|  |  |
| --- | --- |
| **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**

**ДИСЦИПЛИНА: «Методы обработки информации»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. ИУК4-72Б | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Калашников А. С.)  (Подпись) (Ф.И.О.) |
| Проверил: | | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ( Никитенко У. В. )  (Подпись) (Ф.И.О.) |
| Дата сдачи (защиты):  Результаты сдачи (защиты): | | |
|  | - Балльная оценка:  - Оценка: | |
| Калуга, 2023 | | |

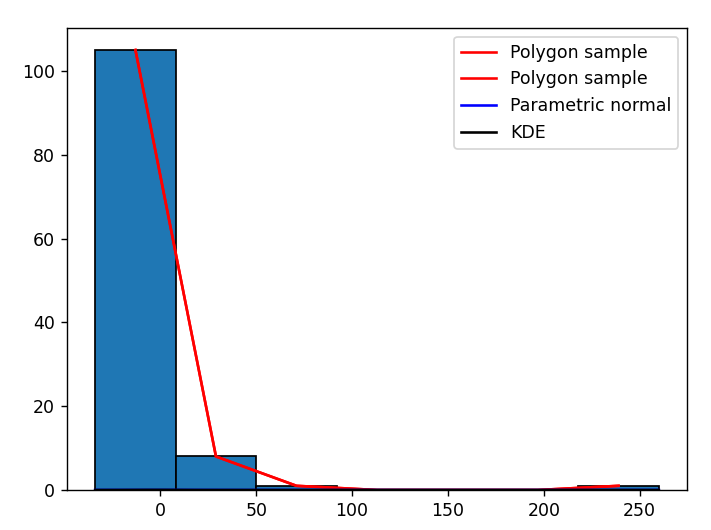
**Цель**: Овладение приемами первичной обработки большой выборки. Выдвижение гипотезы о законе распределения гeнеральной совокупности.

**Вариант №6**

Необходимо разбить выборку на k равных частей. Коэффициент k находим по формуле Стеджерсса k=1+[log2n].

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № промежутка | Границы промежутков | | Ni | Средняя точка  промежутка |
| Ai-1 | Ai |
| 1 | -34 | 8 | 105 | -13 |
| 2 | 8 | 50 | 8 | 29 |
| 3 | 50 | 92 | 1 | 71 |
| 4 | 92 | 134 | 0 | 113 |
| 5 | 134 | 176 | 0 | 155 |
| 6 | 176 | 218 | 0 | 197 |
| 7 | 218 | 260 | 1 | 239 |

Графическая обработка выборки. Исходный код программы представлен в Приложении 1. Построим гистограмму и полигон частот



**Рис**.1 Графическая обработка выборки

**Нахождение выборочных характеристик положения и рассеивания:**

**Характеристики положения:**

**Среднее:** -0.3089

**Медиана:** -3.1474

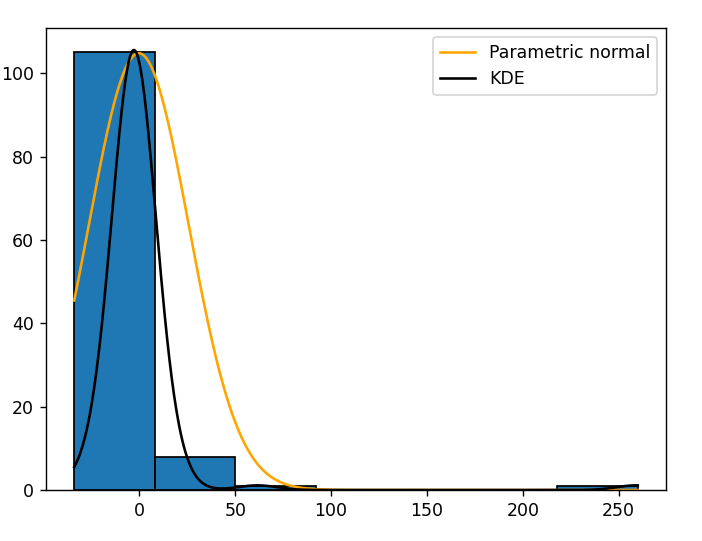
**Мода:** -2.6766

**Характеристики рассеивания:**

**Размах выборки:** 291.6210

**Дисперсия:** 26.0704

**Усреднённая оценка ядерной плотности вероятности**

****

**Рис.2** Усредненная оценка

**Оценка плотности вероятности в средних точках промежутков статистического ряда**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Zi | -13 | 29 | 71 | 113 | 155 | 197 | 239 | ⅀ |
| Ni | 105 | 8 | 1 | 0 | 0 | 0 | 1 | 115 |
| fг(x) | -0.0269 | -0.0020 | -0.0003 | 0.0 | 0.0 | 0.0 | -0.0003 |  |
| fуя(x) | 0.0224 | 0.0011 | 0.0002 | 0 | 0 | 0 | 0 |  |
| fп(x) | 0.0136 | 0.0081 | 0.0004 | 0 | 0 | 0 | 0 |  |
| (fуя-fг)2 | 0.0024 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0024 |
| (fп-fг)2 | 0.0016 | 0.0001 | 0 | 0 | 0 | 0 | 0 | 0.0017 |

**Анализ близости оценок по среднеквадратичным отклонениям**

Исходя из таблицы можно сделать вывод что усреднённая ядерная оценка плотности более близка к логнормальному, а значит предположение о характере распределения выборки (нормальное распределение) можно считать верным.

**Выводы:** в результате выполнения лабораторной работы овладел приемами первичной обработки большой выборки. Выдвижение гипотезы о законе распределения гeнеральной совокупности.

**Приложение №1**

**Листинг программы:**

import matplotlib.pyplot as plt

import numpy as np

import statistics as st

from scipy.stats import gaussian\_kde

def normal(x):

    return 1 / np.sqrt(2\*np.pi) / np.sqrt(np.var(data)) \* np.e\*\*(-1/2 \* ((x - np.mean(data)) / np.sqrt(np.var(data)))\*\*2)

data = list()

with open("Test6.csv", "r+") as input:

    data = [float(item) for item in input.readlines()]

centers = [-13, 29, 71, 113, 155, 197, 239]

buckets = [0] \* 7

for i in range(len(data)):

    if -34<data[i]<8:

        buckets[0] += 1

    elif 8<data[i]<50:

        buckets[1] += 1

    elif 50<data[i]<92:

        buckets[2] += 1

    elif 92<data[i]<134:

        buckets[3] += 1

    elif 134<data[i]<176:

        buckets[4] += 1

    elif 176<data[i]<218:

        buckets[5] += 1

    elif 218<data[i]<260:

        buckets[6] += 1

print(buckets[0])

plt.hist(data, bins=7, edgecolor= "black", range=(-34, 260))

# Нормализованный полигон

plt.plot(centers, [buckets[i] for i in range(len(buckets))],color= "red", label= "Polygon sample")

# Полигон приведённых частот

plt.plot(centers, [buckets[i] for i in range(len(buckets))], color= "red", label= "Polygon sample")

print(f"Mean: {np.mean(data)}")

print(f"Median: {np.median(data)}")

print(f"Mode: {st.mode(data)}")

print(f"R: {max(data) - min(data)}")

print(f"s^2: {np.var(data)}")

print(f"s: {np.sqrt(np.var(data))}")

print(f"V: {np.sqrt(np.var(data)) / np.mean(data) \* 100}%")

x = np.linspace(-34, 260, 700)

y = [(normal(item)) for item in x]

plt.plot(x, y, color= "orange", label= "Parametric normal")

kde = gaussian\_kde(data)

plt.plot(x, [(kde(item)) for item in x], color= "black", label= "KDE")

plt.legend()

plt.show()

fg = [round(buckets[i] / len(data) / -34, 260) for i in

range(len(centers))]

fya = [round(float(kde(i)), 260) for i in centers]

fp = [round(normal(i), 260) for i in centers]

fyag = [round((fya[i] - fg[i])\*\*2, 260) for i in range(len(fg))]

fpg = [round((fp[i] - fg[i])\*\*2, 260) for i in range(len(fg))]

print(fg)

print(fya)

print(fp)

print(fyag)

print(fpg)