

# ЗАДАНИЕ 11

на практические занятия по дисциплине

«Прикладная теория вероятностей и статистика»

## «Статистические тесты. Критерий согласия Пирсона»

### Задание

Реализовать python программу проверки гипотезы о нормальном распределении выборки. Программа должна построить гистограмму исходных данных, теоретическую плотность вероятности, вывести значение статистики, р-значение и принимаемую гипотезу (нулевую или альтернативную) с пороговым значением  $p$  в 0.05.

В качестве исходных данных взять результаты опроса из материалов задания 7-8 в виде csv файла и проверить для каждой категории по своей группе и всему потоку гипотезу о нормальном распределении выборок.

Листинг:

```
import pandas as pd
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt
from tabulate import tabulate

# Зададим уровень значимости
alpha = 0.05

# Считаем данные из всех CSV файлов
data = pd.concat([pd.read_csv("4131.csv"),
                  pd.read_csv("4132.csv"),
                  pd.read_csv("4133.csv"),
                  pd.read_csv("4134.csv"),
                  pd.read_csv("4136.csv")])

# Выделим данные только для файла 4133.csv, который является группой
group_data = pd.read_csv("4133.csv")

# Проверим нормальность выборки по всем данным
for column in data.columns:
    sample = data[column].dropna()
    print("Группа:", column)
    print("Всего наблюдений:", len(sample))
```

```

# Построим гистограмму исходных данных
plt.hist(sample, bins=20, density=True, alpha=0.5)

# Вычислим оценки параметров нормального распределения
loc, scale = stats.norm.fit(sample)

# Вычислим теоретическую плотность вероятности
x = np.linspace(sample.min(), sample.max(), 100)
y = stats.norm.pdf(x, loc=loc, scale=scale)

# Построим теоретическую плотность вероятности
plt.plot(x, y, 'r', linewidth=2)
plt.title(column)
plt.show()

# Выполним тест согласия Пирсона
chi2, p = stats.normaltest(sample)
print(tabulate([["Статистика критерия", chi2], ["p-value", p]],
headers=["", "Значение"]))

    if p < alpha:
        print("Гипотеза о нормальности отвергается на уровне значимости",
alpha)
    else:
        print("Гипотеза о нормальности не отвергается на уровне
значимости", alpha)
        print("=" * 50)

# Проверим нормальность выборки только для группы (файл 4133.csv)
for column in group_data.columns:
    sample = group_data[column].dropna()
    print("Группа:", column)
    print("Всего наблюдений:", len(sample))

    # Построим гистограмму исходных данных
    plt.hist(sample, bins=20, density=True, alpha=0.5)

    # Вычислим оценки параметров нормального распредел
    loc, scale = stats.norm.fit(sample)

    # Вычислим теоретическую плотность вероятности
    x = np.linspace(sample.min(), sample.max(), 100)
    y = stats.norm.pdf(x, loc=loc, scale=scale)

    # Построим теоретическую плотность вероятности
    plt.plot(x, y, 'r', linewidth=2)
    plt.title(column)
    plt.show()

    # Выполним тест согласия Пирсона
    chi2, p = stats.normaltest(sample)
    print(tabulate([["Статистика критерия", chi2], ["p-value", p]],
headers=["", "Значение"]))

    if p < alpha:
        print("Гипотеза о нормальности отвергается на уровне значимости",
alpha)
    else:
        print("Гипотеза о нормальности не отвергается на уровне
значимости", alpha)
        print("=" * 50)

```

Вывод:

/Users/andrey/Documents/PyCharm/pythonProject/bin/python

/Users/andrey/Documents/PyCharm/pythonProject/main.py

Группа: category\_1

Всего наблюдений: 123

Значение

-----

Статистика критерия 7.5981

p-value 0.022392

Гипотеза о нормальности отвергается на уровне значимости 0.05

=====

Группа: category\_2

Всего наблюдений: 123

Значение

-----

Статистика критерия 8.09548

p-value 0.0174618

Гипотеза о нормальности отвергается на уровне значимости 0.05

=====

Группа: category\_3

Всего наблюдений: 123

Значение

-----

Статистика критерия 6.28932

p-value 0.0430816

Гипотеза о нормальности отвергается на уровне значимости 0.05

=====

Группа: category\_4

Всего наблюдений: 123

Значение

-----

Статистика критерия 5.04546

p-value 0.0802403

Гипотеза о нормальности не отвергается на уровне значимости 0.05

=====

Группа: category\_5

Всего наблюдений: 123

Значение

-----

Статистика критерия 7.30849

p-value 0.025881

Гипотеза о нормальности отвергается на уровне значимости 0.05

=====

Группа: category\_1

Всего наблюдений: 23

Значение

-----

Статистика критерия 1.07284

p-value 0.584837

Гипотеза о нормальности не отвергается на уровне значимости 0.05

=====

Группа: category\_2

Всего наблюдений: 23

Значение

-----

Статистика критерия 1.40356

p-value 0.495701

Гипотеза о нормальности не отвергается на уровне значимости 0.05

=====

Группа: category\_3

Всего наблюдений: 23

Значение

-----

Статистика критерия 5.85865

p-value 0.0534331

Гипотеза о нормальности не отвергается на уровне значимости 0.05

=====

Группа: category\_4

Всего наблюдений: 23

Значение

-----

Статистика критерия 2.34047

p-value 0.310293

Гипотеза о нормальности не отвергается на уровне значимости 0.05

=====

Группа: category\_5

Всего наблюдений: 23

Значение

-----

Статистика критерия 0.501969

p-value 0.778035

Гипотеза о нормальности не отвергается на уровне значимости 0.05

=====

Process finished with exit code 0











