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 работа №1

по курсу «Математическая статистика»

на тему: «Гистограмма и эмпирическая функция распределения»

Студент	ИУ7-63Б (Группа)	(Подпись, дата)	Лысцев Н. Д. (И. О. Фамилия)
Преподаватель		(Подпись, дата)	Власов П. А. (И. О. Фамилия)

2024 г.

СОДЕРЖАНИЕ

1 Задание

Цель работы: построение гистограммы и эмпирической функции распределения.

Содержание работы

1. Для выборки объема n из генеральной совокупности X реализовать в виде программы на ЭВМ
   а) вычисление максимального значения M_{max} и минимального значения M_{min};
   б) размаха R выборки;
   в) вычисление оценок $\hat{\mu}$ и S^2 математического ожидания $\mathrm{M}X$ и дисперсии $\mathrm{D}X$;
   г) группировку значений выборки в $m = [\log_2 n] + 2$ интервала;
   д) построение на одной координатной плоскости гистограммы и графика функции плотности распределения вероятностей нормальной случайной величины с математическим ожиданием $\hat{\mu}$ и дисперсией S^2;
   е) 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 Теоретические сведения

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

Пусть $\vec{x} = (x_1, ..., x_n)$ — выборка объема n из генеральной совокупности X.

- M_{min} — минимальное значение выборки $\vec{x}$, определяется по формуле (2.1).

$$M_{min} = x_{(1)} = \min\{x_1, ..., x_n\} \tag{2.1}$$

где $x_{(1)}$ — крайний левый член вариационного ряда выборки $\vec{x}$.

- M_{max} — максимальное значение выборки $\vec{x}$, определяется по формуле (2.2).

$$M_{max} = x_{(n)} = \max\{x_1, ..., x_n\} \tag{2.2}$$

где $x_{(n)}$ — крайний правый член вариационного ряда выборки $\vec{x}$.

- R — размах выборки, определяется по формуле (2.3).

$$R = M_{max} - M_{min} \tag{2.3}$$

- $\hat{\mu}$ — оценка математического ожидания(выборочное среднее), определяется по формуле (2.4).

$$\hat{\mu}(\vec{x}) = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{2.4}$$

- S^2 — оценка дисперсии(исправленная выборочная дисперсия), определяется по формуле (2.5).

$$S^2(\vec{x}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2, \tag{2.5}$$

2.2 Определение эмпирической плотности и гистограммы

Если объем выборки велик ($n > 50$), то данные группируют не только в виде статистического ряда, но и в виде интервального статистического ряда. Для этого выбирается число $m \in \mathbb{N}$ — количество интервалов, а отрезок $J = [x_{(1)}, x_{(n)}]$ разбивают на m равновеликих промежутков. Длина Δ каждого из них определяется по формуле (2.6).

$$\Delta = \frac{|J|}{m} = \frac{x_{(n)} - x_{(1)}}{m}. \tag{2.6}$$

Интервалы определяются равенствами (2.7).

$$\begin{aligned} J_i &= [x_{(1)} + (i-1)\Delta; x_{(1)} + i\Delta], \ \ i = \overline{1, m-1}, \\ J_m &= [x_{(1)} + (m-1)\Delta; x_{(n)}]. \end{aligned} \tag{2.7}$$

Опр. Интервальным статистическим рядом, отвечающим выборке $\vec{x}$, называется таблица вида:

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

Здесь n_i — число элементов выборки $\vec{x}$, попавших в промежуток J_i, $i = \overline{1, m}$.

При выборе числа промежутков используют формулу (2.8).

$$m = [\log_2 n] + 2. \tag{2.8}$$

Пусть для данной выборки $\vec{x}$ построен интервальный статистический ряд.

Опр. Эмпирической плотностью распределения соответствующей выборке $\vec{x}$ называется функция:

$$f_n(x) = \begin{cases} \frac{n_i}{n \cdot \Delta}, & x \in J_i, \ i = \overline{1, m}, \\ 0, & \text{иначе}. \end{cases} \tag{2.9}$$

Опр. График эмпирической функции плотности называется гистограммой.

2.3 Определение эмпирической функции распределения

Пусть $\vec{x} = (x_1, ..., x_n)$ — выборка из генеральной совокупности X.

Обозначим $l(t, \vec{x})$ — число компонент $\vec{x}$, которые меньше, чем t ($t \in \mathbb{R}$).

Опр. Эмпирической функцией распределения, отвечающей выборке $\vec{x}$, называется отображение $F_n : \mathbb{R} \to \mathbb{R}$, заданное формулой:

$$F_n(t) = \frac{l(t, \vec{x})}{n}. \tag{2.10}$$

3 Текст программы