SkillFactory

SkillFactory

Доверительные интервалы*

* confidence intervals

Confidence comes not from always being right, but from not fearing being wrong.

(Peter T. McIntyre)

Оценка параметров ген. совокупности

Можем посчитать: **Выборочные статистики**	**Хотим знать:** **Параметры генеральной совокупности**
s	
	p

Это случайные величины

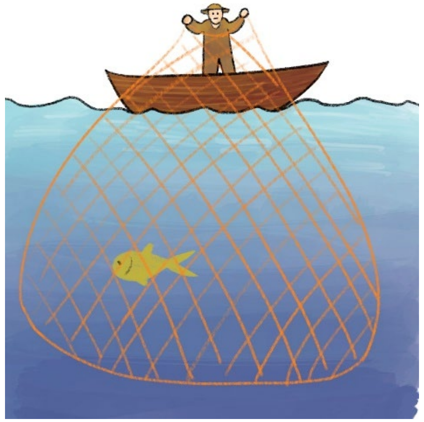
Это константы

Точечные и интервальные оценки

Оценки параметров

Точечные оценки

Интервальные оценки



Доверительный интервал: пример

$$\mu = \bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

X – заработная плата

n=100, σ=25, $\bar{X} = 153$,

$\mu \in (148; 158)$ на 95%-ном уровне доверия

Доверительный интервал для μ

По ЦПТ: $\bar{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$

Приведем к станд. виду: $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx Z \sim N(0, 1)$

Выразим $\bar{X}$: $\bar{X} = \mu + Z \cdot \frac{\sigma}{\sqrt{n}}$

Истинное среднее

Случайная ошибка

Переформулируем: $\bar{X} = \mu \pm z_{\text{крит}} \cdot \frac{\sigma}{\sqrt{n}}$

Выразим μ: $\mu = \bar{X} \pm z_{\text{крит}} \cdot \frac{\sigma}{\sqrt{n}}$

$$\mu \in (\bar{X} - z_{\text{крит}} \cdot \frac{\sigma}{\sqrt{n}} ; \bar{X} + z_{\text{крит}} \cdot \frac{\sigma}{\sqrt{n}})$$

с определенной степенью уверенности

Структура доверительного интервала

$$\mu = \bar{X} \pm z_{\text{крит}} \cdot \frac{\sigma}{\sqrt{n}}$$

Параметр = Оценка $\pm$ Предел погрешности

Предел погреш. = крит.знач. $\times$ ст.откл. оценки

Как найти *z*-критическое?

Пример:

X – цена планшета

$\mu = 175, \sigma = 5, \mathrm{n}=100;$

По ЦПТ: $\bar{X} \approx N(175, \frac{5}{\sqrt{100}})$

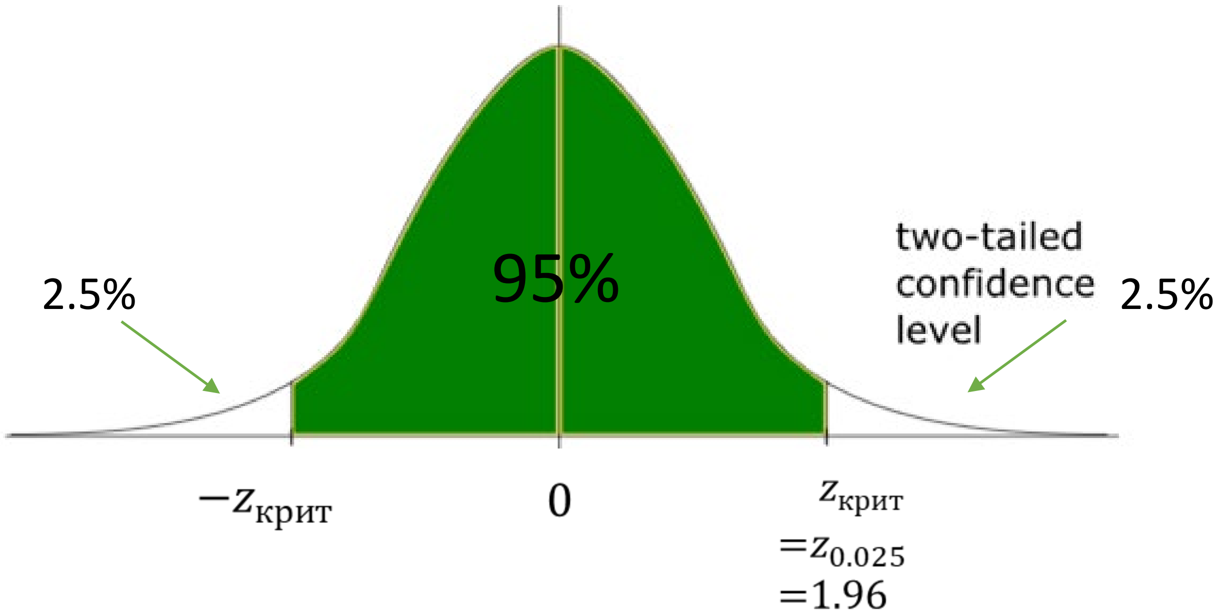
$$\bar{X} = 175 \pm z_{\text{крит}} \cdot \frac{5}{\sqrt{100}}$$

Какое z-критическое

даст интервал для $\bar{X}$ с 95%-ной вероятностью?

Как найти z-критическое?

$$Z \sim N(0,1),$$
$$\bar{X} = 175 \pm z_{\text{крит}} \cdot \frac{5}{\sqrt{100}}$$

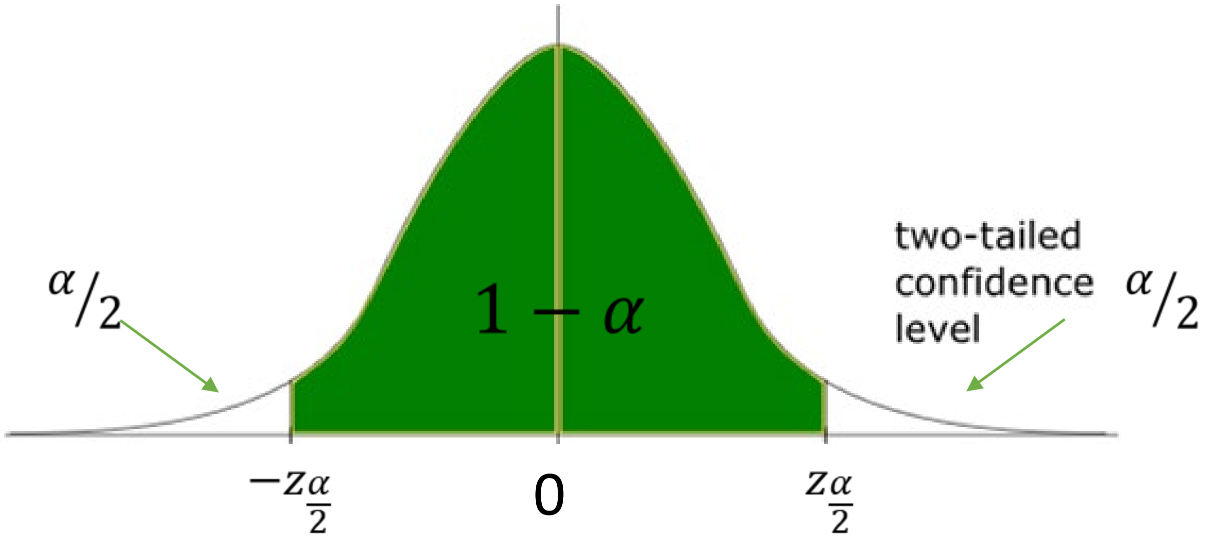


$$z_{\text{крит}} = z_{\frac{0.05}{2}} = \mathbf{z_{0.025}} = 1.96$$

Как найти z-критическое?

$$\bar{X} = 175 \pm 1.96 \cdot \frac{5}{\sqrt{100}} \approx 175 \pm 1$$

$\bar{X} \epsilon (174; 176)$ с вероятностью 95%



$1 - \alpha =$ **уровень доверия**

$\alpha =$ **уровень значимости**

Формула доверительного интервала

В общем виде:

$$\bar{X} = \mu \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

с вероятностью $(1 - \alpha)$

Выразим μ:

$$\boxed{\mu = \bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}$$

на уровне доверия $(1 - \alpha)$

Пример решения задачи

Была опрошена случайная выборка из **36** жителей региона об их затратах на продукты питания за последний месяц. Выборочное среднее оказалось **16 100** рублей.
Допустим, известно, что стандартное отклонение расходов равно **12 000** рублей.

Найдите 95%-ный доверительный интервал для средних расходов жителя данного региона на продукты питания в месяц.

X – расходы на продукты питания
$\bar{X} = 16100, \sigma = 12000, \mathrm{n} = 36$

Пример решения задачи

Условие: X – расходы на продукты питания
$\bar{X} = 16100$, $\sigma = 1200$, n=36

Проверка допущений:
1. Выборка случайна
2. n>30 , значит по ЦПТ: $\bar{X} \approx N(\mu, \frac{\sigma}{\sqrt{n}})$

Решение:

$$\mu = \bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$z_{0.025} = 1.96$$

$\mu = 16100 \pm 1.96 \cdot \frac{12000}{\sqrt{36}} = 16100 \pm 3920$

$\mu \in (12180; 20020)$ на 95%-ном уровне доверия.

Стандартные значения z -крит.

90%	
95%	
99%	

Есть шанс, что мы «промахнемся»

Доверительный интервал строится на данных случайной выборки.

Уровень доверия α = вероятность того, что наш «случайный» интервал НЕ включит истинное значение параметра.

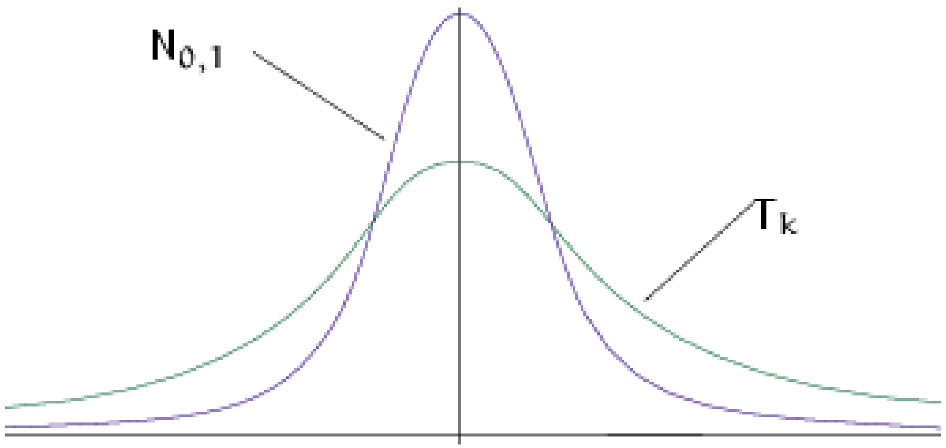
Распределение Стьюдента

- $$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx Z \sim N(0,1)$$

$$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \approx T(k),$$

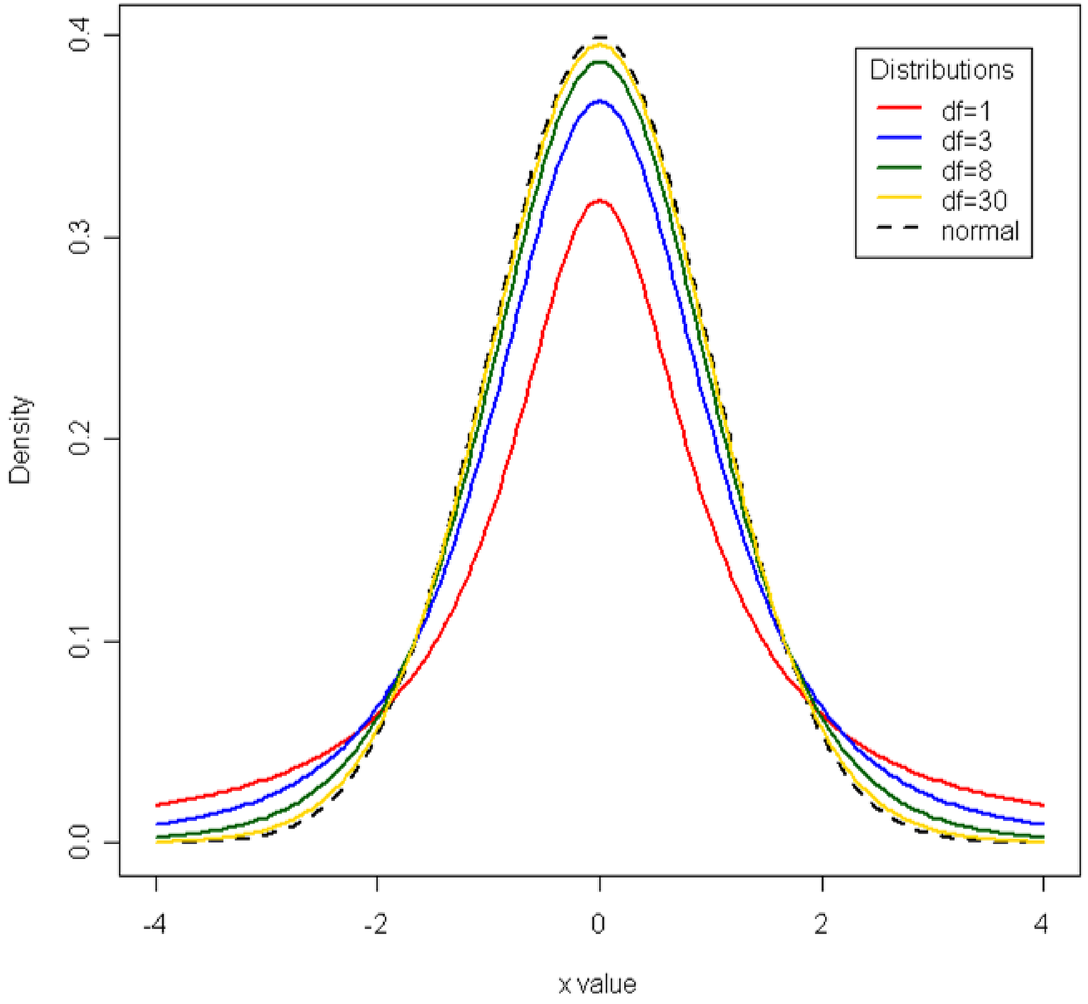
Т-распределение Стьюдента

k-степени свободы (degrees of freedom, df), k=n-1



Распределение Стьюдента

- $$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \approx T(k)$$



Какую формулу использовать

Доверительный интервал для μ

?

$$\mu = \bar{X} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Если σ известно

$$\mu = \bar{X} \pm t_{\frac{\alpha}{2}}(k) \cdot \frac{s}{\sqrt{n}}$$

Если σ неизвестно

Доверительный интервал для $\boldsymbol{p}$

- Если n велико ($n\hat{p} > 5,\ n(1-\hat{p}) > 5$), по ЦПТ:

$$\hat{p} \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

$$\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx Z \sim N(0,1)$$

$$\Downarrow$$

$$p = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Пример решения задачи

- 64% случайной выборки из 550 человек заявили, что поддержат кандидата.

Определите 99%-ный доверительный интервал для пропорции избирателей

p – пропорция избирателей за кандидата
$\hat{p} = 0.64, n = 550$

Пример решения задачи

Условие:

p – пропорция избирателей за кандидата
$\hat{p} = 0.64, n = 550$

Проверка допущений:

1. Выборка взята случайно
2. $n\hat{p} = 550 \cdot 0.64 = 352 > 5,$

$n(1 - \hat{p}) = 0.36 \cdot 550 = 198 > 5$

Решение: $p = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$z_{0.005} = 2.58$

$p = 0.64 \pm 2.58 \sqrt{\frac{0.64 \cdot 0.36}{550}} = 0.64 \pm 0.053$

Мы на 99% уверены, что пропорция избирателей между **0.587 и 0.693.**