

Проверка статистических гипотез

1 Основные понятия.

- **Нулевая гипотеза** (H_0) – утверждение о параметре генеральной совокупности (параметрах генеральных совокупностей) или распределении, которое необходимо проверить.
- **Альтернативная гипотеза** (H_A) – утверждение, противоположное нулевой гипотезе. Выдвигается, но не проверяется.

Все гипотезы можно разделить на двусторонние (ненаправленные) и односторонние (направленные).

Двусторонние альтернативы

$$(H_A : p \neq 0.5)$$

Односторонние альтернативы

- левосторонние ($H_A : p < 0.5$)

- правосторонние ($H_A : p > 0.5$)

- **Уровень значимости** (α) – вероятность отвергнуть верную нулевую гипотезу.

Нулевая гипотеза всегда проверяется на определенном уровне значимости. Например, если мы проверяем нулевую гипотезу на уровне значимости 5%, это означает, что если мы будем проводить аналогичные исследования 100 раз и проверять на основе имеющихся данных интересующую нас нулевую гипотезу, в 5 случаях из 100 мы отвергнем нулевую гипотезу, хотя она будет верной.

Уровень значимости в каком-то смысле является понятием, противоположным уровню доверия. Уровень доверия – вероятность не отвергнуть верную нулевую гипотезу. Имеет место следующее соотношение:

$$\alpha = 1 - \gamma$$

Проверить нулевую гипотезу на уровне значимости 5% и проверить нулевую гипотезу на уровне доверия 95% – это одно и то же.

- **Типы ошибок**

Ошибка I рода – вероятность отвергнуть верную нулевую гипотезу (α).

Ошибка II рода – вероятность не отвергнуть неверную нулевую гипотезу (β).

Мощность критерия – вероятность отвергнуть неверную нулевую гипотезу ($1 - \beta$).

	H_0 верна	H_0 неверна
H_0 отвергается	ошибка I рода	+
H_0 не отвергается	+	ошибка II рода

2 Логика проверки статистических гипотез.

Предположим, что мы проводим исследование, посвященное степени поддержки разных кандидатов в губернаторы в регионе. У нас есть гипотеза, которая утверждает, что доля сторонников кандидата А в регионе равна 0.7. Мы опросили 1000 человек и выяснили, что доля респондентов, которые поддерживают кандидата А, равна 0.5. Можно ли сразу по таким результатам опроса сделать однозначный вывод, что доля сторонников кандидата в регионе не равна 0.7 (ведь $0.5 \neq 0.7$)? Нельзя.

Во-первых, мы уже знаем, что оценки параметра (в данном случае доли), полученные по одной выборке, отличаются от истинного значения параметра генеральной совокупности. Поэтому из того факта, что доля сторонников кандидата А в выборке равна 0.5, не следует, что доля его сторонников по всему региону обязательно равна 0.5.

Во-вторых, нам неизвестно, какая разница между выборочной долей и долей, заявленной в гипотезе, считается «маленькой», то есть достаточной для того, чтобы не отвергнуть нулевую гипотезу. В нашем примере доля сторонников кандидата A в выборке равна 0.5, мы можем считать, что 0.5 сильно отличается от 0.7, поэтому нам следует отвергнуть нулевую гипотезу. А что было бы, если бы выборочная доля была бы 0.6? Или 0.65? Сделали бы мы тогда вывод, что доля сторонников кандидата A в регионе не равна 0.7? Непонятно, потому что неизвестно, что считать сильным отличием, а что просто списывать на неточность оценок, получаемых по выборке.

Для того, чтобы понять, являются ли различия между значением в гипотезе и полученным по выборке, действительно существенными или эти различия – просто следствие того, что оценки по выборке мы получаем с некоторой погрешностью, требуется формальная проверка гипотез. Для разных видов вопросов существуют свои статистические критерии, позволяющие проверять соответствующие им нулевые гипотезы.

Статистический критерий – правило, которое позволяет делать вывод о том, стоит ли на основе имеющихся данных отвергнуть нулевую гипотезу или нет. Обычно для критерия определяется соответствующая ему **статистика** – функция от наблюдений, которая имеет свое распределение. Для того чтобы понять, действительно ли разница между значением параметра в гипотезе и значением оценки, полученной по выборке, является существенной, необходимо сравнить два показателя: наблюдаемое значение статистики и критическое значение статистики.

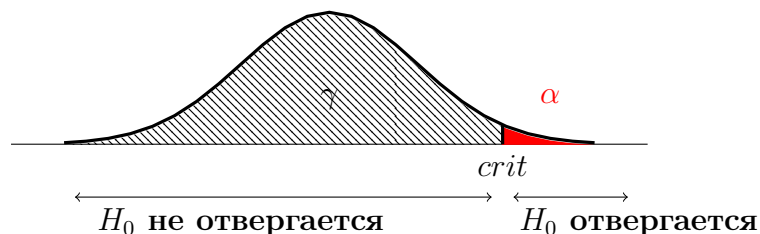
Наблюдаемое значение статистики – значение статистики, которое получается по выборке, на основе имеющихся данных.

Критическое значение – пороговое значение статистики, которое ожидается в случае, если нулевая гипотеза верна. Критическое значение статистики отделяет область типичных значений статистики от **критической области** – области редких значений статистики при условии, что нулевая гипотеза верна. Область типичных значений – область не-отвержения нулевой гипотезы, критическая область – область отвержения нулевой гипотезы. Вид критической области зависит от типа альтернативной гипотезы.

I. Альтернативная гипотеза правосторонняя

Предположим, что мы проверяем некоторую гипотезу на уровне значимости α . Область типичных значений – все значения до критического ($crit$). Критическая область – все значения больше $crit$. Допустим, мы рассчитали наблюдаемое значение статистики. Если оно попадает в область типичных значений, то нулевая гипотеза отвергаться не будет. Если оно попадает в критическую область – в правый «хвост», то нулевая гипотеза должна отвергаться. Почему критическая область находится справа? Когда нулевая гипотеза отвергается, мы отвергаем ее в пользу альтернативы, а правосторонняя альтернатива говорит о том, что интересующий нас параметр *больше* какого-то значения (например, $H_A : p > 0.5$).

распределение статистики при условии, что H_0 верна

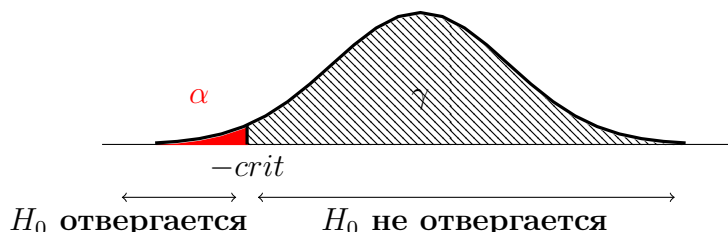


II. Альтернативная гипотеза левосторонняя

Предположим, что мы проверяем некоторую гипотезу на уровне значимости α . Область типичных значений – все значения до критического с отрицательным знаком ($-crit$). Критическая область – все значения меньше $-crit$. Допустим, мы рассчитали наблюдаемое значение

статистики. Если оно попадает в область типичных значений, то нулевая гипотеза отвергаться не будет. Если оно попадает в критическую область – в левый «хвост», то нулевая гипотеза должна отвергаться. Почему критическая область находится слева? Когда нулевая гипотеза отвергается, мы отвергаем ее в пользу альтернативы, а левосторонняя альтернатива говорит о том, что интересующий нас параметр *меньше* какого-то значения (например, $H_A : p < 0.5$).

распределение статистики при условии, что H_0 верна



III. Альтернативная гипотеза двусторонняя

Предположим, что мы проверяем некоторую гипотезу на уровне значимости α . Область типичных значений – все значения от критического с отрицательным знаком до критического с положительным знаком (от $-crit$ до $crit$). Критическая область – все значения больше $crit$ или меньше $-crit$. Допустим, мы рассчитали наблюдаемое значение статистики. Если оно попадает в область типичных значений, то нулевая гипотеза отвергаться не будет. Если оно попадает в критическую область – в левый или правый «хвост», то нулевая гипотеза должна отвергаться. Почему критическая область находится слева и справа? Когда нулевая гипотеза отвергается, мы отвергаем ее в пользу альтернативы, а двусторонняя альтернатива говорит о том, что интересующий нас параметр или *больше* какого-то значения или *меньше* него (например, $H_A : p \neq 0.5$).

распределение статистики при условии, что H_0 верна



3 Алгоритм проверки гипотез

1. Сформулировать нулевую гипотезу (H_0).
2. Сформулировать альтернативную гипотезу (H_A).

NB: Важно указывать, так как от типа альтернативной гипотезы (двусторонняя или односторонняя) зависит критическое значение статистики и критическая область.

3. Выбрать критерий, необходимый для проверки нулевой гипотезы.
4. Определить критическое значение статистики и критическую область.
 - Для двусторонней альтернативы: критическое значение – квантиль уровня $(1 - \frac{\alpha}{2})$
 - Для односторонней альтернативы: критическое значение – квантиль уровня $(1 - \alpha)$
5. Определить наблюдаемое значение статистики.

6. Сравнить наблюдаемое и критическое значения. Если наблюдаемое значение (по модулю) больше критического значения статистики – попадаем в критическую область, следовательно, нулевую гипотезу необходимо отвергнуть.
7. Сделать статистический и содержательный вывод.

Пример статистического вывода: на имеющихся данных, на уровне значимости 5% (уровне доверия 95%) есть основания/нет оснований отвергнуть нулевую гипотезу в пользу альтернативы.

NB: Важно всегда указывать уровень значимости (уровень доверия), на котором проверяется гипотеза, так как без этого уточнения выводы о нулевой гипотезе не имеют большого смысла: на одном уровне значимости гипотеза может быть отвергнута, а при выборе другого уровня значимости – нет. Желательно также прописывать, что выводы делаются на имеющихся данных, так как мы можем отвечать только за те результаты, которые получили по той выборке / выборкам, которые у нас есть, а не за «истинность» выводов вообще.

Пример содержательного вывода: среднее значение индекса политической стабильности для демократических стран не равно 2.

Важно! По результатам проверки статистической гипотезы мы никогда не делаем вывод о том, что нулевая гипотеза верна / должна быть принята. Вопрос об истинности нулевой гипотезы – содержательный вопрос, и если он и проверяется статистически, то с помощью более продвинутых методов и в рамках специально продуманного дизайна исследования. Всё, что мы можем решить по итогам проверки: отвергнуть нулевую гипотезу или нет. Как из того, что события не независимы, автоматически не следует их зависимость, так и из того, что нулевая гипотеза не отвергается, не следует, что она принимается.

4 Проверка гипотезы о равенстве доли числу

Нулевая гипотеза	$H_0 : p = p_0$
Статистический критерий	z - критерий
Критическое значение статистики критерия	двусторонняя альтернатива: $z_{\text{крит}} = z_{(1-\frac{\alpha}{2})}$ односторонняя альтернатива: $z_{\text{крит}} = z_{(1-\alpha)}$
Наблюдаемое значение статистики критерия	$z_{\text{набл}} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}\hat{q}}{n}}},$ где \hat{p} – выборочная доля, $\hat{q} = 1 - \hat{p}$, p_0 – значение доли из гипотезы, n – число элементов в выборке
Проверка	$ z_{\text{набл}} > z_{\text{крит}} \Rightarrow H_0 \text{ отвергается}$ $ z_{\text{набл}} < z_{\text{крит}} \Rightarrow H_0 \text{ не отвергается}$

5 Проверка гипотезы о равенстве среднего значения числу

Требования к данным: 1) данные измерены в количественной шкале; 2) выборка взята из генеральной совокупности, имеющей (примерно) нормальное распределение.

Нулевая гипотеза	$H_0 : \mu = a$
Статистический критерий	критерий Стьюдента для одной выборки (t - критерий)
Критическое значение статистики критерия	двусторонняя альтернатива: $t_{\text{крит}} = t_{(1-\frac{\alpha}{2}, df=n-1)}$ односторонняя альтернатива: $t_{\text{крит}} = t_{(1-\alpha, df=n-1)}$
Наблюдаемое значение статистики критерия	$t_{\text{набл}} = \frac{\bar{x}-a}{\frac{s}{\sqrt{n}}},$ <p>где \bar{x} – выборочное среднее, a – значение среднего из гипотезы, s – выборочное стандартное отклонение, n – число элементов в выборке</p>
Проверка	$ t_{\text{набл}} > t_{\text{крит}} \Rightarrow H_0$ отвергается $ t_{\text{набл}} < t_{\text{крит}} \Rightarrow H_0$ не отвергается

6 Проверка гипотезы о равенстве средних значений

Требования к данным: 1) данные измерены в количественной шкале; 2) выборки взяты из генеральных совокупностей, имеющих (примерно) нормальное распределение.

Требования к описанному ниже варианту критерия Стьюдента для двух выборок:

1) две выборки являются независимыми; 2) дисперсии двух генеральных совокупностей равны.

Нулевая гипотеза	$H_0 : \mu_1 = \mu_2$
Статистический критерий	критерий Стьюдента для двух выборок (t - критерий)
Критическое значение статистики критерия	двусторонняя альтернатива: $t_{\text{крит}} = t_{(1-\frac{\alpha}{2}, df=n_1+n_2-2)}$ односторонняя альтернатива: $t_{\text{крит}} = t_{(1-\alpha, df=n_1+n_2-2)}$
Наблюдаемое значение статистики критерия	$t_{\text{набл}} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{общ}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad s_{\text{общ}} = \sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}}$ <p>где \bar{x}_1 – среднее по первой выборке, \bar{x}_2 – среднее по второй выборке, s_1 – стандартное отклонение первой выборки, s_2 – стандартное отклонение второй выборки, n_1 – число элементов в первой выборке, n_2 – число элементов в первой выборке</p>
Проверка	$ t_{\text{набл}} > t_{\text{крит}} \Rightarrow H_0 \text{ отвергается}$ $ t_{\text{набл}} < t_{\text{крит}} \Rightarrow H_0 \text{ не отвергается}$