26.02.2023 Теория вероятностей и математическая статистика

Тамбовцев Р.

ДЗ №8

Задание.

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. alpha = 0.05.

Решение:

Ввод [90]:
```
import numpy as np # импортируем библиотеку numpy для работы с функциями и массивами
```

1. Посчитаем кол-во человек в группах и общее кол-во.

Ввод [92]:
```
soccer = 8
hock = 9
weightlifter = 11

group = [soccer, hock, weightlifter]

total = soccer + hock + weightlifter
print(total)
```

```
28
```

кол-во групп k

Ввод [93]:
```
k = len(group)
print(k)
```

```
3
```

Рост: 1. Футболистов 2. Хоккеистов 3. Штангистов

Созданим массивы из значений роста

Ввод [94]:
```
y_s = np.array([173, 175, 180, 178, 177, 185, 183, 182], dtype=np.float64)
y_h = np.array([177, 179, 180, 188, 177, 172, 171, 184, 180], dtype=np.float64)
y_f = np.array([172, 173, 169, 177, 166, 180, 178, 177, 172, 166, 170], dtype=np.float64)
```

Проведем однофакторный дисперсионный анализ.

1. Найдем средние значения роста по каждой группе спортсменов:

Ввод [95]:
```
y_s_mean = round(np.mean(y_s),3)
y_h_mean = round(np.mean(y_h),3)
y_f_mean = round(np.mean(y_f),3)
print(y_s_mean,
      y_h_mean,
      y_f_mean)
```

```
179.125 178.667 172.727
```

Значения среднего роста имеют отличия. Установим, что эти отличия статистически значимы. Соберём все значения роста в один массив:

Ввод [96]:
```
y_total = np.concatenate([y_s, y_h, y_f])
y_total
```

Out[96]:
```
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.])
```

определим среднее значение роста по всем трём группам

Ввод [97]:
```
y_mean_total = round(np.mean(y_total),3)
print(y_mean_total)
```

```
176.464
```

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

$$S^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \overline{Y})^2$$

Ввод [98]:
```
S2 = np.sum((y_total - y_mean_total)**2)
round(S2,3)
```

Out[98]:
```
830.964
```

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

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

Ввод [99]:
```
s2_f = ((y_s_mean - y_mean_total)**2) * soccer + ((y_h_mean - y_mean_total)**2) * hock + ((y_f_mean - y_mean_total)**2) * weightli
round(s2_f,3)
```
Out[99]: 253.943

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

$$S^2_{\text{ост}} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2$$

Ввод [100]:
```
S2_ost = np.sum((y_s - y_s_mean)**2) + np.sum((y_h- y_h_mean)**2) + np.sum((y_f - y_f_mean)**2)
round(S2_ost,3)
```
Out[100]: 577.057

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

Ввод [89]:
```
print(S2)
S2_check = s2_f + S2_ost
print(S2_check)
```
```
830.9642879999998
830.9999279999997
```

Найдём общую дисперсию

Ввод [101]:
```
sigma2_general = S2 / (total - 1)
sigma2_general
```
Out[101]: 30.776455111111105

Найдём факторную дисперсию

Ввод [102]:
```
sigma2_f = s2_f / (k - 1)
sigma2_f
```
Out[102]: 126.97155399999988

Найдем остаточную дисперсию:

Ввод [105]:
```
sigma2_ost = S2_ost / (total - k)
sigma2_ost
```
Out[105]: 23.082272800000002

Вычислим F_H:

Ввод [111]:
```
a = 0.05
print(k-1)
print(total-k)
```
```
2
25
```

По таблице Критические точки распределения Фишера–Снедекора определяем $F_{\text{крит}}$: 3,38

Ввод [110]:
```
F_h = sigma2_f / sigma2_ost
F_h
```
Out[110]: 5.500825464639681

Если значение F_H превышает $F_{\text{крит}}$ из таблицы критических точек распределения Фишера-Снедекора для заданного уровня

значимости α двух степеней свободы $df_{\text{межд}} = k - 1$ (относится к числителю соотношения) и $df_{\text{внутр}} = n - k$ (относится к знаменателю),

то выборки имеют разные средние значения.

Так как F_h > Fкрит, различие значениях среднего роста в трех группах статистически значимо.

Также вычислим эмпирическое корреляционное отношение η^2:

Ввод [113]:
```
eta2 = s2_f / S2
eta2
```
Out[113]: 0.3056005073469533

Принято считать, что при значениях η^2 ниже 0.2-0.3 групповые значения

средних не имеют статистически достоверного отличия.

Ввод [ ]:
```
т.к. полученное значение больше 0.3, то считаем, что групповые значения средних имеют статистически достоверные
отличия
```