Часть II. Математическая статистика

Глава 7. Выборочный метод статистического наблюдения

7.1. Генеральная и выборочная совокупности. Как было отмечено во введении, методы эмпирического статистического исследования могут быть разделены на сплошные и несплошные (выборочные). Сплошное статистическое исследование требует полного охвата объекта исследования, всех его элементов без исключения. Однако такое исследование по многим причинам может оказаться или очень трудоёмким, или требующим больших денежных затрат, или, наконец, просто невозможным. В этих случаях используются методы несплошного исследования, которые хорошо себя зарекомендовали в различных областях науки и техники. Существует несколько методов несплошного наблюдения. Среди них наиболее широкое распространение в социологии, педагогике и психологии получил *выборочный метод*. Основными понятиями выборочного метода являются понятия генеральной и выборочной совокупности.

Генеральной совокупностью называют множество объектов, которые являются предметом изучения в пределах, очерченных программой исследования и территориально - временными границами. Любую генеральную совокупность характеризует некоторый, явно задаваемый набор признаков, по значениям которых всегда можно определить, относится данный объект к генеральной совокупности или нет.

Часть объектов генеральной совокупности, выступающих в качестве объектов наблюдения, называется *выборочной совокупностью* или просто *выборкой.* Так как статистическое наблюдение осуществляется именно за элементами выборочной совокупности, то она должна быть построена таким образом, чтобы при минимуме исследуемых объектов удавалось с необходимой степенью гарантии представить всю генеральную совокупность.

Значения различных описательных мер (арифметического среднего, медианы, дисперсии, стандартного отклонения и т.п.), вычисленных для генеральной совокупности, называются *параметрами*. Для выборки те же описательные меры называются *статистиками*. Параметр описывает генеральную совокупность так же, как статистика выборку. Статистики принято обозначать латинскими буквами, а параметры – греческими. Статистику, вычисленную по выборке, можно рассматривать как оценку параметра совокупности. Например, выборочное среднее $\overline{x}$ – оценивает генеральную среднюю μ; выборочная дисперсия S^2 оценивает генеральную дисперсию σ^2; выборочное стандартное отклонение S – генеральное стандартное отклонение σ.

Свойство выборки при определённых условиях становиться более или менее точным отражением всей генеральной совокупности называется *репрезентативностью*. Как и всякое свойство, репрезентативность выборочных данных может быть выражена в достаточной или недостаточной степени. В первом случае статистики дают достоверные оценки генеральных параметров, во втором – недостоверные. Следует заметить, что получение недостоверных оценок не умаляет значения выборочных показателей для характеристики самой выборки. Получение же достоверных оценок значительно расширяет область их применения.

7.2. Систематические и случайные ошибки статистического наблюдения. При сборе статистической информации выборочным методом могут возникать ошибки различного рода. Их принято распределять по двум группам.

К первой группе относятся ошибки, которых можно избежать или свести к минимуму хорошей организацией исследования. Назовём некоторые из них:

1) ошибки методические, которые возникают при применении неправильной методики сбора и обработки материалов;

2) ошибки точности, которые возникают при первичной регистрации фактов, измерении непроверенными, испорченными инструментами; расчёты с недостаточной точностью;

3) ошибки внимания – описки, просчёты, опечатки;

4) ошибки типичности, происходящие оттого, что в выборку включаются объекты, нетипичные для всей генеральной совокупности.

Ошибки типичности могут быть допущены бессознательно, при непонимании того, что в выборку должны привлекаться объекты без учёта величины изучаемого признака, в случайном порядке. Ошибки типичности могут быть причиной ложных выводов, особенно если они совершаются сознательно, при тенденциозном подборе первичных данных.

Надо отбросить необоснованные надежды на то, что перечисленные выше методические ошибки (их ещё называют ошибками регистрации) могут каким-то образом быть обезврежены или учтены последующим применением математических методов.

Ко второй группе относятся так называемые ошибки *репрезентативности*, которые неизбежно возникают, когда по части характеризуют целое. Никакая, даже самая образцовая организационно - методическая работа (за исключением перехода на сплошное изучение генеральной совокупности), не может устранить ошибки репрезентативности. Однако эти ошибки могут быть учтены при последующем применении статистических методов.

В конкретных условиях ошибку репрезентативности можно свести к достаточно малой величине, которую методами математической статистики можно определить на основе анализа выборочных данных и учесть при оценке генерального параметра.

Определять величину ошибок репрезентативности требуется только для выборочных показателей, например, средней арифметической, разности двух выборочных средних, коэффициента корреляции и др., если они используются для оценки параметров генеральной совокупности, и при условии, что организация исследования сводит к минимуму все другие виды ошибок.

7.3. Построение выборки. Существует несколько способов построения выборки. Из них мы рассмотрим здесь только *простой 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 **перечень** элементов генеральной совокупности, который удовлетворяет требованиям полноты, точности, адекватности, удобства работы с ним, отсутствия дублирования единиц наблюдения.

Под **полнотой** подразумевается наличие в перечне всех единиц данной генеральной совокупности, а под **точностью** – точность информации о каждой единице отбора. **Адекватность** предполагает соответствие выборки задачам исследования. Удобство работы с перечнем и выборкой – существенное условие повышения качества исследования.

Различают простой случайный *повторный* отбор и простой случайный *бесповторный* отбор.

При простом случайном *повторном отборе* каждый выбранный объект после изучения и регистрации возвращается в генеральную совокупность, так что любой объект может попасть в выборку несколько раз.

При простом случайном *бесповторном отборе* объекты, отобранные, как и в предыдущем случае, случайно, не возвращаются в генеральную совокупность и не могут повторно попасть в выборку. Это наиболее распространённый способ организации выборки.

Чтобы устранить трудности, возникающие при исследовании больших генеральных совокупностей (они особенно велики в социологических исследованиях), для реализации простого случайного отбора пользуются либо *генератором случайных чисел*, который имеется в любом языке программирования, либо *таблицами случайных чисел*. При организации 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 классифицируются и вычисляются 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. Кратко характеризуем каждую из них.

Систематическая выборка. Из полного списка единиц генеральной совокупности отбирается по одному объекту через интервал, равный шагу отбора – отношению объема генеральной совокупности к объему выборки. Основным достоинством этого метода является то, что он позволяет с помощью простой техники отбора и при небольшом объеме выборки охватить сравнительно большие генеральные совокупности. Однако при этом существует опасность совпадения интервала отбора со скрытой периодичностью распределения признака в генеральной совокупности, что может привести к смещениям. Много проблем создает неопределенность выбора первоначального объекта (начальной точки отбора).

Гнездовая выборка. Выборочные единицы отбираются с помощью одного из указанных выше способов случайного отбора. Единицы отбора представляют собой статистические группы (гнезда), которые целиком или выборочно подвергаются обследованию. При этом методе уменьшаются затраты (по сравнению с простой случайной выборкой равного объема) на организацию процедур отбора, например, при пространственно разбросанной генеральной совокупности. Однако при формировании искусственных гнезд создается трудность отнесения каждого отдельного элемента генеральной совокупности только к одному гнезду и обеспечения приблизительно одинаковых размеров гнезд.

Стратифицированная выборка (районированная выборка, расслоенный отбор). Исследуемая совокупность предварительно разделяется на страты (слои) в соответствии с генеральным распределением известных и значимых для исследования признаков, после чего из каждой страты извлекается выборка. При удачном подразделении совокупности на однородные группы расслоенный отбор дает выигрыш в точности по сравнению с простым случайным. Однако если классификация на страты не совпадает с распределением изучаемых признаков, возможно смещение выборки по этим признакам.

При определении объема выборки в страте используются два подхода.

А. Пропорциональное размещение, при котором объем выборки из страты пропорционален размеру страты в генеральной совокупности.

В. Оптимальное размещение, при котором объем выборки из страты пропорционален в страте среднеквадратичному отклонению признака и обратно пропорционален издержкам на получение выборки. Этот случай выгодно использовать тогда, когда отбор из одних слоев обходится дороже, чем из других. Трудность в том, что при этом необходимо знать дисперсию признака расслоения внутри страты.

Многоступенчатая выборка. Процедура её построения разбивается на ряд этапов (ступеней), на каждом из которых меняется единица отбора. Отбор на каждой из ступеней может осуществляться любым из вышеописанных способов. Если на одной из ступеней выступает географический регион, то существенно уменьшается стоимость полевого исследования. В то же время надо иметь в виду, что при многоступенчатой выборке ошибка выборки, как правило, выше, чем для простого случайного или систематического отбора.

Квотная выборка. Производится разбиение генеральной совокупности на классы согласно нескольким распределениям выбранных признаков. На основе знания статистического объема каждого класса и заданной доли отбора из него определяется «квота» – объем выборки соответствующего класса. Квотный выбор обследуемых объектов из потенциально возможных возлагается на анкетеров или интервьюеров.

Квотная выборка удобна для случаев, когда размер выборки невелик. Уменьшается стоимость организации выборки, поскольку производительность труда анкетеров или интервьюеров выше при самостоятельном выборе лиц для опроса, чем при их поиске по спискам адресатов.

Квотная выборка основана на предположении, что распределения контролируемых признаков в выборке обеспечивают репрезентативность воспроизведения распределения зависимых признаков. Невозможно точно измерить смещения, вызванные неслучайным характером отбора лиц для опроса.

7.4. Понятие выборочного распределения. Оценивание значений параметров генеральной совокупности с помощью соответствующих статистик (μ с помощью $\bar{x}$, σ^2 с помощью S^2, ρ с помощью r и т.д.) производится на основе так называемого *выборочного распределения*. Поясним смысл этого понятия.

Пусть в заданной генеральной совокупности требуется оценить некоторый параметр θ, например, математическое ожидание (μ), медиана (M_e), дисперсия (σ^2), которому соответствует статистика $\overline{\theta}$: $\bar{x}$, $\overline{M_e}$, S^2 соответственно.

Для построения выборочного распределения, соответствующего статистике $\overline{\theta}$, из данной генеральной совокупности мысленно бесконечное число раз случайным образом извлекают выборки одного и того же объёма (n) и для каждой из них вычисляют $\overline{\theta}$ (конкретно $\bar{x}, \overline{M_e}, S^2$ и т.п.). По полученным данным строится распределение частот значений $\overline{\theta}$. Это распределение и называют выборочным распределением статистики $\overline{\theta}$.

При оценке параметра θ с помощью выборочного распределения соответствующей статистики $\overline{\theta}$ это распределение путём несложных преобразований сводят к одному из известных теоретических распределений: стандартизированному нормальному, t распределению Стьюдента, хи-квадрат распределению, распределению Фишера и т.д., речь о которых будет идти в следующей главе.

Остановимся теперь более подробно на вопросе построения и особенностях выборочного распределения среднего арифметического ($\bar{x}$).

Пусть в урне содержится шестьсот небольших, неразличимых на ощупь шаров, помеченных цифрами от 1 до 6: каждая цифра на ста шарах. Это будет нашей генеральной совокупностью. Она характеризуется распределением, приведённым в таблице 7.1, и в графической форме (полигоном относительных частот) на рисунке 7.1.

Таблица 7.1. Распределение случайной величины X

Значения случайной величины, X	1	2	3	4	5	6
Частота возможного появления каждой цифры, n_i	100	100	100	100	100	100
Относительная частота, ω_i	1/6	1/6	1/6	1/6	1/6	1/6

Извлечём теперь из этой совокупности случайным образом 4 шара ($n = 4$). Вычислив среднее арифметическое обозначенных на них цифр, вернем шары в урну, после чего содержимое последней тщательно перемешаем. Операцию повторим достаточно большое число раз, например 100 раз, и построим распределение вычисленных средних арифметических значений (табл. 7.2).

Вместо шаров можно бросать четыре игральные кости, грани которых обозначены цифрами от 1 до 6.

В графической форме это распределение представлено полигоном относительных частот на рисунке 7.2 (а). Как видим, он сильно отличается от полигона относительной частоты самой генеральной совокупности (рис. 7.1).

Таблица 7.2. Распределение средних значений по четырём шарам

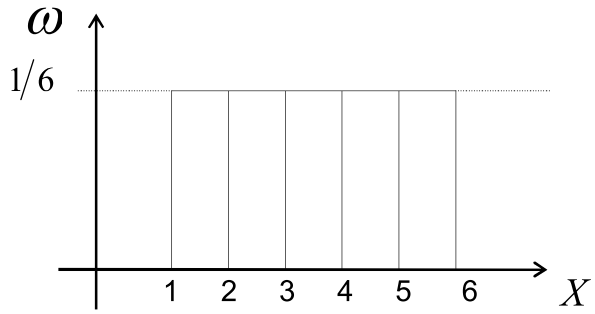


Рис. 7,1. Полигон частостей генеральной совокупности

Суммы цифр на шарах	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Итого
Средние значения, $\overline{x}$	1,0	1,25	1,5	1,75	2,0	2,25	2,5	2,75	3,0	3,25	3,5	3,75	4,0	4,25	4,5	4,75	5,0	5,25	5,5	5,75	6,0	-
Частота, n_i	0	1	1	3	4	5	2	4	13	12	13	14	6	4	11	2	4	1	0	0	0	100
Относительная частота, ω_i	0	0,01	0,01	0,03	0,04	0,05	0,02	0,04	0,13	0,12	0,13	0,14	0,06	0,04	0,11	0,02	0,04	0,01	0	0	0	1

Продолжая извлечение выборок двести, четыреста, шестьсот раз, получим полигоны относительной частоты, изображённые на рисунках 7.2 (б), 7.2 (в) и 7.2 (г). Все они определены на одном и том же множестве значений (1,0; 1,25; 1,5; 1,75; и т.д. до 6,0), представляющем конечную арифметическую прогрессию с разностью (шагом), равной 0,25. С увеличением числа выборок это множество не меняется, сами же полигоны деформируются.

Рис. 7.2(а)

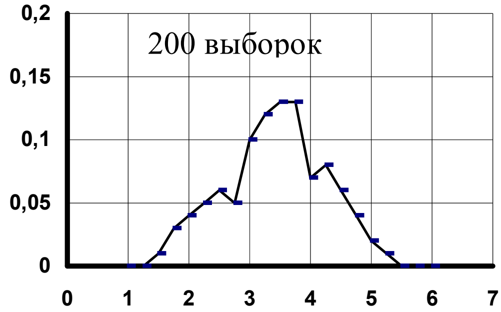


Рис. 7.2(б)

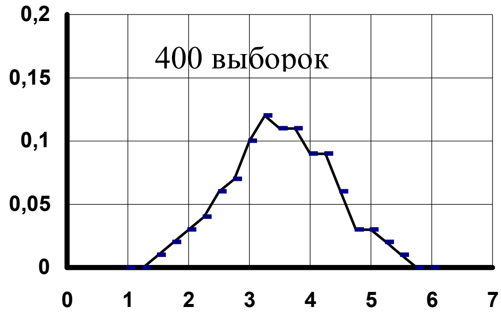


Рис. 7.2 (в)

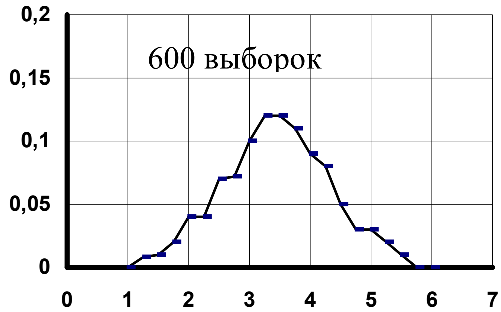


Рис. 7.2 (г)

Предположим теперь, что процесс извлечения выборок ($n = 4$) продолжается бесконечно. Можно предвидеть, что величина деформации полигона будет от шага к шагу ослабевать, а форма его приближаться к некоторому, в определённом смысле предельному, полигону, который и выражает *выборочное распределение.* В данном случае оно содержит 20 звеньев. С увеличением объёма выборок число звеньев будет расти. В частности, если каждый раз извлекать не 4, а 20 шаров, то звеньев будет уже не 20, а 100, и полигон будет больше напоминать гладкую кривую линию.

Одна из основных теорем теории вероятностей, называемая *центральной предельной теоремой*, гласит: если выборки объёма n случайным образом извлекаются из бесконечно большой совокупности с параметрами μ и σ^2, где μ – среднее значение генеральной совокупности, а σ^2 – её дисперсия, то выборочное распределение средних арифметических будет близким к нормальному. При этом среднее значение всех выборочных средних – обозначим его $M[\bar{x}]$ – будет равно μ, а дисперсия – обозначим её $D[\bar{x}]$ – равна $\frac{\sigma^2}{n}$.

Это обстоятельство является одной из причин, объясняющих столь важную роль нормального распределения в статистике.

В рассмотренном нами примере с шарами среднее значение генеральной совокупности (μ), как легко подсчитать, равно 3,5, а дисперсия σ^2 =2,917. В соответствии с центральной предельной теоремой, выборочное распределение средних будет близким к нормальному распределению, среднее значение которого $M[\bar{x}] = 3{,}5$, дисперсия $D[\bar{x}] = \frac{\sigma^2}{n} = \frac{2{,}917}{4} = 0{,}73$, а стандартное отклонение – обозначим его $\sigma[\bar{x}]$ – равно $\frac{\sigma}{\sqrt{n}}$ = 0,85.

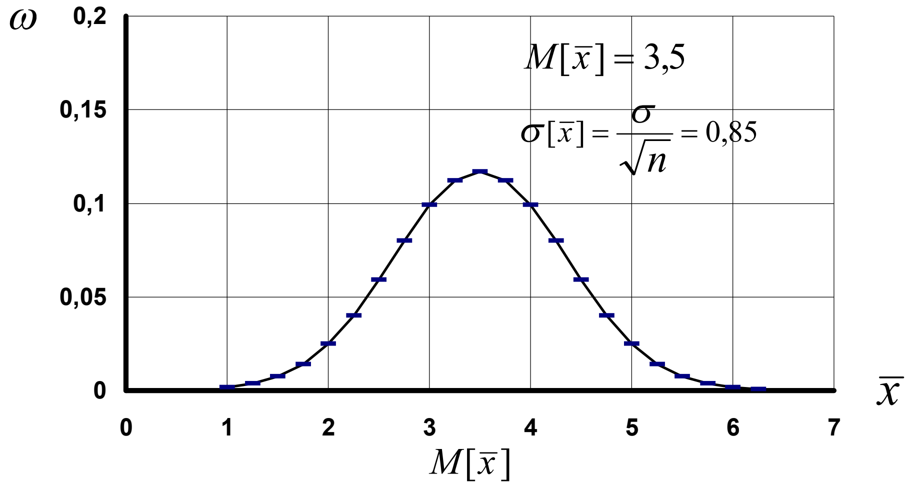


Рис. 7.3. Полигон выборочного распределения для $\bar{x}$

Так как в данном случае мы имеем дело с нормальным распределением, совокупности значений которой заданы в интервальной шкале с шагом l=0,25, то относительная частота $\tilde{\omega}_i$, соответствующая i - му интервалу может быть вычислена по формуле $\tilde{\omega}(x_i) = \frac{n_i}{n} = l \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i - \mu)^2}{2 \cdot \sigma^2}}$, где x_i – середина i -го интервала (см. формулу 6.4).

Так как μ=3,5; $\sigma^2 = 0{,}73$; $\sigma = 0{,}85$; $l = 0{,}25$, то $\tilde{\omega}(x_i) = 0{,}117 \cdot e^{-\frac{(x_i - 3{,}5)^2}{1{,}46}}$.

Пользуясь этой формулой, вычислим $\tilde{\omega}(x_i)$ для значений x_i от 1,0 до 6,0 с шагом $l = 0,25$.

Соответствующее распределение в графической форме приведено на рисунке 7.3.

Дисперсию средних $D[\bar{x}] = \dfrac{\sigma^2}{n}$ часто называют *дисперсией ошибки среднего*, а стандартное отклонение средних $\sigma[\bar{x}] = \dfrac{\sigma}{\sqrt{n}}$ - *стандартной ошибкой среднего*, или *ошибкой репрезентативности среднего*.

Глава 8. Интервальное оценивание параметров

8.1. Точечное оценивание параметров генеральной совокупности

Оценка параметра называется *точечной*, если в качестве его значения принимается одно число, вычисляемое по определённому правилу по данным выборки. Часто в качестве такой оценки принимается значение статистики, соответствующей оцениваемому параметру. Качество точечной оценки характеризуется такими свойствами, как несмещённость и состоятельность. Оценка параметра θ (например, статистика $\overline{\theta}$) называется *несмещённой*, если математическое ожидание выборочного распределения $\overline{\theta}$ равно величине оцениваемого параметра θ $\left(M\left[\overline{\theta}\right]=\theta\right)$.

Как мы видели, независимо от характера генеральной совокупности математическое ожидание выборочных средних $M[\overline{x}]=\mu$ и, следовательно, $\overline{x}$ представляет собой несмещённую оценку μ.

Если выборки извлекаются из нормального или любого другого симметричного распределения случайным образом, то среднее выборочных медиан Me также является несмещённой оценкой генерального среднего, т.е. $M[Me]=\mu$. Если совокупность ещё и унимодальна, то и мода будет несмещённой оценкой для генерального среднего μ: $M[M_0]=\mu$.

Несмещённой оценкой дисперсии σ^2 является «исправленная» выборочная дисперсия $S^2=\frac{1}{n-1}\sum\left(x_i-\overline{x}\right)^2$, так как $M\left[S^2\right]=\sigma^2$. В то же время величина $\frac{1}{n}\sum\left(x_i-\overline{x}\right)^2$ представляет собой отрицательно смещённую оценку σ^2 так как: $M\left[\frac{1}{n}\sum\left(x_i-\overline{x}\right)^2\right]\le\sigma^2$. Отрицательно смещённой является и выборочное стандартное отклонение: $M[S]\le\sigma$. В таблице 8.1 приведены основные параметры и характеристики их точечных оценок.

Таблица 8.1. Основные оценки параметров распределения

Параметр	Характер генеральной совокупности	Оценка	Характеристика оценки
μ	любая совокупность	$\overline{x}$	Несмещённая
μ	симметричная	Me	Несмещённая
μ	симметричная и унимодальная	Mo	Несмещённая
μ	асимметричная	Me, Mo	Смещённые
σ^2	любая совокупность	S^2	Несмещённая
σ	нормальная	S	Отрицательно смещённая

Оценка называется *состоятельной*, если при неограниченном увеличении объёма выборки её среднее арифметическое приближается к значению оцениваемого ею параметра. По-видимому, все несмещённые оценки состоятельны. Состоятельными являются и многие смещённые оценки, например, неисправленная дисперсия $\frac{\sum(x_i - \bar{x})^2}{n}$, стандартное отклонение S.

Непосредственное применение точечного оценивания весьма затруднительно, ведь исследователь не располагает выборочным распределением и его параметрами, в его распоряжении – одна выборка и одно значение соответствующей статистики. А оно даже при большом объёме выборки может существенно отличаться от значения оцениваемого параметра.

8.2. Интервальное оценивание параметров генеральной совокупности. В середине 20 века Дж. Нейман и Э. Пирсон для оценки параметров стали использовать не одно значения соответствующей статистики, а два, образующих так называемый *доверительный интервал.*

Рассмотрим суть этого понятия и метод его использования. Пусть требуется оценить некоторый параметр генеральной совокупности, который мы обозначим буквой θ. Для этого, в соответствии с предложенным Нейманом и Пирсоном методом, вводятся два числа (θ_1 и θ_2), между которыми, как ожидается, находится величина искомого параметра θ.

Очевидно, чем больше разность $\theta_2 - \theta_1$, тем больше вероятность γ того, что интервал (θ_1, θ_2) «накроет» значение оцениваемого параметра θ и, наоборот, чем меньше эта разность, тем меньше вероятность γ. При такой системе оценивания интервал $(\theta_1,\ \theta_2)$ называют *доверительным интервалом,* а вероятность γ выполнения неравенства $\theta_1 \le \theta \le \theta_2$. – *доверительной вероятностью.* Числа θ_1 и θ_2, определяемые по данным выборки, называют *доверительными границами,* θ_1– *нижней (левой) границей,* θ_2– *верхней (правой) границей.*

Практика психолого-педагогических и социологических исследований выработала три основных порога доверительной вероятности: первый при обычной ответственности ($\gamma = 0{,}95$), второй при повышенной ответственности ($\gamma = 0{,}99$), третий при высокой ответственности ($\gamma = 0{,}999$).

Когда говорят, что интервал (θ_1, θ_2) с вероятностью γ, например, с вероятностью 0,99, накрывает θ, то это означает, что если бы исследователи повторяли извлечение выборок и построение доверительных интервалов для θ многократно, то 99% этих интервалов содержали бы значение

θ. Однако в реальных условиях всякий исследователь использует, как правило, лишь одну выборку. Ему неизвестно, будет ли построенный им доверительный интервал содержать искомый параметр. Если доверительная вероятность достаточно велика, исследователю трудно удержаться от соблазна, считая, что этот, только что полученный им конкретный интервал содержит значение параметра. Это служит веской причиной для выбора большого значения γ. Однако если он разумен, то будет допускать незначительную вероятность (0,05; 0,01; 0,001) того, что его утверждение ошибочно. Эту вероятность, равную $1-\gamma$, называют статистической значимостью и обозначают α. Таким образом, $\alpha = 1-\gamma$.

Возникает вопрос, как, условившись об уровне доверительной вероятности γ, или статистической значимости α, и, располагая выборочными значениями статистики, вычислить доверительные границы. Проиллюстрируем этот метод на примере оценивания математического ожидания генеральной совокупности.

8.3. Построение доверительного интервала для математического ожидания. Рассмотрим вопрос о построении доверительного интервала для математического ожидания μ генеральной совокупности в предположении, что известна её дисперсия σ^2. С этой целью, прежде всего, вспомним, что выборочное распределение для $\bar{x}$ является нормальным распределением со средним $M[\bar{x}]=\mu$ и стандартным отклонением $\sigma[\bar{x}] = \dfrac{\sigma}{\sqrt{n}}$

Стандартизируя это распределение подстановкой

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}, \qquad (8.1)$$

получим для случайной величины z стандартизированное нормальное распределение, математическое ожидание которой равно нулю, а стандартное отклонение – единице.

Предположим, что исследователь организовал одну выборку и, вычислив значение $\bar{x}$, преобразовал его, пользуясь формулой (8.1), в значение z, которое мы изобразим точкой на оси Z (рис. 8.1).

Для того чтобы судить о её возможной близости к математическому ожиданию $M[z]$, которое, как мы видели, равно нулю, вспомним (пункт 3.4), что при нормальном распределении случайной величины z интервал $(M[z]-\sigma[z];\ M[z]+\sigma[z])$, или, учитывая, что z стандартизирована, интервал (-1; 1) будет содержать примерно 68 процентов значений случайной величины z (точнее 68,3%) интервал в два раза больший $(-2;\,2)$ – примерно 95 процентов (точнее 95,4%), а интервал $(-3;\,3)$ – 99,7 процентов (точнее, 99,73%) таких значений. Следовательно, вероятность того, что найденное исследователем значение z попадёт в каждый из рассмотренных интервалов, равна соответственно 0,683, 0,954 и 0,9973.

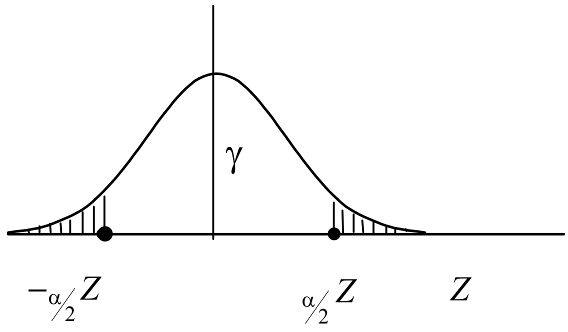


Рис. 8.1. Стандартизированное нормальное распределение

В общем случае для любой вероятности γ может быть указан интервал $\left(-_{\alpha/2}\,z;\ _{\alpha/2}z\right)$, вероятность попадания в который точки z равна $\gamma = 1-\alpha$. Чем больше значение $_{\alpha/2}z$, тем шире интервал и тем больше доверительная вероятность γ. Каждому значению $_{\alpha/2}z$ соответствует своя доверительная вероятность γ, при этом большему значению $_{\alpha/2}z$ соответствует большее значение γ.

Доверительная вероятность γ попадания z в интервал $\left(-_{\alpha/2}\,z;\ _{\alpha/2}z\right)$ численно равна площади незаштрихованной части криволинейной трапеции, которая может быть вычислена интегрированием функции плотности $\varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2}$ на интервале $\left(-_{\alpha/2}\,z;\ _{\alpha/2}z\right)$. В результате интегрирования получим, что $\gamma = 2\frac{1}{\sqrt{2\pi}}\int\limits_0^{_{\alpha/2}z} e^{-\frac{t^2}{2}}dt$. Воспользовавшись функцией Лапласа $\Phi(x) = \int\limits_0^x e^{-\frac{1}{2}t^2}\,dt$, которая хорошо изучена и табулирована (табл. N Приложения), последнее равенство можно записать в виде: $\gamma = 2 \cdot \Phi\left(_{\alpha/2}z\right)$.

Теперь для каждого значения $_{\alpha/2}z$ (в таблице N Приложения им соответствуют данные столбца x) можно найти значение γ (удвоенное значение столбца Ф(x)).

Заметим, что вероятность попадания случайной величины z в каждый из отсекаемых построенным интервалом «хвостов» распределения равна $\alpha/2$, где $\alpha = 1-\gamma$.

Вероятность попадания значения z в область, лежащую выше доверительного интервала, т.е. правее точки ${}_{\alpha/2}Z$, равна $\alpha/2$. Нижнюю границу доверительного интервала, которую мы обозначили $-_{\alpha/2}Z$, можно обозначить ${}_{1-\alpha/2}Z$. Вторая форма записи означает, что вероятность попадания случайной величины z в область, лежащую правее точки ${}_{1-\alpha/2}Z$, равна $1-\alpha/2$. В данном пособии значения верхней доверительной границы для основных уровней статистической значимости α приведены в таблице А Приложения (строка, соответствующая двустороннему критерию $\alpha(2)$). Для удобства зависимость между верхними границами доверительных интервалов и статистической значимостью для трёх порогов надёжности приведена в таблице 8.2.

Таблица 8.2. Три порога надёжности оценок

Пороги	Требования к надежности	Надёжность γ	Статистическая значимость α	Верхняя граница доверительного интервала ${}_{\alpha/2}Z$
1	Обычные	0,95	0,05	1,960
2	Повышенные	0,99	0,01	2,576
3	Высокие	0,999	0,001	3,291

Таким образом, выбрав доверительную вероятность γ $(0<\gamma<1)$ и найдя соответствующее значение ${}_{\alpha/2}Z$, можно построить интервал $\left(-_{\alpha/2}z;\ {}_{\alpha/2}z\right)$, которому с вероятностью γ будет принадлежать значение z, соответствующее случайно составленной выборке объема n, т.е. удовлетворятся двойное неравенство

$$-_{\alpha/2}z < z <_{\alpha/2} z\ . \tag{8.2}$$

Чтобы построить доверительный интервал для величины $\bar{x}$ подставим в неравенство (8.2) значение $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$. В результате получим:

$$-_{\alpha/2}Z < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} <_{\alpha/2} Z, \quad \text{откуда}$$

$$\mu -_{\alpha/2} Z\cdot\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu +_{\alpha/2} Z\cdot\frac{\sigma}{\sqrt{n}}. \tag{8.3}$$

Из последнего неравенства следует, что интервалу

$$(\mu - {}_{\alpha/2}Z \cdot \frac{\sigma}{\sqrt{n}};\ \mu + {}_{\alpha/2}Z \cdot \frac{\sigma}{\sqrt{n}})$$

с вероятностью γ будет принадлежать значение $\bar{x}$ случайно составленной выборки объёма n.

Вернёмся теперь к неравенству $\mu - {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}} \le \bar{x} \le \mu + {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}}$.

Для дальнейшего очень важно, что оно равносильно неравенству

$$\bar{x} - {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}}. \tag{8.4}$$

Действительно, неравенство $\mu - {}_{\alpha/2}Z \cdot \frac{\sigma}{\sqrt{n}} \le \bar{x}$ можно заменить неравенством $\mu \le \bar{x} + {}_{\alpha/2}Z \cdot \frac{\sigma}{\sqrt{n}}$, а неравенство $\bar{x} \le \mu + {}_{\alpha/2}Z \cdot \frac{\sigma}{\sqrt{n}}$ – неравенством $\bar{x} - {}_{\alpha/2}Z \cdot \frac{\sigma}{\sqrt{n}} \le \mu$, откуда и следует неравенство $\bar{x} - {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}}$.

Таким образом, найдя $\bar{x}$ и зная σ и n, можно построить интервал

$$\left(\bar{x} - {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}};\ \bar{x} + {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}}\right), \tag{8.5}$$

который с заданной вероятностью γ содержит математическое ожидание μ генеральной совокупности.

Сопоставляя последнее неравенство с неравенством $\theta_1 \le \theta \le \theta_2$ (пункт 8.2), можно заключить, что при оценке параметра μ левая (нижняя) доверительная граница $\theta_1 = \bar{x} - {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}}$, а правая (верхняя) $\theta_2 = \bar{x} + {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}}$. Этот факт иногда обозначают равенством

$$\mu = \bar{x} \pm {}_{\alpha/2}Z\frac{\sigma}{\sqrt{n}}. \tag{8.6}$$

Величину $m_\mu = \frac{\sigma}{\sqrt{n}}$ называют *ошибкой репрезентативности*, а $\Delta = {}_{\alpha/2}Z \cdot m_\mu = {}_{\alpha/2}Z \cdot \frac{\sigma}{\sqrt{n}}$ – *предельной ошибкой выборки.*

Заметим, что, если генеральная совокупность конечна и содержит N элементов, а объём выборки (n) превышает 4% от N, то ошибка репрезентативности m_{μ} вычисляется по формуле:

$$m_{\mu} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}. \tag{8.7}$$

В качестве примера построим доверительный интервал, соответствующий доверительной вероятности $\gamma = 0{,}95$ для математического ожидания генеральной совокупности, удовлетворяющей условиям центральной предельной теоремы, дисперсия которой равна 9. Предположим, что среднее значение случайной выборки объема 225 человек оказалось равным 24,3. Так как $\gamma = 0{,}95$ соответствует ${}_{\alpha/2}Z = 1{,}96$, то предельная ошибка равна $\Delta = 1{,}96 \cdot \frac{3}{15} \approx 0{,}39$, а доверительный интервал имеет вид (23,91; 24,69). Если генеральная совокупность состояла из 1300 элементов, то ошибку репрезентативности, а следовательно, и предельную ошибку следует умножить на упомянутый выше коэффициент $\sqrt{\frac{N-n}{N-1}}$, который в нашем случае равен $\sqrt{\frac{1300-225}{1300-1}} = 0{,}91$. В результате интервал уменьшается до (23,95; 24.65).

Глава 9. Три важных распределения выборочных статистик

9.1. Понятие степени свободы. Прежде чем знакомиться с другими вероятностными распределениями, введем понятие *степени свободы*, которым в дальнейшем нам придется пользоваться довольно часто. Термин этот широко используется в статистике и механике для характеристики количества свободно варьируемых элементов совокупности. Если совокупность n величин связана линейными соотношениями так, что, зная $n-k$ из них, можно определить остальные k, то число степеней свободы ν считается равным $n-k$.

Пусть, например, для исследования требуется взять 10 объектов с любым развитием изучаемого признака. Так как в данном случае величина признака не имеет никаких ограничений, то число степеней свободы ν=10 – 0 =10. Если же, например, требуется взять десять чисел с условием, что их сумма равна, скажем, 100, то первых девять из них выбираются свободно, а десятое не имеет свободы выбора, и поэтому система имеет девять степеней свободы.

При вычислении средней арифметической никаких ограничений величины признака не имеется. Поэтому число степеней свободы равно числу данных. При вычислении дисперсии, как и при вычислении среднего квадратичного отклонения, для совокупности из n элементов имеется одно ограничение: среднее арифметическое этих данных должно равняться заданному числу. Поэтому число степеней свободы равно числу данных без одной $\nu = n-1$. Под числом степеней свободы ν используемой статистики понимают разность между числом независимых наблюдений, которое равно объему выборки n и числом оцениваемых по выборке параметров k: $\nu = n-k$.

9.2. Распределение Стьюдента. Обратимся теперь к распределению Стьюдента, введенному в начале 20-го века В. С. Госсетом (псевдоним «Стьюдент»), который, рассмотрев случайную величину

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \qquad (9.1)$$

и изучив ее выборочное распределение, показал, как это распределение можно использовать для оценки математического ожидания в случае, когда неизвестна дисперсия генеральной совокупности.

Для построения соответствующего выборочного распределения из заданной генеральной совокупности случайным образом извлечем n элементов и подсчитав для них $\bar{x}$ и S, вычислим значение t. Если эту процедуру мысленно повторить бесконечное число раз, построить соответствующий полигон, сгладить его и, наконец, подобрать масштаб по оси ординат с таким расчётом, чтобы площадь под кривой стала равна единице, то получим t-распределение Стьюдента с $n-1$ степенью свободы ($\nu = n-1$), которое обозначают t_ν.

При построении выборки предполагается, что отдельные наблюдения x_i независимы и распределены (хотя бы приближенно) нормально.

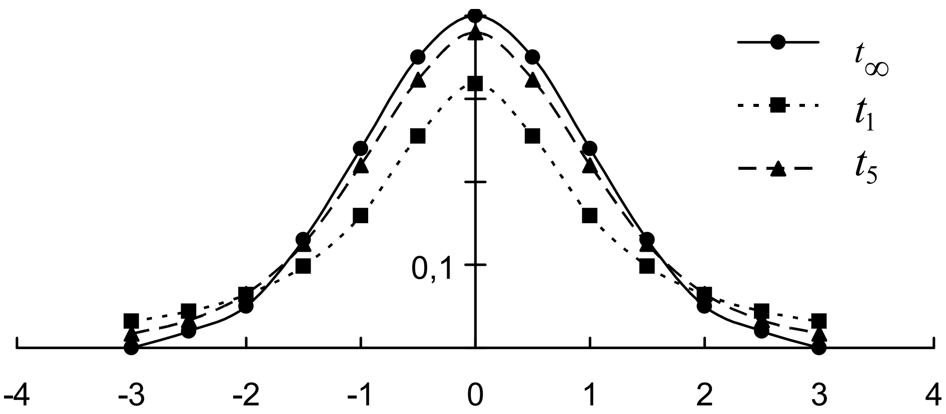


Рис. 9.1. t-распределение с 1 и 5 степенями свободы и стандартное нормальное распределение t_∞

Его вид, как показал Стьюдент, зависит только от числа степеней свободы. Чем меньше оно, тем более пологой будет кривая t - распределение Стьюдента, в результате чего увеличивается вероятность на «хвостах» кривой и уменьшается в центре (рис. 9.1). С увеличением числа степеней свободы распределение сходится к нормальному распределению.

Верхние границы доверительных интервалов t-распределения для основных значений α (0,05; 0,025; 0,01; 0,005) приведены в таблице В Приложения. Так, например, ${}_{0.05}t_5$=2,015, означающее, что вероятность принятия случайной величиной t_5 значения, превышающего число 2,015, равна 0,05. Заметим, что для стандартизированного нормального распределения ${}_{0.05}z$=1,645.

При любом ν t - распределение Стьюдента обладает свойствами:

1) оно симметрично и унимодально;

2) его среднее арифметическое равно нулю;

3) его дисперсия равна $\frac{\nu}{\nu-2}$, а стандартное отклонение $\sqrt{\frac{\nu}{\nu-2}}$;

4) все распределения имеют несколько увеличенный эксцесс.

Располагая распределением Стьюдента, вернемся к вопросу о построении доверительного интервала математического ожидания генеральной совокупности для случая, когда σ^2 неизвестно. Повторяя рассуждения, приведенные в пункте 8.3 предыдущей главы, применительно к t- распределению Стьюдента, получим неравенство $\bar{x} - {}_{\alpha/2}t_{\nu} \cdot \frac{S}{\sqrt{n}} \le \mu \le \bar{x} + {}_{\alpha/2}t_{\nu} \cdot \frac{S}{\sqrt{n}}$, в котором $\bar{x}$ и S соответственно выборочное среднее и стандартное отклонение, а ${}_{\alpha/2}t_{\nu}$ – верхняя граница доверительного интервала распределения Стьюдента, зависящая от числа степеней свободы ν и от выбранного уровня статистической значимости $\alpha = 1-\gamma$, где γ заданная надежность. Соответствующий доверительный интервал будет иметь вид:

$$(\bar{x} - {}_{\alpha/2}t_{\nu} \frac{S}{\sqrt{n}};\ \bar{x} + {}_{\alpha/2}t_{\nu} \frac{S}{\sqrt{n}}). \tag{9.2}$$

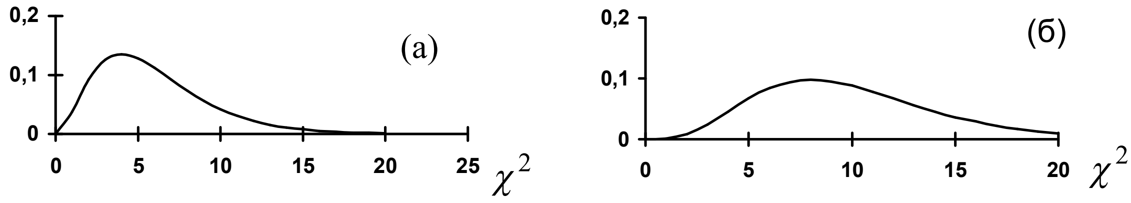
Если, например, γ=0,95 (α=0,05), объем выборки n=25 и вычисленные по выборке S=3 и $\bar{x}$=24,3, то для построения доверительного интервала математического ожидания нужно прежде всего найти значение ${}_{\alpha/2}t_{\nu} = {}_{0.025}t_{24}$=2,064. Пользуясь этими данными, получим, что ошибка репрезентативности $m_{\mu} = \frac{S}{\sqrt{n}} = \frac{3}{\sqrt{25}} = 0{,}6$, откуда предельная ошибка выборки $\Delta = {}_{\alpha/2}t_{\nu} \cdot m_{\mu} = 2{,}064 \cdot 0{,}6 \approx 1{,}24$, а доверительный интервал, определяемый равенством $\mu = 24{,}3 \pm 1{,}24$, имеет вид $(23{,}06\ \ 25{,}54)$.

9.3. Хи-квадрат распределение. χ^2- распределение было введено Р. Хельмертом и К. Пирсоном в конце 19 века. Для его построения из стандартизированной нормально распределенной генеральной совокупности случайным образом извлекаются выборки $x_1, x_2, \cdots, x_n$, для каждой из которых вычисляется случайная величина χ^2, равная сумме квадратов $\sum_{i=1}^{n} x_i^2$.

После регистрации многих тысяч её значений и построения полигона частот его сглаживают и подбирают такой масштаб на оси ординат, чтобы площадь под образовавшейся кривой была равна единице. В результате получается выборочное распределение случайной величины χ^2, которое называют *хи-квадрат распределением с* $\nu = n-1$ *степенями свободы и* обозначают χ^2_{ν}. То же самое распределение можно получить, если выборки объема n извлекать из нормально распределённой генеральной совокупности общего вида с дисперсией σ^2. Однако в этом случае после извлечения выборки объема n, прежде чем составлять суммы квадратов, надо каждое значение выборки подвергнуть стандартному преобразованию $z_i = \frac{x_i - \bar{x}}{\sigma}$. В этом случае χ^2 примет вид:

$$\chi_\nu^2 = \chi_{n-1}^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{\sigma^2} = \frac{(n-1)\cdot S_n^2}{\sigma^2}. \qquad (9.3)$$

На рисунке 9.2 изображены графики χ_ν^2- распределения с 6-ю и 10-ю степенями свободы: χ_6^2 и χ_{10}^2. Для любого целого ν (1,2,3, ...) существует своё χ_ν^2- распределение. Его свойства зависят, как и в распределении Стьюдента, только *от числа степеней свободы* ν.



Кривая распределения χ_6^2 Кривая распределения χ_{10}^2

Рис. 9.2

Приведём некоторые из свойств χ_ν^2 распределения:

1) его среднее значение равно числу степеней свободы;

2) мода χ_n^2 при $n \geq 2$ находится в точке $n-2$;

3) стандартное отклонение χ_n^2 равно $\sqrt{2n}$;

4) с ростом n χ_n^2 стремится к нормальному распределению со средним n и стандартным отклонением $\sqrt{2n}$;

5) если две независимые величины распределены по закону χ^2 с ν_1 и ν_2 степенями свободы, то их сумма имеет χ^2- распределение с $\nu_1 + \nu_2$ степенями свободы.

Используя применявшийся выше принцип обозначения критических точек распределения, обозначим символом ${}_\alpha\chi_n^2$ точку хи-квадрат распределения с n степенями свободы, обладающую тем свойством, что вероятность попадания случайной величины χ_n^2 в область, расположенную правее точки ${}_\alpha\chi_n^2$, равна α. Например, равенство ${}_{0,05}\chi_{10}^2$=18.31 означает, что вероятность принятия случайной величиной χ_{10}^2 значения, превышающего 18,31, равна 0,05.

Значения точек для статистической значимости α, равной 0,99; 0,975; 0,950; 0,05 и 0,025 и 0,01 приведены в таблице С Приложения. Пусть например, надо выделить интервал, вне которого содержится 5% значений случайной переменной χ^2_{10}. Для этого отделим слева и справа части криволинейной трапеции, содержащие по 2,5% площади всей трапеции. Нижняя граничная точка интервала ${}_{0,975}\chi^2_{10} = 3{,}25$, а верхняя – ${}_{0,025}\chi^2_{10} = 20{,}48$, откуда следует, что искомый интервал, содержащий 95% всех значений случайной переменной χ^2_{10}, имеет вид (3,25; 20,48). Хи-квадрат распределение позволяет строить доверительный интервал неизвестной дисперсии нормально распределенной генеральной совокупности в случае, если неизвестно её математическое ожидание.

Действительно, пусть дана нормально распределенная генеральная совокупность, для которой неизвестны ни μ, ни σ^2, и надо построить доверительный интервал ее дисперсии с доверительной вероятностью γ.

Предполагая использовать выборку, состоящую из n элементов, найдем граничные точки искомого интервала χ^2- распределения с $n-1$ степенью свободы: нижнюю (левую) ${}_{1-\alpha/2}\chi^2_{n-1}$ и верхнюю (правую) ${}_{\alpha/2}\chi^2_{n-1}$. Каждая из них отсекает в «хвостах» распределения области, вероятность попадания в каждую из которых равна $\alpha/2$. Вместе они образуют интервал, вероятность попадания в который точки χ^2_{n-1} равна $\gamma = 1-\alpha$. Предположим теперь, что случайная величина $\chi^2_{n-1} == \dfrac{(n-1)\cdot S_n^2}{\sigma^2}$ принадлежит этому интервалу, т.е. удовлетворяет двойному неравенству

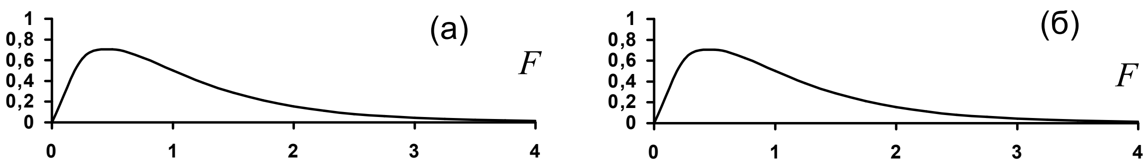
$$ {}_{1-\alpha/2}\chi^2_{n-1} < \frac{(n-1)\cdot S^2}{\sigma^2} < {}_{\alpha/2}\chi^2_{n-1}, \text{ откуда } \frac{{}_{1-\alpha/2}\chi^2_{n-1}\cdot\sigma^2}{n-1} < S^2 < \frac{{}_{\alpha/2}\chi^2_{n-1}\cdot\sigma^2}{n-1}. \quad (9.4)$$

Выразив из каждого неравенства σ^2, и снова объединив неравенства, аналогично тому, как это было сделано в пункте 8.3, получим двойное неравенство $\dfrac{(n-1)\cdot S^2}{{}_{\alpha/2}\chi^2_{n-1}} < \sigma^2 < \dfrac{(n-1)\cdot S^2}{{}_{1-\alpha/2}\chi^2_{n-1}}$, откуда следует, что с надежностью

$$\gamma = 1-\alpha,\ \sigma^2 \text{ будет принадлежать интервалу } \left(\frac{(n-1)S^2}{{}_{\alpha/2}\chi^2_{n-1}}; \frac{(n-1)S^2}{{}_{1-\alpha/2}\chi^2_{n-1}}\right). \quad (9.5)$$

Предположим, что требовалось построить доверительный интервал для дисперсии нормально распределенной генеральной совокупности на уровне статистической значимости $\alpha = 0{,}05$. С этой целью была построена случайная выборка объема $n = 51$ и подсчитана выборочная дисперсия S^2, которая оказалась равной 2. Для построения требуемого доверительного интервала найдем нижнюю критическую точку ${}_{0,975}\chi^2_{50} = 32.36$ и верхнюю критическую точку ${}_{0,025}\chi^2_{50} = 71.42$, откуда, пользуясь выведенными выше неравенствами, получим: $\frac{2 \cdot 50}{71.42} < \sigma^2 < \frac{2 \cdot 50}{32.36}$ или $1.40 < \sigma^2 < 3.09$, и, следовательно, доверительный интервал, оценивающий σ^2, имеет вид (1,40; 3,09).

9.4. F - распределение Фишера. Снова обратимся к стандартному нормальному распределению. Извлечём из него случайным образом два набора значений, которые обозначим $y_1, y_2, \cdots, y_{n+1}$ и $z_1, z_2, \cdots, z_{k+1}$, и составим две случайные величины χ^2_n и χ^2_k. Пользуясь ими, составим случайную величину, называемую *F-отношением* с n и k степенями свободы, равную $\frac{\chi^2_n / n}{\chi^2_k / k}$.



Кривая распределения для $F_{4,4}$ Кривая распределения для $F_{4,25}$

Рис. 9.3

При многократном повторении процедуры случайного выбора $n+1$ и $k+1$ значений нормально распределённой величины и вычисления F получим полигон распределения этой случайной величины. Если его сгладить и подобрать на оси ординат масштаб, при котором площадь под кривой будет равна единице, то получим график распределения случайной величины, называемой *распределением Фишера с* n *и* k *степенями свободы* и обозначаемой $F_{n,k}$.

На рисунке 9.3 изображены кривые распределения для $F_{4,4}$ и $F_{4,25}$.

F - распределение Фишера обладает свойствами:

1) оно унимодально;
2) имеет медиану, равную 1 или меньше;
3) среднее арифметическое равно $k/(k-2)$ для $k \geq 3$.

Правые (верхние) доверительные границы F-распределения Фишера приведены в таблицах D₁ (для $\alpha = 0{,}05$), D₂ (для $\alpha = 0{,}01$) Приложения. Как правило, исследователей интересуют обе границы доверительного интервала. В этом случае, если доверительный интервал соответствует надежности $\gamma = 1-\alpha$, то для нахождения верхней границы интервала надо пользоваться таблицей, соответствующей $\alpha/2$. Например, если надо построить интервал, соответствующий доверительной вероятности 90%, то для построения его верхней границы надо использовать таблицу, соответствующую $\alpha = 0{,}05$. Нижнюю границу вычисляют по правилу:

$$_{1-\alpha}F_{n,k} = \frac{1}{_{\alpha}F_{k,n}}. \tag{11.5}$$

На рисунке 9.4 приведён доверительный интервал F-критерия Фишера, соответствующий доверительной вероятности $\gamma = 0{,}95$, при равномерном распределении критических областей по «хвостам» распределения. Верхняя граничная точка области равна $_{0{,}025}F_{4,10} = 3{,}50$. Нижняя определяется по формуле $_{\alpha}F_{n,k} = \dfrac{1}{_{1-\alpha}F_{k,n}}$ и, следовательно, равна $_{0{,}975}F_{4,10} = \dfrac{1}{_{0{,}025}F_{10,4}} = \dfrac{1}{6} = 0{,}17$.

Распределение Фишера используется, в частности, для проверки утверждения о том, что две независимые случайные выборки извлечены из одной и той же генеральной совокупности. Для этого надо обосновать равенство $\sigma_1^2 = \sigma_2^2$. С этой целью, пользуясь двумя выборками, вычисляют значение $\widetilde{F} = \dfrac{S_1^2}{S_2^2}$ и проверяют, попадает ли вычисленное значение в соответствующий доверительный интервал. Более подробно об этом в пункте 11.5.

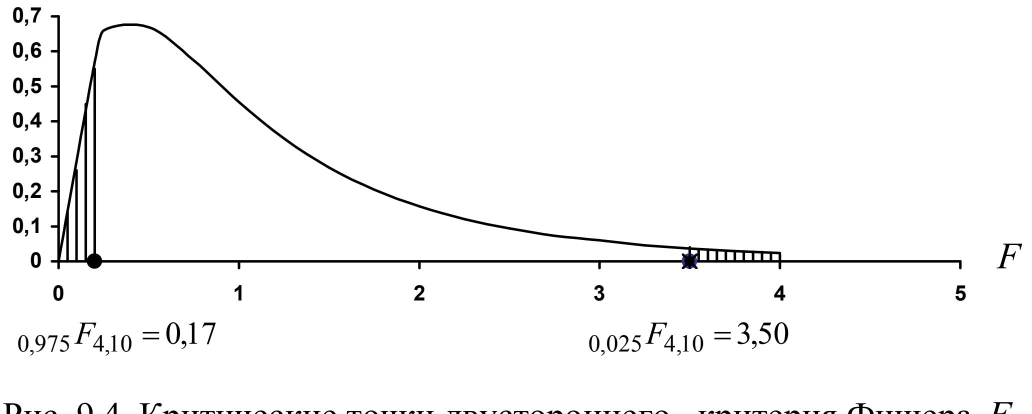


Рис. 9.4. Критические точки двустороннего критерия Фишера $F_{4,10}$

Глава 10. Метод статистических гипотез

10.1. Понятие статистической гипотезы и ее проверки. Теперь мы можем рассмотреть ещё один весьма распространенный метод статистического вывода, получивший название *«Проверка статистических гипотез»*. В его основе, как и при интервальном оценивании, лежат такие понятия, как случайная выборка, выборочное распределение статистики и значения вероятностей, соответствующих гипотезам. Более того, интервальному оцениванию принадлежит ведущая роль при проверке гипотез. Эти два метода оценивания реализуются различными путями, но обычно они приводят к идентичным результатам или результатам, допускающим несложное преобразование одного в другое. Основной по-прежнему остаётся задача: «Какие выводы можно сделать о свойствах генеральной совокупности по выборочным наблюдениям?».

Статистическими гипотезами называют высказывания, касающиеся статистических проблем, например, гипотезу о том, что исследуемая генеральная совокупность распределена по нормальному закону, или что математические ожидания двух нормально распределенных генеральных совокупностей различны и т.п.

Проверяемую гипотезу принято называть *нуль-гипотезой* и обозначать H_0. Наряду с ней каждый раз рассматривается конкурирующая (альтернативная) гипотеза H_1, которая является отрицанием нулевой. Например, если H_0: $\mu_1 = \mu_2$, то H_1: $\mu_1 \neq \mu_2$, или если H_0: $\mu > 100$, то H_1: $\mu \leq 100$, и т.д.

Гипотезы принято подразделять на простые и сложные.

Простой называют гипотезу, содержащую только одно предположение. Например, простой является гипотеза $H_0 : \mu_1 = \mu_2$. Простой будет и гипотеза $H_0 : \sigma^2 = 3$.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез. Например, сложной является гипотеза H_0: $\mu > 0$, состоящая из бесчисленного множества простых гипотез $\mu = b$, где b любое положительное число.

Выдвинутая нуль-гипотеза может быть верной или неверной, поэтому возникает необходимость её проверки. Она в некотором смысле напоминает принятое в математике доказательство от противного. Однако в математической теории предложение отвергается, если оно само или некоторое из его следствий противоречит принятой в этой теории системе аксиом, в статистике же гипотеза отвергается, если она сама или ее следствия противоречат результатам статистического эксперимента. При этом в отличие от математики, в которой предложение отвергается абсолютно, без всяких оговорок, в статистике гипотеза отвергается с определенным уровнем надежности. Саму такую проверку называют статистической.

Механизм проверки статистической гипотезы проиллюстрируем на конкретных примерах.

Пример 1. Пусть имеется некоторая нормально распределенная случайная величина с дисперсией σ^2, и нужно, пользуясь выборочным методом с надежностью γ, выяснить, равно ли ее математическое ожидание μ заданному числу a. В качестве нулевой гипотезы выдвинем предположение H$_0$: $\mu = a$ и выясним, нет ли оснований для её отклонения и принятия альтернативной гипотезы H$_1$: $\mu \neq a$. Проверку предполагается проводить с помощью одной выборки, содержащей n элементов.

Если нуль-гипотеза верна, т.е. $\mu = a$, то, как было показано в пункте 8.3, существует доверительный интервал $\left(a -_{\alpha/2} z \dfrac{\sigma}{\sqrt{n}};\ a +_{\alpha/2} z \dfrac{\sigma}{\sqrt{n}}\right)$, в который с надежностью $\gamma = 1 - \alpha$ должно попасть $\bar{x}$ любой случайной выборки объема n, извлеченной из данной генеральной совокупности. Если же в результате эксперимента окажется, что $\bar{x}$ не попадает в этот интервал, то нулевую гипотезу отвергают с надежностью γ.

При таком применении доверительного интервала его границы называют критическими точками и обозначают ${}_{1-\alpha/2}X = a - z_\gamma \dfrac{\sigma}{\sqrt{n}}$ и ${}_{\alpha/2}X = a + z_\gamma \dfrac{\sigma}{\sqrt{n}}$, где $\alpha = 1 - \gamma$ – вероятность того, что значение $\bar{x}$ попадет в область, лежащую вне выделенного интервала. В этом случае говорят, что нулевая гипотеза отклоняется на уровне значимости α.

Если все значения рассматриваемой случайной величины X подвергнуть стандартизирующему преобразованию $z = \dfrac{x - a}{\sigma / \sqrt{n}}$ то данное в условии задачи нормальное распределение величины X превратится в стандартное нормальное распределение величины Z, а рассмотренный выше интервал – в интервал $\left(-_{\alpha/2}\, z;\ {}_{\alpha/2}z\right)$.

Если теперь построить случайную выборку объема n, а вычисленное для нее $\bar{x}$ преобразовать в $z = \dfrac{\bar{x} - a}{\sigma / \sqrt{n}}$, то принятие или непринятие нуль-гипотезы будет зависеть от того, попадает или не попадает точка z в интервал $\left(-_{\alpha/2}\, z;\ _{\alpha/2}z\right)$.

Если точка z не попадет в интервал $\left(-_{\alpha/2}\, z;\ _{\alpha/2}z\right)$, то, как и прежде, нулевую гипотезу отвергают с надежностью γ, или, как чаще говорят, на уровне статистической значимости α.

Пример 2. Пусть имеется некоторая нормально распределенная совокупность, для которой неизвестны ни дисперсия, ни математическое ожидание и нужно, пользуясь выборочным методом с надежностью γ, выяснить, равна ли ее дисперсия заданному положительному числу a^2. В качестве нулевой гипотезы выдвинем предположение H$_0$: $\sigma^2 = a^2$ и выясним, нет ли оснований для её отклонения и принятия альтернативной гипотезы: H$_1$: $\sigma^2 \neq a^2$. Проверку предполагается проводить с помощью выборочной дисперсии S^2, вычисленной по одной случайной выборке, содержащей n элементов.

Как было показано в пункте 9.3, для любой выборочной дисперсии, вычисленной по случайной выборке объема n, на уровне значимости α будет выполняться неравенство $_{1-\alpha/2}\chi^2_{n-1} < \chi^2_{n-1} < {}_{\alpha/2}\chi^2_{n-1}$, где $\chi^2_{n-1} = \dfrac{(n-1)\cdot S_n^2}{a^2}$. Если верна гипотеза H$_0$: $\sigma^2 = a^2$, то для любой выборочной дисперсии статистика $\chi^2_{n-1} = \dfrac{(n-1)\cdot S_n^2}{a^2}$ на том же уровне значимости должна удовлетворять двойному неравенству $_{1-\alpha/2}\chi^2_{n-1} < \dfrac{(n-1)\cdot S^2}{a^2} < {}_{\alpha/2}\chi^2_{n-1}$. Теперь остается провести эксперимент: выбрать из генеральной совокупности случайным образом n элементов и вычислить S^2. Если при подстановке S^2 хотя бы одно из этих неравенств окажется неверным, то нулевую гипотезу придется отклонить и принять альтернативную, в противном случае у исследователя не будет оснований для отклонения нулевой гипотезы.

10.2. Статистический критерий. Метод, который для каждой выборки позволяет определить, удовлетворяет она нуль-гипотезе или нет, называется *статистическим критерием.* Большинство статистических критериев связано с вычислением по выборке числа (статистики), в соответствии, со значением которого и принимается решение о принятии нуль-гипотезы или её отклонении. Обычно в качестве статистического критерия выбирают такую статистику, выборочное распределение которой хорошо изучено и табулировано.

Так, например, при проверке нулевой гипотезы о равенстве математического ожидания μ заданному числу a в качестве статистического критерия при известном σ принимается статистика $z = \dfrac{\bar{x} - a}{\sigma / \sqrt{n}}$, выборочным распределением которого является табулированное стандартное нормальное распределение.

В рассмотренном выше примере при проверке нулевой гипотезы $H_0 : \mu = a$ в качестве такой статистики, оценивающей параметр μ, была принята выборочная средняя $\bar{x}$, а в качестве статистического критерия при известном σ – статистика $z = \dfrac{\bar{x} - a}{\sigma / \sqrt{n}}$, выборочным распределением которого является стандартное нормальное распределение. Если бы σ было неизвестно, то в качестве статистического критерия пришлось бы воспользоваться статистикой $t = \dfrac{\bar{x} - a}{S / \sqrt{n}}$, выборочным распределением которого является t - распределение Стьюдента.

Аналогично, как мы видели, проверяется нулевая гипотеза о равенстве дисперсии заданному положительному числу a^2: строится случайная выборка объема n, по которой определяется выборочная дисперсия S_n^2 и вычисляется статистика $\chi^2 = \dfrac{(n-1)S^2}{a^2}$. Если значение χ^2 будет принадлежать интервалу $({}_{1-\alpha/2}\chi^2_{n-1},\ {}_{\alpha/2}\chi^2_{n-1})$, то нет оснований отвергать нулевую гипотезу. В противном случае нуль-гипотеза отвергается на уровне α.

В общем случае процедуру проверки гипотезы о параметре можно представить следующим образом. После того как проверяемая гипотеза сформулирована, определяется статистика, позволяющая оценить параметр, о котором идет речь в гипотезе. Затем выбранная статистика преобразуется в некоторую статистику, распределение которой хорошо известно и табулировано, например, стандартное нормальное распределение, распределение Стьюдента, хи-квадрат распределение, или распределение Фишера. В ряде случаев, говоря о *статистическом критерии*, имеют в виду такое выбранное теоретическое распределение. При этом стандартное нормальное распределение $(\mu = 0,\ \sigma = 1)$ называют Z - критерием, распределение Стьюдента – t - критерием, распределение хи-квадрат распределение – χ^2 - критерием, распределение Фишера – F - критерием Фишера.

После выбора определенного критерия K и задания статистической значимости α множество всех значений выбранного критерия разбивается на два непересекающихся подмножества. Одно из них содержит значения критерия, при которых нулевая гипотеза отвергается (эту область называют критической), другое – значения, при которых она принимается (эту область называют *областью принятия гипотезы*). Точки, отделяющие критическую область от области принятия гипотезы, как уже было сказано выше, называют *критическими точками*.

Существуют различные способы обозначения критических точек. Обратимся к тому, который мы ввели и уже использовали выше. Рассмотрим общий случай. Предположим, что на рисунке 10.1 изображён некоторый критерий (выборочное распределение), которое мы пока обозначим буквой K. Символом ${}_{\alpha}K$ обозначим точку оси X, обладающую тем свойством, что вероятность принятия критерием K значения, большего ${}_{\alpha}K$, равна α.

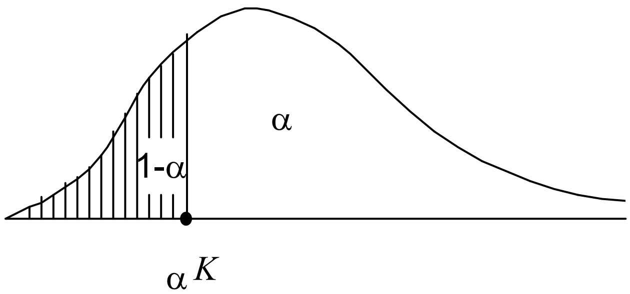


Рис. 10.1. Критическая точка ${}_{\alpha}K$

В каждом конкретном случае критическую точку обозначают буквой, обозначающей используемый критерий с впереди и внизу стоящим индексом, который указывает на вероятность принятия критерием значения (точки), лежащей правее этой критической точки. Например, символ ${}_{0,10}Z$ означает критическую точку Z - критерия, обладающую тем свойством, что вероятность того, что статистика $z = \dfrac{\bar{x} - \mu}{\sigma}$ примет значение, лежащее правее критической точки, будет равна 0,10. Пользуясь таблицей А Приложения легко найти, что ${}_{0,10}Z = 1{,}282$.

Символ ${}_{0.95}\chi^2_{20}$ означает критическую точку χ^2-распределения с 20-ю степенями свободы, обладающую тем свойством, что вероятность того, что критерий $\chi^2 = \dfrac{(n-1)S^2}{a^2}$ примет значение, лежащее правее этой критической точки, равна 0,95. Пользуясь таблицей С Приложения, легко найти, что ${}_{0.95}\chi^2_{20} = 10{,}85$.

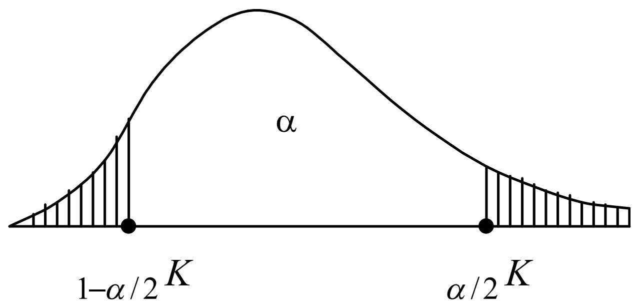


Рис.10.2. Критические точки двустороннего критерия K

Обычно критическая область располагается в «хвостах» критерия и отделяется от области принятия гипотезы двумя критическими точками.

Если уровень значимости проверки гипотезы равен α, то, распределяя возможность ошибки поровну на две критические области, получим две критические точки $_{1-\alpha/2}K$ и $_{\alpha/2}K$. Нап-ример, при статистической значимости $\alpha = 0{,}05$, критические точки можно обозначить $_{0{,}975}K$ и $_{0{,}025}K$.

Преодоление нейтральной полосы между эмпирическим распределением определенного признака и соответствующей гипотетической моделью играет в статистике большую роль. Неотклонённая нуль-гипотеза принимается в качестве *рабочей гипотезы,* так как она может быть правильной и не противоречит материалу наблюдений. Более важно, чем правильность нуль-гипотезы, то, что нет достаточного статистического материала для ее отклонения. Если материал дополнен, то получается новая перепроверка нуль-гипотезы. Она будет приниматься до тех пор, пока новые данные не сделают ее неприемлемой.

Чем больше известно о свойствах генеральной совокупности на основе метода максимального правдоподобия или хотя бы на основе грубых оценок из предыдущих опытов, тем точнее будет вероятностная модель и тем точнее будут результаты оценивания и проверки гипотез.

Для научного метода весьма важно объединение индуктивного и дедуктивного процессов; индукция занимается созданием моделей на основе эмпирических наблюдений, их проверкой и улучшением. Задачей дедукции являются выбор лучших способов вычисления оценок параметров генеральной совокупности и определение статистических распределений этих оценок для случайных выборок.

10.3. Односторонние и двусторонние критерии. Различают одностороннюю (правостороннюю или левостороннюю) и двустороннюю критические области. Зависит это прежде всего от строения альтернативной гипотезы, которая может быть либо *ненаправленной,* либо *направленной.* Альтернативная гипотеза H_1: $\mu \neq 0$, противостоящая нулевой гипотезе H_0: $\mu = 0$ является ненаправленной, поскольку она утверждает только факт неравенства параметра μ нулю и не указывает, в каком направлении (больше или меньше нуля параметр μ). Ситуация существенно меняется, если исследователь в качестве гипотезы альтернативной гипотезе H_0: $\mu = 0$ выдвинет гипотезу H_1: $\mu > 0$. В этом случае альтернативная гипотеза окажется направленной: исследователь как бы утверждает, что μ или равно нулю, или, возможно, больше нуля, но, во всяком случае, не меньше. Поэтому он будет собирать сведения, которые либо подтвердят, что μ больше нуля, либо он по-прежнему будет считать, что $\mu = 0$.

Одно из следствий выдвижения направленной альтернативной гипотезы H_1: $\mu > 0$ состоит в том, что теперь только большие значения $\bar{x}$ будут вызывать необходимость принять решение в пользу H_1, а не H_0: Следовательно, критическая область отклонения H_0 будет лежать в правом конце эмпирического распределения $\bar{x}$ и на уровне статистической значимости α отделяться критической точкой ${}_{\alpha}z$, а не ${}_{\alpha/2}z$, как это было бы при двустороннем Критерии, и будет иметь вид, изображенный на рисунке 10.3.

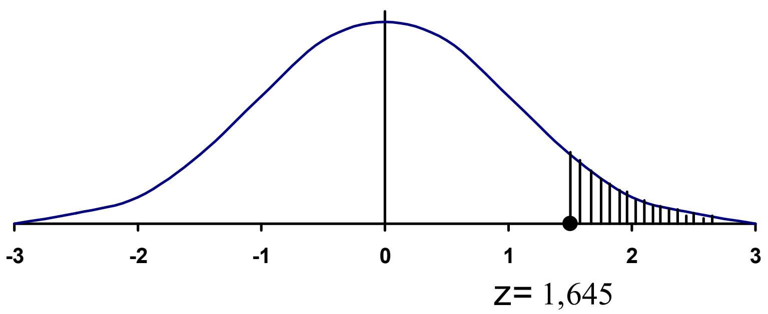


Рис. 10.3. Критическая односторонняя область для проверки гипотезы H_0:μ=0 на уровне статистической значимости α

Критическая односторонняя область для проверки гипотезы $\mu = 0$ против $\mu > 0$ при $n = 200$ и $\alpha = 0.05$. Критическая область в этом случае будет простираться вправо от точки 1,645, которую можно найти, пользуясь таблицей А Приложения.

Так как условиями рассматриваемой гипотезы предусматривается только два соотношения (либо $\mu = 0$, либо $\mu > 0$), то μ, равное минус 0,80, подразумевало бы скорее истинность H_0, а не H_1, даже если бы его появление в выборке $n = 200$ было бы крайне невероятным при $\mu = 0$. Но подобные факты – своеобразие проверок направленных гипотез.

Если цель опыта состоит в том, чтобы установить различие математических ожиданий двух генеральных совокупностей и в качестве нулевой принята гипотеза $H_0 : \mu_1 = \mu_2$, то тип критерия будет зависеть от альтернативной гипотезы. Если у исследователя нет никаких соображений, позволяющих исключить одно из неравенств ($\mu_1 < \mu_2$ или $\mu_1 > \mu_2$), то он воспользуется ненаправленным критерием $H_0 : \mu_1 = \mu_2$ против $H_1 : \mu_1 \neq \mu_2$.

Однако иногда исследователь имеет возможность высказать определенное предположение о знаке ожидаемого различия $\mu_1 < \mu_2$ или, наоборот: $\mu_1 > \mu_2$. В обоих случаях мы должны те отклонения, которые не фиксирует альтернативная гипотеза, отнести к нуль-гипотезе. Если альтернативная гипотеза гласит $\mu_1 > \mu_2$, то соответствующая нуль-гипотеза $\mu_1 \leq \mu_2$.

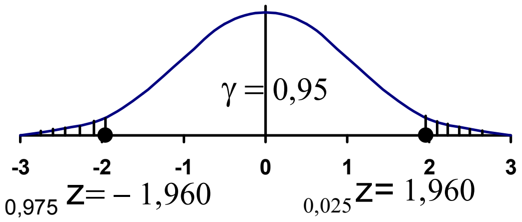


Рис. 10.4. Критические точки двустороннего z-критерия

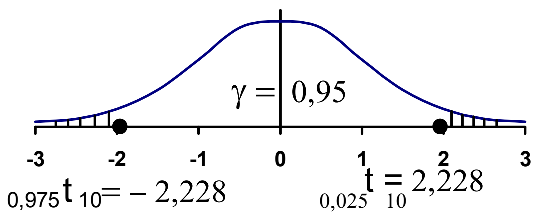


Рис. 10.5. Критические точки двустороннего t-критерия

Рассмотрим для примера двусторонний Z-критерий и уровень значимости $\alpha = 0{,}05$. Значения критических точек ${}_{0,975}Z = -1{,}960$ и ${}_{0,025}Z = 1{,}960$ (рис. 10.4) можно найти, пользуясь таблицей А Приложения.

Аналогично обозначаются критические точки t-критерия Стьюдента при ν степенях свободы (рис. 10.5). Их значения находят по таблице В Приложения. Например, при $\nu = 10$ для $\alpha = 0{,}05$ при двустороннем критерии ${}_{0.975}t_{10} = -2{,}228$ и ${}_{0.025}t_{10} = 2{,}228$.

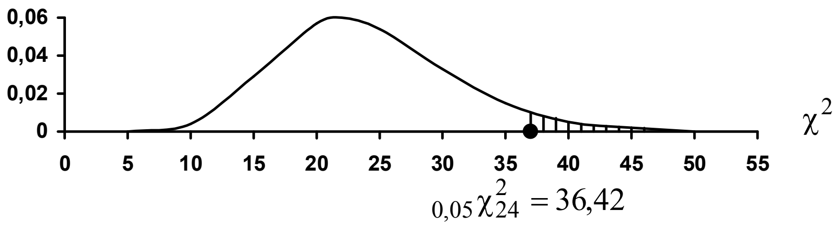


Рис. 10.6. Критическая область одностороннего χ^2_{24} распределения

Рассмотрим теперь χ^2-критерий. При одностороннем критерии критическую точку обозначают ${}_{\alpha}\chi^2_{\nu}$, где ν – число степеней свободы, а α – уровень значимости. Значения этих критических точек для уровней значимости α, равных 0,99; 0,975; 0,95; 0,05; 0,025; 0,01, и значений ν от 1 до

200 приведены в таблице С Приложения. На рисунке 10.6 изображены χ^2_{24} - критерий и критическая точка $_{0,05}\chi^2_{24} = 36{,}42$.

10.4. Ошибки первого и второго рода. Мощность критерия. В итоге статистической проверки гипотезы в двух случаях может быть принято неправильное решение, т.е. могут быть допущены ошибки двух родов.

Ошибка первого рода состоит в том, что будет отвергнута правильная нуль-гипотеза.

Ошибка второго рода состоит в том, что будет принята неправильная нуль-гипотеза.

Вероятность совершить ошибку первого рода (отвергнуть правильную гипотезу) равна *уровню значимости* проверки гипотезы α. Если, например, уровень значимости α принят равным 0,05, то это означает, что в пяти случаях из ста мы рискуем допустить ошибку первого рода (отвергнуть правильную нуль-гипотезу).

Вероятность совершить ошибку второго рода (принять неправильную нуль-гипотезу) принято обозначать буквой β.

Если, например, в результате статистической проверки принято, что новый медикамент лучше, хотя на самом деле он идентичен старому, то это ошибка 1-го рода; если из сравнения вытекает, что оба медикамента одинаковы, хотя на самом деле новый лучше, то имеет место ошибка 2-го рода.

При заданном α и объеме выборки, равном n, значение β, т.е. вероятность совершить ошибку 2-го рода (принять неверную гипотезу), будет тем больше, чем меньше принятое α. Принимать решения при очень малых значениях α и β можно только при очень большом объеме выборки. При малом объеме выборки и малом α вероятность получить значение критерия, попадающее в критическую область, мала, следовательно, мала вероятность зафиксировать в результате статистического эксперимента фактически существующее различие.

Часто к этому условию добавляют еще одно, связанное с понятием мощности критерия. При этом под *мощностью критерия* понимают вероятность попадания критерия в критическую область при условии, что справедлива альтернативная гипотеза. Другими словами, мощность критерия есть вероятность того, что нулевая гипотеза будет отвергнута, если верна конкурирующая гипотеза.

Нетрудно показать, что если вероятность ошибки второго рода равна β, то мощность критерия равна $1-\beta$. Действительно, так как высказывание «принята нулевая гипотеза, причем справедлива конкурирующая» противоположно высказыванию «отвергнута нулевая гипотеза, причем справедлива конкурирующая», то сумма их вероятностей равна единице. Но вероятность первого высказывания, будучи вероятностью ошибки второго рода, равна β. Вероятность же второго высказывания есть мощность критерия, откуда следует, что она равна $1-\beta$.

Пусть мощность $1-\beta$ возрастает; следовательно, уменьшается вероятность β совершить ошибку второго рода. Таким образом, чем *больше* мощность критерия, тем *меньше* вероятность ошибки второго рода.

Итак, если уровень значимости уже выбран, то критическую область следует строить так, чтобы мощность критерия была максимальной.

Ясно, что чем меньше вероятности ошибок первого и второго рода, тем критическая область лучше. Однако при заданном объеме выборки, уменьшить одновременно α и β невозможно: если уменьшать α, то β будет возрастать. Единственный способ одновременного уменьшения вероятностей ошибок первого и второго рода состоит в увеличении объема выборок.

Как же выбрать α наиболее целесообразно? Ответ на этот вопрос зависит от «тяжести последствий» ошибок для каждой конкретной задачи. Например, если ошибка первого рода повлечет большие потери, а второго рода – малые, то следует принять по возможности меньшее значение α.

Так, например, при изготовлении вакцины требуется предельная константа сыворотки. Небезупречные измерения должны быть обнаружены и исключены. Необоснованное принятие нуль-гипотезы «сыворотка в норме» означает опасную ошибку. При этом желательно и β также выбрать возможно меньшее, в то время как отбрасывание хороших результатов не принесет ничего, кроме дополнительных расходов, и в остальном не будет иметь никаких серьезных последствий (т. е. примерно $\alpha = 0{,}10$).

Несмотря на это обстоятельство, исследователи нередко тщательно выбирают только значение α, не обращая внимания на вероятность принятия ошибочной нуль-гипотезы. Такие критерии, заведомо благоприятные для нуль-гипотезы, часто называют консервативными критериями.

Учитывая сложности использования мощности критерия, чаще всего, ссылаясь на правило Неймана, заранее задаются значением α, после чего пытаются сделать β по возможности меньшим, что достигается, прежде всего увеличением объема выборки. Если же объем выборки мал и увеличить его невозможно, то нужно брать уровень значимости не слишком малым, так как и малая выборка, и малый уровень значимости приводят к нежелательному уменьшению мощности критерия. Заметим также, что односторонний критерий имеет большую мощность, чем двусторонний.

Глава 11. Параметрические критерии

11.1 Основные виды статистических критериев. Существует несколько видов статистических критериев. Исторически сложилось так, что первые классические задачи математической статистики были связаны с проверкой гипотез о параметрах, причём решались они в предположении, что исследуемые случайные величины имеют нормальное распределение. Теперь такие задачи и используемые для их решения критерии стали называть *параметрическими*. К ним принадлежат все рассмотренные в предыдущей главе критерии.

Критерии, проверяющие, согласуются ли наблюдаемые распределения с гипотетическими, называют *критериями согласия.* С ними мы познакомимся в главе 14. Критерии согласия относятся к так называемым непараметрическим критериям.

Непараметрическими критериями называются критерии, не использующие предположения о характере рассматриваемого распределения. Непараметрические методы математической статистики – методы непосредственной оценки и проверки гипотез о теоретическом распределении вероятностей и тех или иных его общих свойствах (симметрии, независимости и т.п.) по результатам наблюдений. Название «непараметрические методы» подчеркивает их отличие от классических (параметрических) методов, в которых предполагается, что неизвестное теоретическое распределение принадлежит какому-либо семейству, зависящему от конечного числа параметров (например, семейству нормальных распределений), и которые позволяют по результатам наблюдений оценивать неизвестные значения этих параметров и проверять те или иные гипотезы относительно их значений. Особенность непараметрических методов, в отличие от классических, состоит в независимости от неизвестного теоретического распределения. Очевидно, что оптимальные критерии должны быть нечувствительны к отклонениям от сделанных предположений (например, о нормальности распределения), но чувствительны по отношению к контролируемым отклонениям от нуль-гипотезы.

11.2. Процедура применения параметрических критериев. Приступая к статистическому анализу ситуации, целесообразно сформулировать все те гипотезы, которые существенны, доступны для вас и могут быть проверены с помощью соответствующих критериев.

Полученный числовой материал должен быть подвергнут тщательному анализу, но он не должен служить основанием для выдвижения других гипотез, проверяемых на нем же. Такие гипотезы (из статистического материала) должны выдвигаться с весьма большой осторожностью, и их следует дополнительно проверять, поскольку каждая группа чисел имеет свой случайный экстремум. Риск ошибки в этом случае больше, чем тогда, когда гипотеза выдвигается заранее. Гипотезы, полученные из статистического материала, могут быть полезными в качестве новых гипотез для последующих проверок!

Проанализировав проблему и связанные с ней случайные величины, способы их измерения, шкалы, проверив их при необходимости на нормальность, сформулировав основные статистические гипотезы, надо решить, какими критериями пользоваться (параметрическими или непараметрическими).

Саму проверку статистических гипотез обычно осуществляют в следующей последовательности:

а) формулируют нулевую и альтернативную ей гипотезу (ненаправленную или направленную);

б) высказывают предположения, необходимые для определения выборочного распределения статистики, оценивающей параметр, относительно которого сформулирована гипотеза;

в) определяют статистику, с помощью которой предполагается проверять нулевую гипотезу;

г) определяют вид выборочного распределения статистики при условии верности нулевой гипотезы и тип критерия. При необходимости, если предполагается использовать параметрический критерий, обосновывается нормальность распределения;

д) принимают уровень значимости α и объём выборки n;

е) находят критические точки критерия;

ж) из генеральной совокупности случайным образом извлекают выборку объёма n;

з) вычисляют значение выбранного критерия, а также необходимых сопутствующих статистик;

и) полученное значение критерия сопоставляют с его критическими значениями и принимают решение относительно истинности нулевой гипотезы.

В заключение заметим, что по-прежнему значения статистических критериев, вычисленные на основании выборочных значений (либо по выборке и одному или нескольким параметрам) для лучшего отличия от соответствующих табличных значений (например, t, z, χ^2 или F – распределения), будем иногда отмечать «дугой ($\frown$)» ($\widehat{t}$, $\widehat{z}$, $\widehat{\chi}$).

Рассмотрим теперь два примера принятия решения с помощью параметрических критериев.

Пример 1. Проверяемая гипотеза (μ = a) состоит в том, что математическое ожидание μ генеральной совокупности равно некоторому действительному числу a, Пусть, например, высказано предположение, что средний балл, который ученики района смогут набрать при выполнении ЕГЭ, равен 60. Высказанный прогноз предполагается проверить на уровне статистической значимости $\alpha = 0{,}05$. Отобрав случайным образом 50 одиннадцатиклассников района ($n = 50$), проводят среди них контрольную работу, идентичную ЕГЭ. Затем, убедившись в том, что распределение школьников по числу набранных баллов близко нормальному, вычисляют средний балл и дисперсию выборки. Предположим, что средний выборочный балл $\bar{x}$ оказался равным 55, а дисперсия S^2=225 ($S = 15$). Так как дисперсия генеральной совокупности неизвестна, и пользоваться приходится выборочной дисперсией, то в качестве статистического критерия воспользуемся t- критерием Стьюдента (см. 9.1) $t = \dfrac{\bar{x} - \mu}{S / \sqrt{n}}$.

Подставив вычисленные значения $x = 55$, $S = 15$ и предполагаемое значение μ, равное 60, получим $\hat{t} = \dfrac{55 - 60}{15 / \sqrt{50}} = -2{,}357$.

Задавшись уровнем значимости $\alpha = 0{,}05$ и приняв объем выборки равным 50, найдём критические точки двухстороннего критерия Стьюдента: ${}_{0{,}975}t_{49} = -2{,}011$ и ${}_{0{,}025}t_{49} = 2{,}011$.

Так как $\tilde{t}$ попадает в критическую область, то на уровне значимости $\alpha = 0{,}05$ следует отвергнуть нулевую гипотезу H_0 : $\mu = 60$.

На основании полученных данных можно построить доверительный интервал математического ожидания μ на уровне принятой выше статистической значимости α: $\mu = \bar{x} \pm {}_{0{,}025}t_{49} \cdot \dfrac{S_x}{\sqrt{n}} = 55 \pm 2{,}011 \cdot \dfrac{15}{\sqrt{50}} = 55 \pm 4{,}266$, откуда следует, что доверительный интервал имеет вид (50,7; 59,3).

Замечание. Если бы у исследователя были веские причины считать, что μ не может принимать значения большие 60, то он, сохранив прежнюю нулевую гипотезу $H_0 : \mu = 60$, в качестве альтернативной принял бы гипотезу $H_1 : \mu < 60$. В этом случае критическая область расположилась бы в левом «хвосте» t- распределения с одной критической точкой ${}_{0.05}t_{49}$, равной 1,672. Если вспомнить, что значение $\hat{t} = 2{,}317$, то и в этом случае следует отвергнуть нулевую гипотезу.

Пример 2. Проверяемая гипотеза $(\sigma^2 = a^2)$ состоит в том, что дисперсия равна заданному положительному числу a^2. Пусть, например, многократные проверки разброса годовых оценок пятиклассников при работе по стабильному учебнику показали, что дисперсия системы оценок равна 0,64. Исследователя интересовало, значимо ли (на уровне 0,05) изменилось это значение с переходом на новый учебник. По мнению исследователя, его сокращение при условии сохранения уровня успешности усвоения материала говорило бы в пользу нового учебника.

С этой целью после года работы по новому учебнику была сформирована случайная выборка из 125 человек и подсчитана дисперсия S_x^2, которая оказалась равной 0,54. Какой вывод отсюда можно сделать? Действительно ли с переходом на новый учебник значимо изменилась дисперсия.

Рассмотрим гипотезу $H_0 : \sigma^2 = 0{,}64$ против альтернативной гипотезы – $H_1 : \sigma^2 \neq 0{,}64$. Так как $n = 125$, а $S^2 = 0{,}54$, то в соответствии с выражением (см. 9.3) $\hat{\chi}^2 = \dfrac{(n-1)S^2}{a^2} = \dfrac{124 \cdot 0{,}54}{0{,}64} = 104{,}6$.

Так как, по замыслу исследователя, используется ненаправленная статистика, то критерий будет двусторонним. Его нижняя граничная точка ${}_{0,975}\chi^2_{124} = 91{,}57$, а верхняя – ${}_{0,025}\chi^2_{124} = 152{,}21$. Между этими значениями находится найденное исследователем значение $\tilde{\chi}^2 = 104{,}6$.

Отсюда следует, что на уровне $\alpha = 0{,}05$ нельзя отвергать нулевую гипотезу, т.е. предположение о том, что в генеральной совокупности дисперсия оценок учащихся, обучающихся по новому учебнику, не отличается от такой же дисперсии при работе по стандартному учебнику.

Замечание. Если бы у исследователя были основания к тому, чтобы исключить возможность увеличения дисперсии, то он мог бы сформулировать направленную альтернативную гипотезу H_1: $\sigma^2 < 0.64$. В этом случае мы имели бы односторонний критерий, критическая область которого располагается в левом «хвосте» распределения и с одной критической точкой ${}_{0,95}\chi^2_{124}$=95,70 .Так как вычисленное выше значение $\hat{\chi}^2$=104,6 не попадает в критическую область, то отклонять и такую направленную гипотезу нет основания.

11.3. Проверка ненаправленной гипотезы $\mu_1 = \mu_2$ по независимым выборкам. Проверяется нуль-гипотеза $\mu_1 = \mu_2$ о равенстве средних значений генеральных совокупностей, лежащих в основе обеих выборок, при неизвестных, но равных дисперсиях.

При сопоставлении двух выборок надо учитывать возможность существования связи между их элементами. Например, обе выборки могут быть получены в результате тестирования одной и той же группы испытуемых в двух разных условиях – в начале и в конце некоторого эксперимента. Связь в этом случае выражается в том, что результаты тестирования каждого испытуемого естественно объединить в пары. В определенных условиях, если, например, изучаются ценностные ориентации школьников, связь может возникнуть, когда в одной группе тестируются девушки, а в другой их братья и т. д. В этом пункте мы рассмотрим случай независимых выборок, например, когда два ряда данных получены тестированием учащихся контрольной и экспериментальной групп.

Проверка ненаправленной гипотезы $\mu_1 = \mu_2$ по *независимым* выборкам осуществляется в следующем порядке:

а) формулируется проверяемая гипотеза, которая состоит в том, что разность между средними значениями μ_1 и μ_2 двух совокупностей равна нулю, против альтернативной гипотезы о том, что эта разность отлична от нуля: H_0: $\mu_1 - \mu_2 = 0$, H_1: $\mu_1 - \mu_2 \neq 0$;

б) обосновывается нормальность распределений случайных величин X_1 и X_2, а также равенство их дисперсий $\left(\sigma^2_{X_1} = \sigma^2_{X_2}\right)$;

в) определяется статистика, с помощью которой проверяется нулевая гипотеза H_0 против H_1. В данном случае такой статистикой является:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{S_{\bar{x}_1 - \bar{x}_2}}, \qquad (11.1)$$

где $$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}; \qquad (11.2)$$

где $\bar{x}_1$ и $\bar{x}_2$ – выборочные средние из X_1 и X_2 соответственно. S_1^2 и S_2^2 – несмещённые дисперсии из тех же выборок с общей генеральной дисперсией σ_x^2;

г) если H_0 верна (т.е. $\mu_1 = \mu_2$), то величина t в выражении (11.1) распределена по закону Стьюдента с $\nu = n_1 + n_2 - 2$ степенями свободы;

д) принимают уровень значимости α и объёмы выборок n_1 и n_2;

е) находят критические точки критерия;

ж) из генеральных совокупностей X_1 и X_2 извлекаются *независимые* выборки объёма n_1 и n_2;

з) вычисляют значение выбранного критерия, а также необходимых сопутствующих статистик;

и) полученное значение критерия сопоставляют с его критическими значениями и принимают решение относительно истинности нулевой гипотезы.

Замечание. Нарушение однородности дисперсии $\left(\sigma^2_{X_1} = \sigma^2_{X_2}\right)$ допускается, если $n_1 = n_2$. В этом случае выражение $S_{\bar{x}_1 - \bar{x}_2}$ (11.2) принимает вид:

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2 + S_2^2}{n}}. \tag{11.3}$$

Пример. Сопоставляя изложение одной и той же темы в двух различных учебниках, учитель, работающий в нескольких параллельных классах, отобрал из них случайным образом две группы по 25 человек в каждой и поручил им самостоятельно проработать эту тему: одной группе по первому учебнику, другой – по второму. В конце эксперимента учащимся был предложен тест на усвоение изученного материала. Результаты оценивались количеством правильных ответов (X).

После проверки выборок на нормальность (равенство дисперсий можно было не проверять, так как объемы выборок одинаковы), были получены следующие данные: в первой выборке $n_1 = 25$, $\bar{x}_1 = 7{,}65$, $S_1^2 = 6{,}50$, во второй $n_2 = 25$, $\bar{x}_2 = 6{,}00$, $S_2^2 = 5{,}90$.

Нулевую гипотезу H_0: $\mu_1 = \mu_2$ проверим на уровне значимости $\alpha = 0{,}05$. Из выражения (11.3) $S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{6{,}50 + 5{,}90}{25}} = 0{,}704$, откуда $\hat{t} = \dfrac{7{,}65 - 6{,}00}{0{,}704} = 2{,}34$. Критические точки t-критерия Стьюдента при $\nu = n_1 + n_2 - 2 = 48$ и $\alpha = 0{,}05$ (двусторонний критерий) равны ${}_{0,975}t_{48} = -2{,}02$ и ${}_{0,025}t_{48} = -2{,}02$. Так как $\hat{t} = 2{,}34$ правее верхней критической точки, то гипотезу $H_0: \mu_1 = \mu_2$ следует отклонить на уровне значимости $\alpha = 0{,}05$.

11.4. Сравнение двух выборочных дисперсий по независимым выборкам. В общем случае проверка двусторонней гипотезы $\sigma_1^2 = \sigma_2^2$ по независимым выборкам проводится в следующем порядке:

а) сопоставляются две совокупности (1 и 2) с неизвестными исследователю дисперсиями σ_1^2 и σ_2^2. Проверяется гипотеза H_0: $\sigma_1^2 = \sigma_2^2$ против гипотезы H_1: $\sigma_1^2 \neq \sigma_2^2$;

б) предполагается, что выборка объёмом n_1, случайным образом извлекается из нормальной совокупности со средним μ_1 и дисперсией σ_1^2; вторая случайная выборка, объёмом n_2, независимая от первой, извлекается из второй нормальной совокупности со средним μ_2 и дисперсией σ_2^2. При проверке H_0 значения μ_1 и μ_2 несущественны;

в) для проверки H_0 против H_1 используется критерий отношения двух выборочных дисперсий:

$$F = S_1^2 \big/ S_2^2 \ , \tag{11.4}$$

где S_1^2 и S_2^2 – несмещённые дисперсии, вычисленные по данным выборкам;

г) когда верна гипотеза H_0: $\sigma_1^2 = \sigma_2^2$, выборочное распределение $F = S_1^2 \big/ S_2^2$ представляет собой F - распределение Фишера с $n_1 - 1$ и $n_2 - 1$ степенями свободы.

Дальнейшие действия по общей схеме.

Пример. Сравниваются два разных изложения одной и той же темы в двух учебниках. Исследователя интересует, какое из них ведёт к меньшей дисперсии в оценках.

С этой целью из состава учащихся трёх восьмых классов случайным образом формируются две группы по 20 человек каждая. Отобранные учащиеся самостоятельно изучают интересующую исследователя тему: первая группа по первому учебнику, вторая – по второму. Качество усвоения тестируется в десятибалльной системе, дисперсии сравниваются. Верхнее критическое значение двустороннего критерия Фишера при $\alpha = 0{,}10$ равно ${}_{0,05}F_{19,19} = 2{,}17$. Нижнее значение находится по правилу

$${}_{0,95}F_{19,19} = \frac{1}{{}_{0,05}F_{19,19}} = 0{,}46\,.$$

В результате эксперимента было установлено, что в первой группе дисперсия равна $S_1^2 = 1{,}97$, во второй – $S_1^2 = 1{,}97$, откуда значение критерия $\hat{F} = S_1^2 / S_2^2$ =1,97/4,34=0,45.

Так как значение $\hat{F} = 0{,}45$ меньше нижнего критического значения ${}_{0,95}F_{19,19} = 0{,}46$, то нуль-гипотезу следует отклонить на уровне статистической значимости α = 0,10. Отсюда следует, что на выбранном уровне значимости можно утверждать, что дисперсии различны.

Судя по тому, что $S_1^2 = 1{,}97$, а $S_1^2 = 1{,}97$, можно предположить, что $\sigma_1^2 < \sigma_2^2$. Для того чтобы в этом убедиться надо сформулировать направленную гипотезу H_0: $\sigma_1^2 = \sigma_2^2$, альтернативная гипотеза H_1: $\sigma_1^2 < \sigma_2^2$.

В заключение заметим, что для выборок одинакового объема, как в нашем примере, нуль-гипотезу $H_0 : \sigma_1^2 = \sigma_2^2$ можно проверить с помощью соотношения

$$\hat{t} = \frac{\sqrt{n-1}\left(S_1^2 - S_2^2\right)}{2\sqrt{S_1^2 S_2^2}}, \quad \nu = n-1. \tag{11.5}$$

Применив этот критерий в нашем примере, получим $\hat{t} = \frac{\sqrt{20-1}\left|,97-4{,}34\right|}{2\sqrt{1{,}97 \cdot 4{,}34}} = \frac{4{,}359 \cdot 2{,}37}{5{,}848}$=1,767. В то же время ${}_{0,05}t_{19} = 1{,}729$.

Так как ${}_{0,05}t_{19} = 1{,}729$<1.767=$\hat{t}$, то на уровне $\alpha = 0.10$ нулевую гипотезу надо отклонить (на уровне $\alpha = 0.05$ это уже не так).

При больших и очень больших объемах выборок ($n_1, n_2 > 100$) можно воспользоваться z -критерием:

$$\hat{z} = \frac{\left|S_1 - S_2\right|}{\sqrt{\frac{S_1^2}{2n_1} + \frac{S_2^2}{2n_2}}}. \tag{11.6}$$

Если вычисленное значение статистики $\hat{z}$ превосходит теоретическое z-значение или равно ему, то стандартные отклонения S_1 и S_2 или соответствующие дисперсии различаются значимо, т.е. они неоднородные; в противном случае они считаются однородными.

Глава 12. Непараметрические критерии

12.1. Понятие непараметрического критерия. Во всех рассмотренных в предыдущей главе примерах, связанных с проверкой статистических гипотез, предполагалось, что исследуемые случайные величины имеют нормальное распределение. Там такие методы мы назвали *параметрическими*. Однако реальные эмпирические распределения никогда не бывают в строгом смысле нормальными, в результате чего исследователей, применявших эти методы при решении прикладных задач, не оставляло ощущение опасности ошибки. Все это предопределило бурное развитие в середине 20 века так называемых *непараметрических* методов, позволяющих проверять статистические гипотезы без знания вида распределения.

Применение классических методов, как правило, накладывает на проверяемую совокупность и другие, часто трудно проверяемые условия. Например, классическое сравнение средних величин «по Стьюденту» основано на следующих предположениях:

1) выборки случайны и независимы;

2) признаки измерены в интервальной шкале или шкале отношений;

3) генеральные совокупности должны быть (по крайней мере приближенно) нормально распределены;

4) дисперсии должны быть равны ($\sigma_1^2 = \sigma_2^2$).

Непараметрические же методы, соответствующие критерию Стьюдента, требуют только независимости данных.

Практически единственное требование, соблюдение которого необходимо для применения непараметрических методов, состоит в том, чтобы *все данные или пары данных были выбраны случайным образом и независимо друг от друга из одной и той же генеральной совокупности.* Соблюдение этого требования должно обеспечиваться постановкой и проведением эксперимента.

Поскольку непараметрический критерий, когда его применяют на нормальном распределении, всегда слабее, чем соответствующий параметрический критерий, вводится понятие коэффициента эффективности:

$$E = \frac{n_1 \text{ для параметрического критерия}}{n_2 \text{ для непараметрического критерия}},$$

где n_1 и n_2 объемы выборок, необходимых для получения заданной мощности параметрического и соответственного непараметрического критерия.

Например, если при использовании непараметрического критерия на определенном уровне значимости в среднем требуется выборка объема $n = 100$, то при использовании в тех же целях и при тех же условиях соответствующего параметрического критерия достаточна выборка объема $n = 85$, то коэффициент эффективности E будет равен 0,85. По этому показателю определяют действенность непараметрического критерия, когда он применяется вместо классического критерия на совокупности нормально распределенных данных.

Непараметрические методы предполагают случайную выборку из генеральной совокупности с непрерывным распределением. Они рекомендуются к применению тогда, когда

а) параметрические методы слишком чувствительны к отклонениям от сделанных допущений, или

б) удовлетворение этим допущениям с помощью соответствующих преобразований или с помощью устранения выбросов представляет значительные трудности.

Непараметрические методы, которые отличаются относительно простыми вычислениями, называют также *быстрыми*.

Другим важным свойством непараметрических критериев, кроме *экономичной вычислительной процедуры*, является *применимость без предварительных допущений.* Их недостаток – *малая мощность*, поскольку только часть информации, содержащаяся в данных, используется для принятия статистического решения.

Статистические решения быстрых тестов *консервативны,* т. к. с их помощью труднее отклонить нуль-гипотезу, чем с помощью параметрических критериев, для этого требуется большая выборка или большее число противоречащих альтернативных данных.

В 1962 году Линерт (Lienert) сформулировал следующие рекомендации к применению быстрых непараметрических критериев.

1. Важнейшим применением непараметрических тестов, по его мнению, является *приближенная* проверка значимости для параметрических и непараметрических методов. С помощью таких методов можно определить, выгодно ли вообще проводить проверку значимости с помощью оптимальных критериев. При использовании быстрого критерия есть три возможности:

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.

2. Другой областью применения непараметрических критериев является суждение о значимости данных, полученных в *предварительных* опытах. Результаты предварительных опытов должны быть хорошо обоснованы, если последующие главные испытания призваны обеспечить надежные выводы.

3. И, наконец, непараметрические критерии могут применяться для получения *обоснованного* вывода там, где имеется достаточно большая выборка ($n > 100$). Эта рекомендация основана на том, что при больших значениях n даже слабый критерий должен указать на наличие значимости, когда результат не только статистически, но и практически должен быть значимым.

Среди трех приведенных возможностей применения быстрых критериев наибольшее значение имеет первая, так как здесь экономический эффект проявляется вдвойне: во-первых, приближенные методы проще, во-вторых, вообще не нужно применять сложных и дорогостоящих критериев.

Определенные трудности в применении параметрических методов статистического оценивания возникают из-за требования, чтобы измерение исследуемых случайных величин производилось в интервальной шкале или в шкале отношений. В связи с этим некоторые авторы считают недопустимым применение параметрических методов к анализу данных, «измеренных» в балльных или экспертных оценках, предлагая переводить их в ранговую шкалу. В реальных условиях требования к построению интервальных шкал и шкал отношений не заходят столь далеко. Например, с успехом применяются параметрические методы для статистического анализа результатов тестовых измерений, особенно если тест обеспечивает равномерное распределение баллов по заданиям и нормальное распределение учащихся по количеству набранных ими баллов. Вполне допустимо использование экспертных оценок (в том числе и оценок, выставляемых преподавателями), особенно, если они являются результатом усреднения нескольких независимых первичных оценок. Во всяком случае, перевод таких оценок в ранги порядковой шкалы и применение непараметрических критериев не обеспечивают повышения надежности сделанных выводов.

12.2. Сравнение двух независимых выборок по Колмогорову и Смирнову. Наиболее распространенным и строгим непараметрическим критерием однородности, является критерий Колмогорова (1933) и Смирнова (1939). Он предназначен для сравнения двух независимых выборок (рядов измерений или значений частот) и ответа на вопрос, относятся ли они к одной и той же генеральной совокупности. Он включает в себя проверку всех видов различия распределений, в особенности различия средних положений (среднее значение, медиана), рассеяния, асимметрии и эксцесса, т.е. различия функции распределения.

В качестве статистики служит наибольшая разность между ординатами обеих относительных кривых накопленных частот. При этом (при одинаковых для обеих выборок границах классов) накопленные частоты F_1 и F_2 делятся на соответствующие объемы выборок n_1 и n_2.

Затем вычисляется разность ($F_1/n_1 - F_2/n_2$). Максимум абсолютного значения этой разности и есть искомая статистика D (для более интересного в данном случае двухстороннего критерия):

$$D = \max\left|\left(\frac{F_1}{n_1} - \frac{F_2}{n_2}\right)\right|. \qquad (12.1)$$

Распределение статистики D было табулировано Смирновым в 1948 году.

Для средних и больших объемов выборок $(n_1 + n_2 > 35)$ значение D может быть приближенно заменено выражением

$${}_{\alpha}D = {}_{\alpha}K \cdot \sqrt{\frac{n_1 + n_2}{n_1 n_2}}, \qquad (12.2)$$

где ${}_{\alpha}K$ есть постоянная, зависящая от вероятности ошибки α: ее значения приведены в таблице (12,1).

Таблица 12.1. Значения коэффициента ${}_{\alpha}K$ в зависимости от 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

Если определенное на основании двух выборок значение $\widetilde{D}$ достигает критического значения ${}_{\alpha}D$ или превосходит его, это означает наличие значимой разницы между распределениями или функциями накопленных вероятностей.

Для случая равных объемов выборок ($n_1 = n_2 = n$) приведено ниже несколько строк из таблицы (Massey, 1951), для критических значений

${}_{\alpha}D_n$ в знаменателе указан объем выборки (табл. 12.2). Числитель для нетабулированных значений ${}_{\alpha}D_n$ получается по формуле ${}_{\alpha}K \cdot \sqrt{2n}$, его надо округлить до следующего целого числа. Например, для $\alpha = 0{,}05$ и $n = 10$ в соответствии с таблицей 1.21 получаем $1.36\sqrt{2 \cdot 10} = 6.08$ и округляем до 7, т.е. ${}_{0.05}D_{10}$=7/10.

Таблица 12.2. Отдельные значения ${}_{\alpha}D_n$ (критерий двусторонний)

$n(= n_1 = n_2)$	10	15	20	25	30
$\alpha = 0.05$	7/10	8/15	9/20	10/25	11/30
$\alpha = 0.01$	8/10	9/15	11/20	12/25	13/30

Если вычисленное на основании двух выборок значение $\hat{D}$ равно критическому значению ${}_{\alpha}D_n$ или превосходит его, то имеется значимое различие.

Пример. Сравним два ряда измерений, приведенных в таблице 12.3. О возможных различиях какого-либо вида ничего не известно. Мы проверяем нуль-гипотезу: генеральные совокупности одинаковы, против альтернативной гипотезы: генеральные совокупности имеют различные распределения ($\alpha = 0.05$, критерий двусторонний).

Таблица 12.3. Данные двух выборок

Первый ряд	2,1	3,0	1,2	2,9	0,6	2,8	1,6	1,7	3,2	1,7
Второй ряд	3,2	3,8	2,1	7,2	2,3	3,5	3,0	3,1	4,6	3,2

Распределив приведенные в таблице 12.3 данные по интервалам (табл. 12.4), перейдем к накопленным частотам (табл. 12.5), для которых вычислим отношения F_1 / n_1 и F_2 / n_2, а также модули их разностей.

Таблица 12.4. Интервальное распределение частот

Интервалы	0,0-0,9	1,0-1,9	2,0-2,9	3,0-3,9	4,0-4,9	5,0-5,9	6,0-6,9	7,0-7,9
f_1	1	4	3	2	0	0	0	0
f_2	0	0	2	6	1	0	0	1

Таблица 12.5. Расчет разности $F_1 / n_1 - F_2 / n_2$

Интервалы	0,0-0,9	1,0-1,9	2,0-2,9	3,0-3,9	4,0-4,9	5,0-5,9	6,0-6,9	7,0-7,9
F_1 / n_1	1/10	5/10	8/10	10/10	10/10	10/10	10/10	10/10
F_2 / n_2	0/10	0/10	2/10	8/10	9/10	9/10	9/10	10/10
$F_1 / n_1 - F_2 / n_2$	1/10	5/10	6/10	2/10	1/10	1/10	1/10	0/10

В качестве абсолютно наибольшей разности получаем значение $\widetilde{D} = 6/10$, которое меньше, чем критическое значение ${}_{0.05}D_{10}$=7/10, следовательно, гипотеза об однородности сохраняется: на основании имеющихся выборок нельзя отвергать возможность существования общей генеральной совокупности.

Замечание. Если проверку нулевой гипотезы произвести, используя сырые данные (табл. 12.3), средствами параметрического t - критерия (табл. 14.7 и последующие расчеты), то так как ${}_{0.05}t_{18}$=2,101 и, следовательно, $\widehat{t} < {}_{0.05}t_{18}$, нуль-гипотеза отклоняется на уровне $\alpha = 0{,}10$. Не отклоняется только на уровне $\alpha = 0{,}01$.

Таблица 12.6. Расчет основных параметров выборок , рассмотренных в таблице 12.3.

	$\sum$	$\bar{x}$	S^2	S
1-я выборка	20,8	2,08	0,753	0,868
2-я выборка	36,0	3,60	2,098	1,448

$$S = \sqrt{\frac{S_1^2 + S_2^2}{n}} = \sqrt{\frac{0.753 + 2.098}{10}} = 0.534, \quad \widetilde{t} = \frac{3.60 - 2.08}{0.534} = 2.846.$$

В рассмотренном примере подтвердилось высказанное в пункте 12.1 утверждение, что непараметрический критерий, когда его применяют на нормальном распределении, особенно при малых выборках, всегда слабее, чем соответствующий параметрический критерий

12.3. Сравнение рассеяний двух малых выборок по Пиллаи и Буэнавентуре. Рассеяние двух независимых рядов измерений могут быть сравнены с помощью *размахов* каждого из рядов (R_1, R_2). Для этой цели образуют аналогично F-критерию отношение R_1 / R_2 при $R_1 > R_2$ и проверяют, достигает ли отношение R_1 / R_2 соответствующей границы в табл. 12.7.

Когда, например, ряд измерений (А) с $n_1 = 9$ и ряд измерений (В) с $n_2 = 10$ имеют размах $R_1 = 19$ и $R_2 = 10$, тогда R_1 / R_2=1,9 больше, чем табличное значение 1,82 для $\alpha = 0{,}05$, поэтому нуль-гипотеза отклоняется.

Рассмотренный критерий Пиллаи и Буэнавентуры предполагает приближенно нормальное распределение.

Таблица 12.7. Верхние границы значений F-распределения, основанного на размахах (при $\alpha = 0{,}05$).

n_2 \ n_1	2	3	4	5	6	7	8	9	10
2	12,71	19,08	23,2	26,2	28,6	30,5	32,1	33,5	34,7
3	3,19	4,37	5,13	5,72	6,16	6,53	6,85	7,12	7,33
4	2,03	2,66	3,08	3,38	3,62	3,84	4,00	4,14	4,26
5	5,60	2,05	2,35	2,57	2,75	2,89	3,00	3,11	3,19
6	1,38	1,74	1,99	2,17	2,31	2,42	2,52	2,61	2,69
7	1,24	1,57	1,77	1,92	2,04	2,13	2,21	2,28	2,34
8	1,15	1,43	1,61	1,75	1,86	1,94	2,01	3,08	2,13
9	1,09	1,33	1,49	1,62	1,72	1,79	1,86	1,92	1,96
10	1,05	1,26	1,42	1,54	1,63	1,69	1,78	1,82	1,85

То же самое при $\alpha = 0{,}01$

n_2 \ n_1	2	3	4	5	6	7	8	9	10
2	63,66	95,49	116,1	131	142	153	161	168	174
3	7,37	10,00	11,64	12,97	13,96	14,79	15,52	16,13	16,60
4	3,73	4,79	5,50	6,01	6,44	6,80	7,09	7,31	7,51
5	2,66	3,33	3,75	4,09	4,36	4,57	4,73	4,89	5,00
6	2,17	2,66	2,98	3,23	3,42	3,58	3,71	3,81	3,88
7	1,89	2,29	2,57	2,75	2,90	3,03	3,13	3,24	3,33
8	1,70	2,05	2,27	2,44	2,55	2,67	2,76	2,84	2,91
9	1,57	1,89	2,07	2,22	2,32	2,48	2,50	2,56	2,63
10	1,47	1,07	1,92	2,06	2,16	2,26	2,33	2,38	2,44

12.4. Сравнение средних значений двух малых выборок по Лорду. Для сравнения центров независимых рядов измерений равного объема ($n_1 = n_2 \le 20$) вычисляют разность между средними арифметическими и делят ее на среднее арифметическое размаха (R_1, R_2):

$$u = \frac{|\bar{x}_1 - \bar{x}_2|}{(R_1 + R_2)/2}. \tag{12.3}$$

Если статистика u, аналогичная t-статистике, достигает или превосходит границу, представленную в таблице 12.8, то разность средних значений на соответствующем уровне значима. Критерий предполагает нормальность распределения и равенство дисперсий в табулированной области, он имеет такую же мощность, как и t- критерий.

Таблица 12.8. Границы для сравнения двух средних значений независимых рядов измерений равного объема по Лорду

$n = n_1 = n_2$	Односторонний критерий		Двусторонний критерий	
	0,05	0,01	0,05	0,01
3	0,974	1,715	1,272	2,093
4	0,644	1,047	0,831	1,237
5	0,493	0,772	0,613	0,896
6	0,405	0,621	0,499	0,714
7	0,347	0,585	0,426	0,600
8	0,306	0,459	0,373	0,521
9	0,275	0,409	0,334	0,464
10	0,250	0,371	0,304	0,419
11	0,233	0,340	0,280	0,384
12	0,214	0,315	0,260	0,355
13	0,201	0,294	0,243	0,331
14	0,189	0,276	0,228	0,311
15	0,179	0,261	0,216	0,293
16	0,170	0,247	0,205	0,278
17	0,162	0,230	0,195	0,264
18	0,155	0,225	0,187	0,0,252
19	0,149	0,216	0,179	0,242
20	0,143	0,207	0,172	0,232

Пример. Если нужно сравнить ряды измерений (А): 2, 4, 1, 5 и (В): 7, 3, 4, 6, то т.к. $\bar{x}_1 = 3$, $\bar{x}_2 = 5$, а $R_1 = 5 - 1 = 4$. $R_2 = 7 - 3 = 4$, получится $u = \frac{|3-5|}{(4+4)/2} = 0.5$, что при $n_1 = n_2 = 4$ и двустороннем критерии на уровне $\alpha = 0{,}05$ не дает оснований для отклонения H_0. Обе выборки взяты из общей генеральной совокупности со средним значением μ. Мур составил таблицы, (1957) для этого критерия для неодинаковых объемов при $n_1 + n_2 \le 39$.

Рассмотренный критерий Лорда, как и критерий Пиллаи и Буэнавентуры, предполагает приближенно нормальное распределение.

12.5. U-критерий Уилкоксона, Манна и Уитни. Распределения, используемые в психолого-педагогических исследованиях, редко когда являются, в строгом смысле слова, нормальными. В этом случае для сопоставления независимых выборок часто используют ранговый критерий Манна и Уитни, основанный на так называемом критерии Уилкоксона для независимых выборок, который является непараметрическим аналогом t-критерия 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 (1947), который проверяет нуль-гипотезу о принадлежности двух независимых выборок одной и той же генеральной совокупности, равенстве их функций распределения: $F_1(x) = F_2(x)$. Эта гипотеза включает также *равенство положений,* в частности, равенство значений медиан и равенство значений математических ожиданий $\mu_1 = \mu_2$.

Асимптотическая эффективность U-критерия равна примерно 95%, т.е. при использовании этого критерия для 1000 значений мощность критерия получается такая же, как при использовании t-критерия для 950 значений при условии справедливости нормального распределения. Очевидно, что U-критерий целесообразно применять также для приближенных расчетов или для контроля заключений на основании t-критерия, когда эти заключения выглядят неправдоподобными. Предполагается, что сравниваемые выборки относятся к распределениям одинакового типа. В случае произвольного распределения асимптотическая эффективность U-критерия не может быть ниже 86,4%.

Для вычисления статистики U упорядочивают $(m+n)$ значений объединенной выборки по величине, причем каждому рангу приписывают, к какой из выборок он относится. Пусть сумма рангов первой выборки равна R_1, второй выборки – R_2. Вычисляем

$$U_1 = mn + \frac{m(m+1)}{2} - R_1 \quad \text{и} \quad U_2 = mn + \frac{n(n+1)}{2} - R_2. \qquad (12.4)$$

Полученные результаты проверяются по формуле:

$$U_1 + U_2 = m \cdot n. \qquad (12.5)$$

Искомая статистика есть меньшее из значений U_1 и U_2. Нуль-гипотеза отвергается тогда, когда вычисленное значение U-критерия оказывается меньше критического значения $U_{(m,n;\ a)}$, приведенного в таблицах F_1 и F_2 Приложения или равно ему.

Для достаточно больших выборок (m+n>60) справедлива превосходная аппроксимация:

$$U_{(m,n;\,\alpha)} = \frac{nm}{2} \cdot z \cdot \sqrt{\frac{nm(n+m+1)}{12}}. \qquad (12.6)$$

Значение z для двух или одностороннего критерия может быть определено по таблице А Приложения.

Применение критерия Манна – Уитни проиллюстрируем на конкретном примере. Предположим, что в таблице (12.9) приведены результаты тестирования (в баллах) экспериментальной и контрольной групп. Число замеров в группах может не совпадать.

Таблица 12.9. Баллы, набранные учащимися экспериментальной и контрольной групп (максимально возможное число баллов – 50).

Экспериментальная 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, то воспользуемся U-критерием.

В качестве нуль-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. 12.10). При совпадении двух или более значений набранных баллов ученикам, их набравшим, приписываются средние значения рангов.

Для каждой из групп подсчитывается сумма рангов. В нашем случае сумма рангов экспериментальной группы $-R_1 = 148$, сумма рангов контрольной группы – $R_2 = 108$. Затем, пользуясь формулами (12.4), вычислим

$$\mathrm{U}_1 = mn + \frac{m(m+1)}{2} - R_1 = 10 \cdot 12 + \frac{10 \cdot 11}{2} - 148 = 27 \text{ и}$$

$$\mathrm{U}_2 = mn + \frac{n(n+1)}{2} - R_2 = 10 \cdot 12 + \frac{12 \cdot 13}{2} - 108 = 90.$$

Полученные значения проверим, пользуясь правилом (12.5): $U_1 + U_2 = 93 + 27 = 10 \cdot 12 = n_1 \cdot n_2$.

В качестве проверочной статистики берут меньшее из значений U_1 и U_2. В рассматриваемом примере ею будет значение $U = 27$.

Нуль-гипотеза отвергается, когда вычисленное значение U меньше критического значения, которое определяется по таблицам F_1, F_2 и F_3 Приложения.

Таблица 12.10. 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

В нашем случае при $n_1 = 10$ и $n_2 = 12$ на уровне значимости $\alpha = 0{,}05$ двухстороннего критерия (таблица F_1 Приложения) $U_{крит} = 29$. Так как $U < U_{\text{крит}}$, то нулевую гипотезу следует отвергнуть. Следовательно, судя по качеству выполнения теста, различие между экспериментальной и контрольной группами статистически значимо на уровне $\alpha = 0{,}05$.

Вместо выражения (12.6) в случаях, когда значение α не может быть заранее задано или нет таблиц критических значений $U(m,n;\alpha)$ и когда объемы выборок не слишком малы ($m \geq 8$, $n \geq 8$), при отсутствии в них элементов с одинаковым значением, можно воспользоваться Z-критерием:

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

Полученное значение z сравнивается с таблицами стандартного нормального распределения. Как легко подсчитать, в нашем случае z=2,175, что больше критического значения 1,960 при $\alpha = 0{,}05$, следовательно, как и в предыдущем случае, нулевую гипотезу следует отвергнуть.

Если внутри выборки, но не между двумя выборками, содержатся повторяющиеся значения, то им присваиваются средние ранги.

Если же одинаковые значения встречаются в разных выборках, то применяется скорректированная формула:

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

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 $2{,}18 > 1{,}96$, то при $\alpha = 0{,}05$ нулевую гипотезу следует отвергнуть.

12.6 Упрощенный критерий Тьюки. Две группы измерений отличаются тем больше, чем меньше пересекаются их значения. На этой основе Тьюки (1959) разработал свой упрощенный метод сравнения двух независимых выборок.

Пусть одна группа содержит наибольшее, а другая наименьшее значение; тогда необходимо подсчитать количество:

1) тех значений группы, которые превосходят все значения другой группы;

2) тех значений другой группы, которые меньше всех значений первой группы.

Оба значения (каждое должно быть больше нуля) складываются, и таким образом получается значение статистики Т. Если объемы выборок примерно одинаковы, то критические значения статистики равны соответственно 7, 10 и 13:

7 – для двустороннего критерия на 5%-ном уровне;

10 – для двустороннего критерия на 1%-ном уровне и

13 – для двустороннего критерия на 0,1%-ном уровне.

Двум равным значениям следует приписать число 0,5.

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

$$n_1 \le n_2 \le 3 + 4n_1 / 3 \qquad (12.9)$$

Во всех других случаях рассчитанное значение статистики Т перед сравнением его с числами 7, 10 или 13 должно быть уменьшено на корректирующее число:

$$1, \text{ если } (3 + 4n_1 / 3) < n_2 < 2n_1 \qquad (12.10)$$

$$\text{целую часть от } \frac{n_2 - n_1 + 1}{n_1}, \text{ если } 2n_1 \le n_2. \qquad (12.11)$$

Например, для $n_1 = 7$ и $n_2 = 13$ условие (12.9) не выполняется, так как $3 + \frac{4 \cdot 7}{3} = \frac{37}{3} < 13$. Корректирующее значение в соответствии с (12.10) равно 1.

При $n_1 = 4$ и $n_2 = 14$ в соответствии с (12.11) получаем $\frac{14 - 4 + 1}{4} = \frac{11}{4} = 2{,}75$, т.е. корректирующее значение равно 2.

Если одна выборка превышает другую больше чем на 9 элементов, то для 0,1% - го уровня нужно применять критическое значение 14 вместо 13.

В качестве примера рассмотрим два ряда измерений, упорядоченных в порядке возрастания (табл. 12.11).

Таблица 12.11. Сопоставляются два ряда измерений, упорядоченных в порядке возрастания

1-ый ряд	14,6	14,7	14,8	14,9	15	15,1	15,3	16,1	16,7	17,3
2-ой ряд	13,8	13,9	14,2	14,3	14,4	14,6	14,7	15	–	–

Во втором ряду имеется пять значений, меньших минимального значения первого ряда {13.8; 13,9; 14,2; 14,3; 14,4} и одно совпадающее {14.6}, всего 5,5.

В первом ряду имеются пять значений превышающих максимальное значение второго ряда {15.1;15.3; 16.1; 16,7; 17,3} и одно совпадающее {15,0}, всего 5,5. Статистика Т=5,5+5,5=11. Корректирующее значение равно нулю, так как $\left(n_1 \le n_2 \le 3 + 4n_1 / 3\right)$ 8<10<41/3. Поскольку Т=11>10, нуль-гипотеза (равенство функций распределения, соответствующих обеим выборкам) на1%-ном уровне должна быть отклонена.

12.7. Критерий Мостеллера. Простейший непараметрический критерий для проверки одинаковости распределений двух генеральных совокупностей в 1948 г. предложил Мостеллер. Для проверки нуль-гипотезы (две рассматриваемые генеральные совокупности имеют одинаковое распределение) из данных совокупностей извлекается по одной выборке одинакового объема ($n_1 = n_2 = n$). Нуль-гипотеза при $n > 5$ отклоняется на уровне значимости 0,05 (вероятность ошибки 5%), если

5 наибольших или наименьших значений при $n \le 25$,

6 наибольших или наименьших значений при $n > 25$

содержит одна и та же выборка.

В качестве примера сопоставим результаты тестирования экспериментальной и контрольной групп по одному и тому же тесту.

Экспериментальная группа (баллы): 48, 36, 28, 46, 36, 24, 48, 38, 26, 34, 37, 31, 35, 31, 27, 40, 35, 43, 34, 33, 34, 35, 41, 30, 36. Минимальный интервал, содержащий все эти значения (24, 48), включающий размах 25.

Контрольная группа (баллы): 39, 21 ,50, 31, 26, 36, 24, 16, 20, 22, 16, 30, 25, 43, 27, 30, 38, 19, 31, 31, 32, 35, 31, 20, 31. Минимальный интервал, содержащий все эти значения (16, 50), включающий размах 35.

Контрольная группа содержит 7 элементов (21, 16, 20, 22, 16, 19, 20), каждый из которых меньше любого из элементов экспериментальной группы. В соответствии с критерием Мостеллера нулевую гипотезу об одинаковости распределений следует отвергнуть на пятипроцентном уровне статистической значимости.

12.8. Критерии Розенбаума. В 1954 году Розенбаум предложил два непараметрических критерия: критерий положения и критерий изменчивости.

Критерий положения. Даны две генеральные совокупности, о распределении которых ничего не известно. Проверяется ненаправленная гипотеза о равенстве их медиан, а именно нулевая гипотеза $H_0 : Me_1 = Me_2$ и альтернативная гипотеза $H_1 : Me_1 \ne Me_2$.

Для проверки нуль-гипотезы из данных генеральных совокупностей извлекают две случайных выборки одинакового объема n. Если, по меньшей мере, 5 (для $n \geq 16$; $\alpha = 0{,}05$) или 7 (для $n \geq 20$; $\alpha = 0{,}01$) значений одной выборки лежат выше или ниже размаха другой выборки, то нуль-гипотеза (равенство медиан) с заданной вероятностью ошибки отклоняется; предполагается, что различие размахов случайное; вероятности ошибки даны для одностороннего критерия; для двухстороннего критерия они должны быть удвоены.

В качестве примера вновь обратимся к двум рассмотренным выше выборкам значений набранных баллов в экспериментальной и контрольной группах. Так как семь значений контрольной группы лежат ниже минимального интервала экспериментального ряда, то и на этот раз нулевую гипотезу о равенстве медиан двух генеральных совокупностей следует отвергнуть на уровне статистической значимости $\alpha = 0{,}05$.

Критерий изменчивости. Если, по меньшей мере 7 (для $n \geq 25$; $\alpha = 0{,}05$) или 10 (для $n \geq 51$; $\alpha = 0{,}01$) значений одной выборки (с большим размахом; критерий односторонний) лежат вне размаха другой выборки, то нуль-гипотеза (равенство мер изменчивости – дисперсий) отклоняется с заданной вероятностью ошибки; предполагается, что различие медиан случайное.

Если неизвестно, что обе генеральные совокупности имеют одно и то же положение, то этот критерий проверяет и положение и рассеяние. Для $7 \leq n \leq 24$ необходимо заменить 7 на 6 ($\alpha = 0{,}05$), для $21 \leq n \leq 50$ (или $11 \leq n \leq 20$) нужно заменить 10 на 9 (или на 8).

Обратившись и на этот раз к двум рассмотренным выше выборкам, заметим, что 8 значений контрольной группы лежат вне минимального интервала экспериментальной группы $\{16.\ 17.\ 19.\ 20.\ 20.\ 21.\ 22.\ 50\}$. Отсюда следует, что нуль-гипотезу о равенстве дисперсий следует отклонить с вероятностью ошибки $\alpha = 0{,}05$.

Глава 13. Критерии, основанные на парных сравнениях

13.1. Парные сравнения. До сих пор при описании экспериментов, в которых сравнивалась эффективность двух систем обучения, развития или воспитания, мы приводили примеры двух принципиально различных процедур формирования выборочных совокупностей.

В соответствии с первой из них из генеральной совокупности случайным образом извлекаются две *независимые выборки*, над которыми и выполняются предусмотренные экспериментом действия.

Во втором случае формировались *зависимые выборки*, например, в первой выборке девочки, а во второй их братья. Затем в процессе дальнейшей обработки оперировали парами соответствующих элементов, например «сестра – брат». В этом случае говорят о *парных сравнениях*. К их числу относят и случай, когда строится одна выборка, а сам эксперимент проводится в два этапа, например, в начале эксперимента и в конце, или при сравнении двух тестов на одной и той же группе учащихся.

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, полу, активности, характеру и т.п. элемент. Распределение отобранных «партнёров» по двум группам производится случайным образом.

Ещё более эффективной считается следующая процедура: отбирают группу учащихся и проводят так называемое «сравнение направо-налево» так, чтобы индивидуумы справа и слева образовывали взаимно независимые группы учащихся, и затем случайным образом определяют, где следует применить первый, где второй метод.

Метод парных сравнений считается более точным потому, что с его помощью достигается почти полная однородность экспериментального материала, в результате чего существенно уменьшается или даже полностью исключается рассеяние, которое имеется между двумя сопоставляемыми группами. Заметим, что используемые при парном сравнении выборки, так же, как и кратные выборки, называют связными, или коррелированными.

Разумеется, применение парного сравнения уменьшает число степеней свободы. Действительно, при сравнении средних значений в случае однородных дисперсий имеется $n_1 + n_2 - 2$ степеней свободы. В то же время при кратных выборках число степеней свободы равно числу пар или разностей минус единица, т.е. $(n_1 + n_2)/2 - 1$. Если $n_1 = n_2 = n$, то отношение числа степеней свободы при независимых выборках и кратных выборках равно $(2n-2)/(n-1) = 2/1$. В результате уменьшения числа степеней свободы происходит значительная потеря точности. Однако так как рассеяние между испытуемыми при всех обстоятельствах больше, чем рассеяние между двумя группами этих испытуемых, то выигрыш в точности, который приносит применение связанных выборок, превосходит потери от уменьшения степеней свободы. Вообще общий выигрыш будет тем больше, чем больше отношение этих рассеяний друг к другу.

Взаимнокоррелируемые пары наблюдений получают на основе двух следующих принципов. Известно построение опытов с повторным контролем в одной и той же выборке индивидуумов. Испытуемые проверяются, например, один раз при нормальных условиях и другой раз – в состоянии возбуждения, стресса. При этом необходимо обратить внимание на то, чтобы такие факторы, как упражнение (обучение) или усталость, были исключены. Другой принцип заключается в организации кратных выборок с помощью предварительной проверки одного из измеримых или оцениваемых признаков, который наиболее коррелирован с изучаемым признаком. Испытуемые, например, с помощью предварительного критерия размещаются в ранжированный ряд. Любые два следующих друг за другом в этом ряду индивидуума образуют пару. С помощью случайной процедуры (например, с помощью бросания монеты) определяется, какого партнера к какой выборочной группе отнести.

13.2. Проверка ненаправленной гипотезы $\mu_1 = \mu_2$ по зависимым выборкам. Рассмотрим теперь процедуру проверки гипотезы $\mu_1 = \mu_2$ в случае зависимых выборок. Как отмечалось выше, такая зависимость может быть как естественной, например, если сопоставляемые выборки получены в результате тестирования одной и той же группы испытуемых в двух разных условиях, так и искусственной, определяемой методом построения выборок. Именно этот случай, т.е. случай зависимых выборок и будет рассмотрен в настоящем пункте.

Проверка ненаправленной гипотезы $\mu_1 = \mu_2$ по *зависимым* выборкам осуществляется в следующем порядке:

а) проверяемая гипотеза состоит в том, что разность между средними значениями μ_1 и μ_2 двух совокупностей равна нулю против альтернативной гипотезы о том, что эта разность отлична от нуля. $H_0: \mu_1 - \mu_2 = 0$, $H_1: \mu_1 - \mu_2 \neq 0$;

б) не исключается, что между выборками может существовать отличная от нуля корреляция, и есть возможность «объединить» элементы этих выборок в «пары». В этом случае данные, собранные по двум выборкам, могут быть представлены в виде n - пар ($x_{i1};x_{i2}$), где $i=1,2,\cdots,n$.

Рассматриваемая гипотеза эквивалентна гипотезе о том, что $X_1 - X_2$ обладает нулевым средним в генеральной совокупности;

в) обозначив разности пар наблюдений из двух выборок $x_{i1} - x_{i2}$ символом d_i, получим критерий:

$$t = \frac{\overline{d}}{Sd/\sqrt{n}}, \qquad (13.1)$$

где $\overline{d} = \sum_{i=1}^{n} d_i / n$ – среднее n значений разностей $d_i = x_{i1} - x_{i2}$,

а $S_d = \sqrt{\sum_{i=1}^{n} \left(d_i - \overline{d}\right)^2 / (n-1)}$ – стандартное отклонение разностей d_i.

Подставив эти значения в (13.1) и преобразовав, получим:

$$t = \frac{\overline{d}}{\sqrt{\dfrac{\overline{d^2} - \overline{d}^2}{n-1}}}, \text{ где } \overline{d^2} = \sum d_i^2 / n;$$

г) если верна гипотеза $H_0: \mu_1 - \mu_2 = 0$, то t в выражении (13.1) будет подчиняться t-распределению Стьюдента с $\nu = n-1$ степенями свободы.

Дальнейшие действия по общей схеме.

Пример. Для выяснения изменений, происшедших за пять лет обучения в вузе в уровне умений решать задачи по курсу школьной математики, первокурсникам, а через четыре года тем же пятикурсникам давалось задание из десяти задач. В результате было получено 55 пар значений типа «до-после» и их разностей $d_i = x_{i1} - x_{i2}$. Среднее $\overline{d}$ и стандартное отклонение S_d значений d_i оказались равны:

$$\overline{d} = \sum_{1}^{55} \frac{x_{i1} - x_{i2}}{55} = -1{,}56; \qquad S_d = \sqrt{\sum_{i=1}^{n} \left(d_i - \overline{d}\right)^2 / 54} = 2{,}50.$$

Для проверки $H_0: \mu_1 - \mu_2 = 0$ на уровне значимости $\alpha = 0{,}05$ вычислим значение t в выражении (13.1): $\tilde{t} = \dfrac{\overline{d}}{Sd/\sqrt{n}} = \dfrac{-1{,}56}{2{,}50/\sqrt{55}} = -4{,}63$. Нижняя критическая точка t-распределения с ν, равным 55-1=54, будет равна:

${}_{0,975}t_{54} = -2{,}005$. Так как $\hat{t}$ меньше ${}_{0,975}t_{54}$, то исследователь должен на уровне $\alpha = 0{,}05$ отклонить нулевую гипотезу.

13.3. Проверка двухсторонней гипотезы $\sigma_1^2 = \sigma_2^2$ по зависимым выборкам. В этом случае рекомендуется следующая последовательность действий:

а) проверяемая нуль-гипотеза состоит в том, что две совокупности имеют одинаковую дисперсию:

$$H_0\colon\ \sigma_1^2 = \sigma_2^2 \ \text{ против гипотезы } H_1\colon\ \sigma_1^2 \neq \sigma_2^2;$$

б) предполагается, что две выборки одинакового объёма n, независимость которых не установлена, берутся из нормальных совокупностей с дисперсиями σ_1^2 и σ_2^2;

в) при проверке H_0 против H_1 используется критерий:

$$\hat{t} = \frac{\left|\left(Q_x - Q_y\right)\cdot\sqrt{n-2}\right|}{2\sqrt{Q_xQ_y - \left(Q_{xy}\right)^2}}, \tag{13.2}$$

где $Q_x = (n-1)S_x^2$, $Q_y = (n-1)S_y^2$, а $Q_{xy} = \sum xy - \dfrac{\sum x \sum y}{n}$.

Где S_x^2 и S_y^2 – выборочные дисперсии первой и второй выборок соответственно, n – число пар, объединяющих каждое наблюдение первой выборки с одним наблюдением второй выборки.

Например, для данных, приведенных в таблице

x_i	21	18	20	21	$\sum x = 80$
y_i	26	33	27	34	$\sum y = 120$

$$Q_x = 6,\ Q_y = 50 \text{ и } Q_{xy} = (21\cdot 26 + 18\cdot 33 + 20\cdot 27 + 21\cdot 34) - \frac{80\cdot 120}{4} = -6,$$

получаем $t = \dfrac{\left|(6-50)\cdot\sqrt{4-2}\right|}{2\cdot\sqrt{6\cdot 50 - (-6)^2}} = 1{,}91$. Так как при $\nu - 4 - 2 = 2$, на уровне значимости $\alpha = 0{,}05$, ${}_{0.05}t_2 = 4{,}31$, то есть $\hat{t} < 4.31$. Отсюда следует, что на 5%-ном уровне нуль-гипотеза – связанные дисперсии равны – должна быть принята.

При одностороннем критерии с нуль-гипотезой $\sigma_x^2 = \sigma_y^2$ против $\sigma_x^2 < \sigma_y^2$ критическая граница была равна $t_{2;0,05;\text{одност}} = 2{,}92$.

Формула (13.2) может быть преобразована к виду:

$$t = \frac{S_1^2 - S_2^2}{\sqrt{4S_1^2 S_2^2\left(1 - r_{12}^2\right) / (n-2)}}\ , \qquad (13.3)$$

где, а r_{12}^2 – коэффициент корреляции, найденный по этим сформированным парам (см. главу 16);

г) если верна H_0: $\sigma_1^2 = \sigma_2^2$, выборочное распределение t в выражении (13.3) представляет собой t- распределение Стьюдента с $n-2$ степенями свободы.

Пример. Сравнивается разброс годовых оценок учащихся шестых классов по геометрии с разбросом годовых оценок тех же учащихся по тому же предмету, но уже в седьмом классе. Выборка учащихся шестых классов, состоящая из 60 человек, показала, что $S_1^2 = 0{,}62$. У тех же учащихся через год в седьмом классе $S_2^2 = 0{,}88$. Коэффициент корреляции по парам наблюдений (ученик шестого класса – он же ученик седьмого класса) – $r_{12} = 0{,}80$. Пользуясь выражением (13.3), получим:

$$\widehat{t} = \frac{0{,}62 - 0{,}88}{\sqrt{\frac{4 \cdot 0{,}62 \cdot 0{,}88}{60-2}\left(1 - (0{,}80)^2\right)}} = -2{,}23\,.$$

На уровне значимости $\alpha = 0{,}05$ критическими значениями t- распределения Стьюдента, с которыми сравнивается полученное значение $t = -2{,}23$, являются ${}_{0,975}t_{58} = -2{,}000$ и ${}_{0,025}t_{58} = 2{,}000$.

Так как $\widehat{t}$ лежит ниже нижней критической точки, то нулевую гипотезу о равенстве дисперсий генеральных совокупностей следует отвергнуть и признать, что с переходом в седьмой класс разброс оценок увеличивается.

При неравных дисперсиях производят проверку (приближённую) по критерию Уилкоксона для разностей пар или по знаковому критерию.

13.4. Критерий Уилкоксона для разностей пар. В пункте 13.2 была рассмотрена процедура проверки гипотезы $\mu_1 = \mu_2$ в предположении, что используемые выборки зависимы, а разности рассматриваемых парных значений распределены по нормальному закону. 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 и почти также строго проверяет нормально распределенные разности; его эффективность для больших и малых выборок равна 95%.

С помощью рангового критерия Уилкоксона проверяется нулевая гипотеза, заключающаяся в том, что обе связанные выборки принадлежат к одной генеральной совокупности и, следовательно, равны их *функции распределения* $F_1(x) = F_2(x)$ и математические ожидания или медианы. Соответствующая нуль-гипотеза имеет вид: значение математического ожидания разностей d_i парных наблюдений μ_d равно нулю.

Рассмотрим, приведенные в таблице 13.1, пары значений.

Таблица 13.1. Баллы, набранные учащимися в начале (A) и в конце (B) эксперимента (максимально возможное число баллов – 50).

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

Отбрасывая пары с отдельными равными значениями, в таблице 13.1 таких пар две: (71: 71) и (43; 43), из оставшихся n пар (в нашем примере их 10) значений образуют разности:

$$d_i = x_{i1} - x_{i2} \qquad (13.4)$$

и упорядочивают их абсолютные значения $|d_i|$ по рангам: наименьшее значение получает ранг 1, наибольшее – n. Равным по величине значениям разностей приписывается *средний ранг*. У каждого ранга отмечается знак (положительный или отрицательный), соответствующий знаку разности. Далее образуют суммы положительных и отрицательных рангов $\left(\widehat{R}_p \text{ и } \widehat{R}_n\right)$ и проверяют их с помощью выражения

$$\widehat{R}_p + \widehat{R}_n = n(n+1)/2, \qquad (13.5)$$

где n – число пар со 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 статистики используется меньшая из сумм рангов $\left(\widehat{R}\right)$. В нашем случае $\widehat{R}_p = 2$

Нуль-гипотеза отбрасывается, если вычисленное $\widehat{R}$- значение равно критическому значению $R(n;\ \alpha)$ из таблицы G Приложения. Так как в рассматриваемом случае $R(10;0{,}5)=8$, то нулевая гипотеза отбрасывается.. Для $n>25$ справедлива аппроксимация:

$$R(n;\alpha)=\frac{n(n+1)}{4}-z\cdot\sqrt{\frac{1}{24}n(n+1)(2n+1)}. \qquad (13.6)$$

Необходимое значение z для одно - или двустороннего критерия можно взять из таблицы А Приложения. Тогда вместо (13,6) применяют (когда не могут или не хотят задавать значение α *и* $n>25$) эквивалентную запись.

$$z=\frac{\left|R-\frac{n(n+1)}{4}\right|}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}. \qquad (13.7)$$

Полученное значение $\hat{z}$ сравнивается затем со стандартным нормальным распределением (табл. А Приложения).

Пример. Исследователь сравнивает два метода (два теста), по-разному проверяющих уровень усвоения некоторого материала. Каждый из 10 испытуемых выполняет оба теста. В таблице 13.2 приведены результаты первого (А) и второго (В) тестирования. О нормальности распределения не известно. Нужно проверить нуль-гипотезу: между результатами тестирования нет значимого различия на уровне $\alpha=0{,}05$.

Таблица 13.2. Результаты тестирования десяти учащихся двумя тестами

№ испытуемых	1	2	3	4	5	6	7	8	9	10
А (баллов)	24	45	48	64	71	52	18	44	50	62
В (баллов)	28	36	41	58	71	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		9	3,5	6	1	8
R_p= 42		(+)7	(+)5	(+)3,5		(+)9	(+)3,5	(+)6		(+)8
R_n=3	(-)2	–	–	–	–	–	–	–	(-)1	–
Контроль 42+3=45=9(9+1)/2, следовательно $\hat{R}_n=3$										

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

Так как $\hat{R}$< R(9; 0,05), то нуль-гипотеза должна быть отклонена, т.е. на уровне статистической значимости $\alpha=0{,}05$ можно считать, что тесты дают не одинаковый результат.

Из быстрых непараметрических методов для анализа разностей парных наблюдений рассмотрим максимум-критерий и критерий знаков Диксона и Муда.

13.5. Максимум критерий для разностей пар. Максимум-критерий– очень простой критерий для сравнения двух парных рядов измерений. Следует только заметить, что если 5 абсолютно наибольших разностей имеют одинаковый знак, то разность надежна на 10%-ном уровне; при 6 разностях такого рода разность значима на 5%-ном уровне; при 8 – на 1%-ном уровне и при 11 – на 0,1%-ном уровне. Эти числа – 5, 6, 8 и 11 – справедливы при двустороннем критерии и для объема выборки $n \geq 6$.

При одностороннем критерии эти числа соответствуют 5%-, 2,5%-, 0,5%- и 0,05%-ным границам.

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 (Вальтер, 1958).

Обратимся к предыдущему примеру (табл. 13.2) Последовательность разностей: +15; +13; +7; +6; +6; -3; -4 имеет 7 одинаковых знаков и при двустороннем критерии, как и в предыдущем случае, приводит к отклонению нуль-гипотезы на уровне значимости $\alpha = 0{,}05$.

13.6. Критерий знаков Диксона и Муда. Название критерия происходит оттого, что используются только знаки разностей наблюдаемых значений. Предполагается непрерывность случайной переменной. Критерий служит в первую очередь для быстрой проверки разности положений двух независимых выборок. В отличие от t-критерия и критерия Уилкоксона здесь *не требуется, чтобы пары принадлежали к общей генеральной совокупности;* они могут, например, относительно возраста, пола и т. п. принадлежать разным генеральным совокупностям. Существенно, чтобы результаты отдельных пар были независимы друг от друга. Нуль-гипотеза критерия знаков: 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 значение медианы распределения разностей равно нулю. Доверительные границы для медианы находят по таблице L Приложения. Нуль-гипотеза отклоняется, когда имеется слишком мало или слишком много разностей одного знака, так что не достигнуты или превышены границы таблице L Приложения. Нулевые разности не принимаются во внимание, что, естественно, уменьшает объем выборки. Вероятность появления определенного числа плюсов или минусов определяется на основе биномиального распределения при $p = q = 1/2$. Рассмотрим, например, биномиальное распределение $\varphi(x) = C_6^x (1/2)^x (1-1/2)^{6-x} = \dfrac{C_6^x}{2^6} = 0{,}015625 \cdot C_6^x$, приведенное в таблице 13.3.

Таблица 13.3. Биномиальное распределение $\varphi(x) = C_6^x (1/2)^x (1-1/2)^{6-x}$

x	0	1	2	3	4	5	6
$\varphi(x)$	0,0156	0,0938	0,2344	0,3125	0,2344	0,0938	0,0156

Приведенные в ней данные показывают, что должно быть по меньшей мере 6 пар наблюдений, чтобы результат при двустороннем критерии был надежен на 5%-ном уровне: n=6, x=0 или 6; табулированные значения p=2·0,0156=0,0312 <0,05. Подобным же образом были определены и другие границы в таблицах L и L_1 Приложения.

Эффективность критерия знаков уменьшается с ростом объема выборки от 95% при $n = 6$ до 64% при $n \to \infty$.

Правое граничное значение (*ПГ*) табл. L_1 определяется по формуле: (*n – ЛГ+1*), где (*ЛГ*) – левое граничное значение.

Доверительный интервал (ДИ) для медианы (Me). 95%-ный *ДИ* и 99%-ный *ДИ для Me* получают:

при $n \le 100$ – из табл. L, столбцы 5% и 1% по формуле

$ЛГ \le Me \le 1 + ПГ$.

Например, $n = 60$, 95%-ный *ДИ: 22-е значение* $\le Me \le 39$-е значение;

при $n > 100$ – из табл. L_1, столбцы 5% и 1% по формуле:

$ЛГ \le Me \le n - ЛГ + 1$.

Например, $n = 300$, 95%-ный *ДИ: 133-е значение* $\le Me \le 168$-е значение.

Пример. Обратимся вновь к примеру, рассмотренному в таблице 13.2. В нем мы получили 10 пар (двусторонний критерий на 5%-ном уровне), где одна нулевая разность, 7 положительных и 2 отрицательные разности. Из таблицы L для $n = 9$ получаем границы 2 и 7. Фактические значения совпадают с этими границами, поэтому нет основания на уровне $\alpha = 0{,}05$ отклонять нуль-гипотезу.

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

$$\hat{z} = \frac{|2x - n| - 1}{\sqrt{n}}, \qquad (13.8)$$

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 значений, 8 отрицательных и 24 положительных. Обращаясь к таблице L на уровне 1% 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 формуле (13.8). Учитывая, что в нашем случае $x = 8,\ n = 32$, получим, что $\widehat{z} = \dfrac{|2 \cdot 8 - 32| - 1}{\sqrt{32}} = \dfrac{15}{5.657} = 2.65$. 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.

В 1958 году Дуквортом и Вуатом был предложен основанный на критерии знаков Диксона и Муда быстрый критерий. В нем статистикой T служит абсолютное значение разности между числом положительных и отрицательных рангов. 5%-ный уровень этой разницы равен $2\sqrt{n}$; 10%-ный уровень $-1{,}6 \cdot \sqrt{n}$, где n – общее число знаков заданных разностей.

Если $T > 2\sqrt{n}$ или $T > 1.6\sqrt{n}$, то при двустороннем критерии разница должна рассматриваться как значимая.

В последнем из рассмотренных выше примеров: T=24-8=16, а при двустороннем критерии на уровне $\alpha = 0{,}05$ верхнее критическое значение $T_{крит} = 2\sqrt{n} = 2\sqrt{32} = 2 \cdot 5.657 = 11.31$. Так как $\widehat{T} > T_{крит}$, то нулевую гипотезу о равенстве нулю разности приходится отвергнуть, то есть разница между обеими выборками статистически значима.

13.7. Другие применения критерия знаков для быстрого ориентирования.

1. Сравнение двух независимых выборок. Если необходимо сравнить положение центров двух независимых выборок, то можно не пользоваться методом вычисления средних значений, а сформировать случайным образом выборочные пары, определить их знаки и к ним применить критерий знаков.

2. Проверка принадлежности к одной генеральной совокупности.

*Пример 1.*Может ли последовательность чисел: 13, 12, 11, 9, 12, 8, 13, 12, 11, 11, 12, 10, 13, 11, 10, 14, 10, 10, 9, 11, 11 быть выборкой из генеральной совокупности со средним значением μ=10?

Мы подсчитаем число значений меньших, чем 10 и больших, чем 10, образуем разность и сравним её с критическим значением:

$$\hat{T} = 14 - 3 = 11 > 2 \cdot \sqrt{17} = 8{,}2 .$$

Гипотеза о том, что выборка относится к генеральной совокупности с μ=10, должна быть отклонена $(P < 0{,}05)$.

Пример 2. Принадлежит ли заданная последовательность значений: 24, 27, 26, <u>28, 31,</u> 35, 33, 37, 36, 37, 34, 32, <u>32, 29,</u> 28, <u>28, 31, 28,</u> 26, 25 к одной генеральной совокупности?

Для ответа на этот вопрос Тейлор предложил другую модификацию знакового критерия, позволяющую оценить изменчивость положения центра внутри генеральной совокупности. Вначале определяется выборочная медиана, затем вычисляют, как часто располагается медиана внутри последовательных пар чисел. Это значение обозначим через x^*. Если среднее значение генеральной совокупности изменяется, то x^* мало по сравнению с объёмом выборки n. Нуль-гипотеза о том, что случайная выборка принадлежит одной генеральной совокупности, отклоняется на 5%-ном уровне, если $\left|n - 2x^* - 1\right| \geq 2\sqrt{n-1}$.

Медиана нашей выборки объёмом $n = 20$ равна $x = 29{,}5$. Подчёркнутые пары $\left(x^* = 4\right)$ содержат между собой медиану. Мы получили $n - 2x^* - 1 = 20 - 8 - 1 = 11$ и $2\sqrt{n-11} = 2\sqrt{20-1} = 8{,}7$.

Так как $11 > 8.7$, то следует принять, что наблюдения относятся к двум различным генеральным совокупностям $(P < 0{,}05)$.

Глава 14. Критерии согласия

14.1. Постановка вопроса. До сих пор мы проверяли гипотезы, связанные с оценкой параметров известных или не известных распределений. Теперь обратимся к так называемым *критериям согласия*, которые позволяют устанавливать меру совпадения двух распределений. Особое значение имеет сравнение эмпирического и теоретического (или гипотетического) распределений. Для дальнейшего это важно еще и потому что, как мы видели, применение параметрических критериев требует проверки исследуемой генеральной совокупности на нормальность. Если между ними имеется согласие, то можно заключить, что эмпирическое распределение вызвано теми же причинами, которые лежат в основе теоретического распределения.

В основе построения критериев согласия лежит тот или иной способ определения «расстояния» между эмпирической функцией распределения $F_n^*(x)$ и предполагаемой (теоретической) функции распределения $F(x)$.

Меру отклонения эмпирической функции распределения от теоретической $D = D\{F_n^*, F\}$ можно определить многими способами, в соответствии с которыми получаются различные критерии для проверки интересующей исследователей гипотезы. Например, можно положить

$$D\{F_n^*, F\} = \sup_x \left| F_n^*(x) - F(x) \right|$$

или $D\{F_n^*, F\} = \int\limits_{-\infty}^{+\infty} \left[F_n^*(x) - F(x) \right]^{2k} \cdot g(x) dx$, где $g(x) \geq 0$, а $\int\limits_{-\infty}^{+\infty} g(x) dx < \infty$.

В первом случае для проверки данной гипотезы получим критерий Колмогорова, во втором случае (при $k = 1$) – критерий Мизеса.

14.2. Хи-квадрат критерий. Наиболее употребительной является мера, введенная Пирсоном, приводящая к критерию χ^2. Рассмотрим эту меру. Разобьем множество значений случайной величины на k множеств $S_1, S_2, ..., S_k$ без общих точек. Практически такое разбиение можно осуществить с помощью $(k-1)$ чисел $a_1 < a_2 < \ldots < a_{k-1}$. При этом правый конец каждого интервала исключается из соответствующего множества, а левый – включается.

Пусть p_i, где $i = 1, 2, \ldots, k$, а $\sum\limits_{i=1}^{k} p_i = 1$ – вероятность того, что случайная величина принадлежит множеству S_i и $np_i = \widehat{n}_i$ – теоретическая частота, соответствующая интервалу S_i. В свою очередь, пусть n_i, $i = 1, 2, \ldots, k$ –

количество величин из выборки объема n, попавшие в интервал S_i $\left(\sum_{i=1}^{k} n_i = n,\ \sum_{i=1}^{k} \frac{n_i}{n} = 1\right)$ (табл. 14.1).

Таблица 14.1. Сопоставление эмпирического и теоретического распределений

Интервал	s_1	s_2	…	s_i	…	s_k
Эмпирические частоты	n_1	n_2	…	n_i	…	n_k
Теоретические частоты	$\widehat{n}_1$	$\widehat{n}_2$	…	$\widehat{n}_i$	…	$\widehat{n}_k$

За меру отклонения эмпирической функции распределения от теоретической принимают величину $D = \sum_{i=1}^{k} \frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$.

Расчеты показывают, что распределение этой случайной величины является χ^2-распределением. Обозначив np_i символом $\widetilde{n}_i$, запишем:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - \widehat{n}_i)^2}{\widehat{n}_i}. \tag{14.1}$$

Число степеней свободы $\nu = k - 1 - r$, где k – число интервалов вариационного ряда, а r – число параметров предполагаемого распределения, которые оценены по данным выборки (например, для нормального распределения оценивают два параметра – μ и S^2).

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. χ^2-критерий отвечает на вопрос, случайно или нет расхождение частот. Как любой критерий, χ^2-критерий не доказывает справедливость гипотезы, а лишь с определённой вероятностью α устанавливает её согласие или несогласие с данными наблюдения.

Как мы уже отмечали, критические точки χ^2-распределения приведены в таблице С Приложения. Они находятся по заданному уровню значимости α и числу степеней свободы.

Рассмотрим, например, распределение респондентов по их оценке признака: «очень важный», «важный», «маловажный», «совсем не важный» (табл. 14.2, верхняя строка).

Таблица 14.2. Эмпирические и теоретические частоты распределения респондентов в зависимости от выбранных ими ответов

	Очень важный	Важный	Мало - важный	Совсем не Важный	Итого
n_i	8	12	10	6	36
$\widehat{n}_i$	9	9	9	9	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 14.2.

Нулевая гипотеза H_0 заключается в том, что мера отклонения эмпирического распределения от равномерного равна нулю. Альтернативная гипотеза: отклонение больше нуля (естественно, что оно не может быть отрицательным), и, следовательно, критерий односторонний. При $\alpha = 0{,}05$ и $\nu = 4–1=3$ ${}_{0,05}\chi_3^2$=7,81. В то же время, как показывают расчеты,

$$\widehat{\chi}^2 = \frac{(8-9)^2}{9} + \frac{(12-9)^2}{9} + \frac{(10-9)^2}{9} + \frac{(6-9)^2}{9} = 2{,}22\,.$$

Таким образом, наблюдаемое значение $\widehat{\chi}^2$ не попадает в критическую область, следовательно, данные наблюдений согласуются с нулевой гипотезой и не дают оснований её отвергнуть.

χ^2-критерий часто применяют и для проверки гипотезы об отсутствии различий в мнениях респондентов, сгруппированных по тому или иному признаку.

Пример. Для изучения отношения учащихся к новой форме проведения уроков по физическому воспитанию случайным образом опросили 120 учащихся школы. Результаты опроса приведены в таблице 14.3.

Рассмотрим нулевую гипотезу H_0 об отсутствии в генеральной совокупности значимых различий в мнениях учащихся указанных трёх возрастных групп. Уровень значимости α примем равным 0,05. Для нахождения ожидаемой частоты (частоты, которую имело бы распределение, если бы таких различий не было вовсе) для каждой клетки таблицы перемножим соответствующие маргинальные частоты (частоты, представленные в итоговом столбце и итоговой строке) и разделим на общее число респондентов (120).

Таблица 14.3. 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 представлены в таблице 14.4. $\tilde{\chi}^2$=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 14.4. Схема вычисления χ^2

Ячейки таблицы 58	Частота n_i	Ожидаемая частота $\hat{n}_i$	$n_i - \hat{n}_i$	$(n_i - \hat{n}_i)^2$	$\frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$
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 $\hat{\chi}^2 > {}_{0,05}\chi_6^2$, 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.

Для корректного применения методов, основанных на использовании критерия χ^2, 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 категории, либо обратиться к другому критерию.

14.3. 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 χ^2-критерия. Новым в этом случае является необходимость расчета теоретических частот стандартного нормального распределения. Обычно и в этом случае прибегают к помощи специальной расчетной таблицы.

Обратимся к примеру. Пусть в таблице 14.5 приведено распределение 40 учащихся по количеству набранных при тестировании баллов.

Таблица 14.5. 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 не отличается от нормального, т.е. $\left|F_n^*(x) - F(x)\right| = 0$ – альтернативная гипотеза – $\left|F_n^*(x) - F(x)\right| > 0$. В качестве статистического критерия вновь примем статистику $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{\tilde{n}_i}$, где n_i – экспериментальные, а $\hat{n}_i$ – теоретические частоты.

Для вычисления фактического значения критерия, обратимся к таблице 14.6. При этом учтём, что в соответствии с выборкой, приведенной в таблице 14.5 $\bar{x}$=5,5 и S=1,948.

Заполнение расчетной таблицы 14.6 проведем в следующем порядке:

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

б) пользуясь таблицей N Приложения и четностью функции $\varphi(z) = \dfrac{1}{\sqrt{2\pi}} \cdot e^{-0,5z^2}$, занесем ее значения в пятый столбец таблицы 14.6;

в) для получения теоретической частоты все значения пятого столбца умножим на $R = \dfrac{nb}{s} = \dfrac{40 \cdot 1}{1.948} = 20,53$ (см. раздел). 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.

Таблица 14.6. Расчет теоретических частот и статистики $\hat{\chi}^2$

1	2	3	4	5	6	7	8	9	10
x	n	$x-\bar{x}$	$\dfrac{x-\bar{x}}{s} = z$	$\varphi(z)$	$R\ \varphi(z)$	$\hat{n}$	$n-\tilde{n}$	$(n-\hat{n})^2$	$\dfrac{(n-\hat{n})^2}{\hat{n}}$
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

Как отмечалось выше, число степеней свободы $\nu = k - 1 - r$. Так как k=6. и при вычислении $\varphi(z)$ 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 $\hat{\chi}^2 = 1,74$ не попадает в критическую область, то нуль-гипотеза принимается на уровне статистической значимости $\alpha = 0,05$.

14.4. Критерий согласия Колмогорова–Смирнова. Критерий согласия Колмогорова (1942) и Смирнова (1948) служит для проверки согласия наблюдаемого и теоретического распределений. Это непараметрический критерий; он соответствует χ^2- критерию согласия. В особенности хорошо критерий Колмогорова – Симонова (К–С-критерий) обнаруживает отклонения от нормального закона при малых объемах выборок. Нерегулярность распределения, как правило, лучше устанавливать с помощью χ^2-критерия, а отклонения формы распределения – с помощью К–С-критерия. Этот критерий предполагает непрерывное распределение, хотя он применяется и при дискретных распределениях.

Рассмотрим процедуру применения К–С-критерия. Пусть высказана ненаправленная гипотеза, согласно которой выборка относится к известному распределению $F_0(x)$. Как и при использовании χ^2- критерия, случайные значения выборки объема n распределяют по k интервалам (классам) таким образом, чтобы внутри интервалов наблюдаемые значения распределялись случайным образом независимо друг от друга. Затем, подсчитав наблюдаемые частоты n_i, строят функцию накопленной частоты $N_L = \sum_{i=1}^{L} n_i$. Аналогично поступают и с гипотетической функцией $F_0(x)$. Подсчитав соответствующие теоретические частоты $\hat{n}_i$, строят функцию накопленной частоты $\hat{N}_L = \sum_{i=1}^{L} \hat{n}_i$. Располагая эмпирической и теоретической функциями накопленных частот (кумулятами), составляют распределение абсолютных значений их разностей $\left|\hat{N}_L - N_L\right|$.

С помощью статистики

$$\frac{\max\left|\hat{N}_L - N_L\right|}{n} = D \qquad (14.2)$$

и проверяют нулевую гипотезу о том, что выборка относится к известному распределению $F_0(x)$.

При объеме выборки $n > 35$ нулевую гипотезу проверяют с помощью данных в таблице 14.7 критических значений

Таблица 14.7. Критические значения К–С-критерия для $n > 35$`.

Уровни значимости	0,20	0,15	0,10	0,05	0,01
Границы для D	$1{,}07/\sqrt{n}$	$1{,}14/\sqrt{n}$	$1{,}22/\sqrt{n}$	$1{,}36/\sqrt{n}$	$1{,}63/\sqrt{n}$

Числители в показателях для границ вычисляются по формуле $K_\alpha = \sqrt{-0{,}5 \cdot \ln(\alpha/2)}$, например, при α=0,10; $K_\alpha = \sqrt{(-0.5)\ln(0.05)}$=1,22.

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.

Если для согласования с нормальным распределением оценивались выборочные значения среднего и дисперсии, то выводы на основании таблицы 14.7 будут очень консервативными. Точные границы для К–С-критерия предложены Лиллиефорсом (1967), согласно которому для $n > 30$ справедливы выражения $0{,}805/\sqrt{n}$ $(\alpha = 0.10)$ и $0.886/\sqrt{n}$ $(\alpha = 0.05)$.

Обратимся снова к примеру, рассмотренному в пункте 14.3 о распределение 40 учащихся по количеству набранных при тестировании баллов (табл. 14.5). Данные о n_i и $\widehat{n}_i$ извлечем из таблицы 14.6.

Таблица 14.8. Расчетная таблица для проверки нулевой гипотезы с помощью К–С-критерия

n_i	0	2	5	4	10	8	6	2	1	2
$\widehat{n}_i$	0,57	1,64	3,64	6,15	8,00	8,00	6,15	3,64	1,64	0,57
$\widehat{N}_L$	0	2	7	11	21	29	35	37	38	40
N_L	0,57	2,21	5,85	12,00	20,00	28,00	34,15	37,79	39,43	40
$\lvert \widehat{N}_L - N_L \rvert$	0,57	0,21	1,15	1,00	1,00	1,00	0,85	0,79	1,43	0

$$\widehat{D} = \frac{\max\lvert \widehat{N}_L - N_L \rvert}{n} = \frac{1{,}43}{40} = 0{,}03575.$$

Так как для согласования с нормальным распределением оценивались выборочные значения среднего и дисперсии, то для нахождения границ для К–С-критерия воспользуемся правилом Лиллиефорса. На уровне значимости $\alpha = 0{,}05$ критическая точка одностороннего критерия равна $0{,}886/\sqrt{40}$=0,14. Так как $\widehat{D} < 0.14$,то, как и в случае применения χ^2-критерия (см. пункт 14.3), нуль-гипотеза принимается на 5%-ном уровне значимости.

Глава 15. Выводы относительно долей генеральной совокупности. Таблицы сопряженности

15.1. Доля как параметр, характеризующий совокупность. В ряде случаев результаты педагогического эксперимента оцениваются не системой набранных баллов, а в дихотомической шкале, долей (процентом) испытуемых, выполнивших (или не выполнивших) контрольное задание. В общем случае *долей* в статистике называют отношение числа элементов совокупности (k), обладающих некоторым заданным свойством, к общему числу (n) элементов совокупности, иначе говоря, *доля* – это относительная частота. Долю в генеральной совокупности будем обозначать буквой π, а в выборке буквой p (p=k/n).

С помощью статистики p, вычисленной по выборке, оценивается параметр π. Если оценивается дихотомическая переменная X, равная единице, когда изучаемый элемент обладает рассматриваемой характеристикой, и равная нулю, когда не обладает, то p соответствует π так же, как $\bar{x}$ соответствует μ. Например, пусть измеряемой характеристикой будет «пол», который обозначим X со значениями 1 – для мальчиков и 0 – для девочек. Если в выборке, содержащей n детей, присутствует k мальчиков, то $\sum_{i=1}^{n} x_i = k$. Следовательно, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = k/n = p$, генеральным средним X будет π. Итак, p как оценка π обладает всеми свойствами, присущими $\bar{x}$, как оценки μ.

Покажем, что в данной ситуации дихотомическая переменная X имеет генеральную дисперсию, равную $D[x] = p(1-p)$, где p – доля элементов, которым приписана единица.

Действительно, пусть в совокупности из n элементов k элементов имеют значение 1 и (n - k) – значение 0. Тогда $p = \bar{x} = k/n$. Дисперсия

$$D[x] = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - p)^2 = \frac{1}{n}\cdot\sum_{i=1}^{k}(1-p)^2 + \frac{1}{n}\cdot\sum_{i=1}^{n-k}(0-p)^2 = \frac{k}{n}(1-p)^2 + \frac{n-k}{n}p^2 =$$

$$= p(1-p)^2 + (1-p)p^2 = (1-p)(p - p^2 + p^2) = p(1-p).$$ Откуда $\sigma[x] = \sqrt{p(1-p)}$.

Можно доказать, что при достаточно больших значениях n (np и $n(1-p)$ должно быть больше пяти) выборочное распределение p можно считать нормальным со средним выборочных значений p (M[p]) и стандартным отклонением $\sigma[p] = \sqrt{p(1-p)/n}$.

Простой способ построения доверительного интервала для относительной частоты (доли) генеральной совокупности основан на предположении о нормальном распределении выборочных долей р. Он может быть использован только тогда, когда объем выборки достаточно велик и нет ни слишком больших, ни слишком малых относительных частот, так что np и n(1-p) больше 5. Доверительный интервал тогда в первом приближении определяется следующим выражением:

$$\widehat{p} \pm {}_{\alpha/2}Z \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}. \qquad (15.1)$$

Умножив все значения $\widehat{p}$ на 100%, получим доверительный интервал, выраженный в процентах.

Пример. Предположим, что в выборке из 70 мужчин, систематически курящих, оказалось 42 человека. Тогда относительная частота (доля) курящих $\widehat{p} = 42/70 = 0{,}6$, или 60%. В предположении нормальности распределения числа курящих мужчин 95%-ный доверительный интервал ($\alpha = 0{,}05$) имеет вид $0{,}6 \pm 1{,}96 \cdot \sqrt{\frac{0{,}6(1-0{,}6)}{70}}$ =0,6 ± 0,116, т.е. (0,484; 0,716), или в процентах (48,4%; 71,6%).

15.2. Проверка гипотезы $\pi = a$. Рассматривается гипотеза, заключающаяся в том, что доля π равна a, где a – некоторое число, заключённое между нулём и единицей.

Приступая к проверке H_0: $\pi = a$ против H_1: $\pi \neq a$, вспомним, что при выполнении нуль-гипотезы $\mu = a$, выборочное распределение $\bar{x}$ будет нормальным распределением со средним μ и стандартным отклонением $\sigma_X / \sqrt{n}$. То же самое справедливо для p. Для выборок, объёма n среднее арифметическое выборочного распределения p равно a – доле в генеральной совокупности, а стандартное отклонение:

$$\sigma[p] = \frac{\sigma_\pi}{\sqrt{n}} = \frac{\sqrt{a(1-a)}}{\sqrt{n}} = \sqrt{a(1-a)/n} \ .$$

Если справедлива нулевая гипотеза H_0: π=a и если na и $n(1-a)$ больше пяти, то величина

$$z = \frac{p-a}{o[p]} = \frac{p-a}{\sqrt{a(1-a)/n}} \qquad (15.2)$$

распределена приближенно нормально для случайных выборок объёма n с нулевым средним и единичным стандартным отклонением.

Пример. Исследуется вопрос о «доле» студентов института, которые вынуждены совмещать учёбу с работой. Предварительные (пилотажные) обследования привели исследователя к мысли, что эта доля близка к десяти процентам. Проверяется гипотеза $H_0: \pi = 0{,}1$ против альтернативной гипотезы $H_1: \pi \neq 0{,}1$.

Задавшись уровнем значимости $\alpha = 0{,}05$ и приняв объем выборки равным $n = 100$, вычислим критические точки критерия: $_{0,975}z = -1{,}96$ и $_{0.025}Z = 1.96$.

Отобрав случайным образом 100 студентов и установив, что среди них всего 8 человек совмещают учебу с работой, определили тем самым выборочную долю p, которая оказалась равной 0,08. Пользуясь теперь статистическим критерием (15.2), применимость которого вытекает из того, что ($na = 10 > 5$), вычислим значение статистики:

$$z = \frac{0{,}08 - 0{,}10}{\sqrt{0{,}10 \cdot 0{,}90/100}} = -0{,}67\,.$$

Сравнивая это значение с вычисленными выше критическими значениями стандартизированного нормального распределения, заключаем, что на уровне статистической значимости $\alpha = 0{,}05$ нет оснований отклонять нулевую гипотезу.

15.3. Проверка гипотезы о равенстве долей (относительных частот) генеральной совокупности по независимым выборкам. Рассматриваются две большие совокупности, в которых доли лиц (элементов), обладающих некоторым характеристическим свойством, равны, соответственно, π_1 и π_2. Проверяется гипотеза $H_0: \pi_1 = \pi_2$ против гипотезы $H_1: \pi_1 \neq \pi_2$.

Для проверки нулевой гипотезы из данных совокупностей извлекаются две независимые выборки: из первой – n_1 элементов, из второй – n_2 элементов. Предположим, что число элементов, обладающих характеристическим свойством, в первой выборке оказалось k_1, а во второй – k_2 и, следовательно, доли $p_1 = k_1/n_1$, $p_2 = k_2/n_2$.

Гипотеза H_0 проверяется с помощью статистики:

$$z = \frac{p_1 - p_2}{\sqrt{k(1-k)(1/n_1 + 1/n_2)}}\ , \qquad (15.3)$$

где $k = (k_1 + k_2)/(n_1 + n_2)$ – доля в обеих выборках.

Если нулевая гипотеза верна, и $n_1 p_1$, или $n_1(1-p_1)$ больше пяти, и $n_2 p_2$, или $n_2(1-p_2)$ тоже больше пяти, то z в выражении (15.3) имеет стандартное нормальное распределение.

Пример. Из учащихся шести восьмых классов школы случайным образом отбираются две группы по 50 человек каждая.

Учащиеся первой группы изучают материал по пособию, в котором определение предшествует примерам, учащиеся второй группы – по пособию, в котором определение даётся 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 формуле (15.3). 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.

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

15.4. Проверка гипотезы о равенстве долей (относительных частот) генеральной совокупности по зависимым выборкам. Проверяемая гипотеза та же, что и в разделе 15.3, а именно H_0: $\pi_1 = \pi_2$ против гипотезы H_1: $\pi_1 \neq \pi_2$. Однако на этот раз не требуется, чтобы извлекаемые из генеральной совокупности выборки были независимыми.

Как и во всех методах, связанных с зависимыми выборками, необходимо составить «пары» наблюдений, так что их число в выборках должно быть одинаковым. Так как элементы этих пар могут принимать только два значения, 0 и 1, то 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. 15.1).

В ней n – общее число элементов в выборке; k_1 – число единиц в первой выборке; k_2 – число единиц во второй выборке.

Таблица 15.1. Таблица сопряженности 2х2

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

.

Для проверки H_0 против H_1 можно применить статистику

$$z = \frac{d-a}{\sqrt{d+a}} \quad , \tag{15.4}$$

где a и d – число элементов выборки, которые изменили свою позицию.

Когда справедлива гипотеза H_0: $\pi_1 = \pi_2$, статистика z из уравнения (15.4) распределена приближённо нормально с нулевым средним и единичным стандартным отклонением, при условии, что $a+d > 10$.

Пример. Группе молодых людей ($n = 60$) предлагается указать до и после прослушивания лекции о вреде курения, одобряют они курение или нет. Результаты приведены в таблице (15.2).

Таблица 15.2. Отношение респондентов к курению

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 (15.4)

$$\hat{z} = \frac{10-26}{\sqrt{10+26}} = -2{,}67\,.$$

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.

15.5. Анализ таблиц сопряженности 2х2. Как отмечалось в начале этой главы, относительные частоты (доли) приходится сравнивать в случаях, когда результаты педагогического эксперимента выражаются в дихотомической шкале. Пусть, например, сопоставляются два метода проведения лабораторных занятий. Результаты «измеряются» долей учащихся, успешно справляющихся с заданием в отведенное время. В контрольной группе, занятия в которой проводились традиционными методами, успешно справились с заданием 85 учеников из 100. В контрольной группе, в которой занятия проводились инновационными методами, успешно справились с заданием 77 учеников из 81. Возникает вопрос: являются ли инновационные методы проведения лабораторных занятий более эффективными или успех имеет случайный характер.

Для ответа на этот вопрос обычно проводится классификация n объектов по двум парам признаков на 4 класса. Наблюдаемые частоты a, b, c, d заносят в четырехклеточную таблицу (табл. 15.3). Граничные случаи, когда результат можно отнести к обоим возможным классам, приводят к половинным значениям. Обе выборки из альтернативных данных исследуются на возможность их рассмотрения в виде случайных выборок из одной генеральной совокупности, представленной общей суммой. Другими словами, выясняется, имеют ли отношения a/c и b/d только случайные отклонения от среднего отношения $(a+b)/(c+d)=n_1/n_2$, можно ли рассматривать отношения a/b и c/d как случайные отклонения от среднего отношения $(a+c)/(b+d)$.

Вышеприведенный пример приводит к четырехклеточной таблице (табл. 15.3), предназначенной для выяснения вопроса, является ли более высокий уровень выполнения лабораторных заданий в инновационных условиях случайным.

Нуль-гипотеза: процент учащихся, успешно выполнивших задание, не зависит от метода проведения занятия. Иначе говоря, обе группы учащихся, и контрольная, и экспериментальная, относятся к общей генеральной совокупности. Выполнение этой гипотезы будет означать, что эффект обучения в обоих случаях одинаков.

Таблица 15.3. Сопоставление результатов двух методов проведения лабораторных занятий, измеренных в дихотомической шкале

Группы	Учащиеся		Итого
	выполнили	не выполнили	
Контрольная	$a=85$	$b=15$	100
Экспериментальная	$c=77$	$d=4$	81
Итого	162	19	181

В соответствии с данными, приведенными в таблице 15.4, доля учащихся контрольной группы, своевременно выполнивших задание, равна 0,850, а в экспериментальной – 0,959. Возникает вопрос, как оценить значимость этого различия. С этой целью прежде всего выясним, какими должны были бы быть частоты a, b, c и d (обозначим их $\widehat{a}$, $\widehat{b}$, $\widehat{c}$, $\widehat{d}$), чтобы при тех же маргинальных частотах $a+b=100$, $c+d=81$, $a+c=162$, $b+d=19$ рассматриваемые доли были равны, т.е.

$$\frac{a}{a+b}=\frac{c}{c+d}=\frac{a+c}{a+b+c+d}.$$

Легко показать, что для этого необходимо и достаточно, чтобы каждое из значений $\widehat{a}$, $\widehat{b}$, $\widehat{c}$, $\widehat{d}$ равнялось отношению произведения маргинальных частот, соответствующих данному символу, к общему числу учащихся, например, $\widehat{a}=\dfrac{(a+b)(a+c)}{a+b+c+d}=\dfrac{100\cdot 162}{181}\approx 89{,}5$.

Аналогично вычисляются остальные частоты: $\widehat{b}=10{,}5$; $\widehat{c}=72{,}5$; $\widehat{d}=8{,}5$. Таблица сопряженности 15.4, соответствующая этим значениям, будет иметь вид:

Таблица 15.4. Теоретические частоты, вычисленные в предположении равенства долей учащихся контрольной и экспериментальных групп, выполнивших задание в срок

Группы	Учащиеся		Итого
	выполнили	не выполнили	
Контрольная	$\widehat{a}=89{,}5$	$\widehat{b}=10{,}5$	100
Экспериментальная	$\widehat{c}=72{,}5$	$\widehat{d}=8{,}5$	81
Итого	162	19	181

Нуль-гипотеза о равенстве долей в обеих группах, или независимости обоих альтернативных признаков, проверяется с помощью χ^2- критерия.

$$\chi^2=\frac{(a-\widehat{a})^2}{\widehat{a}}+\frac{(b-\widehat{b})^2}{\widehat{b}}+\frac{(c-\widehat{c})^2}{\widehat{c}}+\frac{(d-\widehat{d})^2}{\widehat{d}}$$

Так как $(a-\widehat{a})^2=(b-\widehat{b})^2=(c-\widehat{c})^2=(d-\widehat{d})^2=(ad-bc)^2/n^2=\Delta^2$,

то
$$\chi^2=\Delta^2\left(\frac{1}{\widehat{a}}+\frac{1}{\widehat{b}}+\frac{1}{\widehat{c}}+\frac{1}{\widehat{d}}\right) \qquad (15.5)$$

Откуда, развернув каждое из значений $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, $\tilde{d}$ и выполнив преобразования, получим:

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}, \tag{15.6}$$

где $n = a + b + c + d$.

Четырёхклеточный χ^2– критерий имеет *только одну степень свободы.* Обе формулы (15.5) и (15.6) могут употребляться только тогда, когда все ожидаемые частоты больше 3, а общее число испытуемых $n > 20$.

При малых значениях n необходимо в (15.6) заменить n на $n-1$. Эта формула применима, если $n_1 > 5$ и $n_2 > n_1/3$.

Нуль-гипотеза о независимости или однородности отклоняется, если значение χ^2, рассчитанное по формулам (15.5), (15.6) по крайней мере, равно критическому значению χ^2 из следующей таблицы.

Таблица 15.5. Критические значения χ_1^2, используемые при работе с четырехклеточными таблицами

Вероятность ошибки α	0,05	0,01	0,001
Двусторонний критерий ($H_0 : \pi_1 = \pi_2,\ H_1 : \pi_1 \neq \pi_2$)	3,841	6,635	10,827
Односторонний критерий ($H_0 : \pi_1 = \pi_2, H_1 : \pi_1 > \pi_2$ *или* $\pi_2 > \pi_1$)	2,706	5,412	9,550

Обычно используется двусторонний критерий.

Пользуясь формулой 15.6 и данными таблицы 15.3, вычислим χ^2.

$$\chi^2 = \frac{181(85 \cdot 4 - 77 \cdot 15)^2}{162 \cdot 19 \cdot 100 \cdot 81} = 4{,}822.$$

Завершая рассмотрение примера, приведенного в начале этого пункта, заметим, что на 5%-ном уровне (при одностороннем критерии) $\chi^2 = 4{,}822 >_{0{,}05} \chi_1^2 = 2{,}706$ и, следовательно, нуль-гипотеза о случайном характере различий в долях успешно справившихся с заданием должна быть отвергнута.

15.6. Проверка гипотезы о независимости классификаций в таблице сопряжённости признаков. Часто производятся наблюдения, допускающие измерение каждого элемента (человека, группы и т.д.) в соответствии с двумя видами классификации. Например, студентов можно классифицировать как по полу (мужской - женский), так и по политической ориентации (правый, левый, центрист).

В таблице сопряжённости 15.6 в качестве гипотетического примера приведена такая классификация 115 человек.

Таблица 15.6. Распределение 115 респондентов по двум показателям: полу и политической ориентации

Пол	Политическая ориентация			Итого
	левые	правые	центристы	
Мужской	25	28	12	65
Женский	18	13	19	50
Итого	43	41	31	115

Если подобную таблицу записать с использованием буквенной символики (табл. 15.7), то для проверки нулевой гипотезы об однородности (пропорциональности) используется χ^2 критерий:

Таблица 15.7. Общий вид таблицы сопряженности 3х2

Первый признак	Второй признак			Итого
	A	B	C	
P	a_{11}	a_{12}	a_{13}	$n_1 = a_{11} + a_{12} + a_{13}$
Q	a_{21}	a_{22}	a_{23}	$n_2 = a_{21} + a_{22} + a_{23}$
Итого	$m_1 = a_{11} + a_{21}$	$m_1 = a_{11} + a_{21}$	$m_3 = a_{13} + a_{23}$	n

$$\chi^2 = n\left(\frac{a_{11}^2}{n_1 m_1} + \frac{a_{12}^2}{n_1 m_2} + \frac{a_{13}^2}{n_1 m_3} + \frac{a_{21}^2}{n_2 m_1} + \frac{a_{22}^2}{n_2 m_2} + \frac{a_{23}^2}{n_2 m_3} - 1\right). \qquad (15.7)$$

В более общем случае, если в таблице k строк и l столбцов, формула приобретает вид:

$$\chi^2 = n\left[\sum_{i=1}^{k}\sum_{j=1}^{l}\frac{a_{ij}^2}{n_i m_j} - 1\right]. \qquad (15.8)$$

В соответствии с приведёнными в таблице 14.9 данными

$$\chi^2 = 115\cdot\left[\frac{25^2}{65\cdot 43} + \frac{28^2}{65\cdot 41} + \frac{12^2}{65\cdot 31} + \frac{18^2}{50\cdot 43} + \frac{13^2}{50\cdot 41} + \frac{19^2}{50\cdot 31} - 1\right] = 6{,}36.$$

Если нуль-гипотеза о независимости двух классификаций верна, то статистика χ^2 в формуле (15.8) имеет распределение хи-квадрат с числом степеней свободы $\nu = (k-1)(l-1)$. В рассматриваемом примере

$\nu = (2-1)(3-1) = 2$. На уровне значимости $\alpha = 0{,}05$ ${}_{0,975}\chi_2^2 = 0{,}051$ ${}_{0,025}\chi_2^2 = 7{,}38$. Так как выборочное значение $\hat{\chi}^2$=6,36 лежит между этими значениями, то на уровне значимости $\alpha = 0{,}05$ нет оснований отклонять нуль-гипотезу о пропорциональности двух классификаций.