Гипотезы

Майнор ИАД, 20 февраля 2020 г.

Денис Деркач

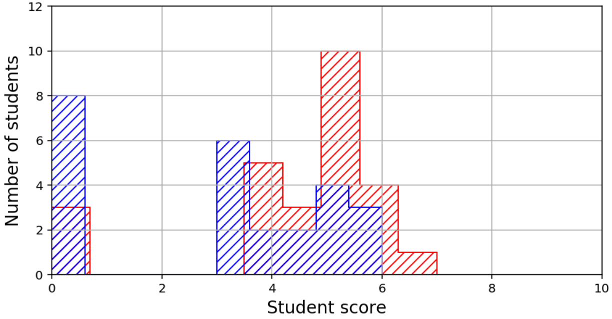
Лаборатория Lambda, ФКН ВШЭ

Оглавление

Понятие статистической гипотезы

Мотивирующий пример

- 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 — определённое 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.

Проверка статистической гипотезы

Определение

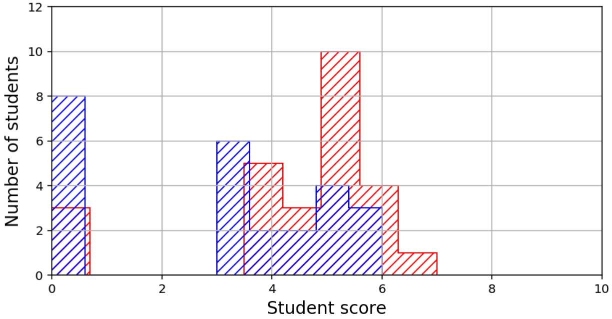
Проверка статистической гипотезы — это процесс принятия решения о том, противоречит ли рассматриваемая статистическая гипотеза наблюдаемой выборке данных.

Всегда рассматривается задача проверки:

- нулевой гипотезы H_0;
- против альтернативной гипотезы H_1.

Пример

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



- H_0 : студенты в обеих группах одинаково знают материал;
- H_1 : студенты знают материал по-разному.

Формулировки и правила

- мы говорим про ”отклонение H_0”, ”принятие H_1” не рассматривается;
- мы говорим ”имеющихся количества данных недостаточно для отклонение H_0”;
- мы тестируем только пару $(H_0; H_1)$ и ничего не знаем о неописанных состояниях системы:
  - пример: контрольная работа была слишком простая и студенты получили максимум.

Статистический критерий

Статистический критерий

Определение

Статистический критерий — правило, по которому на основании реализации выборки $(x_1, x_2, \ldots, x_n)$, принимается или отвергается статистическая гипотеза.

Обычно критерий задаётся при помощи статистики критерия $T(x_1, x_2, .., x_n)$.

Пример: взболтать, но не смешивать

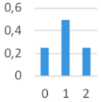
- Джеймс Бонд говорит, что предпочитает мартини взболтанным, но не смешанным.
- Проведём слепой тест: предложим ему n раз пару напитков и спросим, какой из двух он предпочитает
- Выборка: объём из n нулей и единиц (предпочёл взболтанный — 1, смешанный — 0), получается вектор.
- Нулевая гипотеза: Джеймс Бонд выбирает наугад.
- Статистика: число единиц в выборке.

Векторы для n=2:

0	0
0	1
1	0
1	1

Статистика:

0
1
1
2



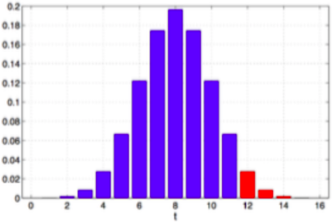
Распределение статистики

Денис Деркач

Статистика шотов

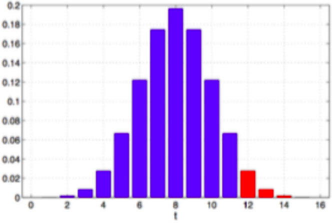
- Если нулевая гипотеза справедлива и Джеймс Бонд не различает напитки, то все исходы равновероятны.
- Пусть $n = 16$ — тогда существует 65535 различных бинарных исходов.

Распределение количества единиц в векторе исходов.

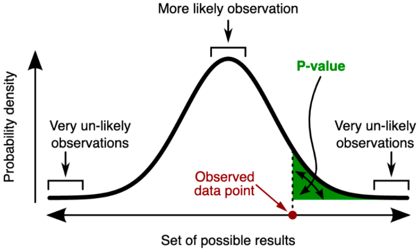


Эксперимент

- Провели эксперимент — Джеймс Бонд выбрал взболтанный мартини в 12 из 16 раз.
- Вероятность того, что он выберет взболтанный мартини 12 раз или больше при условии, что выбирает наугад: $2512/65536 \approx 0.0384$
- Отклоняем ли мы гипотезу о том, что Джеймс Бонд не разбирается в мартини?



p-значение

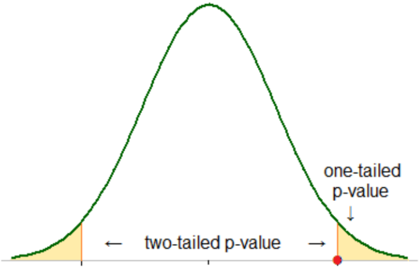


Определение

p-значением называется вероятность того, что статистика может быть такой или более экстремальной при верной нулевой гипотезе.

В примере с Бондом мы уже её подсчитали.

Типы критериев



- Односторонний критерий

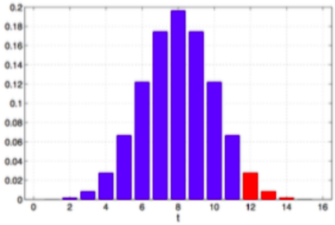
$$H_0 \ : \theta \leq \theta_0 \ \ vs \ \ H_1 \ : \theta > \theta_0.$$

- Двусторонний критерий

$$H_0 \ : \theta = \theta_0 \ \ vs \ \ H_1 \ : \theta \neq \theta_0.$$

Эксперимент

- Провели эксперимент — Джеймс Бонд выбрал взболтанный мартини в 12 из 16 раз.
- Вероятность того, что он выберет взболтанный мартини 12 раз или больше при условии, что выбирает наугад: $2512/65536 \approx 0.0384$
- Отклоняем ли мы гипотезу о том, что Джеймс Бонд не разбирается в мартини?



Критическая область

Нам надо **заранее задать** значения статистики, где гипотеза будет отклонена.

Определение

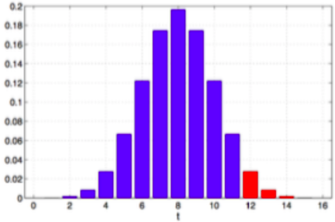
- χ_0 — область принятия гипотезы H_0,
- χ_1 — область ее отклонения (критическая область).

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

$$\mathbb{P}(T \in \chi_1 | \mathbb{H}_0) \leqslant \alpha.$$

Эксперимент

- Провели эксперимент — Джеймс Бонд выбрал взболтанный мартини в 12 из 16 раз.
- Вероятность того, что он выберет взболтанный мартини 12 раз или больше при условии, что выбирает наугад: $2512/65536 \approx 0.0384$
- Отклоняем ли мы гипотезу о том, что Джеймс Бонд не разбирается в мартини?
- Да, если заранее определили уровень значимости $\alpha = 0.05$.



Ход проверки гипотез

1. Формулируем задачу.
2. Формулируем нулевую и альтернативную гипотезу.
3. Задаём критическое значение.
4. Ищем нужный статистический критерий.
5. Ищем правильное распределение статистического критерия.
6. Проводим измерение.
7. Ищем p-значение.
8. Сравниваем его с критическим значением.
9. Делаем вывод.

Проблемы использования p-значений

p-значение очень часто неправильно интерпретируется как что-то полностью описывающее эксперимент.

- Don't base your conclusions solely on whether an association or effect was found to be "statistically significant" (i.e., the p-value passed some arbitrary threshold such as $p < 0.05$).
- Don't believe that an association or effect exists just because it was statistically significant.
- Don't believe that an association or effect is absent just because it was not statistically significant.
- Don't believe that your p-value gives the probability that chance alone produced the observed association or effect or the probability that your test hypothesis is true.
- Don't conclude anything about scientific or practical importance based on statistical significance (or lack thereof).

Don't. Don't. Just…don't. Yes, we talk a lot about don'ts. The *ASA*

Внимательно следите за тем, как вы интерпретируете результаты эксперимента.

Ошибки

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

		Верная гипотеза	
		H_0	H_1
Результат применения	H_0	OK	ошибка 2-го рода
критерия	H_1	ошибка 1-го рода	OK

Примеры

H_0 : not-pregnant!

Мощность теста

Определение

Мощность критерия показывает вероятность попадания значения критерия T в критическую область, когда F — ее истинное распределение.

Чем мощнее критерий, тем лучше для нас. В идеале мы должны использовать самый мощный критерий.

Выбор критерия

- 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.

Из всех критических областей удовлетворяющих заданному уровню значимости выбирается та, для которой вероятность ошибки 2-го рода минимальна.

Параметрические тесты

Описание критерия

- В примере с Бондом мы просчитывали все возможные варианты.
- В реальной жизни это сделать трудно.
- Зато у нас есть некоторые соображения как правильно выбирать критерии под конкретную задачу. Это автоматически будет давать нужные распределения.

Типы критериев

- Одновыборочные: например, среднее равно нулю.
- Двухвыборочные: например, среднее одной выборки равно среднему другой.
- Согласия: например, эта кривая хорошо описывает выборку.
- Распределения: например, эта выборка из нормального распределения.

Критерий Стьюдента

- Пусть $X, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, где (μ, σ^2) неизвестны.
- Задача
$$\mathbb{H}_0 : \mu = \mu_0 \quad vs. \quad \mathbb{H}_1 : \mu \neq \mu_0$$
- Обозначим через S_n^2 выборочную дисперсию. 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 $n-1$ 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)

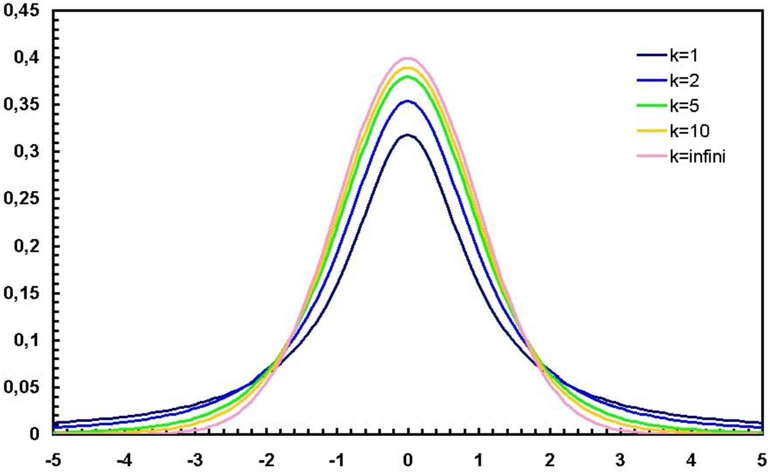


Рис.: https://ru.wikipedia.org/wiki/Распределение_Стьюдента

Критерий Стьюдента (t-test)

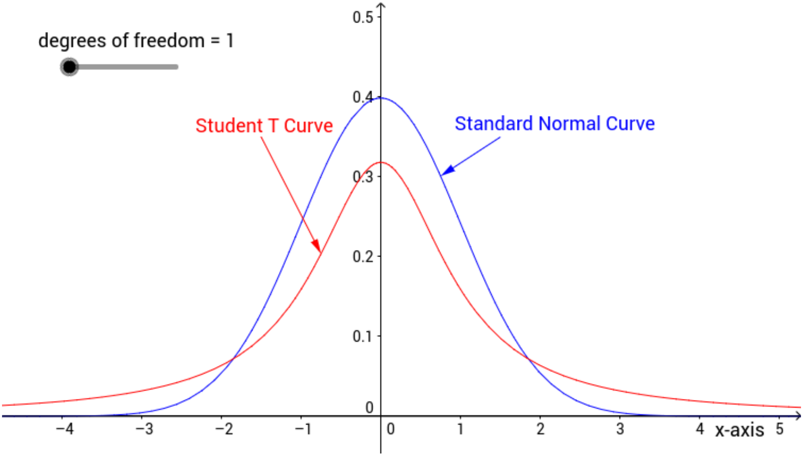


Рис.: `http://tananyag.geomatech.hu/m/53882`

Двухвыборочный t-критерий

- $\{X_i\}_{i=1}^n, \{Y_j\}_{j=1}^m$ — две выборки из нормальных распределений $\mathcal{N}(\mu_X, \sigma_X^2), \mathcal{N}(\mu_Y, \sigma_Y^2)$
- Задача

$$\mathbb{H}_0 : \mu_X = \mu_Y \quad \text{vs.} \quad \mathbb{H}_1 : \mu_X \neq \mu_Y$$

- Механика проверки гипотезы та же, что и раньше.

Непараметрические тесты

Непараметрические тесты

- Параметрические тесты хороши, когда известны распределения выборки.
- Например, t-тест работает для нормального распределения.
- Для остальных случаев есть непараметрические тесты.

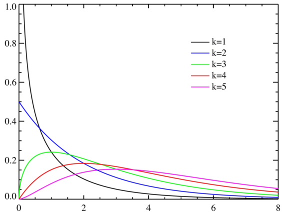
χ^2 тест Пирсона

- Тест согласия с распределением/функцией.
- Составляется как

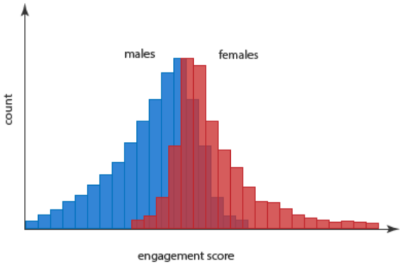
$$\chi^2 = N \sum_{i=1}^{n} \frac{(O_i/N - p_i)^2}{p_i}.$$

здесь N - количество событий всего, O_i - количество событий в этой точке, p_i ожидаемая частота событий.

- распределено как χ^2 с $N-1$ степенью свободы.



Ненормальность данных



- Иногда нам нужно сравнить две выборки, которые явно ненормальны.
- t-тест не работает, что делать?
- Использовать критерий Манна-Уитни-Вилкоксона.

Критерий Манна-Уитни-Вилкоксона

1. Выстроить обе выборки в одну.
2. Присвоить номер по возрастанию.
3. Найти средний номер в выборках (R1 или R2).
4. Подсчитать:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}$$

5. Получить критическое значение (U распределение, для больших семплов, нормальное), сравнить с p-value.

NB: Этот критерий сравнивает распределения, а не средние.

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 $T_1, ..., T_{N!}$.

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

Критерий перестановок

Для критерия перестановок количество более критических значений и есть p-значения.

$$p\text{-value} = \mathrm{P}(T > t_{obs}) = \frac{1}{N!}\sum_{j=1}^{N!}\mathbb{I}(T_j > t_{obs})$$

Критерий перестановок: пример

- Пусть $(X_1, X_2, Y_1) = (1, 9, 3)$.
- 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 $= \mathrm{P}(T > 2) = 4/6$.

Бутстрэп и Гипотезы

- В критерии перестановок мы просто перемешивали события.
- В принципе, мы можем вытаскивать произвольные события с возвращением и получать новые выборки.
- Это будет бутстрэпным тестом.

Финальный слайд

- Тестирование гипотез — центральная задача статистики.
- Для каждой задачи скорее всего есть специальный тест, нужно его просто сформулировать.