ГУАП

КАФЕДРА № 43

ОТЧЕТ   
ЗАЩИЩЕН С ОЦЕНКОЙ

ПРЕПОДАВАТЕЛЬ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ассистент |  |  |  | К. А. Кочин |
| должность, уч. степень, звание |  | подпись, дата |  | инициалы, фамилия |

|  |
| --- |
| ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ |
| «Статистическое оценивание параметров выборки» |
| по курсу: Прикладная теория вероятностей и статистика |
|  |
|  |

РАБОТУ ВЫПОЛНИЛ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СТУДЕНТ гр. № | 4332 |  |  |  | А. А. Лютов |
|  |  |  | подпись, дата |  | инициалы, фамилия |

Санкт-Петербург 2025

**Задание:**

1) Построить вариационные ряды :

1) , объёмом 15, 30, 100, 1000;

2) , объёмом 15, 30, 100, 1000;

3) , объёмом 15, 30, 100, 1000;

4) , объёмом 15, 30, 100, 1000;

2) Вычислить для каждого ряда оценку исходного параметра:

1) Выборочное значение;

2) Полусумма максимума и минимума;

3) Выборочная медиана;

4) Среднее арифметическое с отбросом с обеих сторон вариационного ряда по k членов;

3) Провести эксперимент 50 раз для каждой модели шума и объёма выборки, и построить графики с отметками оценок в виде точек, где ось Y – значение оценок, X – объём выборки.

4) Для каждой комбинации из алгоритма оценки, модули шума и объёма выборки рассчитать среднее значение относительной ошибки и заполнить таблицу.

**Расчёты и графики:**

Для генерации вариационных рядов и оценки этих рядов я написал программу на языке Python, используя библиотеку numpy.

**Lab2.py**

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

# Параметр c

c = 13

def Mean(z):

    return np.mean(z)

def HalfSum(z):

    return (np.min(z) + np.max(z)) / 2

def Median(z):

    return np.median(z)

def TrimmeredMean(z):

    k = 1 if len(z) < 10 else (2 if len(z) < 15 else 3)

    return np.mean(sorted(z)[k:-k])

algorithms = {

    "Выборочное среднее": Mean,

    "Полусумма макс и мин": HalfSum,

    "Медиана": Median,

    "Среднее с отбросом": TrimmeredMean,

}

distributions = [

    ("z=c+N(0,c/100)", lambda n: c + np.random.normal(0, c/100, n)),

    ("z=c+N(0,c/20)", lambda n: c + np.random.normal(0, c/20, n)),

    ("z=c+U(-c/100,c/100)", lambda n: c + np.random.uniform(-c/100, c/100, n)),

    ("z=c+U(-c/20,c/20)", lambda n: c + np.random.uniform(-c/20, c/20, n)),

]

sampleSize = [15, 30, 100, 1000]

experiments = 50

for distName, distFunc in distributions:

    fig, axes = plt.subplots(1, 4, figsize=(16, 4), sharey=True)

    fig.suptitle(f"Model: {distName}")

    for ax, (alg\_name, alg\_func) in zip(axes, algorithms.items()):

        for n in sampleSize:

            estimates = [alg\_func(distFunc(n)) for \_ in range(experiments)]

            ax.scatter([n] \* experiments, estimates, label=f"n={n}", s=5)

        ax.axhline(y=c, color='blue', linestyle='--')

        ax.set\_xscale('log')

        ax.set\_xlabel("Sampling size")

        ax.set\_title(alg\_name)

        ax.legend()

    plt.show()

results = []

for distName, distFunc in distributions:

    for n in sampleSize:

        errorMean = []

        for alg\_name, alg\_func in algorithms.items():

            errorList = []

            for \_ in range(experiments):

                z = distFunc(n)

                x\_hat = alg\_func(z)

                error = abs((c - x\_hat) / c)

                errorList.append(error)

            errorMean.append(np.mean(errorList))

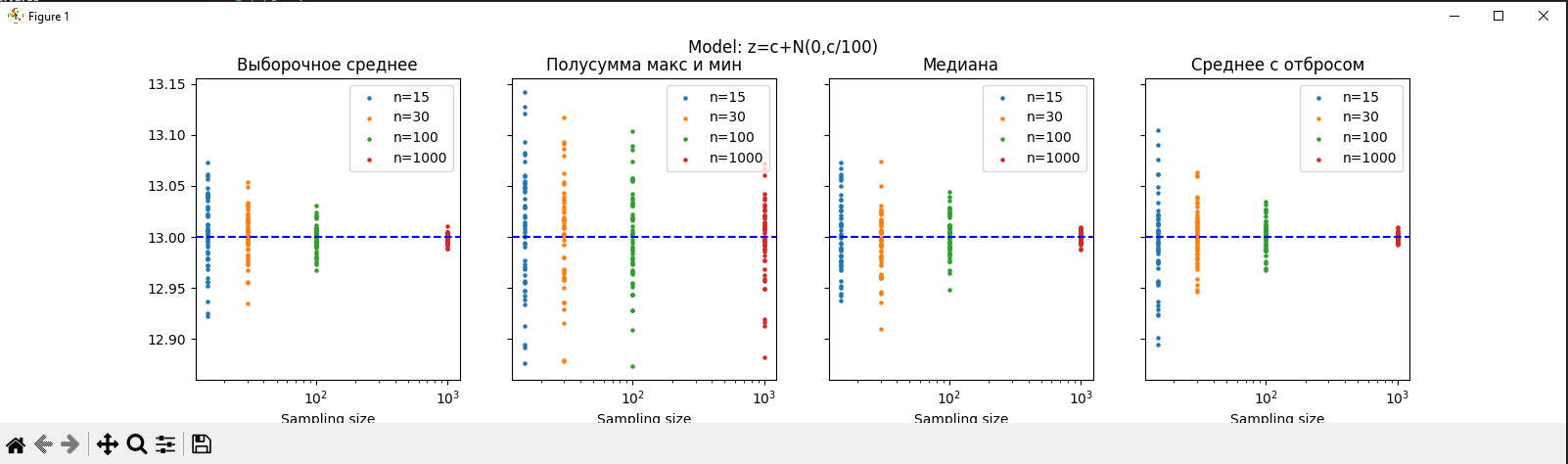
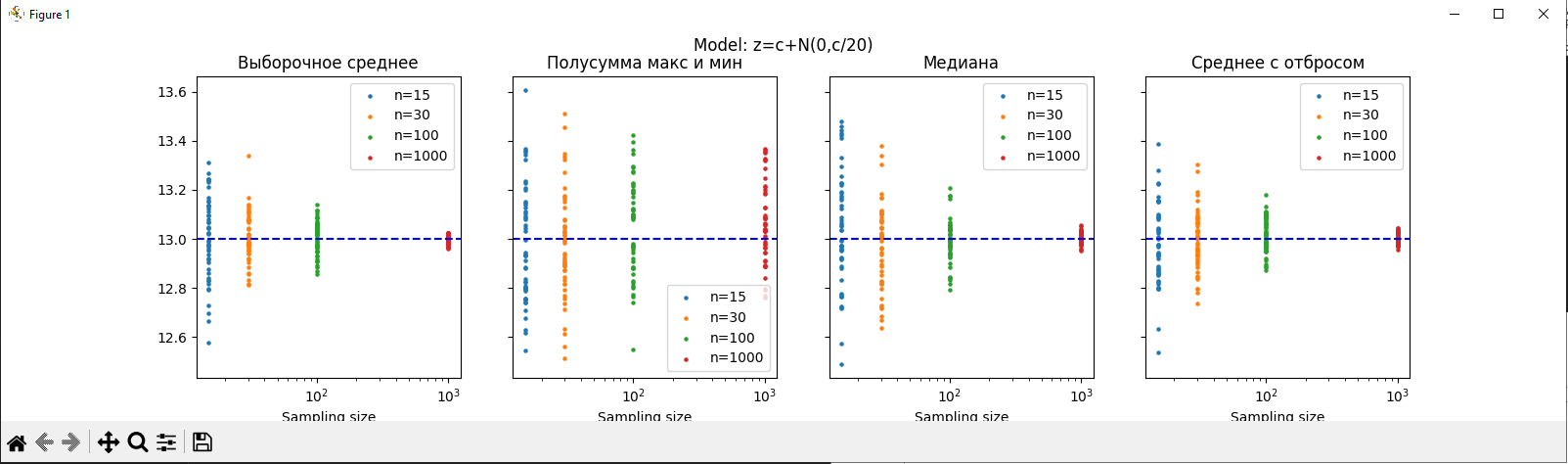
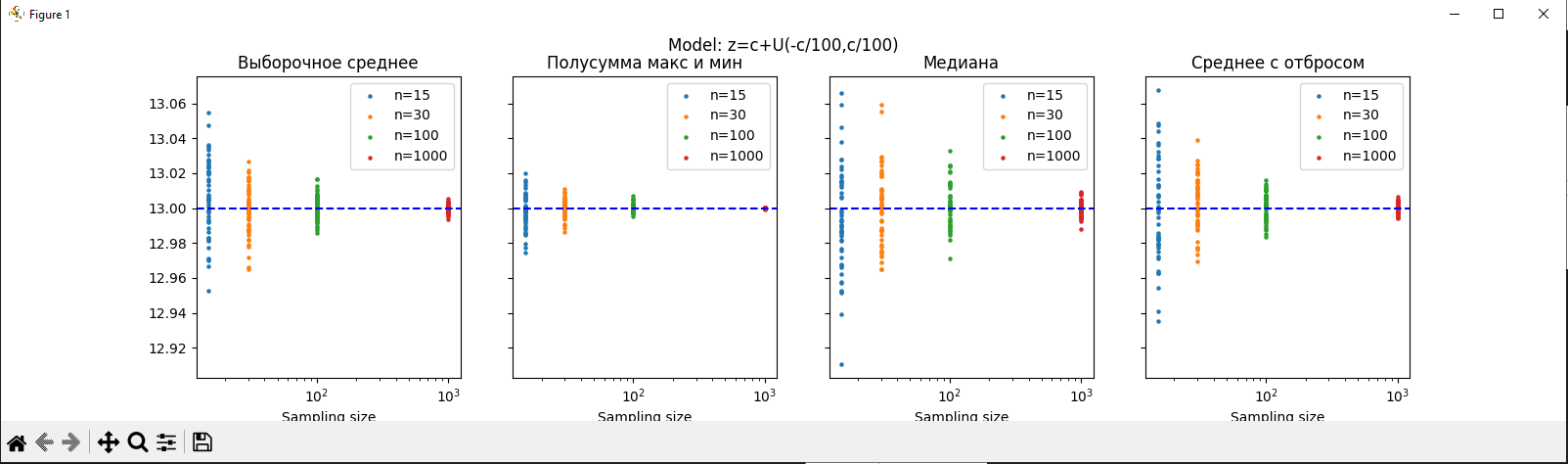
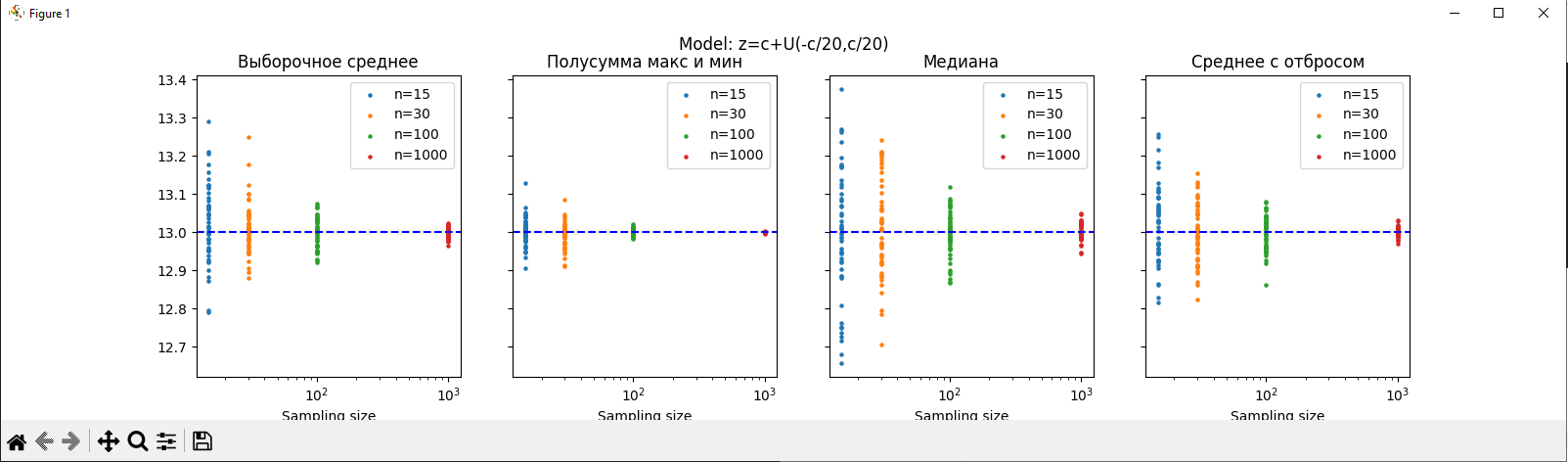
        results.append([distName, n] + errorMean)

columns = ["Модель шума", "Объем выборки", "σx1", "σx2", "σx3", "σx3"]

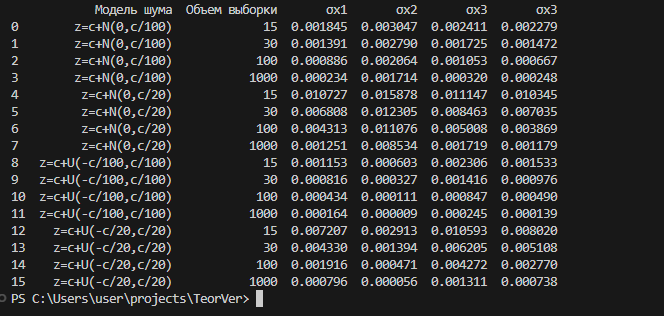
df = pd.DataFrame(results, columns=columns)

print(df)

**Графики оценок вариационных рядов:**

**Таблица средних значений относительных ошибок:**



**Выводы:**

В результате выполнения лабораторной работы я ознакомился с четырьмя характеристиками вариационных рядов, с их преимуществами и недостатками.

Полусумма максимума и минимума – самая простая величина по вычислениям, достаточно неточная, при исходя из опытов показала наиболее нестабильные результаты, позволяет быстро определить характеристику вариационного ряда.

Выборочное среднее значение – более точная характеристика, но учитывает все значения в выборке, поэтому может быть достаточно чувствительно к критическим выбросам.

Среднее значение с отбросом – решает проблему с выбросами, поэтому является наиболее точной характеристикой.

Медиана – наиболее трудоёмкая в плане вычислений характеристика, также не чувствительна к выбросам, может давать неточную характеристику при неравномерном распределнии вероятностей.