AB тесты: статистика и базовые примеры.

Камалов Ирек.

22 июля 2022 г.

План презентации.

- Введение.
- Нормальные выборки.
- Биномиальные распределения.
- Пуассоновские распределения.
- Мультиномиальные распределения.
- Тест Манна-Уитни.
- Байесовское AB тестирование (для бин. распр.)
- Заключение.

Введение.

- AB тесты - стат. тесты для сравнения параметров 2 распред.
- $H_0 : \theta_1 = \theta_2$
  $H_1 : \theta_1 \neq \theta_2 (<>)$
- Одностор. альтернатива - $\theta_1 > \theta_2 (<)$
- Двухстор. альтернатива - $\theta_1 \neq \theta_2$
- Данные обычно удовл. стат. соглашениям - выборки из нез. о.р. сл. в.

Введение.

- Используется для оценки значимости эксп. данных
- Пример - сравниваем конверсию страниц. Мы всегда знаем, какая страница оказалась лучше, но насколько результату можно доверять
- A/B/n тестирование - аналог для n групп

Нормальные выборки.

- Базовый случай: $\xi \sim N(\mu, \sigma^2)$, сравниваем мат. ожидания
- $X_1, X_2, \ldots X_n(\text{нез.}) \sim N(\mu_1, \sigma_1^2)$
  $Y_1, Y_2, \ldots Y_m(\text{нез.}) \sim N(\mu_2, \sigma_2^2)$
- $H_0 : \mu_1 = \mu_2$
  $H_1 : \mu_1 \neq \mu_2 (<>)$
- Примеры:
  1. Сравниваем цены квартир в Москве и в Петербурге
  2. Правда ли, что японские авто служат дольше корейских
  3. Сравниваем зар. плату по разным отраслям

Нормальные выборки.

Т-тесты.

- Используем t-критерии (Стьюдента, Уэлча)
- На этом примере рассм. 2 вида альт. гипотез: одностор. и двухстор.

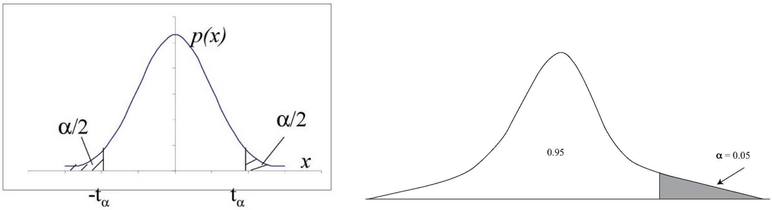


Рис.: Одностор. и двухстор. критические области уровня α.

Бином. распределения.

- ξ - количество успехов в n независимых исп. Бернулли, вер. усп. p
- $X \sim Bin(n_1, p_1), Y \sim Bin(n_2, p_2)$
- $H_0 : p_1 = p_2$
  $H_1 : p_1 \neq p_2 (<>)$
- Примеры:
  1. Конверсия 2 страниц (кол-во кликов посетителей по кнопке)
  2. Эффективность лекарства (делим людей на 2 гр., одна получает вакцину, другая плацебо)
  3. Социология - процент опр. людей в группах

Биномиальные распределения.

Z-test.

- $\xi \sim Bin(n, p) => \xi \xrightarrow{d} N(np, np(1-p))$ (ЦПТ)
- То есть

$$X \xrightarrow{d} N(n_1 p_1, n_1 p_1 q_1), Y \xrightarrow{d} N(n_2 p_2, n_2 p_2 q_2)$$

- При верной H_0 $distr(X/n_1 - Y/n_2) \approx N(0, \sigma^2)$
- Считаем, что n_1, n_2 велики, для дов. интервала остаётся оценить дисперсию

Бином. распределения.

- 2 варианта оценки
  1. pooled prop. $\hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}$
     $\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$
  2. unpooled prop. $\hat{\sigma} = \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$
- Получаем $[-z_{\alpha/2}\hat{\sigma}, +z_{\alpha/2}\hat{\sigma}]$, либо полуинтервал для одностор. альт.
- Считаем реальное $X/n_1 - Y/n_2$ и сравниваем с границами

Бином. распределения.

Точный тест Фишера.

- Используется при небольших выборках
- Составляем факторную таблицу. Для простоты рассм. на примере
- Есть 2 группы - мужчины и женщины, изучаем их склонность к диете
- Данные отражены в табл.

	Мужчины	Женщины	Всего
На диете	1	9	10
Не на диете	11	3	14
Всего	12	12	24

Бином. распределения.

Точный тест Фишера.

	Мужчины	Женщины	Всего
На диете	a=1	b=9	D=10
Не на диете	c=11	d=3	14
Всего	n=12	12	N=24

- При верной H_0 мы можем посчитать вероятность наблюдать такую таблицу
- С помощью комб. получаем
  $p = \binom{a+b}{a}\binom{c+d}{c} \Big/ \binom{N}{a+c} = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{N!\,a!\,b!\,c!\,d!}$
  гипергеометрическое распределение на a с парам. N, D, n

Бином. распределения.

Точный тест Фишера.

	Мужчины	Женщины	Всего
На диете	a=1	b=9	D=10
Не на диете	c=11	d=3	14
Всего	n=12	12	N=24

- Работаем с фиксир. N, D, n
- Стат. значимость набл-ых данных - их вероятность + вероятность более редких таблиц.
- $p - val = P(a = 1|N, D, n) + P(a = 0|N, D, n) = 0.00138$ - одностор альт. "женщины более склонны"
- $p - val = P(a = 1|N, D, n) + P(a = 0|N, D, n)$ $+P(a = 9|N, D, n) + P(a = 10|N, D, n) = 0.0028$ - двухстор. альт. "склонности просто различны"

Пуассоновские распределения.

- Количество успехов в нез. испытаниях Бернулли с очень большим числом n
- $P(\xi = k) = \frac{\lambda^k}{k!} e^{\lambda}$
- $X \sim Pois(\lambda_1), Y \sim Pois(\lambda_2)$
- $H_0 : \lambda_1 = \lambda_2$
  $H_1 : \lambda_1 \neq \lambda_2 (<>)$
- Примеры
  1. Правда ли, что Яндекс.Еда получает больше заказов в неделю чем Delivery Club
  2. Сравниваем кол-во пациентов в день в двух клиниках
  3. Сравниваем частоту транзакций у 2 типов пользователей
  4. Правда ли, что в апреле столько же дождей, что в сентябре

Пуассоновские распределения.

C-тест (conditional)

- Сводим пуасс. распр. к биномиальному: $(X|X+Y=k) \sim Bin(k,p), p = (\frac{n_1}{n_2})(\frac{\lambda_1}{\lambda_2})(1+(\frac{n_1}{n_2})(\frac{\lambda_1}{\lambda_2}))$
- Далее считаем ур. значимости
- $P(X \geqslant k_1|k,p) = \sum_{i=k_1}^{k} \binom{n}{i} p^k (1-p)^{k-i}$ - одностор. альтернатива, k_1 - наблюдаемое значение X
- $2 * min\{P(X \geqslant k_1|k,p), P(X \leqslant k_1|k,p)\}$ - двухстор. альт.

Мультином. распределения.

- Рассм. нез. испытания с k возм. исходами (соотв. k событиям). Их вер-сти $p_1, p_2, ...p_k$
- ξ_i - количество наступлений события A_i среди n исп.
- $P(\xi_1 = m_1, ..., \xi_k = m_k)$ - мультином. распр.
- Имеем 2 мультином. распр. $M(p_1, ...p_k), M(q_1, ...q_k)$
- $H_0 : p_1, ...p_k = q_1, ...q_k$
  $H_1 : p_1, ...p_k \neq q_1, ...q_2$ (только двухстор. альт.)

Мультином. распределения.

Примеры.

1. Курс студентов делят на 2 потока перед экзаменом. Имеем частоты оценок 2, 3, 4, 5 для двух потоков, хотим проверить их однородность
2. Правда ли, что мужчины и женщины совершают одинаковые объёмы покупок по выбранным категориям
3. Одинаковое ли распределение по специальностям на рынке труда в Москве и в Петербурге

Мультином. распределения.

Критерий однородности χ^2.

- $\mu_1, ...\mu_k$; $\nu_1, ...\nu_k$ - наблюдаемые частоты данных событий
- Статистика критерия:

$$\chi^2 = n_1 n_2 \sum_{i=1}^{k} \frac{(\mu_i/n_1 - \nu_i/n_2)^2}{\mu_i + \nu_i}, n_1 = \sum_{i=1}^{k} \mu_i, n_2 = \sum_{i=1}^{k} \nu_i$$

- $\mu_i/n_1 \xrightarrow{d} N(p_i, p_i q_i)$ (ЦПТ), ν_i/n_2 аналог.
- При верной H_0 $\chi^2 \sim \chi^2(k-1)$
- Для α находим кр. конст. через таблицу $\chi^2(k-1)$
- Если наблюдаем χ^2 больше константы, отвергаем H_0

Тест Манна-Уитни.

- Если о распредлении ничего не изветсно, можем исп. ранги в упор. выборке
- $X_1, X_2, ...X_{n_1}(\text{нез.}); Y_1, Y_2, \ldots Y_{n_2}(\text{нез.})$
- $H_0 : F_X(u) = F_Y(u)\ \forall u \in \mathbb{R}$
  $H_1 : F_X(u) \neq F_Y(u)(<>\ \forall u \in \mathbb{R})$
- Основное применение - медицина, психология. Например, сравнить баллы по психол. тесту среди групп.

Тест Манна-Уитни.

- Объединяем $\{X_i\}, \{Y_j\}$ в упор. ряд
- R_1, R_2 - суммы рангов $\{X_i\}, \{Y_j\}$ в данном ряду
- $U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1+1)}{2} - R_1$,
  $U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2+1)}{2} - R_2$,
  если всё вычислено верно, то $U_1 + U_2 = n_1 \cdot n_2$.
- $U = \min\{U_1, U_2\}$
- Для α, n_1, n_2 наход. $crit$, если $U \notin crit$, отвергаем H_0
- $U \xrightarrow{d} N(n_1 n_2/2, n_1 n_2(n_1 + n_2 + 1)/12)$,
  обычно приближ. хорошее уже при $n_1, n_2 \geqslant 20$

Байесовское AB тестирование (для бин. распр.).

- Вновь рассм. 2 бином. сл. в. $X \sim Bin(n_1, p_1), Y \sim Bin(n_2, p_2)$, сравниваем p_1, p_2
- Рассмотрим отдельно некот. $\xi \sim Bin(n, \theta_0)$
- Введём сл. в. θ с распр. $p(\theta)$, которая моделирует параметр θ_0
- Также имеем $p(\theta|\xi = k)$ - апост. распр. данной величины, $P(\xi = k|\theta)$ - ф. правдоподобия
- По ф. Байеса $p(\theta|\xi = k) = \frac{P(\xi=k|\theta)p(\theta)}{\int\limits_{-\infty}^{+\infty} P(\xi=k|\theta)p(\theta)d\theta}$
- Как оценить интеграл?

Байесовское AB тестирование (для бин. распр.).

- По ф. Байеса $p(\theta|\xi = k) = \frac{P(\xi=k|\theta)p(\theta)}{\int\limits_{-\infty}^{+\infty} P(\xi=k|\theta)p(\theta)d\theta}$
- Если взять в качестве $p(\theta)$ сопряж. распр. к $P(\xi = k|\theta)$, то интеграл можно найти аналитически
- Сопряж. распр. $p(\theta)$ инвариантно отн. умнож. на $P(\xi = k|\theta)$ (меняются лишь параметры)
- Для бином. распр. сопряж. является бета-распр. $Beta(a, b), p(x) = \frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}$
- Различные парам. a, b порождают разные априорн. распр., в которых мы будем оценивать параметр

Байесовское AB тестирование (для бин. распр.).

- По ф. Байеса $p(\theta|\xi = k) = \frac{P(\xi=k|\theta)p(\theta)}{\int\limits_{-\infty}^{+\infty} P(\xi=k|\theta)p(\theta)d\theta}$
- При $a = b = 1$, $Beta(a, b) = R[0, 1]$
- Получим $(\theta|\xi = k) \sim Beta(k + 1, n - k + 1)$
- Далее можем смоделировать искомые p_1, p_2

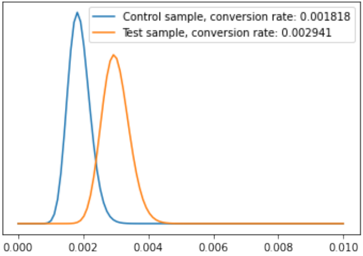
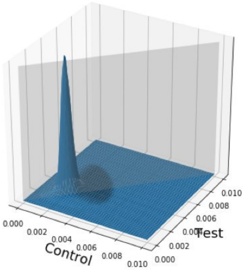
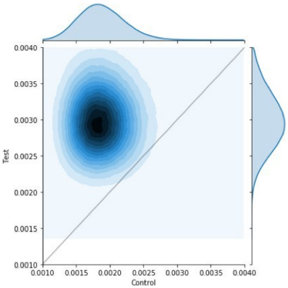


Рис.: Апостериорные распр-ия конверсии.

Байесовское AB тестирование (для бин. распр.).

- Требуется использовать этот рез-тат, чтобы оценить значимость набл. данных
- Для этого строим совместную плотность апост. конверсий





- Вероятность правильного решения - интеграл плотности по полуплоскости OX $y \geqslant x$

Заключение.

- Были рассм. основные мат. модели AB тестов
- К ним были даны базовые примеры из прикл. областей
- Отд. внимание было уделено байесовскому подходу
- Был составлен .ipynb с реализацией моделей
- Не была рассмотрена связь AB тестов с uplift modeling, будет далее