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	<u>ИУК</u>	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 (Подпись)

Проверил: (Никитенко У.В.)

Дата сдачи (защиты):

Результаты сдачи (защиты):

- Балльная оценка:

- Оценка:

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.

№ промежутка	Границы промежутков		n_i	Средняя точка
	a_{i-1}	a_i		промежутка

- 2. Графические характеристики выборки строим гистограмму и полигон приведенных частот. Выдвигаем гипотезу о виде плотности вероятности генерального распределения.
- 3. Находим выборочные характеристики положения и рассеивания.
- 4. Для сравнения с гистограммой и полигоном приведенных частот на одном чертеже постройте графики гистограммной оценки плотности вероятности \hat{f}_{Γ} , параметрической оценки плотности вероятности \hat{f}_{η} , и усредненную ядерную оценку плотности вероятности \hat{f}_{yg} .
- 5. Значения оценок плотности вероятности в средних точках промежутков группированного статистического ряда оформите в виде таблицы.

z_{t}	Σ	
n_i		1
$\widehat{f}_{ ext{r}}(x)$		7
$\widehat{f}_{ys}(x)$		
$\widehat{f}_{\pi}(x)$	_	
$(\widehat{f}_{y\pi}-\widehat{f}_r)^2$		
$(\widehat{f}_{\mathrm{ff}}-\widehat{f}_{\mathrm{r}})^2$		

6. Проанализируйте близость оценок по средним квадратическим отклонениям $\hat{f}_{y_{\mathrm{N}}}$ и \hat{f}_{Π} от \hat{f}_{Γ} .

Вариант 7

Листинг:

```
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='[ructorpamma']
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("N", [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 оценка плотности вероятности",
parametric estimate)
print ("Значения оценок в средних точках:")
print(table)
table = PrettyTable()
table.add column("N", [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:

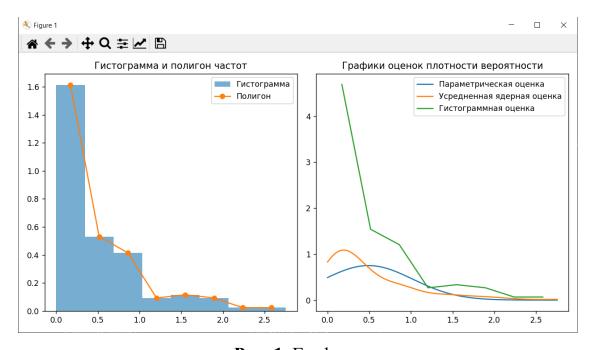


Рис. 1. Графики

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

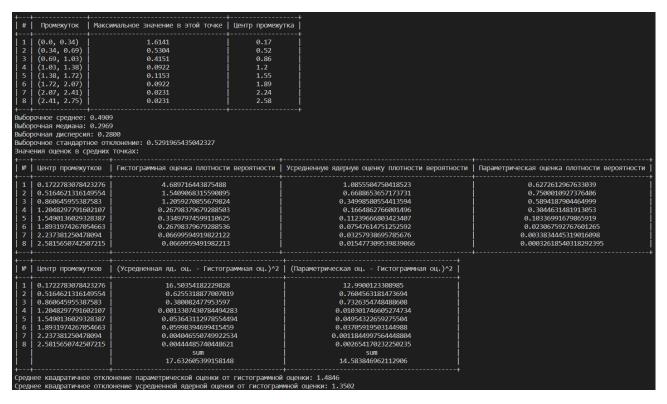


Рис. 2. Вычисления

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