Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное 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 №1

по курсу «Математическая статистика»

на тему: «Гистограмма и эмпирическая функция распределения»

Студент группы ИУ7-63Б ____________ Паламарчук А. Н.
(Подпись, дата) (Фамилия И.О.)

Преподаватель ____________ Саркисян П. С.
(Подпись, дата) (Фамилия И.О.)

Москва — 2025 г.

Содержание

1 Задание

1.1 Цель работы

Цель работы: построение гистограммы и эмпирической функции распределения.

1.2 Содержание работы

1) Для выборки объёма n из генеральной совокупности X реализовать в виде программы на ЭВМ
    1) вычисление максимального значения $M_{\max}$ и минимального значения $M_{\min}$;
    2) размаха R выборки;
    3) вычисление оценок $\hat{\mu}$ и S^2 математического ожидания MX и дисперсии DX;
    4) группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
    5) построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
    6) построение на другой координатной плоскости графика эмпирической функции распределения и функции распределения нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2.
2) Провести вычисления и построить графики для выборки из индивидуального варианта.

2 Теоретическая часть

2.1 Формулы для вычисления величин M_{max} , M_{min}, R, $\hat{\mu}$, S^2

Пусть $\vec{x} = (x_1, \ldots, x_n)$ — выборка из генеральной совокупности X, где n — объём данной выборки. Расположим компоненты x_i, $i = \overline{1, n}$ в порядке неубывания:

$$x_{(1)} \ \leq \ x_{(2)} \ \leq \ \ldots \ \leq \ x_{(n)} \tag{2.1}$$

Вариационным рядом, построенным по выборке $\vec{x} = (x_1, \ldots, x_n)$, называют вектор $(x_{(1)}, \ldots, x_{(n)})$.

Минимальное значение выборки рассчитывается по формуле (2.2); максимальное — (2.3). Размах выборки рассчитывается по формуле (2.4); выборочное среднее — (2.5), исправленная выборочная дисперсия — (2.6).

$$M_{\min} = x_{(1)} = \underset{x_i \in \vec{x}}{min}\ x_i \tag{2.2}$$

$$M_{\max} = x_{(n)} = \underset{x_i \in \vec{x}}{max}\ x_i \tag{2.3}$$

$$R = M_{\max} - M_{\min}. \tag{2.4}$$

$$\hat{\mu}(\vec{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.5}$$

$$S^2(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \tag{2.6}$$

2.2 Эмпирическая плотность и гистограмма

Пусть $\vec{x} = (x_1, \ldots, x_n)$ — реализация выборки из генеральной совокупности X, где n — объём данной выборки.

При большом объеме n выборки значения x_i группируют в интервальный статистический ряд. Для этого отрезок $J = [x_{(1)}, x_{(n)}]$ делят на m равновеликих промежутков по формуле (2.7):

$$J_i = [x_{(1)} + (i-1) \cdot \Delta,\ x_{(1)} + i \cdot \Delta), i = \overline{1; m-1} \tag{2.7}$$

Последний промежуток определяется по формуле (2.8):

$$J_m = [x_{(1)} + (m-1) \cdot \Delta, x_{(n)}] \tag{2.8}$$

Ширина каждого из таких промежутков определяется по формуле (2.9).

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m} \tag{2.9}$$

Интервальным статистическим рядом, отвечающим выборке $\vec{x}$, называют таблицу 2.1, в которой n_i — число элементов выборки, попавших в J_i, $i = \overline{1, m}$

Таблица 2.1 – Интервальный статистический ряд

J_1	...	J_i	...	J_m
n_1	...	n_i	...	n_m

Пусть для выборки $\vec{x} = (x_1, \ldots, x_n)$ построен интервальный статистический ряд $(J_i,\ n_j)$, $i = \overline{1, m}$.

Эмпирической плотностью, отвечающей выборке $\vec{x}$, называют функцию:

$$f_n(x) = \begin{cases} \frac{n_i}{n\Delta}, x \in J_i, i = \overline{1; m} \\ 0 \quad, x \notin J \end{cases} \tag{2.10}$$

где J_i — полуинтервал статистического ряда, n_i — количество элементов выборки, входящих в полуинтервал.

Гистограмма — это график эмпирической функции плотности.

2.3 Эмпирическая функция распределения

Пусть $\vec{x} = (x_1, \ldots, x_n)$ — выборка из генеральной совокупности X, где n — объём данной выборки. Обозначим $l(t, \vec{x})$ — число элементов выборки $\vec{x}$, которые имеют значения меньше t.

Эмпирической функцией распределения, отвечающей выборке $\vec{x}$, называют отображение $F_n : R \to R$, определенное правилом:

$$F_n(t) = \frac{l(t, \vec{x})}{n} \tag{2.11}$$

3 Практическая часть

3.1 Результаты расчетов

Индивидуальный вариант №14

Результаты расчетов для выборки приведены на формулах (3.1), (3.2), (3.3), (3.4), (3.5), (3.6).

$$M_{\min} = 0.09 \quad (3.1)$$

$$M_{\max} = 5.47 \quad (3.2)$$

$$R = 5.38 \quad (3.3)$$

$$\hat{\mu}(\vec{x}) = 3.055 \quad (3.4)$$

$$S^2(\vec{x}) = 1.055 \quad (3.5)$$

$$m = 8 \quad (3.6)$$

На рисунке 3.1 представлены гистограмма и график функции плотности распределения вероятностей нормальной случайной величины с выборочным математическим ожиданием $\hat{\mu}$ и выборочной дисперсией S^2.

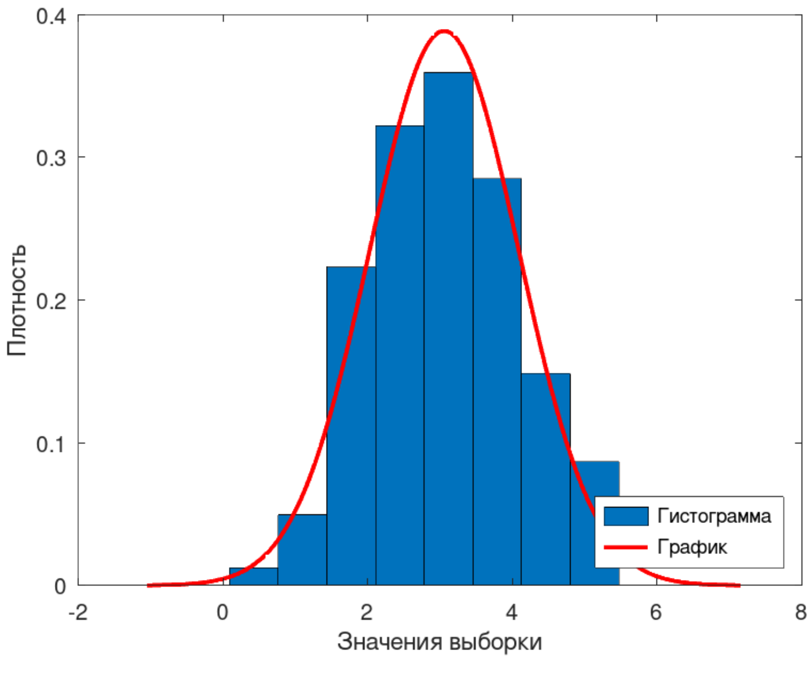

Рисунок 3.1 – Гистограмма и график функции плотности распределения вероятностей.

На рисунке 3.2 представлены график эмпирической функции распределения и функции распределения нормальной случайной величины с выборочным математическим ожиданием $\hat{\mu}$ и выборочной дисперсией S^2.

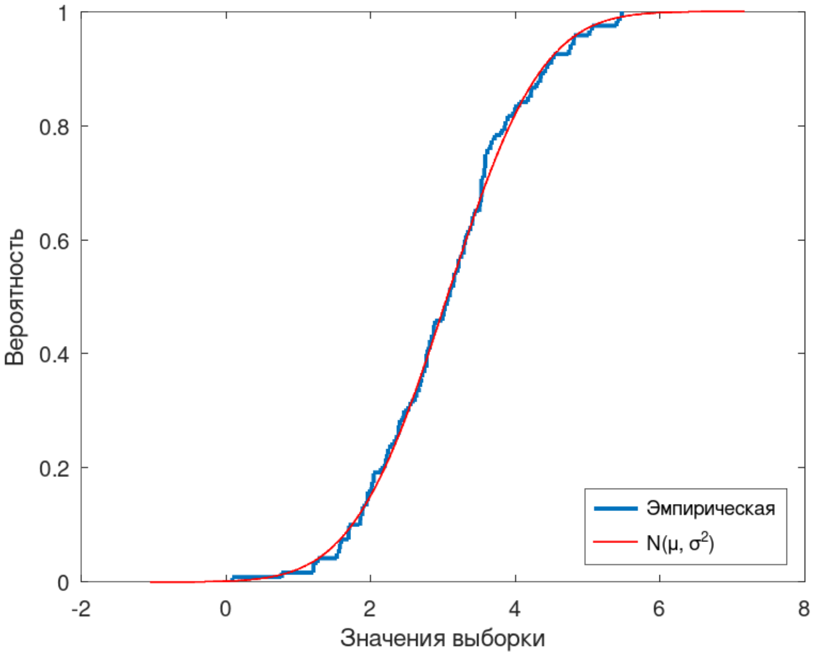

Рисунок 3.2 – Графики эмпирической функции распределения и функции распределения нормальной случайной величины.