

ОСНОВЫ СТАТИСТИКИ

Проверка гипотез. Использование
Z-оценок и t-критерия Стьюдента



Здравствуйте!

Изучив материал, вы будете уметь

- объяснять логику тестирования гипотезы, включая понятие нулевой гипотезы,
- выборочного распределения,
- альфа-уровня и испытательной статистической величины.
- что это означает «отклонять нулевую гипотезу», или «не отклонять нулевую гипотезу». объясните различие между одним – и двусторонними тестами. Определять и объяснять ошибки типа I и типа II.

Что же такое статистическая гипотеза? Статистической гипотезой называют предположение относительно параметра известного распределения или о виде неизвестного распределения. При этом, проверяемую гипотезу (выдвинутую) называют нулевой (основной) и обозначают H_0 . Наряду с нулевой гипотезой H_0 рассматривают ей противоречащую гипотезу, называемую альтернативной (конкурирующей) H_1 . От основного определения предположений отходит ещё одно, не менее важное, – статистическая гипотеза есть изучение генеральной совокупности важных для науки объектов, относительно коих учёными делаются выводы. Ее можно проверить с помощью выборки (части генеральной совокупности).

Приведём несколько примеров статистических гипотез.

1. Успеваемость всего класса, возможно, зависит от уровня образования каждого учащегося.
2. Начальный курс математики в равной степени усваивается как детьми, пришедшими в школу в 6 лет, так и детьми, пришедшими в 7.

Простой гипотезой в статистике называют такое предположение, которое однозначно характеризует определённый параметр величины, взятой учёным. Сложная состоит из нескольких или бесконечного множества простых. Указывается некоторая область или нет точного ответа. Полезно понимать несколько определений гипотез в статистике, чтобы не путать их на практике.

Концепция нулевой гипотезы

Нулевая гипотеза — это теория о том, что есть некие две совокупности, которые не различаются между собой. Однако на научном уровне нет понятия «не различаются», но есть «их сходство равно нулю». От этого определения и было образовано понятие. В статистике нулевая гипотеза обозначается как H_0 . Причём крайним значением невозможного (маловероятного) считается от 0.01 до 0.05 или менее.

Лучше разобраться, что такое нулевая гипотеза, пример из жизни поможет. Педагог в университете предположил, что различный уровень подготовки учащихся двух групп к зачётной работе вызван незначительными параметрами, случайными причинами, не влияющими на общий уровень образования (разница в подготовке двух групп студентов равна нулю). Однако встречно стоит привести пример альтернативной гипотезы — допущения, опровергающего утверждение нулевой теории (H_1). Например: директор университета предположил, что различный уровень в подготовке к зачётной работе у учащихся двух групп вызван применением педагогами разных методик обучения (разница в подготовке двух групп существенна и на то есть объяснение).

Теперь сразу видна разница между понятиями «нулевая гипотеза» и «альтернативная гипотеза». Примеры иллюстрируют эти понятия.

Проверка нулевой гипотезы. Создать предположение — это ещё полбеды. Настоящей проблемой для новичков считается проверка нулевой гипотезы. Именно тут многих и ожидают трудности. Используя метод альтернативной гипотезы, утверждающей нечто обратное нулевой теории, можно сравнить оба варианта и выбрать верный. Так действует статистика. Пусть нулевая



гипотеза H_0 , а альтернативная H_1 , тогда:

$H_0: c = c_0$;

$H_1: c \neq c_0$.

Здесь c – это некое среднее значение генеральной совокупности, которое предстоит найти, а c_0 – данное изначально значение, по отношению к которому проверяется гипотеза. Также есть некоторое число X – среднее значение выборки, по которому определяется c_0 .

«Доверительный» способ проверки.

Существует наиболее действенный способ, с помощью которого нулевая статистическая гипотеза легко проверяется на практике. Он заключается в построении диапазона значений до 95% точности. Для начала понадобится знать формулу расчёта доверительного интервала:

$$X - t * S_x \leq c \leq X + t * S_x,$$

где X – данное изначально число на основе альтернативной гипотезы; t – табличные величины (коэффициент Стьюдента); S_x – стандартная средняя ошибка, которая рассчитывается как $S_x = \sigma/N$, где в числителе стандартное отклонение, а в знаменателе – объём выборки. Итак, предположим ситуацию. До ремонта конвейер в день выпускал 32.1 кг конечной продукции, а после ремонта, как утверждает предприниматель, коэффициент полезного действия вырос, и конвейер, по недельной проверке, начал выпускать 39.6 кг в среднем. Нулевая гипотеза будет утверждать, что ремонт никак не повлиял на КПД конвейера. Альтернативная гипотеза скажет, что ремонт коренным образом изменил КПД конвейера, поэтому производительность его повысилась. По таблице находим $n=7$, $t = 2,447$, откуда формула примет следующий вид: $39,6 - 2,447 * 4,2 \leq c \leq 39,6 + 2,447 * 4,2$; $29,3 \leq c \leq 49,9$. Получается, что значение 32.1 входит в диапазон, следовательно, значение, предложенное альтернативой – 39.6 – не принимается автоматически.

Помните, что сначала проверяется на правильность нулевая гипотеза, а потом – противоположная.

Разновидности отрицания

До этого рассматривался такой вариант построения гипотезы, где H_0 утверждает что-либо, а H_1 это опровергает. Откуда можно было составить подобную систему: $H_0: c = c_0$; $H_1: c \neq c_0$. Но существует ещё два родственных способа опровержения. К примеру, нулевая гипотеза утверждает, что средняя оценка успеваемости класса больше 4.54, а альтернативная тогда скажет, что средняя успеваемость того же класса менее 4.54.

$$H_0: c \geq 4.54; H_1: c < 4.54.$$

И выглядеть в виде системы это будет так:

Обратите внимание, что нулевая гипотеза утверждает, что значение больше или равно, а статистическая – что строго меньше.

Статистическая проверка

Статистическая проверка нулевых гипотез заключается в использовании статистического критерия. Такие критерии подчиняются различным законам распределения. К примеру,



существует F-критерий, который рассчитывается по распределению Фишера. Есть T-критерий, чаще всего используемый на практике, зависящий от распределения Стьюдента. Квадратный критерий согласия Пирсона и т. д.

Область принятия нулевой гипотезы

В алгебре есть понятие «область допустимых значений». Это такой отрезок или точка на оси X, на котором находится множество значений статистики, при которых нулевая гипотеза верна. Крайние точки отрезка – критические значения. Лучи по правую и левую сторону отрезка – критические области. Если найденное значение входит в них, то нулевая теория. Нулевая гипотеза в статистике временами очень изворотливое понятие.

Во время проверки её можно допустить ошибки двух типов:

1. Отвержение верной нулевой гипотезы. Обозначим первый тип как $\alpha=1$.
2. Принятие ложной нулевой гипотезы. Второй тип обозначим как $\alpha=2$.

Стоит понимать, что это не одинаковые параметры, исходы ошибок могут существенно различаться между собой и иметь разные выборки.

Пример ошибок двух типов

Со сложными понятиями легче разобраться на примере. Во время производства некоего лекарства от учёных требуется чрезвычайная осторожность, так как превышение дозы одного из компонентов провоцирует высокий уровень токсичности готового препарата, от которого пациенты, принимающие его, могут умереть. Однако на химическом уровне выявить передозировку невозможно. Из-за этого перед тем как выпустить лекарство в продажу, небольшую его дозу проверяют на крысах или кроликах, вводя им препарат. Если большая часть испытуемых умирает, то лекарство в продажу не допускается, если подопытные живы, то лекарство разрешают продавать в аптеках.

Первый случай: на самом деле лекарство было не токсично, но во время эксперимента была допущена оплошность и препарат классифицировали как токсичный и не допустили в продажу. $\alpha=1$.

Второй случай: в ходе другого эксперимента при проверке другой партии лекарства решено, что препарат не токсичен, и в продажу его допустили, хотя на самом деле препарат был ядовит. $\alpha=2$.

Первый вариант повлечёт за собой крупные финансовые затраты поставщика-предпринимателя, так как придётся уничтожить всю партию лекарства и начинать с нуля.

Вторая ситуация спровоцирует смерть пациентов, купивших и употреблявших это лекарство.

Теория вероятности.

Не только нулевые, но все гипотезы в статистике и экономике разделяют по уровню значимости. Уровень значимости – процент появления ошибок первого рода (отклонение верной нулевой гипотезы).

первый уровень – 5% или 0.05, т. е. вероятность ошибиться 5 к 100 или 1 к 20.

второй уровень – 1% или 0.01, т. е. вероятность 1 к 100.

третий уровень – 0.1% или 0.001, вероятность 1 к 1000.



Критерии проверки гипотезы

Если учёным уже был сделан вывод о правильности нулевой гипотезы, то её необходимо подвергнуть проверке. Это необходимо, чтобы исключить ошибку. Существует основной критерий проверки нулевой гипотезы, состоящий из нескольких этапов:

1. Берётся допустимая ошибочная вероятность $P=0.05$.
2. Подбирается статистика для критерия 1.
3. По известному методу находится область допустимых значений.
4. Теперь вычисляется значение статистики T .
5. Если T (статистика) принадлежит области принятия нулевой гипотезы (как в «доверительном» методе), то предположения считаются верными, а значит, и сама нулевая гипотеза остаётся верной.

Именно так действует статистика. Нулевая гипотеза при грамотной проверке будет принята или отвергнута.

Систематизируем теперь все то, что было сказано выше.

Установлено 5 стадий при проверке гипотез:

Определение нулевой (H_0) и альтернативной гипотезы (H_1) при исследовании. Определение уровня значимости критерия.

Отбор необходимых данных из выборки.

Вычисление значения статистики критерия, отвечающей.

Вычисление критической области, проверка статистики критерия на предмет попадания в критическую область.

Интерпретация достигнутого уровня значимости p и результатов.

Определение нулевой и альтернативной гипотез, уровня статистической значимости.

При проверке значимости гипотезу следует формулировать независимо от используемых при ее проверке данных (до проведения проверки). В таком случае можно получить действительно продуктивный результат.

Всегда проверяют нулевую гипотезу (H_0), которая отвергает эффект (например, разница средних равняется нулю) в популяции.

Например, при сравнении показателей курения у мужчин и женщин в популяции нулевая гипотеза означала бы, что показатели курения одинаковые у женщин и мужчин в популяции.

Затем определяют альтернативную гипотезу (H_1), которая принимается, если нулевая гипотеза неверна. Альтернативная гипотеза больше относится к той теории, которую собираются исследовать. Итак, на этом примере альтернативная гипотеза заключается в утверждении, что показатели курения различны у женщин и мужчин в популяции.

Разницу в показателях курения не уточнили, т.е. не установили, имеют ли в популяции мужчины более высокие или более низкие показатели, чем женщины. Такой подход известен как двусторонний критерий, потому что учитывают любую возможность, он рекомендуется постольку, поскольку редко есть уверенность заранее в направлении какого-либо различия, если таковое существует.

В некоторых случаях можно использовать односторонний критерий для гипотезы, в котором направление эффекта задано. Его можно применить, например, если рассматривать заболевание, от которого умерли все пациенты, не получившие лечения; новый препарат не мог бы ухудшить



положение дел.

Уровень значимости. Важным этапом проверки статистических гипотез является определение уровня статистической значимости, т.е. максимально допускаемой исследователем вероятности ошибочного отклонения нулевой гипотезы.

Получение статистики критерия, определение критической области

После того как данные будут собраны, значения из выборки подставляют в формулу для вычисления статистики критерия. Эта величина количественно отражает аргументы в наборе данных против нулевой гипотезы.

Критическая область. Для принятия решения об отклонении или не отклонении нулевой гипотезы необходимо также определить критическую область проверки гипотезы.

Выделяют 3 вида критических областей:

$$(-\infty, x_{\alpha/2}) \cup (x_{1-\alpha/2}, +\infty)$$

двусторонняя:

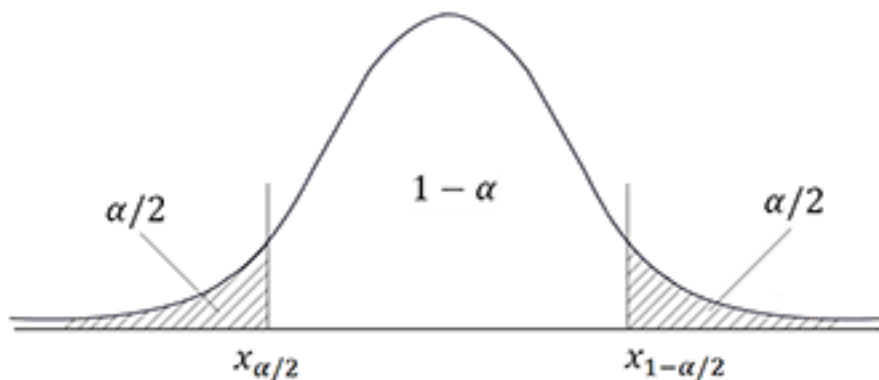


Рис. 10.1. Двусторонняя критическая область

левосторонняя: $(-\infty, x_{\alpha})$

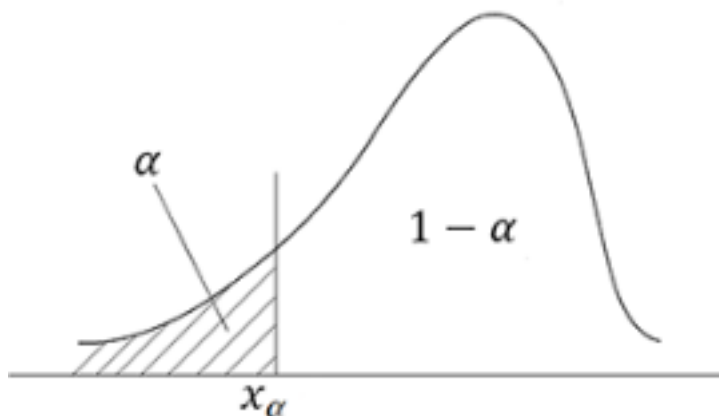


Рис. 10.2 Левосторонняя критическая область



правосторонняя: $(x_{1-\alpha}, +\infty)$

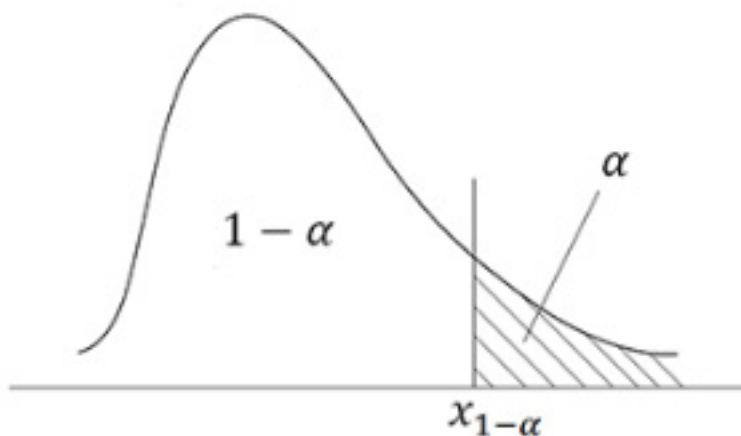


Рис. 10.3 Правосторонняя критическая область

- - заданный исследователем уровень значимости.

Если наблюдаемое значение критерия (K) принадлежит критической области ($K_{кр}$, заштрихованная область на рис.1-3), гипотезу отвергают, если не принадлежит - не отвергают.

Для краткости можно записать и так:

$|K| > K_{кр}$ - отклоняем H_0

$|K| < K_{кр}$ - не отклоняем H_0

Получение значения p (достигнутого уровня значимости).

Все статистики критерия подчиняются известным теоретическим распределениям вероятности. Значение статистики критерия, полученное из выборки, связывают с уже известным распределением, которому она подчиняется, чтобы получить значение p , площадь обоих «хвостов» (или одного «хвоста», в случае односторонней гипотезы) распределения вероятности.

Большинство компьютерных пакетов обеспечивают автоматическое вычисление двустороннего значения p .

Значение p – это вероятность получения нашего вычисленного значения критерия или его еще большего значения, если нулевая гипотеза верна.

Иными словами, p – это вероятность отвергнуть нулевую гипотезу при условии, что она верна.

Нулевая гипотеза всегда относится к популяции, представляющей больший интерес, нежели выборка. В рамках проверки гипотезы мы либо отвергаем нулевую гипотезу и принимаем альтернативу, либо не отвергаем нулевую гипотезу. Подробнее об ошибках при проверке гипотез.

Применение значения p .

Следует решить, сколько аргументов позволят отвергнуть нулевую гипотезу в пользу альтернативной. Чем меньше значение p , тем сильнее аргументы против нулевой гипотезы.



Традиционно полагают, если $p < 0,05$, ($=0,05$) то аргументов достаточно, чтобы отвергнуть нулевую гипотезу, хотя есть небольшой шанс против этого. Тогда можно отвергнуть нулевую гипотезу и сказать, что результаты значимы на 5% уровне.

Напротив, если $p > 0,05$, то аргументов недостаточно, чтобы отвергнуть нулевую гипотезу. Не отвергая нулевую гипотезу, можно заявить, что результаты не значимы на 5% уровне. Данное заключение не означает, что нулевая гипотеза истинна, просто недостаточно аргументов (возможно, маленький объем выборки), чтобы ее отвергнуть.

Уровень значимости (т.е. выбранная «граница отсечки») 5% задается произвольно. На уровне 5% можно отвергнуть нулевую гипотезу, когда она верна. Если это может привести к серьезным последствиям, необходимо потребовать более веских аргументов, прежде чем отвергнуть нулевую гипотезу, например, выбрать значение $= 0,01$ (или 0,001).

Определение результата только как значимого на определенном уровне граничного значения (например, 0, 05) может ввести в заблуждение. Например, если $p = 0,04$, то нулевую гипотезу отвергаем, но если $p = 0,06$, то ее не отвергли бы. Действительно ли они различны? Мы рекомендуем всегда указывать точное значение p , обычно получаемое путем компьютерного анализа.

Проверка гипотез против доверительных интервалов.

Доверительные интервалы и проверка гипотез тесно связаны. Первоначальная цель проверки гипотезы состоит в том, чтобы принять решение и предоставить точное значение p .

Доверительный интервал (ДИ) количественно определяет изучаемый эффект (например, разницу в средних) и дает возможность оценить значение результатов. ДИ предоставляют интервал вероятных значений для истинного эффекта, поэтому его также можно использовать для принятия решения даже без точных значений p .

Например, если бы гипотетическое значение для данного эффекта (например, значение, равное нулю) находилось вне 95% ДИ, можно было бы считать гипотетическое значение неправдоподобным и отвергнуть . В этом случае станет известно, что $p < 0,05$, но не станет известно его точное значение.

Ошибки I и II рода при проверке гипотез, мощность.

Большинство проверяемых гипотез сравнивают между собой группы объектов, которые испытывают влияние различных факторов.

Например, можно сравнить эффективность двух видов лечения, чтобы сократить 5-летнюю смертность от рака молочной железы. Для данного исхода (например, смерть) сравнение, представляющее интерес (например, различные показатели смертности через 5 лет), называют эффектом или, если уместно, эффектом лечения.

Нулевую гипотезу выражают как отсутствие эффекта (например, 5-летняя смертность от рака молочной железы одинаковая в двух группах, получающих разное лечение); двусторонняя альтернативная гипотеза будет означать, что различие эффектов не равно нулю.

Критериальная проверка гипотезы дает возможность определить, достаточно ли аргументов, чтобы отвергнуть нулевую гипотезу. Можно принять только одно из двух решений:

-отвергнуть нулевую гипотезу и принять альтернативную гипотезу,



-остаться в рамках нулевой гипотезы.

Важно: в литературе достаточно часто встречается понятие «принять нулевую гипотезу». Хотелось бы внести ясность, что со статистической точки зрения принять нулевую гипотезу невозможно, т.к. нулевая гипотеза представляет собой достаточно строгое утверждение (например, средние

значения в сравниваемых группах равны $H_0: \bar{X}_1 = \bar{X}_2$).

Поэтому фразу о принятии нулевой гипотезы следует понимать как то, что мы просто остаемся в рамках гипотезы.

Принятие неправильного решения

Возможно неправильное решение, когда отвергают/не отвергают нулевую гипотезу, потому что есть только выборочная информация.

		Верная гипотеза	
		H_0	H_1
Результат применения критерия	H_0	H_0 верно принята	H_0 неверно принята (Ошибка второго рода)
	H_1	H_0 неверно отвергнута (Ошибка первого рода)	H_0 верно отвергнута

Ошибка 1-го рода: нулевую гипотезу отвергают, когда она истинна, и делают вывод, что имеется эффект, когда в действительности его нет. Максимальный шанс (вероятность) допустить ошибку 1-го рода обозначается α (альфа). Это уровень значимости критерия; нулевую гипотезу отвергают, если наше значение p ниже уровня значимости, т. е., если $p < \alpha$.

Следует принять решение относительно значения, а прежде, чем будут собраны данные; обычно назначают условное значение 0,05, хотя можно выбрать более ограничивающее значение, например 0,01.

Шанс допустить ошибку 1-го рода никогда не превысит выбранного уровня значимости, скажем $\alpha = 0,05$, так как нулевую гипотезу отвергают только тогда, когда $p < 0,05$. Если обнаружено, что $p > 0,05$, то нулевую гипотезу не отвергнут и, следовательно, не допустят ошибки 1-го рода.

Ошибка 2-го рода: не отвергают нулевую гипотезу, когда она ложна, и делают вывод, что нет эффекта, тогда как в действительности он существует. Шанс возникновения ошибки 2-го рода обозначается β (бета); а величина $(1-\beta)$ называется мощностью критерия.

Следовательно, мощность – это вероятность отклонения нулевой гипотезы, когда она ложна, т.е.



это шанс (обычно выраженный в процентах) обнаружить реальный эффект лечения в выборке данного объема как статистически значимый.

В идеале хотелось бы, чтобы мощность критерия составляла 100%; однако это невозможно, так как всегда остается шанс, хотя и незначительный, допустить ошибку 2-го рода.

К счастью, известно, какие факторы влияют на мощность и, таким образом, можно контролировать мощность критерия, рассматривая их.

Мощность и связанные факторы

Планируя исследование, необходимо знать мощность предложенного критерия. Очевидно, можно начинать исследование, если есть «хороший» шанс обнаружить уместный эффект, если таковой существует (под «хорошим» мы подразумеваем, что мощность должна быть по крайней мере 70-80%).

Этически безответственно начинать исследование, у которого, скажем, только 40% вероятности обнаружить реальный эффект лечения; это бесполезная трата времени и денежных средств.

Ряд факторов имеют прямое отношение к мощности критерия.

Объем выборки: мощность критерия увеличивается по мере увеличения объема выборки. Это означает, что у большей выборки больше возможностей, чем у незначительной, обнаружить важный эффект, если он существует.

Когда объем выборки небольшой, у критерия может быть недостаточно мощности, чтобы обнаружить отдельный эффект. Эти методы также можно использовать для оценки мощности критерия для точно установленного объема выборки.

Вариабельность наблюдений: мощность увеличивается по мере того, как вариабельность наблюдений уменьшается.

Интересующий исследователя эффект: мощность критерия больше для более высоких эффектов. Критерий проверки гипотез имеет больше шансов обнаружить значительный реальный эффект, чем незначительный.

Уровень значимости: мощность будет больше, если уровень значимости выше (это эквивалентно увеличению допущения ошибки 1-го рода, α , а допущение ошибки 2-го рода, β , уменьшается).

Таким образом, вероятнее всего, исследователь обнаружит реальный эффект, если на стадии планирования решит, что будет рассматривать значение p как значимое, если оно скорее будет меньше 0,05, чем меньше 0,01.

Обратите внимание, что проверка ДИ для интересующего эффекта указывает на то, была ли мощность адекватной. Большой доверительный интервал следует из небольшой выборки и/или набора данных с существенной вариабельностью и указывает на недостаточную мощность.

Проверка множественных гипотез

Часто нужно выполнить критериальную проверку значимости множественных гипотез на наборе данных с многими переменными или существует более двух видов лечения.

Ошибка 1-го рода драматически увеличивается по мере увеличения числа сравнений, что приводит к ложным выводам относительно гипотез. Следовательно, следует проверить только небольшое число гипотез, выбранных для достижения первоначальной цели исследования и



точно установленных априорно.

Можно использовать какую-нибудь форму апостериорного уточнения значения p , принимая во внимание число выполненных проверок гипотез.

Например, при подходе Бонферрони (его часто считают довольно консервативным) умножают каждое значение p на число выполненных проверок; тогда любые решения относительно значимости будут основываться на этом уточненном значении p .

Выводы

Различают простую и сложную статистические гипотезы. Гипотезу называют простой, если она содержит только одно предположение, и сложной, если она состоит из конечного или бесконечного числа простых гипотез.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость проверки гипотезы.

Проверки гипотезы могут быть допущены ошибки первого и второго рода. Ошибка I рода состоит в том, что правильная нулевая гипотеза H_0 отвергается. Ошибка II рода заключается в том, что принимается неправильная нулевая гипотеза.

Вероятность совершить ошибку первого рода принято обозначать через α ; второго рода – β . Вероятность совершить ошибку первого рода называют уровнем значимости.

Обычно уровень значимости принимают равным 0,05; 0,01; 0,005; 0,001. Вероятность не допустить ошибку II рода, т.е. отвергнуть гипотезу H_0 , когда она неверна, называется мощностью критерия.