

6. КРИТЕРИИ ЗНАЧИМОСТИ И ПРОВЕРКА ГИПОТЕЗ

6.1. Введение

В этой главе рассматривается группа статистических методов, которые получили наибольшее распространение в статистических исследованиях, связанных с правоведением. Эти методы применяются всегда, когда предстоит проверить какие-то теоретические предположения, связанные с эффективностью мероприятий, направленных на совершенствование какого-либо процесса. Исследователь выдвигает предположения исходя из анализа конкретного явления с позиций педагогики, физиологии, медицины, психологии или другой области знаний, представителем которой он является. Затем справедливость предположений проверяется на основании данных соответствующего эксперимента, условия которого контролируются.

6.2. Основные понятия и определения

6.2.1. Нулевая гипотеза (нуль-гипотеза) и альтернатива (альтернативная гипотеза)

Статистической гипотезой (или просто гипотезой) называется утверждение о распределении генеральной совокупности, соответствующее некоторым представлениям об изучаемом явлении. В частном случае это может быть утверждение о значениях параметров (σ и μ) нормально распределенной генеральной совокупности.

Предположим, что в эксперименте участвуют две группы заключенных. Одна из них (контрольная) содержится и перевоспитывается по традиционной программе, а для второй (экспериментальная) используются новые методы. Действенность нового комплекса оценивается по различию уровня агрессивности, показанных в этих группах после определенного срока. По полученным данным необходимо проверить следующие утверждения:

1. Среднее значение уровня агрессивности не изменилось, т. е.

$$\mu_1 = \mu_2.$$

Здесь μ_1 и μ_2 – средние значения соответствующих генеральных совокупностей (уровни агрессивности всех аналогичных заключенных, которые могли бы перевоспитываться по традиционной (μ_1) и новой (μ_2) программам).

2. Вариативность агрессивности возросла:

$$\sigma_2 > \sigma_1.$$

Здесь σ_2 и σ_1 — так же, как и в п. 1, значения соответствующих генеральных параметров.

3. Средняя агрессивность возросла на 3 единицы:

$$\mu_1 - \mu_2 = 3$$

Это три различные статистические гипотезы. Конечно, возможные утверждения не ограничиваются приведенным списком. Гипотезы предстоит проверить с помощью какого-то метода — критерия.

Статистические гипотезы обычно рассматривают, генеральные совокупности, одна из которых может представлять собой теоретическую модель (например, нормальное распределение), а о второй судят по выборке из нее. В других случаях обе генеральные совокупности представлены выборками.

При проверке статистических гипотез принят следующий подход. Считается, что получение в результате эксперимента любых новых данных об изучаемом явлении, не согласующихся с данными, имеющимися до проведения эксперимента, — маловероятное событие. В то же время, если взять две выборки, представляющие собой результаты измерения одного и того же признака, и сравнить между собой их характеристики (среднее арифметическое, стандартное отклонение и др.), то окажется, что они практически всегда различаются. Это различие можно рассматривать как обусловленное только действием случайностей. Поэтому первоначально гипотезу всегда можно сформулировать таким образом: между двумя генеральными совокупностями нет ожидаемого различия.

Такая гипотеза называется *нулевой гипотезой*, или нуль-гипотезой. Обратное ей утверждение о том, что в действительности между генеральными совокупностями есть различие, называется *альтернативной гипотезой*, или альтернативой.

Нулевую гипотезу принято обозначать, как H_0 , а альтернативную — H_1 .

Итак, вначале выдвигается нулевая гипотеза о том, что различие между генеральными совокупностями равно нулю. Затем получают выборку или несколько выборок, и если выборочные данные не противоречат нулевой гипотезе, т. е. различие можно объяснить только случайностью выборки, то нулевая гипотеза

сохраняется (принимается). Если же полученные результаты не удается объяснить только действием случайных факторов, то нулевая гипотеза отвергается, а принимается альтернативная гипотеза.

Пусть, например, оценивается эффективность нового метода перевоспитания для заключенных по среднему значению агрессивности в контрольной и экспериментальной группах. Тогда нулевую гипотезу H_0 можно сформулировать так: среднее значение результатов не изменилось, т.е. $\mu_1 = \mu_2$. Для краткости это записывается так: $H_0: \mu_1 = \mu_2$.

Если заранее нельзя сказать, к чему приведет новый метод — к увеличению или уменьшению агрессивности, то альтернативная гипотеза H_1 будет состоять в том, что средние значения генеральных совокупностей неодинаковы: $H_1: \mu_1 \neq \mu_2$.

6.2.2. Ошибки при проверке гипотез

Ошибки, допускаемые при проверке гипотез, удобно разделить на два типа: 1) отклонение гипотезы H_0 , когда она верна, — ошибка первого рода; 2) принятие гипотезы H_0 , когда в действительности верна какая-то другая гипотеза, — ошибка второго рода.

Вероятность ошибки первого рода обозначается α . Величина α называется **уровнем значимости** критерия, по которому проверяется справедливость гипотезы H_0 .

Вероятность ошибки второго рода обозначается β . Ее величина зависит от альтернативной гипотезы H_1 . Рассмотрим для приведенного выше примера следующие две ситуации: 1) в действительности средняя агрессивность возросла на 3 единицы, 2) средняя агрессивность увеличилась на 30 единиц. Ясно, что для одних и тех же условий эксперимента и одинакового уровня значимости α вероятность ошибки второго рода β (принять гипотезу об отсутствии различия) для второй из альтернатив будет меньше.

Вероятности α и β удобно представить, как это сделано в табл. 6.1.

Таблица 6.1

Ошибки при проверке гипотез

	Решение	
	Принять H_0	Принять H_1
Справедлива H_0	Правильное с вероятностью $1 - \alpha$	Ошибочное с вероятностью α
Справедлива H_1	Ошибочное с вероятностью β	Правильное с вероятностью $1 - \beta$

Наглядным способом интерпретации ошибок является их графическое представление.

Предположим, что проверяется гипотеза $H_0: \mu_1 = \mu_0$ о равенстве среднего значения генеральной совокупности заданной величине μ_0 (известной, например, из предыдущих экспериментов).

Для этого берется выборка объема n , находится ее среднее арифметическое $\bar{x}_B$ и по его величине судят о справедливости гипотезы H_0 .

Распределение среднего арифметического $\bar{x}_B$ при условии, что верна гипотеза H_0 , будет $f(\bar{x}_B / \mu_0)$. Это распределение чисто качественно представлено на рис. 6.1.

Распределение среднего арифметического $\bar{x}_B$ при условии, что верна альтернативная гипотеза $H_1: \mu_1 \neq \mu_0$, будет уже другим — $f(\bar{x}_B / \mu_1)$.

Будем считать, что гипотеза H_0 отвергается, если выборочное среднее арифметическое $\bar{x}_B$ окажется больше некоторого значения $K_{\text{критич}}$, т. е. $\bar{x}_B > K_{\text{критич}}$, как показано на рис. 6.1.

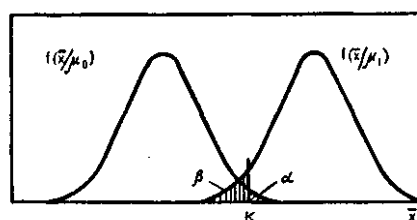


Рис. 6.1. Ошибки первого и второго рода

Область непринятия гипотезы H_0 называется **критической областью** критерия. Она показана на рис. 6.1 наклонной штриховкой. Уровень значимости α будет соответствовать площади критической области.

Вероятность ошибки второго рода β будет равна площади под кривой распределения $f(\bar{x}_B / \mu_1)$, показанной на рис. 6.1. вертикальной штриховкой.

Величина $1 - \beta$ называется **мощностью критерия**.

Следует особо подчеркнуть, что любая гипотеза должна формулироваться, а уровень значимости α задаваться исследователем, всегда до получения экспериментальных данных, по которым эта гипотеза будет проверяться.

При выборе уровня значимости α исследователь исходит из практических соображений, отвечая на вопрос: какую вероятность ошибки он считает допустимой для его конкретной задачи?

Обычно считают достаточным $\alpha = 0,05$ (5%), иногда $\alpha = 0,01$, редко $\alpha = 0,001$.

Между стандартными статистическими критериями и стандартными доверительными интервалами существует тесная связь: если принимается гипотеза о том, что значение параметра (μ , σ) нормально распределенной генеральной совокупности равно фиксированному значению (μ_0 , σ_0) с уровнем значимости α , то это эквивалентно заданию $100(1 - \alpha)\%$ -ного доверительного интервала для данного параметра нормального распределения. Поэтому оба подхода — доверительные интервалы и критерии значимости — в данном случае равноценны. Преимущество доверительных интервалов в том, что они дают представление об истинном значении параметра генеральной совокупности, а недостаток в том, что их трудно построить в более сложных случаях, например при анализе дисперсий (стандартных отклонений).

6.2.3. Критерии значимости

В рассмотренном выше примере (см. п. 6.2.2) при проверке гипотезы об отсутствии различия среднего уровня агрессивности в контрольной и экспериментальной группах можно было бы поступить следующим образом: вычислить средние арифметические результаты в группах и сравнить их между собой. Если окажется, что различие средних арифметических больше, например, 50 единиц, то можно утверждать, что новый комплекс оказался эффективным. Но при этом неизвестно, какие ошибки допускаются при таком утверждении, поэтому невозможно точно доказать наличие или отсутствие различий.

Методы, которые для каждой выборки формально точно определяются, удовлетворяют выборочные данные нулевой гипотезы или нет, называются *критериями значимости*.

Общая схема проверки гипотез

Процедура проверки гипотез обычно проводится по следующей схеме:

Формулируются гипотезы H_0 и H_1 .

Выбирается уровень значимости критерия α .

По выборочным данным вычисляется значение $K_{набл}$ некоторой случайной величины K , называемой статистикой критерия, или просто критерием, который имеет известное стандартное распределение (нормальное, Т-распределение Стьюдента и т.п.)

Вычисляется критическая область и область принятия гипотезы. То есть находится критическое (граничное) значение критерия $K_{критич}$ при уровне значимости α , взятым из соответствующих таблиц.

Найденное значение $K_{набл}$ критерия сравнивается с $K_{критич}$ и по результатам сравнения делается вывод: принять гипотезу или отвергнуть.

♦ Если вычисленное по выборке значение критерия $K_{набл}$ меньше чем $K_{критич}$, то гипотеза H_0 принимается на заданном уровне значимости α .

(В этом случае наблюдаемое по экспериментальным данным различие генеральных совокупностей можно объяснить только случайностью выборки. Однако принятие гипотезы H_0 совсем не означает доказательства равенства параметров генеральных совокупностей. Просто имеющийся в распоряжении статистический материал не дает оснований для отклонения гипотезы о том, что эти параметры одинаковы. Возможно, появится другой экспериментальный материал, на основании которого эта гипотеза будет отклонена.)

♦ Если вычисленное значение критерия $K_{набл}$ больше $K_{критич}$, то гипотеза H_0 отклоняется в пользу гипотезы H_1 при данном уровне значимости α .

(В этом случае наблюдаемое различие генеральных совокупностей уже нельзя объяснить только случайностями и говорят, что наблюдаемое различие значимо (статистически значимо) на уровне значимости α .)

Следует подчеркнуть разницу между статистической значимостью и практической значимостью. Заключение о практической значимости всегда делается человеком, изучающим данное явление. И здесь истинным критерием является опыт и интуиция исследователя, а статистические критерии значимости — лишь формально точный инструмент, используемый в исследовании. Чем больше исследователь знает об изучаемом явлении, тем точнее будет сформулированная им гипотеза и тем точнее будут выводы,

сделанные с помощью критериев значимости.

Замечание 1

Ранее уже подчеркивалось, что уровень значимости и должен выбираться исследователем до получения экспериментальных данных, по которым будет проверяться гипотеза. Но часто с предварительным выбором возникают затруднения. Обычно говорят, что для научных исследований (в том числе и в правоведении) достаточен уровень значимости $\alpha=0,05$, но если выводы, которые предстоит сделать по результатам проверки гипотез, связаны с большой ответственностью, то рекомендуется выбирать $\alpha=0,01$ или $\alpha=0,001$.

Как установить ответственность в трактовке результатов эксперимента и тот риск, который связан с выбором уровня значимости α ? Чтобы не давать прямых ответов на эти непростые вопросы, часто поступают следующим образом: уровень значимости до эксперимента точно не устанавливается, а по экспериментальным данным вычисляется вероятность P того, что критерий (статистика критерия) выйдет за пределы значения, рассчитанного по выборке. Таким образом, P — это экспериментальный уровень значимости. Точное значение P обычно не указывают, а окончательные результаты приводят в следующем виде: 1) если вычисленное значение критерия не превосходит критического значения на уровне значимости $\alpha=0,05$, то различие считается статистически незначимым; 2) если вычисленное по выборке значение критерия превышает критические значения при $\alpha=0,05$, $\alpha=0,01$ или $\alpha=0,001$, то записывают $P<0,05$, $P<0,01$ или $P<0,001$. Это означает, что наблюдаемые различия статистически значимы на уровнях значимости 0,05, 0,01 или 0,001.

Критерии значимости подразделяются на три типа:

1. Критерии значимости, которые служат для проверки гипотез о параметрах распределений генеральной совокупности (чаще всего нормального распределения). Эти критерии называются параметрическими.

2. Критерии, которые для проверки гипотез не используют предположений о распределении генеральной совокупности. Эти критерии не требуют знания параметров распределений, поэтому называются непараметрическими.

3. Особую группу критериев составляют критерии согласия, служащие для проверки гипотез о согласии распределения генеральной совокупности, из которой получена выборка, с ранее принятой теоретической моделью (чаще всего нормальным распределением).

6.2.4. Односторонние и двусторонние критерии

Если цель исследования том, чтобы выявить различие параметров двух генеральных совокупностей, которые соответствуют различным ее естественным условиям (условия жизни, возраст испытуемых и т. п.), то часто неизвестно, какой из этих параметров будет больше, а какой меньше. Например, если интересуются вариативностью результатов в контрольной и экспериментальной группах, то, как правило, нет уверенности в знаке различия дисперсий или стандартных отклонений результатов, по которым оценивается вариативность. В этом случае нулевая гипотеза состоит в том, что дисперсии равны между собой ($H_0 : \sigma_1^2 = \sigma_2^2$), а цель исследования — доказать обратное ($H_1 : \sigma_1^2 \neq \sigma_2^2$), т.е. наличие различия между дисперсиями. При этом допускается, что различие может быть любого знака. Такие гипотезы называются двусторонними.

Но иногда задача состоит в том, чтобы доказать увеличение или уменьшение параметра; например, средний результат в экспериментальной группе выше, чем контрольной. При этом уже не допускается, что различие может быть другого знака. Тогда альтернативная гипотеза $H_1 : \mu_2 > \mu_1$ (или $H_1 : \mu_2 < \mu_1$ а обратное ей утверждение $H_0 : \mu_2 \leq \mu_1$ (или $H_0 : \mu_2 \geq \mu_1$). Такие гипотезы называются односторонними.

Критерии значимости, служащие для проверки двусторонних гипотез, называются двусторонними, а для односторонних — односторонними.

Возникает вопрос о том, какой из критериев следует выбирать в том или ином случае. Ответ на этот вопрос находится за пределами формальных статистических методов и полностью зависит от целей исследования. Ни в коем случае нельзя выбирать тот или иной критерий после проведения эксперимента на основе анализа экспериментальных данных, поскольку это может привести к неверным выводам. Если до проведения эксперимента допускается, что различие сравниваемых параметров может быть как положительным, так и отрицательным, то следует использовать двусторонний критерий. Если же есть дополнительная информация, например, из предшествующих экспериментов, на основании которой можно сделать предположение, что один из параметров больше или меньше другого, то используется односторонний критерий. Когда имеются основания для применения одностороннего критерия, его следует предпочесть двустороннему, потому что односторонний критерий полнее использует информацию об

изучаемом явлении и поэтому чаще дает правильные результаты.

Например, необходимо доказать различие средних значений генеральных совокупностей (средних значений некоторого результата исследований) при двух различных методиках применяемых в контрольной и экспериментальной группах. Если есть данные, что экспериментальная группа покажет в среднем лучший результат, то нужно выдвинуть нулевую гипотезу $H_0: \mu_2 = \mu_1$ против двусторонней альтернативы $H_1: \mu_2 \neq \mu_1$. Различие доказывается по разности средних арифметических результатов в контрольной и экспериментальной группах $(\bar{x}_2 - \bar{x}_1)$. Распределение разности $\bar{x}_2 - \bar{x}_1$ при условии, что верна нулевая гипотеза H_0 схематично представлено на рис. 6.2, а.

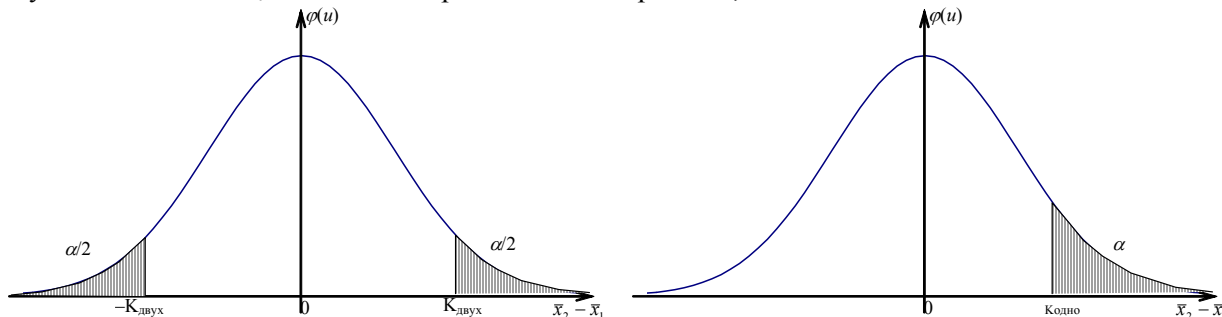


Рис. 6.2. Уровни значимости при двустороннем (а) и одностороннем (б) критериях

Решение об отклонении гипотезы H_0 принимается в том случае, если разность $\bar{x}_2 - \bar{x}_1$ выходит за пределы некоторого значения $K_{\text{двух}}$ (допустимы отклонения в обе стороны от нуля). Ошибка, которая при этом допускается, равна, как известно, уровню значимости α . Но поскольку отклонения возможны в обе стороны, то при симметричном распределении вероятности отклонений, больших $K_{\text{двух}}$ и меньших $K_{\text{двух}}$, будут одинаковы и составят $\alpha/2$.

Если предположить, что в экспериментальной группе будут показаны в среднем более высокие результаты, то можно выдвинуть одностороннюю альтернативу $H_1: \mu_2 > \mu_1$. В этом случае при той же нулевой гипотезе $H_0: \mu_2 = \mu_1$ распределение разности $\bar{x}_2 - \bar{x}_1$ будет таким же, как и для двустороннего критерия (см. рис. 6.2, б). Но теперь представляют интерес только положительные значения разности $\bar{x}_2 - \bar{x}_1$. Решение об отклонении H_0 принимается, когда $\bar{x}_2 - \bar{x}_1$ окажется больше некоторого $K_{\text{одн}}$. При том же уровне значимости α $K_{\text{одн}}$ будет всегда меньше $K_{\text{двух}}$, поэтому нулевая гипотеза будет при одностороннем критерии отклоняться чаще.

Таким образом, двусторонние критерии оказываются более консервативными, чем односторонние.

В этом нет никакого противоречия или доказательства несостоятельности статистических методов. Просто в первом случае, используя двустороннюю гипотезу, мы допускали и отрицательный эффект новой программы. В такой ситуации выводы должны быть более осторожными, чем в случае односторонней гипотезы, когда имеется дополнительная информация, позволяющая сделать предположение о положительном эффекте новой программы, что, естественно, дает возможность сделать более точный вывод. Правда, следует отметить, что если превышение критического значения в каком либо исследовании незначительно, то в достоверности вывода о наличии положительного эффекта можно усомниться. В такой ситуации следует провести дополнительные исследования.

6.3. Критерии согласия

Все рассмотренные ниже критерии значимости являются оптимальными, т. е. обеспечивают наивысшую достоверность статистических выводов только в тех случаях, когда выборки получены из *нормально распределенной генеральной совокупности*. При отклонениях от нормального распределения точность оптимальных критериев существенно падает, поэтому, чтобы уверенно применять оптимальные критерии, необходимо проверить предположение о нормальном распределении генеральной совокупности. Для этого используются критерии согласия. Здесь нулевая гипотеза H_0 представляет собой утверждение о том, что распределение генеральной совокупности, из которой получена выборка, не отличается от нормального. Существует несколько разновидностей критериев согласия. Рассмотрим те из них, которые получили наибольшее распространение на практике.

6.3.1. Предварительная проверка соответствия нормальному распределению

Критерии согласия требуют достаточно большой вычислительной работы, поэтому целесообразно перед тем, как их использовать, проверить с помощью более простых методов соответствие имеющихся экспериментальных данных нормальному распределению. Эти методы, естественно, обладают меньшей

мощностью и позволяют установить только значительные расхождения с нормальным распределением, но если такие расхождения будут установлены, то необходимость в применении более точных, но более сложных критериев, как правило, отпадает.

Для предварительной проверки эмпирического распределения на нормальность можно использовать основные свойства нормального распределения, изложенные ранее. При этом эмпирическое распределение представляется в виде вариационного ряда или гистограммы. Если в качестве параметров μ и σ нормального распределения принять их выборочные оценки $\bar{x}_B$ и S , то для проверки можно использовать следующие свойства нормального распределения: 1) практически все отклонения от среднего значения (99,7 %) должны быть меньше $\pm 3S$; 2) примерно 2/3 всех отклонений (68,3 %) должны быть меньше $\pm S$; 3) половина всех отклонений от среднего значения должна быть меньше $\pm 0,657S$; 4) можно использовать такое свойство нормального распределения, что его коэффициенты асимметрии γ_3 и эксцесса γ_4 равны нулю.

Значения коэффициентов A_s и E_x сравниваются с критическими значениями на уровне значимости α , и если критические значения превышены, то делается вывод о том, что распределение генеральной совокупности, из которой получена выборка, не согласуется с нормальным. В противном случае модель нормального распределения может быть принята. Таблица критических значений A_s и E_x содержится в различных книгах. Здесь не будем подробно останавливаться на этих приближенных критериях. Отметим лишь еще раз, что они могут использоваться только совместно с более точными критериями, рассмотренными ниже.

6.3.2. Критерий согласия χ^2 (хи-квадрат)

Критерий согласия χ^2 разработан лучше других критериев и чаще других используется. Он основан на сравнении эмпирических частот интервалов группировки с теоретическими (ожидаемыми) частотами, рассчитываемыми по формулам нормального распределения.

Условия применения: объем выборки $n \geq 40$, выборочные данные сгруппированы в интервальный вариационный ряд с числом интервалов не менее 7, ожидаемые (теоретические) частоты интервалов не должны быть меньше 5.

Гипотеза H_0 : $f(x) = f^{теор}(x)$ — плотность распределения $f(x)$ генеральной совокупности, из которой взята выборка, соответствует теоретической модели $f^{теор}(x)$ нормального распределения.

Альтернатива H_1 : $f(x) \neq f^{теор}(x)$.

Уровень значимости: α .

Порядок, применения:

1. Формулируется гипотеза, выбирается уровень значимости α .
2. Получается выборка объема $n \geq 40$ независимых наблюдений и представляется эмпирическое распределение в виде интервального вариационного ряда.
3. Рассчитываются выборочные характеристики $\bar{x}_B$ и S . Их используют в качестве генеральных параметров μ и σ нормального распределения, с которым предстоит сравнить эмпирическое распределение.
4. Вычисляются значения теоретических частот $n_i^{теор}$ попадания в i -й интервал группировки. Для этого необходимо вычислить:

$$n_i^{теор} = n \left[\Phi_0 \left(\frac{x_{vi} - \bar{x}_B}{S} \right) - \Phi_0 \left(\frac{x_{ni} - \bar{x}_B}{S} \right) \right]$$

где $\Phi_0(u)$ — функции Лапласа, x_{vi} и x_{ni} — верхняя и нижняя границы i -го интервала группировки.

Если окажется, что вычисленные ожидаемые частоты $n_i^{теор}$ некоторых интервалов группировки меньше 5, то соседние интервалы объединяются так, чтобы сумма их ожидаемых частот была больше или равна 5. Соответственно складываются и эмпирические частоты объединяемых интервалов.

5. Значение χ^2 -критерия рассчитывается по формуле:

$$\chi_{набл}^2 = \sum_{i=1}^k \frac{(n_i - n_i^{теор})^2}{n_i^{теор}}$$

где n_i — эмпирические частоты; $n_i^{теор}$ — ожидаемые (теоретические) частоты; k — число интервалов группировки после объединения.

6. Из таблиц распределения χ^2 находится критическое значение $\chi_{критич}^2$ критерия для уровня

значимости α и числа степеней свободы $\nu=n-3$.

7. Вывод: если $\chi^2_{\text{набл}} \geq \chi^2_{\text{критич}}$ то эмпирическое распределение не соответствует нормальному распределению на уровне значимости α , в противном случае нет оснований отрицать это соответствие.

6.4. Критерии, основанные на нормальном распределении

6.4.1. Сравнение выборочного среднего арифметического с известным средним значением генеральной совокупности

Рассмотрим, как с помощью статистических критериев решить вопрос: значимо ли отличие выборочного среднего значения от среднего значения генеральной совокупности, из которой предположительно взята выборка, или наблюдаемое различие является случайным? Такая постановка вопроса типична для выборочного контроля качества продукции в промышленности, но и при исследованиях в правоведении такой вопрос часто возникает, когда предстоит решить, значимо ли отличается среднее значение признака, полученное по выборке, от среднего значения, известного по результатам многочисленных предыдущих экспериментов.

Применяемый для этих целей t-критерий Стьюдента основан на предположении о нормальности распределения генеральной совокупности, но результаты проверки гипотез удовлетворяют по точности и при небольших отклонениях от нормальности распределения.

Условия применения t-критерия: выборка получена из генеральной совокупности, имеющей приближенно нормальное распределение с параметрами μ и σ .

Гипотеза $H_0: \mu = \mu_0$ – среднее значение μ генеральной совокупности, из которой получена выборка, равно данному значению μ_0 (известному, например, из предыдущих экспериментов).

Альтернатива $H_0: \mu \neq \mu_0$ (двусторонний критерий применяется тогда, когда допускаются отклонения в обе стороны от μ_0).

Уровень значимости: α .

Порядок применения Т-критерия:

1. Принимается предположение о нормальности, формулируются гипотезы H_0 и H_1 задается уровень значимости α .
2. Получают выборку объема n .
3. Вычисляется выборочное среднее арифметическое $\bar{x}_B$ и исправленная выборочная дисперсия S^2 .
4. Определяется значение t-критерия по формуле:

$$t_{\text{набл}} = \frac{|\bar{x}_B - \mu_0|}{S / \sqrt{n}}$$

Величина t имеет при справедливости гипотезы H_0 t-распределение Стьюдента (определенное в гл. 4) с $\nu=n-1$ степенями свободы.

5. По таблицам находится $t_{\text{критич}}$ – критическое значение t-критерия при уровне значимости α и числе степеней свободы $\nu=n-1$. (Таблицы обычно содержат критические значения $t_{\text{критич}}$ для двустороннего критерия.)

6. Делается вывод: если $t_{\text{набл}} \geq t_{\text{критич}}$, то выборочное среднее значимо отличается от μ_0 на уровне значимости α , и в этой ситуации отклоняется гипотеза H_0 , т. е. считается, что выборка взята из другой генеральной совокупности, для которой $\mu \neq \mu_0$. Если $t_{\text{набл}} < t_{\text{критич}}$, то на заданном уровне различие незначимо и сохраняется гипотеза H_0 .

Замечание 2.

При больших объемах выборки ($n \geq 30$), как указано и гл. 4, t-распределение переходит в нормированное нормальное распределение, поэтому при проверке гипотезы вместо t-критерия можно использовать U-критерий, основанный на нормированном нормальном распределении статистики критерия.

В этом случае вычисляют величину

$$U_{\text{набл}} = \frac{|\bar{x}_B - \mu_0|}{S / \sqrt{n}}$$

и сравнивают ее с критическими значениями u_α нормированного нормального распределения. Для стандартных уровней значимости значения u_α приведены в табл. 6.2.

Замечание 1.

Если перед проведением эксперимента известно не только среднее значение μ генеральной совокупности, из которой получена выборка, но и его дисперсия σ , то нет необходимости в вычислении S . Кроме того, при проверке гипотезы вместо t -критерия можно использовать Z -критерий, основанный на нормированном нормальном распределении статистики критерия.

В этом случае вычисляют величину

$$Z_{\text{набл}} = \frac{|\bar{x}_B - \mu_0|}{\sigma / \sqrt{n}}$$

и сравнивают ее с критическими значениями z нормированного нормального распределения. Для стандартных уровней значимости значения z приведены в табл. 6.2.

Таблица 6.2.

Критические значения U -критерия

Уровень значимости, α	Критическое значение, U_α	
	Двусторон ний критерий	Односторо нный критерий
0,05	1,96	1,64
0,01	2,58	2,33
0,001	3,28	3,09

6.4.2. Сравнение двух выборочных средних значений для связанных выборок

Существует много практических задач, в которых две сравниваемые выборки взаимосвязаны в силу особенностей организации эксперимента или просто потому, что этой взаимосвязи нельзя избежать.

В практике медицинских, биологических и педагогических исследований часто используются так называемые парные сравнения. Один из методов таких сравнений заключается в том, что измерения проводятся для одной и той же группы испытуемых до и после применения интересующих исследователя воздействий. Результаты парных сравнений всегда точнее, чем сравнения на независимых группах, и объясняется это тем, что разброс результатов внутри группы испытуемых всегда больше, чем разброс разностей результатов, полученных при повторных измерениях для одних и тех же индивидуумов. Это можно пояснить на следующем простом примере. Допустим, необходимо по измеренному уровню IQ установить влияние на студентов какого-то вида учебной нагрузки. Конечно, можно было бы провести такой эксперимент на двух независимых однородных группах: в одной из них определить среднее значение IQ до чтения курса лекций, а в другой – после. Но и без точных математических доказательств ясно, что выводы будут точнее, если измерения IQ провести у одних и тех же студентов до и после чтения курса лекций. Поэтому парные сравнения всегда выгодно использовать, конечно, если удастся организовать эксперимент так, что будет устранено влияние мешающих факторов (усталость, эффект обучения и т. п.).

При парных сравнениях нельзя использовать рассмотренные выше методы для независимых выборок, поскольку это приведет к большим ошибкам.

Для сравнения средних значений здесь используется модификация T -критерия для связанных выборок. Особенность его в том, что гипотеза формулируется в отношении разностей d , сопряженных пар наблюдений.

Условия применения: $d_i = x_i - y_i$ – разность связанных пар результатов измерения. Делается предположение о нормальном распределении этих разностей в генеральной совокупности с параметрами μ_d, σ_d .

Гипотеза H_0 : $\mu_d = 0$.

Альтернатива H_1 : $\mu_d \neq 0$ (для двустороннего критерия). Можно сформулировать и одностороннюю альтернативу, например, H_1 : $\mu_d > 0$.

Уровень значимости: α .

Порядок применения:

1. Делается предположение о нормальном распределении разностей d_i , формулируется гипотеза H_0 и альтернатива H_1 выбирается уровень значимости α .

2. Получают две выборки объема n , представляющие собой ряды связанных пар наблюдений.

3. Вычисляются среднее арифметическое $\bar{d}_B$ и исправленная выборочная дисперсия S_d^2 .

4. Определяется значение t-критерия:

$$t_{\text{набл}} = \frac{\bar{d}_B}{S_d / \sqrt{n}}. \quad (6.1)$$

5. Из таблицы t-распределения Стьюдента находится $t_{\text{критич}}$ – критическое значение t-критерия при заданном уровне значимости α и числе степеней свободы $\nu = n - 1$.

6. Делается вывод: если $t_{\text{набл}} \geq t_{\text{критич}}$, то наблюдаемое различие значимо на уровне значимости α , в противном случае различие статистически незначимо.

При больших выборках (для $n > 30$) вместо t-критерия можно использовать u-критерий. В этом случае вычисленное по формуле (6.1) значение $t_{\text{набл}}$ сравнивается с критическим значением $u_{\text{критич}}$ нормированного нормального распределения (см. табл. 6.2).

6.4.3. Сравнение двух выборочных дисперсий из нормальных совокупностей

Условия применения F-критерия: обе выборки независимы и получены из нормально распределенных генеральных совокупностей с параметрами μ_x, σ_x и μ_y, σ_y .

Гипотеза $H_0: \sigma_x^2 = \sigma_y^2$.

Альтернатива $H_1: \sigma_x^2 \neq \sigma_y^2$.

(Это двусторонняя гипотеза, поэтому следует применять двусторонний критерий. Если же предположить, что одна из генеральных совокупностей имеет большую дисперсию (обозначим ее σ_1^2), чем другая (σ_2^2), то можно сформулировать одностороннюю гипотезу $H_1: \sigma_x^2 > \sigma_y^2$, и тогда применяется односторонний F-критерий.)

Уровень значимости критерия задается α .

Порядок применения F-критерия:

1. Принимается предположение о нормальности распределения генеральных совокупностей, формулируется гипотеза и альтернатива, назначается уровень значимости α , как указано выше.

2. Получают две независимые выборки из совокупностей X и Y объемом n_x и n_y соответственно.

3. Рассчитываются значения исправленных выборочных дисперсий S_x^2 и S_y^2 . Большую из дисперсий (S_x^2 или S_y^2) обозначают S_1^2 , меньшую — S_2^2 .

4. Вычисляется значение F-критерия по формуле:

$$F_{\text{набл}} = \frac{S_1^2}{S_2^2}$$

5. Сравнивается вычисленное значение $F_{\text{набл}}$ с критическим значением $F_{\text{критич}}$ при заданном уровне значимости α и числе степеней свободы $\nu_1 = n_1 - 1$ и $\nu_2 = n_2 - 1$.

(Критические значения F при уровнях значимости α , равных 0,05, 0,01, 0,001 приведены в таблицах).

Отметим, что в обычно в таблице приведены критические значения *одностороннего* F-критерия. Поэтому если цель исследования доказать, что одна дисперсия больше другой ($H_1: \sigma_1^2 > \sigma_2^2$), то критические значения берутся непосредственно из этой таблицы. Если же применяется двусторонний критерий ($H_1: \sigma_1^2 \neq \sigma_2^2$), то критические значения, взятые из таблиц, соответствуют удвоенным уровням значимости: 0,01, 0,02 и 0,002.

6. Делается вывод: если вычисленное значение $F_{\text{набл}}$ больше или равно критическому $F_{\text{критич}}$, то дисперсии различаются значимо на заданном уровне значимости. К противном случае нет оснований для отклонения нулевой гипотезы о равенстве двух дисперсий.

Следует отметить, что F-критерий очень чувствителен к отклонениям от нормальности распределения генеральной совокупности. Если предположение о нормальном распределении не может быть принято, то F-критерий применять не следует. В этом случае используются непараметрические методы.

F-критерий используется для малых и средних объемов выборки ($n < 100$). Для больших объемов выборки ($n > 100$) при проверке гипотезы о равенстве дисперсий удобнее применять U-критерий. В этом случае вычисляется величина

$$U = \frac{|S_x - S_y|}{\sqrt{\frac{S_x^2}{2n_x} - \frac{S_y^2}{2n_y}}}$$

и сравнивается с критическими значениями U , взятыми из таблиц нормированного нормального распределения. Для стандартных уровней значимости значения u приведены в табл. 6.2.

6.4.4. Сравнение двух выборочных средних значений для независимых выборок

В этом разделе рассматривается очень важный для практики спорта критерий математической статистики, позволяющий получить ответ на вопрос: значимо ли различаются средние значения, полученные по двум независимым выборкам (например, по результатам в контрольной и экспериментальной группах)? Здесь также применяется t -критерий Стьюдента, основанный на предположении, что выборки получены из генеральных совокупностей, имеющих приблизительно нормальное распределение. Кроме того, применение t -критерия отличается при различных предположениях относительно дисперсий этих генеральных совокупностей. В математической статистике обычно рассматриваются случаи известных и неизвестных генеральных дисперсий, но, поскольку на практике генеральные дисперсии, как правило, неизвестны, здесь описывается *только случай неизвестных дисперсий*. При этом возможны следующие варианты предположений:

- 1) обе дисперсии неизвестны, но предполагается, что они равны между собой;
- 2) обе дисперсии неизвестны, и предположение о их равенстве не делается.

Как выбрать подходящий вариант? Конечно, если нет уверенности в равенстве дисперсий, нужно использовать второй вариант, потому что в этом случае требуется меньше знаний о распределении генеральных совокупностей, но всегда платой за это является меньшая точность выводов.

Поэтому обычно поступают следующим образом: вначале по имеющимся выборочным данным проверяют гипотезу о равенстве дисперсий, используя F -критерий, а затем уже выбирают тот или иной вариант t -критерия. Строго говоря, это некорректно с точки зрения математической статистики, поскольку, как уже неоднократно подчеркивалось, критерий должен выбираться до получения экспериментальных данных, и правильнее было бы выбрать предположение о равенстве или неравенстве дисперсий по другим, предварительно полученным экспериментальным данным.

При описанном выше подходе t -критерий применяется следующим образом.

Условия применения: обе выборки независимы и получены из генеральных совокупностей X и Y , имеющих нормальное распределение с параметрами μ_x, σ_x и μ_y, σ_y .

Гипотеза H_0 : $\mu_x = \mu_y$.

Альтернатива H_1 : $\mu_x \neq \mu_y$ или H_1 : $\mu_x > \mu_y$ ($\mu_x < \mu_y$) в зависимости от того, что требуется доказать: простое различие средних значений или то, что одно из них больше другого.

Уровень значимости: α .

Порядок применения:

1. Принимается предположение о нормальности, формулируются гипотеза H_0 и альтернатива H_1 , задается уровень значимости α .

2. Получают две независимые выборки из совокупностей X и Y объемом n_x и n_y .

3. Вычисляются выборочные характеристики S_x и S_y методами, рассмотренными в гл. 3.

4. Используется F -критерий для проверки гипотезы о равенстве генеральных дисперсий, как показано в разделе 6.3.1.

5. По результатам применения F -критерия принимается или не принимается предположение о равенстве дисперсий.

6. Вычисляются значение t -критерия и число степеней свободы v . Применяемые для этого формулы приведены в табл. 6.3, они различаются в зависимости от предположения о дисперсиях и соотношения между объемами выборок n_x и n_y .

7. Из таблицы t -распределения Стьюдента находится $t_{\text{критич}}$ – критическое значение t -критерия при заданном уровне значимости α и числе степеней свободы v .

8. Делается вывод: если $t_{\text{набл}} \geq t_{\text{критич}}$, то выборочные средние значимо различаются на уровне значимости α (вероятность ошибки меньше α). В противном случае различие статистически незначимо.

Формулы для вычисления значения t-критерия

№	Предположения о дисперсиях σ_1^2 и σ_2^2	Объемы выборок n_x и n_y	Формула t-критерия	Стандартная ошибка разности $S_{\bar{x}-\bar{y}}$	Число степеней свободы ν
1	$\sigma_1^2 = \sigma_2^2$	$n_x = n_y = n$	$t_{набл} = \frac{ \bar{x} - \bar{y} }{S_{\bar{x}-\bar{y}}}$	$S_{\bar{x}-\bar{y}} = \sqrt{\frac{S_x^2 + S_y^2}{n}}$	$\nu = 2n - 2$
2	$\sigma_1^2 = \sigma_2^2$	$n_x \neq n_y$		$S_{\bar{x}-\bar{y}} = \sqrt{\frac{n_x + n_y}{n_x n_y}} \sqrt{\frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}}$	$\nu = n_x + n_y - 2$
3	$\sigma_1^2 \neq \sigma_2^2$	$n_x = n_y = n$		$S_{\bar{x}-\bar{y}} = \sqrt{\frac{S_x^2 + S_y^2}{n}}$	$\nu = (n - 1) \frac{(S_x^2 + S_y^2)^2}{S_x^4 + S_y^4}$
4	$\sigma_1^2 \neq \sigma_2^2$	$n_x \neq n_y$		$S_{\bar{x}-\bar{y}} = \sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}$	$\nu = \frac{\left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}\right)^2}{\frac{S_x^4}{n_x^2(n_x - 1)} + \frac{S_y^4}{n_y^2(n_y - 1)}}$