

Решение домашнего задания к уроку 8 “Дисперсионный анализ. Логистическая регрессия”

```
In [1]: import numpy as np
        from scipy import stats
```

1. Задача

Провести дисперсионный анализ для определения того, есть ли различия среднего роста среди взрослых футболистов, хоккеистов и штангистов. Даны значения роста в трех группах случайно выбранных спортсменов:

Футболисты: 173, 175, 180, 178, 177, 185, 183, 182.

Хоккеисты: 177, 179, 180, 188, 177, 172, 171, 184, 180.

Штангисты: 172, 173, 169, 177, 166, 180, 178, 177, 172, 166, 170.

Решение:

$H_0 : \mu_1 = \mu_2 = \mu_3$
 $H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

$\alpha = 0.05$

```
In [2]: y1 = np.array([173, 175, 180, 178, 177, 185, 183, 182])
        y2 = np.array([177, 179, 180, 188, 177, 172, 171, 184, 180])
        y3 = np.array([172, 173, 169, 177, 166, 180, 178, 177, 172, 166, 170])
        print(f'Футболисты: {y1}')
        print(f'Хоккеисты: {y2}')
        print(f'Штангисты: {y3}')
```

Футболисты: [173 175 180 178 177 185 183 182]
Хоккеисты: [177 179 180 188 177 172 171 184 180]
Штангисты: [172 173 169 177 166 180 178 177 172 166 170]

```
In [3]: n = len(y1) + len(y2) + len(y3)
        print(f'Общее число наблюдений: {n}')
```

Общее число наблюдений: 28

```
In [4]: y1_mean = np.mean(y1)
        y2_mean = np.mean(y2)
        y3_mean = np.mean(y3)
        print(f'Среднее значение переменной Y на 1-м уровне: {y1_mean}')
```

Среднее значение переменной Y на 1-м уровне: 179.125
Среднее значение переменной Y на 2-м уровне: 178.66666666666666
Среднее значение переменной Y на 3-м уровне: 172.72727272727272

```
In [5]: y_mean = (y1_mean * len(y1) + y2_mean * len(y2) + y3_mean * len(y3)) / n
        print(f'Среднее значение переменной Y по всем значениям: {y_mean}')
```

Среднее значение переменной Y по всем значениям: 176.46428571428572

```
In [6]: SF = len(y1) * (y1_mean - y_mean)**2 + len(y2) * (y2_mean - y_mean)**2 + len(y3) * (y3_mean - y_mean)**2
        print(f'Сумма квадратов отклонений средних групповых значений от общего среднего значения: {SF}')
```

Сумма квадратов отклонений средних групповых значений от общего среднего значения: 253.9074675324678

```
In [7]: SO = ((y1 - y1_mean)**2).sum() + ((y2 - y2_mean)**2).sum() + ((y3 - y3_mean)**2).sum()
        print(f'Остаточная сумма квадратов отклонений: {SO}')
```

Остаточная сумма квадратов отклонений: 577.0568181818182

```
In [8]: S = SO + SF
        print(f'Сумма квадратов отклонений наблюдений от общего среднего: {S}')
```

Сумма квадратов отклонений наблюдений от общего среднего: 830.964285714286

```
In [9]: DF = SF / (3 - 1)
print(f'Факторная дисперсия: {DF}')
```

Факторная дисперсия: 126.9537337662339

```
In [10]: DO = SO / (n - 3)
print(f'Остаточная дисперсия: {DO}')
```

Остаточная дисперсия: 23.08227272727273

```
In [11]: FH = DF / DO
print(f'Критерий Фишера: {FH}')
```

Критерий Фишера: 5.500053450812598

Табличное значение критерия Фишера: ($\alpha = 0.05, v_1 = 2, v_2 = 25$) $FT = 3.38$

$5.5 > 3.38 \Rightarrow$ выборки имеют статистически значимые различия, принимаем гипотезу H_1 .

```
In [12]: # Проверка
stats.f_oneway(y1,y2,y3)
```

```
Out[12]: F_onewayResult(statistic=5.500053450812596, pvalue=0.010482206918698694)
```

Так как $p < \alpha$, то выборки имеют статистически значимые различия.