

## ЗАДАНИЕ 12

на практические занятия по дисциплине  
«Прикладная теория вероятностей и статистика»

### «Дисперсионный анализ»

#### Задание

Реализовать python программу для проверки наличия связи между номером группы и результатами опроса по каждой категории (данные опросов взять из 7-8 заданий). Для каждой категории программа должна:

- вывести название категории;
- среднее значение по каждой группе;
- среднее значение по всему потоку;
- результаты дисперсионного анализа (значение статистики и p-значение);
- имеется ли связь между номером группы и результатами опросов при уровне значимости (пороговом значении ) 0.05 (т.е. нулевая гипотеза принята или отвергнута?)

#### Листинг:

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

def read_csv_files(file_names):
    data = {}
    for file_name in file_names:
        group_number = file_name.split('.')[0]
        data[group_number] = pd.read_csv(file_name)
    return data

def perform_anova(data, category, alpha=0.05):
    group_means = []
    all_data = []
```

```

for group_number, df in data.items():
    group_mean = df[category].mean()
    group_means.append((group_number, group_mean))
    all_data.extend(df[category].tolist())

overall_mean = np.mean(all_data)

groups = [df[category].tolist() for df in data.values()]
statistic, p_value = f_oneway(*groups)

reject_null_hypothesis = p_value < alpha

return group_means, overall_mean, statistic, p_value,
reject_null_hypothesis

def print_results(data, categories):
    for category in categories:
        print(f"\nКатегория: {category}")
        group_means, overall_mean, statistic, p_value,
        reject_null_hypothesis = perform_anova(data, category)

        print("\nСредние значения по группам:")
        print(tabulate(group_means, headers=["Группа", "Среднее"]))

        print(f"\nСреднее значение по всему потоку: {overall_mean:.2f}")

        print(f"\nРезультаты дисперсионного анализа:")
        print(f"Значение статистики: {statistic:.2f}")
        print(f"p-значение: {p_value:.4f}")

        if reject_null_hypothesis:
            print("Отвергаем нулевую гипотезу: есть связь между номером
группы и результатами опросов")
        else:
            print("Принимаем нулевую гипотезу: связь между номером группы и
результатами опросов отсутствует")

file_names = ["4131.csv", "4132.csv", "4133.csv", "4134.csv", "4136.csv"]
data = read_csv_files(file_names)

categories = data[next(iter(data))].columns.tolist() # получаем список
категорий из первого файла

print_results(data, categories)

```

Вывод:

Скриншот:

```
Run: main x
/Users/andrey/Documents/PyCharm/pythonProject/bin/python /Users/andrey/Documents/PyCharm/pythonProject/main.py

Категория: category_1

Средние значения по группам:
Группа    Среднее
-----
4131      -5.45833
4132      -5.62963
4133      -5.6087
4134      -5.43478
4136      -4.38462

Среднее значение по всему потоку: -5.29

Результаты дисперсионного анализа:
Значение статистики: 0.07
p-значение: 0.9921
Принимаем нулевую гипотезу: связь между номером группы и результатами опросов отсутствует

Категория: category_2
|
Средние значения по группам:
Группа    Среднее
-----
4131      6.58333
4132     -0.111111
4133      1.69565
4134      4.6087
4136      0.461538

Среднее значение по всему потоку: 2.54

Результаты дисперсионного анализа:
Значение статистики: 1.77
```

Вывод полностью:

Категория: category\_1

Средние значения по группам:

Группа Среднее  
-----

4131 -5.45833  
4132 -5.62963  
4133 -5.6087  
4134 -5.43478  
4136 -4.38462

Среднее значение по всему потоку: -5.29

Результаты дисперсионного анализа:

Значение статистики: 0.07

p-значение: 0.9921

Принимаем нулевую гипотезу: связь между номером группы и результатами опросов отсутствует

Категория: category\_2

Средние значения по группам:

| Группа | Среднее |
|--------|---------|
|--------|---------|

| ----- | ----- |
|-------|-------|
|-------|-------|

|      |         |
|------|---------|
| 4131 | 6.58333 |
|------|---------|

|      |           |
|------|-----------|
| 4132 | -0.111111 |
|------|-----------|

|      |         |
|------|---------|
| 4133 | 1.69565 |
|------|---------|

|      |        |
|------|--------|
| 4134 | 4.6087 |
|------|--------|

|      |          |
|------|----------|
| 4136 | 0.461538 |
|------|----------|

Среднее значение по всему потоку: 2.54

Результаты дисперсионного анализа:

Значение статистики: 1.77

p-значение: 0.1393

Принимаем нулевую гипотезу: связь между номером группы и результатами опросов отсутствует

Категория: category\_3

Средние значения по группам:

| Группа | Среднее |
|--------|---------|
|--------|---------|

| ----- | ----- |
|-------|-------|
|-------|-------|

|      |      |
|------|------|
| 4131 | -5.5 |
|------|------|

|      |          |
|------|----------|
| 4132 | -7.33333 |
|------|----------|

|      |          |
|------|----------|
| 4133 | -2.30435 |
|------|----------|

|      |          |
|------|----------|
| 4134 | -3.26087 |
|------|----------|

|      |          |
|------|----------|
| 4136 | -5.11538 |
|------|----------|

Среднее значение по всему потоку: -4.80

Результаты дисперсионного анализа:

Значение статистики: 0.93

р-значение: 0.4504

Принимаем нулевую гипотезу: связь между номером группы и результатами опросов отсутствует

Категория: category\_4

Средние значения по группам:

| Группа | Среднее |
|--------|---------|
|--------|---------|

| ----- | ----- |
|-------|-------|
|-------|-------|

|      |       |
|------|-------|
| 4131 | 8.375 |
|------|-------|

|      |         |
|------|---------|
| 4132 | 6.55556 |
|------|---------|

|      |         |
|------|---------|
| 4133 | 7.08696 |
|------|---------|

|      |         |
|------|---------|
| 4134 | 7.47826 |
|------|---------|

|      |         |
|------|---------|
| 4136 | 8.61538 |
|------|---------|

Среднее значение по всему потоку: 7.62

Результаты дисперсионного анализа:

Значение статистики: 0.30

p-значение: 0.8759

Принимаем нулевую гипотезу: связь между номером группы и результатами опросов отсутствует

Категория: category\_5

Средние значения по группам:

| Группа | Среднее |
|--------|---------|
|--------|---------|

| ----- | ----- |
|-------|-------|
|-------|-------|

|      |          |
|------|----------|
| 4131 | -1.04167 |
|------|----------|

|      |          |
|------|----------|
| 4132 | 0.925926 |
|------|----------|

|      |         |
|------|---------|
| 4133 | 5.21739 |
|------|---------|

|      |         |
|------|---------|
| 4134 | 4.34783 |
|------|---------|

|      |          |
|------|----------|
| 4136 | 0.615385 |
|------|----------|

Среднее значение по всему потоку: 1.92

Результаты дисперсионного анализа:

Значение статистики: 1.29

p-значение: 0.2768

Принимаем нулевую гипотезу: связь между номером группы и результатами опросов отсутствует

Process finished with exit code 0

