Лабораторная работа №3

Выполнил: Сайфуллин Динислам

Вариант 2

Задание 1

Требуется построить доверительный интервал уровня $1-\alpha$ для разности математических ожиданий двух нормальных выборок при неизвестных, но равных дисперсиях.

Имеются две независимые выборки:

- $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu_1, \sigma^2)$
- $Y_1, Y_2, \ldots, Y_m \sim \mathcal{N}(\mu_2, \sigma^2)$

Необходимо построить доверительный интервал для разности математических ожиданий двух нормальных распределений:

$$\tau = \mu_1 - \mu_2$$

При этом дисперсии неизвестны, но предполагаются равными ($\sigma_1^2 = \sigma_2^2 = \sigma^2$).

Оценкой математического ожидания в каждой выборке являются их выборочные средние:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$$

Разность выборочных средних $\bar{X} - \bar{Y}$ является несмещённой оценкой параметра τ.

Так как дисперсии равны, для оценки общей дисперсии используется объединённая оценка дисперсии:

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

где:

- $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ — выборочная дисперсия выборки X,
- $S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$ — выборочная дисперсия выборки Y.

Таким образом, стандартная ошибка разности средних равна:

$$\mathrm{SE}(\bar{X} - \bar{Y}) = \sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}$$

Так как X_i и Y_j имеют нормальное распределение, то разность средних также имеет нормальное распределение.

После нормировки на стандартную ошибку разности получаем статистику:

$$T = \frac{(\bar{X} - \bar{Y}) - \tau}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Эта статистика имеет распределение Стьюдента с $n+m-2$ степенями свободы:

$$T \sim t_{n+m-2}$$

Доверительный интервал уровня доверия $1-\alpha$ для параметра τ строится на основе критических значений распределения Стьюдента.

Интервал имеет вид:

$$\left((\bar{X} - \bar{Y}) - t_{1-\alpha/2} \cdot S_p\sqrt{\frac{1}{n} + \frac{1}{m}}, \quad (\bar{X} - \bar{Y}) + t_{1-\alpha/2} \cdot S_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

где:

- $t_{1-\alpha/2}$ — квантиль уровня $1-\alpha/2$ распределения Стьюдента с $n+m-2$ степенями свободы,
- $\bar{X}, \bar{Y}$ — выборочные средние,
- S_p — объединённая оценка стандартного отклонения.

Таким образом, доверительный интервал симметрично охватывает оценку разности средних.

Реализация на Python

In [1]:
```
import numpy as np
from scipy import stats

mu1 = 2
mu2 = 1
sigma1 = 1
sigma2 = 1
true_tau = mu1 - mu2
alpha = 0.05

def simulate_confidence(n, m):
    coverages = 0

    for _ in range(1000):
        X = np.random.normal(mu1, np.sqrt(sigma1), n)
        Y = np.random.normal(mu2, np.sqrt(sigma2), m)

        mean_X = np.mean(X)
        mean_Y = np.mean(Y)
        var_X = np.var(X, ddof=1)
        var_Y = np.var(Y, ddof=1)

        Sp2 = ((n - 1) * var_X + (m - 1) * var_Y) / (n + m - 2)
        se = np.sqrt(Sp2 * (1/n + 1/m))

        t_crit = stats.t.ppf(1 - alpha/2, df=n + m - 2)

        lower = (mean_X - mean_Y) - t_crit * se
        upper = (mean_X - mean_Y) + t_crit * se

        if lower <= true_tau <= upper:
            coverages += 1

    coverage_rate = coverages / 1000
    return coverage_rate

coverage_25 = simulate_confidence(25, 25)
coverage_10000 = simulate_confidence(10000, 10000)

print(f"Покрытие при n = 25: {coverage_25}")
print(f"Покрытие при n = 10000: {coverage_10000}")
```

```
Покрытие при n = 25: 0.953
Покрытие при n = 10000: 0.952
```

Задание 2

Пусть $X_1, X_2, \ldots, X_n$ — выборка из **распределения Лапласа** с неизвестным параметром сдвига μ и единичным масштабом:

$$f(x;\mu) = \frac{1}{2} e^{-|x-\mu|}, \quad x \in \mathbb{R}.$$

Нужно построить **асимптотический** доверительный интервал уровня $1-\alpha$ для медианы этого распределения.
Так как для Лапласа медиана совпадает с μ, будем строить доверительный интервал для μ.

В качестве оценки медианы используем **выборочную медиану**:

$$\hat{M}_n = \begin{cases} X_{(k+1)}, & \text{если } n = 2k+1, \\ \dfrac{X_{(k)} + X_{(k+1)}}{2}, & \text{если } n = 2k. \end{cases}$$

Здесь $X_{(1)} \le \cdots \le X_{(n)}$ — упорядоченные наблюдения.

Случайная величина $F(X)$ имеет непрерывное распределение, и по теореме о выборочном квантиле:

$$\sqrt{n}\left(\hat{M}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{4f(\mu)^2}\right).$$

Для распределения Лапласа с $f(x) = \frac{1}{2} e^{-|x-\mu|}$ в точке $x = \mu$:

$$f(\mu) = \frac{1}{2} \quad \Longrightarrow \quad \frac{1}{4f(\mu)^2} = \frac{1}{4\,(1/2)^2} = 1.$$

Следовательно,

$$\sqrt{n}\,(\hat{M}_n - \mu) \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{или эквивалентно} \quad \hat{M}_n \approx \mathcal{N}\Big(\mu, \frac{1}{n}\Big).$$

Используя асимптотику нормального приближения:

$$z_{1-\alpha/2} = \text{квантиль } \mathcal{N}(0, 1),$$

получаем

$$\hat{M}_n \pm z_{1-\alpha/2}\sqrt{\frac{1}{n}} \quad \Longrightarrow \quad \left(\hat{M}_n - z_{1-\alpha/2}\frac{1}{\sqrt{n}}, \ \hat{M}_n + z_{1-\alpha/2}\frac{1}{\sqrt{n}}\right).$$

Реализация на Python

In [2]:
```
def simulate_laplace(n):
    true_mu = 2
    coverages = 0

    for _ in range(1000):
        sample = np.random.laplace(loc=true_mu, scale=1, size=n)
        median = np.median(sample)

        se = np.sqrt(1 / (2 * n))
        z_crit = stats.norm.ppf(1 - alpha/2)

        lower = median - z_crit * se
        upper = median + z_crit * se

        if lower <= true_mu <= upper:
            coverages += 1

    coverage_rate = coverages / 1000
    return coverage_rate

coverage_25 = simulate_laplace(25)
coverage_10000 = simulate_laplace(10000)

print(f"Покрытие при n = 25: {coverage_25}")
print(f"Покрытие при n = 10000: {coverage_10000}")
```

```
Покрытие при n = 25: 0.806
Покрытие при n = 10000: 0.839
```