|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования Российской Федерации**  Калужский филиал  федерального государственного бюджетного  образовательного учреждения высшего образования  ***«Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»***  ***(КФ МГТУ им. Н.Э. Баумана)*** |

**ФАКУЛЬТЕТ** ***ИУК «Информатика и управление»***

**КАФЕДРА** \_\_***ИУК4 «Программное обеспечение ЭВМ, информационные технологии»***

**ЛАБОРАТОРНАЯ РАБОТА №2**

**«Систематизация, графическое представление статистических данных, выборочные числовые характеристики на основе большой выборки»**

**ДИСЦИПЛИНА: «Методы обработки информации»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. ИУК4-72Б | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Карельский М.К. )  (Подпись) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Никитенко У.В. )  (Подпись) |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | |
|  | - Балльная оценка:  - Оценка: | |

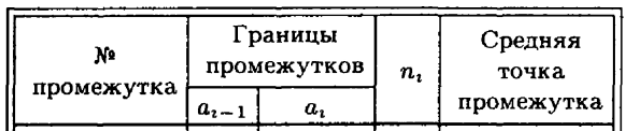
Калуга, 2023

**Цель:** овладение приемами первичной обработки большой выборки. Выдвижение гипотезы о законе распределения генеральной совокупности.

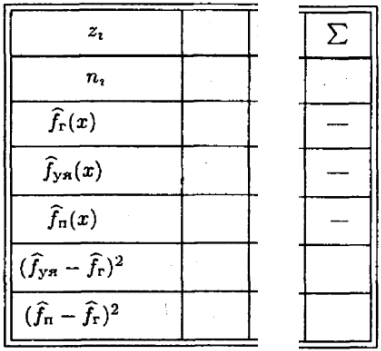
**Задание:**

Для обработки преподавателем выдается случайных чисел. Эти числа хранятся в файле TestNN.csv.

1. Выборка подвергается обработке и оформляется в виде таблицы.



1. Графические характеристики выборки – строим гистограмму и полигон приведенных частот. Выдвигаем гипотезу о виде плотности вероятности генерального распределения.
2. Находим выборочные характеристики положения и рассеивания.
3. Для сравнения с гистограммой и полигоном приведенных частот на одном чертеже постройте графики гистограммной оценки плотности вероятности , параметрической oцeнки плотности вероятности , и усредненную ядерную оценку плотности вероятности .
4. Значения оценок плотности вероятности в средних точках промежутков группированного статистического ряда оформите в виде таблицы.



1. Проанализируйте близость оценок по средним квадратическим отклонениям и от .

**Вариант 7**

**Листинг:**

import numpy as np

import matplotlib.pyplot as plt

from scipy.stats import norm, gaussian\_kde

import pandas as pd

from prettytable import PrettyTable

data = pd.read\_csv('Test7.csv', header=None)

values = data.iloc[:, 0]

num\_bins = 8

plt.figure(figsize=(10, 5))

plt.subplot(1, 2, 1)

plt.hist(values, bins=num\_bins, density=True, alpha=0.6, label='Гистограмма')

hist, bin\_edges = np.histogram(values, bins=num\_bins, density=True, range=(values.min(), values.max()))

bin\_centers = (bin\_edges[:-1] + bin\_edges[1:]) / 2

plt.plot(bin\_centers, hist, marker='o', linestyle='-', label='Полигон')

plt.title("Гистограмма и полигон частот")

plt.legend()

edges = [(round(bin\_edges[i], 2), round(bin\_edges[i + 1], 2)) for i in range(num\_bins)]

table = PrettyTable()

table.add\_column("#", [i for i in range(1, num\_bins + 1)])

table.add\_column("Промежуток", edges)

table.add\_column("Максимальное значение в этой точке", list(map(lambda value: round(value, 4), hist)))

table.add\_column("Центр промежутка", list(map(lambda center: round(center, 2), bin\_centers)))

print(table)

plt.subplot(1, 2, 2)

x = np.linspace(values.min(), values.max(), 100)

plt.plot(x, norm.pdf(x, loc=values.mean(), scale=values.std()), label='Параметрическая оценка')

kde = gaussian\_kde(values)

plt.plot(x, kde(x), label='Усредненная ядерная оценка')

hist\_estimate = hist / (bin\_centers[1] - bin\_centers[0])

plt.plot(bin\_centers, hist\_estimate, label='Гистограммная оценка')

plt.legend()

plt.title('Графики оценок плотности вероятности')

mean\_value = np.mean(values)

median\_value = np.median(values)

variance = np.var(values)

std\_deviation = np.std(values)

print(f"Выборочное среднее: {mean\_value:.4f}")

print(f"Выборочная медиана: {median\_value:.4f}")

print(f"Выборочная дисперсия: {variance:.4f}")

print(f"Выборочное стандартное отклонение: {std\_deviation}")

parametric\_estimate = norm.pdf(bin\_centers, loc=values.mean(), scale=values.std())

kde\_estimate = kde(bin\_centers)

table = PrettyTable()

table.add\_column("№", [i for i in range(1, num\_bins + 1)])

table.add\_column("Центр промежутков", bin\_centers)

table.add\_column("Гистограммная оценка плотности вероятности", hist\_estimate)

table.add\_column("Усредненную ядерную оценку плотности вероятности", kde\_estimate)

table.add\_column("Параметрическая oцeнка плотности вероятности", parametric\_estimate)

print("Значения оценок в средних точках:")

print(table)

table = PrettyTable()

table.add\_column("№", [i for i in range(1, num\_bins + 1)])

table.add\_column("Центр промежутков", bin\_centers)

mse\_parametric = (parametric\_estimate - hist\_estimate) \*\* 2

mse\_kde = (kde\_estimate - hist\_estimate) \*\* 2

table.add\_column("(Усредненная яд. оц. - Гистограммная оц.)^2", mse\_parametric)

table.add\_column("(Параметрическая оц. - Гистограммная оц.)^2", mse\_kde)

table.add\_row(['', '', 'sum', 'sum'])

table.add\_row(['', '', sum(mse\_parametric), sum(mse\_kde)])

print(table)

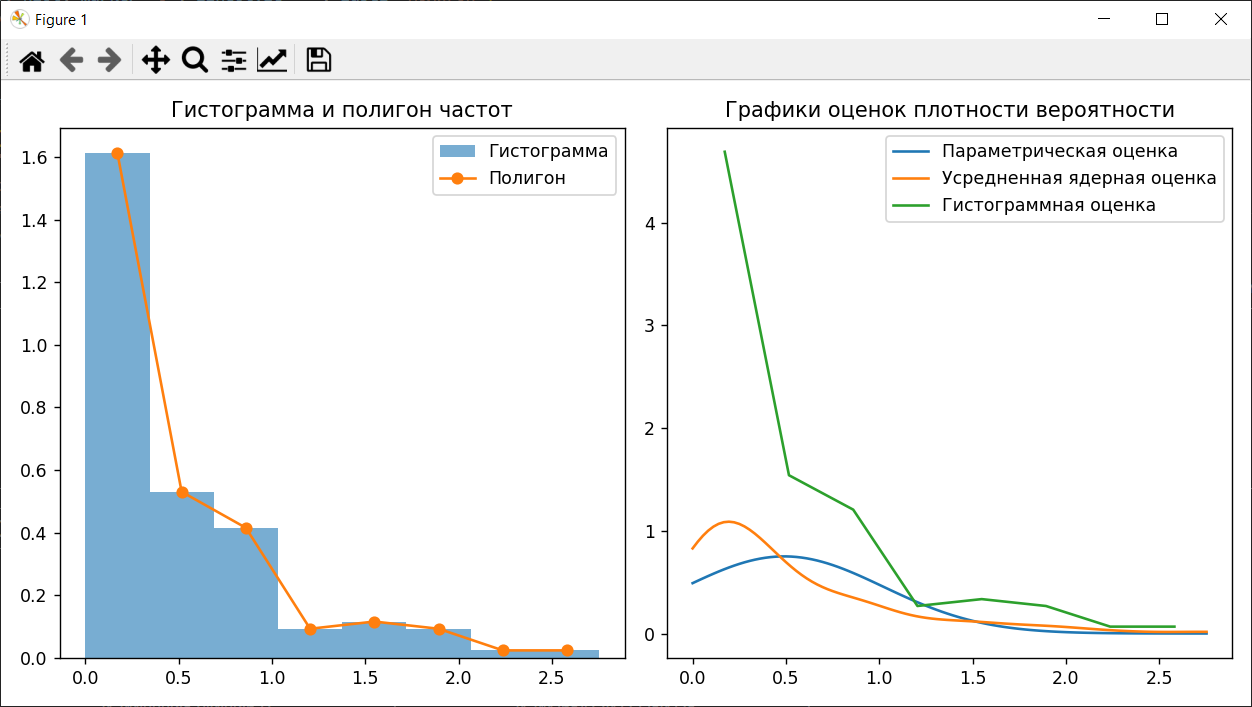
print(f"Среднее квадратичное отклонение параметрической оценки от гистограммной оценки: {np.sqrt(np.mean(mse\_parametric)):.4f}")

print(f"Среднее квадратичное отклонение усредненной ядерной оценки от гистограммной оценки: {np.sqrt(np.mean(mse\_kde)):.4f}")

plt.tight\_layout()

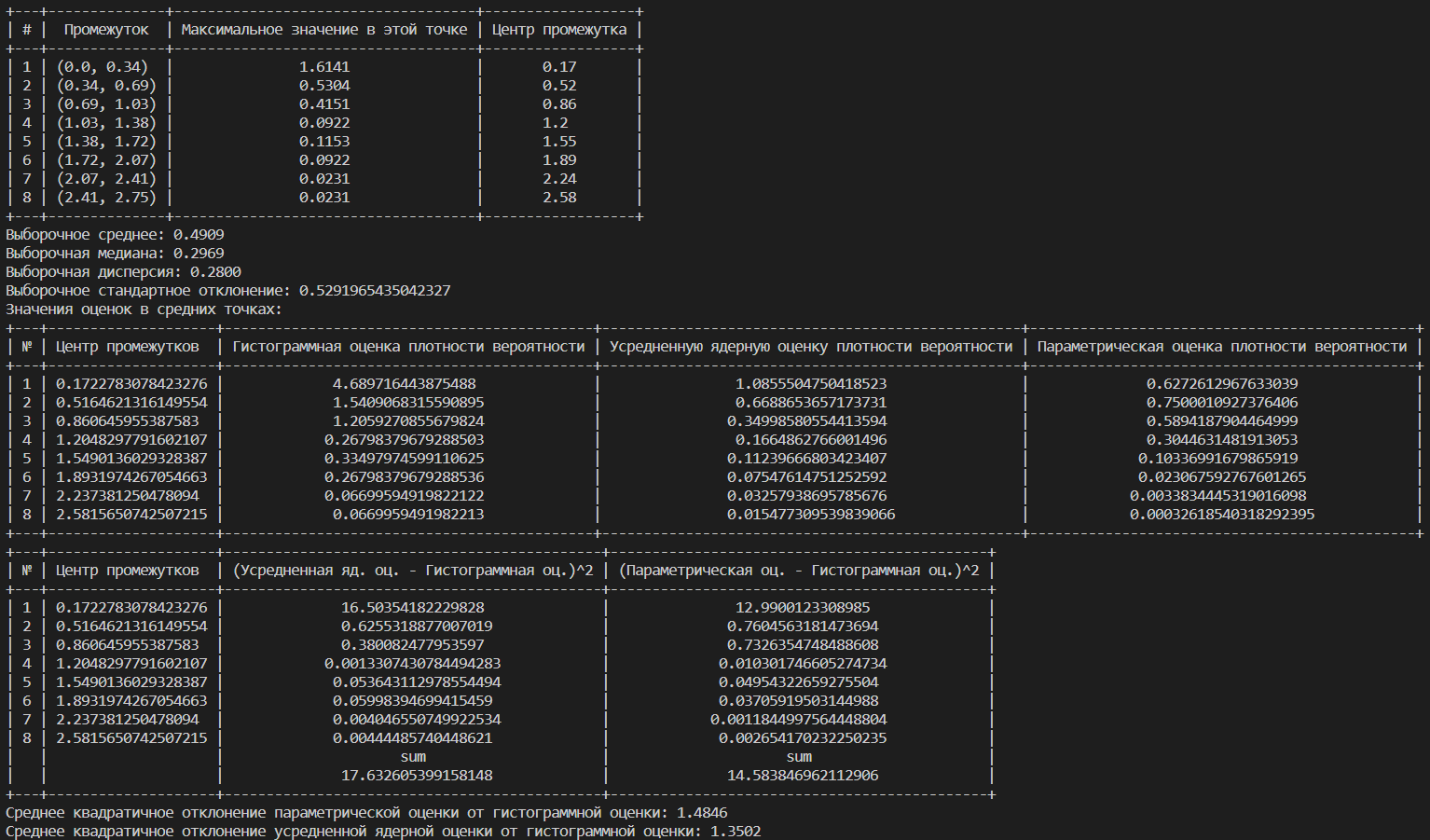
plt.show()

**Результат:**



**Рис. 1.** Графики

Исходя из графиков, можно сделать предположение о геометрическом распределении подборки.



**Рис. 2.** Вычисления

**Вывод:** в ходе выполнения лабораторной работы были получены практические навыки первичной обработки большой выборки.