средние ранги изучаемых выборок;  $z_{\alpha/[m(m-1)]}$  — квантиль стандартного нормального распределения;  $m$  — количество сравниваемых выборок;  $N$  — общее количество вариант в комплексе;  $n_u, n_v$  — объёмы сравниваемых выборок,  $u = 1, \dots, m-1, v = 2, \dots, m; u < v$ .

Если необходимо провести сравнение обработок с контролем, то критерий Данна приобретает следующий вид: принять альтернативную гипотезу (эффект обработки  $U$  отличается от контроля) на критическом уровне значимости  $\alpha$  в том случае, если:

$$|R_{\bullet u} - R_{\bullet contr}| \geq z_{\alpha/[2(m-1)]} \sqrt{\frac{N(N+1)(n_u + n_{contr})}{12n_u n_{contr}}}$$

(двухсторонняя гипотеза), (35)

$$|R_{\bullet u} - R_{\bullet contr}| \geq z_{\alpha/(m-1)} * \sqrt{\frac{N(N+1)(n_u + n_{contr})}{12n_u n_{contr}}}$$

(односторонняя гипотеза), (36)

где:  $R_{\bullet contr}$  — средний ранг контрольной выборки;  $n_{contr}$  — объём контрольной выборки.

**Пример 23.** Используем данные, представленные в примере 22.

$H_0$ : эффекты любых 2 обработок равны.

$H_1$ : эффекты некоторых обработок различаются между собой.

Критический уровень значимости  $\alpha = 0,05$ .

Так как выборки имеют различный объём, то используем критерий множественных сравнений Данна (34):  $z_{\alpha/[m(m-1)]} = z_{0,05/6} = 2,394$ ;

$$\begin{aligned} |R_{\bullet 1} - R_{\bullet 2}| &= |11 - 6,07| = \\ &= 4,93 < 2,394 \sqrt{\frac{22 \cdot 23 \cdot (8+7)}{12 \cdot 8 \cdot 7}} = 8,05; \end{aligned}$$

$$\begin{aligned} |R_{\bullet 1} - R_{\bullet 3}| &= |11 - 17,5| = \\ &= 6,5 < 2,394 \sqrt{\frac{22 \cdot 23 \cdot (8+7)}{12 \cdot 8 \cdot 7}} = 8,05; \end{aligned}$$

$$\begin{aligned} |R_{\bullet 2} - R_{\bullet 3}| &= |6,07 - 17,5| = \\ &= 11,43 > 2,394 \sqrt{\frac{22 \cdot 23 \cdot (7+7)}{12 \cdot 7 \cdot 7}} = 8,31. \end{aligned}$$

Таким образом, статистически значимо ( $p < 0,05$ ) различаются 2 и 3 обработки.

Теперь проведём анализ данных, приняв вторую выборку (контрольные алкоголизованные крысы)

как контрольную и применив одностороннюю гипотезу (36):  $z_{\alpha/(m-1)} = z_{0,05/2} = 1,96$ ;

$$\begin{aligned} |R_{\bullet 1} - R_{\bullet contr}| &= |11 - 6,07| = \\ &= 4,93 < 1,96 \sqrt{\frac{22 \cdot 23 \cdot (8+7)}{12 \cdot 8 \cdot 7}} = 6,59; \\ |R_{\bullet 3} - R_{\bullet contr}| &= |17,5 - 6,07| = \\ &= 11,43 > 1,96 \sqrt{\frac{22 \cdot 23 \cdot (7+7)}{12 \cdot 7 \cdot 7}} = 6,80. \end{aligned}$$

Таким образом, препарат статистически значимо ( $p < 0,05$ ) увеличивает количество заходов в светлый отсек крыс с развившейся алкогольной зависимостью.

**Пример 24.** Несколько изменим данные примера 22 так, чтобы выборки были одинакового объёма (табл. 16).

Таблица 16

Table 16

Интактные крысы		Контрольные алкоголизованные крысы		Крысы, получавшие триметазидин	
Количество заходов	Ранг	Количество заходов	Ранг	Количество заходов	Ранг
1	8,5	0	3	2	13,5
1	8,5	0	3	3	17
0	3	2	13,5	2	13,5
1	8,5	0	3	5	20
2	13,5	1	8,5	4	19
1	8,5	1	8,5	7	21
3	17	0	3	3	17

$H_0$ : эффекты любых 2 обработок равны.

$H_1$ : эффекты некоторых обработок различаются между собой.

Критический уровень значимости  $\alpha = 0,05$ .

Так как выборки имеют равный объём, то может быть использован критерий множественных сравнений Ньюмена–Кейлса.

Рассчитаем необходимые промежуточные показатели:  $R_1 = 3 + 4 \cdot 8,5 + 13,5 + 17 = 67,5$ ;  $R_2 = 4 \cdot 3 + 8,5 \cdot 2 + 13,5 = 42,5$ ;  $R_3 = 2 \cdot 13,5 + 2 \cdot 17 + 19 + 20 + 21 = 121$ .

Расположим полученные суммы в порядке нарастания:  $R_2 = 42,5$ ;  $R_1 = 67,5$ ;  $R_3 = 121$ .

Проведем сравнения (32):

$$\begin{aligned} q_{3,2} &= \frac{R_3 - R_2}{\sqrt{\frac{n^2 l (nl + 1)}{12}}} = \frac{121 - 42,5}{\sqrt{\frac{7^2 \cdot 3(7 \cdot 3 + 1)}{12}}} = \\ &= 4,78 > q_{\infty; 3; 0,05} = 3,314; \end{aligned}$$

$$q_{3,1} = \frac{R_3 - R_1}{\sqrt{\frac{n^2 l(nl + 1)}{12}}} = \frac{121 - 67,5}{\sqrt{\frac{7^2 \cdot 2(7 \cdot 2 + 1)}{12}}} = 4,83 > q_{\infty;2;0,05} = 2,772;$$

$$q_{1,2} = \frac{R_1 - R_2}{\sqrt{\frac{n^2 l(nl + 1)}{12}}} = \frac{67,5 - 42,5}{\sqrt{\frac{7^2 \cdot 2(7 \cdot 2 + 1)}{12}}} = 2,26 < q_{\infty;2;0,05} = 2,772.$$

Таким образом, на уровне значимости  $p < 0,05$  в 2 случаях из трёх следует отвергнуть нулевую гипотезу и принять  $H_1$  о том, что эффекты обработок неравны. Следовательно, можно полагать, что изучаемый препарат увеличивает количество заходов в светлый отсек животных с алкогольной зависимостью по сравнению как с алкоголизированными, так и с интактными крысами.

Теперь проведём анализ данных, приняв вторую выборку (контрольные алкоголизированные крысы) как контрольную. Применив критерий Даннета (33), получаем:

$$q_{contr,3} = \frac{|R_{contr} - R_3|}{\sqrt{\frac{n^2 l(nl + 1)}{6}}} = \frac{|42,5 - 121|}{\sqrt{\frac{7^2 \cdot 3(3 \cdot 7 + 1)}{6}}} = 3,38 > q_{\infty;3;0,05} = 2,21;$$

$$q_{contr,1} = \frac{|R_{contr} - R_1|}{\sqrt{\frac{n^2 l(nl + 1)}{6}}} = \frac{|42,5 - 67,5|}{\sqrt{\frac{7^2 \cdot 3(3 \cdot 7 + 1)}{6}}} = 1,08 < q_{\infty;3;0,05} = 2,21.$$

Таким образом, для первой группы нулевую гипотезу следует принять, а для третьей – отвергнуть на уровне  $p < 0,05$ .

*Непараметрический аналог дисперсионного анализа повторных измерений. Критерий Фридмана–Кендела–Бэбингтона Смита*

При проведении фармакологических исследований часто приходится сравнивать физиологические показатели одного и того же животного в нескольких временных точках или при нескольких различных воздействиях, т. е. сравнивать зависимые выборки. В таких случаях при соблюдении необходимых условий (количественная шкала, нормальное распределение данных) обработку результатов экспериментов проводят с помощью дисперсионного анализа повторных измерений. Однако, если наблюдаются значительные отклонения от нормальности распределения или для оценки признака использована порядковая шкала,

необходимо использовать критерий Фридмана–Кендела–Бэбингтона Смита [3, 5, 17], являющийся непараметрическим аналогом дисперсионного анализа повторных измерений.

Предположим, что мы имеем  $n$  объектов, у каждого из которых в  $m$  временных точках измерен некий показатель в количественной или порядковой шкале. Все случайные ошибки взаимно независимы и извлечены из одной непрерывной совокупности.  $\eta_j$  – неизвестные эффекты обработки ( $j = 1, \dots, m$ ).

Для проверки  $H_0: \eta_1 = \eta_2 = \dots = \eta_m$  против альтернативы о том, что не все  $\eta$  равны между собой сделаем следующее.

1. Проранжируем  $m$  измерений каждого объекта в порядке нарастания. Обозначим  $r_{ij}$  ранги  $X_{ij}$  совместно ранжировке  $X_{i1}, \dots, X_{im}$ , где:  $i = 1, \dots, n$  номера изучаемых объектов,  $j$  – номера обработок.

2. Рассчитаем:  $R_j = \sum_{i=1}^n r_{ij}$  – сумма рангов, присвоенных  $j$ -й обработке;  $R_{.j} = \frac{R_j}{n}$  – средний ранг  $j$ -й обработки;  $R_{..} = \frac{m+1}{2}$ .

3. Рассчитаем:

$$S = \frac{12n}{m(m+1)} \sum_{j=1}^m (R_{.j} - R_{..})^2 = \left[ \frac{12}{nm(m+1)} \sum_{j=1}^m R_j^2 \right] - 3n(m+1). \quad (37)$$

На уровне значимости  $\alpha$ : отклоним  $H_0$ , если  $S \geq s_{m,n,\alpha}$ ;

примем  $H_0$ , если  $S < s_{m,n,\alpha}$ ,

где:  $s_{m,n,\alpha}$  удовлетворяет уравнению  $P_0\{S \geq s_{m,n,\alpha}\} = \alpha$  (см. таблицу распределения статистики Фридмана [3, 5, 7]).

Если у объекта в разных точках встречаются одинаковые значения измерений, то им присваивается средний ранг. В этом случае  $S$  заменяется на  $S'$ :

$$S' = \frac{12 \sum_{j=1}^m (R_j - nR_{..})^2}{nm(m+1) - [1 / (m-1)] \sum_{i=1}^n \left\{ \left( \sum_{j=1}^{g_i} t_{i,j}^3 \right) - m \right\}}, \quad (38)$$

где:  $g_i$  – число групп равных (связанных) измерений у объекта  $i$ ;  $t_{i,j}$  – объём  $j$ -й связанной группы объекта  $i$ ; наблюдения неравные другому для этого объекта рассматриваются как группы объёмом 1. Если вообще нет связанных вариантов, то  $S'$  сводится к  $S$ .

Так как таблицы распределения статистики Фридмана составлены для относительно небольшого диапа-

зона значений  $m, n$ , то широко применяют различные аппроксимации.

При  $n \geq 13$  и  $m \geq 20$  статистика Фридмана хорошо аппроксимируется распределением  $\chi^2$  с числом степеней свободы  $m - 1$ . В этом случае  $H_0$  отвергают на критическом уровне значимости  $\alpha$  и принимают  $H_1$  в том случае, если  $S \geq \chi_{m-1, \alpha}^2$ .

Для  $7 \leq m \leq 19$  и  $n \geq 13$  можно рассчитать показатель  $F = \frac{(n-1) \cdot S}{n(m-1) - S}$ .  $H_0$  отвергают на критическом

уровне значимости  $\alpha$ , если  $F \geq F_{v_1, v_2, \alpha}$ , где  $F_{v_1, v_2, \alpha}$  –  $\alpha$ -квантиль распределения Фишера–Снедекора с  $v_1 = m - 1$  и  $v_2 = (m - 1)(n - 1)$  степенями свободы.

В других случаях наиболее эффективна аппроксимация, предложенная [20]. Для  $m \leq 7$  и  $n \geq 8$  статистику Имана–Давенпорта рассчитывают следующим образом

$$J = \frac{1}{2}[(m-1)F + S], \quad (39)$$

$H_0$  отклоняют на критическом уровне значимости  $\alpha$ , если  $J \geq J_\alpha$ , где:

$$J_\alpha = \frac{1}{2}[(m-1)F_{v_1, v_2, \alpha} + \chi_{(m-1), \alpha}^2], \quad (40)$$

и  $v_1 = m - 1, v_2 = (m - 1)(n - 1)$ .

Если  $m \geq 6$  и  $2 \leq v \leq 6$  и  $m = 5, n = 6$ , то вместо  $J$  и  $J_\alpha$  используют:

$$J^* = \frac{1}{2}[(m-1)(n-1)F + S]; \quad (41)$$

$$J_\alpha^* = \frac{1}{2}[(m-1)(n-1)F + S]. \quad (42)$$

**Пример 25.** На 5 крысах изучали влияние вещества X на неустойчивость походки, вызываемую этанолом. Нарушения оценивали в порядковой шкале (табл. 7 [10]) до и через 30 и 60 мин после введения вещества (табл. 17).

Таблица 17

**Влияние вещества X на неустойчивость походки, вызываемую этанолом**

Table 17

**Effect of substance X on gait instability caused by ethanol**

Крыса	До введения		30 мин после введения		60 мин после введения	
	Баллы	Ранг	Баллы	Ранг	Баллы	Ранг
1	4	3	0	1	1	2
2	3	2,5	1	1	3	2,5
3	3	2,5	2	1	3	2,5
4	4	2,5	3	1	4	2,5
5	2	2	2	2	2	2

$H_0$ : Нарушения походки во всех 3 временных точках не различаются, т. е. вещество не влияет на изучаемый показатель.

$H_1$ : Под влиянием изучаемого вещества нарушения походки изменяются.

Критический уровень значимости  $\alpha = 0,05$ .

Рассчитаем вспомогательные величины:

$$R_1 = 12,5; R_{.1} = 2,5; R_2 = 6; R_{.2} = 1,2; R_3 = 11,5; R_{.3} = 2,3; R_{..} = 2.$$

Так как у одного объекта в разных точках встречаются одинаковые оценки, используем формулу 38:

$$S' = \frac{12 \sum_{j=1}^m (R_j - nR_{..})^2}{nm(m+1) - [1/(m-1)] \sum_{i=1}^n \left\{ \left( \sum_{j=1}^m t_{i,j}^3 \right) - m \right\}}$$

$$S' = \frac{12(6,25 + 16 + 2,25)}{5 \cdot 3 \cdot 4 - [1/2] \cdot (6 + 6 + 6 + 24)} = 7,54;$$

$$S = 8,244 > s_{4;3;0;0,17} = 8,2.$$

Следовательно, нулевую гипотезу следует отклонить и принять альтернативную, т. е. влияние вещества на неустойчивость походки статистически доказано ( $p < 0,039$ ).

**Пример 26.** У 3 крыс с острым экспериментальным инфарктом миокарда регистрировали эхокардиограмму до введения вещества А через 5, 10 и 20 мин после инъекции. Рассчитывали фракции выброса (табл. 18).

Таблица 18

Table 18

Крыса	До введения		5 мин после введения		10 мин после введения		20 мин после введения	
	ФВ	Ранг	ФВ	Ранг	ФВ	Ранг	ФВ	Ранг
1	63	1	75	4	68	3	64	2
2	58	2	65	4	59	3	57	1
3	68	2	75	4	71	3	67	1

$H_0$ : Фракции выброса во всех 4 временных точках не различаются, т. е. вещество не влияет на изучаемый показатель.

$H_1$ : Под влиянием изучаемого вещества фракция выброса изменяется.

Критический уровень значимости  $\alpha = 0,05$ .

Предварительный анализ показал, что распределение вариант значительно отличается от нормального. В связи с этим для анализа эффективности действия вещества был выбран критерий Фридмана.

Рассчитаем вспомогательные величины:

$$R_1 = 5; R_{.1} = 1,7; R_2 = 12; R_{.2} = 4; R_3 = 9; R_{.3} = 3; R_4 = 4; R_{.4} = 1,3; R_{..} = 2,5.$$

Используем формулу 37:

$$S = \frac{12 \cdot 3}{4 \cdot (4 + 1)} (0,8^2 + 1,5^2 + 0,5^2 + 1,2^2) =$$

$$= \frac{36}{20} (0,64 + 2,25 + 0,25 + 1,44) = \frac{36 \cdot 4,58}{20} = 8,244;$$

$$S = 8,244 > s_{4;3;0,017} = 8,2.$$

Следовательно, нулевую гипотезу следует отклонить и принять альтернативную, т. е. влияние вещества на фракцию выброса статистически доказано ( $p < 0,017$ ).

Результаты обработки данных с помощью критерия Фридмана представляют в виде медиан выборок, нижнего и верхнего квартилей, одновременно указывают показатель Фридмана, величину ошибки I рода ( $p$ ).

После проведения анализа с помощью критерия Фридмана неизбежно возникает вопрос, в каких точках эффекты обработки статистически значимо отличаются друг от друга.

При сравнении всех эффектов обработки друг с другом возможно использовать следующий метод, основанный на ранговых суммах Фридмана:

1. Вычислить  $m(m-1)/2$  модулей ранговых сумм  $|R_i - R_j|$ ,  $i < j$ , где  $i = 1, \dots, m-1$ ;  $j = i+1, \dots, m$ .

2. На критическом уровне значимости  $\alpha$  отвергнуть нулевую гипотезу о равенстве эффектов обработки ( $\eta_i = \eta_j$ ) и принять альтернативную ( $\eta_i \neq \eta_j$ ), если  $|R_i - R_j| \geq r_{\alpha, m, n}$ , где  $r_{\alpha, m, n}$  удовлетворяет соотношению  $P_0\{|R_i - R_j| < r_{\alpha, m, n}\} = 1 - \alpha$  (значения  $r_{\alpha, m, n}$  см. таблицу критических значений для множественных сравнений, основанных на ранговых суммах Фридмана [3]).

Для попарного сравнения после обработки полученных данных с помощью критерия Фридмана можно также использовать непараметрическую модификацию критерия Ньюмена–Кейлса (формула 32) [5, 17].

Во многих исследованиях сравнивать попарно все эффекты обработки нет необходимости, достаточно сравнить контрольную выборку со всеми остальными. В этом случае можно использовать следующий алгоритм, основанный на сравнении ранговых сумм Фридмана (предположим, что контролем является первая выборка):

1. Вычислить  $m-1$  модулей разностей ранговых сумм  $|R_i - R_1|$ ,  $i = 2, \dots, m$ ,  $R_1$  — сумма рангов контрольной выборки,  $R_i$  — сумма рангов выборки, сравниваемой с контрольной.

2. При двухстороннем критерии следует отвергнуть нулевую гипотезу ( $\eta_1 = \eta_i$ ) и принять альтернативную ( $\eta_1 \neq \eta_i$ ), если  $|R_i - R_1| \geq r_{\alpha, m-1, n}^*$ , где  $r_{\alpha, m-1, n}^*$  удовлетворяет соотношению  $P_0\{|R_i - R_1| < r_{\alpha, m-1, n}^*, i = 2, \dots, m\} = 1 - \alpha$  (см. таблицу критических значений для двухсторонних множественных сравнений обработок с контролем, основанных на ранговых суммах Фридмана).

3. При использовании одностороннего критерия  $r_{\alpha, m-1, n}^*$  заменяют на  $r_{\alpha, m-1, n}^{**}$  (см. таблицу критических значений для односторонних множественных сравнений обработок с контролем, основанных на ранговых суммах Фридмана).

Для сравнения контрольной выборки с остальными после обработки данных с помощью критерия Фридмана можно также использовать непараметрический вариант критерия Даннета (формула 33) [5, 17].

*Пример 27.* Используем данные примера 26. Проведём попарное сравнение эффектов обработки, используя модули разностей ранговых сумм:

$$|R_1 - R_2| = |5 - 12| = 7 < r_{0,049;4;3} = 8;$$

$$|R_1 - R_3| = |5 - 9| = 4 < r_{0,049;4;3} = 8;$$

$$|R_1 - R_4| = |5 - 4| = 1 < r_{0,049;4;3} = 8;$$

$$|R_2 - R_3| = |12 - 9| = 3 < r_{0,049;4;3} = 8;$$

$$|R_2 - R_4| = |12 - 4| = 8 = r_{0,049;4;3} = 8;$$

$$|R_3 - R_4| = |9 - 4| = 5 < r_{0,049;4;3} = 8.$$

Таким образом, на основании проведённых расчётов можно сделать вывод о том, что статистически значимые различия на уровне  $p = 0,049$  наблюдаются только между второй и четвёртой временными точками.

Проведём такой же анализ с помощью критерия Ньюмена–Кейлса. Проведём сортировку ранговых сумм в порядке возрастания:  $R_4 = 4$ ;  $R_1 = 5$ ;  $R_3 = 9$ ;  $R_2 = 12$ . Рассчитаем показатели для критерия Ньюмена–Кейлса (формула 32):

$$q = \frac{R_4 - R_2}{\sqrt{\frac{nl(l+1)}{12}}} = \frac{12 - 4}{\sqrt{\frac{3 \cdot 4(4+1)}{12}}} =$$

$$= 3,58 < q_{0,05;4;\infty} = 3,633.$$

Так как остальные сравниваемые величины находятся между этими двумя, дальнейшие расчёты не нужны. Различия между временными точками, при обработке критерием Ньюмена–Кейлса статистически незначимы.

При таком плане эксперимента как в примере 24, исследователя, прежде всего, интересуют различия по сравнению с исходным уровнем. В связи с этим мы проведём именно такой анализ, используя двухстороннюю гипотезу:

$$|R_2 - R_1| = |12 - 5| = 7 = r_{0,046;3;3}^* = 7;$$

$$|R_3 - R_1| = |9 - 5| = 4 < r_{0,046;3;3}^* = 7;$$

$$|R_4 - R_1| = |4 - 5| = 1 < r_{0,046;3;3}^* = 7;$$

Таким образом, на 5-й минуте после введения изучаемый препарат статистически значимо ( $p < 0,0185$ ) увеличивает фракцию выброса.



Проведём такой же анализ с помощью критерия Даннета (формула 33):

$$q'_2 = \frac{|R_{contr} - R_2|}{\sqrt{\frac{n l(l+1)}{6}}} = \frac{|5 - 12|}{\sqrt{\frac{3 \cdot 4(4+1)}{6}}} =$$

$$= \frac{7}{3,16} = 2,214 < q'_{0,05;4;\infty} = 2,35;$$

$$q'_3 = \frac{|R_{contr} - R_3|}{\sqrt{\frac{n l(l+1)}{6}}} = \frac{|5 - 9|}{\sqrt{\frac{3 \cdot 4(4+1)}{6}}} =$$

$$= \frac{4}{3,16} = 1,265 < q'_{0,05;4;\infty} = 2,35;$$

$$q'_4 = \frac{|R_{contr} - R_4|}{\sqrt{\frac{n l(l+1)}{6}}} = \frac{|5 - 4|}{\sqrt{\frac{3 \cdot 4(4+1)}{6}}} =$$

$$= \frac{1}{3,16} = 0,316 < q'_{0,05;4;\infty} = 2,35.$$

Таким образом, критерий Даннета не показывает статистически значимой разницы между исходным уровнем фракции выброса и этим показателем в других временных точках после введения вещества. Можно полагать, что непараметрические варианты критериев Ньюмена–Кейлса и Даннета являются менее мощными, чем множественные сравнения, основанные

на ранговых суммах Фридмана. Непараметрические модификации критериев Ньюмена–Кейлса и Даннета следует применять в тех случаях, когда таблицы критерия Фридмана не содержат нужных показателей.

Итак, мы разобрали методы статистической обработки порядковых и количественных (при невозможности применения параметрических методов) данных при различном дизайне постановки эксперимента.

### Глоссарий

Квантиль в математической статистике — значение, которое заданная **случайная** величина не превышает с фиксированной **вероятностью**;

$z_\alpha$  — квантиль стандартного нормального распределения с критической вероятностью  $\alpha$ ;

$\nu$  — число степеней свободы;

Логнормальное распределение — распределение случайной величиной, при котором десятичные или натуральные логарифмы вариант (наблюдений) распределены по нормальному закону;

Медианой (Me) выборки называется значение варианты, делящей ранжированный ряд наблюдений на две равные части (приходящееся на середину ряда);

$H_0$  — нулевая гипотеза, которая предполагает, что сравниваемые величины не различаются;

$H_1$  — альтернативная гипотеза, которая предполагает наличие статистически значимых различий.

Ошибка I рода — вероятность того, что отвергнутая  $H_0$  является правильной.

### СВЕДЕНИЯ ОБ АВТОРАХ

**Цорин Иосиф Борисович**

*Автор, ответственный за переписку*

e-mail: tsorin@yandex.ru

ORCID ID: 0000-0002-3988-7724

SPIN-код: 4015-3025

д. б. н., в. н. с. лаборатории

фармакологического скрининга ФГБНУ

«НИИ фармакологии имени В.В. Закусова»,

Москва, Россия

**Tsorin Iosif B.**

*Corresponding author*

e-mail: tsorin@yandex.ru

ORCID ID: 0000-0002-3988-7724

SPIN code: 4015-3025

D. Sci. in Biology, Leading researcher of labora-

tory of pharmacological screening FSBI «Zakusov

Institute of Pharmacology», Moscow, Russia

### Литература / References

1. Халафян А.А. STATISTICA 6. Математическая статистика с элементами теории вероятностей. — М.: Бином; 2010. — 496 с. [Halafian A.A. STATISTICA 6. Matematicheskaya statistika s elementami teorii veroyatnostej. Moscow: Binom; 2010. (In Russ).].

2. Петри А., Сэбин К. Наглядная медицинская статистика. Издание 3-е. — М.: «ГЭОТАР-МЕДИА»; 2015. — 216 с. [Petrie A, Sabin C. Medical statistics at a glance. Third edition. — Wiley-Blackwell. A John Wiley & Sons, Ltd., Publication. 2009.].

3. Холлендер М., Вулф Д.А. Непараметрические методы статистики. — М.: «Финансы и статистика»; 1983. — 518 с. [Hollander MH, Wolfe DA. Nonparametric statistical methods. John Wiley and Sons / New York, London, Sydney, Toronto. 1973.].

4. Лакин Г.В. Биометрия. — М.: «Высшая школа»; 1990. — 352 с. [Lakin GV. Biometriya. Moscow: Vysshaya shkola; 1990. (In Russ).].

5. Гланц С. Медико-биологическая статистика. — М.: Практика; 1998. — 459 с. [Glantz St.A. Primer of biostatistics. Fourth Edition. McGraw-Hill: Health Professions Division. New York; 1994.].

6. Рунион Р. Справочник по непараметрической статистике. — М.: «Финансы и статистика»; 1982. — 198 с. [Runyon RP. Nonparametric statistics. A contemporary approach: Addison-Wesley Publishing Company; 1977.].

7. Ликеш И., Ляга И. Основные таблицы математической статистики. — М.: «Финансы и статистика»; 1985. — 356 с. [J. Likes, J. Laga. Zakladni Staticke Tabulky. Praha: 1978.].

8. Янко Я. Математико-статистические таблицы. — М.: Госстатиздат; 1961. — 243 с. [Yanko Ya. Mathematic-statistic tables. Praha; 1958.].

9. Iman RL. An approximation to the exact distribution of the wilcoxon-mann-whitney rank sum test statistic. *Communications in Statistics – Theory and Methods*. 1976;5(7):587-598.
10. Анохина И.П., Колик Л.Г. Методические основы поиска фармакологических средств для лечения алкоголизма и оценке их эффективности. В кн.: Руководство по проведению доклинических исследований лекарственных средств. Часть первая. 2013. – М.: ФГБУ «НЦЭСМП»; – С.310-333. [Anohina IP, Kolik LG. Methodological foundations of the search for pharmacological agents for the alcoholism treatment and evaluating their effectiveness. In book: Guidelines for conducting preclinical research of drugs. Pt. I. 2013. Moscow: FSBI “SCEEMP”; p. 310–333. (In Russ).].
11. Hodges JL, Lehmann EL. Estimates of location based on ranks tests. *Ann. Math. Statist.* 1963;34:598-611.
12. Moses LE. One sample limits of some two-sample rank tests. *J. Amer. Statist. Ass.* 1964;59:645-651.
13. Цорин И.Б. Статистическая обработка качественных (категориальных) данных в фармакологических исследованиях // *Фармакокинетика и фармакодинамика*. – 2019. – № 3. – С. 3–18. [Tsorin IB. Statistical processing of qualitative (nominal) data in pharmacological research. *Farmakokinetika i farmakodinamika*. 2019;(3):3–18. (In Russ).]. DOI: 10.24411/2587-7836-2019-10050
14. Большев Л.Н. Смирнов Н.В. Таблицы математической статистики. Издание 3-е. М.: «Наука»; 1983. – 416 с. [Bolshev LN, Smirnova NV. Tablicy matematicheskoy statistiki. Izdanie 3-e. Moscow: «Nauka»; 1983. (In Russ).].
15. Iman RL. Use of a t-statistic as an approximation to the exact distribution of the wilcoxon signed ranks test statistic. *Communications in Statistics*. 1974;3(8): 795-806.
16. Iman RL, Davenport JM. New approximation to the exact distribution of the Kruskal-Wallis test statistic. *Communications in Statistics – Theory and Methods*. 1976;5(14):1335–1348.
17. Кобзарь А.И. Прикладная математическая статистика. М.: Физматлит; 2006. – 816 с. [Kobzar AI. Applied mathematical statistics. Moscow: Fizmatlit; 2006. (In Russ).].
18. Dannet CW. New tables for multiple comparisons with a control. *Biometrics*. 1964;20:482–4491.
19. Dunn OJ. Multiple comparisons using rank sums. *Technometrics*. 1964;6:241–252.
20. Iman RL, Davenport JM. Approximation of the critical region of the fbietkan statistic. *Communications in Statistics – Theory and Methods*. 1980;9(6):571–595.