***Тема №9. Методы сравнения двух групп медицинских данных***

Часто встречающейся и достаточно сложной задачей является сравнение данных, полученных в процессе наблюдений или экспериментов. К примеру, это может быть ситуация сравнения данных обследования одних и тех же пациентов до терапии с аналогичными данными после нее.

Допустим, что удается заметить какие-либо численные различия в характеристиках сравниваемых данных. Первым делом возникает вопрос: возникли ли эти различия случайным образом, или они обусловлены воздействием определенного фактора. Последние будут систематически повторяться в дальнейшем при воспроизведении условий эксперимента или наблюдения – такие различия являются **статистически значимыми.**

**Выбор подходящего метода сравнения наборов данных определяется несколькими факторами: характером сравниваемых признаков (качественные, порядковые или количественные), числом сопоставляемых групп, зависимостью или независимостью выборок, а также видом распределения признака.**

В данной методичке пойдет речь лишь о **сравнении двух групп** данных.

Выборки являются **независимыми**, если набор объектов исследования в каждую из групп осуществлялся независимо от того, какие объекты исследования включены в другую группу. Так, в частности, происходит при рандомизации, когда распределение объектов происходит случайным образом. Примером сравнения независимых выборок может служить сопоставление данных анализа крови в группе пациентов с аналогичными показателями в группе здоровых.

Группы являются **зависимыми (связанными)** в динамических исследованиях, когда изучаются одни и те же объекты в разные моменты времени. Например, показатели анализа крови у одних и тех же пациентов до и после лечения.

Классические критерии (методы) сравнения делятся на две основные группы: **параметрические** и непараметрические.

Параметрические критерии – критерии, основанные на оценке параметров распределения, к которым относятся среднее арифметическое и среднее квадратическое отклонение. Они **применимы только при выполнении трех условий: 1) данные являются количественными; 2) данных достаточно много в каждой из сравниваемых групп (n>30); 3) численные данные каждой группы подчиняются нормальному распределению.**

Если нарушается выполнение хотя бы одного из вышеперечисленных трех условий, то использование параметрических критериев невозможно. В этом случае для сравнения данных используются непараметрические критерии.

**Непараметрические критерии** не основаны на оценке параметров распределения и вообще **не требуют, чтобы данные подчинялись какому-то определенному типу распределения.** Непараметрические критерии дают более грубые оценки, чем параметрические, но являются более универсальными. Параметрические методы более точны, но лишь в случае, если правильно определено распределение совокупности.

Перед тем как перейти к рассмотрению статистических критериев, введем понятия **нулевой** и **альтернативной гипотез**.

Нулевая гипотеза - , подразумевается допущение об отсутствии того или иного события или явления.

Альтернативная гипотеза - , подразумевается допущение о наличии события или явления.

Обе гипотезы обязательно должны иметь взаимоисключающеесодержание.

Нулевая гипотеза не может быть отвергнута, если ее вероятность окажется выше некого наперед заданного уровня α, достаточно близкого к 0, т.е. . Эта величина α носит название уровень значимости нулевой гипотезы.

Альтернативная гипотеза может быть принята лишь в том случае, если ее вероятность достигнет некого наперед заданного уровня β или превзойдет его, т.е. . Эта величина β – уровень доверительной вероятности.

 – это некий порог; он соответствует «уровням безошибочных прогнозов», т.е. вероятностям 0.95, 0.99 и 0.999 (область практически достоверных событий). Соответственно, α очерчивает область практически невозможных событий с порогами вероятностей 0.05, 0,01 и 0.001, соответственно.

α – это вероятность с которой мы допускаем вероятность ошибиться. В медицинских исследованиях вероятность ошибки не должна превышать 5%, т.е. α=0,05.

Поскольку и – взаимоисключающие гипотезы, то их суммарная вероятность равна единице. Следовательно, рост вероятности одной из гипотез автоматически приводит к снижению вероятности другой. Например, если , это означает то, что будет выполняться условие . И в этом случае нулевая гипотеза может быть отвергнута как событие практически невозможное, а альтернативная должна быть принята как событие практически достоверное. Если же , то . И в этой ситуации нулевая гипотеза не может быть отвергнута, а альтернативная не может быть принята.

Например, в процессе исследования ставится задача доказать наличие статистически значимых различий между результатами наблюдений в опытной и контрольной группах. Это значит, что данные, полученные при применении того или иного статистического критерия, должны позволить отвергнуть нулевую гипотезу об отсутствии указанных различий.

**Параметрические критерии**

Заключение о случайности или неслучайности различий между выборочными совокупностями при использовании параметрических критериев осуществляется на основании сравнения параметров распределений.

Как мы обсуждали ранее, в случае нормального распределения признака рассчитываются два основных параметра: среднее арифметическое значение и среднее квадратическое значение . Единственное, следует заметить, что эти параметры могут быть рассчитаны только в случае, если данные являются количественными (иначе параметры посчитать невозможно), и данных достаточно много (иначе точность определения величин очень мала).

В силу этого для использования параметрических критериев необходимо выполнение трех условий: 1) данные являются количественными; 2) данных достаточно много в каждой из сравниваемых групп (n>30); 3) численные данные каждой группы подчиняются нормальному распределению.

Для того, чтобы определить, является ли нормальным исследуемое распределение, используются критерии Шапиро-Уилка и Колмогорова-Смирнова.

Основными представителями параметрических критериев являются: критерий Стьюдента и критерий Фишера.

**Критерий Стьюдента (*t*-критерий)** – критерий, основанный на сравнении средних значений выборок.

**Двухвыборочный критерий Стьюдента** – метод сравнения, применяемый для проверки гипотезы о равенстве генеральных средних *двух независимых, несвязанных выборок*. Например, в таком случае могут быть контрольная и опытная группы, состоящие из разных пациентов, количество которых в группах может быть различно.

**Парный критерий Стьюдента** – метод сравнения, применяемый для сравнения *зависимых, связанных* данных, когда одна и та же группа объектов порождает числовой материал для проверки гипотез о средних. Например, измеряется содержание лейкоцитов у здоровых животных, а затем у тех же самых животных после облучения определенной дозой излучения.

**Критерий Фишера** основан на сравнении средних квадратических отклонений.

На практике часто встречается ситуация, когда факторы, влияющие на состояние изучаемых объектов, вызывают не только и даже не столько сдвиг этих характеристик на числовой оси, сколько усиливают или ослабляют их межиндивидуальное разнообразие.

Количественным индикатором этих изменений является различие средних квадратических отклонений. Однако, как всякая выборочная числовая характеристика, среднее квадратическое отклонение **–** величина случайная. Следовательно, наблюдаемое различие тоже может оказаться случайным.

**Критерий Стьюдента для независимых выборок**

Рассмотрим выборку объемом – пусть среднее вариант этой выборки равно М1, среднеквадратичное отклонение . И выборку объемом со средним М2, среднеквадратичным отклонением . При этом М1≠М2, а выборки подчиняются нормальному закону распределения. Обозначим разницу средних значений выборок .

Нам необходимо доказать, что разница между выборочными средними не могла быть получена случайным образом. Наблюдаемая разница средних выходит за пределы возможных случайных колебаний. Это можем утверждать если вероятность сравняется с некоторым порогом или превысит его.

α – это вероятность с которой мы допускаем вероятность ошибиться. В медицинских исследованиях вероятность ошибки не должна превышать 5%, т.е. α=0,05.

Проверка гипотез производится при помощи критерия Стьюдента, обозначаемого символом :

где - стандартная ошибка или мера отклонения наблюдаемой разницы выборочных средних от теоретически возможной, «генеральной». Формально величина *t* показывает, во сколько раз разница выборочных средних превышает свою собственную случайную вариацию.

**В случае независимых выборок** критерий *t* рассчитывается следующим образом:

При этом, как для первой, так и для второй выборки стандартная ошибка m рассчитывается по формуле:

Полученное значение критерия *t* сравнивают со стандартным табличным значением t-критерия Стьюдента для выбранного уровня значимости и числа степеней свободы .

**Если , нулевая гипотеза не может быть отвергнута**, и различие выборочных средних считается **«статистически незначимым»** (при этом обязательно указывается при каком уровне значимости это имеет место).

**Если , то это означает что величина *d* оказалась за пределами своих собственных случайных колебаний.** Такое различие называют **«статистически значимым».**

Достоверность в статистическом смысле обозначает, что полученное различие предсказуемо: при повторении эксперимента или наблюдения в тех же условиях оно будет воспроизводиться с вероятностью β или более.

Например, при сравнении двух групп критерий *tкр* равен 2, и, если полученное значение *t* больше 2, то различие статистически значимо и это можно утверждать с вероятностью безошибочного прогноза, равной 95% (при *tкр = 3* и более – с вероятностью безошибочного прогноза – 99%). Величина критерия менее 2 свидетельствует об отсутствии статистической значимости различий сравниваемых показателей.

**Критерий Стьюдента для зависимых выборок**

Рассмотрим теперь **случай зависимых выборок**. Это такие массивы данных, в которых каждому числовому значению одной выборки обязательно соответствует парное, причинно и следственно связанное, значение другой выборки. Это имеет место, когда какие-либо характеристики состояния организма регистрируются до некоторого воздействия на него и после или при разных вариантах воздействия, но обязательно у одних и тех же людей. Простейший пример, когда у некой группы людей измерили частоту пульса и величину артериального давления, потом попросили сделать, скажем, 20 приседаний, и провели те же измерения повторно. Понятно, что реакция сердечно-сосудистой системы каждого человека будет весьма индивидуальной, причем результаты измерений, полученные «после того», будут находиться в причинной и «исторической» связи с исходным состоянием «до того», т.е. в зависимости от них.

В этом случае при обнаружении ненулевой разницы выборочных средних результатов «до того» *и «после того» также рассчитывается критерий*

Однако, величина рассчитывается иным образом:

- это разница парных вариант: , квадрат которой, как видно из формулы, суммируется по всем парам.

Величина в данном случае зависит от того, насколько однородно будет изменяться измеряемая характеристика у разных объектов исследуемой группы. Действительно, если различие в каждой паре значений, полученных «до» и «после», будет нестабильно ( примерно с одинаковой вероятностью будет иметь то положительный, то отрицательный знак) или малосущественно (достаточно часто будут появляться нулевые парные разницы), то разница выборочных средних , естественно, будет стремиться к нулю. *При этом*  непременно окажется больше нуля, даже в том крайнем случае, когда среди всех сравниваемых пар будет только одна единственная ненулевая разница. Напротив, если все парные различия будут иметь один и тот же знак (будут однонаправленными), то выборочные средние «до» и «после» существенно разойдутся на числовой оси и, соответственно, величина d окажется достаточно велика. Это приведет к снижению *и, следовательно, увеличению критерия Стьюдента.*

Проверка справедливости гипотез при этом производится так же, как и для независимых выборок:

* **если , то различие выборочных средних признается статистически значимым;**
* **если , разница признается незначимой.**

Различие лишь в том, что число степеней свободы для определения табличного значения в данном случае составляет , где *n* – число сравниваемых пар.

Упомянем также о ситуации, когда для установления достоверности различия средних результатов никаких расчетов с *применением критерия Стьюдента просто не требуется*. Это возможно в ситуации, когда максимальное значение одного из сравниваемых выборочных вариационных рядов заведомо меньше минимального значения другого вариационного ряда. Иными словами, значения обоих выборок занимают совершенно разные, не накладывающиеся друг на друга даже частично области на числовой оси. Если такое имеет место, то критерий Стьюдента лишь подтвердит заведомую достоверность различий средних значений сравниваемых выборок. Однако, такая «экспресс-оценка» достоверности возможна лишь в том случае, если сравниваемые выборки достаточно представительны – имеют объем порядка полутора десятков значений или более.

**Критерий Фишера –** критерий сравнения выборочных дисперсий.

На практике часто встречается ситуация, когда факторы, влияющие на состояние изучаемых объектов, вызывают не только и даже не столько сдвиг этих характеристик на числовой оси, сколько усиливают или ослабляют их межиндивидуальное разнообразие.

Количественным индикатором этих изменений является различие выборочных дисперсий. Однако, как всякая выборочная числовая характеристика, выборочная дисперсия **–** величина случайная. Следовательно, наблюдаемое различие дисперсий тоже может оказаться случайным. Таким образом, к выборочным оценкам дисперсии полностью приложимы все те рассуждения, о которых шла речь при обсуждении источников различия выборочных средних.

Дисперсия имеет распределение (распределение Пирсона), поэтому для ее анализа критерий Стьюдента неприменим. Для того, чтобы приблизить распределение к нормальному рассматривают разность логарифмов сравниваемых дисперсий, которая обозначается символом *Z*:

Величина *Z* имеет нормальное распределение и, соответственно, к ней может быть применен критерий Стьюдента.

На практике часто рассматривают отношение *F* большей из сравниваемых дисперсий к меньшей (следуя свойствам логарифмов):

*Полученная величина критерия сравнивается с критическим табличным значением.* И также как в предыдущих рассуждениях, нулевая гипотеза либо отвергается и различие выборочных дисперсий считается статистически достоверным, либо делается вывод, что нулевую гипотезу отвергнуть нельзя и разница выборочных дисперсий находится в границах практически возможных случайных колебаний.

**Если на** независимых выборках **была обнаружена** достоверность **различия дисперсий, то их средние значения** нельзя сравнивать **по *t*- критерию Стьюдента!**

Рассмотренный метод сравнения мер вариации и его модификации являются основой чрезвычайно мощного и информативного метода математико-статистического анализа данных, получившего название **дисперсионный** анализ.

Непараметрические критерии

Параметрические критерии обладают высокой информативностью, поскольку позволяют не только обнаружить достоверность различий, но и точно, конкретно демонстрируют их характер и степень. Однако, при всех несомненных достоинствах параметрические критерии обладают и рядом существенных недостатков – **ограничениями их применимости**. Самый серьезный из них - допущение о нормальности распределения сравниваемых величин. Втрое ограничение - непригодность таких критериев к выборкам малого объема (<10-15 измерений). На таких выборках параметры распределения (средние, дисперсии) могут резко измениться от добавления или убавления даже одного единственного числа. Третье – высокая чувствительность к артефактам, которые оказывают сильное слияние на параметры распределения, вызывая сдвиг средних значений в ту или иную сторону. В результате может «всплыть» различие, которого на самом деле нет или наоборот – оказаться «зашумленной» действительная разница. Влияние артефактов особенно велико на малых выборках. Специфика же медицинской работы состоит в том, что из-за сложности исследуемых процессов и явлений они, как правило, имеют дело именно с выборками малого объема, имеющими неизвестный закон распределения, часто полученными в результате достаточно грубых измерений, «нашпигованными» артефактами.

Для извлечения содержательной информации из числовых массивов такого рода были разработаны **непараметрические критерии**. Это критерии, применение которых не требует пересчета массивов исходных данных в компактно заменяющие их параметры распределения - средние значения, дисперсии или стандартные отклонения и т.д. – и их последующее сравнение.

Как следствие, не только теряет силу требование «нормальности» генеральной совокупности, но и, более того, закон распределения сравниваемых величин вообще не играет никакой роли. Особые, достаточно простые, способы преобразования исходных данных делают эту группу критериев еще и практически нечувствительными к артефактам. В результате, непараметрические критерии успешно работают даже на чрезвычайно малых выборках при наличии грубых измерений и грубых ошибок.

Рассмотрим критерии Манна-Уитни и Вилкоксона.

Критерий Манна-Уитни и критерий Вилкоксона – **критерии ранговые**, т.е. основанные на сравнении сумм рангов, полученных тем или иным образом из сравниваемых выборочных распределений. В данном конкретном случае рангом называется порядковый **номер** числа в ранжированном (расставленном в порядке возрастания) массиве данных – чем больше число, тем выше его ранг. При этом, если числа не повторяются, то их ранги в точности соответствуют их порядковым номерам. Если же некое число повторяется несколько раз, то всем им приписывается средний ранг. Продемонстрируем, как все это происходит и выглядит. Допустим, мы получили следующий вариационный ряд данных *x*:

*5.6 11.7 -3.5 6.3 8 7.4 0.5 8 3 3.1 15.2 3.1 8 6.7 111 4.4*

Здесь числа представлены в том порядке, как они были получены.

Расставим их в порядке возрастания и припишем порядковые номера, а также ранги *R*:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *16* |
| *x* | *-3.5* | *0.5* | *3* | *3.1* | *3.1* | *4.4* | *5.6* | *6.3* | *6.7* | *7.4* | *8* | *8* | *8* | *11.7* | *15.2* | *111* |
| *R* | *1* | *2* | *3* | *4.5* | *4.5* | *6* | *7* | *8* | *9* | *10* | *12* | *12* | *12* | *14* | *15* | *16* |

Из приведенного примера хорошо видно, что при ранжировании происходит «линеаризация данных» **-** сглаживание их резких колебаний за счет того, что ранг числа не зависит от его абсолютной величины и разницы с соседними вариантами. Например, последнее число 111 чуть ли не на порядок превышает ближайшее к нему 15.2. Тем не менее, ранг его всего на 1 выше, чем у предпоследнего числа.

Ранговые критерии для сравнения выборочных совокупностей делятся на две группы – для независимых и зависимых выборок.

**Критерий Манна-Уитни – ранговый критерий для сравнения независимых выборок.**

***Рассмотрим этот критерий на примере.*** Допустим, получены следующие данные о величине ЧСС в двух группах детей 2-3 и 4-5 лет:

*x*(2-3 года): 102, 87, 105, 110, 99, 90 (*nx=6*)

*y*(4-5 лет): 98, 100, 88, 92, 83, 95, 100, 92, 85, 94 (*ny=10*)

Сначала выборки смешивают и ранжируют как одну совокупность:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| № | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *16* |
| *x* | *83* | *85* | *87* | *88* | *90* | *92* | *92* | *94* | *95* | *98* | *99* | *100* | *100* | *102* | *105* | *110* |
| *R* | *1* | *2* | *3* | *4* | *5* | *6.5* | *6.5* | *8* | *9* | *10* | *11* | *12.5* | *12.5* | *14* | *15* | *16* |

Полученные ранги приписывают числам исходных рядов и подсчитывают их суммы:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x(2-3 года)* | *102* | *87* | *105* | *110* | *99* | *90* |  |  |  |  |  |
| *Rx* | *14* | *3* | *15* | *16* | *11* | *5* |  |  |  |  |  |
| *y(4-5 лет)* | *98* | *100* | *88* | *92* | *83* | *95* | *100* | *92* | *85* | *94* |  |
| *Ry* | *10* | *12.5* | *4* | *6.5* | *1* | *9* | *12.5* | *6.5* | *2* | *8* |  |

Далее полученные суммы включают в специальную формулу для подсчета критерия *U:*

В нашем примере получаем , .

В качестве берут меньшее из полученных значений (т.е. 17) и сравнивают его с критическими значениями, взятыми из специальной таблицы: для и для . Поскольку , нулевую гипотезу отвергнуть нельзя даже для и различие уровней ЧСС следует признать статистически незначимым.

Чтобы говорить о статистически значимых различиях, должно выполняться условие .

**Критерий парных сравнений** Вилкоксона**– ранговый критерий для сравнения зависимых выборок.**

Рассмотрим его на примере. У 10 здоровых взрослых людей измеряли кровяное давление после введения кофеина и плацебо. Получены следующие данные для «верхнего», систолического давления СД:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x(кофеин) | 126 | 145 | 137 | 116 | 137 | 157 | 125 | 139 | 143 | 163 |
| y(плацебо) | 121 | 143 | 115 | 120 | 135 | 157 | 115 | 130 | 153 | 160 |

Возникает вопрос, можно ли на основании этих данных полагать, что кофеин оказывает физиологическое действие.

Вначале значения одного ряда строго попарно вычитают из значений другого с учетом знака разницы *d*. Вычтем нижний ряд из верхнего:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x(кофеин) | 126 | 145 | 137 | 116 | 137 | 157 | 125 | 139 | 143 | 163 |
| y(плацебо) | 121 | 143 | 115 | 120 | 135 | 157 | 115 | 130 | 153 | 160 |
| *d* | 5 | 2 | 22 | -4 | 2 | 0 | 10 | 9 | -10 | 3 |

Далее разницы ранжируют по известным правилам, но при этом не учитывают знак разницы (т.е. ранжируют по модулю). Нулевую пару отбрасывают.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *d* | 2 | 2 | 3 | -4 | 5 | 9 | 10 | -10 | 22 |
| *R* | 1.5 | 1.5 | 3 | 4 | 5 | 6 | 7.5 | 7.5 | 9 |

Отдельно суммируют ранги для положительных и отрицательных разниц. В нашем случае получаем: , . В качестве значения критерия *Tz* берут меньшую сумму независимо от знака, т.е. *Tz* =11,5. Сравниваем это значение с «критиче­ским» из специальной таблицы, входом в которую является число сравниваемых пар, но лишь тех, которые не дают нулевые разницы. В нашем случае таковых 9. Тогда *Tкр* = 6 для и *Tкр* =2 для. Поскольку даже для первого уровня значимости, различий уровней СД нулевую гипотезу отвергнуть нельзя и различия не являются статистически значимыми (р<0,05). Иными словами, у нас нет пока оснований утверждать, что действие кофеина носит исключительно физиологический характер.

Смысл теста состоит в следующем. Если бы мы имели бесконечно большой ряд случайных разниц, то число и величина положительных разниц равнялись бы числу отрицательных и, соответственно, суммы их рангов были бы равны. На конечном и ограниченном числовом массиве опять же чисто случайно может иметь место «перекос» в сторону преимущественно положительных или отрицательных разниц. Это обстоятельство и учитывается в критических значениях критерия.

*Tкр* – это граница между практически возможными и практически невозможными значениями критерия. Соответственно, если , то полученная нами сумма рангов с достаточно высокой вероятностью могла возникнуть чисто случайно и о сдвиге одного числового ряда относительно другого ничего определенного сказать нельзя. Это недостоверное различие. Если же , то наблюдаемое различие положительных и отрицательных разниц не могло быть получено случайным образом. Это означает, что смещение значений в сопоставляемых числовых рядах объясняется действием какой-то систематически действующей, неслучайной причины, т.е. носит статистически достоверный (устойчивый и прогнозируемый) характер.

Как было показано выше, пары, имеющие одинаковые числовые значения и, соответственно, дающие нулевые разницы, исключаются из рассмотрения. И если таких случаев много, то «жесткость» критерия нарастает, поскольку *Tкр* тем меньше, чем меньше сравниваемых пар. Соответственно, увеличивается число ситуаций, когда нулевую гипотезу отвергнуть невозможно, и различие будет считаться незначимым. Более того, если число пар окажется меньше 6, то критерий Вилкоксона вообще перестанет «работать»: **6 - минимальное число пар**, для которого еще существует *Tкр*. Для меньшего числа его просто невозможно рассчитать. А подобные ситуации в медико-биологической практике возникают довольно часто, поскольку многие измерения неизбежно приходится выполнять с достаточно высокой степенью грубости, и вероятность появления совпадающих значений здесь все еще весьма высока.

**Критерий согласия Пирсона (критерий )**

Критерии согласия позволяют определить степень соответствия эмпирических и теоретических распределений вероятностей, а также двух эмпирических распределений, полученных, например, в «контроле» и «опыте» или в различных вариантах «опыта» или «наблюдения». Этот критерий позволяет проверить гипотезу о схожести фактического, полученного на практике, распределения вероятностей случайной величины и теоретического.

Данный критерий может использоваться, например, для сравнения частот встречаемости качественных или порядковых признаков в выборочных совокупностях.

Критерий Пирсона записывается следующим образом:

где k – число классов ряда распределения, – фактические (наблюдаемые) частоты встречаемости случайной величины в каждом i-ом классе (в виде целых чисел), – теоретически ожидаемые (вычисленные) частоты для данного класса, – разница между ними.

Таким образом, представляет собой вовсе не квадрат какого-то числа, а суммупо всем классам распределения данной случайной величины (от 1-го до k-го) величины квадратов разницы фактических и теоретических частот в каждом классе, отнесенных к теоретическим частотам для этих же классов.

Допустим, что по каждому классу распределения , т.е. фактические (наблюдаемые) и ожидаемые (вычисленные) частоты идеально совпадают. Тогда и, соответственно, . Понятно, что такого рода ситуация может иметь место только в том случае, когда форма эмпирического распределения абсолютно идентична форме теоретической модели, рассчитанной по эмпирическим данным.

Допустим теперь, что хотя бы для одного из сравниваемых классов эмпирического и теоретического распределений, то есть для какого-то одного из них. Поскольку мы имеем дело с «суммой квадратов», то автоматически станет больше нуля. Чем больше будет таких различий для разных классов и чем значительнее будут сами различия, тем больше будет «набегать» сумма квадратов. Следовательно, при различии наблюдаемых и ожидаемых частот сравниваемых распределений может принимать любые положительные значения, вплоть до бесконечности. Нетрудно представить, что чем менее схожей будет форма сравниваемых распределений, тем большие числовые значения будет принимать и, само собой, наоборот. Иными словами, является мерой сходства/различия формы сравниваемых распределений вероятностей.

Технология использования критерия «Хи-квадрат» чрезвычайно проста. По приведенной выше формуле подсчитывается «экспериментальное» значение «Хи-квадрат» , которое сравнивают с табличным или «критическим» значением , взятым сообразно наличному числу степеней свободы). Далее, как всегда, проверяется выполнение двух неравенств:

- если , то разница наблюдаемых (фактических) и ожидаемых (теоретических) частот сравниваемых распределений незначительна и не выходит за рамки ее собственных случайных колебаний, не превышает критического порога «возможного». Это не позволяет отвергнуть «нулевую гипотезу», согласно которой имеющие место различия частот носят случайный характер;

- если , то разница наблюдаемых (фактических) и ожидаемых (теоретических) частот сравниваемых распределений столь велика, что выходит за рамки ее собственных случайных колебаний. Самопроизвольное появление такого значения «Хи-квадрат» относится к разряду невозможных событий. Следовательно, различие форм эмпирического и теоретического распределений обусловлено действием некого систематически действующего фактора, и эмпирическое распределение на принятом уровне значимости не может рассматриваться в качестве случайной модификации теоретической модели.

Однако, при использовании критерия Пирсона существуют следующие ограничения:

* во-первых, объем выборочной совокупности должен быть не менее 50;
* во-вторых, допускается сравнение только абсолютных, а не относительных частот, т.е. количества значений случайной величины, попадающей в каждый класс распределения «в штуках»
* в-третьих, если в теоретическом (вычисленном) распределении встречается класс, в котором число значений **менее пяти**, то его еще до начала вычисления критерия объединяют с соседним, складывая их частоты. При этом такое же сокращение числа классов, независимо от фактического количества частот в них, производят и в эмпирическом распределении.

Отметим, что в современных программных пакетах математико-статистической обработки данных SPSS и Statistica все операции, необходимые для расчета статистических критериев автоматизированы. Главной задачей пользователя является правильный выбор статистического критерия в каждом конкретном случае. Программа выдает полный отчет о результатах расчетов с указанием уровня значимости нулевой гипотезы. Подробное использование этих статистических программ изложено в электронных и печатных руководствах пользователя.