

Мат. стат. ликзбез, часть 2

Интеллектуальный анализ данных, 20!7

Спасибо К. В. Воронцову, МФТИ, Data Factory Яндекса, O.D.S. и кофеину.

Малютин Евгений Алексеевич



Рис.: Закреп в паблосе "Всё про A/B тесты"

Формальные определения

- выборка: $X^n = (X_1, \dots, X_n), X \sim P$
- нулевая гипотеза: $H_0 : P \in \omega$
- альтернатива: $H_1 : P \notin \omega$
- статистика: $T(X^n)$,
 $T(X^n) \sim F(x)$ при H_0
 $T(X^n) \approx F(x)$ при H_1
- проверка:
 - $T(X) = t$
 - $P(T(X) \geq t) = \text{p-value}$
 - $\text{p-value} \leq \alpha - H_0$ отвергается

В прошлых сериях...

Ошибки I и II рода

	H_0 верна	H_0 неверна
H_0 принимается	H_0 верно принята	Ошибка II рода
H_0 отвергается	Ошибка I рода	H_0 верно отвергнута

Рис.: Ошибки

- $P(H_0 \text{ отвергнута} | H_0 \text{ верна}) = P(p \leq \alpha | H_0) \leq \alpha$
- $row = P(\text{отвергаем } H_0 | H_1) = 1 - P(\text{принимаем } H_0 | H_1)$

Статистическая и практическая значимость

- Интерес представляет не p , а размер эффекта – степень отклонения данных от нулевой гипотезы

Статистическая и практическая значимость

- Интерес представляет не p , а размер эффекта – степень отклонения данных от нулевой гипотезы
- Примеры:
 - Вероятность верного предсказания
 - Вероятность выздоровления пациента минус вероятность плацебо
 - Увеличение среднего чека интернет-магазина на программе лояльности... минус издержки

Размер по выборке:

- Оценка размера по выборке – случайная величина

Размер по выборке:

- Оценка размера по выборке – случайная величина
- p_value показывает, с какой вероятностью эта оценка получилась случайно

Размер по выборке:

- Оценка размера по выборке – случайная величина
- p_value показывает, с какой вероятностью эта оценка получилась случайно
- p_value зависит не только от эффекта, но и от размера выборки

Размер по выборке:

- Оценка размера по выборке – случайная величина
- p_value показывает, с какой вероятностью эта оценка получилась случайно
- p_value зависит не только от эффекта, но и от размера выборки
- по мере увеличения n H_0 может сначала приниматься, а потом начать отвергаться

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.
 - Занимались меньше 20 минут vs занимались в районе час за 3 года. Спортивные набрали, в среднем, меньше ($p_value < 0.001$)

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.
 - Занимались меньше 20 минут vs занимались в районе час за 3 года. Спортивные набрали, в среднем, меньше ($p_value < 0.001$)
 - меньше на **150 грамм**

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.
 - Занимались меньше 20 минут vs занимались в районе час за 3 года.
Спортивные набрали, в среднем, меньше ($p_value < 0.001$)
 - меньше на **150 грамм**
- Пример 2: Гормональный препарат "Премарин облегчающий симптомы менопаузы

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.
 - Занимались меньше 20 минут vs занимались в районе час за 3 года.
Спортивные набрали, в среднем, меньше ($p_value < 0.001$)
 - меньше на **150 грамм**
- Пример 2: Гормональный препарат "Премарин облегчающий симптомы менопаузы"
 - Испытание было досрочно прервано, т.к. прием вел к побочным эффектам в 0.07% случаев...

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.
 - Занимались меньше 20 минут vs занимались в районе час за 3 года. Спортивные набрали, в среднем, меньше ($p_value < 0.001$)
 - меньше на **150 грамм**
- Пример 2: Гормональный препарат "Премарин облегчающий симптомы менопаузы"
 - Испытание было досрочно прервано, т.к. прием вел к побочным эффектам в 0.07% случаев...
 - увеличению риска рака(0.07%), инфаркта (0.07%) и инсульта (0.08%), потенциально – тысячи смертей в одних только США

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.
 - Занимались меньше 20 минут vs занимались в районе час за 3 года. Спортивные набрали, в среднем, меньше ($p\text{-value} < 0.001$)
 - меньше на **150 грамм**
- Пример 2: Гормональный препарат "Премарин облегчающий симптомы менопаузы"
 - Испытание было досрочно прервано, т.к. прием вел к побочным эффектам в 0.07% случаев...
 - увеличению риска рака(0.07%), инфаркта (0.07%) и инсульта (0.08%), потенциально – тысячи смертей в одних только США
- Пример 3: Замедление симптомов Альцгеймера

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.
 - Занимались меньше 20 минут vs занимались в районе час за 3 года. Спортивные набрали, в среднем, меньше ($p_value < 0.001$)
 - меньше на **150 грамм**
- Пример 2: Гормональный препарат "Премарин облегчающий симптомы менопаузы"
 - Испытание было досрочно прервано, т.к. прием вел к побочным эффектам в 0.07% случаев...
 - увеличению риска рака(0.07%), инфаркта (0.07%) и инсульта (0.08%), потенциально – тысячи смертей в одних только США
- Пример 3: Замедление симптомов Альцгеймера
 - Разница в 13 пунктов IQ, но большое p_value

Клёвые примеры:

- Пример 1: физические упражнения и вес у женщин.
 - Занимались меньше 20 минут vs занимались в районе час за 3 года. Спортивные набрали, в среднем, меньше ($p_value < 0.001$)
 - меньше на **150 грамм**
- Пример 2: Гормональный препарат "Премарин облегчающий симптомы менопаузы"
 - Испытание было досрочно прервано, т.к. прием вел к побочным эффектам в 0.07% случаев...
 - увеличению риска рака(0.07%), инфаркта (0.07%) и инсульта (0.08%), потенциально – тысячи смертей в одних только США
- Пример 3: Замедление симптомов Альцгеймера
 - Разница в 13 пунктов IQ, но большое p_value
 - возможно, исследование стоит продолжить для большего n



Рис.: Шо то – дверь, шо это – дверь...

Хотим проверить идею

- здравый смысл (объяснить маме)

Хотим проверить идею

- здравый смысл (объяснить маме)
- провести опрос

Хотим проверить идею

- здравый смысл (объяснить маме)
- провести опрос
- фокус-группа

Хотим проверить идею

- здравый смысл (объяснить маме)
- провести опрос
- фокус-группа
- A/B тест

Метрики

- бизнес-метрики

Метрики

- бизнес-метрики
- (деньги, DAU, d-30 retention)

Метрики

- бизнес-метрики
- (деньги, DAU, d-30 retention)
- промежуточные метрики

Метрики

- бизнес-метрики
- (деньги, DAU, d-30 retention)
- промежуточные метрики
- (ctr, ctp и пр)

Метрики

- бизнес-метрики
- (деньги, DAU, d-30 retention)
- промежуточные метрики
- (ctr, ctp и пр)
- оффлайн-тестирование

Метрики

- бизнес-метрики
- (деньги, DAU, d-30 retention)
- промежуточные метрики
- (ctr, ctp и пр)
- оффлайн-тестирование
- (исторические данные: "средняя позиция клика")

Важные моменты

- Стратификация и рандомизация

Важные моменты

- Стратификация и рандомизация
- Связанные выборки

Важные моменты

- Стратификация и рандомизация
- Связанные выборки
- Проведение нескольких экспериментов сразу

Важные моменты

- Стратификация и рандомизация
- Связанные выборки
- Проведение нескольких экспериментов сразу
- Устойчивость?

Важные моменты

- Стратификация и рандомизация
- Связанные выборки
- Проведение нескольких экспериментов сразу
- Устойчивость?
 - swap

Важные моменты

- Стратификация и рандомизация
- Связанные выборки
- Проведение нескольких экспериментов сразу
- Устойчивость?
 - swap
 - обратный эксперимент

Важные моменты

- Стратификация и рандомизация
- Связанные выборки
- Проведение нескольких экспериментов сразу
- Устойчивость?
 - swap
 - обратный эксперимент
 - A/A-тестирование

Важные моменты

- Стратификация и рандомизация
- Связанные выборки
- Проведение нескольких экспериментов сразу
- Устойчивость?
 - swap
 - обратный эксперимент
 - A/A-тестирование
 - размер выборки