

Практическое задание №8

1. 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.

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

$$H_1: \exists i, j \in \{1, 2, 3\}, i \neq j, \mu_i \neq \mu_j$$

Для подтверждения гипотезы используем F-критерий Фишера

$$k_1 = 3 - 1 = 2$$

$$k_2 = (8 + 9 + 11) - 3 = 25$$

Анализ проведёт для доверительной вероятности 0,95

$$F_{\text{таб}}(2, 25, 0.05) = 3,38$$

$$\sigma_{\text{ost}}^2 = \frac{S_{\text{ost}}^2}{n - k}$$

$$S_{\text{ost}}^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

```
In [1]: import numpy as np
```

```
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])
```

```
In [3]: y = np.array([y1, y2, y3], dtype=object)
```

```
In [4]: y_means = [el.mean() for el in y]
```

```
In [14]: print('Средние значения по выборкам:\nфутболисты -', y_means[0], '\nhоккеисты -', y_means[1], '\nштангисты -', y_means[2])
Средние значения по выборкам:
футболисты - 179.125
хоккеисты - 178.66666666666666
штангисты - 172.72727272727272
```

```
In [22]: SS_ost = np.array([np.array([(yij-y_means[i])**2 for yij in y[i]]).sum() for i in range(len(y))]).sum()
print('Остаточная сумма квадратов отклонений -', SS_ost)
Остаточная сумма квадратов отклонений - 577.0568181818182
```

```
In [23]: Var_ost = SS_ost/(28-3)
print('Остаточная дисперсия -', Var_ost)
Остаточная дисперсия - 23.082227272727273
```

$$S_F^2 = \sum_{i=1}^k (\bar{y}_i - \bar{Y})^2 * n_i$$

$$\sigma_F^2 = \frac{S_F^2}{k - 1}$$

```
In [34]: Y_mean = np.array([y[i].sum() for i in range(len(y))]).sum()/28
print('Среднее общей совокупности -', Y_mean)

Среднее общей совокупности - 176.46428571428572

In [35]: SS_fac = np.array([np.array([(y_means[i] - Y_mean)**2]*len(y[i])) for i in range(len(y))]).sum()
print('Сумма квадратов отклонений факторная -', SS_fac)

Сумма квадратов отклонений факторная - 253.9074675324678

In [36]: Var_fac = SS_fac/(3-1)
print('Факторная дисперсия -', Var_fac)

Факторная дисперсия - 126.9537337662339

In [37]: F = Var_fac/Var_ost
print('Критерий расчётный -', F)

Критерий расчётный - 5.500053450812598
```

$$F_{\text{таб}} = 3,38 < F_{\text{расч}} = 5,5$$

Рост спортсменов зависит от спорта (гипотеза H_1)

(различие между группами статистически значимое с доверительной вероятностью $p = 0,95$)

Подтверждение расчётов с помощью библиотеки `scipy`

```
In [38]: from scipy import stats

In [39]: stats.f_oneway(y1,y2,y3)

Out[39]: F_onewayResult(statistic=5.500053450812596, pvalue=0.010482206918698694)
```

p-value позволяет сделать вывод, что различие между группами становится статистически значимым при доверительной вероятности $p < 0,99$