

Урок 8. Дисперсионный анализ. Логистическая регрессия

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.

Решение,

$$K = 3$$

$$n_1 = 8 \text{ (футболисты)}$$

$$n_2 = 9 \text{ (хоккеисты)}$$

$$n_3 = 11 \text{ (штангисты)}$$

$$n = n_1 + n_2 + n_3 = 8 + 9 + 11 = 28$$

1) Найдем среднее в каждой группе \bar{y}_i

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

$$\bar{y}_1 = 179.125$$

$$\bar{y}_2 = 178.667$$

$$\bar{y}_3 = 172.727$$

$$(173 + 175 + 180 + 178 + 177 + 185 + 183 + 182) / 8 = 179.125$$

$$(177 + 179 + 180 + 188 + 177 + 172 + 171 + 184 + 180) / 9 \approx 178.667$$

$$(172 + 173 + 169 + 177 + 166 + 180 + 178 + 177 + 172 + 166 + 170) / 11 \approx 172.727$$

2) Найдем \bar{Y} - среднее значение переменной Y по всем значениям;

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} = \frac{1}{n} \sum_{i=1}^k \bar{y}_i \cdot n_i$$

$$\bar{Y} = 176,464$$

$$(179,125 \times 8 + 178,667 \times 9 + 172,727 \times 11) / 28 \approx 176.464$$

3) Найдем S^2 - сумму квадратов отклонений наблюдений от общего среднего;

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

$$S^2 \approx 831$$

4) Найдем S_F^2 - сумму квадратов отклонений средних групповых значений от общего среднего значения \bar{Y} ;

$$S_F^2 = \sum_{i=1}^k (\bar{y}_i - \bar{Y})^2 \cdot n_i \quad S_F^2 = 253,943$$

$$(179,125 - 176,464)^2 \times 8 + (178,667 - 176,464)^2 \times 9 + (172,727 - 176,464)^2 \times 11 \approx 253,9$$

5) Найдем $S_{ост}^2$ - остаточную сумму квадратов отклонений

$$S_{ост}^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$$S_{ост}^2 = (173 - 179,125)^2 + (175 - 179,125)^2 + \dots + (170 - 172,727)^2$$

$$S_{ост}^2 = 577,057$$

$$S^2 = S_F^2 + S_{ост}^2$$

$$S^2 = 253,943 + 577,057 \approx 831$$

6) Наблюдаемый критерий Фишера F_H

$$F_H = \frac{S_F^2}{S_{ост}^2}$$

Вычисляем факторную дисперсию:

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

$$S_F^2 = 253,943 / (3-1) = 126,972$$

Вычисляем остаточную дисперсию:

$$S_{ост}^2 = \frac{S_{ост}^2}{n-k} = \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

$$S_{ост}^2 = 577,057 / (28-3) = 23,082$$

$$F_H = 126,972 / 23,082 = 5,501$$

7) Найдем значение $F_{крит}$ в таблице критических точек распределения Фишера-Снедекора для заданного уровня значимости $\alpha = 0,05$ и двух степеней свободы:

$$df_{факт} = k-1 = 3-1 = 2$$

$$df_{выбр} = n-k = 28-3 = 25$$

$$F_{крит} = 3,38$$

$$5,501 > 3,38$$

\Rightarrow различие средних роста между группами статистически значимо