

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,
8         ↪ 0.06, -1.59, 0.61, 0.62, -0.05, 1.56]
9
10 # 1. Вариационный ряд
11 data_sorted = sorted(data)
12
13 # 2. Экстремальные значения
14 min_val = data_sorted[0]
15 max_val = data_sorted[-1]
16
17 # 3. Размах
18 range_val = max_val - min_val
19
20 # 4. Числовые характеристики
21 mean = sum(data_sorted) / len(data_sorted) # Выборочное среднее
22
23 # Выборочная дисперсия
24 variance = sum((x - mean) ** 2 for x in data_sorted) / len(data_sorted)
25
26 # Исправленная дисперсия
27 corrected_variance = sum((x - mean) ** 2 for x in data_sorted) /
28 ↪ (len(data_sorted) - 1)
29
30 # Выборочное СКО
31 std_dev = variance ** 0.5
32
33 # Исправленное СКО
34 corrected_std_dev = corrected_variance ** 0.5
35
36 # 5. Статистический ряд (распределение частот)
37 unique_values = sorted(set(data_sorted))
38 frequencies = [data_sorted.count(x) for x in unique_values]
39
40 # Статистический ряд
41 statistical_series = list(zip(unique_values, frequencies))
```

```

41 # 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 # Эмпирическая функция для всех значений выборки
46 empirical_values = [empirical_distribution(x, data_sorted) for x in data_sorted]
47
48 # Аналитическая функция распределения (предполагаем равномерное распределение)
49 def analytical_distribution(x, min_val, max_val):
50     if x < min_val:
51         return 0
52     elif x > max_val:
53         return 1
54     else:
55         return (x - min_val) / (max_val - min_val)
56
57 # Аналитическая функция для всех значений выборки
58 analytical_values = [analytical_distribution(x, min_val, max_val) for x in
59 ↪ data_sorted]
60
61 # 7. Построение группированного (интервального) ряда
62 n = len(data_sorted)
63 k = math.ceil(1 + math.log2(n)) # Количество интервалов по формуле Стерджесса
64 intervals = np.linspace(min_val, max_val, k + 1) # k интервалов
65
66 # Группированный ряд
67 histogram_data = [0] * k
68 for x in data_sorted:
69     for i in range(1, len(intervals)):
70         if intervals[i - 1] <= x < intervals[i]:
71             histogram_data[i - 1] += 1
72             break
73
74 # Нормируем частоты
75 relative_frequencies = [freq / len(data_sorted) for freq in histogram_data]
76
77 # Полигон частот
78 polygon_x = [(intervals[i] + intervals[i + 1]) / 2 for i in range(len(intervals)
79 ↪ - 1)]
80 polygon_y = relative_frequencies
81
82 # 8. Построение графиков
83 plt.figure(figsize=(16, 8))
84
85 # Эмпирическая и аналитическая функции распределения
86 plt.subplot(1, 2, 1)
87 plt.step(data_sorted, empirical_values, where='post', label='Эмпирическая функция
88 ↪ распределения')
89 plt.plot(data_sorted, analytical_values, linestyle='--', color='orange',
90 ↪ label='Аналитическая функция распределения')
91 plt.xlabel('Значение выборки')
92 plt.ylabel('F(x)')
93 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),
96     ↪ align='edge', alpha=0.7, label='Гистограмма')
97     plt.plot(polygon_x, polygon_y, marker='o', color='red', label='Полигон частот')
98     plt.xlabel('Интервалы')
99     plt.ylabel('Относительная частота')
100    plt.title('Гистограмма и полигон частот')
101    plt.grid()
102    plt.legend()
103
104    plt.tight_layout()
105    plt.show()
106
107    # Вывод результатов
108    print("Вариационный ряд:", data_sorted)
109    print("Статистический ряд (значение, частота):", statistical_series)
110    print("Экстремальные значения: min =", min_val, ", max =", max_val)
111    print("Размах:", range_val)
112    print("Математическое ожидание:", mean)
113    print("Выборочная дисперсия:", variance)
114    print("Исправленная дисперсия:", corrected_variance)
115    print("Выборочное среднее квадратическое отклонение:", std_dev)
116    print("Исправленное среднее квадратическое отклонение:", corrected_std_dev)
117
118    # Вывод интервалов значений для гистограммы
119    print("\nИнтервалы значений для гистограммы:")
120    for i in range(len(intervals) - 1):
121        print(f"Интервал {i + 1}: ({intervals[i]:.2f}, {intervals[i + 1]:.2f}) с
122        ↪ частотой {relative_frequencies[i]:.2f}")

```

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 среднееквадратическое отклонение: 0.9825704809325384
- Исправленное среднееквадратическое 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

