

Практическое задание 8

In [1]:

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import warnings
5 import scipy
6 from scipy import stats
7 from sympy import symbols, diff
8 import math
9 warnings.filterwarnings('ignore')
```

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.

In [2]:

```
1 fut = np.array([173, 175, 180, 178, 177, 185, 183, 182], dtype=np.float64)
2 hoc = np.array([177, 179, 180, 188, 177, 172, 171, 184, 180], dtype=np.float64)
3 sht = np.array([172, 173, 169, 177, 166, 180, 178, 177, 172, 166, 170])
```

In [3]:

```
1 n1 = len(fut)
2 n2 = len(hoc)
3 n3 = len(sht)
4 n = n1 + n2 + n3
5 print(n)
```

28

Всего три группы:

In [4]:

```
1 k = 3
```

Проведем однофакторный дисперсионный анализ. Сначала найдем средние зарплаты для каждой профессии:

In [5]:

```
1 fut_mean = np.mean(fut)
2 print(fut_mean)
3
```

179.125

In [6]:

```
1 hoc_mean = np.mean(hoc)
2 print(hoc_mean)
```

178.66666666666666

In [7]:

```
1 sht_mean = np.mean(sht)
2 print(sht_mean)
```

172.72727272727272

Видно, что средние зарплаты разнятся. Установим, что это отличие статистически значимо. Для этого сначала соберем все значения заработных плат в один массив:

In [8]:

```
1 y_all = np.concatenate([fut, hoc, sht])
2 y_all
```

Out[8]:

```
array([173., 175., 180., 178., 177., 185., 183., 182., 177., 179., 180.,
       188., 177., 172., 171., 184., 180., 172., 173., 169., 177., 166.,
       180., 178., 177., 172., 166., 170.])
```

Найдем среднее значение заработной платы по всем значениям:

In [9]:

```
1 y_mean = np.mean(y_all)
2 print(y_mean)
```

176.46428571428572

Найдем S^2 — сумму квадратов отклонений наблюдений от общего среднего:

In [10]:

```
1 s2 = np.sum((y_all - y_mean)**2)
2 s2
```

Out[10]:

830.9642857142854

Найдем S_F^2 - сумму квадратов отклонений средних групповых значений от общего среднего:

In [11]:

```
1 s2_f = ((fut_mean - y_mean)**2) * n1 + ((hoc_mean - y_mean)**2) * n2 + ((sht_mean - y_r
2 s2_f
```

Out[11]:

253.9074675324678

Найдем $S_{\text{ост}}^2$ — остаточную сумму квадратов отклонений:

In [12]:

```
1 s2_residual = ((fut - fut_mean)**2).sum() + ((hoc - hoc_mean)**2).sum() + ((sht - sht_r
2 s2_residual
```

Out[12]:

577.0568181818182

Удостоверимся, что соблюдается равенство $S^2 = S_F^2 + S_{\text{ост}}^2$:

In [13]:

```
1 print(s2)
2 print(s2_f + s2_residual)
```

830.9642857142854

830.964285714286

Найдем общую дисперсию:

In [14]:

```
1 sigma2_general = s2 / (n - 1)
2 sigma2_general
```

Out[14]:

30.776455026455015

Найдем факторную дисперсию:

In [15]:

```
1 sigma2_f = s2_f / (k - 1)
2 sigma2_f
```

Out[15]:

126.9537337662339

Найдем остаточную дисперсию:

In [16]:

```
1 sigma2_residual = s2_residual / (n - k)
2 sigma2_residual
```

Out[16]:

23.08227272727273

Вычислим F_H :

In [17]:

```
1 F_h = sigma2_f / sigma2_residual
2 F_h
```

Out[17]:

5.500053450812598

Найдем значение $F_{\text{крит}}$ в [таблице критических точек распределения Фишера-Снедекора](https://www.matburo.ru/tv/table_fisher.pdf) (https://www.matburo.ru/tv/table_fisher.pdf) для заданного уровня значимости $\alpha = 0.05$ и двух степеней свободы:

$$df_{\text{межд}} = k - 1 = 3 - 1 = 2 \text{ и } df_{\text{внутр}} = n - k = 28 - 3 = 25.$$

Для данных значений $F_{\text{крит}} = 3.38$. Так как $F_H > F_{\text{крит}}$, различие среднего роста среди взрослых футболистов, хоккеистов и штангистов статистически значимо.

Проверим данные расчеты с помощью библиотеки питон

In [18]:

```
1 stats.f_oneway(fut, hoc, sht)
```

Out[18]:

F_onewayResult(statistic=5.500053450812596, pvalue=0.010482206918698694)

Мы видим, что статистика рассчитана правильно. Кроме того, также видно, что $\alpha > p\text{-value}$. Следовательно верна альтернативная гипотеза, т.е. различие среднего роста среди взрослых футболистов, хоккеистов и штангистов статистически значимо