

## Roc-кривая и А5

$$H_0: \mu_T = \mu_C$$

$$H_A: \mu_T > \mu_C \text{ MDE}$$

$$\lambda = P(H_A | H_0)$$

$$\beta = P(H_0 | H_A)$$

$$n = f(\lambda, \beta, \text{MDE}, \sigma_T^2, \sigma_C^2)$$

$$H_0 = \{y = 0\}$$

$$H_A = \{y = 1\}$$

→ критика корректирующейся  $\alpha$  (...)

→ хотим маленькой  $\alpha$  (...)

$$Z = \frac{\bar{x} - \bar{y}}{se(\bar{x} - \bar{y})}$$

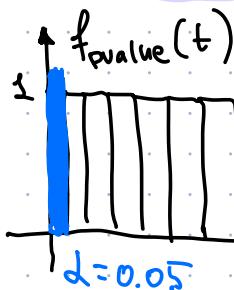
$H_0$  верна

$H_A$  верна

$Z_{obs}$

VS

$Z_{crit}$



pvalue

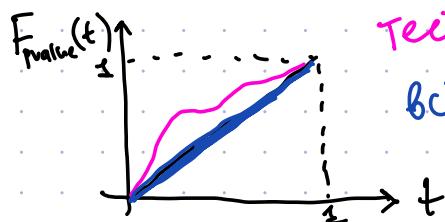
VS

$\lambda$

P-е pvalue

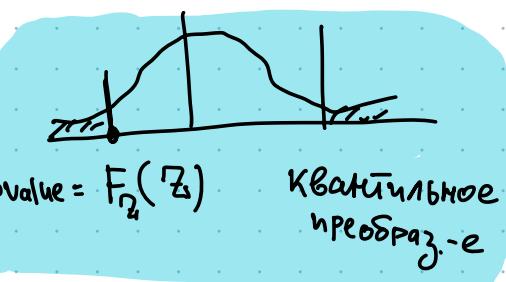
P-е pvalue

A-A test



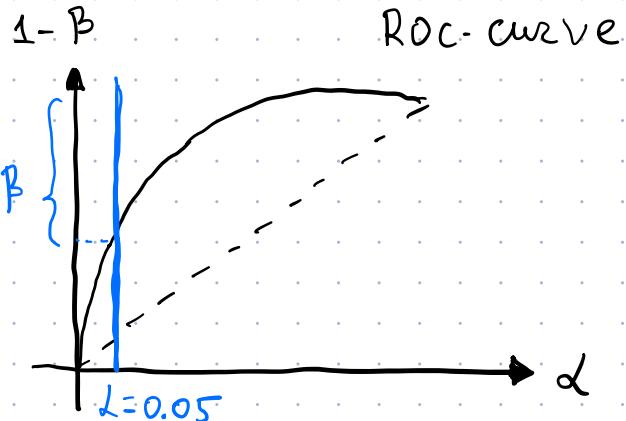
Тест разбивается

всё ок



$$pvalue = F_D(Z_t)$$

Квантильное преобраз.-е



$$h =$$

	0	1
0	TN	FN
1	FP	TP

$$FPR \quad TPR = \text{Recall}$$

$$\lambda$$

$$1 - \beta$$

## Ratio - метрика

$$CTR = \frac{\sum_{u \in \text{Users}} \text{click}_u}{\sum_{u \in \text{Users}} \text{views}_u} = \frac{\sum_{u \in \text{Users}} \text{views}_u \cdot \text{userCTR}_u}{\sum_{u \in \text{Users}} \text{views}_u}$$

click и view коррелируют  $\uparrow \text{views} \Rightarrow \text{click} \uparrow$

средняя оценка  
оценки разела

$$S_i = \frac{\sum_{u \in \text{Users}} \sum_{j=1}^{N_u} Y_{ij}}{\sum_{u \in \text{Users}} N_u}$$

$$\bar{Y} = \frac{\sum_i S_i}{\sum_i N_i} = \frac{\sum_i S_i / n}{\sum_i N_i / n} = \frac{\bar{S}}{\bar{N}}$$

Наго наутибс  
работат с  
откомпенси  
среднис

балансировка АБ

	out	not out
treat		
control		

$\hat{P}_T$  - freq. out of treatment

$\hat{P}_C$  - freq. out of control

$$RR = \frac{P_T}{P_C} \quad \text{Risk Ratio}$$

$$\text{eff} = 1 - RR \quad \exists \text{ эфф. балансир}$$

CI 95%

RR

$$\bar{Y} = \frac{\bar{S}}{\bar{N}} \underset{\text{asy}}{\sim} \frac{N(\dots)}{N(\dots)}$$

правиль касар-TO

....

## Дельта-метод

$$\hat{\theta} \xrightarrow{\text{asy}} N(\theta, \hat{\sigma}^2)$$

$g(t)$  - однород. кнр.

$$g(\hat{\theta}) \xrightarrow{\text{asy}} N(g(\theta), \hat{\sigma}^2 \cdot g'(\hat{\theta})^2)$$

$[k \times 1]$

$$\hat{\theta} \xrightarrow{\text{asy}} N(\theta, \hat{\Sigma})$$

$g(t_1, t_2, \dots, t_k)$

$$g(\hat{\theta}) \xrightarrow{\text{asy}} N(g(\theta), \nabla \hat{g}^T \hat{\Sigma} \nabla \hat{g})$$

→ маленькая энч.

→ мы в окр. ист. стн.

Одномерный  
дельта-метод

многомерный  
дельта-метод

## Допоза 1: ввод

$$\hat{RR} = \frac{\hat{P}_T}{\hat{P}_C} - \text{оценка кнр. из заб. от ЗК.}$$

$$h_T = h_C = h$$

$$\hat{Var}(\hat{p}) = \hat{\Sigma} = \begin{pmatrix} \frac{\hat{P}_T(1-\hat{P}_T)}{h_T} & 0 \\ 0 & \frac{\hat{P}_C(1-\hat{P}_C)}{h_C} \end{pmatrix}$$

$$\hat{p} = \begin{pmatrix} \hat{P}_T \\ \hat{P}_C \end{pmatrix}$$

$$g(t_1, t_2) = \frac{t_1}{t_2} \quad \nabla g = \begin{pmatrix} \frac{1}{t_2} \\ -\frac{t_1}{t_2^2} \end{pmatrix} \quad \nabla \hat{g} = \begin{pmatrix} \frac{1}{\hat{P}_C} \\ -\frac{\hat{P}_T}{\hat{P}_C^2} \end{pmatrix}$$

$$\hat{Var}(\hat{RR}) = \hat{g}^T \cdot \hat{\Sigma} \cdot \nabla \hat{g} = \left( \frac{\hat{P}_T(1-\hat{P}_T)}{\hat{P}_C \cdot h} \right) - \left( \frac{(1-\hat{P}_C) \cdot \hat{P}_T}{\hat{P}_C \cdot h} \right) \cdot \begin{pmatrix} \frac{1}{\hat{P}_C} \\ -\frac{\hat{P}_T}{\hat{P}_C^2} \end{pmatrix} =$$

$$= \frac{\hat{P}_T(1-\hat{P}_T)}{\hat{P}_c^2 \cdot n} + \frac{\hat{P}_T^2(1-\hat{P}_c)}{n \cdot \hat{P}_c^3} = \frac{\hat{P}_T}{n \hat{P}_c^2} \left( (1-\hat{P}_T) + \frac{(1-\hat{P}_c)}{\hat{P}_c} \right)$$

Доказа 2: проверять

$$\theta = \ln RR \quad \hat{\theta} = \ln \hat{RR} \xrightarrow{a.s.} N \dots$$

$$\Rightarrow CI \text{ для } \theta$$

$\Rightarrow$  ожидаемый, вероятно, эквивалентный

$$\hat{\theta} = \ln \hat{P}_T - \ln \hat{P}_c$$

$$\ln \hat{P} \approx \ln p + \frac{1}{p} (\hat{p} - p)$$

$$\text{Var}(\ln \hat{p}) \approx \frac{1}{p^2} \cdot \text{Var}(\hat{p}) = \frac{p(1-p)}{p^2 \cdot n} = \frac{1-p}{p} \cdot \frac{1}{n}$$

$$\text{Var}(\ln \hat{p}) = \frac{1-\hat{p}}{\hat{p}} \cdot \frac{1}{n}$$

$$P(RR \in [\exp(\hat{\theta}_L); \exp(\hat{\theta}_R)])$$

$$\exp\left(\hat{\theta} \pm 1.96 \cdot \sqrt{\frac{1-\hat{P}_T}{\hat{P}_T} \cdot \frac{1}{n} + \frac{1-\hat{P}_c}{\hat{P}_c} \cdot \frac{1}{n}}\right)$$

! Давно ли хорошо проверить  
какой лиже !!

Резюме:

$$X_1, \dots, X_n \sim \text{iid}$$

$$Y_1, \dots, Y_n \sim \text{iid}$$

$$\frac{\mu_Y}{\sigma_X} = 2$$

$$\frac{Y}{X}$$

Δ-метод позволяет  
находить CI

## Линеаризация

$$R = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} = \frac{\bar{x}}{\bar{y}}$$

Если для класса пригодить преобразование

$$F(x, y) :$$

a) Ratio-метрика стала нотоизмеркой

b) направляемость и зависимость сохраняются

$$\begin{array}{ccc} X_1, \dots, X_n & \bar{X} & \uparrow \\ \ln X_1, \dots, \ln X_n & \bar{\ln X} & \downarrow \end{array}$$

пересеч

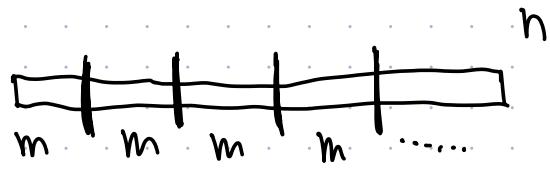
$$\text{click} = w \cdot \text{views}$$

$$X = w \cdot Y$$

$$\text{linearize Click} = \text{click} - \hat{w} \cdot \text{views}$$

Z<sub>i</sub>-test

## Бакетизация



$$\bar{c}_{tr} \sim N(c_{tr}, \dots)$$

$\sigma^2$  - Внутренний  
шум

$\sigma^2$  - внешний

## Стратифициация

$$\Sigma = \frac{\hat{\Delta}}{se(\hat{\Delta})}$$

$L$  страт

$$i_e = 1, \dots, N_e$$

$$i_e = 1, \dots, n_e$$

$$w_e = \frac{N_e}{N}$$

$$\sum_{e=1}^L N_e = N$$

$$\sum_{e=1}^L n_e = n$$

верт. соб.

бюджетка

$$\mu = \frac{1}{N} \sum_{e=1}^L \sum_{i=1}^{N_e} x_{ie} = \sum_{e=1}^L \frac{N_e}{N} \cdot \frac{1}{N_e} \sum_{i=1}^{N_e} x_{ie}$$

*We* *Me*

$$= \sum_{e=1}^L w_e \cdot m_e$$

$$\sigma^2 = \frac{1}{N} \sum_{e=1}^L \sum_{i=1}^{N_e} (x_{ie} - \mu)^2 = \sum_{e=1}^L \frac{N_e}{N} \cdot \frac{1}{N_e} \sum_{i=1}^{N_e} (x_{ie} - m_e + m_e - \mu)^2 =$$

*We* *Me*

$$= \sum_{e=1}^L w_e \cdot \underbrace{\frac{1}{n_e} \sum_{i=1}^{n_e} (x_{ie} - \mu_e)^2}_{\sigma_e^2} + \sum_{e=1}^L w_e \cdot \frac{1}{n_e} \sum_{i=1}^{n_e} (\mu_e - \mu)^2 = \\ n_e \cdot (\mu_e - \mu)^2$$

$$= \underbrace{\sum_{e=1}^L w_e \cdot \sigma_e^2}_{\text{внутригрупповая дисперсия}} + \underbrace{\sum_{e=1}^L w_e \cdot (\mu_e - \mu)^2}_{\text{межгрупповая дисперсия}}$$

Вывод:

$$\forall e \quad \forall e \quad x_{1e}, \dots, x_{ne}$$

$$n_1 + \dots + n_L = n$$

$w_e$  — это знато, что было  
дано классу найти и с  
так, чтобы дисперсия  $\downarrow$

$$\bar{x}_s = \sum_{e=1}^L w_e \cdot \bar{x}_e \quad \bar{x}_e = \frac{1}{n_e} \sum_{i=1}^{n_e} x_{ie} \quad \sum_{e=1}^L w_e = 1$$

$$\mathbb{E}(\bar{x}_s) = \sum_{e=1}^L w_e \cdot \frac{1}{n_e} \sum_{i=1}^{n_e} \mathbb{E}(x_{ie}) = \sum_{e=1}^L w_e \mu_e = \mu$$

$\downarrow \text{де}$

$n_e \cdot \mu_e$

$$\text{Var}(\bar{x}_s) = \sum_{e=1}^L w_e^2 \cdot \frac{1}{n_e^2} \cdot \sum_{i=1}^{n_e} \text{Var}(x_{ie}) = \sum_{e=1}^L \frac{w_e^2 \cdot \sigma_e^2}{n_e}$$

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} = \sum_{e=1}^L \frac{w_e}{n} \sigma_e^2 + \sum_{e=1}^L \frac{w_e}{n} \cdot (\mu_e - \mu)^2$$

11

$$\begin{cases} \text{Var}(\bar{x}_s) = \sum_{e=1}^L \frac{w_e^2 \sigma_e^2}{n_e} \rightarrow \min_{w_1, \dots, w_L} \\ n_1 + \dots + n_L = n \end{cases}$$

$$L = \text{Var}_{\bar{x}_s} - \lambda(n_1 + \dots + n_L - n) \Rightarrow n_e^* =$$

$$\boxed{\frac{w_e \sigma_e^2}{\sum_{k=1}^L w_k \sigma_k^2} \cdot n}$$

$$w_e = \frac{n_e}{n}$$

$w_e^*$

### Гипотезы

тех. соб.	$N$	стата	бсро	бгс.	нч.	цена набл.
ббочка	$n$	1	$N_1$	$n_1$	$P_1$	$c_1$
$X_1, \dots, X_n$		2	$N_2$	$n_2$	$P_2$	$c_2$
		$\Sigma$	$N$	$n$	$P$	

$$\text{Var}(\hat{P}) = \frac{P(1-P)}{n} \cdot \cancel{\frac{n-n}{n}}$$

$$P = \frac{N_1}{N} \cdot P_1 + \frac{N_2}{N} \cdot P_2$$

стата зависимость  
от груп от группы

$$\hat{P} = \frac{N_1}{N} \cdot \hat{P}_1 + \frac{N_2}{N} \cdot \hat{P}_2$$

$$TC = c_1 \cdot h_1 + c_2 \cdot h_2$$

Цель: минимизировать ббочки  $n_i$ :

минимизировать  $TC \rightarrow \min$  ~  $\text{Var}(\hat{P}) \rightarrow \min$

$$\begin{cases} \text{Var}(\hat{P}) = \frac{N_1^2}{N^2} \cdot \frac{P_1(1-P_1)}{h_1} + \frac{N_2^2}{N^2} \cdot \frac{P_2(1-P_2)}{h_2} \rightarrow \min_{h_1, h_2} \\ c_1 h_1 + c_2 h_2 = TC \end{cases}$$

фиксированные:  $N_1, N_2, N, \underbrace{\sigma_1^2, \sigma_2^2}_{\text{предварительно оценки}}, TC, c_1, c_2$   
 переменные:  $n_1, n_2$

$$L = \frac{N_1^2}{N^2} \cdot \frac{\sigma_1^2}{n_1} + \frac{N_2^2}{N^2} \cdot \frac{\sigma_2^2}{n_2} + \lambda (TC - c_1 n_1 - c_2 n_2)$$

$$\begin{cases} \frac{\partial L}{\partial n_1} = 0 \\ \frac{\partial L}{\partial n_2} = 0 \\ \frac{\partial L}{\partial \lambda} = 0 \end{cases} \quad \begin{cases} \frac{N_1^2}{N^2} \cdot \sigma_1^2 \cdot \left(-\frac{1}{n_1^2}\right) - \lambda c_1 = 0 \\ \frac{N_2^2}{N^2} \cdot \sigma_2^2 \cdot \left(-\frac{1}{n_2^2}\right) - \lambda c_2 = 0 \\ TC = c_1 n_1 + c_2 n_2 \end{cases}$$

$$\frac{N_1}{N_2} \cdot \frac{\sigma_1^2}{\sigma_2^2} \cdot \frac{n_2}{n_1} = \sqrt{\frac{c_1}{c_2}}$$

$$n_2 = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{N_2}{N_1} \cdot \sqrt{\frac{c_1}{c_2}} \cdot n_1$$

$$TC = c_1 \cdot n_1 + c_2 \cdot \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{N_2}{N_1} \cdot \sqrt{\frac{c_1}{c_2}} \cdot n_1$$

$$n_1^* = \frac{TC}{c_1 + \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{N_2}{N_1} \cdot \sqrt{c_2 c_1}} ; \quad n_2^* = \dots$$

♥ Соңғы жемісінде страту  
 ♥ страту < соңғы білескенің ғана.  
 (Combine strat.)

$$h_1^* = \frac{\frac{TC}{\sqrt{C_1}} \cdot g_1^j \cdot N_1}{\sqrt{C_1} \cdot g_1^j \cdot N_1 + \sqrt{C_2} \cdot g_2^j \cdot N_2}$$

$$h_j^* = \frac{\sum_{e=1}^L N_e^j \cdot g_e^j \cdot \frac{TC}{\sqrt{C_j}}}{\sum_{e=1}^L \sqrt{C_e} \cdot N_e \cdot g_e^j}$$

— штогаға  
рекомендация