

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
```

# Лабораторная работа №1

## Задание 3

**{Фамилия Имя}, {Номер группы}, Вариант {Номер варианта}, ({Дата})**

## Задание

По данным выборки ( $n = 100$ ) требуется:

1. Составить вариационный ряд и список вариантов
2. Определить количество интервалов, шаг интервала, составить интервальный вариационный ряд и найти среднее по интервалу.
3. Составить ряд распределения частот интервального вариационного ряда и построить гистограмму частот.
4. Составить ряд распределения относительных частот интервального вариационного ряда и построить гистограмму относительных частот.
5. Составить эмпирическую функцию распределения и построить график эмпирической функции распределения.
6. Найти основные числовые характеристики вариационного ряда:
  - Выборочное среднее -  $\bar{x}_g$
  - Выборочную дисперсию -  $S^2$
  - Стандартное отклонение (Среднеквадратическое отклонение) -  $S$
  - Коэффициент вариации -  $CV$
7. Пояснить смысл полученных результатов

## Данные

```
lb_author = 'Фамилия, Имя'
lb_num = 1
lb_variant = 59
lb_exercise_num = 3
print(f'Фамилия, Имя: {lb_author}\nНомер лабораторной работы:
```

```
{lb_num}\nВариант лабораторной работы: {lb_variant}\nНомер задания лабораторной работы: {lb_exercise_num}')
```

Фамилия, Имя: Фамилия, Имя  
Номер лабораторной работы: 1  
Вариант лабораторной работы: 59  
Номер задания лабораторной работы: 3

```
task = pd.read_json('input/Данные к заданию №3.json')  
data = pd.Series(task['Данные'][lb_variant - 1])  
print(data.tolist())
```

```
[31.2, 30.7, 31.3, 30.6, 30.2, 28.9, 30.3, 30.8, 32.7, 27.6, 29.5, 31.0,  
32.2, 29.7, 31.3, 30.7, 33.1, 29.2, 28.2, 26.5, 28.3, 28.9, 30.7, 27.3,  
27.5, 34.8, 35.3, 32.3, 28.6, 31.4, 29.0, 31.1, 28.5, 30.4, 27.2, 27.1,  
29.5, 30.8, 28.2, 36.6, 31.7, 28.6, 30.0, 25.9, 25.6, 29.4, 32.5, 32.6,  
25.6, 28.9, 28.2, 27.6, 30.0, 30.4, 26.9, 29.8, 31.6, 33.0, 30.7, 28.6,  
28.0, 29.8, 35.2, 28.3, 29.6, 31.9, 30.7, 28.3, 32.2, 29.1, 25.4, 29.5,  
28.0, 32.2, 30.9, 27.5, 30.9, 32.4, 29.7, 32.2, 30.6, 32.1, 33.2, 32.0,  
32.2, 34.9, 27.6, 27.0, 28.8, 33.0, 31.0, 27.6, 30.9, 28.5, 32.1, 31.8,  
32.7, 29.8, 30.4, 26.6]
```

## Всего элементов ряда

```
data_len = len(data)  
data_len
```

```
100
```

## Пункт 1

Составить вариационный ряд и список вариантов.

## Вариационный ряд

```
data_sort = data.sort_values()  
print(data_sort.tolist())
```

```
[25.4, 25.6, 25.6, 25.9, 26.5, 26.6, 26.9, 27.0, 27.1, 27.2, 27.3, 27.5,  
27.5, 27.6, 27.6, 27.6, 27.6, 28.0, 28.0, 28.2, 28.2, 28.2, 28.3, 28.3,
```

```
28.3, 28.5, 28.5, 28.6, 28.6, 28.6, 28.8, 28.9, 28.9, 28.9, 29.0, 29.1,
29.2, 29.4, 29.5, 29.5, 29.5, 29.6, 29.7, 29.7, 29.8, 29.8, 29.8, 30.0,
30.0, 30.2, 30.3, 30.4, 30.4, 30.4, 30.6, 30.6, 30.7, 30.7, 30.7, 30.7,
30.7, 30.8, 30.8, 30.9, 30.9, 30.9, 31.0, 31.0, 31.1, 31.2, 31.3, 31.3,
31.4, 31.6, 31.7, 31.8, 31.9, 32.0, 32.1, 32.1, 32.2, 32.2, 32.2, 32.2,
32.2, 32.3, 32.4, 32.5, 32.6, 32.7, 32.7, 33.0, 33.0, 33.1, 33.2, 34.8,
34.9, 35.2, 35.3, 36.6]
```

## Варианты, $x_i$

```
unique = sorted(data.unique())
print(unique)
```

```
[25.4, 25.6, 25.9, 26.5, 26.6, 26.9, 27.0, 27.1, 27.2, 27.3, 27.5, 27.6,
28.0, 28.2, 28.3, 28.5, 28.6, 28.8, 28.9, 29.0, 29.1, 29.2, 29.4, 29.5,
29.6, 29.7, 29.8, 30.0, 30.2, 30.3, 30.4, 30.6, 30.7, 30.8, 30.9, 31.0,
31.1, 31.2, 31.3, 31.4, 31.6, 31.7, 31.8, 31.9, 32.0, 32.1, 32.2, 32.3,
32.4, 32.5, 32.6, 32.7, 33.0, 33.1, 33.2, 34.8, 34.9, 35.2, 35.3, 36.6]
```

## Пункт 2

Определить количество интервалов, шаг интервала и составить интервальный вариационный ряд.

### Количество интервалов, $k$

$$k = \lceil \log_2 n \rceil + 1$$

```
k = int(np.ceil(np.log2(len(data_sort)) + 1))
print(f'Количество интервалов: {k}')
```

Количество интервалов: 8

### Шаг интервала, $\Delta$

$$\Delta = \frac{\max(x) - \min(x)}{k}$$

```
delta = (data_sort.max() - data_sort.min()) / k
print(f'Шаг интервала: {delta}')
```

Шаг интервала: 1.4000000000000004

## Интервальный вариационный ряд

```
intervals = pd.Series([data_sort.min() + i * delta for i in range(k+1)])  
print(intervals.tolist())
```

```
[25.4, 26.799999999999997, 28.2, 29.6, 31.0, 32.4, 33.8, 35.2, 36.6]
```

## Среднее по интервалу

```
intervals_mean = [(i.left + i.right)/2 for i in  
pd.Series(data_sort).value_counts(bins=8).sort_index().index]  
intervals_mean
```

```
[26.094, 27.5, 28.9, 30.3, 31.7, 33.099999999999994, 34.5,  
35.900000000000006]
```

## Пункт 3

Составить ряд распределения частот интервального вариационного ряда и построить гистограмму частот.

## Распределение частот интервального вариационного ряда

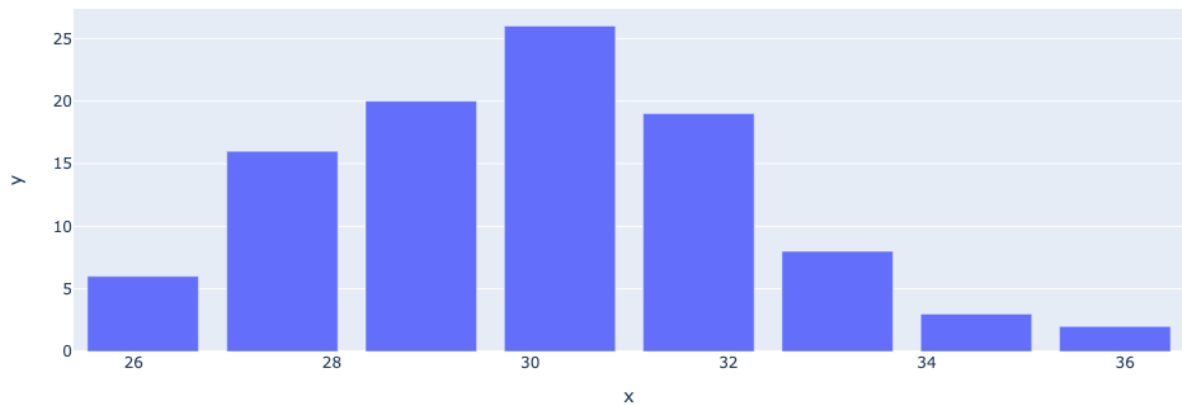
```
freq_intervals =  
pd.Series(data_sort).value_counts(bins=8).sort_index().values  
freq_intervals = pd.DataFrame(freq_intervals, index=intervals_mean,  
columns=["Частота"]).T  
freq_intervals
```

	26.094	27.500	28.900	30.300	31.700	33.100	34.500	35.900
Частота	6	16	20	26	19	8	3	2

## Гистограмма частот вариационного ряда

```
fig = px.bar(x=intervals_mean, y=freq_intervals.iloc[0],  
title='Гистограмма частот вариационного ряда')  
fig.show()
```

Гистограмма частот вариационного ряда



## Пункт 4

Составить ряд распределения относительных частот интервального вариационного ряда и построить гистограмму относительных частот.

### Распределение относительных частот интервального вариационного ряда

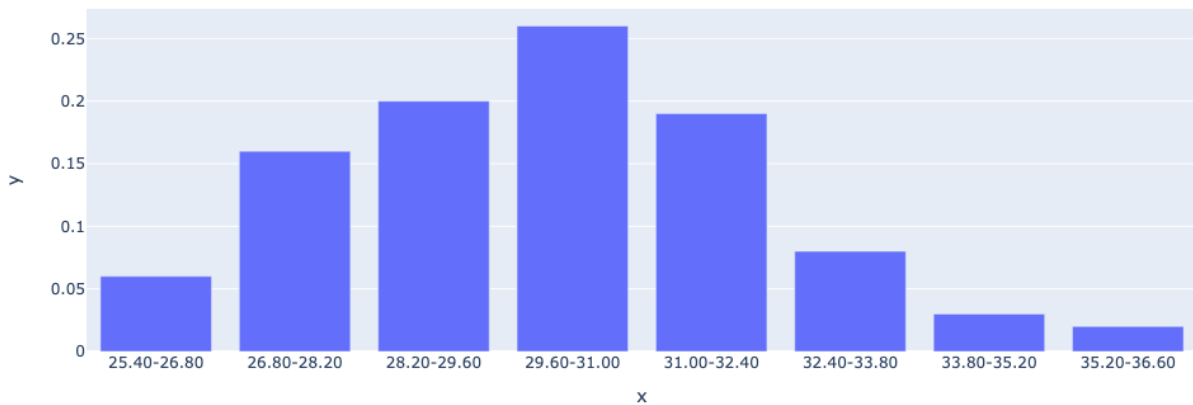
```
rel_intervals_freq =
pd.Series(data_sort).value_counts(bins=8).sort_index().values /
len(data_sort)
rel_intervals_freq = pd.DataFrame(rel_intervals_freq, index=
[f'{intervals[i]:.2f}-{intervals[i+1]:.2f}' for i in
range(intervals.count()-1)], columns=["Относительная частота"]).T
rel_intervals_freq
```

	25.40- 26.80	26.80- 28.20	28.20- 29.60	29.60- 31.00	31.00- 32.40	32.40- 33.80	33.80- 35.20	35.20- 36.60
Относительная частота	0.06	0.16	0.2	0.26	0.19	0.08	0.03	0.02

### Гистограмма распределения относительных частот интервального вариационного ряда

```
fig = px.bar(x=[f'{intervals[i]:.2f}-{intervals[i+1]:.2f}' for i in
range(intervals.count()-1)], y=rel_intervals_freq.values.tolist()[0],
title='Гистограмма относительных частот вариационного ряда')
fig.show()
```

Гистограмма относительных частот вариационного ряда



## Пункт 5

Составить эмпирическую функцию распределения и построить график эмпирической функции распределения.

Эмпирическая функция распределения  $F^*$

```
emp_func = rel_intervals_freq.iloc[0].cumsum()
emp_func.name = "F*"
emp_func = pd.DataFrame(emp_func, index=[f'{intervals[i]:.2f}-{intervals[i+1]:.2f}' for i in range(intervals.count()-1)]).T
emp_func
```

	25.40- 26.80	26.80- 28.20	28.20- 29.60	29.60- 31.00	31.00- 32.40	32.40- 33.80	33.80- 35.20	35.20- 36.60
F*	0.06	0.22	0.42	0.68	0.87	0.95	0.98	1.0

```
def emp_func_latex(emp_func, nums):
    print(f'$$\nF^*(x) = \n\begin{{cases}}')
    for i in range(len(emp_func.values.tolist()[0])):
        if i == 0:
            print(f'0, && x \le {unique[0]:.2f} \\\n')
        elif i == len(emp_func.values.tolist()[0]) - 1:
            print(f'{emp_func.values.tolist()[0][i-1]:.2f}, && {nums[i-1]:.2f} < x \le {nums[i]:.2f} \\\n')
            print(f'1.0, && x > {nums[i]:.2f} \\\n')
        else:
            print(f'{emp_func.values.tolist()[0][i-1]:.2f}, && {nums[i-1]:.2f} < x \le {nums[i]:.2f} \\\n')
    print('\n\end{{cases}}\n$')

```

```
emp_func_latex(emp_func, intervals)
```

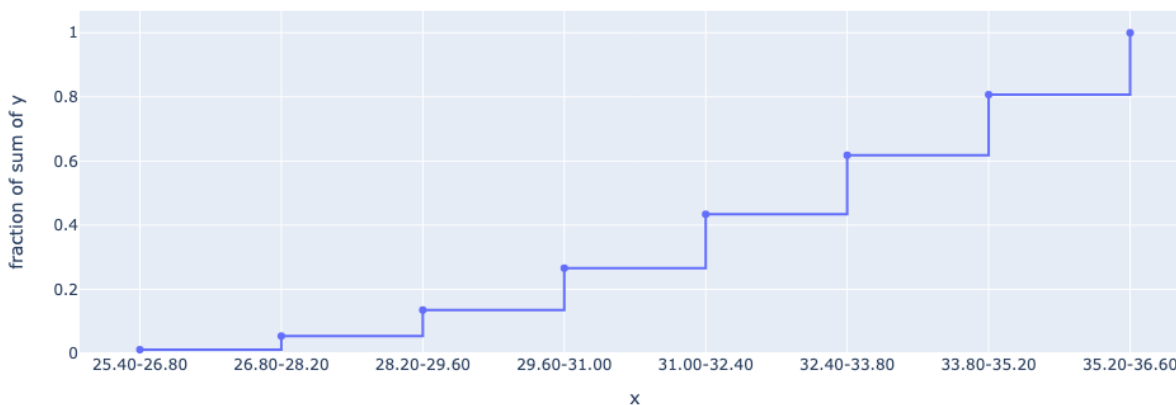
```
$$
F^*(x) =
\begin{cases}
0, & x \leq 25.40 \\
0.06, & 25.40 < x \leq 26.80 \\
0.22, & 26.80 < x \leq 28.20 \\
0.42, & 28.20 < x \leq 29.60 \\
0.68, & 29.60 < x \leq 31.00 \\
0.87, & 31.00 < x \leq 32.40 \\
0.95, & 32.40 < x \leq 33.80 \\
0.98, & 33.80 < x \leq 35.20 \\
1.0, & x > 35.20
\end{cases}
$$
```

$$F^*(x) = \begin{cases} 0, & x \leq 25.40 \\ 0.06, & 25.40 < x \leq 26.80 \\ 0.22, & 26.80 < x \leq 28.20 \\ 0.42, & 28.20 < x \leq 29.60 \\ 0.68, & 29.60 < x \leq 31.00 \\ 0.87, & 31.00 < x \leq 32.40 \\ 0.95, & 32.40 < x \leq 33.80 \\ 0.98, & 33.80 < x \leq 35.20 \\ 1.0, & x > 35.20 \end{cases}$$

## График эмпирической функции распределения

```
fig = px.ecdf(x=[f'{intervals[i]:.2f}-{intervals[i+1]:.2f}' for i in
range(intervals.count()-1)], y=emp_func.values.tolist()[0], markers=True,
title='График эмпирической функции распределения')
fig.show()
```

График эмпирической функции распределения



```
# def plot_cdf_func(x, y, title, c=1):
#     fig, ax = plt.subplots(figsize=(10, 5))
#     min_y, max_y = min(y), max(y)
#     arrow_length = ((max_y - min_y) / len(y)) * c
#     for idx in range(len(y) - 1):
#         dx = x[idx] - x[idx + 1]
#         ax.arrow(x=x[idx + 1], y=y[idx], dx=dx, dy=0, color="blue",
# head_width=.05, head_length=arrow_length)
#     ax.set_title(title)
#     ax.set_ylabel("F*")
#     ax.set_xlabel("x")
#     plt.plot()
#
#
# plot_cdf_func([f'{intervals[i]:.2f}-{intervals[i+1]:.2f}' for i in
# range(intervals.count()-1)], emp_func.values.tolist()[0], 'График
# эмпирической функции распределения')
```

## Пункт 6

Найти основные числовые характеристики вариационного ряда

**Выборочное среднее,  $\bar{x}$**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

```
mean = np.mean(data)
print(f'Выборочное среднее: {mean:.2f}')
```

Выборочное среднее: 30.11



## Выборочная дисперсия, $s_x^2$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

```
var = np.var(data, ddof=1)
print(f'Выборочная дисперсия: {var:.2f}')
```

Выборочная дисперсия: 5.17

## Стандартное отклонение (Среднеквадратическое отклонение), $s_x$

$$s_x = \sqrt{s^2}$$

```
std = np.std(data, ddof=1)
print(f'Среднеквадратическое отклонение: {std:.2f}')
```

Среднеквадратическое отклонение: 2.27

## Коэффициент вариации, $CV$

$$CV = \frac{s}{\bar{x}} \cdot 100$$

```
cv = (std / mean) * 100
print(f'Коэффициент вариации: {cv:.2f}')
```

Коэффициент вариации: 7.55