Интервальные оценки в статистике

Разберём **пошагово** построение **доверительного интервала для математического ожидания μ** при **известном стандартном отклонении σ**.

Условия задачи:

Пусть:

- $X_1, X_2, \ldots, X_n$ — независимые и одинаково распределённые случайные величины,
- $X_i \sim N(\mu, \sigma^2)$,
- σ — **известно**,
- μ — **неизвестно**, и мы хотим построить для него **доверительный интервал с уровнем надёжности** $1 - \alpha$.

1. Выборочная характеристика

Рассмотрим выборочное среднее:

Рассмотрим **выборочное среднее**:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Из свойств нормального распределения:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

2. Стандартизуем выборочное среднее

Определим стандартизованную величину:

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

Это означает, что:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

где $z_{\alpha/2}$ — квантиль стандартного нормального распределения, то есть:

$$P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

3. Преобразуем неравенство

Перепишем неравенство, выразив μ:

$$P\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

4. Финальная формула доверительного интервала для μ:

$$\boxed{\mu \in \left[\bar{X} \; - \; z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{X} \; + \; z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]}$$

Это и есть **доверительный интервал уровня $1 - \alpha$.**

5. Пример

Пусть:

Пусть:

- $\bar{X} = 100$ — средняя зарплата,
- $\sigma = 15$ — известное стандартное отклонение,
- $n = 36$ — объём выборки,
- $\alpha = 0.05$ → уровень доверия 95%,
- $z_{0.025} \approx 1.96$.

Подставим:

$$\mu \in \left[100 - 1.96 \cdot \frac{15}{\sqrt{36}}, \quad 100 + 1.96 \cdot \frac{15}{\sqrt{36}}\right] = [95.1, 104.9]$$

Ответ: **доверительный интервал для среднего μ: [95.1, 104.9]**

📌 Напоминание:

Уровень доверия	α	$z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.576

Теперь выведем **доверительный интервал для математического ожидания μ**, когда **сигма неизвестна**, т.е. переходим от нормального к **t-распределению Стьюдента**.

Условия задачи:

Пусть:

- $X_1, X_2, \ldots, X_n$ — независимые и одинаково распределённые случайные величины,
- $X_i \sim N(\mu, \sigma^2)$,
- μ — **неизвестно**,
- σ — **тоже неизвестно** (в отличие от предыдущего случая).

1. Выборочные характеристики

- Среднее:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Выборочная дисперсия (несмещённая):

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- Стандартная ошибка среднего:

$$\mathrm{SE} = \frac{S}{\sqrt{n}}$$

2. Проблема: σ неизвестно ⇒ нельзя использовать Z-распределение

Раньше мы писали:

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

Теперь мы не знаем σ, и заменяем её на S. Но тогда **распределение становится не нормальным!**

3. Что делать: используем t-распределение Стьюдента

Факт из теории (доказанный через нормальное и хи-квадрат распределения):

Если $X_i \sim N(\mu, \sigma^2)$, то:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

где:

- S — выборочное стандартное отклонение,
- t_{n-1} — t-распределение с $n-1$ степенью свободы.

4. Построим интервал с уровнем доверия $1 - \alpha$

Из свойства симметричного t-распределения:

$$P\left(-t_{\alpha/2,n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2,n-1}\right) = 1 - \alpha$$

Умножим неравенство на $S/\sqrt{n}$:

$$P\left(\bar{X} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

5. Финальная формула доверительного интервала:

$$\boxed{\mu \in \left[\bar{X} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}},\ \bar{X} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}\right]}$$

6. Пример

Пусть:

- $\bar{X} = 75$,
- $S = 12$,
- $n = 16$,
- Уровень доверия: $95\% \Rightarrow \alpha = 0.05$,

- Степени свободы: $df = 15$,
- $t_{0.025,15} \approx 2.131$ (из таблицы).

Тогда:

$$\text{SE} = \frac{S}{\sqrt{n}} = \frac{12}{4} = 3$$

$$\mu \in [75 - 2.131 \cdot 3,\ 75 + 2.131 \cdot 3] = [68.6, 81.4]$$

Ответ: доверительный интервал: $\mu \in [68.6, 81.4]$

Теперь рассмотрим доверительный интервал для доли (вероятности успеха) — это важный случай, особенно в опросах, тестировании гипотез, A/B тестах и т.п.

Цель

Пусть:

- p — неизвестная вероятность успеха (доля),
- $\hat{p} = \frac{x}{n}$ — выборочная доля успехов, где:
  - x — число "успешных" наблюдений,
  - n — объём выборки.

Нужно построить **доверительный интервал для p** с доверительным уровнем $1 - \alpha$.

1. Мотивация: биномиальное распределение

Случайная величина $X \sim Bin(n, p)$ описывает число успехов в n испытаниях.

Выборочная доля:

$$\hat{p} = \frac{X}{n}$$

Имеет математическое ожидание и дисперсию:

$$E[\hat{p}] = p, \quad \text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

2. Центральная предельная теорема (ЦПТ)

По ЦПТ, при больших n распределение p приближается к нормальному:

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

Поэтому можно использовать нормальное приближение для построения доверительного интервала.

3. Формула доверительного интервала для p:

Если n достаточно велико (обычно $np \geq 10,\ n(1-p) \geq 10$), то:

$$\boxed{p \in \left[\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]}$$

где:

- $\hat{p} = \frac{x}{n}$ — выборочная доля,
- $z_{\alpha/2}$ — квантиль стандартного нормального распределения.

4. Пример

Пусть:

- $x = 47$ респондентов из $n = 100$ поддерживают инициативу,
- Уровень доверия: 95% → $z_{0.025} \approx 1.96$.

Вычисляем:

$$\hat{p} = \frac{47}{100} = 0.47$$

$$\mathrm{SE} = \sqrt{\frac{0.47 \cdot (1-0.47)}{100}} = \sqrt{\frac{0.2491}{100}} \approx 0.0499$$

$$\text{Погрешность} = 1.96 \cdot 0.0499 \approx 0.0978$$

$$\Rightarrow p \in [0.47 - 0.0978,\ 0.47 + 0.0978] = [0.3722,\ 0.5678]$$

Ответ: доверительный интервал для доли: $p \in [0.372, 0.568]$

5. Условия применимости нормального приближения:

- $n\hat{p} \geq 10$,
- $n(1-\hat{p}) \geq 10$.

Если условия не выполняются, можно:

- использовать точный биномиальный интервал (Кловера-Пирсона),
- или коррекцию Уилсона, которая лучше работает при малых n.

Альтернативный вид записи

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

Интервал Уилсона (Wilson score interval)

Идея:

Это улучшенный приближённый интервал, который корректирует смещение z-интервала. **Особенно хорош при малых n и долях, близких к 0 или 1.**

Формула интервала Уилсона:

$$\text{центр} = \frac{\hat{p} + \frac{z^2}{2n}}{1 + \frac{z^2}{n}}, \quad \text{половина ширины} = \frac{z}{1 + \frac{z^2}{n}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}$$

Тогда сам интервал:

$$\left[\frac{\hat{p} + \frac{z^2}{2n} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}\right]$$

Теперь выведем и объясним, **как строится доверительный интервал для дисперсии генеральной совокупности σ^2**, когда она **оценивается по нормальной выборке**, а μ может быть известным или неизвестным.

Цель:

Построить доверительный интервал для генеральной дисперсии σ^2:

1. Выборочная дисперсия

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

где:

- $\bar{X}$ — выборочное среднее,
- S^2 — **несмещённая** оценка дисперсии σ^2.

2. Ключевой факт из теории:

Если $X_i \sim N(\mu, \sigma^2)$, то:

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}}$$

Это фундаментальный результат — **относительно нормальной выборки** отношение S^2 к σ^2 (с масштабом) имеет **хи-квадрат распределение**.

3. Построим доверительный интервал

Используем квантильные значения хи-квадрат распределения:

- $\chi^2_{\alpha/2,\, n-1}$,
- $\chi^2_{1-\alpha/2,\, n-1}$.

Из распределения:

$$P\left(\chi^2_{1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2}\right) = 1 - \alpha$$

Преобразуем неравенство и выразим σ^2:

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right) = 1 - \alpha$$

4. Финальная формула доверительного интервала для σ^2:

$$\boxed{\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,\, n-1}}, \quad \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\, n-1}} \right]}$$

А если нужен интервал для стандартного отклонения σ, то просто извлекаем корень:

$$\boxed{\sigma \in \left[\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2,\, n-1}}}, \quad \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\, n-1}}} \right]}$$

5. Пример

Пусть:

- $n = 10$,
- $S^2 = 4.0$,
- Уровень доверия $95\% \Rightarrow \alpha = 0.05$,
- $\chi^2_{0.025,\, 9} = 19.02$,
- $\chi^2_{0.975,\, 9} = 2.70$.

Подставим в формулу:

$$\sigma^2 \in \left[\frac{9 \cdot 4}{19.02}, \frac{9 \cdot 4}{2.70} \right] = [1.89, 13.33]$$

6. Условия применимости

1. Должна быть нормальная выборка $X_i \sim N(\mu, \sigma^2)$,

2. Без нормальности распределения это не работает корректно — хи-квадрат распределение перестаёт быть точным.

Теперь разберём, как построить доверительный интервал для стандартного отклонения σ, если объём выборки большой и распределение неизвестно или не совсем нормальное.

Когда применяется?

- Объём выборки **большой** ($n > 30$, желательно $n > 100$).
- Распределение **не обязано быть нормальным**.
- Используем **Центральную предельную теорему** для приближения.

Цель:

Построить доверительный интервал для **стандартного отклонения** $\boldsymbol{\sigma}$ по большой выборке $X1, \ldots, Xn$:

- S — выборочное стандартное отклонение,
- n — размер выборки,
- σ — истинное стандартное отклонение.

Идея построения:

По ЦПТ, при большом n, выборочная дисперсия S^2 будет примерно нормально распределена. Но прямого нормального распределения у S нет, поэтому делаем замену через ошибку среднего и применяем логарифмическое приближение или используем **асимптотический интервал**.

Мы знаем, что точный результат:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

А при больших nnn, по **ЦПТ**, распределение χ^2_{n-1} приближается к нормальному:

$$\chi^2_{n-1} \approx N(n-1,\ 2(n-1))$$

Отсюда:

- Стандартное отклонение: $\sqrt{2(n-1)}$,
- Используем это в z-форме:

$$\frac{\chi^2_{n-1} - (n-1)}{\sqrt{2(n-1)}} \sim N(0,1)$$

Тогда доверительный интервал:

$$I_\gamma \left(\frac{(n-1) \cdot S^2}{(n-1) + z_{\alpha/2} \cdot \sqrt{2(n-1)}} \, ; \, \frac{(n-1) \cdot S^2}{(n-1) - z_{\alpha/2} \cdot \sqrt{2(n-1)}} \right)$$