Проверка гипотез. Часть 1

Статистические гипотезы и статистические критерии. Характеристики критериев. Тест Вальда. Лемма Неймана-Пирсона. Тест Стьюдента.

ПМИ ФКН ВШЭ, 6 октября 2018 г.

Алексей Артемов[1,2]

[1] Сколтех [2] НИУ ВШЭ

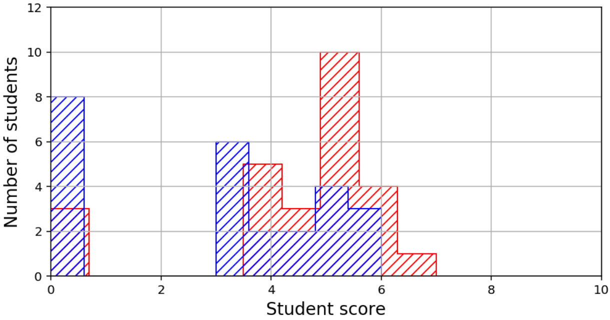
Содержание лекции

- Понятие статистической гипотезы. Нулевая гипотеза и альтернатива.
- Ошибки первого и второго рода. Мощность критерия.
- Как построить нулевую гипотезу.
- Критерий Вальда. P-значение.
- Критерий на основе отношения правдоподобия.
- Критерий Неймана-Пирсона.
- Критерий Стьюдента.

Понятие статистической гипотезы

Мотивирующие приложения

- В первой группе — 26 студентов, а во второй — 24. Средний балл за тест №1 в первой группе 4.4 балла, во второй — 2.81.



- Что можно **спрашивать** об этих данных?

Понятие статистической гипотезы

Определение

Статистическая гипотеза — определенное предположение о распределении вероятностей, лежащем в основе наблюдаемой выборки данных.

- **Простая гипотеза** однозначно определяет функцию распределения на рассматриваемом множестве.
  - **Пример:** $\theta = \theta_0$ – простая гипотеза.
- **Сложная гипотеза** утверждает принадлежность распределения к некоторому множеству распределений на рассматриваемом множестве.
  - **Пример:** $\theta > \theta_0$ или $\theta < \theta_0$ – сложная гипотеза.

Понятие статистической гипотезы

Определение

Проверка статистической гипотезы — это процесс принятия решения о том, противоречит ли рассматриваемая статистическая гипотеза наблюдаемой выборке данных.

- Всегда рассматривается задача проверки гипотезы $\mathbb{H}_0$ против альтернативы $\mathbb{H}_1$ (или же задача $(\mathbb{H}_0, \mathbb{H}_1)$).

Статистический критерий и его характеристики

Односторонние и двусторонние критерии

Определение

Статистический критерий — строгое математическое правило, по которому не отвергается или отвергается статистическая гипотеза.

В зависимости от типа статистической гипотезы выделяют односторонние и двусторонние статистические критерии:

- Односторонний критерий

$$\mathbb{H}_0 : \theta \leqslant \theta_0 \quad \text{vs} \quad \mathbb{H}_1 : \theta > \theta_0.$$

- Двусторонний критерий

$$\mathbb{H}_0 : \theta = \theta_0 \quad \text{vs} \quad \mathbb{H}_1 : \theta \neq \theta_0.$$

Параметрические и непараметрические критерии

Определение

Параметрический критерий — критерий , предполагающий, что выборка порождена распределением из заданного параметрического семейства. В частности, существует много критериев, предназначенных для анализа выборок из нормального распределения.

Определение

Непараметрический критерий — критерий, не опирающийся на дополнительные предположения о распределении.

Критическая область

- Статистический критерий — это правило, которое для каждой реализации выборки должно приводить к одному из двух решений: принять гипотезу $\mathbb{H}_0$ или отклонить ее (принять ее альтернативу $\mathbb{H}_1$).
- В связи с этим каждому критерию соответствует некоторое разбиение выборочного пространства χ на два взаимно дополняющих множества χ_0 и χ_1.
- χ_0 состоит из тех реализаций выборки x, для которых $\mathbb{H}_0$ не отвергается, а χ_1 из тех, для которых $\mathbb{H}_0$ отвергается (принимается $\mathbb{H}_1$).

Критическая область

Определение

В определениях предыдущего слайда

- χ_0 – область принятия гипотезы $\mathbb{H}_0$,
- χ_1 – область ее отклонения – **критическая область.**

- Таким образом любой критерий проверки гипотезы $\mathbb{H}_0$ однозначно задается соответствующей критической областью χ_1.

Общий принцип принятия решений

Определение

Общий принцип принятия решений состоит в следующем:

- Если в эксперименте наблюдается **маловероятное** при справедливости гипотезы $\mathbb{H}_0$ событие, то считается, что гипотеза $\mathbb{H}_0$ **не согласуется** с данными и в этом случае она отклоняется (обнуляется — ср. «нулевая гипотеза»).
- В противном случае считается что данные **не противоречат** $\mathbb{H}_0$ и $\mathbb{H}_0$ принимается.

Уровень значимости

- В соответствии с общим принципом принятия решения критическую область χ_1 выбирают так, чтобы была мала вероятность $\mathrm{P}(\mathbf{X}^\ell \in \chi_1 | \mathbb{H}_0)$.

Определение

Говорят, что критерий **имеет уровень значимости** α, если

$$\mathrm{P}(\mathbf{X}^\ell \in \chi_1 | \mathbb{H}_0) \leqslant \alpha.$$

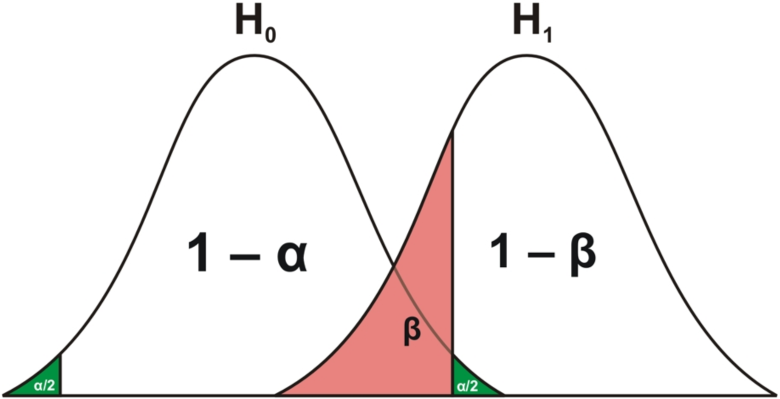
Ошибки при проверке гипотез

- Следуя любому критерию мы можем принять правильное решение, либо совершить одну из двух ошибок — первого или второго рода.

		Верная гипотеза	
		$\mathbb{H}_0$	$\mathbb{H}_1$
Результат применения критерия	$\mathbb{H}_0$	OK	ошибка 2-го рода β
	$\mathbb{H}_1$	ошибка 1-го рода α	OK

- В соответствии с военной терминологией, ошибка 1-го рода — ложная тревога; ошибка 2-го рода — пропуск цели

Ошибки при проверке гипотез



Мощность (чувствительность) критерия

- Пусть критерий имеет критическую область χ_1, а $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$ — множество всех допустимых распределений выборки $\mathbf{X}^\ell$.
- При этом $\mathcal{F}_0$ — множество распределений, удовлетворяющих гипотезе $\mathbb{H}_0$, а $\mathcal{F}_1$ — соответственно $\mathbb{H}_1$.

Определение

Функционал

$$W(F) = W(F; \chi_1) = \mathrm{P}(\mathbf{X}^\ell \in \chi_1 | F), \qquad F \in \mathcal{F}$$

называется **функцией мощности критерия.**

- Другими словами, мощность критерия показывает вероятность попадания значения выборки $\mathbf{X}^\ell$ в критическую область χ_1, когда F — ее истинное распределение.

Связь мощности и ошибок различных родов

Через функцию мощности легко выразить вероятности обоих типов ошибок, свойственных нашему критерию.

Определение

- $W(F)$ — **вероятность ошибки первого рода** при $F \in \mathcal{F}_0$.
- $1 - W(F)$ — **вероятность ошибки второго рода** при $F \in \mathcal{F}_1$.

Размер критерия

Определение

Размер критерия:

$$\alpha = \sup_{F \in \mathcal{F}_0} W(F).$$

- Берется наихудшая из всех процедур, которые могли породить выборку $\mathbf{X}^\ell$ согласно нулевой гипотезе.
- Отсюда легко видеть что если размер критерия не превосходит α, то его уровень значимости равен α.
- Для простых гипотез ($\mathcal{F}_0 = \{F_0\}$), естественно,

$$\alpha = W(F_0) = \mathrm{P}(\mathbf{X}^\ell \in \chi_1 | F_0).$$

Условно-экстремальная постановка

- Логично стремление построить критерий так, чтобы свести к минимуму вероятности ошибок обоих типов.
- Однако при фиксированном объеме выборки сумма вероятностей ошибок обоих типов не может быть сделана сколь угодно малой.
- Поэтому руководствуются рациональным принципом выбора критической области.

Определение (Условно-экстремальная постановка)

Из всех критических областей, **удовлетворяющих заданному уровню значимости ($\leqslant \alpha$)**, выбирается та, для которой **вероятность ошибки 2-го рода минимальна** ($\beta \to \inf$).

Подсчет мощности данного критерия

Пример

- Пусть $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, где σ — известно.
- Рассмотрим 2 гипотезы: $\mathbb{H}_0 : \mu \leqslant 0; \quad \mathbb{H}_1 : \mu > 0.$
- Рассмотрим критерий — $\mathbb{H}_0$ отклоняется, если $T = \overline{X}_n > c$.
- Тогда критическая область $\chi_1 = \{\mathbf{X}^\ell : T(\mathbf{X}^\ell) > c\}$.
- Значит

$$W(\mu) = \mathrm{P}(\overline{X}_n > c) = \mathrm{P}\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} > \frac{\sqrt{n}(c-\mu)}{\sigma}\right)$$

$$= \mathrm{P}\left(Z > \frac{\sqrt{n}(c-\mu)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}(c-\mu)}{\sigma}\right),$$

где Z — стандартная нормальная величина.

Подсчет мощности данного критерия

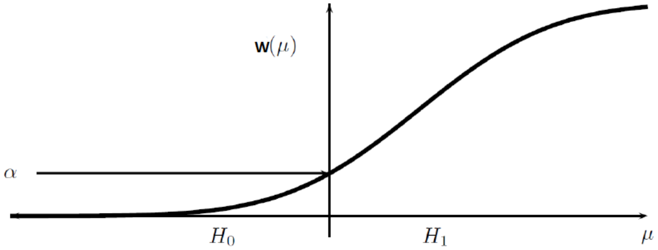


Рис.: $W(\mu)$

- Отсюда легко видеть, что размер критерия равен

$$W(0) = \sup_{\mu \in \mathbb{R}} \left\{ 1 - \Phi\left(\frac{\sqrt{n}(c-\mu)}{\sigma} \right) \right\} = 1 - \Phi\left(\frac{\sqrt{n}c}{\sigma} \right).$$

Статистический критерий и его характеристики

Пример

Чтобы размер критерия равнялся α необходимо, чтобы $c = \frac{\sigma\Phi^{-1}(1-\alpha)}{\sqrt{n}}$.

Определение

Наиболее мощный критерий — критерий, который имеет максималь мощность относительно гипотезы $\mathbb{H}_1$ (т.е максимальную $1-W(F)$ при $F \in \mathcal{F}_1$) среди всех критериев размера α.

Статистический критерий и его характеристики

- В конкретных задачах наиболее мощный критерий не всегда достижим, поэтому в реальных задачах часто приходится ограничиваться более умеренными требованиями.
- Минимальным таким требованием является требование несмещенности.

Определение

Статистический критерий называется несмещенным, если при любом альтернативном распределении данных мы должны попадать в критическую область с большей вероятностью, нежели при нулевой гипотезе.

Статистический критерий и его характеристики

В случае выборок большого объема важным также является условие состоятельности.

Определение
Статистический критерий называется состоятельным, если в случае истинности альтернативной гипотезы при большом числе наблюдений мы будем попадать в критическую область с вероятность близкой к 1 (т.е., отклоняя нулевую гипотезу, мы будем принимать правильное решение).

Нулевая гипотеза: построение

- Hypothesis: “Adding water to toothpaste protects against cavities.”

Нулевая гипотеза: построение

- Hypothesis: “Adding water to toothpaste protects against cavities.”
- Null hypothesis ($\mathbb{H}_0$): “Adding water does not make toothpaste more effective in fighting cavities.”
- Hypothesis: “The evidence produced before the court proves that this man is guilty.”

Нулевая гипотеза: построение

- Hypothesis: “Adding water to toothpaste protects against cavities.”
- Null hypothesis ($\mathbb{H}_0$): “Adding water does not make toothpaste more effective in fighting cavities.”
- Hypothesis: “The evidence produced before the court proves that this man is guilty.”
- Null hypothesis ($\mathbb{H}_0$): “This man is innocent.”

Нулевая гипотеза: построение

- Hypothesis: “Adding water to toothpaste protects against cavities.”
- Null hypothesis ($\mathbb{H}_0$): “Adding water does not make toothpaste more effective in fighting cavities.”
- Hypothesis: “The evidence produced before the court proves that this man is guilty.”
- Null hypothesis ($\mathbb{H}_0$): “This man is innocent.”
- Ситуация «по умолчанию» — похожа на ситуацию «презумпции невиновности».

Нулевая гипотеза и фальсифицируемость

- “Фальсифицируемость (принципиальная опровержимость утверждения, опровергаемость, критерий Поппера) — критерий научности эмпирической или иной теории, претендующей на научность.”[Википедия]
- **Научная теория не может быть принципиально неопровержимой.**
- Неопровержимость — признак лженауки (точнее, объявление нефальсифицируемой теории научной — признак лженауки).
- “Если **нельзя замыслить, придумать опыт,** в результате которого гипотеза может оказаться не верна, — это антинаучная@#$%, как бы красиво она ни звучала.”

Нулевая гипотеза и фальсифицируемость

- Являются ли фальсифицируемыми утверждения:

Нулевая гипотеза и фальсифицируемость

- Являются ли фальсифицируемыми утверждения:
  - “поддержка партии власти — 65%”?

Нулевая гипотеза и фальсифицируемость

- Являются ли фальсифицируемыми утверждения:
    - “поддержка партии власти — 65%”?
    - “все лебеди — белые”?

Нулевая гипотеза и фальсифицируемость

- Являются ли фальсифицируемыми утверждения:
    - “поддержка партии власти — 65%”?
    - “все лебеди — белые”?
    - “гороскоп определяет судьбу человека”?

Нулевая гипотеза и фальсифицируемость

- Являются ли фальсифицируемыми утверждения:
  - “поддержка партии власти — 65%”?
  - “все лебеди — белые”?
  - “гороскоп определяет судьбу человека”?
  - “рисунок кожного рельефа ладоней определяет черты характера человека”?

Нулевая гипотеза и фальсифицируемость

- Являются ли фальсифицируемыми утверждения:
  - “поддержка партии власти — 65%”?
  - “все лебеди — белые”?
  - “гороскоп определяет судьбу человека”?
  - “рисунок кожного рельефа ладоней определяет черты характера человека”?
  - … you name it
- Для любопытных: `http://lurkmore.to/Критерий_Поппера`

Действия в задачах проверки гипотез

1. Сформулировать (математически) нулевую гипотезу $\mathbb{H}_0$.
2. Сформулировать (математически) альтернативу.
3. Выбрать уровень значимости α.
4. Получить выборку $\mathbf{X}^\ell$.
5. Подсчитать значение статистики $T = T(\mathbf{X}^\ell)$.
6. Построить критическую область $\mathcal{R}_\alpha$.
7. На основе шагов 5—6 сделать вывод **о согласии $\mathbb{H}_0$ с данными $\mathbf{X}^\ell$.**

Типы статистических критериев

Типы статистических критериев

После краткого введения в теорию проверки гипотез пререйдем непосредственно к разлизным статистическим критериям. В данной лекции будут рассмотретны следующие статистические критерии:

- Критерий Вальда.
- P-значение.
- Тестирование на основе доверительного интервала.
- Критерий на основе отношения правдоподобия.
- Критерий Неймана-Пирсона для случая двух простых гипотез.
- Критерий Стьюдента.

Критерий Вальда (Z-test)

Пусть:

- θ — скалярный параметр;
- $\widehat{\theta}$ — его оценка;
- $\widehat{se}$ — оценка стандартной ошибки оценки $\widehat{\theta}$.

Гипотеза:

$$\mathbb{H}_0 : \theta = \theta_0 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq \theta_0.$$

Определение (Критерий Вальда размера α)

Если $\widehat{\theta}$ — асимптотически нормальная оценка параметра θ, т.е.

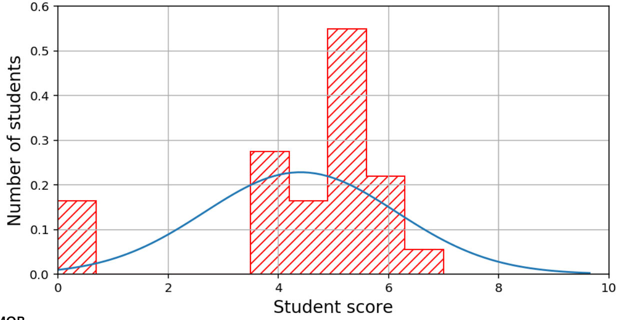
$$W = \frac{\widehat{\theta} - \theta_0}{\widehat{se}} \to \mathcal{N}(0, 1), \qquad n \to \infty,$$

то гипотеза $\mathbb{H}_0$ отклоняется, если $|W| > z_{\alpha/2}$.

Критерий Вальда (Z-test)

Рассмотрим выборку:

```
group1 = [
    6, 4, 4, 0, 5, 5, 7, 5, 4.5, 5,
    4, 4, 4.5, 0, 5, 5, 5, 5, 4, 5,
    0, 6, 5, 6, 4.5, 6]
```



Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу

$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$

Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу
$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$
- Как получить $\widehat{\theta}$ — оценку среднего?

Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу

$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$

- Как получить $\widehat{\theta}$ — оценку среднего?
  $\widehat{\theta} = 4.4$
- Как получить $\widehat{se}$ — оценку стандартной ошибки оценки $\widehat{\theta}$?

Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу
$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$
- Как получить $\widehat{\theta}$ — оценку среднего?
$\widehat{\theta} = 4.4$
- Как получить $\widehat{se}$ — оценку стандартной ошибки оценки $\widehat{\theta}$?
$\widehat{se} = 1.748/\sqrt{26} = 0.342$
- Как получить $\widehat{se}$, если — оценка стандартной ошибки оценки $\widehat{\theta}$.

Критерий Вальда (Z-test)

- Пусть $X_i \sim \mathcal{N}(\theta, \sigma^2)$. Проверим гипотезу
$$\mathbb{H}_0 : \theta = 5 \quad \text{vs.} \quad \mathbb{H}_1 : \theta \neq 5.$$
- Как получить $\widehat{\theta}$ — оценку среднего?
$\widehat{\theta} = 4.4$
- Как получить $\widehat{se}$ — оценку стандартной ошибки оценки $\widehat{\theta}$?
$\widehat{se} = 1.748/\sqrt{26} = 0.342$
- Как получить $\widehat{se}$, если — оценка стандартной ошибки оценки $\widehat{\theta}$.
Например, бутстреп! (Для более сложных задач — информация Фишера, дельта-метод, ...)
- При $\alpha = 0.05$, поскольку $z_{\alpha/2} = 1.96$, имеем
$$W = \frac{\widehat{\theta} - \theta}{\widehat{se}} = \frac{4.4 - 5}{0.342} = -1.738$$

Теорема

Асимптотически размер критерия Вальда равен α, то есть

$$W(F) = \mathrm{P}(|W| > z_{\alpha/2}|f) \to \alpha, \;\; n \to \infty, \;\; f \in \mathcal{F}_0.$$

Доказательство.

При условии, что $\theta = \theta_0$, в силу асимптотической нормальности оценки выполнено $\frac{\widehat{\theta}-\theta_0}{\widehat{se}} \rightsquigarrow \mathcal{N}(0,1)$. Следовательно, вероятность отклонить основную гипотезу, когда она на самом деле верна, равняется:

$$\mathrm{P}(|W| > z_{\alpha/2}|f) = \mathrm{P}\left(\frac{|\widehat{\theta} - \theta_0|}{\widehat{se}} > z_{\alpha/2}|f\right) \to$$
$$\to \mathrm{P}(|Z| > z_{\alpha/2}) = \alpha.$$

$\square$

Пример (сравнение средних значений)

- Пусть $X_1, ..., X_m$ и $Y_1, ..., Y_n$ — две независимые выборки из генеральных совокупностей.
- Средние значения равны μ_1 и μ_2 соответственно.
- $\widehat{s}_1^2$ и $\widehat{s}_2^2$ — выборочные дисперсии.
- Положим $\delta = \mu_1 - \mu_2$.
- Проверим гипотезу $\mathbb{H}_0 : \delta = 0$ vs. $\mathbb{H}_1 : \delta \neq 0$
- Построим статистики

$$\widehat{\delta} = \overline{X}_m - \overline{Y}_n; \qquad \widehat{se} = \sqrt{\frac{\widehat{s}_1^2}{m} + \frac{\widehat{s}_2^2}{n}}.$$

- Гипотеза $\mathbb{H}_0$ отвергается, если $|W| > z_{\alpha/2}$, где

$$W = \frac{\widehat{\delta} - 0}{\widehat{se}} = \frac{\overline{X}_m - \overline{Y}_n}{\sqrt{\frac{\widehat{s}_1^2}{m} + \frac{\widehat{s}_2^2}{n}}}.$$

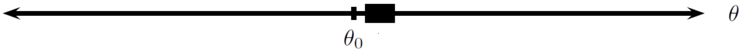
Доверительные интервалы

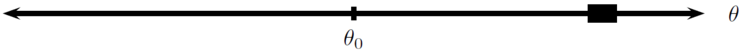
Теорема

Критерий Вальда размера α отклоняет гипотезу $\mathbb{H}_0 : \theta = \theta_0$ в пользу $\mathbb{H}_1 : \theta \neq \theta_0$ если и только если $\theta_0 \notin C$, где

$$C = (\widehat{\theta} - \widehat{se} z_{\alpha/2}, \widehat{\theta} + \widehat{se} z_{\alpha/2})$$

Таким образом, тестирование гипотезы эквивалентно проверке, попало ли значение Θ_0 в доверительный интервал.





P-значение

Определение

Пусть для каждого $\alpha \in (0, 1)$ имеется критерий размера α для некоторой статистики (функции от выборки) $T(\mathbf{X}^\ell)$ с критической областью $\mathcal{R}_\alpha$. Тогда

$$p - \text{value} = \inf\{\alpha : T(\mathbf{X}^\ell) \in \mathcal{R}_\alpha\}.$$

- Таким образом, p-value - это наименьший уровень значимости, на котором еще можно отклонить $\mathbb{H}_0$.
- Чем меньше p-value - тем вероятнее, что $\mathbb{H}_0$ надо отклонить.

P-значение, rules of thumb

- Типичные значения для p-value:
  - $p < 0.01 \to \mathbb{H}_0$ — заведомо не верна
  - $p \sim 0.01 - 0.05 \to \mathbb{H}_0$ — не верна
  - $p \sim 0.05 - 0.10 \to \mathbb{H}_0$ — скорее не верна
  - $p > 0.1 \to$ ничего определенного о гипотезе $\mathbb{H}_0$ сказать нельзя
- Большое p-value не является подтверждением гипотезы $\mathbb{H}_0$. Большое p-value появляется, если:
  - $\mathbb{H}_0$ — верна
  - $\mathbb{H}_0$ — неверна, но мощность критерия недостаточна

P-значение, примеры из статей

- “Our original co-occurrence-based semantic model predicts voxel responses significantly better than the simplified model (paired t-test across all voxels; $t = 170$, $p < 1e - 16$).”

P-значение, примеры из статей

- “Our original co-occurrence-based semantic model predicts voxel responses significantly better than the simplified model (paired t-test across all voxels; $t = 170$, $p < 1e - 16$).”
- “However, we did not find a significant correlation between dextrality and PrAGMATiC generalization scores (Pearson’s $r = -0.20$, p-value$= 0.66$ for the left hemisphere; $r = -0.06$, p-value $= 0.90$ for the right).”

P-значение, примеры из статей

- "Our original co-occurrence-based semantic model predicts voxel responses significantly better than the simplified model (paired t-test across all voxels; $t = 170$, $p < 1e - 16$)."
- "However, we did not find a significant correlation between dextrality and PrAGMATiC generalization scores (Pearson's $r = -0.20$, p-value$= 0.66$ for the left hemisphere; $r = -0.06$, p-value $= 0.90$ for the right)."
- "At least four dimensions explained a significant amount of variance ($P < 0.001$, Bonferroni-corrected bootstrap test) in all but one subject; in the last subject only three dimensions were significant (Extended Data Fig. 2)".

Теорема

- Пусть критерий размера α, построенный для статистики $T(\mathbf{X}^\ell)$, имеет вид: $\mathbb{H}_0$ отвергается, если $T(\mathbf{X}^\ell) > c_\alpha$
- Гипотезе $\mathbb{H}_0$ соответствует семейство распределений $\mathcal{F}_0$.
- Тогда

$$p\text{-value} = \sup_{f \in \mathcal{F}_0} \mathrm{P}(T(\mathbf{X}^\ell) \geqslant T(\mathbf{x}^\ell) | f),$$

  где $\mathbf{X}^\ell$ – реализация выборки $\mathbf{x}^\ell$. Если $\mathcal{F}_0 = f$, то

$$p\text{-value} = \mathrm{P}(T(\mathbf{X}^\ell) \geqslant T(\mathbf{x}^\ell) | f).$$

- То есть p-value — это вероятность (при выполнении гипотезы $\mathbb{H}_0$) того, что статистика $T(\mathbf{X}^\ell)$ примет значение больше либо равное тому, которое реализовалось в опыте (реализация $\mathbf{X}^\ell$)

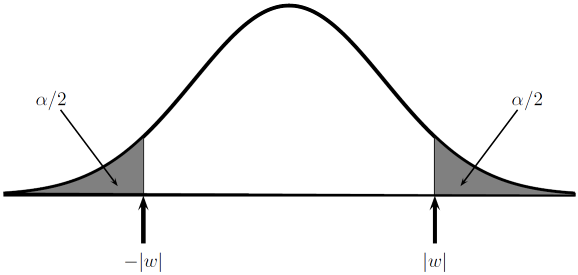
P-значение для критерия Вальда

Теорема

Пусть $w = \frac{(\widehat{\theta} - \theta_0)}{\widehat{se}}$ — наблюдаемое значение статистики Вальда W. Тогда:

$$p\text{-value} = \mathrm{P}(|W| > |w| | f) \simeq \mathrm{P}(|Z| > |w|) = 2\Phi(-|w|),$$

где $Z \sim N(0, 1)$, $\quad f \in \mathcal{F}_0$.



Пример (Равенство значений холестерина в крови)

Здесь как и в примере про сравнение средних значений:

$$W = \frac{\widehat{\delta} - 0}{\widehat{se}} = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{216.2 - 195.3}{\sqrt{5^2 + 2.4^2}} = 3.78.$$

Пусть $Z \sim N(0,1)$, тогда

$$p - value = \mathrm{P}(|Z| > 3.78) = 2 \cdot \mathrm{P}(Z < -3.78) = 0.0002$$

.

Критерий на основе отношения правдоподобия

Определение

Рассмотрим две конкурирующие гипотезы

$$\mathbb{H}_0 : f \in \mathcal{F}_0 \quad \text{vs.} \quad \mathbb{H}_1 : f \in \mathcal{F}_1$$

Пусть $\widehat{f}$ — ОМП и $\widehat{f}_0$ — ОМП при $f \in \mathcal{F}_0$ Статистика отношения правдоподобия имеет вид:

$$\lambda = 2 \log \frac{\sup_{f \in \mathcal{F}} \mathcal{L}(f)}{\sup_{f \in \mathcal{F}_0} \mathcal{L}(f)} = 2 \log \frac{\mathcal{L}(\widehat{f})}{\mathcal{L}(\widehat{f}_0)}.$$

Распределение хи-квадрат (χ^2)

Определение (распределение хи-квадрат(χ^2))

Пусть $Z_1, ..., Z_k$ - независимые стандартно нормально распределенн
случайные величины. $V = \sum_{i=1}^{k} Z_i^2$, тогда $V \sim \chi_k^2$ - хи-квадрат с k степенями свободы

$$F(V) = \frac{v^{\frac{k}{2}-1} e^{-\frac{V}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})}, \quad \mathbb{E}(V) = k, \quad \mathbb{V}(V) = 2k$$

$\chi_{k,\alpha}^2 = F^{-1}(1-\alpha)$ - верхняя квантиль, F - функция распределения, т.е.

$$\mathrm{P}(\chi_k^2 > \chi_{k,\alpha}^2) = \alpha.$$

Теорема

Допустим, что $f = (\theta_1, \ldots, \theta_q, \theta_{q+1}, ..., \theta_r)$. Пусть $\mathcal{F}_0 = \{f : (\theta_{q+1}, \ldots, \theta_r) = (\theta_{0,q+1}, ..., \theta_{0,r})\}$.
Пусть λ - критерий на основе отношения правдоподобия. При гипотезе $\mathbb{H}_0 : f \in \mathcal{F}_0$

$$\lambda(\mathbf{X}^\ell) \rightsquigarrow \chi^2_{q,\alpha},$$

где q - размерность F за вычетом размерности $\mathcal{F}_0$. p-value для критерия равно $\mathrm{P}(\chi^2_q > \lambda)$.

Пример

Пусть $f = (\theta_1, \ldots, \theta_5)$, необходимо проверить, что $\theta_4 = \theta_5 = 0$. Тогда у предельного распределения имеется 3 степени свободы.

Пример (Горох Менделя $1/3$)

Пример. Горох Менделя. Два типа: круглые желтые зерна и сморщенные зеленые зерна. Имеется 4 типа потомков: круглые желтые, сморщенные желтые, круглые зеленые и сморщенные зеленые. Количество потомков каждого типа образуют мультиномиальное распределение с вероятностью $p = (p_1, p_2, p_3, p_4)$.

Из теории следует, что

$$p = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right).$$

В опыте получено, что $n = 556, X = (315, 101, 108, 32)$.

Пример (Горох Менделя $2/3$)

Статистика отношений правдоподобия для

$$\mathbb{H}_0 : p = p_0 \quad vs. \quad \mathbb{H}_1 : p \neq p_0$$

принимает вид:

$$\lambda = 2\log\frac{\mathcal{L}(\widehat{p})}{\mathcal{L}(\widehat{p_0})} = 2\sum_{j=1}^{4} X_j \log\frac{\mathcal{L}(\widehat{p})}{\mathcal{L}(\widehat{p_0})} = 2\cdot(315\log(\frac{315/556}{9/16}) +$$

$$+101\log(\frac{101/556}{3/16}) + 108\log(\frac{108/556}{3/16}) + 32\log(\frac{32/556}{1/16})) = 0.48$$

Пример (Горох Менделя $3/3$)

При гипотезе $\mathbb{H}_1$ 4 параметра. Так как сумма параметров должна равняться 1, то размерность пространства параметров равна 3. При гипотезе $\mathbb{H}_0$ свободных параметров нет, значит количество степеней свободы равно 3 и χ_3^3 является предельным распределением.

$$p - value = \mathrm{P}(\chi_3^2 > 0.48) = 0.92$$

Замечание

Как правило, и критерий χ^2, и критерий отношения правдоподобий дают примерно одинаковые результаты при условии, что размер выборки достаточно большой.

Критерий Неймана-Пирсона для случая двух простых гипотез

Лемма (Неймана-Пирсона)

$\mathbb{H}_0 : \theta = \theta_0 \quad vs. \quad \mathbb{H}_1 : \theta = \theta_1$

Статистика Неймана-Пирсона:

$$T = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = \frac{\prod_{i=1}^{n} f(x_i; \theta_1)}{\prod_{i=1}^{n} f(x_i; \theta_0)}. \tag{1}$$

Допустим, что $\mathbb{H}_0$ отвергается при $T > k$. Выберем k так, что $\mathrm{P}_{\theta_0}(T > k) = \alpha$.

Тогда, критерий Неймана-Пирсона (на основе статистики (1)) будет иметь наибольшую мощность $W(\theta_1)$ среди всех критериев размера α.

Критерий Стьюдента (t-test)

Определение

Случайная величина имеет распределение Стьюдента (t-распределен
с k степенями свободы, если:

$$f(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})(1+\frac{t^2}{k})^{\frac{k+1}{2}}}$$

При $k \to \infty$ t-распределение стремится к стандартному нормальному
распределению. При $k = 1$ t-распределение совпадает с распределен
Коши.

t-критерий используют, когда распределение данных близко к нормальному, а размер выборки невелик.

Критерий Стьюдента (t-test)

Теорема (Критерий Стьюдента (t-test))

Пусть $X_1, \dots, X_n \sim N(\mu, \sigma^2)$, где параметры (μ, σ^2) неизвестны.

$$\mathbb{H}_0 : \mu = \mu_0 \quad vs. \quad \mathbb{H}_1 : \mu \neq \mu_0$$

Обозначим через S_n^2 выборочную дисперсию. Тогда статистика критерия:

$$T = \frac{\sqrt{n}(\overline{X_n} - \mu_0)}{S_n}$$

Основная гипотеза отвергается, если $|T| > t_{n-1,\alpha/2}$, где $t_{n-1,\alpha/2}$ – квантиль распределения Стьюдента с $\mathbf{n-1}$ степенями свободы.

При больших n выполняется $T \sim \mathcal{N}(0,1)$, то есть при больших n t-критерий эквивалентен критерию Вальда.

Критерий Стьюдента (t-test)

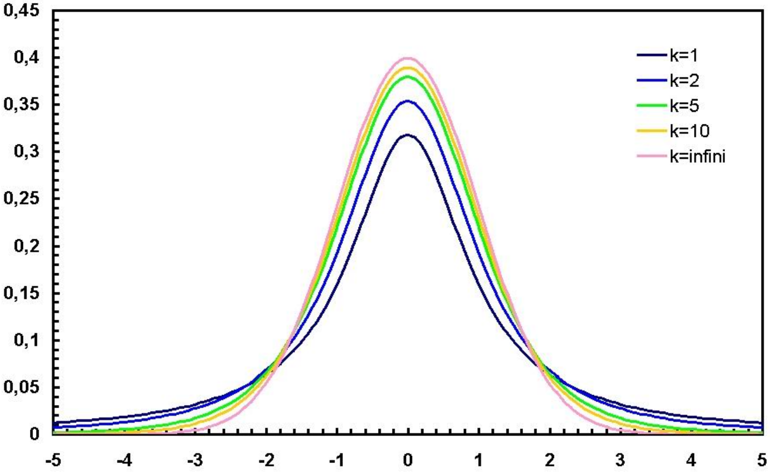


Рис.: https://ru.wikipedia.org/wiki/Распределение_Стьюдента

Критерий Стьюдента (t-test)

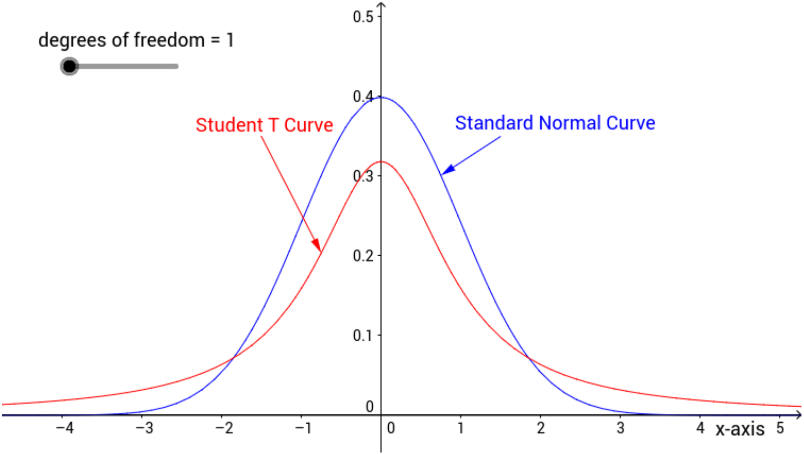


Рис.: `http://tananyag.geomatech.hu/m/53882`

Критерий перестановок

Определение (Постановка задачи)

Критерий перестановок применяется для проверки того, отличаются ли распределения.
Пусть $X_1, ...X_m \sim F_X$ и $Y_1, ..., Y_n \sim F_Y$ - две независимые выборки. Требуется решить:

$$\mathbb{H}_0 : F_X = F_Y \quad vs. \quad \mathbb{H}_1 : F_X \neq F_Y$$

Критерий перестановок - ”точный” в том смысле, что он не использует предположения об асимптотической сходимости к нормальному распределению.

Критерий перестановок:

1. Обозначим через $T(x_1, ..., x_m, y_1, ..., y_n)$ некоторую тестовую статистику, например, $T(X_1, ..., X_m, Y_1, ..., Y_n) = |\overline{X}_m - \overline{Y}_n|$.
2. Положим $N = m + n$ и рассмотрим все $N!$ перестановок объединенной выборки $X_1, ..., X_m, Y_1, ..., Y_n$.
3. Для каждой из перестановок подсчитаем значение статистики T.
4. Обозначим эти значения $T_1, ..., T_{N!}$.

Теорема (Критерий перестановок)

Если $\mathbb{H}_0$ верна, то при фиксированных упорядоченных значениях $\{X_1, ..., X_m, Y_1, ..., Y_n\}$ значение статистики T распределены равном на множестве $T_1, ..., T_{N!}$.

Теорема

Обозначим как перестановочное распределение статистики T такое, согласно которому:

$$P_0(T = T_i) = \frac{1}{N!}, \quad i = 1, ..., N!$$

Пусть t_{obs} - значение статистики, которое было получено в опыте. Тогда:

$$p - value = \mathrm{P}(T > t_{obs}|f) = \frac{1}{N!}\sum_{j=1}^{N!}\mathbb{I}(T_j > t_{obs}), \quad f \in \mathcal{F}_0$$

Пример

Допустим, что $(X_1, X_2, Y_1) = (1, 9, 3)$. Пусть $T(X_1, X_2, Y_1) = |\overline{X} - \overline{Y}| = 2$, тогда

Перестановка	Значение T	Вероятность
(1,9,3)	2	1/6
(9,1,3)	2	1/6
(1,3,9)	7	1/6
(3,1,9)	7	1/6
(3,9,1)	5	1/6
(9,3,1)	5	1/6

$$p - value = \mathrm{P}(T > 2) = 4/6$$