**Домашняя работа №8**

**Задача №1**

**Код на Python:**

*#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.***import** numpy **as** np  
  
**class** Sample:  
 **def** \_\_init\_\_(self,data,name):  
 self.data = np.array(data)  
 self.name = name  
 self.length = self.data.size  
 self.mean = np.mean(self.data)  
 self.tss = np.sum((self.data - self.mean) \*\* 2)  
  
footbal = Sample([173, 175, 180, 178, 177, 185, 183, 182], **'футболисты'**)  
hockey = Sample([177, 179, 180, 188, 177, 172, 171, 184, 180], **'хокеисты'**)  
weightlift = Sample([172, 173, 169, 177, 166, 180, 178, 177, 172, 166, 170], **'штангисты'**)  
  
overall = Sample(np.concatenate([footbal.data,hockey.data,weightlift.data]),**'общее'**)  
  
print(footbal.\_\_dict\_\_)  
print(hockey.\_\_dict\_\_)  
print(weightlift.\_\_dict\_\_)  
print(overall.\_\_dict\_\_)

**Вывод:**

{'data': array([173, 175, 180, 178, 177, 185, 183, 182]), 'name': 'футболисты', 'length': 8, 'mean': 179.125, 'tss': 118.875}

{'data': array([177, 179, 180, 188, 177, 172, 171, 184, 180]), 'name': 'хокеисты', 'length': 9, 'mean': 178.66666666666666, 'tss': 227.99999999999997}

{'data': array([172, 173, 169, 177, 166, 180, 178, 177, 172, 166, 170]), 'name': 'штангисты', 'length': 11, 'mean': 172.72727272727272, 'tss': 230.1818181818182}

{'data': 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]), 'name': 'общее', 'length': 28, 'mean': 176.46428571428572, 'tss': 830.9642857142854}

**Пояснение:**

*Видим, что средние значения незначительно отличаются друг от друга и от общего среднего в том числе. Т.к. количество наблюдений не достаточно велико мы не можем использовать статику хи-квадрат поэтому будем использовать статистику Фишера в предположении о нормальности ошибок измерения.*

*Сформулируем H\_0 гипотезу о равенстве средних групповых значений.*

*Чтобы проверить эту гипотезу, расчитаем наблюдаемое значение статистики Фишера со степенями свободы dfвнеш = k - 1 = 3 - 1 = $ и df\_внутр = n - k = 28 - 3 = 25.*

*Для этого найдём оценку дисперсии фактороной и остаточной дисперсий.*

**Код продолжение:**  
  
ESS = sum([footbal.length \* (footbal.mean - overall.mean) \*\* 2, hockey.length \* (hockey.mean - overall.mean) \*\* 2,weightlift.length \* (weightlift.mean - overall.mean) \*\* 2])  
RSS = sum([footbal.tss, hockey.tss, weightlift.tss])  
TSS = overall.tss  
  
var\_f = ESS/(3 - 1)  
var\_r = RSS/ (overall.length - 3)  
  
print(TSS, ESS + RSS)  
print(**'var\_f'**,var\_f)  
print(**'var\_r'**,var\_r)  
print(**'F\_obs'**,var\_f / var\_r)  
  
  
**from** scipy.stats **import** f  
  
print(**'Уровень надёжности'**, f.cdf(5.5, 2, 25))  
print(**'Уровень значимости'**, 1 - f.cdf(5.5, 2, 25))

**Вывод:**

830.9642857142854 830.964285714286

var\_f 126.9537337662339

var\_r 23.08227272727273

F\_obs 5.500053450812598

*Найдем какому уровню значимости соответствует данное значение статистики.*

**Код продолжение:**

**from** **scipy.stats** **import** f print('Уровень надёжности', f.cdf(5.5, 2, 25)) print('Уровень значимости', 1 - f.cdf(5.5, 2, 25))

**Вывод:**

Уровень надёжности 0.9895174039896039

Уровень значимости 0.010482596010396139

*Таким образом, при уровне значимости большем, чем 0.0105 гипотеза H0 не будет отвергаться, при уровне значимости меньшем, чем 0.0105 гипотеза H0 будет отвергаться.*