Семинар 8

Упр. Выборка $n=50$ чел. $X_i \to$ 1 любит сыр; 0 не любит сыр.

предп: X_i незав одинак. распр.

$$\sum_{i=1}^{n} X_i = 40$$

а) постройте 90% дов инт для p (асимп) (с $Var(\hat{p})$ в знам)

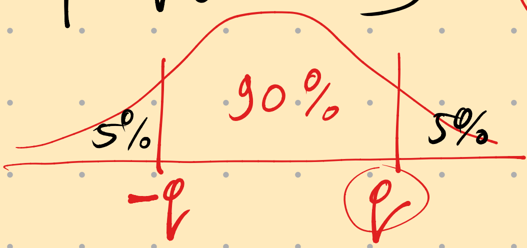
б) — // — (с $\widehat{Var}(\hat{p})$ в знам) лёгкая

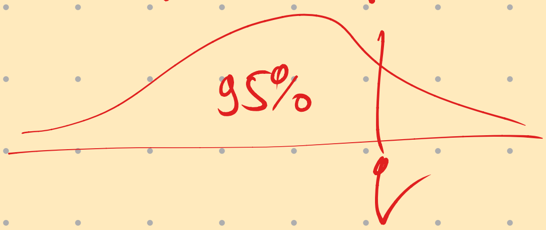
б) $$\left[\hat{p} - q\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\ \hat{p} + q\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

ЦПТ
• подмена в знам Var на $\widehat{Var}$

q – это квантиль

$$\hat{p} = \frac{40}{50} = 0{,}8$$






```
from scipy import stats
stats.norm.ppf(0.95)
```

$N(0;1)$

$$q = 1.64$$

$$\left[0{,}8 - 1{,}64\cdot\sqrt{\frac{0{,}8\cdot 0{,}2}{50}};\ 0{,}8 + 1{,}64\cdot\sqrt{\frac{0{,}8\cdot 0{,}2}{50}}\right]$$

$$[0{,}71;\ 0{,}89]$$

а) $$p \in \left[\frac{q^2 + 2\hat{p}n - \sqrt{D}}{2(n+q^2)};\ \frac{q^2 + 2\hat{p}n + \sqrt{D}}{2(n+q^2)}\right]$$

$$D = (q^2 + 2\hat{p}n)^2 - 4(n+q^2)\cdot n\hat{p}^2$$

• ЦПТ

в знаме ост-ки $Var(\hat{p})$

$\mathcal{D} = 93.9$ $\qquad \frac{\sqrt{\mathcal{D}}}{2(n+q^2)} = 0.09$

$q = 1.64$
$\hat{p} = 0.8$
$n = 50$

$$\frac{q^2 + 2\hat{p}n}{2(n+q^2)} = 0.78$$

$$[0.78 - 0.09;\ 0.78 + 0.09] = [0.69;\ 0.88]$$

$\boxed{\text{Задача 2}}$

X_i	W_i
1	0
1	1
0	0
0	1
0	1
⋮	⋮

$n = 100$

$X_i \to 1 / 0$ [ответил „да" про вопрос про сыр]

$W_i \to 1 / 0$ [посмотрел рекламу до анкетир]

$(X_i, W_i) \sim$ один. расп

набл. незав.

(X_i, W_i) зав

(X_i, W_i) и (X_j, W_j) незав

таблица сопряжённости

	$W_i = 0$	$W_i = 1$
$X_i = 0$	20	10
$X_i = 1$	20	50

кол-ва

contingency table

две выборки

$(W_i = 1)$ $\quad Q_1 = 1,\ Q_2 = 1,\ Q_3 = 0 \dots$ $\quad n_1$ набл.

$(W_i = 0)$ $\quad R_1 = 0,\ R_2 = 1 \dots$ $\quad n_0$ набл.

все Q_i все R_j незав

все $Q_i \sim$ один. расп.
все $R_j \sim$ один. рас

$$p_1 = P(X_i = 1 \mid W_i = 1) \qquad p_0 = P(X_i = 1 \mid W_i = 0)$$

$$p_1 = P(Q_i = 1) \qquad p_0 = P(R_i = 1)$$

$\boxed{\text{асу CI } p_1 - p_0\ ?}$

легко: → берём наивную оценку $\hat{p}_1 - \hat{p}_0$
→ применяем ЦПТ
с $Var \to \hat{Var}$

а) $\hat{p}_1 - \hat{p}_0$ на языке (X_i, W_i):

n_1 − кол-во $W_i = 1$
$n_1 = \sum W_i$!!

$\hat{}\ \ \hat{}\ \ \sum X_i \ \ \sum X_i$

$$p_1 - p_0 = \frac{\sum_{W_i=1} X_i}{n_1} - \frac{\sum_{W_i=0} X_i}{n_0}$$

на языке (Q_i) и (R_j)

$$\hat{p}_1 - \hat{p}_0 = \bar{Q} - \bar{R}$$

$$\bar{Q} = \frac{\sum Q_i}{n_1} \qquad \bar{R} = \frac{\sum R_j}{n_0}$$

б) $\operatorname{Var}(\hat{p}_1 - \hat{p}_0 \mid n_1, n_0) \overset{?}{=}$

в) сост. $\widehat{\operatorname{Var}}(\hat{p}_1 - \hat{p}_0 \mid n_1, n_0)$

plug-in estimator

$$\left[\operatorname{plim} \frac{\widehat{\operatorname{Var}}(\hat{p}_1 - \hat{p}_0)}{\operatorname{Var}(\hat{p}_1 - \hat{p}_0)} = 1\right]$$

$$\operatorname{Var}(\hat{p}_1 - \hat{p}_0 \mid n_1, n_0) \overset{?}{=} \operatorname{Var}(\hat{p}_1 \mid n_1, n_0) + \operatorname{Var}(\hat{p}_0 \mid n_1, n_0) - 2\cancel{\operatorname{Cov}(\hat{p}_1, \hat{p}_0 \mid n_1, n_0)}^{\,0}$$

$$= \frac{1}{n_1} \cdot \operatorname{Var}(X_i \mid W_i = 1) + \frac{1}{n_0} \cdot \operatorname{Var}(X_i \mid W_i = 0) =$$

$$= \frac{p_1 \cdot (1 - p_1)}{n_1} + \frac{p_0 \cdot (1 - p_0)}{n_0}$$

← не знаем!! в наст-ей дисперсии настоящую p_1 и p_0

в) $\widehat{\operatorname{Var}}(\hat{p}_1 - \hat{p}_0 \mid n_1, n_0) = \dfrac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \dfrac{\hat{p}_0(1 - \hat{p}_0)}{n_0}$ // ранее $\boxed{\dfrac{p(1-p)}{n}}$

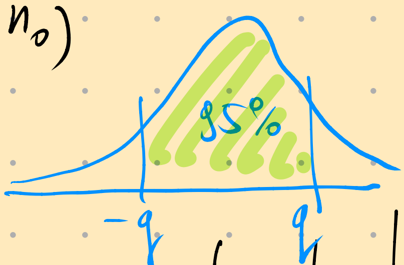
сост. оц.

ЦПТ

лемма Слуцкого

$$\frac{(\hat{p}_1 - \hat{p}_0) - E(\hat{p}_1 - \hat{p}_0 \mid n_1, n_0)}{\sqrt{\widehat{\operatorname{Var}}(\hat{p}_1 - \hat{p}_0 \mid n_1, n_0)}} \to N(0, 1)$$

$$-q \le \frac{\hat{p}_1 - \hat{p}_0 - (p_1 - p_0)}{se(\hat{p}_1 - \hat{p}_0)} \le q$$



se = standard error

станд. ошибка

$$se(\hat\theta) = \sqrt{\hat{Var}(\hat\theta | \ldots)}$$ ← корень из оценки дисперсии

$$p_1 - p_0 \in \left[\hat p_1 - \hat p_0 - q \cdot se(\hat p_1 - \hat p_0);\ \hat p_1 - \hat p_0 + q \cdot se(\hat p_1 - \hat p_0)\right]$$

$$\hat p_1 - \hat p_0 = \frac{5}{6} - \frac{1}{2} = \frac{5-3}{6} = \frac{2}{6} = \frac{1}{3} = 0{,}33\ldots$$

кол-ва:

	$W_i = 0$	$W_i = 1$
$X_i = 0$	20	10
$X_i = 1$	20	50

$$\hat p_1 = \frac{50}{60} = \hat p(X_i = 1 | W_i = 1)$$
$$\hat p_0 = \frac{20}{40} = \hat p(X_i = 1 | W_i = 0)$$

$$se(\hat p_1 - \hat p_0) = \sqrt{\frac{\hat p_1 \cdot (1 - \hat p_1)}{n_1} + \frac{\hat p_0 (1 - \hat p_0)}{n_0}} = 0{,}09\ldots$$

асимп. CI 95% $\quad q = 1{,}96 \quad [0{,}33 - 1{,}96 \cdot 0{,}09 ;\ 0{,}33 + 1{,}96 \cdot 0{,}09]$

Задача.

X_i → 🙂 доволен, 😐, 🙁 недов.

X_i ~ незав. один. распр.

$P(X_i = 🙂) = p_+ \quad P(X_i = 🙁) = p_- \quad P(X_i = 😐) = 1 - p_+ - p_-$

X_i
🙂
🙁
😐
🙂

	🙂	😐	🙁
кол-во	30	20	10

настоящие не знаем

$\theta = p_+ - p_-$ „разность"

n – всего людей
n_+ – сколько 🙂
n_- – сколько 🙁
n_0 – сколько 😐

!! а) наивная точечная $\hat\theta = \hat p_+ - \hat p_- = \frac{n_+}{n} - \frac{n_-}{n} =$

б) сост. оц-ку $\hat{Var}(\hat\theta) = \frac{30-10}{60}$

[шокан] δ $\quad Var(\hat\theta)$

$= 0{,}33\ldots$

2) одз 95 % CI для θ.

д) $Var(\hat p_+ - \hat p_-) = Var\left(\frac{\sum G_i - \sum B_i}{n}\right) = ⊛$

$\boxed{\text{разложение в сумму индикаторов}}$

$$G_i = \begin{cases} 1, \text{ если } X_i = 🙂 \\ 0, \text{ если } X_i \neq 🙂 \end{cases}$$

$$B_i = \begin{cases} 1, \text{ если } X_i = 🙁 \\ 0, \text{ если } X_i \neq 🙁 \end{cases}$$

X_i	G_i	B_i
🙂	1	0
🙁	0	1
😐	0	0
⋮	⋮	⋮

зав

$G_1 \quad B_1$ незав

незав

$G_2 \quad B_2$

$G_3 \quad B_3$

зав

$G_i \cdot B_i \equiv 0$

$$⊛ = Var\left(\frac{\sum(G_i - B_i)}{n}\right) = \frac{1}{n^2}\sum Var(G_i - B_i) =$$

$$= \frac{n}{n^2} \cdot Var(G_1 - B_1) = \frac{1}{n} \cdot Var(G_1 - B_1) = \frac{p_+ + p_- - (p_+ - p_-)^2}{n}$$

несм. незав. дисп.

вер X_1	p_+ = 🙂	p_0 = 😐	p_- = 🙁
G_1	1	0	0
B_1	0	0	1
$G_1 - B_1$	1	0	−1
$(G_1 - B_1)^2$	1	0	1

$$Var(G_1 - B_1) = E((G_1 - B_1)^2) - (E(G_1 - B_1))^2 =$$

$$= 1 \cdot p_+ + 1 \cdot p_- - (p_+ - p_-)^2$$

$\ldots \hat{p}_+ + \hat{p}_- - (\hat{p}_+ - \hat{p}_-)^2$

б) сост $\widehat{Var}(\hat{p}_+ - \hat{p}_-) = \frac{\hat{p}_+ + \hat{p}_- - (\hat{p}_+ - \hat{p}_-)^2}{n}$!!

$$se(\hat{p}_+ - \hat{p}_-) = \sqrt{\frac{\hat{p}_+ + \hat{p}_- - (\hat{p}_+ - \hat{p}_-)^2}{n}} = $$

$$\approx 0.10$$

$\hat{p}_+ = \frac{30}{60}$

$\hat{p}_- = \frac{10}{60}$

$n = 60$

$$\hat{\theta} = \hat{p}_+ - \hat{p}_-$$

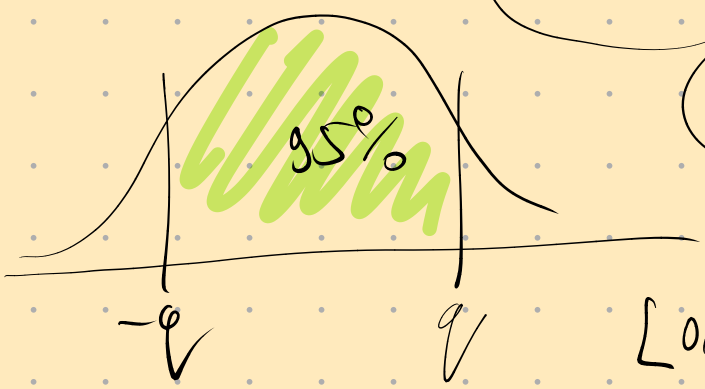
2) ЦПТ + лемма Слуцкого

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

asy CI

$$\theta \in \left[\hat{\theta} - q \cdot se(\hat{\theta}); \hat{\theta} + q \cdot se(\hat{\theta})\right]$$

$$-q \leq \frac{\hat{\theta} - \theta}{se(\hat{\theta})} \leq q$$



$$[0{,}33 - 1{,}96 \cdot 0{,}10; 0{,}33 + 1{,}96 \cdot 0{,}10]$$