

Федеральное государственное автономное образовательное учреждение  
высшего образования «Национальный исследовательский университет  
ИТМО»

Дисциплина «Теория вероятностей»

**Отчёт**  
**по практической работе №6**

Выполнил:  
Гаврилин Олег Сергеевич

Группа: 2.8

Преподаватель:  
Селина Елена Георгиевна

Санкт-Петербург 2024г.

Задание: каждый студент получает выборку из 20 чисел. Необходимо определить следующие статистические характеристики: вариационный ряд, экстремальные значения и размах, оценки математического ожидания и среднеквадратического отклонения, эмпирическую функцию распределения и её график, гистограмму и полигон приведенных частот группированной выборки. Для расчета характеристик и построения графиков нужно написать программу на одном из языков программирования. Листинг программы и результаты работы должны быть представлены в отчете по практической работе. Стандартные функции статистики использовать нельзя.

Код:

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 import math
4
5 # Исходные данные: выборка из 20 чисел
6 data = [-0.03, 0.73, -0.59, -1.59, 0.38, 1.49, 0.14, -0.62, -1.59, 1.45, -0.38,
7         -1.49, -0.15, 0.63, 0.06, -1.59, 0.61, 0.62, -0.05, 1.56]
8
9 # 1. Вариационный ряд
10 data_sorted = sorted(data)
11
12 # 2. Экстремальные значения
13 min_val = data_sorted[0]
14 max_val = data_sorted[-1]
15
16 # 3. Размах
17 range_val = max_val - min_val
18
19 # 4. Числовые характеристики
20 mean = sum(data_sorted) / len(data_sorted) # Выборочное среднее
21
22 # Выборочная дисперсия
23 variance = sum((x - mean) ** 2 for x in data_sorted) / len(data_sorted)
24
25 # Исправленная дисперсия
26 corrected_variance = sum((x - mean) ** 2 for x in data_sorted) / (len(data_sorted) -
27     ↪ 1)
28
29 # Выборочное СКО
30 std_dev = variance ** 0.5
31
32 # Исправленное СКО
33 corrected_std_dev = corrected_variance ** 0.5
34
35 # 5. Статистический ряд (распределение частот)
36 unique_values = sorted(set(data_sorted))
37 frequencies = [data_sorted.count(x) for x in unique_values]
38
39 # Статистический ряд
40 statistical_series = list(zip(unique_values, frequencies))
41
42 # 6. Эмпирическая функция распределения
```

```

42 def empirical_distribution(x, data_sorted):
43     return sum(1 for value in data_sorted if value <= x) / len(data_sorted)
44
45 # 7. Аналитическая функция распределения (предполагаем равномерное распределение)
46 def analytical_distribution(x, min_val, max_val):
47     if x < min_val:
48         return 0
49     elif x > max_val:
50         return 1
51     else:
52         return (x - min_val) / (max_val - min_val)
53
54 # 8. Построение группированного (интервального) ряда
55 n = len(data_sorted)
56 k = math.ceil(1 + math.log2(n)) # Количество интервалов по формуле Стерджесса
57 intervals = np.linspace(min_val, max_val, k + 1) # k интервалов
58
59 # Группированный ряд
60 histogram_data = [0] * k
61 for x in data_sorted:
62     for i in range(1, len(intervals)):
63         if intervals[i - 1] <= x < intervals[i]:
64             histogram_data[i - 1] += 1
65             break
66
67 # Нормируем частоты
68 relative_frequencies = [freq / len(data_sorted) for freq in histogram_data]
69
70 # Полигон частот
71 polygon_x = [(intervals[i] + intervals[i + 1]) / 2 for i in range(len(intervals) -
72     ↪ 1)]
73
74 polygon_y = relative_frequencies
75
76
77 # 9. Построение графиков
78 plt.figure(figsize=(16, 8))
79
80 # Эмпирическая и аналитическая функции распределения
81 plt.subplot(1, 2, 1)
82 plt.step(data_sorted, [empirical_distribution(x, data_sorted) for x in data_sorted],
83     ↪ where='post', label='Эмпирическая функция распределения')
84 plt.plot(data_sorted, [analytical_distribution(x, min_val, max_val) for x in
85     ↪ data_sorted], linestyle='--', color='orange', label='Аналитическая функция
86     ↪ распределения')
87 plt.xlabel('Значение выборки')
88 plt.ylabel('F(x)')
89 plt.title('Эмпирическая и аналитическая функции распределения')
90 plt.grid()
91 plt.legend()
92
93 # Гистограмма и полигон частот
94 plt.subplot(1, 2, 2)
95 plt.bar(intervals[:-1], relative_frequencies, width=np.diff(intervals), align='edge',
96     ↪ alpha=0.7, label='Гистограмма')

```

```

90 plt.plot(polygon_x, polygon_y, marker='o', color='red', label='Полигон частот')
91 plt.xlabel('Интервалы')
92 plt.ylabel('Относительная частота')
93 plt.title('Гистограмма и полигон частот')
94 plt.grid()
95 plt.legend()
96
97 plt.tight_layout()
98 plt.show()
99
100 # 10. Вывод аналитической функции распределения
101 print("Аналитическая функция распределения F(x):")
102 print(f"F(x) = 0, если x < {min_val}")
103 print(f"F(x) = 1, если x > {max_val}")
104 print(f"F(x) = (x - {min_val}) / ({max_val} - {min_val}), если {min_val} <= x <=
    ↪ {max_val}")
105
106 # Вывод результатов
107 print("\nВариационный ряд:", data_sorted)
108 print("Статистический ряд (значение, частота):", statistical_series)
109 print("Экстремальные значения: min =", min_val, ", max =", max_val)
110 print("Размах:", range_val)
111 print("Математическое ожидание:", mean)
112 print("Выборочная дисперсия:", variance)
113 print("Исправленная дисперсия:", corrected_variance)
114 print("Выборочное среднее квадратическое отклонение:", std_dev)
115 print("Исправленное среднее квадратическое отклонение:", corrected_std_dev)
116
117 # Вывод интервалов значений для гистограммы
118 print("\nИнтервалы значений для гистограммы:")
119 for i in range(len(intervals) - 1):
120     print(f"Интервал {i + 1}: ({intervals[i]:.2f}, {intervals[i + 1]:.2f}) с частотой
        ↪ {relative_frequencies[i]:.2f}")
121
122 # 11. Вывод значений для каждой точки
123 print("\nЗначения эмпирической и аналитической функций распределения для каждого
    ↪ значения:")
124 for x in data_sorted:
125     print(f"x = {x:.2f}, F(x) (эмпирическая) = {empirical_distribution(x,
        ↪ data_sorted):.4f}, F(x) (аналитическая) = {analytical_distribution(x,
        ↪ min_val, max_val):.4f}")
126

```

## Результаты вычислений

### Вариационный ряд

[-1.59, -1.59, -1.59, -1.49, -0.62, -0.59, -0.38, -0.15, -0.05, -0.03,  
0.06, 0.14, 0.38, 0.61, 0.62, 0.63, 0.73, 1.45, 1.49, 1.56]

## Статистический ряд

Значение	Частота
-1.59	3
-1.49	1
-0.62	1
-0.59	1
-0.38	1
-0.15	1
-0.05	1
-0.03	1
0.06	1
0.14	1
0.38	1
0.61	1
0.62	1
0.63	1
0.73	1
1.45	1
1.49	1
1.56	1

## Экстремальные значения

$\min = -1.59, \max = 1.56$

## Размах

3.1500000000000004

## Числовые характеристики

- Математическое ожидание:  $-0.020499999999999997$
- Выборочная дисперсия:  $0.9654447499999999$
- Исправленная дисперсия:  $1.016257631578947$
- Выборочное среднеквадратическое отклонение:  $0.9825704809325384$
- Исправленное среднеквадратическое отклонение:  $1.008096042834683$

## Интервалы значений для гистограммы

Интервал	Частота
$(-1.59, -1.06)$	0.20
$(-1.06, -0.54)$	0.10
$(-0.54, -0.01)$	0.20
$(-0.01, 0.51)$	0.15
$(0.51, 1.03)$	0.20
$(1.03, 1.56)$	0.10

Значения аналитической функции распределения  $F(x)$  для каждого значения  $x$ :

$$F(x) = \begin{cases} 0, & \text{если } x < -1.59, \\ 0.0000, & \text{если } -1.59 \leq x < -1.49, \\ 0.0317, & \text{если } -1.49 \leq x < -0.62, \\ 0.3079, & \text{если } -0.62 \leq x < -0.59, \\ 0.3175, & \text{если } -0.59 \leq x < -0.38, \\ 0.3841, & \text{если } -0.38 \leq x < -0.15, \\ 0.4571, & \text{если } -0.15 \leq x < -0.05, \\ 0.4889, & \text{если } -0.05 \leq x < -0.03, \\ 0.4952, & \text{если } -0.03 \leq x < 0.06, \\ 0.5238, & \text{если } 0.06 \leq x < 0.14, \\ 0.5492, & \text{если } 0.14 \leq x < 0.38, \\ 0.6254, & \text{если } 0.38 \leq x < 0.61, \\ 0.6984, & \text{если } 0.61 \leq x < 0.62, \\ 0.7016, & \text{если } 0.62 \leq x < 0.63, \\ 0.7048, & \text{если } 0.63 \leq x < 0.73, \\ 0.7365, & \text{если } 0.73 \leq x < 1.45, \\ 0.9651, & \text{если } 1.45 \leq x < 1.49, \\ 0.9778, & \text{если } 1.49 \leq x < 1.56, \\ 1.0000, & \text{если } x \geq 1.56. \end{cases}$$

