```
import pandas as pd
import numpy as np
import scipy.stats as st
```

Лабораторная работа №3

Задание 1

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

Данные

В одном из городов распределение кол-ва детей в семье имеет следующий вид:

```
data = pd.DataFrame(
    {
        'Количество детей': list(range(6)),
        'Количество семей': [110, 112, 54, 18, 5, 2]
    }
)
alpha = 0.05
n = data["Количество семей"].sum()

data.T
```

	0	1	2	3	4	5
Количество детей	0	1	2	3	4	5
Количество семей	110	112	54	18	5	2

```
data = data.rename(columns={'Количество детей': 'X', 'Количество семей': 'n'})
data
```

	X	n
0	0	110
1	1	112
2	2	54
3	3	18
4	4	5

	X	n
5	5	2

```
print(f"Всего семей: {n}")
```

```
Всего семей: 301
```

С помощью критерия Пирсона проверить гипотезу о том, что результаты получены из распределения Пуассона генеральной совокупности.

Шаг 1: Формулировка гипотезы

H_0: Случайная величина имеет распределения Пуассона.

H_1: Случайная величина не имеет распределения Пуассона.

Шаг 2: Определение уровня значимости и степеней свободы

Уровень значимости:

```
print(f"Уровень значимости: {alpha}")
```

```
Уровень значимости: 0.05
```

Степени свободы:

```
df = (data["n"] >= 5).sum() - 2
df
```

```
3
```

Шаг 3: Расчет значения статистики критерия Пирсона

Вычислим ожидаемые частоты в каждом интервале, если бы выборка имела распределение Пуассона.

$$P(i) = \frac{\lambda^i \cdot e^{-\lambda}}{i!}$$

$$n' = nP_i$$

Считаем λ

$$\bar{x} = \frac{1}{n} \cdot \sum x_i \cdot n_i$$

$$\lambda \approx \bar{x}$$

```
lmbda =  (data["X"] * data["n"]).sum() / n
lmbda
```

```
1.0099667774086378
```

Считаем ожидаемые частоты

```
def expected_freq(x):
    p = st.poisson.pmf(x, mu=lmbda)
    return n * p

data["n'"] = data["X"].apply(expected_freq)
data["n'"]
```

```
0    109.633555
1    110.726248
2     55.914916
3     18.824069
4      4.752921
5      0.960058
Name: n', dtype: float64
```

Теперь вычислим значение статистики критерия Пирсона

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

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

```
1.2568488786645524
```

Шаг 4: Определение критического значения и принятие решения

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

```
7.814727903251179
```

```
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 принимает нулевую гипотезу $H_0$ и
отвергаем альтернативную гипотезу $H_1$. Это означает, что результаты
измерений получены из распределения Пуассона с параметром $\lambda =
{lmbda}$."))
```

Вывод

Таблица распределения хи-квадрат (χ^2) со степенями свободы 3 и уровнем значимости 0.05 даёт критическое значение $\chi^2 = 7.814727903251179$.

Так как вычисленное значение статистики критерия Пирсона ($\chi^2 = 1.2568488786645524$) не превышает критическое значение ($\chi^2 = 7.814727903251179$), то мы принимает нулевую гипотезу H_0 и отвергаем альтернативную гипотезу H_1. Это означает, что результаты измерений получены из распределения Пуассона с параметром $\lambda = 1.0099667774086378$.