```python
import pandas as pd
import numpy as np
import scipy.stats as st
from IPython.display import display, Markdown
```

Лабораторная работа №3

Задание 3

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

Данные

Для заданного интервального выборочного ряда проверить гипотезу: закон распределения генеральной совокупности является показательным при уровне значимости α

Замечание
Начальное значение – это левая граница первого интервала, шаг – длина каждого интервала; число интервалов определяется длиной таблицы.

```python
freq = pd.Series([156, 91, 51, 42, 15, 20, 5, 4, 4, 0, 2], name="f")
n = sum(freq)
alpha = 0.05
min_x = 0
h = 1.3
```

```python
Markdown(f'$n = {n}$; $\\alpha={alpha}$; min_x $= {min_x}$; $h = {h}$')
```

$n = 390$; $\alpha = 0.05$; min_x $= 0$; $h = 1.3$

```python
data = pd.DataFrame()
data["x_i"] = [min_x + h * i for i in range(len(freq))]
data["x_(i+1)"] = [min_x + h * (i + 1) for i in range(len(freq))]
data["f"] = freq
data
```

	x_i	x_(i+1)	f
0	0.0	1.3	156
1	1.3	2.6	91

	x_i	x_(i+1)	f
2	2.6	3.9	51
3	3.9	5.2	42
4	5.2	6.5	15
5	6.5	7.8	20
6	7.8	9.1	5
7	9.1	10.4	4
8	10.4	11.7	4
9	11.7	13.0	0
10	13.0	14.3	2

Шаг 1: Формулировка гипотезы

- H_0: закон распределения генеральной совокупности является показательным.
- H_1: закон распределения генеральной совокупности не является показательным.

Шаг 2: Вычисление характеристик выборки.

Вычисление среднего значения

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

```
def mean_interval_row(xi, xj, f):
    x_mean = (xi + xj) / 2
    return (x_mean * f).sum() / f.sum()

mean = mean_interval_row(data["x_i"], data["x_(i+1)"], data["f"])
mean
```

```
2.6133333333333333
```

Вычисление дисперсии

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

```
def var_interval_row(xi, xj, f):
    x_mean = (xi + xj) / 2
    mean_square = ((x_mean ** 2) * f).sum() / f.sum()
    mean = mean_interval_row(xi, xj, f)
    return mean_square - mean ** 2
```

```
variance = var_interval_row(data["x_i"], data["x_(i+1)"], data["f"])
variance
```

```
5.873655555555557
```

Вычисление стандартного отклонения

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

```
std = np.sqrt(variance)
std
```

```
2.4235625751268643
```

Шаг 3: Проверка гипотезы

Для проверки гипотезы о показательном распределения воспользуемся критерием согласия хи-квадрат. Нужно вычислить теоретические (ожидаемые) частоты попадания значений в каждый интервал, если бы распределение было равномерным. Для этого воспользуемся формулой:

$$P_i = P(x_i < X < x_{i+1}) = e^{-\lambda x_i} - e^{-\lambda x_{i+1}}$$

$$f' = fP_i$$

$$\lambda = 1/\bar{x}$$

```
lmbda = 1 / mean
lmbda
```

```
0.3826530612244898
```

```
p = st.expon.cdf(data["x_(i+1)"], scale=1/lmbda) -
st.expon.cdf(data["x_i"], scale=1/lmbda)

assert np.isclose(p.sum(), 1, rtol=.01), "Сумма теоретических оснований
должна быть равна 1."

data["f'"] = p * data["f"].sum()
data["f'"]
```

```
0     152.848836
1      92.944306
2      56.517565
3      34.367196
4      20.898001
5      12.707655
6       7.727269
7       4.698797
8       2.857244
9       1.737433
10      1.056498
Name: f', dtype: float64
```

Шаг 4: Вычисление значения статистики критерия

Вычислим значение статистики критерия χ^2:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - f_i^*)^2}{f_i^*} = \frac{(\text{наблюдаемая частота} - \text{ожидаемая частота})^2}{\text{ожидаемая частота}}$$

```
chi2_value = ((data["f"] - data["f'"])**2 / data["f'"]).sum()
chi2_value
```

```
12.292391902659396
```

Шаг 5: Определение критического значения и принятие решения

Степени свободы:

```
# один параметр у показательного распр. lambda
df = (data["f"] >= 5).sum() - 2
df
```

```
5
```

Критическое значение

```
chi2_critical = st.chi2.ppf(1-alpha, df=df)
chi2_critical
```

```
11.070497693516351
```

```python
from IPython.display import Markdown, display

display(Markdown("## Вывод"))
display(Markdown(f"Таблица распределения хи-квадрат ($\chi^2$) со степенями свободы {df} и уровнем значимости {alpha} даёт критическое значение $\chi^2 = {chi2_critical}$."))
display(Markdown(f"Так как вычисленное значение статистики критерия Пирсона ($\chi^2 = {chi2_value}$) превышает критическое значение ($\chi^2 = {chi2_critical}$), то мы отвергаем нулевую гипотезу $H_0$ и приниамаем альтернативную гипотезу $H_1$. Это означает, что закон распределения генеральной совокупности не является показательным."))
```

Вывод

Таблица распределения хи-квадрат (χ^2) со степенями свободы 5 и уровнем значимости 0.05 даёт критическое значение $\chi^2 = 11.070497693516351$.

Так как вычисленное значение статистики критерия Пирсона ($\chi^2 = 12.292391902659396$) превышает критическое значение ($\chi^2 = 11.070497693516351$), то мы отвергаем нулевую гипотезу H_0 и приниамаем альтернативную гипотезу H_1. Это означает, что закон распределения генеральной совокупности не является показательным.