

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

Лабораторная работа №3

Задание 2

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

Данные

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

Замечание

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

```
freq = pd.Series([2, 9, 30, 61, 92, 107, 75, 40, 11, 4], name="f")
n = sum(freq)
alpha = 0.05
min_x = -2.9
h = 1.3
```

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

$n = 431$; $\alpha = 0.05$; $\min_x = -2.9$; $h = 1.3$

```
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	-2.9	-1.6	2
1	-1.6	-0.3	9

	x_i	x_(i+1)	f
2	-0.3	1.0	30
3	1.0	2.3	61
4	2.3	3.6	92
5	3.6	4.9	107
6	4.9	6.2	75
7	6.2	7.5	40
8	7.5	8.8	11
9	8.8	10.1	4

Шаг 1: Формулировка гипотезы

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

Шаг 2: Вычисление характеристик выборки.

Вычисление среднего значения

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

```
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
```

3.8096287703016247

Вычисление дисперсии

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

```
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
```

```
4.503567379589899
```

Вычисление стандартного отклонения

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

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

```
2.122161016414612
```

Шаг 3: Проверка гипотезы о нормальности распределения

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

$$P_i = P(x_i < X < x_{i+1}) = \Phi\left(\frac{x_{i+1} - \bar{x}}{S}\right) - \Phi\left(\frac{x_i - \bar{x}}{S}\right)$$

$$f' = fP_i$$

```
p = st.norm.cdf((data["x_(i+1)"] - mean) / std) - st.norm.cdf((data["x_i"]
- mean) / std)
```

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

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

0	1.989322
1	9.051664
2	28.600693
3	62.783364
4	95.779742
5	101.564606
6	74.862071
7	38.350836

```
8      13.650848
9      3.374717
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
```

```
1.2613069812716673
```

Шаг 5: Определение критического значения и принятие решения

Степени свободы:

```
# два параметра у норм. распр. мат. ожидание и стандартное отклонение
df = (data["f"] >= 5).sum() - 3
df
```

```
5
```

Критическое значение

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

```
11.070497693516351
```

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

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 ($\chi^2 = 1.2613069812716673$) не превышает критическое значение ($\chi^2 = 11.070497693516351$), то мы принимаем нулевую гипотезу H_0 и отвергаем альтернативную гипотезу H_1 . Это означает, что закон распределения генеральной совокупности является нормальным.