Лекция 9.

Классические дов. интервалы

→ для отношения σ_2^2/σ_1^2 — пример исп-ия распр-ия Фишера.

→ для разницы мат-х ожид-ий $\boxed{\mu_2-\mu_1}$

Задача.

Есть две нез-х выборки

запись: $X_1,\dots,X_{n_X}$, $Y_1,\dots,Y_{n_Y}$ $\qquad \bar X$, $\bar Y$

запись

x_i	g_i
x_1	a
x_2	a
$\vdots$	b
$\vdots$	a
$\vdots$	$\vdots$
x_n	a

$\bar X_a$, $\bar X_b$

Предположения

набл. (X_i, g_i) незав.

жёсткие!

$$(x_i \mid g_i=a) \sim N(\mu_a;\sigma_a^2)$$

$$(x_i \mid g_i=b) \sim N(\mu_b;\sigma_b^2)$$

Цель: точный CI для $\dfrac{\sigma_a^2}{\sigma_b^2}$?

Упр. $\bar X_a = \dfrac{\sum_{g_i=a} x_i}{n_a}$ $\qquad n_a$ — кол-во набл. с $g_i=a$

несмещ. выбор. дисперсия $\hat\sigma_a^2 = \dfrac{\sum_{g_i=a}(x_i-\bar x_a)^2}{n_a-1}$

n_a, n_b фиксир.

$$\begin{cases} E(\bar x_a) = \mu_a \\ E(\hat\sigma_a^2) = \sigma_a^2 \end{cases}$$

$$\frac{\sum_{g_i=a}(x_i-\bar x_a)^2}{\sigma_a^2} \Bigg|_{n_a\text{-фикс.}} \sim \chi^2_{n_a-1}$$

$$\left.\frac{\sum_{j_i=b}(x_i-\bar{x}_b)^2}{\sigma_b^2}\right|_{n_b} \sim \chi^2_{n_b-1}$$

незав

Опр. $R \sim F_{d_1,d_2}$

$$R = \frac{\chi^2_{d_1}/d_1}{\chi^2_{d_2}/d_2}$$ незав

$$R = \frac{\sum_{j_i=a}(x_i-\bar{x}_a)^2 / \left(\sigma_a^2(n_a-1)\right)}{\sum_{j_i=b}(x_i-\bar{x}_b)^2 / \left(\sigma_b^2(n_b-1)\right)}$$

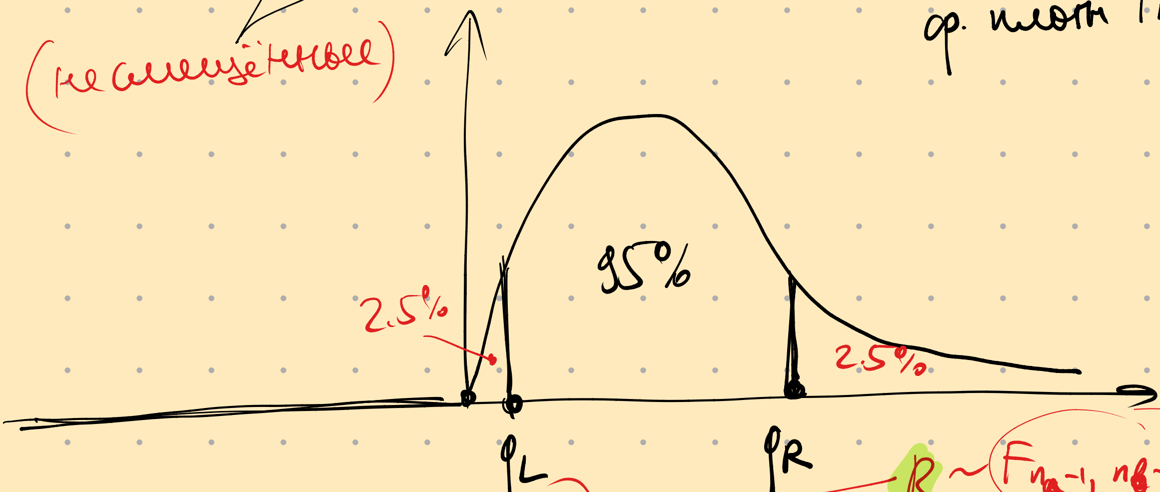
$$R = \frac{\left(\frac{\sum_a (x_i-\bar{x}_a)^2}{n_a-1}\right)}{\left(\frac{\sum_b (x_i-\bar{x}_b)^2}{n_b-1}\right)} \cdot \frac{\sigma_b^2}{\sigma_a^2} \Bigg| \sim F_{n_a-1,\, n_b-1}$$

n_a, n_b фиксир.

$$\frac{\hat{\sigma}_a^2}{\hat{\sigma}_b^2} \cdot \frac{\sigma_b^2}{\sigma_a^2} \sim F_{n_a-1,\, n_b-1}$$

(несмещённые)

ф. плотн $F_{n_a-1,\, n_b-1}$



$R \sim F_{n_a-1,\, n_b-1}$

$$P\left(q_L \le \frac{\hat{\sigma}_a^2}{\hat{\sigma}_b^2} \cdot \frac{\sigma_b^2}{\sigma_a^2} \le q_R\right) = 0.95$$

95% CI $\boxed{q_L \cdot \frac{\hat{\sigma}_b^2}{\hat{\sigma}_a^2} \le \frac{\sigma_b^2}{\sigma_a^2} \le q_R \cdot \frac{\hat{\sigma}_b^2}{\hat{\sigma}_a^2}}$

вариант формулы

экв. q_L, q_R $F_{n_b-1,\, n_a-1}$

$$\left[\frac{1}{q_R} \cdot \frac{\hat{\sigma}_b^2}{\hat{\sigma}_a^2} \le \frac{\sigma_b^2}{\sigma_a^2} \le \frac{1}{q_L} \cdot \frac{\hat{\sigma}_b^2}{\hat{\sigma}_a^2}\right]$$

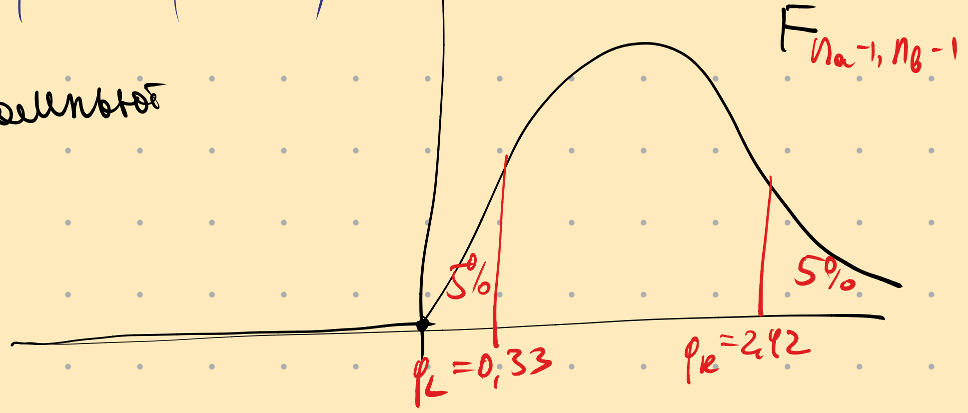
Пример

	число набл.	выб. средние	несм. оц. диспер.
по a:	10	5	40
по b:	20	30	45

90% CI для $\frac{\sigma_b^2}{\sigma_a^2}$

$F_{9,19}$ (!)

по табл / компьютер



реализация интервала:

$$\left[0.33 \cdot \frac{45}{40};\ 2.42 \cdot \frac{45}{40}\right]$$

альтерн. $\left(F_{d_1,d_2}\right)^{-1} \sim F_{d_2,d_1}$

Интервал для разницы $\mu_1 - \mu_2$

стиль 1

x_i	g_i
2	a
3	a
1	b
3	a

стиль 2:

$x_1 \dots x_{n_x}$

$y_1 \dots\dots y_{n_y}$

Асимптотич. CI при мягких предп-х и $n \to \infty$

$\boxed{\text{типовой проект}}$

$$\frac{\hat\theta - E(\hat\theta)}{\sqrt{\hat{Var}(\hat\theta)}} \xrightarrow[n\to\infty]{dist} N(0;1)$$

хотим: интервал для $\mu_a - \mu_b$

① набл. независимы.

② $(x_i \mid g_i = a) \sim Law_a$

$(x_i \mid g_i = b) \sim Law_b$

$E(x_i \mid g_i = a) \in \mathbb{R}$
$Var(x_i \mid g_i = a) \in \mathbb{R}$

$E(x_i \mid g_i = b) = \mu_b \in \mathbb{R}$
$Var(x_i \mid g_i = b) = \sigma_b^2 \in \mathbb{R}$

$\bar x_a - \bar x_b - (\mu_a - \mu_b)$ $\qquad \theta = \mu_a - \mu_b$

t $N(0;1)$

$\xrightarrow{} N(0;1)$ $\hat\theta = \bar x_a - \bar x_b$

$\sqrt{\widehat{Var}(\bar x_a - \bar x_b)}$

далее n_a, n_b считаем фиксир-ми:

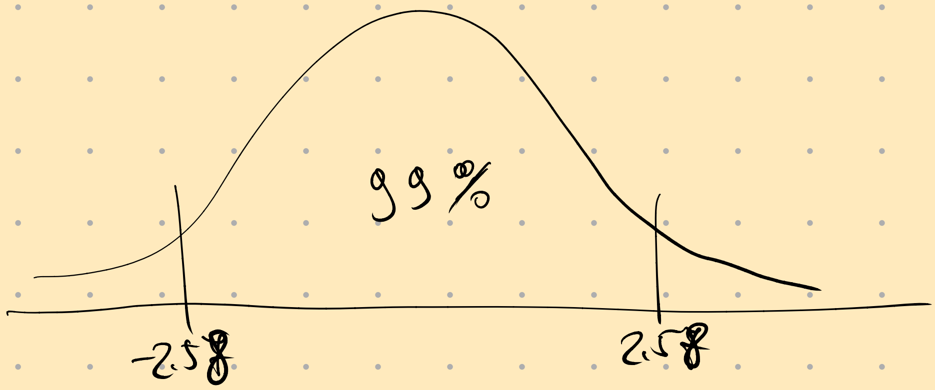
$$\widehat{Var}(\bar x_a - \bar x_b) = \widehat{Var}(\bar x_a) + \widehat{Var}(\bar x_b) = \widehat{Var}\left(\frac{\sum_{g_i=a} x_i}{n_a}\right) + \widehat{Var}\left(\frac{\sum_{g_i=b} x_i}{n_b}\right) = \frac{1}{n_a}\cdot\hat\sigma_a^2 + \frac{1}{n_b}\cdot\hat\sigma_b^2$$

$$Var(\bar x_a - \bar x_b) = \frac{\sigma_a^2}{n_a} + \frac{\sigma_b^2}{n_b}$$

$\hat\sigma_a^2, \hat\sigma_b^2$ — несм. оценки дисперсий

$$\hat\sigma_a^2 = \frac{\sum_a (x_i - \bar x_a)^2}{n_a - 1}$$

$$\frac{\bar x_a - \bar x_b - (\mu_a - \mu_b)}{\sqrt{\frac{\hat\sigma_a^2}{n_a} + \frac{\hat\sigma_b^2}{n_b}}} \xrightarrow{dist} N(0;1)$$



99% asy CI для $\mu_a - \mu_b$

$$[\bar x_a - \bar x_b - 2.58 \cdot se(\bar x_a - \bar x_b);\ \bar x_a - \bar x_b + 2.58 \cdot se(\bar x_a - \bar x_b)]$$

$$se(\bar x_a - \bar x_b) = \sqrt{\frac{\hat\sigma_a^2}{n_a} + \frac{\hat\sigma_b^2}{n_b}}$$

! Рассказ про довер-ный интервал, к-рый лучше не использовать! Он нужен для перехода к более реалистичному!

$\boxed{\text{Интервал для } \mu_a - \mu_b \text{ с помощью } t\text{-статистики}}$

* посылки:
- набл. незав
- $(x_i | g_i = a) \sim N(\mu_a; \sigma^2)$

• $(x_i \mid y_i = b) \sim N(\mu_b; \sigma^2)$

жёсткие предп!

это уже совсем жёстко!!

далее опять n_a, n_b зафиксир.

$\bar{x}_a$ — несм. оценка для μ_a

$\bar{x}_b$ — несм. оценка для μ_b.

несм.

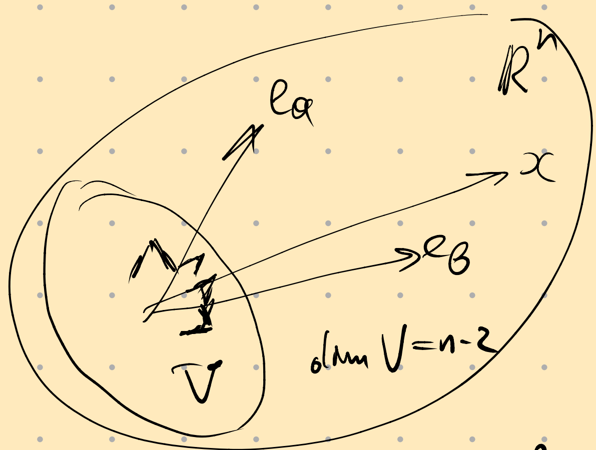
$$\hat{\sigma}^2 = \frac{\sum_a (x_i - \bar{x}_a)^2 + \sum_b (x_i - \bar{x}_b)^2}{(n_a - 1) + n_b - 1}$$

$$E(\hat{\sigma}^2) = \sigma^2$$

$$\frac{\sum_a (x_i - \bar{x}_a)^2 + \sum_b (x_i - \bar{x}_b)^2}{\sigma^2} \sim \chi^2_{n_a + n_b - 2}$$

$n = n_a + n_b$

x	e_a	e_b
x_1	1	0
⋮	1	0
⋮	0	1
⋮	1	0
⋮	0	1
⋮	⋮	⋮
x_n	1	0



$$\sum_a (x_i - \bar{x}_a)^2 + \sum_b (x_i - \bar{x}_b)^2 = \| proj(x, V) \|^2$$

[продолжаем делать непр. жёг]

Упр. $\bar{x}_a - \bar{x}_b - (\mu_a - \mu_b) \sim N\left(0; \frac{\sigma^2}{n_a} + \frac{\sigma^2}{n_b}\right)$

$$Var(\bar{x}_a - \bar{x}_b) = Var(\bar{x}_a) + Var(\bar{x}_b) = \frac{\sigma^2}{n_a} + \frac{\sigma^2}{n_b}$$

независ.

$$t\text{-распр} = \frac{N(0;1)}{\sqrt{\chi^2_d / d}} = \frac{\left(\bar{x}_a - \bar{x}_b - (\mu_a - \mu_b)\right) \Big/ \sqrt{\frac{\sigma^2}{n_a} + \frac{\sigma^2}{n_b}}}{\sqrt{\frac{\sum_a (x_i - \bar{x}_a)^2 + \sum_b (x_i - \bar{x}_b)^2}{\sigma^2} \Big/ n_a + n_b - 2}}$$

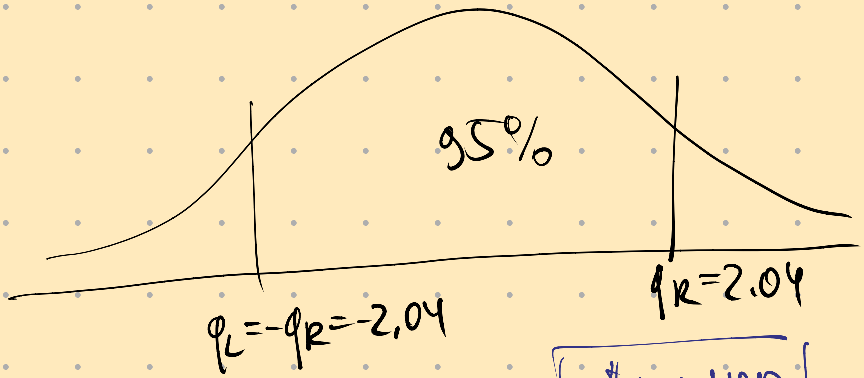
$d = n_a + n_b - 2$

$$= \frac{\bar{x}_a - \bar{x}_b - (\mu_a - \mu_b)}{\sqrt{\hat{\sigma}^2 \cdot \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}} \sim t_{n_a+n_b-2}$$

$$\hat{\sigma}^2 = \frac{\sum_a (x_i - \bar{x}_a)^2 + \sum_b (x_i - \bar{x}_b)^2}{n_a + n_b - 2}$$

$(x_i | g_i = a) \sim N(\mu_a; \sigma^2)$

$(x_i | g_i = b) \sim N(\mu_b; \sigma^2)$



t_{28}

Пример:

	число набл	выб средние	несм. оц. диспер.
по a:	10	5	40
по b:	20	30	45

$$\mu_a - \mu_b \in \left[\bar{x}_a - \bar{x}_b - q_R \cdot se(\bar{x}_a - \bar{x}_b);\ \bar{x}_a - \bar{x}_b + q_R \cdot se(\bar{x}_a - \bar{x}_b)\right]$$

тоже на других предп

$\bar{x}_a - \bar{x}_b = -25$

$$se(\bar{x}_a - \bar{x}_b) = \sqrt{\left(\frac{1}{10} + \frac{1}{20}\right) \cdot \left(\frac{(10-1)\cdot 40 + (20-1)\cdot 45}{10+20-2}\right)}$$

хорошо: малые выборки

минус: $\sigma_a^2 = \sigma_b^2 = \sigma^2$ не бывает!!

предп.: $\sigma_b^2 \neq \sigma_a^2$

$$R = \frac{\bar{x}_a - \bar{x}_b - (\mu_a - \mu_b)}{\sqrt{\frac{\hat{\sigma}_a^2}{n_a} + \frac{\hat{\sigma}_b^2}{n_b}}}$$

- на какое χ^2 распр-ие похож знаменатель?
- вместо $N(0;1)$ предельного распр-ия будем исп-ть

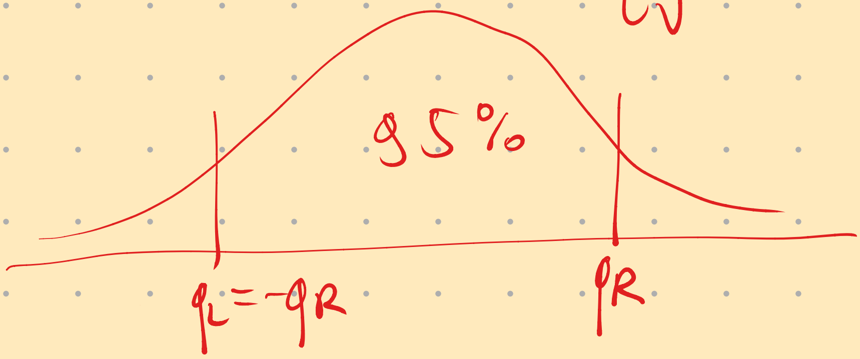
$\llcorner se(\bar{x}_a - \bar{x}_b)$

t-распределения с ν степ. своб.

Статистика Уэлча / тест Уэлча / Welch t-test

асу CI $\mu_a - \mu_b$ $\left[\bar{x}_a - \bar{x}_b - q_R \cdot se(\bar{x}_a - \bar{x}_b);\ \bar{x}_a - \bar{x}_b + q_R \cdot se(\bar{x}_a - \bar{x}_b)\right]$

t_ν



ν?