

Проверка статистических гипотез

Часть 2

Аббакумов Вадим Леонардович

Версия 4

Раздел 2

Технологии проверки статистических гипотез

Основные понятия

Определение (напоминание)

- Статистическая гипотеза – утверждение о свойствах распределения вероятностей случайной величины *(или случайного вектора)*.
- Проверку гипотез на основе выборочных статистических данных называют статистической проверкой гипотез.

Не умеем!

Гипотеза справедлива с вероятностью 0.86

Выбираем из двух гипотез!

- Гипотеза принимается или отвергается
 - заменяем на
 - Выбираем между двумя статистическими гипотезами.
-

Основная и альтернативная гипотезы

- Основная гипотеза
- Альтернативная (конкурирующая) гипотеза
- Стандартное обозначение H_0 и H_1 .
- Если не уточняется, о какой гипотезе идет речь, то подразумевается основная гипотеза.
- Чаще всего (но не всегда) одна гипотеза утверждает, что предположение верно, другая — что нет.

Например,

$$H_0 : EX = EY$$

$$H_1 : EX \neq EY$$

$$H_2 : EX \geq EY$$

$$H_3 : EX \leq EY$$

Например,

Y - продажи до рекламной акции,

X - продажи после рекламные акции

Не рассматриваем рекламную кампанию,
после которой продажи упали

Пепси-Кола

Хрущев, Майкл Джексон, Бритни Спирс,
Бэкхем

Разное отношение к основной и альтернативной гипотезам

- Неточно говорить «...основная гипотеза принята...» или «...выбрана основная гипотеза...».
- Правильно говорить
 - «основная гипотеза отвергнута...» и
 - «основная гипотеза не отвергнута...».
- Так как обычно проверяют лишь достаточное условие.

Комментарий 1:

- Гипотеза: число делится на 6 нацело.
 - Умеем проверять только делимость числа на 2 нацело.
-

Комментарий 2

Часто гипотеза очень сложная, включает в себя несколько переменных.

- Например,
- компания изобретатель производит лекарство (бренд)
- компания последователь выпускает "дженерик" (generic) - копию бренда
- Компания проверяет, что эффект у дженерика такой же, как у бренда.

Недостаточно не отвергнуть гипотезу

- Основная гипотеза: у лекарств одинаковый эффект.
- Дело касается здоровья людей, и не отвергнуть гипотезу недостаточно.
- Надо дополнительно проверить отсутствие побочных эффектов у тех, кто страдает заболеванием «x1», «x2», и так далее...

Даже дозировка важна

В феврале 1917 года истек срок американского патента компании Bayer на аспирин

Грипп "испанка" 1918 года стал самой смертельной пандемией в истории человечества (в абсолютных цифрах)

Симптомы передозировки аспирина и симптомы гриппа очень схожи. Вопрос: как много людей умерло в 1918 от гриппа, и как много людей умерло в 1918 году от больших доз аспирина (до 30 граммов тогда против 1-4 грамма сейчас)?

Заключение

- Неточно говорить, что основная гипотеза принята (*проблемы на собеседовании и при общении с коллегами*).
 - Точнее говорить, что (основная) гипотеза не отвергнута
 - Хотя действуем так, как будто принята...
-

Ошибки первого и второго рода

- Ошибка первого рода совершается, если отвергается основная гипотеза, когда на самом деле она верна.
- Ошибка второго рода совершается, если отвергается конкурирующая гипотеза, когда она верна.

Аналогия

- В больнице врач принимает решение, направлять пациента на операцию, или нет.
-

Гипотеза: нужна срочная операция

	Гипотеза верна	Гипотеза не верна
Гипотеза принята	+	Ошибка 2 рода
Гипотеза отвергнута	Ошибка 1 рода	+

Неприятные ответы

- Может ли врач свести частоту (вероятность) ошибок первого рода к нулю?
- Может ли врач свести частоту (вероятность) ошибок второго рода к нулю?

Есть исключения

- Вакцинация
- Операцию (маленькую и "превентивную") делают всем
- Тем самым исключают ошибки первого рода.

в медицине и анализе данных невозможно
избежать как ошибок первого рода,
так и ошибок второго рода

Ошибка первого рода (обычно) опаснее

Уровень значимости α

- Уровень значимости ограничивает сверху долю ошибок первого рода.
- Начиная с Фишера в качестве уровня значимости чаще всего используют 0.005, 0.01, 0.05.
- Ожидается, что одна проверка из 200, 100, 20 будет давать неверный результат (на большой выборке).

Для новичков!

- Почему $\alpha = 0.005, 0.01, 0.05$?
- В эконометрике чаще всего уровень значимости равен 0.05
- В генетике — 0.01
- Выбор уровня значимости – большая проблема! Зависит, например, от числа наблюдений!

Смотрите литературу

Significance testing has been the favored statistical tool in some experimental social sciences (over 90% of articles in the Journal of Applied Psychology during the early 1990s).

[https://en.wikipedia.org/wiki/
Statistical_hypothesis_testing](https://en.wikipedia.org/wiki/Statistical_hypothesis_testing)

-
- «медицинский» пример
 - На что влияет выбор уровня значимости?
 - Проектирование атомной электростанции
 - Трелевочный трактор
 - Генетики изменили уровень значимости с 0.05 на 0.01
-

Задача.

- Вместо врача рассмотрим банковского служащего, принимающего решение, выдавать заем или нет.
-
- Как будут интерпретироваться статистические понятия в этом случае?

Состоятельный критерий

Когда ошибка второго рода мала?

- Как добиться того, чтобы вероятность ошибки второго рода была малой?
- Ошибку очень сложно вычислять.
- Применяйте состоятельные критерии.
- Ошибку можно уменьшить, если увеличить число анализируемых наблюдений.
- Необходимы большие выборки.

Ошибка второго рода и мощность критерия

Мощность = $1 - P\{\text{ошибка 2 рода}\}$

- Вместо минимизации вероятности ошибки 2 рода максимизируем мощность
- Причины технические, так проще формулы

Дополнительно

- Проверить гипотезу по малой выборке возможно. (граница между большой и маленькой выборкой лежит где-то между 30 и 60 наблюдений).
- Но
- Платой за малый размер будет неприемлемо большая вероятность ошибки второго рода.
- Большинство практиков игнорируют ошибку второго рода.
- Это неверно.
- Профессиональные статистики при работе с малыми выборками увеличивают уровень значимости (например до 0.15 или 0.2), чтобы сделать вероятности ошибок сопоставимыми.

Статистика критерия или тестовая статистикой

- Статистика критерия или тестовая статистика, это функция от наблюдений, которая измеряет, насколько данные соответствуют гипотезе.
- Изредка она важна в приложениях, например, коэффициент корреляции. В таких конкретных случаях мы будем ее обсуждать.
- Статистика критерия необходима для обоснования статистического критерия.

Интерпретация статистики критерия

- "Маленькие" значения статистики критерия указывают, что данные «ведут себя» в соответствии с гипотезой. В этом случае гипотеза не отвергается.
- "Большие" значения статистики критерия указывают, что данные не соответствуют гипотезе, противоречат ей. В этом случае гипотеза отвергается.

P - значение

$T_{\text{экспериментальное}}$

Предположим, что основная гипотеза верна.

P — значение = вероятность, что статистика критерия превысит $T_{\text{экспериментальное}}$

Синонимы: p-value, значимость (Significance).

Пример

В экспериментах по обнаружению бозона Хиггса требовалось преодолеть порог «пять сигм» (что означает наличие р-значения 0,0000003).

<https://habr.com/ru/company/skillfactory/blog/510688/>

Анализ мощности. Перспективно, но сложно.

Начать можно со статьи

A Gentle Introduction to Statistical Power and
Power Analysis in Python

<https://machinelearningmastery.com/statistical-power-and-power-analysis-in-python/>

Алгоритм проверки статистических гипотез -1

1. Имеются n наблюдений , то есть n чисел, столбец в таблице данных.
2. Заранее задается уровень значимости α . Обычно это одно из чисел 0.005, 0.01, 0.05.

Алгоритм проверки статистических гипотез -2

3. Проверяются условия, при которых критерий можно применять.

Условия – Из учебника или справочника.

Несколько важных критериев будет рассмотрено далее

Алгоритм проверки статистических гипотез -3

4. Находится р-значение

5.

Если $p < \alpha$ - основную гипотезу отвергаем,
если $p > \alpha$ - основную гипотезу не отвергаем.

При этом

α – уровень значимости

p - p-value.

Определение статистического критерия.
Наконец-то

Алгоритм, описанный выше, и все его
математическое обоснование

называется

статистическим критерием или
статистическим тестом

Комментарии

- Наблюдения не обязательно являются числами.
 - Выбор того статистического критерия, который решает прикладную задачу – важная и сложная задача
-

Проверка условий применимости

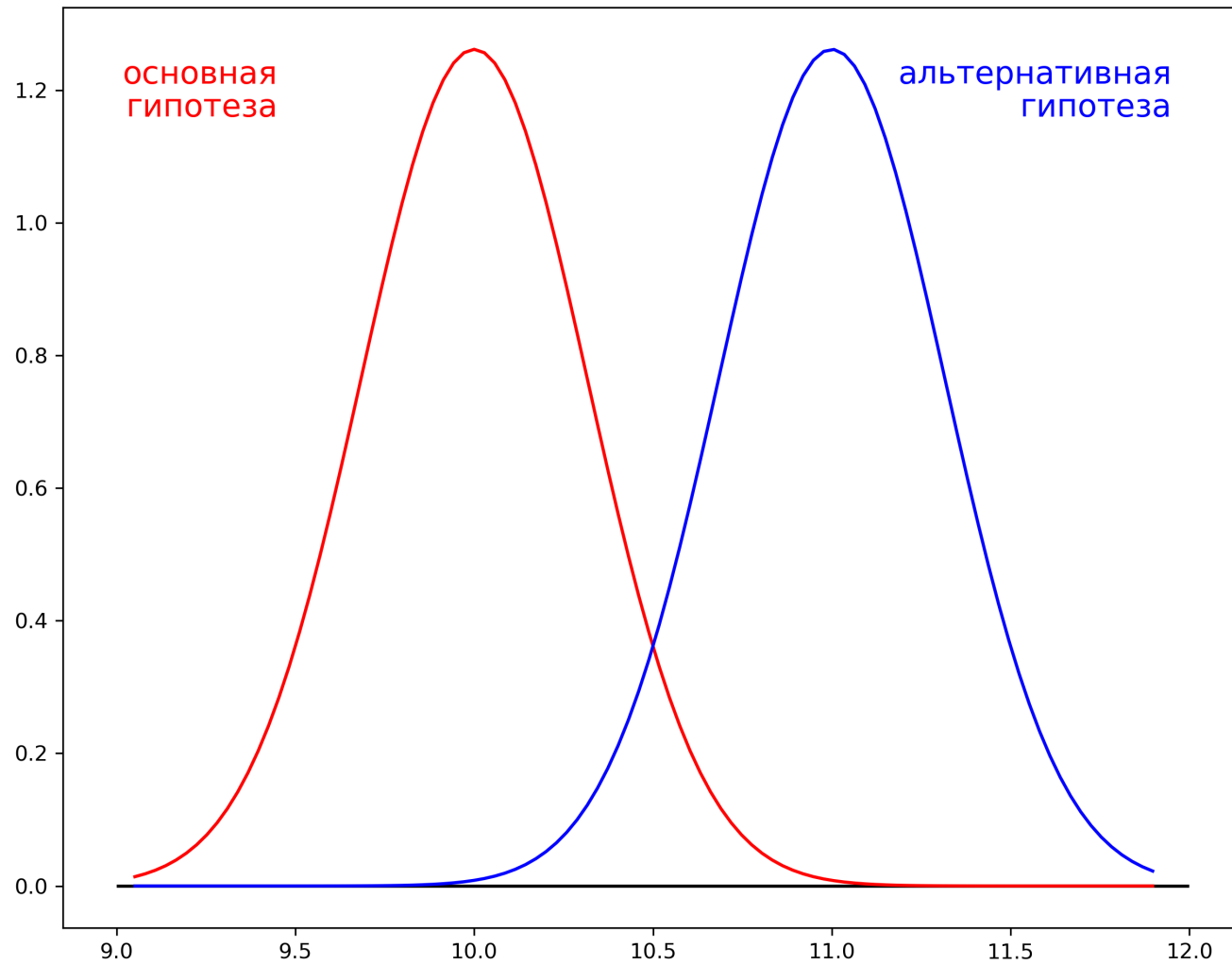
- Например, для применения t – критерия Стьюдента или для проверка гипотезы независимости с помощью критерия Пирсона надо проверить близость распределений переменных к нормальному.

Пример

- Нормальное распределение с дисперсией 1
- Имеется n наблюдений
- $H_0 : EX = 10$
- $H_1 : EX = 11$

Напоминание из теории вероятностей

- Среднее арифметическое n независимых одинаково распределенных случайных величин с общим нормальным распределением $N(a, b)$ имеет
- нормальное распределение $N(a, b/n)$



Вопросы:

- Где на графике ошибка первого рода,
- Где ошибка второго рода?
- Где p -значение?
- Ошибки первого и второго рода связаны прямой или обратной зависимостью?
-

Интерпретация статистики критерия

- В статистике жестко прописано, что именно задавать в качестве основной гипотезы.
- Примеры.