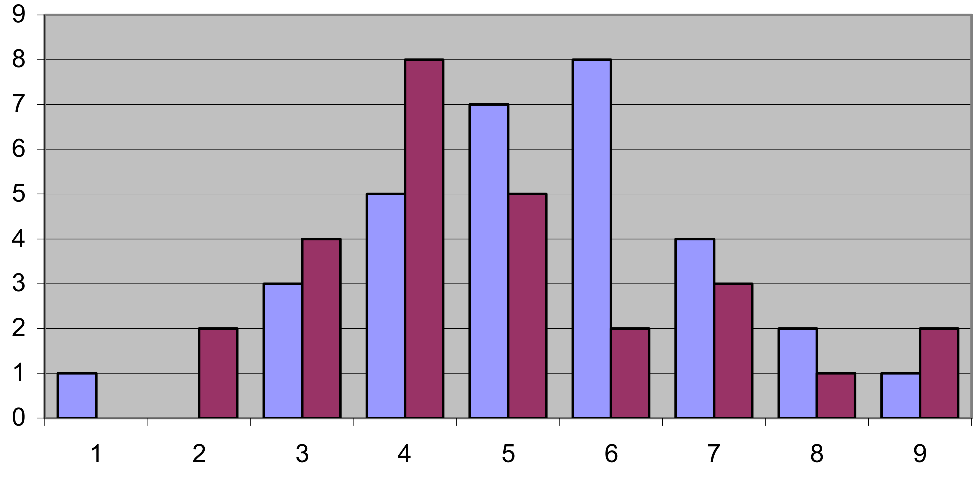
Р.А. Майер, Н.Р. Колмакова, А.В. Ванюрин

СТАТИСТИЧЕСКОЕ СОПРОВОЖДЕНИЕ ПЕДАГОГИЧЕСКОГО ЭКСПЕРИМЕНТА

Учебное пособие



Красноярск 2008

Федеральное агентство по образованию
ГОУ ВПО «Красноярский государственный педагогический университет им. В.П. Астафьева»

Р.А. Майер, Н.Р. Колмакова, А.В. Ванюрин

СТАТИСТИЧЕСКОЕ СОПРОВОЖДЕНИЕ ПЕДАГОГИЧЕСКОГО ЭКСПЕРИМЕНТА

Учебное пособие

Красноярск 2008

ББК 74.00
М14
Печатается по решению редакционно-издательского совета ГОУ ВПО КГПУ «Красноярский государственный педагогический университет им. В.П. Астафьева»

Рецензенты:
Доктор физико-математических наук, профессор
М. В. Носков (СФУ)
Кандидат педагогических наук, доцент
Е. Н. Васильева (ИПКРО)

М14 Майер Р. А., Колмакова Н. Р., Ванюрин А. В. Статистическое сопровождение педагогического эксперимента: учебное пособие; Краснояр. гос. пед. ун-т им. В. П. Астафьева. – Красноярск, 2008. – 88 с.

Адресовано специалистам в области педагогических наук, не имеющих специального математического образования, а также студентам педагогических вузов, чья профессиональная деятельность может потребовать использования статистических методов в подготовке и проведении экспериментального или прикладного исследования. В определенной степени это может быть отнесено к проведению тестирования, различного рода анкетирований, а также исследованию статистических гипотез и выяснению надежности результатов проведенного исследования.

ISBN 978-5-85981-228-8

ББК 74.00

Оглавление

Введение

В последние десятилетия статистические методы, давно уже используемые во многих гуманитарных науках, в частности в психологии, медицине, социологии, стали, наконец, применяться и в педагогических исследованиях. К сожалению, как показал проведенный Д. А. Новиковым анализ соответствующих диссертационных исследований (1), их авторы далеко не всегда понимают смысл выполняемых действий, допуская в своих рассуждениях и расчетах принципиальные ошибки, сводящие на нет всю доказательную часть исследования. Так, например, далеко не всегда показатели, выбранные для определения эффективности рассматриваемой системы (их в дальнейшем будем назвать оценочными показателями), адекватно отражают достигаемые системой результаты. Чаще всего это происходит тогда, когда ограничиваются одним оценочным показателем, который к тому же при внимательном рассмотрении часто оказывается комплексной оценкой, что сразу выводит исследование за рамки научной строгости. Крайне редко в этих целях используются так называемые векторные оценки, компонентами которых служат независимые одномерные показатели, с помощью их легче всего адекватно оценить рассматриваемое педагогическое явление. Редко когда исследуется качество инструментов, используемых для измерения оценочных показателей (тестов, контрольных работ, анкет, интервью и т.д.), часто не объясняются мотивы выбора того или иного статистического критерия и его допустимость. Наконец, чрезвычайно редко описывается генеральная совокупность, на которую с заданной надежностью могут быть распространены выводы диссертационного исследования, и механизм построения выборок. Много некорректных интерпретаций и выводов.

Глава 1. Оценочные показатели и их измерение

1.1. Некоторые особенности организации педагогического эксперимента. Прежде чем приступить к рассмотрению технических вопросов использования статистических методов в педагогических исследованиях, уточним понятие педагогического эксперимента и связанную с ним терминологию, которая, к сожалению, не отличается единообразием.

Как известно, классическая схема статистического эксперимента, разновидностью которого по существу является педагогический эксперимент, заключается в следующем: пусть требуется сравнить два метода обработки некоторой системы объектов на предмет выяснения их влияния на одну из характеристик этой системы. Из-за невозможности применить два метода обработки к одной и той же системе объектов, строят вторую систему, во всех отношениях, по всем свойствам подобную первой. На этих двух системах и проверяют данные методы: на одной – первый, на другой – второй. В педагогических исследованиях систему объектов, к которой применяют разработанный исследователем метод, называют экспериментальной группой, а вторую, в которой обычно реализуются традиционные методы, – контрольной. После реализации рассматриваемых методов изменившиеся системы измеряют на предмет сопоставления значений исследуемой характеристики (качества усвоения учебного материала, приобретенных личностных качеств и т.д.). Затем методами математической статистики выясняется, на каком уровне надежности обнаруженное различие в этих значениях можно считать не случайным, а порожденным различием применявшихся методов.

Самым сложным моментом в этой процедуре является построение экспериментальной и контрольной групп, равных по составу и основным свойствам. Здесь нельзя обойтись обеспечением равенства одного, двух или трех свойств, например, равенства групп по возрастному составу, полу, успеваемости или одному из личностных качеств испытуемых. Надо обеспечить совпадение и многих других характеристик, влияние которых на ход и результаты эксперимента не исключено, например, таких как трудолюбие, навыки самостоятельной работы, мотивация, ценностные ориентации и многие другие качества испытуемых.

В классической схеме статистического эксперимента для этого строится так называемая *генеральная совокупность* объектов, к которым применимы оба сопоставляемых метода и из которой *случайным образом* извлекаются две выборки требуемого объема. В педагогических исследованиях они используются в качестве экспериментальной и контрольной групп. Равенство построенных выборок по содержанию и свойствам обеспечивается автоматически случайным характером их отбора. В силу той же причины выборки подобны генеральной совокупности. В этом случае неслучайный характер различий трактуется как возможность распространения результатов эксперимента на всю генеральную совокупность.

В педагогических исследованиях применение классической схемы педагогического эксперимента в большинстве случаев затрудняется необходимостью собирать отобранных участников эксперимента в одном месте. Естественно, что в этих условиях исследователю приходится ограничиваться небольшой по численности генеральной совокупностью. В этой ситуации важно детально описывать генеральную совокупность, чтобы читатель мог судить о применимости предлагаемой педагогической системы в конкретных условиях.

На практике часто используют *упрощенную схему педагогического эксперимента*, в которой уже не используются ни понятие генеральной совокупности, ни процедура случайного выбора. В ней все начинается с попытки исследователя уравнять «вручную» уже сложившиеся или искусственно сформированные экспериментальную и контрольную группы по всем качествам, могущим, по мнению исследователя, оказать влияние на ход и результаты эксперимента.

Хотя характер применения методов математической статистики на заключительном этапе эксперимента в случае применения упрощенной схемы тот же, что и в случае использования классической схемы, однако вывод формулируется в несколько иной форме, а именно как указание надежности того, что различие в оценках исследуемой характеристики экспериментальной и контрольной групп не случайно (закономерно). Иногда в этой формулировке используется термин *достоверно*, с помощью которого вывод формулируется как указание надежности, с которой можно утверждать достоверность различий в оценках исследуемой характеристики экспериментальной и контрольной групп, или еще короче – как достоверность этих различий на заданном уровне надежности.

Истинная, а не формальная, вычисленная методами математической статистики надежность в упрощенной схеме в значительной степени зависит от тщательности и аккуратности, с которой исследователь обосновывает исходное равенство экспериментальной и контрольной групп.

В заключение отметим два важных момента, которые касаются и классической и упрощенной схемы эксперимента.

1. Как известно любому опытному учителю, начальное равенство групп по составу не гарантирует в дальнейшем их равенство в учебе и поведении. Поэтому, если эксперимент длится сравнительно большой промежуток времени, а исследователь по-настоящему заинтересован в установлении истины, он должен следить за учебой и поведением всех участвующих в эксперименте групп.

2. Многие из перечисленных выше проблем, связанных с одушевленностью участников педагогического эксперимента хотя бы частично могут быть сняты в системе дистанционного эксперимента, не требующего реального объединения членов каждой из выборок, в результате чего ослабляются внутригрупповые влияния, а также личностное влияние учителя на ход и результаты эксперимента.

1.2. Способы построения случайных выборок. Существует несколько способов построения выборок. В педагогических исследованиях чаще всего используется простой бесповторный, случайный отбор, при котором объекты изучения отбираются из генеральной совокупности без предварительного учёта наличия или отсутствия у них изучаемого признака, и тем самым реализуется равенство шансов попадания единиц отбора в выборочную совокупность. В основе процедуры простого случайного отбора лежит пронумерованный *перечень* элементов генеральной совокупности, который обладает свойствами полноты, точности, адекватности, удобства работы с ним, отсутствия дублирования единиц наблюдения. Сам отбор осуществляется с помощью либо *генератора случайных чисел*, который имеется в любом языке программирования, либо *таблицы случайных чисел*. Для обеспечения бесповторного отбора приходится пропускать номера, которые встречаются в таблице во второй раз. К основным достоинствам простых случайных выборок относится простота предварительной информации о генеральной совокупности (перечень её элементов), а также то, что легко проверяются ошибки. При больших выборках простой случайный отбор неудобен, а часто просто непригоден. В этом случае применяются несколько упрощенные варианты простого случайного отбора, среди которых наиболее распространены: *систематическая выборка, гнездовая выборка, стратифицированная выборка, многоступенчатая и квотная выборки.* Каждая из них имеет свои достоинства и свои недостатки. Со всеми этими видами выборок можно познакомиться в главе 7 нашей книги «Теория и практика статистического анализа» (2).

При построении выборок надо иметь в виду, что в статистике используется два их вида: *независимые* и *зависимые* выборки. Если, например, сравниваются результаты выполнения теста одними и теми же людьми в начале эксперимента и по его окончании, то сопоставляемые выборки зависимы. Зависимыми будут также выборки, в одной из которых девочки, а во второй их братья. Такие сравнения называют *парными.* Метод *парных* *сравнений* используется и в других случаях, когда связь между элементами формируемых выборок строится в определенном смысле искусственно. Построив одну общую случайную выборку с четным числом элементов, ее преобразуют в две, подбирая очередному элементу наиболее близкий по возрасту, или полу, или какому-либо другому важному для исследователя показателю элемент. Распределение отобранных «партнёров» по двум группам производится отдельно для каждой пары случайным образом (например, с помощью бросания монеты). При такой организации «пар» практически вдвое уменьшается число независимых элементов сопоставления (в главе второй их число мы будем называть степенью свободы системы) и, следовательно, теряется часть информации, но, не смотря на это, можно существенно уменьшить или даже полностью исключить рассеяние, которое имеется между двумя сопоставляемыми группами. Заметим, что при этом выигрыш в точности, который приносит применение связанных выборок, превосходит потери от уменьшения числа степеней свободы.

1.3. Оценочные показатели в педагогических исследованиях. Важным условием успешности применения статистических методов в педагогических исследованиях является включение в исследуемую систему независимых *оценочных показателей*, которые в совокупности более или менее адекватно характеризовали бы состояние системы. Вводить и использовать их следует по возможности раньше, а не только на заключительном этапе исследования. При этом наряду с показателями, оценивающими ожидаемые позитивные изменения, нужны показатели, фиксирующие возможные отрицательные последствия нововведений, а также показатели, фиксирующие мнения респондентов об отдельных качествах разрабатываемой педагогической системы.

Изначально полезно ввести генеральную совокупность объектов исследования, на которых исследователь предполагает распространить результаты исследования.

Предположим, что исследуется вопрос о том, какое влияние на усвоение учащимися элементов комбинаторики, теории вероятностей и статистики оказывает использование информационных технологий.

Учитывая различия в возможностях и характере использования информационных технологий в разных разделах математики и при использовании разных форм математической деятельности, в рассматриваемом случае нужно ввести, как минимум, шесть приведенных ниже оценочных показателей:

J_1 – владение основными, предусмотренными программой курса, понятиями комбинаторики;

J_2 – владение основными, предусмотренными программой курса, понятиями теории вероятностей;

J_3 – владение основными, предусмотренными программой курса, понятиями статистики;

J_4 – умение решать стандартные задачи комбинаторики;

J_5 – умение решать стандартные задачи теории вероятностей;

J_6 – умение решать стандартные задачи статистики.

Если исследователь предполагает, что разработанная им методика использования информационных технологий будет содействовать лучшему пониманию мировоззренческого значения теории вероятностей, ее связи со статистикой, а также повышению интереса к рассматриваемому разделу математики, то естественно ввести еще как минимум три оценочных показателя, например:

J_7 – уровень понимания связи, существующей между теорией вероятностей и статистикой;

J_8 – уровень понимания мировоззренческого значения теории вероятностей;

J_9 – уровень интереса, сформированного у учащихся к вопросам, рассматриваемым в комбинаторике, теории вероятностей и статистике.

Если в процессе обсуждения разрабатываемой методики высказывалось предположение, что она создаст дополнительные трудности по сравнению с методикой, применяемой в контрольной группе, можно ввести оценочный показатель:

J_{10} – мнение учащихся о трудностях, связанных с изучением комбинаторики, теории вероятностей и статистики и т.д.

Так как эксперимент, выясняющий верность сформулированной гипотезы исследования длится, как минимум, семестр, то по причине указанной в первом пункте этой главы следовало бы ввести показатель (или показатели), оценивающий деловую активность и психологическую атмосферу каждой из участвующих в эксперименте групп.

Всю совокупность таких показателей принимают в качестве развернутой оценки состояния разрабатываемой педагогической системы. В статистике такую оценку называют *векторной.* При ее использовании часто возникает вопрос о возможности свертывания такой оценки в одну, позволяющей выразить состояние системы одним числом, которое в этом случае называют *комплексной оценкой.* Так, во многих спортивных состязаниях победитель выявляется по сумме очков, баллов, набранных на отдельных этапах состязания или в отдельных играх, а в многоборье – в отдельных видах спорта.

На практике такие комплексные оценки встречаются довольно часто, и, очевидно, без них не обойтись, хотя способы их определения нередко кажутся весьма сомнительными. Обычно же в повседневной жизни такие комплексные оценки являются либо результатом определенных общественных соглашений, либо устанавливаются каким-либо нормативным актом определенного директивного органа.

Однако надо иметь в виду, что даже в таких, в некотором смысле бытовых ситуациях, подобный переход от развернутой, векторной оценки к комплексной может оказаться не очень желательным, так как его выполнение неизбежно ведет к потере информации. Тем более нежелателен он в научных исследованиях, особенно когда «свертываются» разнородные величины, например, оценки за разные виды деятельности. К тому же надо иметь в виду трудности интерпретации комплексных оценок, а также то, что за ними часто скрываются негативные моменты вновь вводимой системы.

1.4. Измерение оценочных показателей. Использование оценочных показателей требует разработки соответствующей системы *измерения* различного рода качеств, свойственных исследуемому педагогическому явлению.

Первые методы измерения, разработанные человечеством, были сравнительно просты: выбиралась единица измерения и подсчитывалось, сколько раз она содержится в измеряемом объекте. Система образующихся при этом чисел и привела к понятию числовой шкалы. Постепенно процедура измерения и понятие шкалы усложнялись. Так, Цельсий для построения шкалы измерения температур, нанеся на тонкую трубку с ртутью две черты, соответствующие температуре замерзающей и температуре кипящей воды, разделил образовавшийся отрезок на сто равных частей. Аналогичным способом было получено много различных шкал для измерения многих физических величин.

Несколько иным способом была построена шкала твердость минералов. Для этого был взят набор 10 эталонных минералов для определения относительной твердости методом царапания. За 1 принят тальк, за 2 – гипс, за 3 – кальцит и так далее до 10, которому соответствовал алмаз – самый твердый минерал, оставлявший царапину па любом другом минерале. В результате оказалось возможным любому минералу однозначно приписать определенную твердость. Если, например, исследуемый минерал царапает кварц (7), но не царапает топаз (8), то его твердость принимается равной 7. Аналогично построены шкалы силы ветра Бофорта и землетрясений Рихтера. Со временем стали строиться шкалы измерений более далеких от физики объектов, вплоть до явлений общественного характера.

Этот процесс в конечном счете привел к современному понятию измерения и шкалы измерения. Теперь под *измерением* понимают процедуру, с помощью которой объекты исследования отображаются в некоторую математическую систему. В общественных науках в качестве таких систем чаще всего используются числовые множества, а сам процесс приписывания числовых значений качественным показателям называют *квантификацией.*

Как известно, существуют прямые и косвенные процедуры измерения. Прямые – когда выбирается единица измерения и подсчитывается, сколько раз она содержится в объекте измерения. Косвенные – когда измеряют вспомогательные величины, а потом, пользуясь связями, существующими между «измеряемой» величиной и измеренными вспомогательными величинами, вычисляют требуемую величину. В педагогике тоже применяются как прямые, так и косвенные процедуры. Прямые – когда измеряют рост, остроту зрения, время выполнения задания, число допущенных ошибок и т.д. Однако более распространены косвенные методы измерения, например, измерение уровней знаний, умений и навыков, которые проявляются в ответах учеников, качестве и способах решения ими определенных классов задач, успешности выполнения определенных видов деятельности.

Чем сложнее явление, тем труднее провести его измерение. Сравнительно легко измерить остроту зрения, труднее – механическую память, еще труднее – выносливость, настойчивость, волю. С другой стороны, чем глубже проникает педагогика в структуру связей между психической и поведенческой деятельностью человека, тем больше появляется возможностей для оценки внутренних факторов на основе внешней наблюдаемой деятельности.

Важным средством таких измерений служит *тест* – система заданий стандартной формы, выполнение которых должно выявить определенные свойств личности, интересы, мотивацию, объем и качество знаний и т.д.

1.5. Шкалы измерений. В процессе применения к полученным числовым значениям вычислительных процедур, не искажающих смысла моделируемого явления, важно знать, каким образом, с помощью какой процедуры измерения были получены эти числа. В связи с этим под *шкалой измерения* стали понимать алгоритм, с помощью которого осуществляется квантификация. Приписываемые же объектам числа называют *шкальными значениями* этих объектов. Единственное требование, предъявляемое к шкальным значениям, состоит в том, что рассматриваемые эмпирические отношения должны переходить в соответствующие им числовые отношения.

Вводя тот или иной оценочный показатель, исследователь должен продумать правило вычисления его числовых значений и соответствующую ему шкалу измерений. Существует несколько основных типов шкал, каждому из которых соответствует своё множество допустимых преобразований. Назовем основные применяемые в статистике шкалы по мере убывания их вычислительных возможностей (мощности). Ими являются шкала отношений, интервальная шкала, шкала порядка или ранговая шкала и, наконец, шкала наименований или номинальная шкала.

Шкалы отношений. В них основным способом квантификации (приписывания измеряемым объектам числовых значений) является классическая процедура, в соответствии с которой задается некоторая единица измерения и определяется, сколько раз она или некоторая ее часть «содержится» в измеряемом объекте. Важно заметить, что при этом числу нуль (началу отсчета) соответствует полное отсутствие измеряемого качества. Таким образом, если некоторое явление описывается шкалой отношений, то последняя определяется выбранной единицей измерения. В процессе исследования приходится иногда менять единицу измерения. При этом преобразуется шкала, оставаясь шкалой отношений. Если единицу измерения шкалы отношений умножить на положительное число k, то все шкальные значения измеряемых величин разделятся на это число; если положительное число k больше 1, то уменьшатся; если же меньше 1, то увеличатся. Например, если в шкале для измерения масс единицу измерения в один грамм заменить единицей в один килограмм, то все шкальные оценки уменьшатся в тысячу раз. Аналогично изменяется и *дистанция* между любой парой величин, т.е. разность их шкальных значений. Отсюда следует, что при умножении единицы измерения на положительное число k отношение шкальных значений двух величин, как и отношение дистанций двух пар величин, не изменяется, являясь, как говорят в математике, инвариантами преобразований шкалы отношений. Отсюда в свою очередь следует, что, если дистанции между двумя парами величин были равны до рассмотренного выше преобразования шкалы отношений, то они будут равны и после такого преобразования. Это свойство называют законом *сохранения равенства дистанций*.

Заметим также, что в шкале отношений над шкальными значениями можно выполнять и осмысленно интерпретировать все арифметические операции и вычислять величины, важные для анализа изучаемого явления.

Интервальные шкалы. Интервальные шкалы отличаются от шкал отношений тем, что в них начало отсчета выбирается произвольно. Таким образом, в ней не существует ни естественного начала отсчета, ни естественной единицы измерения. Переход от одной интервальной шкалы к другой осуществляется изменением единицы измерения и сдвигом начала отсчета. Если единицу измерения умножить на положительное число k, то все шкальные значения измеряемой величины (шкальные значения), как и в предыдущем случае, разделятся на число k. Если после этого начало отсчета сдвигается на L единиц новой шкалы в положительном ее направлении. то от всех полученных после выполнения первого шага значений надо отнять число L. Если сдвиг осуществлялся влево, то число L прибавляется.

К таким шкалам относится, например, шкала температур по Цельсию. При переходе от нее к шкале Кельвина все шкальные значения умножаются на единицу, и к ним прибавляется одно и то же число 273.

Легко проверить, что в интервальной шкале уже не выполняется закон постоянства отношений. Пусть, например, два тела нагреты: первое до 60^0C, а второе до 30^0C, и, следовательно, отношение температур первого ко второму равно 2. При переходе в шкалу Кельвина мы получим значения 333^0K и 303^0K, отношение которых равно 1.1. В то же время в шкале интервалов выполняется закон *сохранения равенства дистанций*.

Интервальную шкалу с фиксированной единицей измерения (рост в сантиметрах, возраст в годах, доход в рублях, стаж в годах) часто называют *шкалой разностей*.

Порядковые шкалы. Измерение в таких шкалах возможно тогда, когда исследователь может обнаружить в предметах различие степеней признака или свойства. На этой основе формируются классификационные группы, при обозначении которых используется свойство упорядоченности чисел. Чем сильнее у членов квалификационной группы выражено исследуемое качество, тем большее число ему приписывается (иногда принимается обратный порядок). Порядковая шкала не только задает некоторую классификацию на множестве объектов, но и устанавливает определенный порядок между квалификационными группами. Приписываемые группам числовые значения обычно называют *уровнями*. В результате такого измерения каждому измеряемому объекту присваивается шкальная оценка (уровень), равная номеру соответствующей классификационной группы. Результатом измерений является нестрогое упорядочение объектов.

В такой шкале, как мы видели выше, «измеряется» твердость минералов, сила ветра и землетрясений. Шкалы порядка широко используются в педагогике, психологии, медицине и других науках, не столь точных, как, скажем, физика или химия. В таких шкалах фиксируются квалификационные разряды, возрастные группы, эмоциональные состояния, уровни утомляемости, успешность обучения.

Часто, когда совокупность не очень большая, в такой системе измерения удается ранжировать не группы, а сами члены совокупности, приписывая каждому индивиду свой порядковый номер (ранг). Такая шкала только упорядочивает объекты, как правило, не образуя классификационные группы.

Естественно, что в порядковых шкалах уже не имеют смысла арифметические операции над шкальными значениями, например, сложение, вычитание или деление твердости минералов, уровни квалификации или спортивных достижений. Тем более не имеет смысла среднее арифметическое уровней. В них уже нельзя говорить о том, на сколько или во сколько раз одна измеряемая величина больше (меньше) другой величины.

Шкалы наименований (номинальные, классификационные шкалы). При построении шкалы наименований предметы группируются в классы (по некоторому признаку). Каждому классу дается наименование и числовое обозначение. Шкалы наименований применяются, если в качестве моделируемых эмпирических отношений выступают лишь отношения равенства и неравенства между объектами. Например, по полу, месту жительства, национальности и т. п.

Глава 2. Статистические критерии

2.1. Использование различных шкал измерения в педагогических исследованиях. В педагогических исследованиях используются все перечисленные выше шкалы измерений. Естественно, что наиболее предпочтительными являются шкалы отношений или хотя бы интервальные шкалы, прежде всего из-за их больших вычислительных возможностей. К сожалению, воспользоваться ими можно далеко не всегда. Для этого нужно иметь, как минимум, достаточно устойчивую, не меняющуюся в процессе исследования единицу измерения.

В общем случае принято выделять следующие виды оценочных показателей, измеряемых в шкале отношений:

– временные (время выполнения действия, операции, время реакции, время, затрачиваемое на исправление ошибки, и т.д.);

– скоростные (производительность труда, скорость реакции, движения и т.д. – величины, обратные времени);

– точностные (величина ошибки, количество ошибок, вероятность ошибки, вероятность точной реакции, действия и т.д.);

– информационные (объем заучиваемого материала, перерабатываемой информации, объем восприятия и т.д.).

К ним можно также отнести количество правильных, точно сформулированных, не вызывающих сомнений ответов на серию коротко и точно сформулированных вопросов заданий и т.д. К их числу могут быть отнесены и результаты выполнения контрольных работ, тестов, если их построить по тому же принципу, а также результаты обработки анкет (процент учащихся, давших положительные ответы на тот или иной вопрос) и т.д.

В педагогических исследованиях вопрос о возможности отнесения результатов измерения к шкале отношений или интервальной шкале во многом определяется качеством тестов, используемых для измерения оценочных показателей. Для того чтобы выполненные с его помощью измерения можно было считать выполненными в шкале отношений, необходимо, чтобы тест был не только достаточно надежным, дискриминативным и валидным (адекватным выбранному оценочному показателю), но и однородным по сложности используемых в нем тестовых заданий. Часто в качестве средства такого выравнивания применяется балльная (весовая) система оценки тестовых заданий, при которой им приписываются баллы (веса), пропорциональные времени, которое в процессе разработки теста затрачивали на их выполнение специально привлекаемые для этого испытуемые. В такой системе оценочный показатель измеряется количеством набранных баллов, соответствующих правильно выполненным тестовым заданиям. Такие баллы, иногда называемые *количественными*, и играют в этом случае роль единицы измерения.

В тех случаях, когда построить такой тест не удается, измерение приходится проводить в шкале порядка. В частности, повсеместно распространенная система школьных отметок в баллах (их будем называть *уровневыми баллами*) является одной из многих возможных уровневых систем (пятибалльная, десятибалльная и т.д.). Все такие системы в общем случае должны быть отнесены к шкале порядка, ибо в ней не выполняется равенство дистанций между уровнями: превосходство отличника над хорошистом существенно отличается от превосходства хорошиста над троечником или троечника над двоечником, особенно если, как это часто бывает, оценки используются в воспитательных целях.

Следует иметь в виду, что использование порядковой шкалы в педагогических исследованиях связано с серьезными трудностями. И дело здесь не только в известной необъективности отметок, но и в свойствах самой шкалы порядка. Как отмечалось выше, в этой шкале ничего нельзя сказать о равномерности или неравномерности интервалов между соседними значениями оценок. Поэтому в уровневой шкале недопустимы утверждения типа «знания учащихся в экспериментальных классах в среднем на 0,5 балла выше, чем в контрольных», или что «эффективность экспериментальной методики в 2,6 раза выше контрольной».

В номинальных шкалах шкальным значениям (числам) не приписывается никаких свойств, кроме тождества и различия. Даже при таком простейшем измерении к построению шкалы надо подходить с большой осторожностью. Формируемые классы должны иметь педагогическую значимость. Исследователь должен решить, что он будет классифицировать, какие категории при этом будут исследоваться.

2.2. Агрегированные оценки. Как правило, в любом педагогическом эксперименте участвует значительное число учеников, учителей, образовательных учреждений и т.д. В результате измерения оценочных показателей этих участников получается набор их *индивидуальных оценок.* Понятно, что сравнивать между собой и анализировать одновременно все индивидуальные оценки нецелесообразно, так как всегда существует их разброс, обусловленный неконтролируемым различием участников эксперимента. Поэтому для того, чтобы получить обозримое число характеристик и сгладить влияние неизбежно возникающих колебаний, используют так называемые *агрегированные* (коллективные, групповые) *оценки.* Правило вычисления агрегированных оценок, если исключить их бытовое использование, зависит от шкалы, в которой были получены индивидуальные оценки. Получение агрегированных оценок на основании индивидуальных является их преобразованием, и преобразование это следует выполнять корректно. Приведем некоторые корректные процедуры агрегирования для наиболее распространенных в педагогических исследованиях показателей.

Для абсолютных величин, измеренных в шкале отношений, наиболее типичным является вычисление среднего арифметического по группе. Эта процедура вполне корректна, и обычно ее реализация не вызывает затруднений. Пусть, например, в процессе измерения в шкале отношений получено 20 индивидуальных оценок, представленных в таблице 2.1.

Таблица 2.1

Время в минутах, которое затратили ученики первого класса при чтении заданного текста

2,3	10,5	6,1	1,1	5,9	4,8	7,2	8,5	7,9	9,8
5,6	6,5	6,9	4,3	6,2	9,2	11,3	7,1	2,7	13,1

Для вычисления среднего арифметического некоторой совокупности чисел надо их сумму разделить на их количество. В рассмотренном выше примере среднее арифметическое, обозначаемое символом $\bar{x}$, равно $\frac{137}{20} = 6{,}85$. Иногда в качестве агрегированной оценки полученных индивидуальных оценок принимается их *медиана* – число, разделяющее выборку на две части, равные по количеству входящих в них чисел, в одной из которых все числа не больше, а во второй – не меньше этого разделяющего числа. В рассмотренном случае медиана расположена между числами 6,5 и 6,9. Упрощая расчет, положим ее равной середине отрезка, образованного этими числами, а именно 6,7.

В ряде случаев в качестве агрегированной оценки, характеризующей группу, используется выраженная в процентах *доля* учащихся, справившихся (не справившихся) с заданием. В общем случае *долей* в статистике называют отношение k/n, где k – число учащихся, справившихся (не справившихся) с заданием, а n – общее число испытуемых. Обозначив долю буквой p, получим, что p=k/n. При выражении доли в процентах ее значение умножают на 100. Долю в генеральной совокупности обозначают буквой π.

Принципиально иначе агрегируются показатели, измеренные в шкале порядка. Если имеется набор индивидуальных уровневых баллов, то агрегированной оценкой группы будет число ее членов, получивших тот или иной уровневый балл. Пусть, например, при подведении итогов учебной деятельности оценку «отлично» получили 4 ученика, оценку «хорошо» – 11 учеников, оценку «удовлетворительно» – 12 учеников и оценку «неудовлетворительно» – 2 ученика. В этом случае агрегированной оценкой уже не может служить широко распространенный в школьной практике средний балл. В этом случае такой агрегированной «оценкой» может служить именно приведенное выше указание числа учащихся, получивших каждый из принятых в системе баллов. В рассмотренном примере такая агрегированная «оценка» может быть представлена таблицей 2.2.

Таблица 2.2

Агрегированная «оценка», характеризующая учебные успехи класса

Уровневые баллы	1	2	3	4	5
Количество учеников	0	2	12	11	4

Иногда в качестве агрегированной оценки индивидуальных уровневых баллов в шкале порядка вместо неприемлемого в этом случае среднего арифметического («среднего уровневого балла») применяется понятие *медианы.* Правда, при использовании порядковых шкал, имеющих малое число «разрядов» – «порядковых баллов», медиана малоинформативна, что хорошо видно при рассмотрении приведенного выше примера (см. табл. 2.2), в котором медианой служит оценка «удовлетворительно».

2.3 Основные понятия описательной статистики. Обратимся теперь к рассмотрению простейших фактов так называемой описательной статистики. При этом, прежде всего, рассмотрим случай, когда измерения выполняются в интервальной шкале, или шкале отношений. В этом случае результатом индивидуальных измерений будет служить система действительных чисел, $\langle x_1, x_2, ..., x_n \rangle$, которую называют *числовой выборкой*. В таблице 2.1 приведен пример такой выборки.

При изображении числовой выборки точками числовой оси легко заметить, что располагаются они, как правило, неравномерно, с разной плотностью в разных местах оси. Зависимость плотности расположения точек выборки от места, занимаемого ими на числовой оси, называют *выборочным распределением.* Иногда, если хотят подчеркнуть, что выборочные значения получены из опыта, говорят об *эмпирическом* распределении.

Разбив отрезок, содержащий все точки выборки, на конечное количество равных «элементарных» отрезков и посчитав, сколько выборочных значений попадает в каждый из построенных интервалов, получают интервальное распределение выборочных значений. В таблице 2.3 выражено интервальное распределение данных, приведенных в таблице 2.1.

Таблица 2.3

Интервальное распределение элементов выборки, приведенной в таблице 2.1

Интервалы	0 – 1,9	2 – 3,9	4 – 5,9	6 – 7,9	8 – 9,9	10 – 11,9	12 – 13,9
Частота	1	2	4	7	3	2	1

Для того чтобы сделать это распределение более наглядным, на выделенных «элементарных» отрезках строят прямоугольники с высотами, а следовательно, и площадями, пропорциональными числу точек выборки, оказавшихся в их основаниях. Полученную диаграмму в статистике называют *гистограммой*. В рассматриваемом примере (см. таблицу 2.3) получается гистограмма, изображенная на рисунке 2.1.

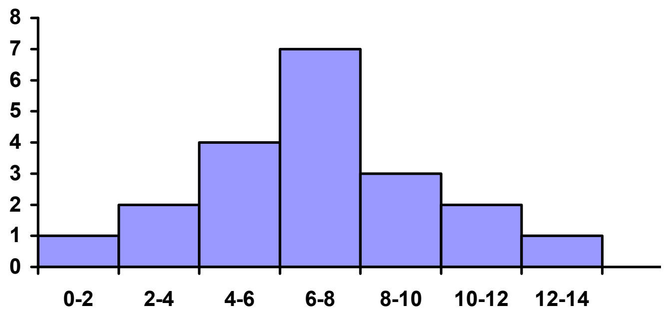


Рис. 2.1

Если число элементарных отрезков достаточно велико, а сами они малы, то, проведя через середины верхних оснований построенных прямоугольников плавную кривую, получают так называемую *кривую выборочного распределения.*

Выборочные распределения обладают рядом числовых характеристик, которые называют *статистическими параметрами распределения*. К статистическим параметрам, характеризующим положение точек выборки на числовой оси, относятся уже упоминавшиеся ранее *среднее арифметическое и медиана.* Заметим только, что среднее арифметическое всех членов генеральной совокупности принято называть *математическим ожиданием* и обозначать буквой μ (мю). Среднее арифметическое выборки с увеличением ее объема неограниченно приближается к μ и поэтому может быть использовано для оценки математического ожидания.

К статистическим параметрам, характеризующим разброс выборочных значений около среднего арифметического, относятся *дисперсия* и *стандартное отклонение*. Дисперсия выборки обозначается символом s^2 и вычисляется по формуле $s^2 = \frac{1}{n-1}\left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2\right]$, а стандартное отклонение s – извлечением квадратного корня из дисперсии.

Заметим, что при вычислении дисперсии элементов всей генеральной совокупности, обозначаемой символом σ^2, пользуются той же формулой, за исключением того, что сумма делится не на разность n-1, а на само число n, что, вообще говоря, при достаточно больших значениях n не существенно.

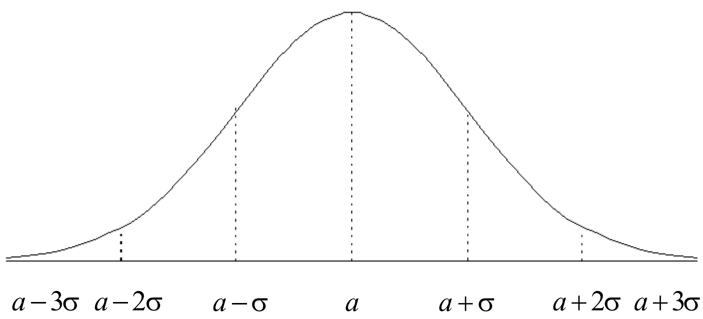


Рис. 2.2

Особый интерес представляют распределения, в которых плотность точек максимальна в окрестности среднего арифметического и монотонно снижается до нуля по мере удаления как вправо, так и влево от него. Если при этом зависимость плотности $\varphi(x)$ от координаты x выражается формулой Гаусса $\varphi(x) = \frac{1}{s\sqrt{2\pi}} \cdot e^{-\frac{(x-\bar{x})^2}{2s^2}}$, где $\bar{x}$ – среднее арифметическое, а s – стандартное отклонение, то такие распределения называют *нормальными*. График подобного распределения приведен на рисунке 2.2. Существуют методы, позволяющие с заданной надежностью ответить на вопрос: подчиняется ли рассматриваемое распределение нормальному закону. Соответствующие методы, называемые *критериями согласия*, приведены в главе четвертой настоящего пособия. Часто в приблизительной нормальности распределения убеждаются визуально, построив его гистограмму или кривую распределения:

Обращаясь к случаю, когда оценочный показатель измеряется в долях (см. пункт 5 главы 1), заметим только, что доля р, вычисленная по выборке, оценивает соответствующий параметр π генеральной совокупности, так же, как среднее арифметическое $\bar{x}$ оценивает математическое ожидание μ. Более того, р, как оценка π, обладает всеми свойствами, присущими среднему арифметическому $\bar{x}$ как оценке математического ожидания. В частности, можно доказать, что при достаточно больших значениях n, при которых каждое из значений np и $n(1-p)$ больше пяти, выборочное распределение p можно считать нормальным со средним выборочных значений p и стандартным отклонением $\sigma=\sqrt{p(1-p)/n}$.

Обратимся теперь к работе с оценками, полученными в уровневой шкале. Как уже говорилось в предыдущей главе, агрегированной оценкой в этом случае служит таблица, в которой указаны принятые в данной системе измерения уровни и соответствующее каждому из них число учащихся, достигших этого уровня (см., например, таблицу 1.2). Часто такие таблицы по аналогии с рассмотренными выше таблицами интервального распределения называют таблицами распределения уровневой случайной величины. Заметим, что если известны агрегированные оценки нескольких совокупностей измеренных в одной и той же уровневой шкале (в одной и той же системе уровней), то нетрудно построить агрегированную оценку их объединения. Для этого надо суммировать количество респондентов, достигших каждого из допустимых уровней.

2.4. Проблемы распространения результатов эксперимента на всю генеральную совокупность. Как известно, при экспериментальном обосновании эффективности разработанной педагогической системы приходится сопоставлять результаты выполнения членами экспериментальной и контрольной групп тестовых заданий (или других средств измерения) по каждому оценочному показателю. При этом сопоставляются либо сами числовые выборки $\langle x_1, x_2,...,x_r \rangle$ и $\langle y_1, y_2,..., y_l \rangle$, полученные в результате измерений соответствующего оценочного показателя, либо связанные с ними параметры – средние арифметические, медианы, дисперсии и так далее. На основании таких сопоставлений уже можно сделать первые предварительные выводы об эффективности разработанной педагогической системы. Пусть, например, при выполнении заключительного проверочного теста среднее значение величин, характеризующих заданный оценочный показатель, в экспериментальной группе оказался равен 14,6, а в контрольной группе – 13,7, откуда следует, что экспериментальная группа справилась с контрольным заданием лучше, чем контрольная. Возникает важный вопрос: с какой надежностью результаты эксперимента можно распространить на всю генеральную совокупность. Иначе говоря, с какой вероятностью р можно утверждать, что, отобрав случайным образом из той же генеральной совокупности другие две группы и проведя с ними ту же работу, мы вновь получим, что экспериментальная группа справится с заданием лучше, чем контрольная. Без ответа на этот вопрос весь эксперимент теряет свою доказательную силу. На неформальном уровне, представив себе гистограммы сопоставляемых выборок на одной числовой оси, естественно предположить, что упомянутая выше вероятность повышается с увеличением разности средних арифметических сопоставляемых выборок и уменьшением разброса значений каждой из них. В педагогических исследованиях, как правило, стремятся обеспечить 95-процентную надежность $P = 0.95$, которая означает, что при ста проверках несовпадение с выводами эксперимента может произойти не более чем в пяти случаях.

2.5. Понятие статистического критерия. Более строго вопрос о возможности распространения выводов эксперимента на всю генеральную совокупность решается с помощью так называемых *статистических критериев*. Каждый критерий представляет собой либо формулу, либо словесно сформулированное правило, позволяющее по выборкам вычислять его значение. Каждому критерию соответствует своя *теоретическая кривая распределения*, которую можно представить в виде непрерывной кривой, простирающейся над числовой осью и обладающей в общем случае следующими свойствами (см. рис. 2.3):

– площадь криволинейной трапеции, образуемая кривой распределения и числовой осью, равна единице;

– вероятность того, что критерий примет значение, большее z, равна площади части криволинейной трапеции, расположенной справа от прямой, проходящей через точку z, перпендикулярно числовой оси.

Саму точку z в этом случае принято обозначать символом ${}_{\alpha}K$, где буква K обозначает используемый критерий, а индекс α – упомянутую выше вероятность, с которой критерий может принять значение, большее z. Вид кривой распределения определяется свойствами критерия и числом степеней свободы (ν), которое равно числу *независимых* переменных, образующих выборку (выборки). Для вычисления ν от общего числа переменных, входящих в выборки, надо отнять число связей, налагаемых на них критерием. Обычно правило вычисления числа степеней свободы приводится вместе с критерием. С помощью кривой распределения критерия вычисляются так называемые *критические точки критерия*, соответствующие некоторым стандартным значениям надежности, с которой выводы эксперимента могут быть распространены на всю генеральную совокупность. Обычно для каждого уровня надежности P=100*p определяются две критические точки, такие, что вероятность попадания значения критерия между ними равна p. Эту область называют *областью принятия гипотезы.* Точки, лежащие вне этого интервала, образуют так называемую *критическую область*. Попадание в нее вычисленного значения критерия ведет к отклонению исследуемой гипотезы с вероятностью α, равной 1-p (α=1- p). Обычно критические точки выбираются с та-

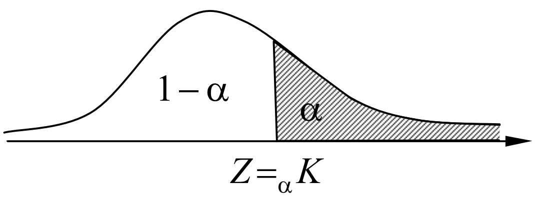


Рис. 2.3

ким расчетом, чтобы вероятности попадания в каждую из двух частей критической области были равны $\alpha/2$. В этом случае правая (верхняя) критическая точка может быть обозначена символом ${}_{\alpha/2}K$, а левая (нижняя) – символом ${}_{1-\alpha/2}K$. В частности, если $\alpha = 0{,}05$, что соответствует 95-процентной надежности, то верхнюю и нижнюю критические точки можно записать соответственно в виде: ${}_{0.025}K$ и ${}_{0.975}K$.

Применение статистических критериев чаще всего осуществляется в форме, напоминающей «доказательство методом от противного». С этой целью для каждого оценочного показателя формулируется так называемая нулевая гипотеза о том, что в масштабах всей генеральной совокупности разрабатываемая педагогическая система никаких преимуществ не дает и в общем случае результаты экспериментальной группы не отличаются от результатов контрольной группы. Эту гипотезу обозначают символом H_0. Одновременно формулируется так называемая альтернативная гипотеза, в определенном смысле обратная нулевой гипотезе. Ее обозначают символом H_1. Если, например, нулевая гипотеза заключается в том, что математическое ожидание μ_1, соответствующее экспериментальной методике, равно математическому ожиданию μ_2, соответствующего традиционной методике, что записывается как $H_0 : \mu_1 = \mu_2$., то альтернативной гипотезой будет гипотеза $H_1 : \mu_1 \neq \mu_2$.

Затем проводится сам педагогический эксперимент, цель которого, – опровергнуть нулевую гипотезу. Сформировав соответствующим образом экспериментальную и контрольную группы и реализовав в экспериментальной группе свою педагогическую систему, исследователь с помощью тестов или других средств измерения для каждой из этих групп находит индивидуальные значения анализируемого оценочного показателя. С помощью полученных числовых выборок исследователь вычисляет значение используемого критерия. Если это значение попадает в критическую область, то нулевая гипотеза отклоняется с заданной выше надежностью

$P = p \cdot 100\%$ или, как принято говорить в статистике, на заданном уровне статистической значимости $\alpha = 1 - P$, где р – упомянутая выше вероятность. Отсюда в свою очередь следует возможность распространения на выбранном уровне статистической значимости (надежности) полученных в эксперименте различий на всю генеральную совокупность. При использовании упрощенной схемы педагогического эксперимента говорят о неслучайном характере обнаруженных различий. В противном случае делают вывод, что на заданном уровне статистической значимости (надежности) нет основания для отклонения нулевой гипотезы, а, следовательно, и для распространения выводов эксперимента на всю генеральную совокупность.

Объем генеральной совокупности, если он существенно превышает объемы выборок, практически не влияет ни на значения критерия, ни на значения критических точек. Если же объем выборок n превышает $1/25$ объема генеральной совокупности, то критические значения следует умножать на $\sqrt{\frac{N-n}{N-1}}$, где N – ориентировочный объем генеральной совокупности, или разделить на него вычисленное значение критерия.

До сих нор мы рассматривали наиболее распространенный случай, когда критическая область критерия распадается на две равновероятные для него части, расположенные в концах кривой распределения критерия. Этот случай соответствует ненаправленной альтернативной гипотезе, в которой лишь утверждается факт неравенства параметров, например, гипотеза $H_1 : \mu_1 \neq \mu_2$ как альтернативная гипотезе $H_0 : \mu_1 = \mu_2$.

Ситуация существенно меняется, если исследователь в качестве гипотезы альтернативной гипотезе $H_0 : \mu_1 = \mu_2$ выдвинет гипотезу $H_1 : \mu_1 > \mu_2$. В этом случае альтернативная гипотеза окажется направленной: исследователь как бы утверждает, что μ_1 или равно или, возможно, больше нуля, но во всяком случае не меньше μ_2. Поэтому он каждый случай, когда

$\bar{x}_1 < \bar{x}_2$ будет относить в пользу нулевой гипотезы. В этом случае критическая область отклонения нулевой гипотезы будет лежать в правом (левом) конце критерия. В таблицах приложения тип альтернативной гипотезы помечен символами: (2), если она ненаправленная, и (1), если направленная.

2.6. Основные типы статистических критериев. Статистические критерии, оценивающие возможность распространения выводов исследования на всю генеральную совокупность, принято делить на два основных типа в зависимости от того, сопоставляются статистические параметры эмпирических распределений (их средние арифметические, медианы, дисперсии и т.п.) или сами эти распределения.

Критерии, в которых сопоставляются статистические параметры эмпирических распределений, называют *параметрическими.* Критерии, в которых сопоставляются сами, выборочные распределения – *непараметрическими*.

Сразу заметим, что для применения параметрических критериев необходимо, чтобы измерения проводились в шкале отношений или интервальной шкале, а распределения результатов измерения были близки нормальному или некоторому другому, хорошо изученному и табулированному распределению. Эти требования, привычные для физики и других естественных наук, создают серьезные трудности при использовании параметрических критериев в гуманитарных науках. Не случайно последние годы в педагогических исследованиях все чаще стали использоваться непараметрические критерии. Этому есть существенные оправдания, однако, как показывает анализ некоторых работ, где применяются непараметрические критерии, при более аккуратной организации тестирования вполне можно было бы воспользоваться параметрическими критериями.

В основе непараметрических критериев лежат различные способы измерения расстояния между двумя распределениями. При использовании непараметрических критериев не имеет значения ни шкала, в которой проводились измерения, ни вид полученных распределений. Однако не следует думать, что непараметрические критерии вообще не чувствительны к процедурам измерения и особенностям полученных распределений. Все «выбросы», существенно нарушающие нормальность распределения, затрудняют интерпретацию результатов эксперимента. Следует также иметь в виду, что для достижения одной и той же надежности непараметрические критерии требуют выборки существенно большего объема, чем параметрические критерии.

В заключение заметим, что непараметрические методы рекомендуются к применению тогда, когда параметрические методы слишком чувствительны к отклонениям от сделанных допущений. Или же удовлетворение этим допущениям с помощью соответствующих преобразований или с помощью устранения выбросов представляет значительные трудности.

Наконец, надо иметь в виду, что в любом случае, в котором применимы параметрические критерии, могут быть применены и непараметрические критерии, для чего придется от сопоставления параметров перейти к сопоставлению самих выборочных распределений. Нужно также помнить, что для получения прежней надежности придется увеличить объем выборок или согласиться со снижением уровня надежности. Все сказанное относится и к так называемым быстрым критериям, которые, как правило, построены на основе непараметрических статистик и обладают еще меньшей надежностью.

В заключение заметим, что все виды статистических критериев применимы как на заключительном этапе педагогического эксперимента, так и на его начальном этапе, когда доказывается совпадение экспериментальной и контрольной групп.

Глава 3. Параметрические критерии

3.1. t-критерий Стьюдента для *независимых* выборок. При сравнении независимых выборок одним из наиболее часто применяемых *параметрических* критериев является t-критерий Стьюдента, выражаемый формулой $t = \dfrac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$, в которой $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_2} \right)}$, где n_1 и n_2 – объемы экспериментальной и контрольной групп, $\bar{x}_1$ и $\bar{x}_2$ – их выборочные средние, а S_1^2 и S_2^2 – их выборочные дисперсии.

Кривая распределения критерия Стьюдента зависит от числа степеней свободы $\nu = n_1 + n_2 - 2$. При достаточно больших значениях ν эта кривая практически неотличима от кривой нормального распределения с вершиной в начале координат. С уменьшением значений ν вершина кривой опускается, а «хвосты» поднимаются. Площадь же под кривой остается постоянной.

При использовании t-критерия Стьюдента рекомендуется строить выборки одинаковой численности или выравнивать их численность, исключая случайным образом лишние элементы из большей выборки. Легко проверить, что если $n_1 = n_2 = n$, то выражение $S_{\bar{x}_1 - \bar{x}_2}$ принимает вид: $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{S_1^2 + S_2^2}{n}}$.

Критические значения (точки) t-критерия Стьюдента приведены в Приложении. Например, при двухстороннем критерии (когда критическая область располагается по обе стороны от области принятия нулевой гипотезы), на уровне статистической значимости $\alpha = 0{,}05$ и числе степеней свободы $\nu = 24$, верхнее критическое значение ${}_{0.025}t$ будет равно 2,064. В силу симметричности кривой распределения относительно прямой, проходящей через начало отсчета перпендикулярно числовой оси, нижняя критическая точка критерия $_{0.975}t$ будет равно -2,064. В результате определяются две области, одна из которых заключена между числами минус 2,064 и 2,064, а другая, критическая, состоит из чисел, не принадлежащих этому интервалу. Если вычисленное значение критерия $t_{выч}$ попадает в критическую область, т.е. $t_{выч} \le -2{,}064$ или $t_{выч} \ge 2{,}064$, то нулевую гипотезу отклоняют на уровне статистической значимости $\alpha = 0{,}05$ (95% надежности). Если же $t_{выч}$ попадает в интервал (-2,064; 2,064), то это означает, что нет оснований для отклонения нулевой гипотезы на заданном выше уровне значимости.

Пример. Предположим, что исследователь разработал новый метод обучения младших школьников чтению. В качестве оценочного показателя он принял время чтения заданного текста в минутах (x). Обозначим буквой μ_1 – среднее для генеральной совокупности время чтения тестового задания, при обучении детей пользуясь традиционной методикой, а буквой μ_2 – при использовании новой методики. Пользуясь этими обозначениями, нулевую гипотезу можно записать в виде $H_0: \mu_1 = \mu_2$, а альтернативную ей, ненаправленную гипотезу, в виде $H_1: \mu_1 \ne \mu_2$. Для доказательства преимущества нового метода надо экспериментом прежде всего опровергнуть нулевую гипотезу и только потом обосновывать, что $\mu_1 > \mu_2$.

С этой целью исследователь, прежде всего, определил в качестве генеральной совокупности детей шестилетнего возраста, посещающих детские сады города N, не имеющих нарушений речи, не умеющих читать, но знающих все буквы алфавита и проявляющих личное желание научиться читать, таких детей в детских садах города оказалось около 360. Отобрав из них *случайным образом* 40 ребятишек, исследователь сформировал из них тоже *случайным образом* четыре группы по 10 человек в каждой. Две из этих групп он использовал в качестве экспериментальных, а две другие – в качестве контрольных. Полагаясь на то, что случайный отбор обеспечивает примерно одинаковый состав отобранных групп, исследователь приступил к обучению. Детей экспериментальных групп он обучал, руководствуясь разработанной им методикой. Детей контрольных групп – руководствуясь традиционной методикой. Обучение во всех группах длилось одинаковое время и примерно с одинаковой интенсивностью. После завершения обучения испытуемым был предложен тест для чтения. В таблице 3.1 приведено время (в минутах), затрачиваемое на его чтение каждым испытуемым экспериментальной и контрольной групп.

Таблица 3.1.

Результаты тестирования на быстроту чтения (в минутах) учащихся экспериментальной и контрольной групп.

1	Экспериментальная группа	0,4	1,1	1,2	1,3	1,5	1,7	1,9	2	2,1	2,1	2,1	2,3	2,3	2,4	2,6	2,8	2,9	3,2	3,4	3,5
2	Контрольная группа	0,8	1,3	1,6	1,7	2,1	2,2	2,3	2,5	2,6	2,6	2,8	2,9	2,9	3,4	3,4	3,6	3,7	4,3	4,4	4,6

Прежде чем вычислять выборочные средние и дисперсии, выясним, можно ли при сопоставлении данных воспользоваться параметрическим критерием. Одно условие такого применения выполняется: все измерения выполнены в шкале отношений. Остается проверить близость образовавшихся распределений нормальному распределению. Для этого, распределив испытуемых по интервалам времени, построим таблицу 3.2, в каждой ячейке которой указано число испытуемых, время чтения которых соответствует указанному интервалу.

Таблица 3.2.

Распределение испытуемых по интервалам времени

Интервал	0,0 0,4	0,5 0,9	1,0 1,4	1,5 1,9	2,0 2,4	2,5 2,9	3,0 3,4	3,5 3,9	4,0 4,4	4,5 5,0	Итого
Экспер.	1	0	3	3	7	3	2	1	0	0	20
Контр.	0	1	1	2	3	6	2	2	2	1	20

Из данных, приведенных в таблице 3.2, следует, что распределение испытуемых экспериментальной группы почти нормально. Достаточно близко к нормальному распределению и распределение испытуемых контрольной группы. При сомнении можно обратиться к одному из критериев согласия, приведенному в главе 4 пособия. Так как сопоставляемые группы независимы, в качестве оценочного критерия можно применить указанный выше t-критерий Стьюдента.

Расчеты показывают, что среднее арифметическое экспериментальной группы $\bar{x} = 2{,}14$, а контрольной группы $\bar{y} = 2{,}785$. Выборочные дисперсии в этих группах равны, соответственно, $S_x^2 = 0{,}647$ и $S_y^2 = 1{,}074$. Эти данные позволяют, пользуясь формулой $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2 + S_2^2}{n}}$, получить $S_{\bar{x}_1 - \bar{x}_2} = 0{,}293$, откуда $t = \frac{2{,}785 - 2{,}14}{0{,}293} = 2{,}20$.

Воспользовавшись таблицей В Приложения, приведем в таблице 3.3 критические значения двухстороннего t-критерия Стьюдента при $\nu = n_1 + n_2 - 2 = 38$ степенях свободы.

Таблица 3.3.

Критические значения двухстороннего t-критерия Стьюдента при 38 степенях свободы

Надежность, в %%	80	90	95	98	99
Статистическая значимость, α	0,20	0,10	0,05	0,02	0,01
Верхние критические значения	1,303	1,684	2,021	2,423	2,704

Так как вычисленное значение критерия, равное 2,20, больше верхнего критического значения, соответствующего 95-процентной надежности, меньше следующего критического значения 2,423, то нулевую гипотезу отклоняем на уровне статистической значимости $\alpha = 0{,}05$ или, что то же самое, с 95-процентной надежностью.

Учитывая явное преимущество экспериментальной группы, можно признать, что выводы, сделанные из проведенного эксперимента, могут быть с 95-процентной надежностью распространены на всю генеральную совокупность.

Замечание 1. Так как отобранное число испытуемых (n=40) превосходит одну двадцать пятую часть генеральной совокупности (N=360), критические точки можно было умножить на коэффициент $\varepsilon = \sqrt{\dfrac{N-n}{N-1}}$, который в нашем случае равен $\varepsilon = \sqrt{\dfrac{360-40}{360-1}}$=0,944. При этом получились бы значения, приведенные в таблице 3.4.

Таблица 3.4.

Критические значения двухстороннего t-критерия Стьюдента при 38 степенях свободы с поправкой на коэффициент ε

Надежность, в %%	80	90	95	98	99
Статистическая значимость, α	0,20	0,10	0,05	0,02	0,01
Верхние критические значения	1,230	1,590	1,908	2,287	2,553

Новые критические точки в нашем случае не меняют выводы исследования. Но так бывает не всегда. Иногда такой переход позволяет перешагнуть очередной порог надежности.

Замечание 2. Если рассмотренная процедура кажется вам слишком громоздкой, то можно обратиться к упрощенной схеме эксперимента. В этом случае, не утруждая себя созданием генеральной совокупности и построением выборок, сразу начинают с формирования экспериментальной и контрольной групп одинакового состава. При доказательстве такого равенства надо проверить все характеристики, которые могут оказать существенное влияние на ход эксперимента, а не одну, как это часто практикуется в диссертационных исследованиях. В дальнейшем действия полностью совпадают с рассмотренными выше, за исключением завершающего этапа: в этом случае поправочный коэффициент уже не применяется, а вычисленное значение критерия сопоставляется непосредственно с соответствующими критическими точками критерия. Если оно попадает в критическую область, то различие в значениях экспериментальной и контрольной групп признается достоверным на заданном уровне значимости. В противном случае объявляется, что различие вызвано случайными причинами.

3.2. t-критерий Стьюдента для *зависимых* выборок. Рассмотрим теперь параметрический критерий, часто применяемый при сравнении зависимых выборок и тоже называемый критерием Стьюдента.

В этом случае основными элементами становятся не одиночные значения результатов измерения, а пары значений взаимосвязанных членов данных выборок, в данном случае их разности $d_i = x_{i1} - x_{i2}$. Часто элементами этих разностей являются результаты измерения каждого респондента вначале и в конце эксперимента. Естественно, что в этом случае число степеней свободы уменьшается в два раза. Затем, следуя общей схеме, вычисляется среднее арифметическое этих разностей $\bar{d} = \sum_{i=1}^{n} d_i / n$. Обозначив буквой μ_1 среднее значение этих разностей при распространении исследуемой методики на всю генеральную совокупность, а буквой μ_2 значение такой же величины при использовании традиционной методики, нулевую гипотезу можно записать в виде $H_0: \mu_1 = \mu_2$, а альтернативную гипотезу в виде $H_1: \mu_1 \neq \mu_2$. Используя среднее значение разностей, которое мы обозначили $\bar{d}$, и стандартное отклонение $S_d = \sqrt{\sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2 / (n-1)}$, получим величину $t = \dfrac{\bar{d}}{S_d / \sqrt{n}}$, которая и является t-критерием Стьюдента для *зависимых* выборок.

Доказано, что если верна нулевая гипотеза, то t в последнем выражении будет подчиняться t-распределению Стьюдента с числом степеней свободы $\nu = n - 1$, и поэтому критическими точками t-критерия Стьюдента в случае зависимых выборок будут те же, что и при независимых выборках, и, следовательно можно обращаться к одной и той же таблице В Приложения.

Пример. Для выяснения изменений, происходящих за пять лет обучения в вузе в уровне умений студентов решать задачи по курсу школьной математики, первокурсникам, а через четыре года тем же пятикурсникам давалось задание из десяти задач. Естественно, что при этом из рассмотрения исключались работы первокурсников, не дошедших до пятого курса, и пятикурсники, не писавшие эту контрольную на первом курсе. В результате было получено 55 пар значений типа «до-после», по ним была вычислена сумма образовавшихся разностей $\sum_{i=1}^{i=55}(x_{i2}-x_{i1})=86$, в которой x_{i1} – оценка, полученная i-м респондентом на первом курсе, а x_{i2} – оценка того же респондента, полученная на пятом курсе. Разделив полученную сумму на численность n=55, получим, что среднее значение $\overline{d}=0{,}71$, откуда следует, что студенты, попавшие в выборку на пятом курсе, выполнили тестовую контрольную лучше, чем на первом. Остается выяснить устойчивость полученного результата, уровень надежности, на котором этот вывод может быть распространен на всю генеральную совокупность.

Для этого, пользуясь формулой $S_d=\sqrt{\sum_{i=1}^{n}\left(d_i-\overline{d}\right)^2 \big/ (n-1)}$, найдем, что $S_d=2{,}47$. Для проверки $H_0:\mu_1-\mu_2=0$ на уровне значимости $\alpha=0{,}05$ вычислим значение t по формуле $\widetilde{t}=\dfrac{\overline{d}}{Sd/\sqrt{n}}$. Оно будет равно $\widetilde{t}=\dfrac{0{,}71}{2{,}47/\sqrt{55}}=2{,}13$. Верхняя критическая точка t-распределения с числом степеней свободы ν=55-1=54 будет равна: ${}_{0{,}025}t_{54}=2{,}005$. Так как вычисленное значение критерия больше критического, то, оставаясь в рамках упрощенной схемы педагогического эксперимента, исследователь должен на уровне $\alpha=0{,}05$ отклонить нулевую гипотезу и, следовательно, признать различие между экспериментальной и контрольной группами достоверным.

3.3. F-критерий Фишера. В некоторых педагогических исследованиях приходится выяснять, в какой степени рассматриваемая педагогическая система влияет на дифференциацию учащихся по уровню знаний: увеличивается или уменьшается дисперсия их оценок. Как и раньше, при этом приходится рассматривать отдельно случай, когда сопоставляемые выборки независимы и когда они зависимы. Ограничимся рассмотрением первого из этих случаев.

Используя символ σ^2 для обозначения дисперсии всех элементов генеральной совокупности, обозначим символом σ_1^2 дисперсию генеральной совокупности, к элементам которой применена экспериментальная методика, и символом σ_2^2, если применена традиционная методика. Приняв такую символику, нулевую гипотезу можно записать в виде: H_0: $\sigma_1^2 = \sigma_2^2$, а альтернативную в виде H_1: $\sigma_1^2 \neq \sigma_2^2$.

Для проверки нулевой гипотезы H_0 против альтернативной гипотезы H_1 используется критерий Фишера $F_{\nu_1 \cdot \nu_2}$, представляющий отношение двух выборочных дисперсий: S_1^2 и S_2^2, а именно $F_{\nu_1 \cdot \nu_2} = S_1^2 / S_2^2$, где ν_1 и ν_2 – степени свободы каждой из сопоставляемых выборок.

Когда верна гипотеза H_0: $\sigma_1^2 = \sigma_2^2$, выборочное распределение $F = S_1^2 / S_2^2$ критерия Фишера представляет собой F-распределение Фишера со степенями свободы $\nu_1 = n_1 - 1$ и $\nu_2 = n_2 - 1$, где n_1 и n_2 – численность выборок. При $\nu_1 = 4$ и $\nu_2 = 25$ – кривая распределения Фишера принимает вид, изображенный на рисунке 3.1.

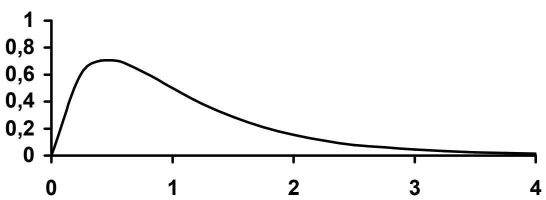


Рис. 3.1

Верхние критические точки этого распределения приведены в таблицах D_1 и D_2 Приложения, в которых степени свободы обозначены соответственно буквами k и n. Например, при n=4 и k=25 на уровне статистической значимости $\alpha = 0{,}05$ верхнее критическое значение критерия Фишера равно 2,8. Нижнее критическое значение при тех же условиях вычисляется по правилу: ${}_{1-\alpha}F_{n,k} = \dfrac{1}{{}_{\alpha}F_{k,n}}$, откуда следует, что для его вычисления нужно сначала, пользуясь той же таблицей D_1, найти верхнее критическое значение критерия Фишера при n=25 и k=4. Оно равно 5,7. Теперь остается найти число, обратное 5,7, которое будет в этом случае равно 1/5,7=0,18, оно и будет нижней критической точкой.

Пример. Сравниваются два разных изложения одной и той же темы в двух учебниках. Исследователя интересует, какое из них ведёт к меньшей дисперсии в получаемых учащимися оценках. Важный вопрос о средних значениях оставим в этом примере без рассмотрения.

С этой целью исследователь из состава учащихся пяти восьмых классов (125 человек), которые в данном случае составляют генеральную совокупность, случайным образом сформировал две группы по 20 человек каждая. Отобранные учащиеся самостоятельно изучали предложенную им тему: первая группа – по первому учебнику, вторая – по второму. Качество усвоения тестируется в десятибалльной системе. По результатам тестирования вычисляются выборочные средние и дисперсии.

Предположим, что в результате эксперимента установлено, что в первой группе дисперсия равна $S_1^2 = 4{,}34$, а во второй – $S_2^2 = 2{,}36$. Какой вывод должен быть сделан в этом случае? Критерий проверки даёт следующий результат: $F = S_1^2 \big/ S_2^2 = 4{,}34/2{,}36 = 1{,}839$.

Верхнее критическое значение двустороннего критерия Фишера при $\alpha = 0{,}05$ и $n_1 = n_2 = 20$, обозначаемое символом ${}_{0,025}F_{19,19}$, равно 2,1.

Так как количество отобранных элементов (40) составляет более чем одну двадцать пятую часть от генеральной совокупности (125), то вводим поправочный коэффициент $\sqrt{\dfrac{N-n}{N-1}}$, который в данном случае равен $\sqrt{\dfrac{125-40}{125-1}} = 0{,}825$. Умножив на него верхнее критическое значение, получим 1,733. Так как вычисленное значение $F = 1{,}839$ больше верхнего критического значения 1,733 и, следовательно, попадает в правую часть критической области, то нулевую гипотезу можно отклонить на уровне значимости $\alpha = 0{,}05$, и, следовательно, на этом уровне можно утверждать, что изложение, принятое в первом учебнике, ведёт к большей дисперсии, чем изложение, принятое во втором учебнике. Этот вывод можно распространить на всю генеральную совокупность.

3.4. Проверка гипотезы о равенстве долей по *независимым* выборкам. Рассматривается генеральная совокупность, в которой двумя методами может формироваться некоторое свойство. Обозначим символами π_1 и π_2 доли лиц (элементов) генеральной совокупности, у которых рассматриваемое качество может быть сформировано соответственно первым и вторым методами. Проверяется нулевая гипотеза $H_1 : \pi_1 = \pi_2$ против альтернативной гипотезы $H_1 : \pi_1 \neq \pi_2$.

Для проверки нулевой гипотезы из генеральной совокупности извлекаются две независимые выборки, первая из которых содержит n_1, а вторая – n_2 элементов. Если k_1 и k_2 – число элементов первой и второй выборки, обладающих характеристическим свойством, то $p_1 = k_1/n_1$, $p_2 = k_2/n_2$ – их доли в этих выборках. Гипотеза H_0 проверяется с помощью критерия: $Z = \dfrac{p_1 - p_2}{\sqrt{k(1-k)(1/n_1 + 1/n_2)}}$, где $k = (k_1 + k_2)/(n_1 + n_2)$ – доля элементов, обладающих исследуемым свойством в объединенной выборке. Если нулевая гипотеза верна и доли p_1 и p_2 не слишком малы и не слишком велики так, что каждая из величин $n_1 p_1$, $n_1(1 - p_1)$, $n_2 p_2$ и $n_2(1 - p_2)$ больше пяти, то z в приведенном выше статистическом критерии имеет стандартное нормальное распределение, критические точки которого приведены в таблице А Приложения.

Пример. Исследователь выдвинул гипотезу о том, что учащиеся восьмых классов усваивают новое понятие лучше, если его определению предшествует рассмотрение примеров, а не наоборот. С этой целью он, убедившись в примерном равенстве учебных групп по составу, предложил учащимся двух первых групп (n_1=50) изучить новое для них понятие по пособию, в котором определение предшествует примерам, а учащимся двух других групп (n_2=50) – по пособию, в котором определение дается после рассмотрения примеров. Затем учащимся обеих выборок предлагается один и тот же тест на проверку усвоения понятия.

Проверка показала, что в первой выборке усвоили понятие 35 человек, во второй – 25 человек, т.е. $p_1 = 0{,}70$ и $p_2 = 0{,}50$. Задавшись уровнем значимости $\alpha = 0{,}05$, вычислим Z по приведенному выше критерию. Для этого, прежде всего, найдём $k = 60/100 = 0{,}60$, после чего $Z = \dfrac{0{,}70 - 0{,}50}{\sqrt{0{,}60 \cdot 0{,}40 \cdot 0{,}04}} = 2{,}04$. Величина $z = 2{,}04$ превышает критическое значение ${}_{0{,}025}z = 1{,}96$, откуда на уровне значимости $\alpha = 0{,}05$ следует, что нулевую гипотезу необходимо отвергнуть. Руководствуясь упрощенной схемой педагогического эксперимента, можно сделать вывод о достоверности различий между экспериментальной и контрольной группами.

Если бы этот пример был не гипотетическим, мы могли бы сделать вывод, что лучше сначала дать определение, а затем приводить примеры.

3.5. Проверка гипотезы о равенстве долей по *зависимым* выборкам. Проверяемая гипотеза та же, что и в предыдущем пункте, а именно H_0: $\pi_1 = \pi_2$ против гипотезы H_1: $\pi_1 \neq \pi_2$. Однако на этот раз не требуется, чтобы сопоставляемые группы были независимыми.

Как и во всех методах, связанных с зависимыми выборками, необходимо составлять «пары» наблюдений. Поэтому их число в выборках должно быть одинаковым. Так как при измерении долями (да, нет) элементы этих пар можно обозначить цифрами 0 и 1, легко понять, что возможны пары только четырёх видов: (0,0) – пусть их будет c штук; (0,1) – пусть их будет a штук; (1,0) – d штук; (1,1) – b штук. Это можно выразить таблицей сопряжённости (табл. 3.5).

Таблица 3.5

Таблица сопряженности 2x2

		Первая выборка		Итого
		0	1	
Вторая выборка	1	a	b	k_2
	0	c	d	$n-k_2$
Итого		$n-k_1$	k_1	n

В ней n – общее число элементов в каждой из выборок, т.е. число пар; k_1 – число единиц в первой выборке; k_2 – число единиц во второй выборке.

Для проверки нулевой гипотезы H_0 против альтернативной гипотезы H_1 можно применить статистический критерий $Z = \frac{d-a}{\sqrt{d+a}}$, где a и d – число элементов выборки, которые изменили свою позицию с (1;0) на (0;1), и наоборот.

Когда справедлива гипотеза H_0: $\pi_1 = \pi_2$, и $a+d > 10$ считают, что распределение приведенного выше статистического критерия Z приближённо нормально с нулевым средним и единичным стандартным отклонением. Его критические точки находятся по той же таблице А Приложения.

Пример. Группе молодых людей ($n = 60$) предлагается до и после прослушивания лекции о вреде курения указать, одобряют они курение или нет. Результаты приведены в таблице (3.6).

Таблица 3.6.

Отношение респондентов к курению до и после прослушивания лекции о вреде курения

Выборка 2: после лекции	Выборка 1: до лекции		Итого
	одобряют	не одобряют	
Не одобряют	$a = 26$	$b = 8$	34
Одобряют	$c = 16$	$d = 10$	26
Итого	42	18	60

Так как $d + a = 36$ больше десяти, то применяем критерий $\hat{z} = \frac{10-26}{\sqrt{10+26}} = -2{,}67$. Значение $\tilde{z} = -2{,}67$ лежит значительно ниже критического $_{0,025}t = -1{,}96$. Следовательно, гипотезу о том, что в рассматриваемых совокупностях доли людей, одобряющих курение до и после лекции, равны, можно отклонить на уровне значимости 0,05. Легко посчитать, что до лекции курение одобряли 42/60=0,7 (70%), а после лекции – 34/60=0,57 (57%) респондентов. Следовательно, можно утверждать (с надежностью 95%), что после чтения подобной лекции доля респондентов, одобряющих курение, снижается. Естественно, что этот вывод может быть распространен только на совокупность людей, из которой отбиралась выборка и при условии повторения этой же лекции тем же лектором в идентичных условиях.

Глава 4. Непараметрические критерии

4.1. Критерий Колмогорова-Смирнова для *независимых* выборок. Одним из наиболее распространенных и строгих непараметрических критериев является критерий Колмогорова-Смирнова (К-С – критерий). Он предназначен для сравнения двух независимых выборок (рядов измерений или значений частот) и ответа на вопрос о статистической значимости различий между ними.

Принятое в этом критерии правило нахождения расстояния между двумя выборками поясним на примере распределения учащихся по заданным интервалам времени, приведенном в таблице 4.1. Числовые данные, приведенные в таблице, называют частотами.

Таблица 4.1

Распределение испытуемых по интервалам времени

Интервал	0,0 0,4	0,5 0,9	1,0 1,4	1,5 1,9	2,0 2,4	2,5 2,9	3,0 3,4	3,5 3,9	4,0 4,4	4,5 5,0	Итого
Экспер.	1	1	3	4	9	3	2	1	0	1	25
Контр.	0	0	2	4	4	6	2	2	3	2	25

В общем случае частоты первой строки обозначим символами n_{i1}, а частоты второй строки – символами n_{i2}, в которых первый индекс является номером столбца, а второй – номером строки. Так, например, частота $n_{3,2}$ равна 2.

Основываясь на данных таблицы 4.1, составим таблицу так называемых *накопленных* частот 4.2, для чего к частоте каждой ячейки таблицы 4.1 прибавим все частоты предшествующих ячеек той же строки таблицы 4.1. Такие распределения называют кумулятами – термин, происходящий от понятия аккумулировать (накапливать) что-либо.

Таблица 4.2.

Распределение накопленных частот

Интервал	0,0 0,4	0,5 0,9	1,0 1,4	1,5 1,9	2,0 2,4	2,5 2,9	3,0 3,4	3,5 3,9	4,0 4,4	4,5 5,0
Экспер. $N_{i,1}$	1	1	4	7	18	21	23	24	24	25
Контр. $N_{i,2}$	0	0	2	6	10	16	18	20	23	25

Накопленные частоты обозначим $N_{i,1}$ и $N_{i,2}$, например, в таблице 4.2 $N_{5,2}=10$.

Разделив все значения кумулят $N_{i,1}$ на n_1, а $N_{i,2}$ на n_2, получим значения *относительных частот кумулят*, представленные в таблице 4.3.

Таблица 4.3.

Интервальное распределение относительных частот кумулят $\frac{N_{i.1}}{n_1}$ и $\frac{N_{i.2}}{n_2}$

Интервалы	0,0-0,4	0,5-0,9	1,0-1,4	1,5-1,9	2,0-2,4	2,5-2,9	3,0-3,4	3,5-3,9	1,0-4,4	4,5-5,0
Эксперим.	0,04	0,08	0,20	0,40	0,72	0,84	0,92	0,96	0,96	1,00
Контрольн.	0,00	0,00	0,08	0,24	0,40	0,64	0,72	0,80	0,92	1,00
Модули разностей	0,04	0,08	0,12	0,16	0,32	0,20	0,20	0,16	0,04	0,00

Максимум абсолютных значений разностей относительных частот сопоставляемых кумулят и принимается в качестве «расстояния» между рядами. В рассмотренном примере это расстояние равно 0,32. Критерий Колмогорова-Смирнова (D) строится в соответствии с таким понятием расстояния, а именно: критерием К-С. для двух выборок, распределенных по одной и той же системе интервалов, называют максимум абсолютного значения разности относительных частот их кумулят, т.е.

$$D = \max\left\{\left|\frac{N_{i.1}}{n_1} - \frac{N_{i.2}}{n_2}\right|\right\}.$$

Критические значения, с которыми должно сопоставляться вычисленное значение критерия при средних и больших объемах выборок $(n_1 + n_2 > 35)$, могут быть приближенно вычислены по формуле ${}_\alpha D_{крит} = {}_\alpha K \cdot \sqrt{\frac{2}{n_1 + n_2}}$, где ${}_\alpha K$ есть постоянная, зависящая от статистической значимости α, а значит и от надежности $P = 1 - \alpha$. Ее значения приведены в таблице 4.4.

Таблица 4.4

Значения коэффициента $_{\alpha}K$ в зависимости от уровня статистической значимости α

α	0,20	0,15	0,10	0,05	0,01	0,001
$_{\alpha}K$	1,07	1,14	1,22	1,36	1,63	1,95

При $n_1 = n_2 = n$ критические значения вычисляются по формуле $_{\alpha}D_{крит} = {}_{\alpha}K \cdot \sqrt{\frac{1}{n}}$. Их значения для $\alpha = 0{,}10$ *или* $0{,}05$ представлены в таблице К. Приложения. В нашем случае, когда $n_1 = n_2 = 25$, $_{\alpha}D_{крит} = {}_{\alpha}K \cdot \sqrt{\frac{1}{25}} = {}_{\alpha}K \cdot 0{,}2$. Соответствующие этому случаю критические значения приведены в таблице 4.5.

Таблица 4.5

Критические значения критерия Колмогорова-Смирнова при $n_1 = n_2 = 25$

Статистическая значимость α	0,20	0,15	0,10	0,05	0,01
Критические значения	0,214	0,228	0,244	0,272	0,326

Если определенное на основании двух выборок значение $\tilde{D}_{выч}$ достигает критического значения $_{\alpha}D_{крит}$ или превосходит его, то это означает наличие значимого различия между рассматриваемыми распределениями и возможность распространения с соответствующей надежностью экспериментальных выводов на всю генеральную совокупность. В противном случае распространять выводы, полученные на основе проведенного эксперимента, на всю генеральную совокупность с принятой надежностью нельзя.

Сравнивая вычисленное значение критерия Колмогорова-Смирнова, которое равно 0,32, с критическими значениями, приведенными в таблице 4.5, легко видеть, что наибольшее критическое значение, которое оно превосходит, равно 0,272. Этому значению, как легко видеть, соответствует статистическая значимость $\alpha = 0{,}05$.

Отсюда следует, что отвергнуть нулевую гипотезу и, следовательно, признать возможным распространение выводов исследования на всю генеральную совокупность, пользуясь критерием Колмогорова-Смирнова, можно на уровне надежности 95%.

4.2. Применение критерия Колмогорова-Смирнова в качестве критерия согласия. *Критерии согласия чаще всего используют при* сравнении эмпирического и теоретического (или гипотетического) распределений. Построен он по тому же принципу, который использовался в предыдущем случае, с опорой на введенное в нем правило вычисления расстояния между двумя распределениями. Отличается оно лишь тем, что вместо двух эмпирических рядов в нем рассматривается только один такой ряд. В качестве же второго ряда рассматриваются количества членов той же эмпирической выборки, но распределенных по интервалам в соответствии со сделанным теоретическим предположением об истинном характере эмпирического распределения. В результате получается таблица, аналогичная таблице 4.1, с тем отличием, что в ней суммы частот в обеих строках одинакова и равна объему выборки, и, кроме того, в строке, посвященной теоретическим частотам, допускаются дробные значения. В результате критерий приобретает вид $D = \max\left\{\left|\frac{N_{i1}}{n} - \frac{N_{i2}}{n}\right|\right\}$ или, как его чаще записывают

$D = \frac{\max\left|\hat{N}_L - N_L\right|}{n}$, где $\hat{N}_L$ – значения кумулят теоретических частот, а N_L – значения кумулят эмпирических частот.

Рассмотрим, например, распределение 36 респондентов по их оценке признака: «очень важный», «важный», «маловажный», «совсем не важный». Соответствующие частоты приведены в таблице 4.6.

Таблица 4.6

Эмпирические частоты распределения респондентов в зависимости от выбранных ими ответов

	Очень важный	Важный	Мало - важный	Совсем не важный	Итого
$n_{i.1}$	8	12	10	6	36

Проверим гипотезу о том, что в генеральной совокупности значения этого признака распределены равномерно. Теоретическое распределение получим, если предположим, что респонденты с одинаковой вероятностью могли выбрать любую из четырёх групп. Тогда ожидаемая частота будет равна $36/4 = 9$, что и отмечено в нижней строке таблицы 4.7.

Таблица 4.7

Эмпирические и теоретические распределения респондентов в зависимости от выбранных ими ответов

	Очень важный	Важный	Мало - важный	Совсем не важный	Итого
Эмпирич.	8	12	10	6	36
Теоретич.	9	9	9	9	36

В соответствии с общим методом строим таблицу кумулят 4.8.

Таблица 4.8

Кумуляты эмпирического и теоретического распределений респондентов в зависимости от выбранных ими ответов

	Очень важный	Важный	Маловажный	Совсем не важный	Итого
Эмпирич.	8	20	30	36	36
Теоретич.	9	18	27	36	36
Модуль разности	1	2	3	0	0

В отличие от предыдущего случая, нет необходимости переходить к таблице распределения относительных частот кумулят. В соответствии с приведенной формулой, выражающей критерий согласия Колмогорова-Смирнова – достаточно найти наибольшее абсолютное значение разности кумулят теоретического и эмпирических распределений. В нашем примере оно равно 3.

Обращаясь к К-С – критерию согласия Колмогорова-Смирнова, получим, что D=3/36=0,083. Критические значения подсчитываются по тем же правилам, что и в случае работы с общим критерием Колмогорова-Смирнова (см. пункт 4.2). На уровне $\alpha = 0{,}05$, при $n = 36$ критическое значение равно 0,227. Так как вычисленное значение существенно ниже критического, то нет оснований на этом уровне значимости отклонять нулевую гипотезу о совпадении в генеральной совокупности этих двух распределений.

В особенности хорошо критерий Колмогорова-Смирнова обнаруживает отклонения от нормального закона распределения при малых объемах выборок. Нерегулярность распределения, как правило, лучше устанавливать с помощью так называемого χ^2- критерия, а отклонения формы распределения – с помощью К-С – критерия. Этот критерий предполагает непрерывное распределение, хотя он применяется и при дискретных распределениях.

Критические границы для малых выборок на уровне значимости $\alpha = 0{,}10$ и $\alpha = 0{,}05$ приведены в таблице К Приложения.

Наблюдаемое $\widehat{D}$-значение значимо на соответствующем уровне, если оно достигает табличного значения или превосходит его.

4.3. Хи-квадрат критерий. К числу наиболее простых непараметрических критериев согласия (и не только «согласия») относится так называемый хи-квадрат критерий, обозначаемый как χ^2-критерий.

Чаще всего χ^2 критерий используется для проверки нулевой гипотезы о близости (совпадении) эмпирического распределения некоторому теоретическому распределению. Правило «измерения расстояния» между такими двумя распределениями опишем сначала на примере.

Рассмотрим два интервальных распределения выборок по интервалам: $S_1, S_2, \ldots, S_k$. Эмпирическое распределение задано своими относительными частотами n_1/n, n_2/n …, n_k/n, где $n=n_1+n_2+\ldots+n_k$, а теоретическое распределение – вероятностями попадания теоретической случайной величины в заданные интервалы: $p_1, p_2, \ldots, p_k$ (см. таблицу 4.9).

Таблица 4.9

Сопоставление относительных частот эмпирического распределения и вероятностей теоретического распределения

Интервал	S_1	S_2	…	S_k	Сумма
Эмпирические частоты	n_1/n	n_2/n	…	n_k/n	1
Теоретические частоты	p_1	p_2	…	p_k	1

В качестве хи-квадрат критерия принимается мера отклонения эмпирической функции распределения от теоретической, измеряемая величиной $D = \sum_{i=1}^{k} \frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2$, откуда следует $D = \sum_{1}^{k} \frac{(n_i - np_i)^2}{np_i}$.

Наконец, обозначив np_i символом $\tilde{n}_i$, получим: $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i}$.

Расчеты показывают, что распределение значений хи-квадрат критерия является χ^2– распределением Пирсона с числом степеней свободы $\nu = k - 1 - r$, где k – число интервалов, а r – число параметров предполагаемого распределения, которые оценены по данным выборки (например, для нормального распределения оценивают два параметра: $\bar{x}$ и S^2), и имеет вид, изображенный на рисунке 4.1.

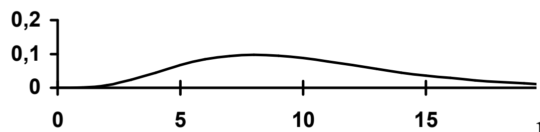


Рис. 4.1

Критические точки приведены в таблице С Приложения, например, при $\alpha = 0{,}05$ и $\nu = 10$ верхняя критическая точка ${}_{0,025}\chi^2_{10}$=20,48, а нижняя – ${}_{0,975}\chi^2_{10} = 3{,}25$.

Популярность χ^2- критерия определяется, прежде всего, тем, что его применение не требует предварительного знания закона распределения исследуемой случайной величины, которая к тому же может принимать как непрерывные, так и дискретные значения, причём измеренные хотя бы на номинальном уровне.

Как правило, эмпирические и теоретические частоты будут различаться. Наблюдаемые несовпадения могут иметь различную природу. Они могут быть следствием малого числа наблюдений, способа группировки или иных причин. Но расхождения могут возникнуть и вследствие неверной гипотезы о характере распределения. Хи-квадрат критерий отвечает на вопрос, случайно или нет расхождение частот. Как любой критерий, χ^2-критерий не доказывает справедливость гипотезы, а лишь с определённой вероятностью $p = 1 - \alpha$ устанавливает её согласие или несогласие с данными наблюдения.

Обратимся теперь к рассмотренному выше примеру равномерности распределения респондентов по их оценке признака: «очень важный», «важный», «маловажный», «совсем не важный» (данные, приведенные в верхней строке таблицы 4.7). Там невозможность отклонения нулевой гипотезы на уровне $\alpha = 0{,}05$ была обнаружена с помощью К-С - критерия. Теперь в тех же целях обратимся к хи-квадрат критерию.

Для удобства воспроизведем эту таблицу еще раз под номером 4.10.

Таблица 4.10

Эмпирические и теоретические частоты распределения респондентов в зависимости от выбранных ими ответов

	Очень важный	Важный	Мало - важный	Совсем не важный	Итого
Эмпирическая	8	12	10	6	36
Теоретический	9	9	9	9	36

Значение хи-квадрат критерия вычислим, пользуясь формулой

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - \tilde{n}_i)^2}{\tilde{n}_i} = \frac{(8-9)^2}{9} + \frac{(12-9)^2}{9} + \frac{(10-9)^2}{9} + \frac{(6-9)^2}{9} = 2{,}22$$

Нулевая гипотеза H_0 заключается в том, что мера отклонения эмпирического распределения от равномерного распределения равна нулю. Альтернативная гипотеза: отклонение больше нуля (естественно, что оно не может быть отрицательным), и, следовательно, критерий односторонний. При $\alpha = 0{,}05$ и $\nu = 4 - 1 = 3$ ${}_{0,05}\chi^2_3 = 7{,}81$. Таким образом, наблюдаемое значение χ^2 (оно равно 2,22) не попадает в критическую область, следовательно, данные наблюдений согласуются с нулевой гипотезой и не дают оснований её отвергать.

Заметим, что для корректного применения методов, основанных на использовании хи-квадрат критерия, исследователь должен обеспечить выполнение следующих условий: выборку необходимо получить из независимых наблюдений, данные могут быть измерены на любом уровне, но ни одна из ожидаемых частот не должна быть слишком мала (минимум 5).

Если же частоты оказываются меньше 5, то необходимо либо уменьшить степень дробности группировки признаков, объединив соседние интервалы, либо обратиться к другому критерию.

Хи-квадрат критерий часто применяют и для проверки гипотезы об отсутствии различий в мнениях респондентов, сгруппированных по тому или иному признаку. Имеется в виду пропорциональность числа респондентов сопоставляемых групп по выбранным ими оценкам. (Обычно это и есть теоретическая гипотеза).

Пример. Для изучения отношения учащихся к новой форме проведения уроков по физическому воспитанию *случайным образом* опросили 120 учащихся школы. Результаты опроса приведены в таблице 4.12.

Рассмотрим нулевую гипотезу H_0 об отсутствии в генеральной совокупности значимых различий в мнениях учащихся указанных трёх возрастных групп (когда численности респондентов, выбравших ту или иную оценку в этих группах, пропорциональны между собой). Уровень значимости α примем равным 0,05. Для нахождения теоретической (ожидаемой) частоты (т.е. частоты, которую имело бы распределение, если бы выполнялось условие пропорциональности) надо перемножить соответствующие маргинальные частоты (частоты, представленные в итоговом столбце и итоговой строке) и разделить полученное произведение на общее число респондентов (120).

Таблица 4.11

Отношение учащихся к новой форме проведения уроков по физическому воспитанию

	К л а с с ы			Всего
	V – VII	VIII, IX	X, XI	
Весьма положительно	(а) 5	(б) 8	(в) 18	31
Положительно	(г) 7	(д) 10	(е) 15	32
С сомнением	(ж) 11	(з) 6	(и) 6	23
Отрицательно	(к) 15	(л) 10	(м) 9	34
Всего	38	34	48	120

Например, ожидаемая частота для клетки (а) равна $\frac{31 \cdot 38}{120} = 9{,}8$; для клетки (д) – $\frac{32 \cdot 34}{120} = 9{,}1$; для клетки (м) – $\frac{34 \cdot 48}{120} = 13{,}6$.

Результаты вычислений значения χ^2- критерия представлены в таблице 4.12. $\widetilde{\chi}^2$=12,603. Число степеней свободы в подобных случаях вычисляется по формуле $\nu = (r-1)(c-1)$, где r – число строк, а c – число столбцов таблицы. В данном конкретном случае $\nu = (4-1)(3-1) = 6$. По таблице С Приложения находим $_{0,05}\chi^2_6$=12,59.

Таблица 4.12

Схема вычисления χ^2

Ячейки таблицы 58	Частота n_i	Ожидаемая частота $\widehat{n}_i$	$n_i - \widehat{n}_i$	$(n_i - \widehat{n}_i)^2$	$\frac{(n_i - \widehat{n}_i)^2}{\widehat{n}_i}$
а	5	9,8	- 4,8	23,04	2,351
б	8	8,8	- 0,8	0,64	0,073
в	18	12,4	5,6	31,36	2,529
г	7	10,1	- 3,1	9,61	0,951
д	10	9,1	0,9	0,81	0,089
е	15	12,8	2,2	4,84	0,378
ж	11	7,3	3,7	13,69	1,875
з	6	6,5	- 0,5	0,25	0,038
и	6	9,2	- 3,2	10,24	1,113
к	15	10,8	4,2	17,64	1,633
л	10	9,6	0,4	0,16	0,017
м	9	13,6	- 4,6	21,16	1,556
Итого	120	120	0	—	12,603

Так как $\widehat{\chi}^2 > {}_{0,05}\chi^2_6$, то на уровне значимости $\alpha = 0{,}05$ нужно отвергнуть гипотезу о том, что нет различий в отношении к новой форме занятий среди учащихся различных классов.

4.4. Сравнение эмпирического распределения со стандартным нормальным распределением. На практике большое значение имеет проверка эмпирического распределения на «нормальность». И в этом случае удобно прибегнуть к помощи критерия хи - квадрат. Новым в этом случае является необходимость расчета теоретических частот нормального распределения. Обычно и в этом случае прибегают к помощи специальной расчетной таблицы.

Обратимся к примеру. Пусть в таблице 4.13 приведено распределение 40 учащихся по количеству набранных при тестировании баллов.

Таблица 4.13

Эмпирическое распределение учащихся по числу набранных баллов

Количество набранных баллов	1	2	3	4	5	6	7	8	9	10	Итого
Количество учащихся	0	2	5	4	10	8	6	2	1	2	40

Нулевая гипотеза H_0 заключается в том, что эмпирическое распределение $F(x)$ не отличается от нормального F_n^*. В качестве статистического критерия вновь примем статистику $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - \widehat{n}_i)^2}{\widetilde{n}_i}$, где n_i – экспериментальные, а $\widehat{n}_i$ – теоретические частоты.

Для вычисления теоретической частоты, соответствующей i-му интервалу распределения, можно воспользоваться формулой

$$\widetilde{n}_i = \frac{nb}{s} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(x_i - \bar{x})^1}{2s^2}}$$, где:

$\widetilde{n}_i$ – теоретическая частота, соответствующая i-му интервалу распределения;

n – объём группы, число объектов исследования;

l – шаг, длина частного интервала, или расстояние от c_i до c_{i+1};

$\bar{x}$ – выборочное среднее;

s – выборочное стандартное отклонение;

x_i – середина i-го интервала.

Пользуясь данными таблицы 4.13, примем длины интервалов равными 1, а в качестве их центров – точки 1, 2, 3, 4 и т д. до 10, объем группы n=40, кроме того, вычисления показывают, что $\bar{x} = 5{,}5$, $s = 1{,}948$. Вычисления можно оформить в виде таблицы 4.14. Ее заполнение проведем в следующем порядке:

а) значения z_i, вычисленные по формуле $z_i = \frac{x_i - \bar{x}}{s} = \frac{x_i - 5.5}{1.948}$, занесем в столбец 4 таблицы;

б) пользуясь таблицей N Приложения и четностью функции $\varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-0,5z^2}$, занесем ее значения в пятый столбец таблицы 4.14;

в) для получения теоретической частоты все значения пятого столбца умножим на $R = \frac{nb}{s} = \frac{40 \cdot 1}{1.948} = 20,53$. Результаты занесем в шестой столбец;

г) малые крайние интервалы, так как у них $\tilde{n}$<5, объединим с соседними интервалами, что приведет к уменьшению числа интервалов до k=6. Окончательные теоретические частоты приведены в седьмом столбце таблицы.

Таблица 4.14

Расчет теоретических частот и критерия χ^2

1	2	3	4	5	6	7	8	9	10
x	n	$x-\bar{x}$	$\frac{x-\bar{x}}{s}=z$	$\varphi(z)$	$R\ \varphi(z)$	$\hat{n}$	$n-\tilde{n}$	$(n-\hat{n})^2$	$\frac{(n-\hat{n})^2}{\hat{n}}$
1	0	-4,5	2,31	0,028	0,574				
2	2	-3,5	1,80	0,079	1,638				
3	5	-2,5	1,28	0,176	3,644	5,856	1,144	1,309	0,223
4	4	-1,5	0,77	0,297	6,149	6,149	–2,149	4,618	0,751
5	10	-0,5	0,26	0,386	7,996	7,996	2,004	4,016	0,502
6	8	0,5	0,26	0,386	7,996	7,996	0,004	0	0
7	6	1,5	0,77	0,297	6,149	6,149	–0,149	0,022	0,149
8	2	2,5	1,28	0,176	3,644	5,856	–0,856	0,733	0,125
9	1	3,5	1,80	0,079	1,638				
10	2	4,5	2,31	0,028	0,574				
	40	0	12,84	1,929	40	40	0		1,75

Вторая половина таблицы, начиная с восьмого столбца, посвящена вычислению критерия хи-квадрат: ищется сумма отношений $\frac{(n-\hat{n})^2}{\hat{n}}$.

Как видим, она оказалась равной 1,75. Это и есть значение критерия.

Как отмечалось выше, число степеней свободы $\nu = k - 1 - r$. Так как k=6 и при вычислении $\widetilde{n}_i$ учитываются два параметра ($\bar{x}$ и S) и, следовательно, $r = 2$, то число степеней свободы равно $6 - 1 - 2 = 3$. Пользуясь таблицей С Приложения, получим, что на уровне значимости α=0,05 верхняя критическая точка χ^2- распределения будет равна: ${}_{0,05}\chi^2_3$=7,81.

Так как в соответствии с таблицей $\widehat{\chi}^2 = 1{,}74$ не попадает в критическую область, то нуль-гипотеза принимается на уровне статистической значимости $\alpha = 0{,}05$.

Глава 5. Ранговые и быстрые критерии

5.1. **Критерий Уилкоксона для разностей пар.** В пункте 3.2 была рассмотрена процедура проверки гипотезы $\mu_1 = \mu_2$ в предположении, что используемые выборки зависимы, а разности рассматриваемых парных значений распределены по нормальному закону (t-критерий Стьюдента для зависимых выборок). Если условие нормальности не выполняется, то обращаются к критерию Уилкоксона для разностей пар, который может быть применен также и к ранжированным данным. Он требует значительно меньшего объема вычислений по сравнению с t- критерием Стьюдента и почти также строго проверяет нормально распределенные разности; его эффективность для больших и малых выборок равна 95%.

С помощью критерия Уилкоксона проверяется нулевая гипотеза, заключающаяся в том, что *функции распределения* обеих связанных выборок равны $F_1(x) = F_2(x)$, и, следовательно, равны их математические ожидания или медианы.

Процедуру применения критерия Уилкоксона проиллюстрируем на данных, приведенных в таблице 5.1,

Таблица 5.1

Баллы, набранные учащимися в начале (А) и в конце (В) эксперимента (максимально возможное число баллов – 100)

Номера испытуемых	1	2	3	4	5	6	7	8	9	10	11	12
На начало (А)	28	36	41	58	71	37	12	36	53	49	43	64
На конец (В)	24	45	48	64	71	52	18	44	50	62	43	77

Для применения критерия Уилкоксона, прежде всего, отбрасывают пары с равными значениями в начале и в конце эксперимента. В таблице 5.1 таких пар две: (71: 71) и (43; 43). В дальнейшем работа ведется с такой «урезанной» таблицей, число пар в которой и принимается в расчет при определении числа степеней свободы. Число оставшихся пар обозначим n (в нашем примере n=10).

Пользуясь данными первых двух строк «урезанной» таблицы (см. таблицу 5.2), заполняют третью строку их разностями $d_i = x_{i1} - x_{i2}$. В четвертой строке записываются ранги абсолютных величин этих разностей. Наименьшее значение получает ранг 1, наибольшее – n (в нашем примере 10). Равным по величине значениям разностей приписывается *средний ранг*. В пятой строке каждому рангу приписывается знак (положительный или отрицательный), соответствующий знаку разности.

Таблица 5.2

Вычисление значения критерия Уилкоксона

Номера испытуемых	1	2	3	4	6	7	8	9	10	12
На начало (A)	28	36	41	58	37	12	36	53	49	64
На конец (B)	24	45	48	64	52	18	44	50	62	77
A – B = d_i	- 4	9	7	6	15	6	8	- 3	13	13
Ранги абсолютных значений d_i	2	7	5	3,5	10	3,5	6	1	8,5	8,5
Ранг с поправкой на знак d_i	- 2	7	5	3,5	10	3,5	6	- 1	8,5	8,5

Далее образуют суммы абсолютных значений положительных и отрицательных рангов $(\widehat{R}_p \text{ и } \widehat{R}_n)$ и проверяют их с помощью выражения $\widehat{R}_p + \widehat{R}_n = n(n+1)/2$, где n – число пар с разными значениями. В нашем случае $\widehat{R}_p = 52$ и $\widehat{R}_n = 3$, откуда $\widehat{R}_p + \widehat{R}_n = 55$, что соответствует значению $n = 10$.

В качестве вычисленного значения критерия Уилкоксона используется меньшая из этих двух сумм, её обозначают буквой $(\widehat{R})$. В нашем случае она равна 3 ($R = 3$).

Критические (нижние) значения критерия Уилкоксона $R(n;\alpha)$ приведены в таблице G Приложения. Нулевая гипотеза отклоняется на уровне α, если вычисленное значение $\widehat{R}$ меньше или равно критическому значению $R(n;\alpha)$. Так как в рассматриваемом случае $R(10;\,0{,}05) = 8$, а $R(10;\,0{,}01) = 3$, то нулевая гипотеза отклоняется на уровне $\alpha = 0{,}05$. Для

$n > 25$ справедлива аппроксимация $R(n;\alpha) = \frac{n(n+1)}{4} - z \cdot \sqrt{\frac{1}{24} n(n+1)(2n+1)}$,

где z – значение Z-критерия (см. табл. А Приложения) при соответствующем значении α.

Пример. Исследователь сравнивает два метода (два теста), по-разному проверяющих уровень усвоения некоторого материала. Каждый из 10 испытуемых выполняет оба теста. В таблице 5.3 приведены результаты первого (А) и второго (В) тестирования. О нормальности распределения неизвестно. Нужно проверить нулевую гипотезу об отсутствии значимого различия на уровне $\alpha = 0{,}05$ между результатами тестирования.

Таблица 5.3

Результаты тестирования десяти учащихся двумя тестами

№ испытуемых	1	2	3	4	5	6	7	8	9	10
А (баллов)	24	45	48	64	68	52	18	44	50	62
В (баллов)	28	36	41	58	68	37	12	36	53	49
А – В = d_i	- 4	9	7	6	0	15	6	8	- 3	13
Ранг модуля d_i	2	7	5	3,5	0	9	3,5	6	1	8
Ранг с поправкой на знак d_i	-2	7	5	3,5	0	9	3,5	6	-1	8

Сумма положительных рангов равна R_p= 42, сумма модулей отрицательных рангов R_n=3. Так как n=9, то проверка дает: 42+3=45=9(9+1)/2. Следовательно, вычисленное значение критерия Уилкоксона равно $\hat{R}_n = 3$.

При двустороннем критерии на уровне $\alpha = 0{,}05$ R(9; 0,05)=5.

Так как $\hat{R}$< R(9; 0,05), то нулевую гипотезу следует отклонить на уровне статистической значимости $\alpha = 0{,}05$, откуда следует, что тесты дают не одинаковый результат.

5.2. U-критерий Уилкоксона, Манна и Уитни. Этот критерий в определенном смысле является развитием рассмотренного выше критерия Уилкоксона для независимых выборок, который является непараметрическим аналогом t-критерия Стьюдента для сравнения двух средних значений *непрерывных* распределений. Эта непрерывность, строго говоря, никогда на практике не выполняется, так как все результаты измерений являются округленными числами.

U-критерий Уилкоксона, Манна и Уитни, который проверяет нулевую гипотезу о совпадении двух независимых выборок из одной и той же генеральной совокупности, равенстве их функций распределения $F_1(x)=F_2(x)$ включает также *равенство положений,* в частности, равенство значений медиан и равенство значений математических ожиданий $\mu_1=\mu_2$. Если сравниваемые выборки относятся к распределениям одинакового типа, то его эффективность равна примерно 95%. В случае же произвольного распределения эффективность U-критерия не может быть ниже 86%.

Вычисленное значение U-критерия Уилкоксона, Манна, Уитни обозначается символом $\widetilde{U}$. Для его вычисления ранжируют объединенную совокупность значений первой и второй выборок $(m+n)$ по величине, причем каждому рангу приписывают знак, показывающий, к какой из выборок он относится. Пусть сумма рангов первой выборки с числом элементов m равна R_1, а второй выборки, с числом элементов n, равна – R_2. Пользуясь этими значениями, вычисляют значения

$$U_1=mn+\frac{m(m+1)}{2}-R_1 \quad \text{и} \quad U_2=mn+\frac{n(n+1)}{2}-R_2.$$

Правильность расчетов проверяется формулой $U_1+U_2=m\cdot n$.

В качестве вычисляемого значения U-критерия принимается меньшее из значений U_1 и U_2. $\widetilde{U}=\min(U_1,U_2)$.

Критические значения (нижние), обозначаемые символом $U(m,n;\alpha)$, приведены в таблицах F_1 и F_2 Приложения.

Нулевая гипотеза отвергается тогда, когда вычисленное значение U-критерия оказывается меньше критического значения $U_{(m,\ n;\ a)}$, или равно ему.

Для достаточно больших выборок (m+n>60) справедлива превосходная аппроксимация $U_{(m,n;\,\alpha)} = \frac{nm}{2} - z \cdot \sqrt{\frac{nm(n+m+1)}{12}}$, в которой значение z для двухстороннего или одностороннего критерия может быть определено по таблице А Приложения.

Применение U-критерия Уилкоксона, Манна, Уитни проиллюстрируем на конкретном примере. Предположим, что в таблице 5.4 приведены результаты тестирования (в баллах) экспериментальной (m=10) и контрольной (n=12) групп.

Таблица 5.4

Баллы, набранные учащимися экспериментальной и контрольной групп (максимально возможное число баллов – 50)

Эксперимент. группа	24	26	28	34	36	36	38	46	48	50	-	-
Контрольная группа	16	18	20	21	22	24	26	30	31	36	39	44

Необходимо определить статистическую значимость различий между этими группами. Так как каждое из этих распределений далеко от нормального распределения, то воспользуемся U-критерием.

В качестве нулевой гипотезы H_0 принимается предположение, что между группами нет различий. Альтернативная гипотеза H_1 – между результатами экспериментальной и контрольной групп есть статистически значимые различия.

Для проверки нулевой гипотезы ранжируем учащихся экспериментальной и контрольной групп по набранным баллам как единый массив: чем меньше набрано баллов, тем меньше ранг (табл. 5.5). При совпадении двух или более значений набранных баллов каждому ученику, попавшему в эту группу, приписывается одно и то же значение, равное среднеарифметическому соответствующих рангов.

Таблица 5.5

Результаты тестирования в экспериментальной и контрольной группах (максимально возможное число баллов – 50)

Экспериментальная группа (1)				Контрольная группа (2)			
порядк. Номер учеников	результ. тестир. (баллы)	ранги	средн. значен. рангов	порядк. номер учеников	результ. тестир. (баллы)	ранги	средн. значен. рангов
1	24	6;7	6,5	1	16	1	1
2	26	8;9	8,5	2	18	2	2
3	28	10	10	3	20	3	3
4	34	13	13	4	21	4	4
5	36	14;15;16	15	5	22	5	5
6	36	14;15;16	15	6	24	6;7	6,5
7	38	17	17	7	26	8;9	8,5
8	46	20	20	8	30	11	11
9	48	21	21	9	31	12	12
10	50	22	22	10	36	14;15;16	15
				11	39	18	18
				12	44	19	19
Итого	366	—	148	Итого	327		108

Подсчитав для каждой из групп сумму рангов, получим для экспериментальной и контрольной группы $-R_1 = 148$ и $R_2 = 108$. Затем, пользуясь приведенными выше формулами, вычислим

$$U_1 = mn + \frac{m(m+1)}{2} - R_1 = 10 \cdot 12 + \frac{10 \cdot 11}{2} - 148 = 27 \text{ и}$$

$$U_2 = mn + \frac{n(n+1)}{2} - R_2 = 10 \cdot 12 + \frac{12 \cdot 13}{2} - 108 = 90.$$

Равенство $U_1 + U_2 = 93 + 27 = 10 \cdot 12 = n_1 \cdot n_2$ подтверждает правильность проведенных расчетов.

В соответствии со сказанным выше, вычисленным значением U-критерия будет меньшее из значений U_1 и U_2, т.е. $\widetilde{U} = 27$.

Для нахождения критических значений U-критерия обратимся к таблицам F_1, F_2 и F_3 Приложения. В нашем случае при $n_1 = 10$ и $n_2 = 12$ на уровне значимости $\alpha = 0{,}05$ двухстороннего критерия (таблица F_1 Приложения) $U(10{,}12; 0{,}05) = 29$. Так как $\widetilde{U} < U(10{,}12;\ 0{,}05)$, то нулевую гипотезу следует отклонить на уровне $\alpha = 0{,}05$. Следовательно, судя по качеству выполнения теста, различие между экспериментальной и контрольной группами статистически значимо на уровне $\alpha = 0{,}05$.

Если объемы выборок превышают приведенные в таблицах Приложения и в выборках отсутствуют элементы с одинаковыми значениями, можно воспользоваться Z-критерием:

$$z = \frac{\left| U - \frac{mn}{2} \right|}{\sqrt{\frac{mn\left(m+n+1\right)}{12}}}. \tag{12.7}$$

В нашем случае эта формула не применима, так как одинаковые значения баллов встречаются в разных выборках. В этом случае применяется скорректированная формула:

$$Z = \frac{\left| U - \frac{n \cdot m}{2} \right|}{\sqrt{\left[\frac{n \cdot m}{S \cdot (S-1)}\right] \cdot \left[\frac{S^3 - S}{12} - \sum_{i=1}^{r} \frac{t_i^3 - t_i}{12}\right]}}, \tag{12.8}$$

где $S = n + m$. Через t_i обозначается число значений, которые имеют одинаковый ранг. Число r означает, что имеется r групп значений с одинаковыми рангами.

В нашем примере:

$t_1 = 3$: три значения 36 с рангом 8;

$t_2 = 2$: два значения 24 с рангом 16,5;

$t_3 = 2$: два значения 26 с рангом 14,5;

$\sum_{i=1}^{3} \frac{t_i^3 - t_i}{12} = \frac{27-3}{12} + \frac{8-2}{12} + \frac{8-2}{12} = 3{,}0$, кроме того $S = 10 + 12 = 22$.

$$Z = \frac{\left| 27 - \frac{10 \cdot 12}{2} \right|}{\sqrt{\frac{10 \cdot 12}{22 \cdot 21} \cdot \left[\frac{22^3 - 22}{12} - 3\right]}} = \frac{33}{15{,}14} = 2{,}18 \cdot$$

Так как критическое значение Z-критерия при $\alpha = 0{,}05$ равно 1,96 (см. табл. А Приложения), и $2{,}18 > 1{,}96$, то мы вновь приходим к необходимости отвергнуть нулевую гипотезу.

5.3. Быстрые критерии. Как мы могли убедиться, рассмотренные выше непараметрические критерии, обеспечивающие достаточно высокий уровень эффективности, не отличаются простотой вычислений. Наряду с ними разработано большое число непараметрических критериев, в которых ценой снижения эффективности достигается сравнительная простота вычислений. Такие критерии принято называть *быстрыми критериями*. Их недостаток – *малая мощность*, поскольку только часть информации, содержащаяся в данных, используется для принятия статистического решения. Быстрые критерии называют иногда *консервативными*, т.к. с их помощью труднее, чем с помощью других критериев, отклонить нулевую гипотезу, конечно, если она в действительности не имеет места, и поэтому приходится брать выборки с несколько большим, чем обычно, числом элементов.

Чаще всего к помощи быстрых непараметрических критериев прибегают тогда, когда хотят выяснить, выгодно ли вообще проводить проверку значимости с помощью более мощных критериев.

При этом возможны три случая:

а) результат может быть отчетливо значимым, проверка по более сильному критерию не нужна, так как цель проверки может быть достигнута и с помощью более слабого критерия;

б) результат может быть абсолютно незначимым, т. е. никакую значимость определить не удается; в этом случае проверка с помощью более сильного критерия также ни к чему;

в) результат может быть слабо значимым, но иметь тенденцию к значимости; в этом, и только в этом, случае последующая проверка с помощью оптимальных критериев возможна, хотя и требует большой осторожности.

5.4. Упрощенный критерий Тьюки. Две группы измерений отличаются тем больше, чем меньше пересекаются их значения. На этой основе Тьюки разработал свой упрощенный метод сравнения двух независимых выборок.

Для этого, прежде всего, записывают элементы каждой из сопоставляемых выборок в порядке возрастания.

Критерий Тьюки применим в случае, если одна из сопоставляемых выборок – обозначим ее буквой А – содержит числовые значения, превосходящие каждое из значений второй группы, а вторая – обозначим ее буквой В – числовые значения, меньшие каждого из значений первой группы. Иначе говоря, числовые значения группы А сдвинуты по отношению к числовым значениям группы В вправо. В этом случае подсчитывают:

1) количество k_1 тех значений группы А, которые превосходят все значения группы В. В случае, если наибольший элемент группы В совпадает с некоторым элементом группы А, то к числу k_1 следует прибавить 0,5;

2) количество k_2 тех значений группы В, которые меньше всех значений группы А. В случае, если наименьший элемент группы А совпадает с некоторым элементом группы В, то к числу k_2 следует прибавить 0,5.

Оба значения (каждое должно быть больше нуля) складываются, и, таким образом, получается вычисленное значение критерия Тьюки, который обозначают буквой Т. Следовательно, $T=k_1+k_2$.

Если объемы выборок примерно одинаковы, то критические значения критерия Тьюки равны, соответственно, 7, 10 и 13:

7 – для двустороннего критерия на уровне $\alpha = 0{,}05$;

10 – для двустороннего критерия на уровне $\alpha = 0{,}01$ и

13 – для двустороннего критерия на уровне $\alpha = 0{,}001$.

Если мы обозначим объемы выборок через n_1 и n_2 при $n_1 \le n_2$, то критерий справедлив при не слишком сильно отличающихся объемах выборок, а именно при $n_1 \le n_2 \le 3 + 4n_1/3$.

Замечание. Во всех других случаях рассчитанное значение статистики T перед сравнением его с числами 7, 10 или 13 должно быть уменьшено на корректирующее число, а именно на 1, если $(3+4n_1/3) < n_2 < 2n_1$, или на целую часть от числа $\frac{n_2 - n_1 + 1}{n_1}$, если $2n_1 \le n_2$.

Например, для $n_1 = 7$ и $n_2 = 13$ условие $n_1 \le n_2 \le 3 + 4n_1/3$ не выполняется, так как $3 + \frac{4 \cdot 7}{3} = \frac{37}{3} < 13$. Корректирующее значение, в соответствии с вышеизложенным, равно 1.

При $n_1 = 4$ и $n_2 = 14$ $\frac{14 - 4 + 1}{4} = \frac{11}{4} = 2{,}75$, т.е. корректирующее значение равно 2.

Наконец, если одна выборка превышает другую больше чем на 9 элементов, то на уровне $\alpha = 0{,}001$, нужно применять критическое значение 14 вместо 13.

В качестве примера рассмотрим два ряда измерений, упорядоченных в порядке возрастания (табл. 5.6).

Таблица 5.6

Сопоставляются два ряда измерений, упорядоченных в порядке возрастания

Ряд А	14,6	14,7	14,8	14,9	15	15,1	15,3	16,1	16,7	17,3
Ряд В	13,8	13,9	14,2	14,3	14,4	14,6	14,7	15	–	–

В первом ряду (ряд А) имеются пять значений, превышающих максимальное значение второго ряда, а именно: 15.1;15.3; 16.1; 16,7; 17,3, и одно совпадающее 15,0, Откуда k_1=5,5.

Во втором ряду (ряд В) имеется пять значений, меньших минимального значения первого ряда, а именно: 13.8; 13,9; 14,2; 14,3; 14,4 – и одно совпадающее – 14.6. Следовательно, k_2=5,5.

Критерий Тьюки T=5,5+5,5=11. Корректирующее значение равно нулю, так как $3 + 4n_1/3 = 3 + \frac{4 \cdot 8}{3} 13\frac{2}{3}$ и, следовательно, выполняется условие

$(n_1 \le n_2 \le 3 + 4n_1/3)$. Поскольку T=11>10, нулевая гипотеза (равенство функций распределения, соответствующих обеим выборкам) должна быть отклонена на уровне $\alpha = 0{,}01$.

5.5. Максимум-критерий для зависимых выборок. Максимум-критерий применяется для сравнения двух парных рядов измерения. Для его применения следует, расположив их разности в порядке убывания абсолютных значений, подсчитать количество абсолютно наибольших разностей, имеющих одинаковый знак. Число k, выражающее это количество, и является значением рассматриваемого критерия.

При $n \ge 6$ критическими значениями двухстороннего максимум-критерия являются числа 5, 6, 8 и 11.

При двустороннем критерии	$\alpha = 0{,}10$	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}001$
При одностороннем критерии	$\alpha = 0{,}05$	$\alpha = 0{,}025$	$\alpha = 0{,}005$	$\alpha = 0{,}0005$
Значения k критерия	5	6	8	11

Если имеются разности, равные по абсолютному значению, но с разными знаками, то для увеличения статистической надежности их располагают таким образом, чтобы уменьшить размеры случайных последовательных (серий) разностей с одинаковыми знаками. Максимум-критерий служит для независимой проверки t-критерия однако не заменяя его.

Обратимся к примеру, рассмотренному в пункте первом этой главы, посвященному критерию Уилкоксона. В таблице 5.1 этого пункта приведены значения двух зависимых выборок. Воспроизведем эту таблицу под очередным номером 5.7.

Таблица 5.7

Баллы, набранные учащимися в начале (А) и в конце (В) эксперимента (максимально возможное число баллов – 100)

Номера испытуемых	1	2	3	4	5	6	7	8	9	10	11	12
На начало (А)	28	36	41	58	71	37	12	36	53	49	43	64
На конец (В)	24	45	48	64	71	52	18	44	50	62	43	77
$B-A = d_i$	-4	9	7	6	0	15	6	8	-3	13	0	13

Расположив неравные нулю разности в порядке убывания их абсолютных значений, получим последовательность +15;+13;+13,+9,+8, +7; +6; +6; -4; -3. Так как она имеет 8 одинаковых по знаку значений, то при двустороннем критерии, как и в пункте 5.1, заключаем, что нулевая гипотеза о совпадении соответствующих функций распределения должна быть отклонена на уровне статистической значимости $\alpha = 0{,}01$.

5.6. Критерий знаков Диксона и Муда. Название критерия происходит оттого, что используются только знаки разностей наблюдаемых значений. В определенном смысле он является упрощением критерия Уилкоксона, а, следовательно, и t-критерия Стьюдента для парных сравнений. При использовании этого критерия, отбросив нулевые разности, подсчитывают число положительных и число отрицательных разностей. Обозначив символом n_1 меньшее из этих чисел, а n_2 – большее, образуем упорядоченную пару (n_1, n_2), которая и представляет вычисленное значение критерия знаков. Критическими значениями также являются упорядоченные пары чисел, первое из которых называют левой границей критерия (*ЛГ*) а второе – правой границей критерия (*ПГ*). Значения левых граничных точек приведены в таблицах L и L_1 Приложения, в которых n означает число неравных нулю разностей. Соответствующие правые граничные значения при n<100 вычисляются по правилу ПГ=n–ЛГ, а при n>100 – по правилу ПГ=n– ЛГ+1.

Нулевая гипотеза критерия знаков предполагает, что разности парных наблюдений в среднем не отличаются от нуля; ожидается, что около половины разностей будут меньше нуля, а другая половина – больше.

Критерий знаков проверяет также гипотезу о том, что медиана распределения разностей равна нулю. Нулевая гипотеза отклоняется, когда имеется слишком мало или слишком много разностей одного знака, т.е. если не достигнута левая граница, т.е. n_1<ЛГ, или превышена правая граница, т.е. n_2>ПГ. Если не выполнены оба условия, то нет оснований отклонять нулевую гипотезу.

Пример. Обратимся вновь к таблице 5.7. В ней мы получили две нулевые разности, 8 положительных и 2 отрицательных разности. Следовательно, n=10, n_1=2, n_2=8, а сам критерий представляет упорядоченную пару (2; 8). В соответствии с таблицей L, при двустороннем критерии и n=10 на уровне $\alpha = 0{,}05$ получим в качестве граничных точки ЛГ=2 и ПГ=8. Так как фактические значения совпадают с этими границами, т.е ни одна из них не выходит за пределы отрезка, образованного критическими точками, то нет основания на уровне $\alpha = 0{,}05$ отклонять нулевую гипотезу.

Полученный результат свидетельствует о сравнительной слабости рассматриваемого критерия. В отличие от распространенных случаев, когда эффективность критериев с увеличением объема выборок увеличивается, эффективность критерия знаков снижается с 95% при $n = 6$ до 64% при неограниченном увеличении значений n.

Не слишком маленькие выборки разностей (n>30) проверяют также просто по нормальному распределению с помощью формулы:

$$\hat{z} = \frac{|2x - n| - 1}{\sqrt{n}},$$

где x – наблюдаемая частота более редких знаков, а n – число пар, уменьшенное на число нулевых разностей.

Пусть, например, при сравнении 36 пар значений было получено 4 нулевых значения, 8 отрицательных и 24 положительных. Обращаясь к таблице L на уровне $\alpha = 0{,}01$, при $n = 32$ имеем границы 9 и 23. Полученные же частоты 8 и 24 не попадают в этот интервал, т.е. попадают в критическую область, лежащую вне этого интервала. Следовательно, нулевую гипотезу о равенстве нулю разностей следует отклонить на уровне статистической значимости $\alpha = 0{,}01$. Вместо таблицы можно было обратиться к приведенной выше формуле. Учитывая, что в нашем случае $x = 8,\ n = 32$, получим, что $\hat{z} = \frac{|2 \cdot 8 - 32| - 1}{\sqrt{32}} = \frac{15}{5.657} = 2{,}65$. Так как на уровне $\alpha = 0{,}01$ при двухстороннем z-критерии критическое значение равно 2,576, что меньше вычисленного значения, то нулевую гипотезу следует отклонить.

Приложение

Статистические таблицы

В таблицах числа, записанные в строке m, означают вероятности того, что табулируемая случайная величина примет значение, лежащее правее соответствующей критической точки.

Если ненаправленная (двухсторонняя) гипотеза проверяется на уровне статистической значимости α, то нижнюю критическую точку следует искать в столбце, соответствующем значению $m = 1 - \alpha / 2$, а верхнюю – в столбце, помеченном $m = \alpha / 2$. Например, если $\alpha = 0{,}05$, то нижнюю критическую точку следует искать в столбце $m = 0{,}975$, а верхнюю – в столбце $m = 0{,}025$.

В таблицах, соответствующих симметричным критериям, например, стандартному нормальному распределению (z-критерий) или распределению Стьюдента (t-критерий), приводятся только верхние критические точки.

Таблица А

Значения верхних критических точек **стандартного нормального распределения** для различных уровней значимости: $\alpha(1)$ при одностороннем критерии и $\alpha(2)$ при двустороннем

m	0,20	0,10	0,05	0,025	0,01	0,005
$\alpha(1)$	0,20	0,10	0,05	0,025	0,01	0,005
$\alpha(2)$	0,40	0,20	0,10	0,05	0,02	0,01
$Z_{\text{Кр}}$	0,842	1,282	1,645	1,960	2,326	2,576

Таблица В

Значения верхних критических точек **t-распределения Стьюдента** для различных уровней значимости: $\alpha(1)$ при одностороннем критерии и $\alpha(2)$ при двустороннем. В первом столбце – число степеней свободы

m	0,10	0,05	0,025	0,01	0,005
$\alpha(1)$	0,10	0,05	0,025	0,01	0,005
$\alpha(2)$	0,20	0,10	0,05	0,02	0,01
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1.699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2.021	2,423	2,704
50	1,299	1,676	2,009	2,403	2,678
60	1,296	1,671	2,000	2,390	2,660
80	1,292	1,664	1,990	2,374	2,639
100	1,290	1,660	1,984	2,364	2,626
120	1,289	1,658	1,980	2,358	2,617
200	1,286	1,653	1,972	2,345	2,601
∞	1,282	1,645	1,960	2,326	2,576

Таблица С

Значения критических точек **хи-квадрат распределения** для различных уровней значимости, α. В первом столбце – число степеней свободы

m	0,99	0,975	0,95	0,05	0,025	0,01
1	0,0002	0,0010	0,0039	3,84	5,02	6,63
2	0,0201	0,0506	0,1026	5,99	7,38	9,21
3	0,115	0,216	0,352	7,81	9,35	11,34
4	0,297	0,481	0,711	9,49	11,14	13,28
5	0,554	0,831	1,15	11,07	12,83	15,09
6	0,872	1,24	1,64	12,59	14,45	16,81
7	1,24	1,69	2,17	14,07	16,01	18,48
8	1,65	2,18	2,73	15,51	17,53	20,09
9	2,09	2,70	3,33	16,92	19,02	21,67
10	2,56	3,25	3,94	18,31	20,48	23,21
11	3,05	3,82	4,57	19,68	21,92	24,73
12	3,57	4,40	5,23	21,023	23,34	26,22
13	4,11	5,01	5,89	22,36	24,74	27,69
14	4,66	5,63	6,57	23,68	26,12	29,14
15	5,23	6,26	7,26	25,00	27,49	30,58
16	5,81	6,91	7,96	26,30	28,85	32,00
17	6,41	7,56	8,67	27,59	30,19	33,41
18	7,01	8,23	9,39	28,87	31,53	34,81
19	7,63	8,91	10,12	30,14	32,85	36,19
20	8,26	9,59	10,85	31,41	34,17	37,57
22	9,54	10,98	14,04	33,92	36,78	40.29
24	10,86	12,40	15,66	36,42	39,36	42,98
26	12,20	13,84	17,29	38,88	41,92	45,64
28	13,56	15,31	18,94	41,34	44,46	48,28
30	14,95	16,79	20,60	43,77	46,98	50,89
35	18,51	20,57	24,8	49,80	53,20	57,34
40	22,16	24,43	29,05	55,76	59,34	63,69
50	29,71	32,36	37,69	67,50	71,42	76,15
60	37,48	40,48	46,46	79,08	83,30	88,38
80	53,54	57,15	64,28	101,88	106,63	112,33
100	70,06	74,22	82,36	124,34	129,56	135,81
120	86,92	91,57	100,62	146,57	152,21	158,95
150	112,7	118,0	128,3	179,61	185,8	193,2
200	156,4	162,7	174,8	234,0	241,1	249,4

Таблица D_1

Значения верхних критических точек **$F_{n,k}$ распределения Фишера** на уровне значимости: $\alpha = 0{,}05$. Значения нижней граничной точки вычисляются по соответствующему значению верхней по правилу ${}_{1-\alpha}F_{n,k} = \dfrac{1}{{}_{\alpha}F_{k,n}}$

k \ n	1	2	3	4	5	6	8	10	12	14	20	30	50	∞
3	10,1	9,6	9,3	9,1	9,0	8,9	8,8	8,8	8,7	8,7	8,7	8,6	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	6,0	6,0	5,9	5,9	5,8	5,7	5,7	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,8	4,7	4,7	4,6	4,6	4,5	4,4	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,1	4,1	4,0	3,9	3,8	3,8	3,7	3,7
7	5,6	4,7	4,4	4,1	4,0	3,9	3,7	3,6	3,6	3,5	3,4	3,4	3,3	3,2
8	5,3	4,6	4,1	3,8	3,7	3,6	3,4	3,3	3,3	3,2	3,2	3,1	3,0	2,9
9	5,1	4,3	3,6	3,6	3,5	3,4	3,2	3,1	3,1	3,0	2,9	2,9	2,8	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	3,1	2,9	2,9	2,8	2,7	2,7	2,6	2,5
11	4,8	4,0	3,6	3,4	3,2	3,1	3,0	2,9	2,8	2,7	2,7	2,6	2,5	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,9	2,8	2,7	2,6	2,5	2,5	2,4	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,8	2,7	2,6	2,6	2,5	2,4	2,3	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,7	2,6	2,5	2,5	2,4	2,3	2,2	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,6	2,5	2,5	2,4	2,3	2,3	2,2	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,6	2,5	2,4	2,4	2,3	2,2	2,1	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,6	2,5	2,4	2,3	2,2	2,2	2,1	2,0
18	4,4	3,5	3,2	2,9	2,8	2,7	2,5	2,4	2,3	2,3	2,2	2,1	2,0	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,5	2,4	2,3	2,3	2,1	2,1	2,0	1,9
20	4,3	3,5	3,1	2,9	2,7	2,6	2,4	2,3	2,3	2,2	2,1	2,0	2,0	1,8
30	4,2	3,3	2,9	2,7	2,5	2,4	2,3	2,2	2,1	2,0	1,9	1,8	1,8	1,6
50	4,0	3,2	2,8	2,6	2,4	2,3	2,1	2,0	1,9	1,9	1,8	1,7	1,6	1,4
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,9	1,8	1,7	1,7	1,6	1,5	1,3	1,0

Таблица D_2

Значения верхних критических точек $\mathbf{F_{n,k}}$ – **распределения Фишера** на уровне значимости: $\alpha = 0{,}01$

k \ n	1	2	3	4	5	6	8	10	12	14	20	30	50	∞
3	34,1	30,8	29,5	28,7	28,2	27,9	27,5	27,2	27,1	26,9	26,7	26,5	26,4	26,1
4	21,2	18,8	16,7	16,0	15,5	15,2	14,8	14,7	14,4	14,2	14,0	13,8	13,7	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,3	10,1	9,9	9,8	9,6	9,4	9,2	9,0
6	13,4	10,9	9,8	9,2	8,8	8,5	8,1	7,9	7,7	7,6	7,4	7,2	7,1	6,9
7	12,3	9,6	8,5	7,9	7,5	7,2	6,8	6,6	6,4	6,3	6,1	5,9	5,9	5,7
8	11,3	8,7	7,6	7,0	6,6	6,4	6,0	5,8	5,7	5,6	5,4	5,2	5,1	4,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,5	5,3	5,1	5,0	4,8	4,6	4,5	4,3
10	10,0	7,9	6,6	6,0	5,6	5,4	5,1	4,9	4,7	4,6	4,4	4,3	4,1	3,9
11	9,7	7,2	6,2	5,7	5,3	5,1	4,7	4,5	4,4	4,3	4,1	3,9	3,8	3,6
12	9,3	6,9	6,0	5,4	5,1	4,8	4,5	4,3	4,2	4,1	3,9	3,7	3,6	3,4
13	9,1	6,7	5,7	5,2	4,9	4,6	4,3	4,1	4,0	3,9	3,7	3,5	3,4	3,2
14	8,9	6,5	5,6	5,0	4,7	4,5	4,1	3,9	3,8	3,7	3,5	3.3	3,2	3,0
15	8,7	6,4	5,4	4,9	4,6	4,3	4,0	3,8	3,6	3,5	3,3	3,2	3,1	2,9
16	8,5	6,2	5,3	4,8	4,4	4,2	3,9	3,7	3,5	3,5	3,3	3,1	3,0	2,8
17	8,4	6,1	5,2	4,7	4,3	4,1	3,8	3,6	3,5	3,4	3,2	3,0	2,9	2,7
18	8,3	6,0	5,1	4,6	4,2	4,0	3,7	3,5	3,4	3,3	3,1	2,9	2,8	2,6
19	8,2	5,9	5,0	4,5	4,2	3,9	3,6	3,4	3,3	3,2	3,0	2,9	2,7	2,5
20	8,1	5,8	4,9	4,4	4,1	3,9	3,6	3,4	3,2	3,1	2,9	2,8	2,6	2,4
30	7,6	5,4	4,5	4,0	3,7	3,5	3,2	3,0	2,8	2,7	2,5	2,4	2,2	2,0
50	7,2	5,1	4,2	3,7	3,4	3,2	2,9	2,7	2,6	2,5	2,3	2,1	1,9	1,7
∞	6,6	4,6	3,8	3,3	3,0	2,8	2,5	2,3	2,2	2,1	1,9	1,7	1,5	1,0

Таблица F_1.

Критические значения одностороннего критерия **Манна – Уитни** U_{Kp} – при $\alpha = 0{,}025$; двустороннего критерия – при $\alpha = 0{,}05$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10
1	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-			
3	-	-	-							
4	-		-	0						
5	-	-	0	1	2					
6	-	-	1	2	3	5				
7	-	-	1	3	5	6	8			
8	-	0	2	4	6	8	10	13		
9	-	0	2	4	7	10	12	15	17	
10	-	0	3	5	8	11	14	17	20	23
11	-	0	3	6	9	13	16	19	23	26
12	-	1	4	7	11	14	18	22	26	29
13	-	1	4	8	12	16	20	24	28	33
14	-	1	5	9	13	17	22	26	31	36
15	-	1	5	10	14	19	24	29	34	39
16	-	1	6	11	15	21	26	31	37	42
17	-	2	6	11	17	22	28	34	39	45
18	-	2	7	12	18	24	30	36	42	48
19	-	2	7	13	19	25	32	38	45	52
20	-	2	8	14	20	27	34	41	48	55
21		3	8	15	22	29	36	43	50	58
22		3	9	16	23	30	38	45	53	61
23		3	9	17	24	32	40	48	56	64
24		3	10	17	25	33	42	50	59	67
25		3	10	18	27	35	44	53	62	71
26		4	11	19	28	37	46	55	64	74
27		4	11	20	29	38	48	57	67	77
28		4	12	21	30	40	50	60	70	80
29		4	13	22	32	42	52	62	73	83
30		5	13	23	33	43	54	65	76	87
31		5	14	24	34	45	56	67	78	90
32		5	14	24	35	46	58	69	81	93
33		5	15	25	37	48	60	72	84	94
34		5	15	26	38	50	62	74	87	99
35		6	16	28	39	51	54	77	89	103
36		6	16	28	40	53	66	79	92	106
37		6	17	29	41	55	68	81	95	109
38		6	17	30	43	56	70	84	98	112
39	0	7	18	31	44	58	72	86	101	115
40	0	7	18	31	45	59	74	89	103	119

Таблица F_1 (продолжение)

Критические значения одностороннего критерия **Манна – Уитни** U_{Kp} при $\alpha = 0{,}025$; двустороннего критерия при $\alpha = 0{,}05$

n_2 \ n_1	11	12	13	14	15	16	17	18	19	20
11	30	-	-	-	-	-	-	-	-	-
12	33	37	-	-	-	-	-			
13	37	41	45							
14	40	45	50	55						
15	44	49	54	59	64					
16	47	53	59	64	70	75				
17	51	57	63	69	75	81	87			
18	55	61	67	74	80	86	93	99		
19	58	65	72	78	85	92	99	106	113	
20	62	69	76	83	90	98	105	112	119	127
21	65	73	80	88	96	103	111	119	126	134
22	69	77	85	93	101	109	117	125	133	141
23	73	81	89	98	106	115	123	132	140	149
24	76	85	94	102	111	120	129	138	147	156
25	80	89	98	107	117	126	135	145	154	163
26	83	93	102	112	122	132	141	151	161	171
27	87	97	107	117	127	137	147	158	168	178
28	90	101	111	122	132	143	154	164	175	186
29	94	105	116	127	138	149	160	171	182	193
30	98	109	120	131	143	154	166	177	189	200
31	101	113	125	136	148	160	172	184	196	208
32	105	117	129	141	153	166	178	190	203	215
33	108	121	133	146	159	171	184	197	210	222
34	112	125	138	151	164	177	190	203	217	230
35	116	129	149	156	169	183	196	210	224	237
36	119	133	147	161	174	188	202	216	231	245
37	123	137	151	165	180	194	209	223	238	252
38	127	141	156	170	185	200	215	230	245	259
39	130	145	160	175	190	206	221	236	252	267
40	134	149	165	180	196	211	227	243	258	274

Таблица F_2

Критические значения одностороннего **критерия Манна – Уитни** U_{Kp} при $\alpha = 0{,}01$; двустороннего критерия при $\alpha = 0{,}02$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10
1	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	1	1
4	-		-	-	-	1	1	2	3	3
5	-	-	-	-	1	2	3	4	5	6
6	-	-	-	1	2	3	4	6	7	8
7	-	-	-	1	3	4	6	7	9	11
8	-	-	-	2	4	6	7	9	11	13
9	-	-	1	3	5	7	9	11	14	16
10	-	-	1	3	6	8	11	13	16	19
11	-	-	1	4	7	9	12	15	18	22
12	-	-	2	5	8	11	14	17	21	24
13	-	-	2	5	9	12	16	20	23	27
14	-	-	2	6	10	13	17	22	26	30
15	-	-	3	7	11	15	19	24	28	33
16	-	-	3	7	12	16	21	26	31	36
17	-	-	4	8	13	18	23	28	33	38
18	-	-	4	9	14	19	24	30	36	41
19	-	1	4	9	15	20	26	32	38	44
20	-	1	5	10	16	22	28	34	40	47

Таблица F_3

Критические значения одностороннего критерия **Манна – Уитни** U_{Kp} при $\alpha = 0,05$; двустороннего критерия при $\alpha = 0,10$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10
1	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-			
3	-	-	-	-						
4	-		-	1						
5	-	-	1	2	4					
6	-	-	2	3	5	7				
7	-	-	2	4	6	8	11			
8	-	1	3	5	8	10	13	15		
9	-	1	4	6	9	12	15	18	21	
10	-	1	4	7	11	14	17	20	24	27
11	-	1	5	8	12	16	19	23	27	31
12	-	2	5	9	13	17	21	26	30	34
13	-	2	6	10	15	19	24	28	33	37
14	-	3	7	11	16	21	26	31	36	41
15	-	3	7	12	18	23	28	33	39	44
16	-	3	8	14	19	25	30	36	42	48
17	-	3	9	15	20	26	33	39	45	51
18	-	4	9	16	22	28	35	41	48	55
19	0	4	10	17	23	30	37	44	51	58
20	0	4	11	18	25	32	39	47	54	62
21	0	5	11	19	26	34	41	49	57	65
22	0	5	12	20	28	36	44	52	50	58
23	0	5	13	21	29	37	46	54	63	72
24	0	6	13	22	30	39	48	57	66	75
25	0	6	14	23	32	41	50	60	69	79

Таблица G

Критические значения критерия **Уилкоксона** для разностей пар

n	Двусторонний критерий			Односторонний критерий Кр		n	Двусторонний критерий			Односторонний критерий	
	5%	1%	0,1%	5%	1%		5%	1%	0,1%	5%	1%
6	0			2		36	208	171	130	227	185
7	2			3	0	37	221	182	140	241	198
8	3	0		5	1	38	235	194	150	256	211
9	5	1		8	3	39	249	207	161	271	224
10	8	3		10	5	40	264	220	172	286	238
11	10	5	0	13	7	41	279	233	183	302	252
12	13	7	1	17	9	42	294	247	195	319	266
13	17	9	2	21	12	43	310	261	207	336	281
14	21	12	4	25	15	44	327	276	220	353	296
15	25	15	6	30	19	45	343	291	233	371	312
16	29	19	8	35	23	46	361	307	246	389	328
17	34	23	11	41	27	47	378	322	260	407	345
18	40	27	14	47	32	48	396	339	274	426	362
19	46	32	18	53	37	49	415	355	289	446	379
20	52	37	21	60	43	50	434	373	304	466	397
21	58	42	25	67	49	51	453	390	319	486	416
22	65	48	30	75	55	52	473	408	335	507	434
23	73	54	35	83	62	53	494	427	351	529	454
24	81	61	40	91	69	54	514	445	368	550	473
25	89	68	45	100	76	55	536	465	385	573	493
26	98	75	51	110	84	56	557	484	402	595	514
27	107	83	57	119	92	57	579	504	420	618	535
28	116	91	64	130	101	58	602	525	438	642	556
29	126	100	71	140	110	59	625	546	457	666	578
30	137	109	78	151	120	60	648	567	476	690	600
31	147	118	86	163	130	61	672	589	495	715	623
32	159	128	94	175	140	62	697	611	515	741	646
33	170	138	102	187	151	63	721	634	535	767	669
34	182	148	111	200	162	64	747	657	556	793	693
35	195	159	120	213	173	65	772	681	577	820	718

Таблица К

Критические точки для К–С-критерия для малых выборок

n	$_{0.10}D$	$_{0.05}D$	n	$_{0.10}D$	$_{0.05}D$	n	$_{0.10}D$	$_{0.05}D$
3	0,636	0,708	14	0,314	0,349	25	0,238	0,264
4	0,565	0,624	15	0,304	0,338	26	0,233	0,259
5	0,509	0,563	16	0,295	0,327	27	0,229	0,254
6	0,468	0,519	17	0,286	0,318	28	0,225	0,250
7	0,436	0,483	18	0,278	0,309	29	0,221	0,245
8	0,410	0,454	19	0,271	0,301	30	0,218	0,242
9	0,387	0,430	20	0,265	0,294	31	0,214	0,238
10	0,369	0,409	21	0,259	0,287	32	0,211	0,234
11	0,352	0,391	22	0,253	0,281	33	0,208	0,231
12	0,338	0,375	23	0,247	0,275	34	0,205	0,227
13	0,325	0,361	24	0,242	0,269	35	0,202	0,224

Таблица L

Границы для критерия знаков (ЛГ- левая, ПГ- правая).

n – число ненулевых разностей

$\alpha(1)$	0,025		0,01		0,005		$\alpha(1)$	0,025		0,01		0,005	
$\alpha(2)$	0,05		0,02		0,01		$\alpha(2)$	0,05		0,02		0,01	
n	*ЛГ*	*ПГ*	*ЛГ*	*ПГ*	*ЛГ*	*ПГ*	*n*	*ЛГ*	*ПГ*	*ЛГ*	*ПГ*	*ЛГ*	*ПГ*
5	0	5	0	5	0	5	48	17	31	16	32	15	33
6	1	5	0	6	0	6	49	18	31	16	33	16	33
7	1	6	1	6	0	7	50	18	32	17	33	16	34
8	1	7	1	7	1	7	51	19	32	17	34	16	35
9	2	7	1	8	1	8	52	19	33	18	34	17	35
10	2	8	1	9	1	9	53	19	34	18	35	17	36
11	2	9	2	9	1	10	54	20	34	19	35	18	36
12	3	9	2	10	2	10	55	20	35	19	36	18	37
13	3	10	2	11	2	11	56	21	35	19	37	18	38
14	3	11	3	11	2	12	57	21	36	20	37	19	38
15	4	11	3	12	3	12	58	22	36	20	38	19	39
16	4	12	3	13	3	13	59	22	37	21	38	20	39
17	5	12	4	13	3	14	60	22	38	21	39	20	40
18	5	13	4	14	4	14	61	23	38	21	40	21	40
19	5	14	5	14	4	15	62	23	39	22	40	21	41
20	6	14	5	15	4	16	63	24	39	22	41	21	42
21	6	15	5	16	5	16	64	24	40	23	41	22	42
22	6	16	6	16	5	17	65	25	40	23	42	22	43
23	7	16	6	17	5	18	66	25	41	24	42	23	43
24	7	17	6	18	6	18	67	26	41	24	43	23	44
25	8	17	7	18	6	19	68	26	42	24	44	23	45
26	8	18	7	19	7	19	69	26	43	25	44	24	45
27	8	19	8	19	7	20	70	27	43	25	45	24	46
28	9	19	8	20	7	21	71	27	44	26	45	25	46
29	9	20	8	21	8	21	72	28	44	26	46	25	47
30	10	20	9	21	8	22	73	28	45	27	46	26	47
31	10	21	9	22	8	23	74	29	45	27	47	26	46
32	10	22	9	23	9	23	75	29	46	27	48	26	40
33	11	22	10	23	9	24	76	29	47	28	48	27	49
34	11	23	10	24	10	24	77	30	47	28	49	27	50
35	12	23	11	24	10	25	78	30	48	29	49	28	50
36	12	24	11	25	10	26	79	31	48	29	50	28	61
37	13	24	11	26	11	26	80	31	49	30	50	29	51
38	13	25	12	26	11	27	82	32	50	31	51	29	53
39	13	26	12	27	12	27	84	33	51	31	53	30	54
40	14	26	13	27	12	28	86	34	52	32	54	31	55
41	14	27	13	28	12	29	88	35	53	33	55	32	56
42	15	27	14	28	13	29	90	36	54	34	56	33	57
43	15	28	14	29	13	30	92	37	55	35	57	34	58
44	16	28	14	30	14	30	94	38	56	36	58	35	59
45	16	29	15	30	14	31	96	38	58	37	59	35	61
46	16	30	15	31	14	32	98	39	59	38	60	36	62

Таблица L_1

Левые границы двустороннего критерия знаков

n	0,05	0,01	*n*	0,05	0,01	*n*	0,05	0,01
101	41	38	139	58	54	177	75	71
102	41	38	140	58	55	178	76	72
103	42	38	141	59	55	179	76	72
104	52	39	142	59	56	180	77	73
105	42	39	143	60	56	181	77	73
106	43	40	144	60	57	182	78	74
107	43	40	145	61	57	183	78	74
108	44	41	146	61	57	184	79	75
109	44	41	147	62	58	185	79	75
110	45	42	148	62	58	186	80	75
111	45	42	149	63	59	187	80	76
112	46	42	150	63	59	188	81	76
113	46	43	151	63	60	190	82	77
114	47	43	152	64	60	192	82	78
115	47	44	153	64	61	194	83	79
116	47	44	154	65	61	196	84	80
117	48	45	155	65	62	198	85	81
118	48	45	156	66	62	200	86	82
119	49	46	157	66	62	210	91	86
120	49	46	158	67	63	220	95	91
121	50	46	159	67	63	230	100	96
122	50	47	160	68	63	240	105	100
123	51	47	161	68	64	250	110	105
124	51	48	162	69	65	260	114	109
125	52	48	163	69	65	270	119	114
126	52	49	164	69	66	280	124	118
127	52	49	165	70	66	290	128	123
128	53	49	166	70	66	300	133	128
129	53	50	167	71	67	350	157	151
130	54	50	168	71	67	400	180	174
131	54	51	169	72	68	450	204	198
132	55	51	170	72	68	500	228	221
133	55	52	171	73	69	550	252	245
134	56	52	172	73	69	600	276	268
135	56	53	173	74	70	700	324	316
136	57	53	174	74	70	800	372	364
137	57	53	175	75	71	900	421	411
138	58	54	176	75	71	1000	469	459

Таблица N

Значения ординат стандартной нормальной кривой $\varphi(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-0,5x^2}$ и интегралов вероятностей $\Phi(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_0^x e^{-0,5t^2}dt$

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
0,00	**0,3989**	**0,0000**	**0,10**	**0,3970**	**0,0398**	**0,20**	**0,3910**	**0,0793**
0,01	0,3989	0,0040	0,11	0,3965	0,0438	0,21	0,3902	0,0832
0,02	0,3989	0,0080	0,12	0,3961	0,0478	0,22	0,3894	0,0871
0,03	0,3988	0,0120	0,13	0,3956	0,0517	0,23	0,3885	0,0910
0,04	0,3986	0,0160	0,14	0,3951	0,0557	0,24	0,3876	0,0948
0,05	0,3984	0,0199	0,15	0,3945	0,0596	0,25	0,3867	0,0987
0,06	0,3982	0,0239	0,16	0,3939	0,0636	0,26	0,3857	0,1026
0,07	0,3980	0,0279	0,17	0,3932	0,0675	0,27	0,3847	0,1064
0,08	0,3977	0,0319	0,18	0,3925	0,0714	0,28	0,3836	0,1103
0,09	0,3973	0,0359	0,19	0,3918	0,0753	0,29	0,3825	0,1141
0,30	**0,3814**	**0,1179**	**0,40**	**0,3683**	**0,1554**	**0,50**	**0,3521**	**0,1915**
0,31	0,3802	0,1217	0,41	0,3668	0,1591	0,51	0,3503	0,1950
0,32	0,3790	0,1255	0,42	0,3653	0,1628	0,52	0,3485	0,1985
0,33	0,3778	0,1293	0,43	0,3637	0,1664	0,53	0,3467	0,2019
0,34	0,3765	0,1331	0,44	0,3621	0,1700	0,54	0,3448	0,2054
0,35	0,3752	0,1368	0,45	0,3605	0,1736	0,55	0,3429	0,2088
0,36	0,3739	0,1406	0,46	0,3589	0,1772	0,56	0,3410	0,2123
0,37	0,3725	0,1443	0,47	0,3572	0,1808	0,57	0,3391	0,2157
0,38	0,3712	0,1480	0,48	0,3555	0,1844	0,58	0,3372	0,2190
0,39	0,3697	0,1517	0,49	0,3538	0,1879	0,59	0,3352	0,2224
0,60	**0,3332**	**0,2257**	**0,70**	**0,3123**	**0,2580**	**0,80**	**0,2897**	**0,2881**
0,61	0,3312	0,2291	0,71	0,3101	0,2611	0,81	0,2874	0,2910
0,62	0,3292	0,2324	0,72	0,3079	0,2642	0,82	0,2850	0,2939
0,63	0,3271	0,2357	0,73	0,3056	0,2673	0,83	0,2827	0,2967
0,64	0,3251	0,2389	0,74	0,3034	0,2703	0,84	0,2803	0,2995
0,65	0,3230	0,2422	0,75	0,3011	0,2734	0,85	0,2780	0,3023
0,66	0,3209	0,2454	0,76	0,2989	0,2764	0,86	0,2756	0,3051
0,67	0,3187	0,2486	0,77	0,2966	0,2794	0,87	0,2732	0,3078
0,68	0,3166	0,2517	0,78	0,2943	0,2823	0,88	0,2709	0,3106
0,69	0,3144	0,2549	0,79	0,2920	0,2852	0,89	0,2685	0,3133
0,90	**0,2661**	**0,3159**	**1,00**	**0,2420**	**0,3413**	**1,10**	**0,2179**	**0,3643**
0,91	0,2637	0,3186	1,01	0,2396	0,3438	1,11	0,2155	0,3665
0,92	0,2613	0,3212	1,02	0,2371	0,3461	1,12	0,2131	0,3686
0,93	0,2589	0,3238	1,03	0,2347	0,3485	1,13	0,2107	0,3708
0,94	0,2565	0,3264	1,04	0,2323	0,3508	1,14	0,2083	0,3729
0,95	0,2541	0,3289	1,05	0,2299	0,3531	1,15	0,2059	0,3749
0,96	0,2516	0,3315	1,06	0,2275	0,3554	1,16	0,2036	0,3770
0,97	0,2492	0,3340	1,07	0,2251	0,3577	1,17	0,2012	0,3790
0,98	0,2468	0,3365	1,08	0,2227	0,3599	1,18	0,1989	0,3810

Таблица N (продолжение)

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
0,99	0,2444	0,3389	1,09	0,2203	0,3621	1,19	0,1965	0,3830
1,20	**0,1942**	**0,3849**	**1,30**	**0,1714**	**0,4032**	**1,40**	**0,1497**	**0,4192**
1,21	0,1919	0,3869	1,31	0,1691	0,4049	1,41	0,1476	0,4207
1,22	0,1895	0,3888	1,32	0,1669	0,4066	1,42	0,1456	0,4222
1,23	0,1872	0,3907	1,33	0,1647	0,4082	1,43	0,1435	0,4236
1,24	0,1849	0,3925	1,34	0,1626	0,4099	1,44	0,1415	0,4251
1,25	0,1826	0,3944	1,35	0,1604	0,4115	1,45	0,1394	0,4265
1,26	0,1804	0,3962	1,36	0,1582	0,4131	1,46	0,1374	0,4279
1,27	0,1781	0,3980	1,37	0,1561	0,4147	1,47	0,1354	0,4292
1,28	0,1758	0,3997	1,38	0,1539	0,4162	1,48	0,1334	0,4306
1,29	0,1736	0,4015	1,39	0,1518	0,4177	1,49	0,1315	0,4319
1,50	**0,1295**	**0,4332**	**1,60**	**0,1109**	**0,4452**	**1,70**	**0,0940**	**0,4554**
1,51	0,1276	0,4345	1,61	0,1092	0,4463	1,71	0,0925	0,4564
1,52	0,1257	0,4357	1,62	0,1074	0,4474	1,72	0,0909	0,4573
1,53	0,1238	0,4370	1,63	0,1057	0,4484	1,73	0,0898	0,4582
1,54	0,1219	0,4382	1,64	0,1040	0,4495	1,74	0,0878	0,4591
1,55	0,1200	0,4394	1,65	0,1023	0,4505	1,75	0,0863	0,4599
1,56	0,1182	0,4406	1,66	0,1006	0,4515	1,76	0,0848	0,4608
1,57	0,1163	0,4418	1,67	0,0989	0,4525	1,77	0,0833	0,4616
1,58	0,1145	0,4429	1,68	0,0973	0,4535	1,78	0,0818	0,4625
1,59	0,1127	0,4441	1,69	0,0957	0,4545	1,79	0,0804	0,4633
1,80	**0,0790**	**0,4641**	**1,90**	**0,0656**	**0,4713**	**2.00**	**0,0540**	**0,4772**
1,81	0,0775	0,4649	1,91	0,0644	0,4719	2,02	0,0519	0,4783
1,82	0,0761	0,4656	1,92	0,0632	0,4726	2,04	0,0498	0,4793
1,83	0,0748	0,4664	1,93	0,0620	0,4732	2,06	0,0478	0,4803
1,84	0,0734	0,4671	1,94	0,0608	0,4738	2,08	0,0459	0,4812
1,85	0,0721	0,4678	1,95	0,0596	0,4744	2,10	0,0440	0,4821
1,86	0,0707	0,4686	1,96	0,0584	0,4750	2,12	0,0422	0,4830
1,87	0,0694	0,4693	1,97	0,0573	0,4756	2,14	0,0404	0,4838
1,88	0,0681	0,4699	1,98	0,0562	0,4761	2,16	0,0387	0,4846
1,89	0,0669	0,4706	1,99	0,0551	0,4767	2,18	0,0371	0,4854
2.20	**0,0355**	**0,4861**	**2.40**	**0,0224**	**0,4918**	**2.60**	**0,0136**	**0,4953**
2,22	0,0339	0,4868	2,42	0,0213	0,4922	2,62	0,0129	0,4956
2,24	0,0325	0,4875	2,44	0,0203	0,4927	2,64	0,0122	0,4959
2,26	0,0310	0,4881	2,46	0,0194	0,4931	2,66	0,0116	0,4961
2,28	0,0297	0,4887	2,48	0,0184	0,4934	2,68	0,0110	0,4963
2,30	0,0283	0,4893	2,50	0,0175	0,4938	2,70	0,0104	0,4965
2,32	0,0270	0,4898	2,52	0,0167	0,4941	2,72	0,0099	0,4967
2,34	0,0258	0,4904	2,54	0,0158	0,4945	2,74	0,0093	0,4969
2,36	0,0246	0,4909	2,56	0,0151	0,4948	2,76	0,0088	0,4971
2,38	0,0235	0,4913	2,58	0,0143	0,4951	2,78	0,0084	0,4973

Таблица N (продолжение)

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
2,80	**0,0079**	**0,4974**	**3,00**	**0,00443**	**0,49865**			
2,82	0,0075	0,4976	3,10	0,00327	0,49903	4,00	0,0001338	0,499968
2,84	0,0071	0,4977	3,20	0,00238	0,49931			
2,86	0,0067	0,4979	3,30	0,00172	0,49952	4,50	0,0000160	0,499997
2,88	0,0063	0,4980	3,40	0,00123	0,49966			
2,90	0,0060	0,4981	3,50	0,00087	0,49977	5,00	0,0000015	0,49999997
2,92	0,0056	0,4982	3,60	0,00061	0,49984			
2,94	0,0053	0,4984	3,70	0,00042	0,49989			
2,96	0,0050	0,4985	3,80	0,00029	0,49993			
2,98	0,0047	0,4986	3,90	0,00020	0,49995			

Библиографический список

1. Новиков, Д. А. Статистические методы в педагогических исследованиях / Д. А. Новиков. – М.: МЗ-Пресс, 2004. – 66 с.

2. Майер, Р.А. Теория и практика статистического анализа в психолого-педагогических и социологических исследованиях / Р.А. Майер, Н.Р. Колмакова, А.В. Ванюрин. – Красноярск, РИО КГПУ. – 2005. – 350 с.

3. Майер, Р.А. Статистические методы в психолого-педагогических и социологических исследованиях / Р.А. Майер, Н.Р. Колмакова. – Красноярск, Изд-во КГПУ, – 2004. – 149 с.

Библиографический список

1. Новиков, Д. А. Статистические методы в педагогических исследованиях / Д. А. Новиков. – М.: МЗ-Пресс, 2004. – 66 с.

2. Майер, Р.А. Теория и практика статистического анализа в психолого-педагогических и социологических исследованиях / Р.А. Майер, Н.Р. Колмакова, А.В. Ванюрин. – Красноярск, РИО КГПУ. – 2005. – 350 с.

3. Майер, Р.А. Статистические методы в психолого-педагогических и социологических исследованиях / Р.А. Майер, Н.Р. Колмакова. – Красноярск, Изд-во КГПУ, – 2004. – 149 с.