Проверка статистических гипотез

Содержание

Понятие статистической гипотезы

Поскольку статистика как метод исследования имеет дело с данными, в которых интересующие исследователя закономерности искажены различными случайными факторами, большинство статистических вычислений сопровождается проверкой некоторых предположений или гипотез об источнике этих данных [3, 8, 10, 48].

Статистическая гипотеза – это предположение о свойствах случайных величин или событий, которое мы хотим проверить по имеющимся данным. Приведем примеры ***статистических гипотез*** в научных исследованиях.

- Исследователь хочет проверить, влияет ли дисциплина «Введение в специальность», читаемая на первом курсе, на мотивацию обучения студентов. Здесь проверяется нулевая гипотеза H_0 «Читаемый курс влияет на мотивацию обучения студентов». Альтернативная гипотеза H_1 – «Читаемый курс не влияет на мотивацию обучения студентов». Для проверки гипотезы H_0 необходимо произвести измерение уровня мотивации до начала чтения курса и после чтения курса и сравнить их результаты, используя один из подходящих критериев.
- Необходимо выявить отношение студентов к целесообразности увеличения часов, отводимых на одну из учебных дисциплин. Выдвигается нулевая гипотеза H_0 «Количество часов целесообразно увеличить». В качестве альтернативной гипотезы H_1 может выступать одна из двух гипотез, либо гипотеза «Количество часов необходимо уменьшить», либо «Количество часов оставить без изменения». Для проверки гипотезы необходимо дважды провести измерение мнения студентов путем анкетирования – до изучения и после изучения дисциплины и сравнить результаты анкетирования между собой.
- Успеваемость класса стохастически (вероятностно) зависит от уровня обучаемости учащихся и др.

Подчеркнём ещё раз, что выдвижение гипотез и выбор метода её проверки возлагается на исследователя.

Нулевая гипотеза – это основное проверяемое предположение, которое обычно формулируется как отсутствие различий, отсутствие влияния фактора, отсутствие эффекта, равенство нулю значений выборочных характеристик и т. п. Примером нулевой гипотезы в педагогике является утверждение о том, что различие в результатах выполнения двумя группами учащихся одной и той же контрольной работы вызвано лишь случайными причинами.

Другое проверяемое предположение (не всегда строго противоположное или обратное первому) называется **конкурирующей** или **альтернативной** гипотезой. Для упомянутого выше примера гипотезы H_0 в педагогике одна из возможных альтернатив H_1 будет определена так: уровни выполнения работы в двух группах учащихся различны, и это различие определяется влиянием неслучайных факторов, например тех или других методов обучения.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость проверить её, так как проверку производят статистическими методами, то данная проверка называется статистической.

При проверке статистических гипотез возможны ***ошибки*** (ошибочные суждения) двух видов:

– можно отвергнуть нулевую гипотезу, когда она на самом деле верна (так называемая **ошибка первого рода**);

– можно принять нулевую гипотезу, когда она на самом деле не верна (так называемая **ошибка второго рода**).

Величина φ, называемая ***мощностью критерия***, представляет собой вероятность отклонения неверной нулевой гипотезы, то есть вероятность правильного решения. Мощность критерия – вероятность попадания критерия в критическую область при условии, что справедлива альтернативная гипотеза. Чем больше φ, тем вероятность ошибки второго рода меньше.

Ошибка, состоящая в принятии нулевой гипотезы, когда она ложна, качественно отличается от ошибки, состоящей в отвержении гипотезы, когда она истинна. Эта разница очень существенна вследствие того, что различна значимость этих ошибок. Проиллюстрируем вышесказанное на следующем примере.

Пример. Процесс производства некоторого медицинского препарата весьма сложен. Несущественные на первый взгляд отклонения от технологии вызывают появление высокотоксичной побочной примеси. Токсичность этой примеси может оказаться столь высокой, что даже такое её количество, которое не может быть обнаружено при обычном химическом анализе, может оказаться опасным для человека, принимающего это лекарство. В результате, прежде чем выпускать в продажу вновь произведенную партию, её подвергают исследованию на токсичность биологическими методами. Малые дозы лекарства вводятся некоторому количеству подопытных животных, например мышей, и результат регистрируют. Если лекарство токсично, то все или почти все животные гибнут. В противном случае доля выживших велика.

Исследование лекарства может привести к одному из возможных способов действия: выпустить партию в продажу (a_1), вернуть партию поставщику для доработки или, может быть, для уничтожения (a_2).

Ошибки двух видов, связанные с действиями a_1 и a_2, совершенно различны, различна и важность избежания их. Сначала рассмотрим случай, когда применяется действие a_1, в то время когда предпочтительнее a_2. Лекарство опасно для пациента, в то время как оно признано безопасным. Ошибка этого вида может вызвать смерть пациентов, употребляющих этот препарат. Это ошибка первого рода, так как нам важнее ее избежать.

Рассмотрим случай, когда предпринимается действие a_2, в то время когда a_1 является более предпочтительным. Это означает, что вследствие неточностей в проведении эксперимента партия нетоксичного лекарства классифицировалась как опасная. Последствия ошибки могут выражаться в финансовом убытке и в увеличении стоимости лекарства. Однако случайное отвержение совершенно безопасного лекарства, очевидно, менее нежелательно, чем пусть даже изредка происходящие гибели пациентов. Отвержение нетоксичной партии лекарства – ошибка второго рода.

Допустимая вероятность ошибки первого рода (α) может быть равна 5% или 1% (0,05 или 0,01).

Уровень значимости – это вероятность ошибки первого рода при принятии решения (вероятность ошибочного отклонения нулевой гипотезы).

Альтернативные гипотезы принимаются тогда и только тогда, когда опровергается нулевая гипотеза. Это бывает в случаях, когда различия, скажем, в средних арифметических экспериментальной и контрольной групп настолько значимы (статистически достоверны), что риск ошибки отвергнуть нулевую гипотезу и принять альтернативную не превышает одного из трёх принятых ***уровней значимости*** статистического вывода:

- первый уровень – 5% (p = 95%); где допускается риск ошибки в выводе в пяти случаях из ста теоретически возможных таких же экспериментов при строго случайном отборе испытуемых для каждого эксперимента;
- второй уровень – 1%, то есть соответственно допускается риск ошибиться только в одном случае из ста;
- третий уровень – 0,1%, то есть допускается риск ошибиться только в одном случае из тысячи.

Последний уровень значимости предъявляет очень высокие требования к обоснованию достоверности результатов эксперимента и потому редко используется. В педагогических исследованиях, не нуждающихся в очень высоком уровне достоверности, представляется разумным принять 5%-ный уровень значимости.

Статистика критерия (K) – некоторая функция от исходных данных, по значению которой проверяется нулевая гипотеза. Чаще всего статистика критерия является числовой функцией, но она может быть и любой другой функцией, например многомерной функцией.

Всякое правило, на основе которого отклоняется или принимается нулевая гипотеза, называется **критерием** для проверки данной гипотезы. Статистический критерий (**критерий**) – это случайная величина, которая служит для проверки статистических гипотез.

Критическая область – совокупность значений критерия, при котором нулевую гипотезу отвергают. **Область принятия нулевой гипотезы** (область допустимых значений) – совокупность значений критерия, при котором нулевую гипотезу принимают. При справедливости нулевой гипотезы вероятность того, что статистика критерия попадает в область принятия нулевой гипотезы, должна быть равна $(1 - \alpha)$.

Число степеней свободы у какого-либо параметра определяют как число опытов, по которым рассчитан данный параметр, минус количество одинаковых значений, найденных по этим опытам независимо друг от друга.

Отдельно выделим и рассмотрим понятие гипотезы научного исследования.

Гипотеза научного исследования – методологическая характеристика исследования, научное предположение, выдвигаемое для объяснения какого-либо явления и требующее проверки на опыте для того, чтобы стать достоверным научным знанием. От простого предположения гипотеза отличается рядом признаков, к ним относят:

– соответствие фактам, на основе которых и для обоснования которых она создана;

– проверяемость;

– приложимость к возможно более широкому кругу явлений;

– относительная простота.

В гипотезе органически сливаются два момента: выдвижение некоторого положения и последующее логическое и практическое доказательство.

Психолого-педагогическая или медико-биологическая гипотеза (научное предположение о преимуществе того или иного метода) в процессе статистического анализа переводится на язык статистической науки и заново формулируется по меньшей мере в виде двух статистических гипотез.

Возможны два типа гипотез: первый тип – **описательные** гипотезы, в которых описываются причины и возможные следствия. Второй тип – **объяснительные**, в них даётся объяснение возможным следствиям из определённых причин, а также характеризуются условия, при которых эти следствия обязательно последуют, то есть объясняется, в силу каких факторов и условий будет данное следствие. Описательные гипотезы не обладают предвидением, а объяснительные обладают таким свойством. Объяснительные гипотезы выводят исследователей на предположения о существовании определённых закономерных связей между явлениями, факторами и условиями.

Гипотезы в научных исследованиях могут предполагать, что одно из средств (или группа их) будет более эффективным, чем другие средства. Здесь гипотетически высказывается предположение о сравнительной эффективности средств, способов, методов, форм обучения.

Более высокий уровень гипотетического предсказания состоит в том, что автор исследования высказывает гипотезу о том, что какая-то система мер будет не только лучше другой, но и из ряда возможных систем она кажется оптимальной с точки зрения определённых критериев. Такая гипотеза нуждается в ещё более строгом и оттого более развёрнутом доказательстве.

Этапы проверки статистической гипотезы

Для проверки выдвинутых гипотез разрабатывается специальная процедура (правило), называемая критерием проверки гипотезы. Все критерии проверки нулевых гипотез построены по единому принципу. Исходными данными для проверки являются обычно две случайные выборки, полученные в результате измерений.

Для разных критериев предложены на основе выполненных измерений процедуры вычисления специальных величин, которые называются **статистиками** (наблюдаемые значения статистики). Они различны для различных критериев и зависят от объёмов случайных выборок. Правила их определения будут изложены ниже применительно к различным критериям.

Проверка статистических гипотез складывается из следующих шагов:

– формулируется задача исследования в виде статистической гипотезы;

– выбирается статистическая характеристика гипотезы;

– анализируются возможные ошибочные решения, и оцениваются их последствия;

– формулируются испытуемая (H_0) и альтернативная (H_1) гипотезы;

– задается уровень значимости α, и определяется критическое значение статистической характеристики (критерия $K_{крит}$);

– вычисляется фактическое (экспериментальное) значение статистической характеристики, сравнивается с критическим значением, принимается решение относительно испытуемой гипотезы ($K_{набл}$).

В зависимости от особенностей исследований проверка попадания экспериментального значения критерия в критическую область может иметь различный характер: правосторонний, левосторонний или двусторонний.

1) ***правосторонний критерий***:

- если $K_{эксп} \geq K_{крит}$, то H_0 отклоняется (наблюдаемое экспериментальное значение критерия принадлежит критической области);
- если $K_{эксп} < K_{крит}$, то H_0 принимается (наблюдаемое значение критерия принадлежит области принятия критерия).

Ниже приведён рис. 1, иллюстрирующий данный случай при условии, что решение принимается с 5%-ным уровнем значимости;

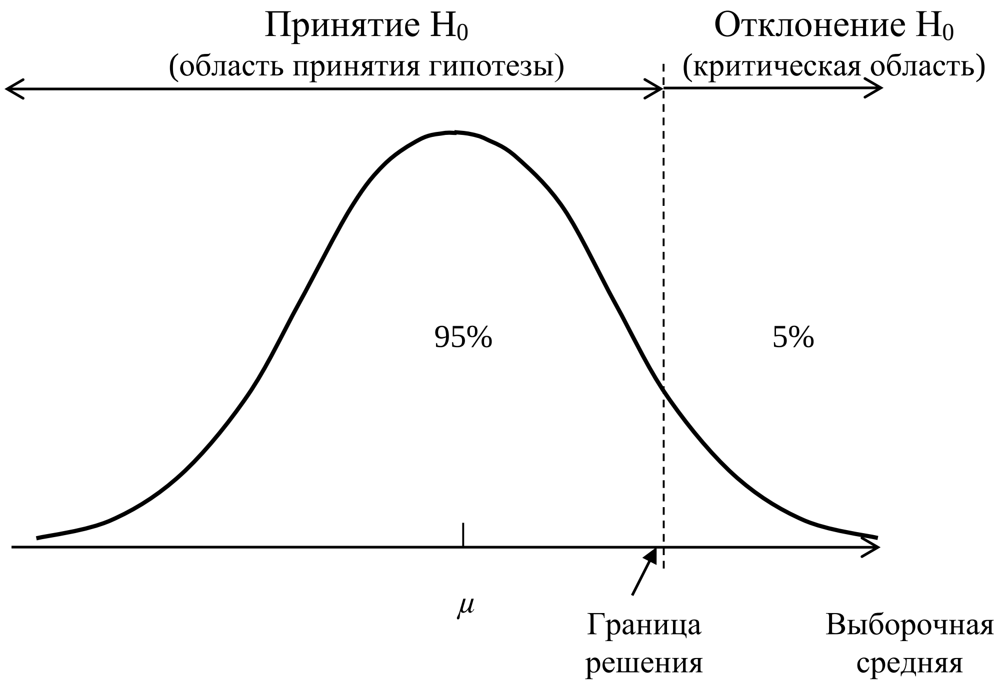


Рис. 1

2) ***левосторонний критерий***:

- если $K_{эксп} \leq K_{крит}$, то H_0 отклоняется (наблюдаемое значение критерия принадлежит критической области);
- если $K_{эксп} > K_{крит}$, то H_0 принимается;

3) ***двусторонний критерий***:

- если $K_{эксп} \neq K_{крит}$, то H_0 отклоняется (экспериментальное значение критерия принадлежит критической области);
- если $K_{эксп} = K_{крит}$, то H_0 принимается.

Данный случай проиллюстрирован на рис. 2 при условии, что решение принимается с 5%-ным уровнем значимости.

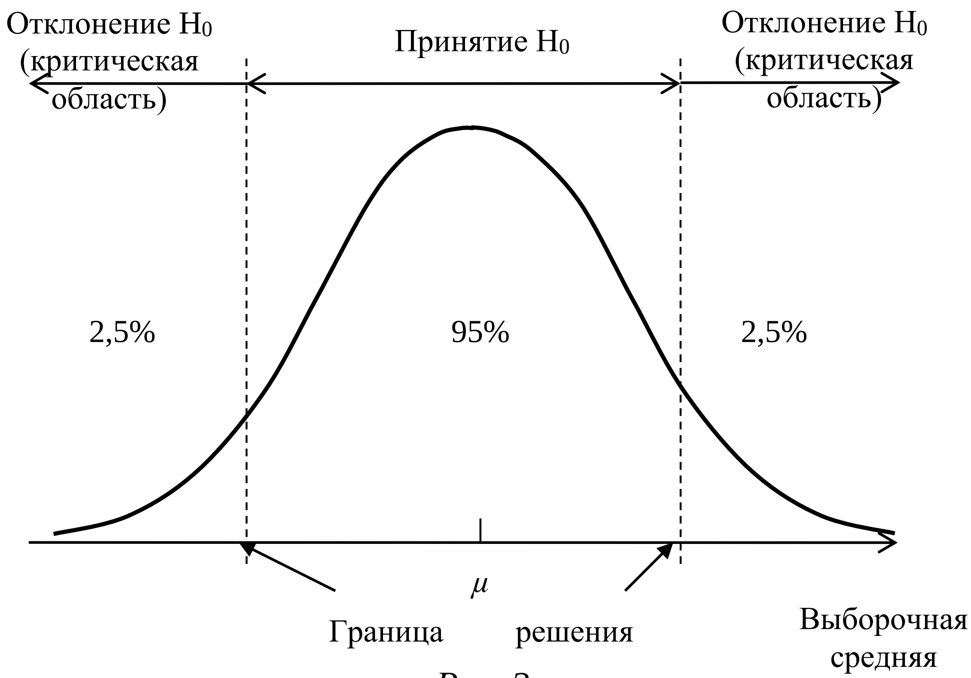


Рис. 2

Замечание. Возможен и другой подход к проверке гипотез. Он состоит в том, что сравниваются между собой не значения статистик $K_{эксп}$ и $K_{крит}$, а вероятности, соответствующие этим статистикам $\alpha_{эксп}$ и $\alpha_{крит}$. В этом случае после определения $K_{набл}$, вычисляется $\alpha_{набл}$, соответствующее этому $K_{эксп}$, и если $\alpha_{\text{набл}} > \alpha_{\text{крит}}$, то гипотеза H_0 принимается, если же $\alpha_{\text{эксп}} \leq \alpha_{\text{крит}}$, то H_0 отвергается.

Сказанное иллюстрируется рис. 3.

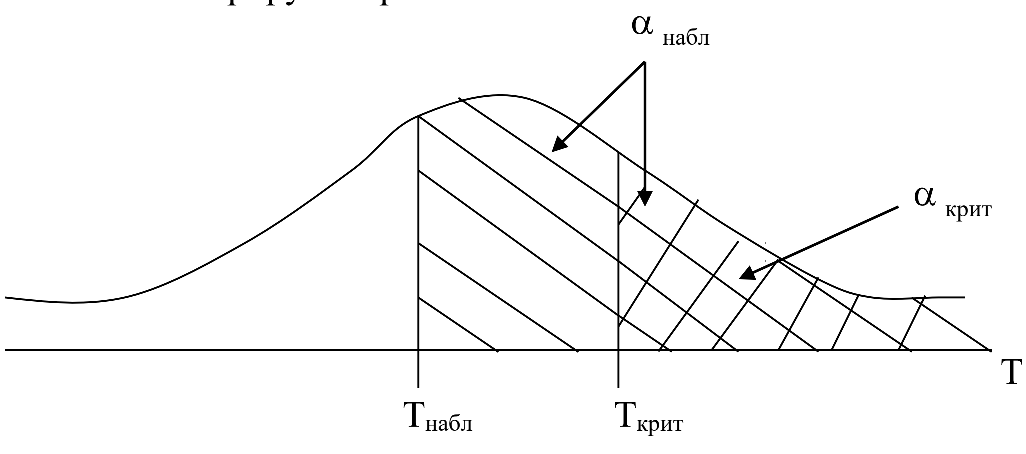


Рис. 3. Принцип отклонения H_0 (второй вариант)

Исторически сложилось так, что в лабораторных и научных исследованиях принято низшим уровнем статистической значимости считать 5%-ный уровень ($\alpha = 0{,}05$), достаточным – 1%-ный ($\alpha = 0{,}01$) и высшим 0,1%-ный ($\alpha = 0{,}001$).

Таким образом, пока статистический уровень значимости не достигнет $\alpha = 0{,}05$, мы ещё не имеем права отклонить нулевую гипотезу, а в дальнейшем будем придерживаться следующего правила отклонения гипотезы об отсутствии различий и принятия гипотезы о статистической достоверности различий.

Правило отклонения нулевой гипотезы H_0 и принятия альтернативной гипотезы H_1: если экспериментальное (эмпирическое) значение критерия равняется критическому значению, соответствующему $\alpha \geq 0{,}05$, или превышает его, то H_0 отклоняется, но мы ещё не можем определённо принять H_1. Если экспериментальное (эмпирическое) значение критерия равняется критическому значению, соответствующему $\alpha \leq 0{,}01$, или превышает его, то H_0 отклоняется и принимается H_1.

В современных статистических пакетах на ЭВМ используются не стандартные уровни значимости, а уровни, подсчитываемые непосредственно в процессе работы с соответствующим статистическим методом. Эти уровни, обозначенные буквой *P*, могут иметь различное числовое выражение в интервале от 0 до 1, например 0,7; 0,23; 0,012. Понятно, что в первых двух случаях полученные уровни значимости слишком велики, и говорить о том, что результат значим, нельзя. В последнем случае результаты значимы на уровне 12 тысячных. Это достоверный результат.

При проверке статистических гипотез с помощью статистических пакетов программа выводит на экран вычисленное значение уровня значимости *P* и подсказку о возможности принятия или неприятия нулевой гипотезы.

Если вычисленное значение *P* превосходит выбранный уровень *α,* то принимается нулевая гипотеза, а в противном случае – альтернативная гипотеза, чем меньше вычисленное значение *P*, тем более исходные данные противоречат нулевой гипотезе.

Статистические критерии в научных исследованиях

Одной из основных задач статистического анализа является совместный анализ нескольких выборок с помощью параметрических или непараметрических критериев [3, 7–13, 15, 23, 30, 31, 33, 35, 37, 43, 45, 48, 50]. Если вид или функция распределения выборки заданы, то в этом случае используются **параметрические критерии**: критерий Стьюдента (*t*) для сравнения выборок по средним значениям, критерий Фишера (*F*) для сравнения выборок по их дисперсиям. Использование параметрических критериев статистики без предварительной проверки вида распределения может привести к определённым ошибкам в ходе проверки рабочей гипотезы. В свою очередь, **непараметрические критерии статистики** свободны от допущения о законе распределения выборок и базируются на предположении о независимости наблюдений, такие, как критерий знаков, двухвыборочный критерий Вилкоксона, критерий Ван дер Вардена, критерий Спирмена, выбор которых, хотя и не требует большого числа членов выборки и знаний, вида распределения, но всё же зависит от целого ряда условий.

Параметрические критерии в научных исследованиях

В группу **параметрических критериев** методов математической статистики входят методы для вычисления описательных статистик, проверки гипотез о принадлежности двух выборок одной совокупности и т. п. Эти методы основываются на предположении о том, что распределение выборок подчиняется нормальному (гауссовому) закону распределения. Среди параметрических критериев статистики нами будут рассмотрены критерий Стьюдента и Фишера.

Возможности и ограничения параметрических и непараметрических критериев.

ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

1. Позволяют прямо оценить различия в дисперсиях (F-критерий Фишера).
2. Позволяют прямо оценить различия в средних, полученных в двух выборках (t-критерий Стьюдента).
3. Позволяют выявить тенденции изменения признака при переходе от условия к условию (дисперсионный однофакторный анализ), но лишь при условии нормального распределения признака.
4. Позволяют оценить взаимодействие двух и более факторов в их влиянии на изменения признака (двухфакторный дисперсионный анализ).
5. Экспериментальные данные должны отвечать двум, а иногда трём условиям:

а) значения признака измерены по интервальной шкале;

б) распределение признака является нормальным;

в) в дисперсионном анализе должно соблюдаться требование равенства дисперсий в ячейках комплекса.

6. Математические расчёты довольно сложны.
7. Если условия, перечисленные в п. 5, выполняются, параметрические критерии оказываются несколько более мощными, чем непараметрические.

Чтобы определить, имеем ли мы дело с нормальным распределением, можно применять следующие методы:

1) в пределах осей можно нарисовать полигон частоты (эмпирическую функцию распределения) и кривую нормального распределения на основе данных исследования. Исследуя формы кривой нормального распределения и графика эмпирической функции распределения, можно выяснить те параметры, которыми последняя кривая отличается от первой;

2) вычисляется среднее, медиана и мода, и на основе этого определяется отклонение от нормального распределения. Если мода, медиана и среднее арифметическое друг от друга значительно не отличаются, мы имеем дело с нормальным распределением. Если медиана значительно отличается от среднего, то мы имеем дело с асимметричной выборкой;

3) эксцесс кривой распределения должен быть равен 0. Кривые с положительным эксцессом значительно «вертикальнее» кривой нормального

распределения. Кривые с отрицательным эксцессом являются более покатистыми по сравнению с кривой нормального распределения;

4) после определения среднего значения распределения частоты и стандартного отклонения находят следующие четыре интервала распределения и сравнивают их с действительными данными ряда:

а) $\bar{x} \pm 0,3\sigma$ – к интервалу должно относиться около 25% частоты совокупности,

б) $\bar{x} \pm 0,7\sigma$ – к интервалу должно относиться около 50% частоты совокупности,

в) $\bar{x} \pm 1,1\sigma$ – к интервалу должно относиться около 75% частоты совокупности,

г) $\bar{x} \pm 3\sigma$ – к интервалу должно относиться около 100% частоты совокупности.

Непараметрические критерии в научных исследованиях

Для доказательства эффективности какого-либо воздействия необходимо выявить статистически значимую тенденцию в смещении (сдвиге) показателей. Для решения подобных задач исследователь может использовать ряд непараметрических критериев различия.

НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

1. Позволяют оценить средние тенденции, например ответить на вопрос, чаще ли в выборке X встречаются более высокие, а в выборке Y более низкие значения признака (критерии Q, U, φ^* и др.).

2. Позволяют оценить различия в диапазонах вариативности признака (критерий φ^*).

3. Позволяют выявить тенденции изменения признака при переходе от условия к условию при любом распределении признака (критерии тенденций L и S).

4. Возможность оценить взаимодействие двух и более факторов в их влиянии на изменения признака отсутствует.

5. Экспериментальные данные могут не отвечать ни одному из этих условий:

а) значения признака могут быть представлены в любой шкале, начиная от шкалы наименований;

б) распределение признака может быть любым, и совпадение его с каким-либо теоретическим законом распределения необязательно и не нуждается в проверке;

в) требование равенства дисперсий отсутствует.

6. Математические расчеты по большей части просты и занимают мало времени (за исключением критериев χ^2 и λ).

7. Если условия, перечисленные в п. 5, не выполняются, непараметрические критерии оказываются более мощными, чем параметрические, так как они менее чувствительны к «засорениям».

λ-критерий Колмогорова – Смирнова

Критерий Колмогорова – Смирнова (критерий согласия) предназначен для сопоставления двух распределений:

а) эмпирического с теоретическим, например, равномерным или нормальным;

б) одного эмпирического распределения с другими эмпирическим распределением.

Критерий позволяет найти точку, в которой сумма накопленных расхождений между двумя распределениями является наибольшей, и оценить достоверность этого расхождения.

Замечание. Если в методе χ^2 сопоставляются частоты двух распределений отдельно по каждому разряду, то здесь сопоставляем сначала частоты по первому разряду, потом по сумме первого и второго разрядов, потом по сумме первого, второго и третьего разрядов и т. д. Таким образом, сопоставляются всякий раз накопленные к данному разряду частоты.

Если различия между двумя распределениями существенны, то в какой-то момент разность накопленных частот достигнет критического значения и мы сможем признать различия статистически достоверными. В формулу данного критерия λ включается эта разность. Чем больше эмпирическое значение $\lambda_{эмп}$, тем более существенны различия.

Гипотезы. Нулевая гипотеза H_0: различия между двумя распределениями недостоверны (судя по точке максимального накопленного расхождения между ними). Альтернативная гипотеза H_1: различия между двумя распределениями достоверны (судя по точке максимального накопленного расхождения между ними).

Ограничения критерия

1. Критерий требует, чтобы выборка была достаточно большой. При сопоставлении двух эмпирических распределений необходимо, чтобы $n_{1,2} \geq 50$. Сопоставление эмпирического распределения с теоретическим иногда допускается при $n > 5$ (Ван дер Варден Б. Л., 1960; Гублер Е. В., 1978).

2. Разряды должны быть упорядочены по нарастанию или убыванию какого-либо признака. Они обязательно должны отражать какое-то однонаправленное его изменение. Например, можем за разряды принять дни недели, 1, 2, 3-й месяцы после прохождения курса терапии, повышение температуры тела, усиление чувства недостаточности и т. п.

Замечание. Отметим, что мы не можем говорить об однонаправленном изменении признака при сопоставлении категорий «очередность рождения», «национальность», «специфика полученного образования» и т. п. Эти данные представляют собой номинативные шкалы: в них нет никакого однозначного однонаправленного изменения признака. Итак, мы не можем накапливать частоты по разрядам, которые отличаются лишь качественно и не представляют собой шкалы порядка. Во всех тех случаях, когда разряды представляют собой не упорядоченные по возрастанию или убыванию какого-либо признака категории, необходимо применять метод χ^2.

Приведём алгоритм для эмпирического и равномерного распределений.

АЛГОРИТМ

1. Занести в таблицу наименования разрядов и соответствующие им эмпирические частоты (первый столбец).

2. Подсчитать относительные эмпирические частоты (частости) для каждого разряда по формуле: $f^*_{эмп} = f_{эмп} / n$, где $f_{эмп}$ – эмпирическая частота по данному разряду; n – общее количество наблюдений. Занести результаты во второй столбец.

3. Подсчитать накопленные эмпирические частости Σf^*_j по формуле $\Sigma f^*_j = \Sigma f^*_{j\text{-}1} + f^*_j$ – это частость, накопленная на предыдущих разрядах; где j – порядковый номер разряда; f^*_j – эмпирическая частость данного j-го разряда. Занести результаты в третий столбец таблицы.

4. Подсчитать накопленные теоретические частости для каждого разряда по формуле $\Sigma f^*_{теор\,j} = \Sigma f^*_{теор\,j\text{-}1} + f^*_{теор\,j}$, где $\Sigma f^*_{теор\,j\text{-}1}$ – теоретическая частость, накопленная на предыдущих разрядах; j – порядковый номер разряда; $f^*_{теор\,j}$ – теоретическая частость данного разряда. Занести результаты в третий столбец таблицы.

5. Вычислить разности между эмпирическими и теоретическими накопленными частостями по каждому разряду (между значениями третьего и четвёртого столбцов).

6. Записать в пятый столбец абсолютные величины полученных разностей, без их знака. Обозначить их как d.

7. Определить по пятому столбцу наибольшую абсолютную величину разности – $d_{\max.}$, то есть $\lambda_{эмп} = d_{\max}$.

8. Определить критические значения критерия $\lambda_{крит}$ для данного количества наблюдений n по статистической таблице Колмогорова – Смирнова из Приложения.

Если $\lambda_{эмп} < \lambda_{крит}$, то принимается нулевая гипотеза H_0 на выбранном уровне значимости α, иначе – принимается альтернативная гипотеза.

Пример. Сопоставление эмпирического распределения с теоретическим.

В выборке здоровых лиц мужского пола, студентов технических и военно-технических вузов в возрасте от 19 и до 22 лет, средний возраст – 20 лет, проводился тест Люшера в восьмицветном варианте. Установлено, что жёлтый цвет испытуемыми чаще предпочитается, чем отвергается. Можно ли утверждать, что распределение жёлтого цвета по восьми позициям у здоровых испытуемых отличается от равномерного распределения? Рассмотрим таблицу эмпирической частоты попадания жёлтого цвета на каждую из восьми позиций (n = 102) (табл. 1).

Таблица 1

Разряды	Позиции жёлтого цвета								Сумма
	1	2	3	4	5	6	7	8	
Эмпирические частоты	24	25	13	8	15	10	9	8	102

Решение. Сформулируем гипотезы. Нулевая гипотеза H_0: эмпирическое распределение жёлтого цвета по восьми позициям не отличается от равномерного распределения. H_1: эмпирическое распределение жёлтого цвета по восьми позициям отличается от равномерного распределения.

Приступим к расчётам, постепенно заполняя результатами таблицу расчёта критерия λ. Занесём в таблицу наименования (номера) разрядов и соответствующие им эмпирические частоты (первый столбец таблицы ниже). Рассчитаем эмпирические частости[1] f^* по формуле $f^*_j = f_j / n$, где f_j – частота попадания жёлтого цвета на данную позицию; n – общее количество наблюдений; j – номер позиции по порядку, полученные результаты записаны во второй столбец (табл. 2).

Таблица 2

Расчёт критерия при сопоставлении распределения выборов жёлтого цвета с равномерным распределением.

Позиция жёлтого цвета	Эмпирические частоты	Эмпирическая частость	Накопленная эмпирическая частость	Накопленная теоретическая частость	Разность
1	24	0,235	0,235	0,125	0,110
2	15	0,147	0,382	0,250	0,132
3	13	0,128	0,510	0,375	0,135
4	8	0,078	0,588	0,500	0,088
5	15	0,147	0,735	0,625	0,110
6	10	0,098	0,833	0,750	0,083
7	9	0,088	0,921	0,875	0,046
8	8	0,079	1,000	1,000	0,000
Сумма	**102**	**1,000**			

Следующие результаты поясняют получение значений в третьем столбце. Для этого суммируют эмпирические частости f^* и тем самым получают накопленные эмпирические частости Σf^*. Например, для первого разряда накопленная эмпирическая частость будет равна эмпирической частости первого разряда, $\Sigma f^*_1 = 0{,}235$[8]. Для второго разряда накопленная эмпирическая частость будет представлять собой сумму эмпирических частостей первого и второго разрядов: $\Sigma f^*_{1+2} = 0{,}235 + 0{,}147 = 0{,}382$ и т. д.

Результаты четвёртого столбца таблицы получены посредством формулы $f^*_{теор} = 1 / k$, где k – количество разрядов (в данном случае позиций цвета).

1 Относительная частота, или частость, – это частота, отнесённая к общему количеству наблюдений; в данном случае это частота попадания жёлтого цвета на данную позицию, отнесённая к количеству испытуемых.

Например, для 1-го разряда теоретическая частость равна: $f^*_{теор} = 1 / 8 = 0{,}125$. Эта теоретическая частость относится ко всем восьми разрядам.

Для второго разряда накопленная теоретическая частость представляет собой сумму теоретических частостей первого и второго: $f^*_{теор\ 1+2} = 0{,}125 + 0{,}125 = 0{,}250$ и т. д.

Далее по пятому столбцу определяем, какая из абсолютных величин разности является наибольшей: в данном случае $d_{\max} = 0{,}135$, то есть $\lambda_{эмп} = 0{,}135$.

По статистической таблице Колмогорова – Смирнова из Приложения определим критические значения критерия $\lambda_{крит.}$:

$$\lambda_{крит} = \begin{cases} 1{,}36/\sqrt{n}, & \alpha = 0{,}05, \\ 1{,}63/\sqrt{n}, & \alpha = 0{,}01. \end{cases}$$

Для данного примера при $n = 102$ имеем: $\lambda_{крит} = \begin{cases} 1{,}36/\sqrt{102} = 0{,}135, & \alpha = 0{,}05, \\ 1{,}63/\sqrt{102} = 0{,}161, & \alpha = 0{,}01. \end{cases}$

Очевидно, что чем больше различаются распределения, тем больше и различия в накопленных частостях. Поэтому нам не составит труда распределить зоны значимости и незначимости по соответствующей оси (рис. 4):

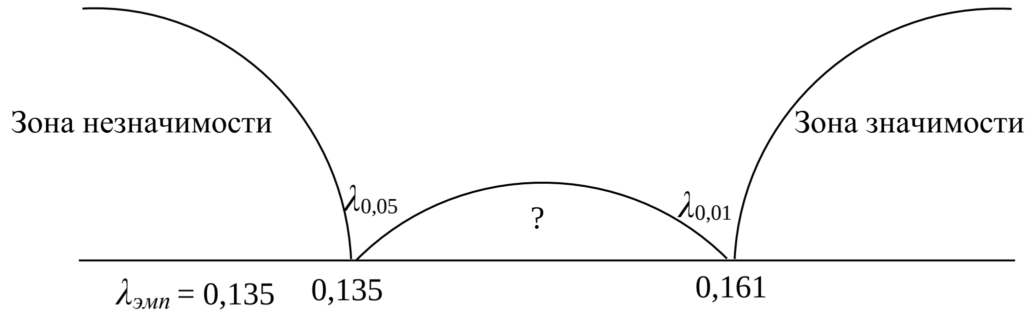


Рис. 4

Таким образом, $\lambda_{эмп} = 0{,}135$ и $\lambda_{эмп} = \lambda_{крит\ 0{,}05}$, тогда на уровне значимости 0,05 принимается гипотеза H_1, то есть распределение жёлтого цвета по восьми позициям отличается от равномерного распределения.

Рассмотрим алгоритм для двух эмпирических распределений.

АЛГОРИТМ

1. Занести в таблицу наименования разрядов и соответствующие им эмпирические частоты, полученные в первом распределении (первый столбец) и во втором распределении (второй столбец).

Замечание. Последовательность выборок может быть выбрана произвольно, так как расхождения между ними оцениваются по абсолютной величине разностей.

2. Подсчитать эмпирические частости по каждому разряду для первого распределения по формуле $f^*_{эмп} = f_{эмп} / n_1$, где $f_{эмп}$ – эмпирическая частота в

данном разряде; n_1 – количество наблюдений в выборке. Занести эмпирические частости первого распределения в третий столбец.

3. Подсчитать эмпирические частости по каждому разряду для второго распределения по формуле $f^*_{эмп} = f_{эмп} / n_2$, где $f_{эмп}$ – эмпирическая частота в данном разряде; n_2 – количество наблюдений во второй выборке. Занести эмпирические частости второго распределения в четвертый столбец таблицы.

4. Подсчитать накопленные эмпирические частости для первого распределения по формуле $\Sigma f^*_j = \Sigma f^*_{j\text{-}1} + f^*_j$, где $\Sigma f^*_{j\text{-}1}$ – частость, накопленная на предыдущих разрядах; j – порядковый номер разряда; $f^*_{j\text{-}1}$ – частость данного разряда. Полученные результаты записать в пятый столбец.

5. Подсчитать накопленные эмпирические частости для второго распределения по той же формуле и записать результат в шестой столбец.

6. Подсчитать разности между накопленными частостями по каждому разряду. Записать в седьмой столбец абсолютные величины разностей, без их знака. Обозначить их как d.

7. Определить по седьмому столбцу наибольшую абсолютную величину разности.

8. Подсчитать эмпирическое значение критерия $\lambda_{эмп}$ по формуле:

$$\lambda_{эмп} = d_{\max} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}},$$

где n_1 – количество наблюдений в первой выборке; n_2 – количество наблюдений во второй выборке.

9. Определить на уровне статистической значимости α критическое значение критерия $\lambda_{крит}$ по статистической таблице Колмогорова – Смирнова из Приложения.

Если $\lambda_{эмп} < \lambda_{крит}$, то принимается нулевая гипотеза H_0 на выбранном уровне значимости α, иначе – принимается альтернативная гипотеза.

Пример. Сопоставлялись два эмпирических распределения: данные, полученные в предыдущем примере, с данными обследования Х. Кларом 800 испытуемых (Klar H., 1974, p. 67). Х. Клар показал, что жёлтый цвет является единственным цветом, распределение которого по восьми позициям не отличается от равномерного. Для получения эмпирических частот Х. Клар использовал метод процентов. Результаты сопоставления представлены в табл. 3.

Таблица 3

Эмпирические частоты попадания жёлтого цвета на каждую из восьми позиций

Разряды позиции жёлтого цвета	1	2	3	4	5	6	7	8	Сумма
Эмпирические частоты	98	113	116	87	91	112	97	86	800

Решение. Сформулируем гипотезы. Нулевая гипотеза H_0: эмпирические распределения жёлтого цвета по 8 позициям в отечественной выборке и выборке Х. Клара не различаются. H_1: эмпирические распределения жёлтого цвета по 8 позициям в отечественной выборке и выборке Х. Клара отличаются друг от друга.

В нашем случае первой будем считать отечественную выборку, второй – выборку Клара и сопоставлять только накопленные эмпирические частости по каждому разряду (табл. 4).

Таблица 4

Расчет критерия при сопоставлении эмпирических распределений жёлтого цвета в отечественной выборке ($n_1 = 102$) и выборке Клара ($n_2 = 800$).

Позиция жёлтого цвета	Эмпирические частоты		Эмпирические частости		Накопленные эмпирические частости		Разность
	f_1	f_2	f_1^*	f_2^*	Σf_1^*	Σf_2^*	$\Sigma f_1^* - \Sigma f_2^*$
1	24	98	0,235	0,123	0,235	0,123	0,112
2	15	113	0,147	0,141	0,382	0,264	**0,118**
3	13	116	0,128	0,145	0,510	0,409	0,101
4	8	87	0,078	0,109	0,588	0,518	0,070
5	15	91	0,147	0,114	0,735	0,632	0,103
6	10	112	0,098	0,140	0,833	0,772	0,061
7	9	97	0,088	0,121	0,921	0,893	0,028
8	8	86	0,079	0,107	1,000	1,000	0
Сумма	102	800	1,000	1,000			

Максимальная разность между накопленными эмпирическими частостями составляет 0,118 и падает на второй разряд. Имеем по формуле:

$$\lambda_{эмп} = d_{\max} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} = 0{,}118 \cdot \sqrt{\frac{102 \cdot 800}{102 + 800}} = 1{,}12 .$$

По статистической таблице Колмогорова – Смирнова из Приложения определяем уровень статистической значимости полученного значения $\alpha = 0{,}16$. Построим для наглядности «ось значимости» (рис. 5).

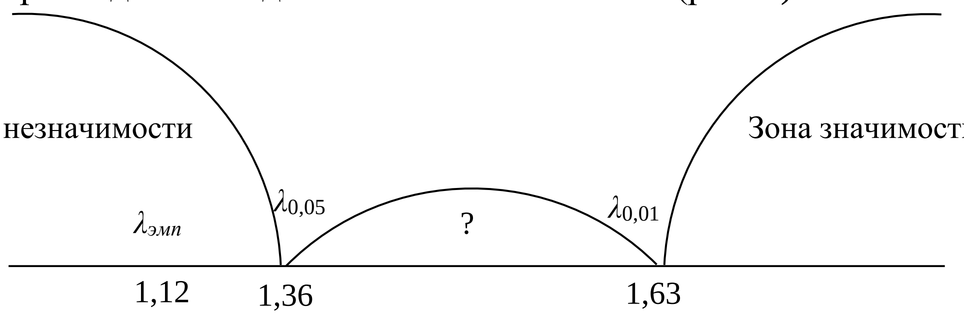


Рис. 5

Таким образом, $\lambda_{эмп} < \lambda_{крит}$, и на уровне значимости 0,05 принимается нулевая гипотеза H_0, то есть эмпирические распределения жёлтого цвета по 8 позициям в отечественной выборке и выборке Х. Клара совпадают.

Следовательно, распределения жёлтого цвета в двух выборках не различаются, но в то же время они по-разному соотносятся с равномерным распределением: у Х. Клара отличий от равномерного распределения не обнаружено, а в отечественной выборке различия обнаружены (с надёжностью 95%).

Критические значения критерия Колмогорова – Смирнова

n	d_{max}		n	d_{max}	
	α = 0,05	α = 0,01		α = 0,05	α = 0,01
5	0,6074	0,7279	50	0,1921	0,2302
10	0,4295	0,5147	60	0,1753	0,2101
15	0,3507	0,4202	70	0,1623	0,1945
20	0,3037	0,3639	80	0,1518	0,1820
25	0,2716	0,3255	90	0,1432	
30	0,2480	0,2972	100	0,1358	
40	0,2147	0,2574	n>100	$1{,}36 / \sqrt{n}$	$1{,}63 / \sqrt{n}$

Критерий χ^2 (хи-квадрат)

Критерий χ^2 (хи-квадрат) применяется для сравнения распределений объектов двух совокупностей на основе измерений (по любой шкале) в двух **независимых** выборках.

Гипотезы. Нулевая гипотеза H_0: $p_1 \leq p_2$, альтернативная гипотеза H_1: $p_1 > p_2$, например, нулевая гипотеза о равенстве вероятностей верного выполнения некоторого задания учащимися контрольных и экспериментальных классов.

Замечание. При проверке нулевых гипотез не обязательно, чтобы значения вероятностей p_1 и p_2 были известны, так как гипотезы только устанавливают между ними некоторые соотношения (равенство, больше или меньше).

Предположим, что состояние изучаемого свойства (например, выполнение определенного задания) измеряется у каждого объекта по шкале наименований, имеющей только две взаимоисключающие категории (например: «выполнено верно» – «выполнено неверно»). По результатам измерения состояния изучаемого свойства у объектов двух выборок составляется четырехклеточная таблица 2×2.

Таблица

	Категория 1	Категория 2	
Выборка 1	Q_{11}	Q_{12}	$Q_{11} + Q_{12} = n_1$
Выборка 2	Q_{21}	Q_{22}	$Q_{21} + Q_{22} = n_2$
	$Q_{11} + Q_{21}$	$Q_{12} + Q_{22}$	$n_1 + n_2 = N$

В этой таблице O_{ij} – число объектов в i выборке, попавших в j категорию по состоянию изучаемого свойства; i ={1;2} – число выборок; j={1;2} – число категорий; N – общее число наблюдений, равное O_{11} + O_{12} + O_{21} + O_{22} или n_1+n_2.

Для проверки нулевой гипотезы по данным таблицы 2×2 (табл. 5) находим эмпирическое значение критерия по формуле:

$$\chi^2_{эмп} = \frac{N\left(\left|Q_{11} \cdot Q_{22} - Q_{12} \cdot Q_{21}\right| - N/2\right)^2}{n_1 \cdot n_2 \cdot \left(Q_{11} + Q_{21}\right) \cdot \left(Q_{12} + Q_{22}\right)}, \qquad (1)$$

где n_1, n_2 — объёмы выборок, $N = n_1 + n_2$ — общее число наблюдений.

Замечание. В иных источниках формула (1) имеет вид: $\chi^2_{\text{эмп}} = \sum_{i=1}^{2} \frac{\left(m_i - m_i^*\right)^2}{m_i^*}$,

где m_i – эмпирические частоты, m^*_i – теоретические частоты.

По статистической таблице критических значений критерия χ^2 из Приложении находим критическое значение $\chi^2_{крит} = \chi_{1\text{-}2\alpha}$ на уровне значимости α и с одной степенью свободы. В силу того, что критерий правосторонний, то применяем правила вывода для правостороннего критерия: если $\chi^2_{эмп} < \chi_{1\text{-}2a}$, то на уровне значимости α принимается нулевая гипотеза, в ином случае – альтернативная.

Ограничения критерия

1. Сумма объёмов двух выборок меньше 20.
2. В таблице 2×2, составленной на основе экспериментальных данных, хотя бы одна из абсолютных частот меньше или равна 5.

АЛГОРИТМ

1. Составить по исходным данным таблицу вида как табл. 5.
2. Сформулировать соответствующие гипотезы, выбрать уровень значимости α.
3. Определить эмпирическое значение критерия «хи-квадрат» $\chi^2_{эмп}$ по формуле (38).
4. Определить критические значения критерия «хи-квадрат» по одноимённой статистической таблице из Приложения.

Если $\chi^2_{эмп} < \chi^2_{крит}$, то на уровне значимости α принимается нулевая гипотеза, в ином случае – альтернативная.

Пример. Проводился эксперимент, направленный на выявление лучшего из учебников, написанных двумя авторскими коллективами в соответствии с целями обучения алгебры и начала анализа и содержанием программы X класса. Для проведения эксперимента методом случайного отбора были выбраны две области. Учащиеся первой области (20 классов) обучались по учебнику № 1, учащиеся второй области (18 классов) обучались по учебнику № 2. Рассмотрим методику сравнения ответов учителей экспериментальных школ двух районов на один из вопросов анкеты: «Доступен ли учебник в целом для самостоятельного чтения и помогает ли он усвоить материал, который учитель не объяснял в классе?». (Ответ: «да – нет».)

Решение. Отношение учителей к изучаемому свойству учебников измерено по шкале наименований, имеющей две категории: «да», «нет». Обе выборки учителей случайные и независимые. Ответы 20 учителей первого района и 15 учителей второго района распределим на две категории и запишем в форме таблицы 2×2.

Таблица 6

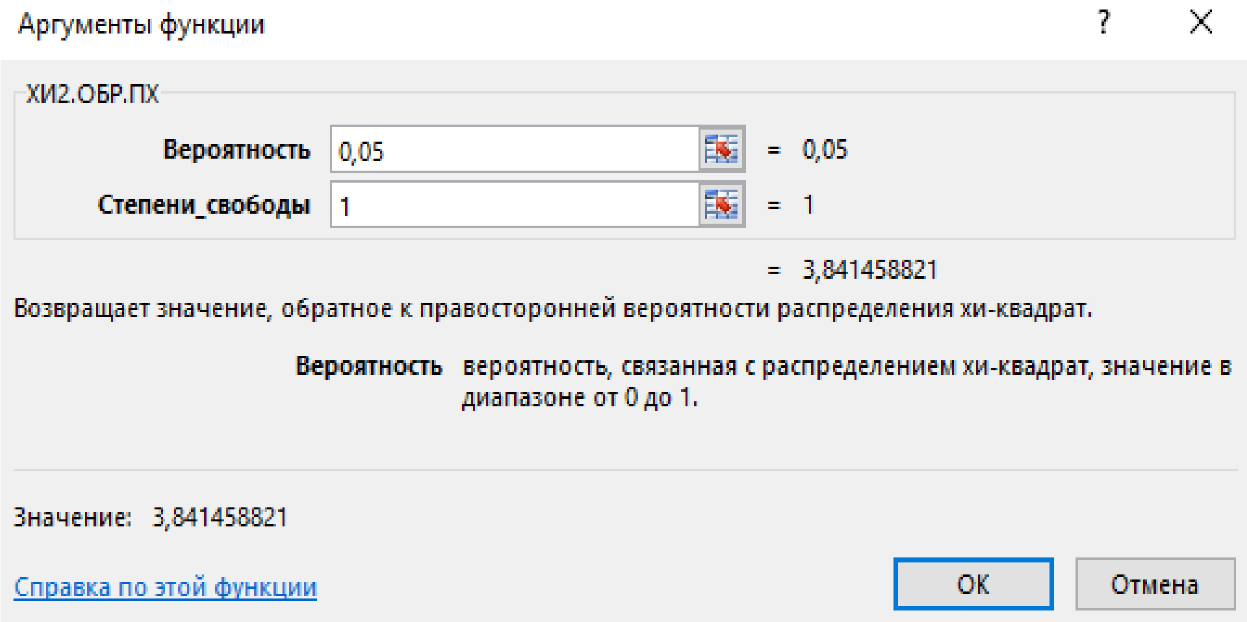
	Да	Нет	
Первый район	$Q_{11} = 15$	$Q_{12} = 5$	$Q_{11} + Q_{12} = n_1 = 20$
Второй район	$Q_{21} = 11$	$Q_{22} = 7$	$Q_{21} + Q_{22} = n_2 = 18$
	$Q_{11} + Q_{21} = 26$	$Q_{12} + Q_{22} = 12$	$n_1 + n_2 = N = 38$

Отметим, что все значения в табл. 6 не меньше 5, что соответствует условиям использования критерия «хи-квадрат».

Сформулируем нулевую гипотезу об одинаковой доступности учебников № 1 и № 2 для самостоятельного чтения учащимися и применим формулу (1):

$$\chi^2_{эмп} = \frac{38\left(\left|15\cdot 7 - 11\cdot 5\right| - 38/2\right)^2}{20\cdot 18\cdot(15+11)\cdot(5+7)} = 0,32.$$

С помощью статистической функции **ХИ2.ОБР.ПХ** для одной степени свободы и уровня значимости α=0,05 найдем $\chi^2_{крит}$ = 3,84.



Таким образом, $\chi^2_{эмп} < \chi^2_{крит}$, так как 0,32 < 3,84, и на уровне значимости 0,05 принимается нулевая гипотеза, то есть результаты проведенного опроса двух экспериментальных оюластей позволяют предполагать об одинаковой доступности учебников № 1 и № 2 для самостоятельного чтения учащимися.

Пример. Изучается частота появления сцепленных полом мутаций у дрозофилы при подкормке солями железа и без подкормки. В опытах было получено 2756 культур с применением подкормки, 805 культур без подкормки.

Среди первых мутации получены в 357 культурах, в 2399 культурах мутаций не было. Среди вторых мутации были в 80 культурах, а в 725 культурах мутаций не наблюдалось.

Решение. С помощью критерия «хи-квадрат» проверяется нулевая гипотеза о том, что число мутаций не изменяется при наличии подкормки, то есть частота мутаций как в группе получавших подкормку, так и в группе не получавших одинакова. Рассчитаем ожидаемое (теоретическое) число культур с мутациями и без мутаций в каждой группе.

Группы	Число культур		Всего
	давшие мутации	не давшие мутации	
С подкормкой	357	2399	2756
Без подкормки	80	725	805
Всего	**437**	**3124**	**3561**

Найдём теоретически ожидаемое число культур с мутациями и без мутаций в группе, получавшей подкормку (первая строка в таблице, содержащая числовые данные): $\frac{2756 \cdot 437}{3561} = 338$ и $\frac{2756 \cdot 3124}{3561} = 2418$. Затем вычислим теоретически ожидаемые частоты и для культур без подкормки (вторая строка в таблице, содержащая числовые данные): $\frac{805 \cdot 437}{3561} = 99$ и $\frac{805 \cdot 3124}{3561} = 706.$

Фактически полученные и теоретически ожидаемые частоты представим в новой таблице.

Группы	Число культур				Всего
	давшие мутации		не давшие мутации		
	фактические частоты	теоретические частоты	фактические частоты	теоретические частоты	
С подкормкой	357	338	2399	2418	2756
Без подкормки	80	99	725	706	805
Всего	**437**		**3124**		**3561**

Найдём эмпирическое значение критерия.

$$\chi^2_{эмп} = \sum_{i=1}^{4} \frac{\left(m_i - m_i^*\right)^2}{m_i^*} = \frac{(357-338)^2}{338} + \frac{(2399-2418)^2}{2418} + \frac{(80-99)^2}{99} + \frac{(725-706)^2}{706} = 5{,}38.$$

Найдём критическое значение критерия с помощью статистической функции **ХИ2.ОБР.ПХ** $\chi^2_{крит} = \chi^2\left(0{,}05;\, 3\right) = 7{,}82.$

Таким образом, $\chi^2_{эмп} < \chi^2_{крит}$, так как 5,38 < 7,82, и на уровне значимости 0,05 принимается нулевая гипотеза, то есть влияние подкормки на изменение количества мутаций нельзя считать полностью доказанным.

Замечание. Применение критерия «хи-квадрат» возможно и в том случае, когда объекты двух выборок из двух совокупностей по состоянию изучаемого свойства распределяются более чем на две категории. Например, учащиеся экспериментальных и контрольных классов распределяются на четыре категории в соответствии с отметками (в баллах: 2, 3, 4, 5), полученными учащимися за выполнение некоторой контрольной работы.

Результаты измерения состояния изучаемого свойства у объектов каждой выборки распределяются на C категорий. На основе этих данных составляется таблица $2 \times C$, в которой два ряда (по числу рассматриваемых совокупностей) и C колонок (по числу различных категорий состояния изучаемого свойства, принятых в исследовании).

Таблица 6

	Категория 1	Категория 2		Категория C	
Выборка 1	Q_{11}	Q_{12}	…	Q_{1C}	n_1
Выборка 2	Q_{21}	Q_{22}	…	Q_{2C}	n_2
	$Q_{11} + Q_{21}$	$Q_{12} + Q_{22}$	…	$Q_{1C} + Q_{2C}$	$n_1 + n_2 = N$

На основе данных ранее представленной табл. 13 можно проверить нулевую гипотезу о равенстве вероятностей попадания объектов первой и второй совокупностей в каждую из i (i = 1, 2, ..., C) категорий, то есть проверить выполнение всех следующих равенств: $p_{11} = p_{21}$, $p_{12} = p_{22}$, ..., $p_{1c} = p_{2c}$.

Возможна, например, проверка гипотезы о равенстве вероятностей получения отметок «5», «4», «3» и «2» за выполнение учащимися контрольных и экспериментальных классов некоторого задания.

Замечание. Для проверки нулевой гипотезы на основе данных таблицы $2 \times C$ подсчитывается эмпирическое значение критерия χ^2 по формуле:

$$\chi^2_{эмп} = \frac{1}{n_1 \cdot n_2} \cdot \sum_{i=1}^{C} \frac{\left(n_1 \cdot Q_{2i} - n_2 \cdot Q_{1i}\right)^2}{Q_{1i} + Q_{2i}}, \qquad (2)$$

где n_1 и n_2 – объёмы выборок.

Эмпирическое значение критерия $\chi^2_{эмп}$, полученное на основе экспериментальных данных, сравнивается с критическим значением $\chi^2_{крит} = \chi_{1-2\alpha}$, которое определяется по одноимённой статистической таблице на уровне значимости α и степенью свободы $k = C - 1$. Критерий правосторонний, поэтому на уровне значимости α: если $\chi^2_{эмп} < \chi_{1-2a}$, то принимается нулевая гипотеза, если $\chi^2_{эмп} > x_{1-2a}$, то принимается альтернативная гипотеза. Последнее означает, что распределение объектов на C категорий по состоянию изучаемого свойства различно в двух рассматриваемых совокупностях.

АЛГОРИТМ

1. Заполнить с учётом исходных данных табл. 6.

2. Сформулировать соответствующие гипотезы.

3. Определить эмпирическое значение критерия «хи-квадрат» $\chi^2_{эмп}$ по формуле (2): $\chi^2_{эмп} = \frac{1}{n_1 \cdot n_2} \cdot \sum_{i=1}^{C} \frac{\left(n_1 \cdot Q_{2i} - n_2 \cdot Q_{1i}\right)^2}{Q_{1i} + Q_{2i}}$.

4. Определить критические значения критерия «хи-квадрат» по одноимённой статистической таблице из Приложения.

Если $\chi^2_{эмп} < \chi^2_{крит}$, то на уровне значимости α принимается нулевая гипотеза, в ином случае – альтернативная.

Пример. Рассмотрим методику сравнения результатов письменной работы, проверявшей усвоение одного из разделов курса учащимися первого и второго районов. Методом случайного отбора были составлены две выборки объёмом по 50 человек каждая соответственно из учащихся каждого из районов, писавших работу. В соответствии со специально разработанными критериями оценки выполнения работы каждый ученик мог попасть в одну из четырех категорий: «плохо», «посредственно», «хорошо» и «отлично». Результаты выполнения работы двумя выборками учащихся используем для проверки альтернативной гипотезы о том, что учебник № 1 способствует лучшему усвоению проверяемого раздела курса, то есть учащиеся первого экспериментального района в среднем будут получать более высокие оценки, чем учащиеся второго района.

Решение. Результаты выполнения работы учащимися обеих выборок запишем в виде таблицы 2×4.

	Плохо	Посред-но	Хорошо	Отлично	
Выборка 1	$Q_{11} = 3$	$Q_{12} = 19$	$Q_{13} = 18$	$Q_{14} = 10$	$n_1 = 50$
Выборка 2	$Q_{21} = 9$	$Q_{22} = 24$	$Q_{23} = 12$	$Q_{24} = 5$	$n_2 = 50$
	$Q_{11} + Q_{21} = 12$	$Q_{12} + Q_{22} = 43$	$Q_{11} + Q_{21} = 30$	$Q_{11} + Q_{21} = 15$	$n_1 + n_2 = N = 100$

Сформулируем нулевую гипотезу: учащиеся первого экспериментального района и учащиеся второго района в среднем будут получать одинаковые оценки.

В соответствии с условиями использования критерия χ^2 найдем эмпирическое значение критерия по формуле (2):

$$\chi^2_{эмп} = \frac{1}{n_1 \cdot n_2} \cdot \left(\frac{\left(n_1 \cdot Q_{21} - n_2 \cdot Q_{11}\right)^2}{Q_{11} + Q_{21}} + \frac{\left(n_1 \cdot Q_{22} - n_2 \cdot Q_{12}\right)^2}{Q_{12} + Q_{22}} + \frac{\left(n_1 \cdot Q_{23} - n_2 \cdot Q_{13}\right)^2}{Q_{13} + Q_{23}} + \frac{\left(n_1 \cdot Q_{24} - n_2 \cdot Q_{14}\right)^2}{Q_{14} + Q_{24}} \right) =$$

$$= \frac{1}{50 \cdot 50} \cdot \left(\frac{(50 \cdot 9 - 50 \cdot 3)^2}{12} + \frac{(50 \cdot 24 - 50 \cdot 19)^2}{43} + \frac{(50 \cdot 12 - 50 \cdot 18)^2}{30} + \frac{(50 \cdot 5 - 50 \cdot 10)^2}{15} \right) =$$

$$= \frac{50^2}{50^2} \cdot \left(\frac{(9-3)^2}{12} + \frac{(24-19)^2}{43} + \frac{(12-18)^2}{30} + \frac{(5-10)^2}{15} \right) = 3 + \frac{25}{43} + \frac{6}{5} + \frac{5}{3} = 6{,}45.$$

С помощью статистической функции **ХИ2.ОБР.ПХ** находим критическое значение критерия χ^2 для степени свободы k = 4 – 1 = 3 и уровня значимости α = 0,05 найдём $\chi^2_{крит}$ = 7,82.

Таким образом, $\chi^2_{эмп} < \chi^2_{крит}$, так как 6,45 < 7,815, и на уровне значимости 0,05 принимаем нулевую гипотезу, то есть учащиеся первого и второго районов в среднем будут получать одинаковые оценки, то есть учебник № 1 не способствует лучшему усвоению проверяемого раздела курса.

Литература

1. Гласс Дж. Статистические методы в педагогике и психологии / Дж. Гласс, Дж. Стенли. – М.: Прогресс, 1976. – 496 с.
2. Гланц С. Медико-биологическая статистика / С. Гланц. - М.: Практика, 1998. - 459 с.
3. Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие для вузов / В. Е. Гмурман. – М.: Высш. шк., 2003. – 479 с.
4. Новиков Д. А. Статистические методы в педагогических исследованиях (типовые случаи) / Д. А. Новиков. – М.: МЗ-Пресс, 2004. – 67 с.
5. Новиков Д. А. Статистические методы в медико-биологическом эксперименте (типовые случаи) / Д. А. Новиков, В. В. Новочадов. – Волгоград: Изд-во ВГМУ, 2005. – 84 с.
6. Шилова З. В. Теория вероятностей и математическая статистика: учебное пособие / З. В. Шилова, О. И. Шилов. – Киров: Изд-во ВГГУ, 2015. – 158 с.