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

Лабораторная работа №3

Задание 4

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

Данные

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

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

```
freq = pd.Series([38, 37, 40, 41, 43, 42, 31, 45, 34, 24], name="f")
n = sum(freq)
alpha = 0.05
min_x = 8.4
h = 2.2
```

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

$n = 375$; $\alpha = 0.05$; min_x $= 8.4$; $h = 2.2$

```
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	8.4	10.6	38
1	10.6	12.8	37

	x_i	x_(i+1)	f
2	12.8	15.0	40
3	15.0	17.2	41
4	17.2	19.4	43
5	19.4	21.6	42
6	21.6	23.8	31
7	23.8	26.0	45
8	26.0	28.2	34
9	28.2	30.4	24

Шаг 1: Формулировка гипотезы

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

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
```

```
18.9512
```

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
```

```
36.450285226666665
```

Вычисление стандартного отклонения

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

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

```
6.0374071609149125
```

Шаг 3: Проверка гипотезы

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

$$P_i = P(x_i < X < x_{i+1}) = \frac{1}{b^* - a^*}$$

$$\text{при } a^* \leq x_j \leq b^*$$

$$f' = fP_i$$

$$a^* = \bar{x} - \sqrt{3} \cdot S$$

$$b^* = \bar{x} + \sqrt{3} \cdot S$$

```
a = mean - np.sqrt(3) * std
b = mean + np.sqrt(3) * std
a, b
```

```
(8.494104051315205, 29.408295948684795)
```

```
p = st.uniform.cdf(data["x_(i+1)"], loc=a, scale=b-a) -
st.uniform.cdf(data["x_i"], loc=a, scale=b-a)

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

```
array([0.1006922 , 0.10519173, 0.10519173, 0.10519173, 0.10519173,
       0.10519173, 0.10519173, 0.10519173, 0.10519173, 0.05777397])
```

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

```
0    37.759574
1    39.446898
2    39.446898
3    39.446898
4    39.446898
5    39.446898
6    39.446898
7    39.446898
8    39.446898
9    21.665240
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
```

```
4.301717542071091
```

Шаг 5: Определение критического значения и принятие решения

Степени свободы:

```
# два параметр у равномерного распр. a, b
df = (data["f"] >= 5).sum() - 3
df
```

```
7
```

Критическое значение

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

```
14.067140449340169
```

```
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 ($\chi^2 = {chi2_value}$) не превышает критическое значение ($\chi^2 = {chi2_critical}$), то мы принимаем нулевую гипотезу $H_0$ и отвергаем альтернативную гипотезу $H_1$. Это означает, что закон распределения генеральной совокупности является равномерным."))
```

Вывод

Таблица распределения хи-квадрат (χ^2) со степенями свободы 7 и уровнем значимости 0.05 даёт критическое значение $\chi^2 = 14.067140449340169$.

Так как вычисленное значение статистики критерия Пирсона ($\chi^2 = 4.3017175420710\,91$) не превышает критическое значение ($\chi^2 = 14.067140449340169$), то мы принимаем нулевую гипотезу H_0 и отвергаем альтернативную гипотезу H_1. Это означает, что закон распределения генеральной совокупности является равномерным.