Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

ОТЧЕТ

по лабораторной работе № 2
по курсу «Математическая статистика»
на тему: «Интервальные оценки»
Вариант № 1

Студент ИУ7-63Б (Группа) _______________ (Подпись, дата) Авдейкина В. П. (И. О. Фамилия)

Преподаватель _______________ (Подпись, дата) Власов П. А. (И. О. Фамилия)

2024 г.

СОДЕРЖАНИЕ

1 Теоретические сведения

1.1 Определения

Рассматривается вторая основная задача математической статистики.

Дано: случайная величина X, закон распределения которой известен с точностью до неизвестного параметра θ.

Требуется: оценить значение θ.

Интервальной оценкой уровня γ (с коэффициентом доверия γ) параметра θ называют пару статистик $\underline{\theta}(\vec{X})$ и $\bar{\theta}(\vec{X})$ таких, что

$$P\{\, \underline{\theta}(\vec{X}) < \theta < \bar{\theta}(\vec{X}) \,\} = \gamma. \tag{1.1}$$

Доверительным интервалом уровня γ (γ–доверительным интервалом) называется интервал

$$(\underline{\theta}(\vec{x}), \bar{\theta}(\vec{x})), \tag{1.2}$$

отвечающий выборочным значениям $\underline{\theta}(\vec{X})$ и $\bar{\theta}(\vec{X})$.

Статистика $T(\vec{X}, \theta)$ называется *центральной*, если закон ее распределения не зависит от θ.

1.2 Формулы

Будем предполагать, что

1) X — случайная величина, закон распределения которой зависит от неизвестного параметра θ;

2) $\gamma \in (0; 1)$ — выбранный уровень доверия;

3) $T(\vec{X}, \theta)$ — центральная статистика;

4) T является непрерывной;

5) функция $F_T(t)$ распределения статистики T является монотонно возрастающей на множестве $\{t \in \mathbb{R} :\ 0 < F_T(t) < 1\}$;

6) T является монотонно возрастающей функцией параметра θ;

Выберем значения α_1, $\alpha_2 \in \mathbb{R}$ так, чтобы $\alpha_1 + \alpha_2 = 1 - \gamma$, при этом $\alpha_1 \geq 0$, $\alpha_2 \geq 0$. Пусть q_{α_1}, $q_{1-\alpha_2}$ — квантили соответствующих уровней случайной величины T. Тогда в соответствии с предположениями 1–6 получаем:

$$\gamma = P\{\, T^{-1}(\vec{X}, q_{\alpha_1}) < \theta < T^{-1}(\vec{X}, q_{1-\alpha_2})\,\}. \tag{1.3}$$

Пусть $X \sim N(\theta, \sigma^2)$, где θ — неизв., σ^2 — изв.

Тогда для построения γ–доверительного интервала для математического ожидания θ используется следующая центральная статистика и ее закон распределения:

$$T(\vec{X}, \theta) = \frac{\theta - \bar{X}}{\sigma}\sqrt{n} \sim N(0, 1). \tag{1.4}$$

В результате построения γ–доверительного интервала для математического ожидания θ получаются формулы для вычисления границ этого интервала ($U_{\alpha_1}, U_{1-\alpha_2}$ — квантили соответствующих уровней):

$$\underline{\theta}(\vec{X}) = \bar{X} + \frac{\sigma U_{\alpha_1}}{\sqrt{n}}, \tag{1.5}$$

$$\bar{\theta}(\vec{X}) = \bar{X} + \frac{\sigma U_{1-\alpha_2}}{\sqrt{n}}. \tag{1.6}$$

При $\alpha_1 = \alpha_2$ формулы 1.5, 1.6 принимают вид:

$$\underline{\theta}(\vec{X}) = \bar{X} - \frac{\sigma U_{\frac{1+\gamma}{2}}}{\sqrt{n}}, \tag{1.7}$$

$$\bar{\theta}(\vec{X}) = \bar{X} + \frac{\sigma U_{\frac{1+\gamma}{2}}}{\sqrt{n}}. \tag{1.8}$$

В случае, если $X \sim N(\theta, \sigma^2)$, θ — неизв., σ^2 — неизв., для построения γ–доверительного интервала для математического ожидания θ используется следующая центральная статистика и ее закон распределения:

$$T(\vec{X}, \theta) = \frac{\theta - \bar{X}}{S(\vec{X})}\sqrt{n} \sim St(n-1), \tag{1.9}$$

где $S^2(\vec{X}) = \frac{1}{n-1}\sum\limits_{i-1}^{n}(X_i - \bar{X})^2$ — исправленная выборочная дисперсия.

Тогда по аналогии с предыдущим случаем получаем в результате построения γ–доверительного интервала для математического ожидания θ формулы для вычисления границ этого интервала ($t_{\alpha_1}, t_{1-\alpha_2}$ — квантили соответствующих уровней):

$$\underline{\theta}(\vec{X}) = \bar{X} + \frac{S(\vec{X})t_{\alpha_1}}{\sqrt{n}}, \tag{1.10}$$

$$\bar{\theta}(\vec{X}) = \bar{X} + \frac{S(\vec{X})t_{1-\alpha_2}}{\sqrt{n}}. \tag{1.11}$$

При $\alpha_1 = \alpha_2$ формулы 1.10, 1.11 принимают вид:

$$\underline{\theta}(\vec{X}) = \bar{X} - \frac{S(\vec{X})t_{\frac{1+\gamma}{2}}}{\sqrt{n}}, \tag{1.12}$$

$$\bar{\theta}(\vec{X}) = \bar{X} + \frac{S(\vec{X})t_{\frac{1+\gamma}{2}}}{\sqrt{n}}. \tag{1.13}$$

В случае, если $X \sim N(m, \theta)$, m — неизв., θ — неизв., для построения γ–доверительного интервала для дисперсии θ используется следующая центральная статистика и ее закон распределения:

$$T(\vec{X}, \theta) = \frac{S^2(\vec{X})}{\theta})(n-1) \sim \chi^2(n-1). \tag{1.14}$$

По аналогии с предыдущим случаем получаем в результате построения γ–доверительного интервала для дисперсии σ^2 формулы для вычисления границ этого интервала ($h_{\alpha_1}, h_{1-\alpha_2}$ — квантили соответствующих уровней):

$$\underline{\theta}(\vec{X}) = \frac{(n-1)S^2(\vec{X})}{h_{1-\alpha_2}}, \tag{1.15}$$

$$\bar{\theta}(\vec{X}) = \frac{(n-1)S^2(\vec{X})}{h_{\alpha_1}}. \tag{1.16}$$

При $\alpha_1 = \alpha_2: \quad h_{\alpha_1} = h_{\frac{1-\gamma}{2}},\ h_{1-\alpha_2} = h_{\frac{1+\gamma}{2}}$.

2 Текст программы

На листинге 2.1 представлена реализованная в ходе лабораторной работы программа (на языке Octave).

Листинг 2.1 – Исходный текст программы

```
X_ = csvread("./data/x.csv");
X_ = sort(X_);
n_max = length(X_);

n = n_max;

begin = 1;

X = X_(1:n);
mu = sum(X) / n;

if (n > 1)
    s_square = sum((X - mu).^2) / (n - 1);
else
    s_square = 0;
endif

printf("Введите уровень доверия (вещественное от 0 до 1 включительно):\n\t");
gamma = str2double(input("gamma = ", "s"));

printf("\nТочечные оценки MX и DX:\n");
printf("\tmu       = %+.4f\n\tS_square = %+.4f\n", mu, s_square);

s = sqrt(s_square);

t = tinv((1 + gamma) / 2, n - 1);
n_sqrt = sqrt(n);

mu_low = mu - s * t / n_sqrt;
mu_high = mu + s * t / n_sqrt;

printf("\nНижняя и верхняя границы для gamma-доверительного интервала для\
        \nматематического ожидания MX:\n");
```

```
printf("\tmu_low  = %+.4f\n\tmu_high = %+.4f\n", mu_low, mu_high);

h_low = chi2inv((1 + gamma) / 2, n - 1);
h_high = chi2inv((1 - gamma) / 2, n - 1);
sigma_low = (n - 1) * s_square / h_low;
sigma_high = (n - 1) * s_square / h_high;

printf("\nНижняя и верхняя границы для gamma-доверительного интервала для\
        \nматематического ожидания DX:\n");
printf("\tsigma_low  = %+.4f\n\tsigma_high = %+.4f\n", sigma_low, sigma_high);

mu_n = zeros(n - begin + 1, 1) + mu;
mu_N = zeros(n - begin + 1, 1);
mu_low_N = zeros(n - begin + 1, 1);
mu_high_N = zeros(n - begin + 1, 1);

for n_i = begin:n
    t_i = tinv((1 + gamma) / 2, n_i - 1);
    mu_N(n_i - begin + 1) = sum(X(1:n_i)) / n_i;
    mu_low_N(n_i - begin + 1) = mu_N(n_i - begin + 1) - s * t_i / sqrt(n_i);
    mu_high_N(n_i - begin + 1) = mu_N(n_i - begin + 1) + s * t_i / sqrt(n_i);
endfor;

figure;
p1 = plot((begin:n), mu_n, "Color", "black", "linestyle", "-", "linewidth", 2);
hold on;
p2 = plot((begin:n), mu_N, "Color", "magenta", "linestyle", "--", "linewidth", 2);
p3 = plot((begin:n), mu_low_N, "Color", "red", "linestyle", "-.", "linewidth", 2);
p4 = plot((begin:n), mu_high_N, "Color", "blue", "linestyle", ":", "linewidth", 2);
hold off;
xlabel("Объем выборки n");
grid on;
```

```
s_square_n = zeros(n - begin + 1, 1) + s_square;
s_square_N = zeros(n - begin + 1, 1);
sigma_low_N = zeros(n - begin + 1, 1);
sigma_high_N = zeros(n - begin + 1, 1);

for n_i = begin:n
    s_square_N(n_i - begin + 1) = sum((X(1:n_i) - mu).^2) / (n_i
        - 1);
    h_low_i = chi2inv((1 + gamma) / 2, n_i - 1);
    h_high_i = chi2inv((1 - gamma) / 2, n_i - 1);
    sigma_low_N(n_i - begin + 1) = (n_i - 1) * s_square_N(n_i -
        begin + 1) / h_low_i;
    sigma_high_N(n_i - begin + 1) = (n_i - 1) * s_square_N(n_i -
        begin + 1) / h_high_i;
endfor;

figure;
p1 = plot((begin:n), s_square_n, "Color", "black", "linestyle",
    "-", "linewidth", 2);
hold on;
p2 = plot((begin:n), s_square_N, "Color", "magenta",
    "linestyle", "--", "linewidth", 2);
p3 = plot((begin:n), sigma_low_N, "Color", "red", "linestyle",
    "-.", "linewidth", 2);
p4 = plot((begin:n), sigma_high_N, "Color", "blue", "linestyle",
    ":", "linewidth", 2);
hold off;
xlabel("Объем выборки n");
grid on;
```

3 Результаты расчетов для выборки из индивидуального варианта

На листинге 3.1 представлены результаты расчетов для выборки из индивидуального варианта №1.

Листинг 3.1 – Результаты расчетов для выборки из индивидуального варианта

```
Введите уровень доверия (вещественное от 0 до 1 включительно):
        gamma = 0.9

Точечные оценки MX и DX:
        mu       = -1.6046
        S_square = +1.0341

Нижняя и верхняя границы для gamma-доверительного интервала для
математического ожидания MX:
        mu_low  = -1.7585
        mu_high = -1.4507

Нижняя и верхняя границы для gamma-доверительного интервала для
математического ожидания DX:
        sigma_low  = +0.8460
        sigma_high = +1.2979
```

На рисунках 3.1, 3.2, 3.3, 3.4 представлены графики для выборки из индивидуального варианта (при построении графиков $\gamma = 0.9$).

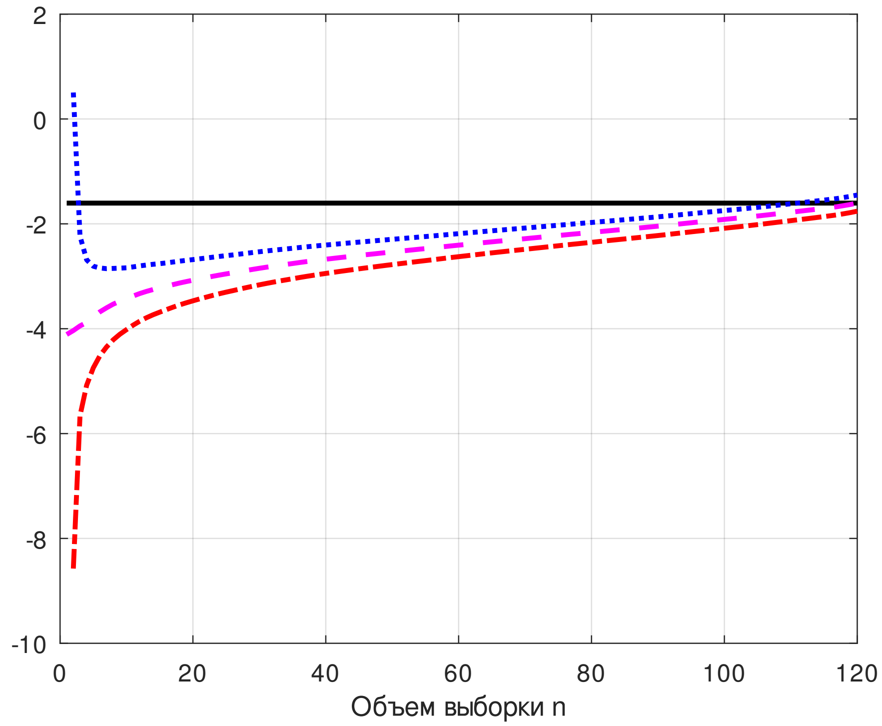

Рисунок 3.1 — Прямая $y = \hat{\mu}(\vec{x}_N)$ и графики функций $y = \hat{\mu}(\vec{x}_n)$, $y = \underline{\mu}(\vec{x}_n)$ и $y = \overline{\mu}(\vec{x}_n)$ как функций объема n выборки, где n изменяется от 1 до N

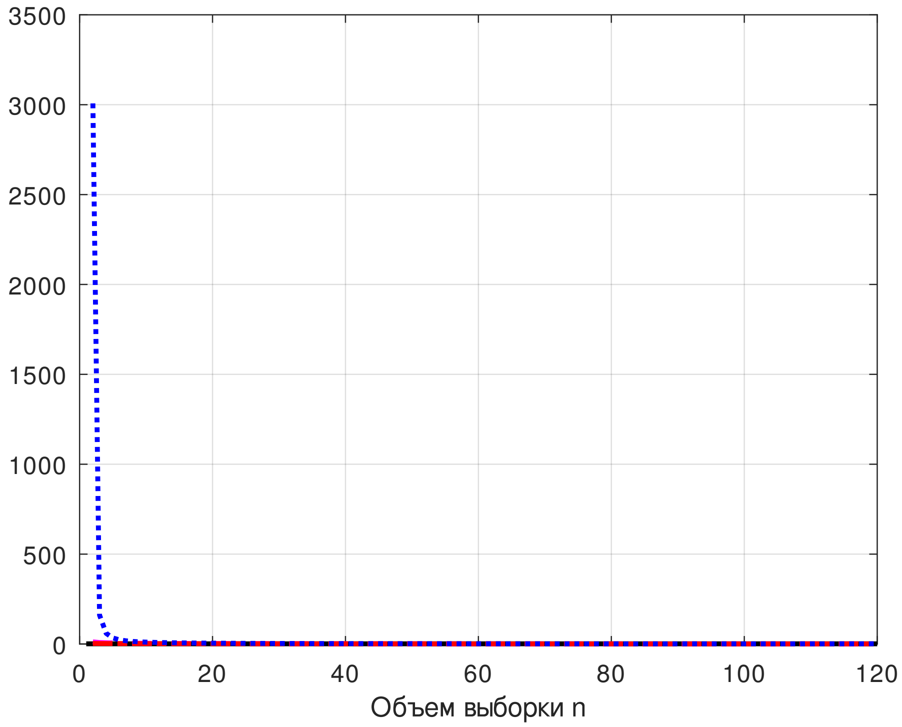

Рисунок 3.2 — Прямая $z = S^2(\vec{x}_N)$ и графики функций $z = S^2(\vec{x}_n)$, $z = \underline{\sigma}^2(\vec{x}_n)$ и $z = \overline{\sigma}^2(\vec{x}_n)$ как функций объема n выборки, где n изменяется от 1 до N

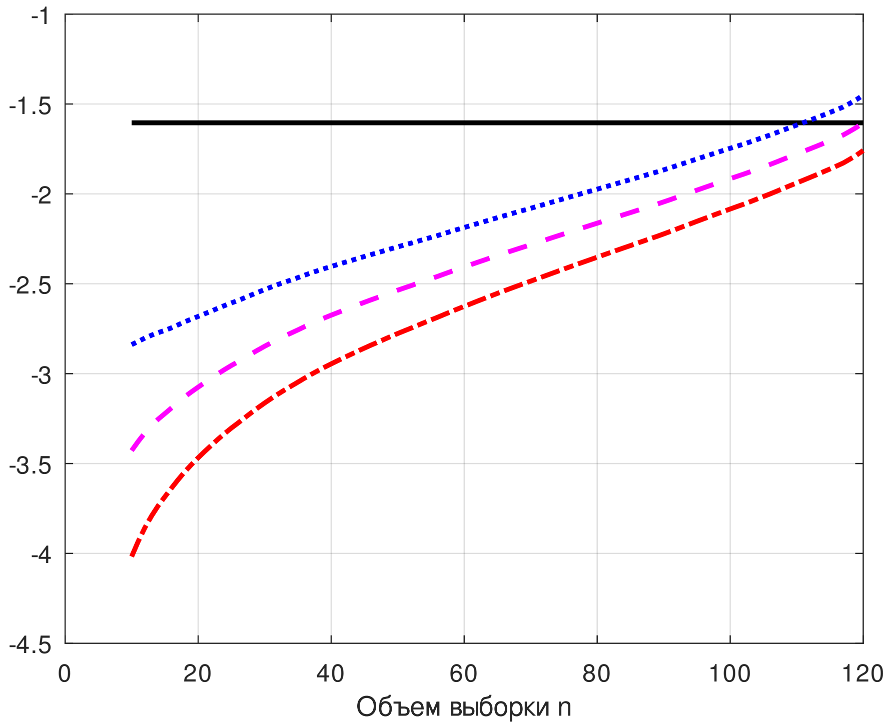

Рисунок 3.3 — Прямая $y = \hat{\mu}(\vec{x}_N)$ и графики функций $y = \hat{\mu}(\vec{x}_n)$, $y = \underline{\mu}(\vec{x}_n)$ и $y = \overline{\mu}(\vec{x}_n)$ как функций объема n выборки, где n изменяется от 10 до N

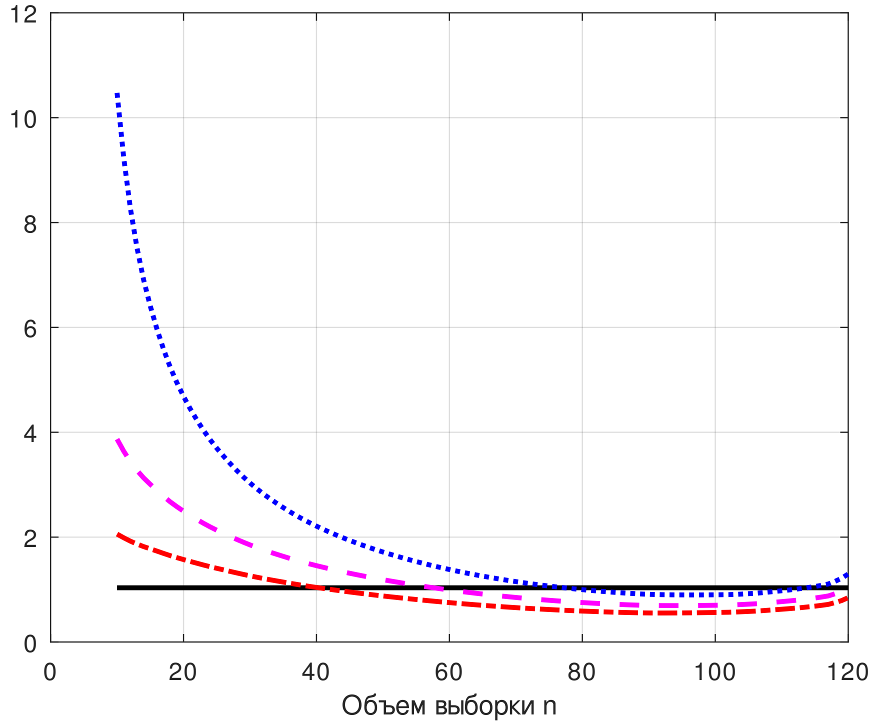

Рисунок 3.4 — Прямая $z = S^2(\vec{x}_N)$ и графики функций $z = S^2(\vec{x}_n)$, $z = \underline{\sigma}^2(\vec{x}_n)$ и $z = \overline{\sigma}^2(\vec{x}_n)$ как функций объема n выборки, где n изменяется от 10 до N