

Данные группируются по признаку-фактору. Затем по каждой группе рассчитывается среднее значение. Задача состоит в том, чтобы увидеть, есть ли связь между признаками или нет; прямая связь или обратная; линейная или нелинейная.

Так как в основание группировки положен непрерывный количественный признак, то число групп определяют одновременно с размером интервала.

Когда совокупность единиц более или менее однородна (вариация по группировочному признаку мала), прибегают к равным интервалам, размер которых приблизительно определяется по формуле Стёрджесса:

$$n = 1 + 3,322 \log n = 1 + 3,322 \log(25) = 6$$

Ширина интервала составит:

$$h = \frac{X_{max} - X_{min}}{n} = \frac{58,5 - 25}{6} = 5,583 \approx 5,6$$

Для каждого значения ряда подсчитаем, какое количество раз оно попадает в тот или иной интервал. Для этого сортируем ряд по возрастанию.

X	Интервал	Количество	Y ₁	Y ₂
25	25 - 30.6	1	31.4	11.1
27	25 - 30.6	2	40.5	10.7
27. 2	25 - 30.6	3	45.9	10.7
27. 7	25 - 30.6	4	41.3	8.3
28. 3	25 - 30.6	5	47.5	10.4
28. 6	25 - 30.6	6	45.4	8.3
28. 7	25 - 30.6	7	46.3	10.7
29	25 - 30.6	8	43.5	8.3
29. 9	25 - 30.6	9	31	11
32. 6	30.6 - 36.2	1	42.5	8.3
34	30.6 - 36.2	2	41	8
34	30.6 - 36.2	3	38.9	9.3
35. 2	30.6 - 36.2	4	40	11.6
36	30.6 - 36.2	5	44.7	8
39.	36.2 - 41.8	1	45.6	11

2				
40. 8	36.2 - 41.8	2	44.7	8
41. 2	36.2 - 41.8	3	53.8	16
41. 9	41.8 - 47.4	1	48.5	12.1
43. 1	41.8 - 47.4	2	44.7	8
43. 6	41.8 - 47.4	3	49.7	13.8
45	41.8 - 47.4	4	48.7	14
51. 2	47.4 - 53	1	52.3	11.5
52. 2	47.4 - 53	2	52.3	15.3
53. 9	53 - 58.6	1	56	22
58. 5	53 - 58.6	2	55.2	25

Аналитическая группировка по X и Y₁.

Группы	№	Кол-во, n _i	ΣX	X _{ср} = ΣX _i / n _i	ΣY	Y _{ср} = ΣY _i / n _i
25 - 30.6	3,4,6,11,7,10,5,9, 1	9	251.4	27.93	372.8	41.42
30.6 - 36.2	23,14,17,2,22	5	171.8	34.36	207.1	41.42
36.2 - 41.8	24,21,25	3	121.2	40.4	144.1	48.03
41.8 - 47.4	12,20,18,8	4	173.6	43.4	191.6	47.9
47.4 - 53	16,19	2	103.4	51.7	104.6	52.3
53 - 58.6	13,15	2	112.4	56.2	111.2	55.6
Итого		25	933.8		1131.4	

Аналитическая группировка по X и Y₂.

Группы	№	Кол-во, n_j	$\sum X$	$X_{cp} = \sum X_j / n_j$	$\sum Y$	$Y_{cp} = \sum Y_j / n_j$
25 - 30.6	3,4,6,11,7,10,5,9,1	9	251. 4	27.93	89.5	9.94
30.6 - 36.2	23,14,17,2,22	5	171. 8	34.36	45.2	9.04
36.2 - 41.8	24,21,25	3	121. 2	40.4	35	11.67
41.8 - 47.4	12,20,18,8	4	173. 6	43.4	47.9	11.98
47.4 - 53	16,19	2	103. 4	51.7	26.8	13.4
53 - 58.6	13,15	2	112. 4	56.2	47	23.5
Итого		25	933. 8		291. 4	

Итоговая аналитическая группировка.

Группы	№	Кол-во, n_j	X	Y1	Y2
25 - 30.6	3,4,6,11,7,10,5,9,1	9	27.9 3	41.42	9.94
30.6 - 36.2	23,14,17,2,22	5	34.3 6	41.42	9.04
36.2 - 41.8	24,21,25	3	40.4	48.03	11.67
41.8 - 47.4	12,20,18,8	4	43.4	47.9	11.98
47.4 - 53	16,19	2	51.7	52.3	13.4
53 - 58.6	13,15	2	56.2	55.6	23.5
Итого		25			

Исходные данные

X	f_1	f_2	Итого
27.93	41.42	9.94	51.36
34.36	41.42	9.04	50.46
40.4	48.03	11.67	59.7

43.4	47.9	11.98	59.88
51.7	52.3	13.4	65.7
56.2	55.6	23.5	79.1
Итого	286.67	79.53	

Тогда имеем 2 группы, для которых необходимо рассчитать групповую среднюю и внутригрупповые дисперсии.

1. Находим средние значения каждой группы.

$$\bar{x}_j = \frac{\sum x_{ij} f_{ij}}{\sum f_{ij}}$$

$$\bar{x}_1 = \frac{27.93 \cdot 41.42 + 34.36 \cdot 41.42 + \dots + 56.2 \cdot 55.6}{286.67} = 43.35$$

$$\bar{x}_2 = \frac{27.93 \cdot 9.94 + 34.36 \cdot 9.04 + \dots + 56.2 \cdot 23.5}{79.53} = 45.18$$

2. Находим среднее квадратическое каждой группы.

$$\sigma_j^2 = \frac{\sum (x_{ij} - \bar{x}_j)^2}{\sum f_{ij}}$$

$$\sigma_1^2 = \frac{(27.93 - 43.35)^2 \cdot 41.42 + (34.36 - 43.35)^2 \cdot 41.42 + \dots + (56.2 - 43.35)^2 \cdot 55.6}{286.67} = 92.24$$

$$\sigma_2^2 = \frac{(27.93 - 45.18)^2 \cdot 9.94 + (34.36 - 45.18)^2 \cdot 9.04 + \dots + (56.2 - 45.18)^2 \cdot 23.5}{79.53} = 97.37$$

Результаты расчета сведем в таблицу:

Номер группы	Групповая средняя	Внутригрупповая дисперсия
1	43.35	92.24
2	45.18	97.37

3. Внутригрупповая дисперсия характеризует изменение (вариацию) изучаемого (результативного) признака в пределах группы под действием на него всех факторов, кроме фактора, положенного в основание группировки:

Среднюю из внутригрупповых дисперсий рассчитаем по формуле:

$$\bar{\sigma}^2 = \frac{\sum \sigma_j^2 \cdot f_j}{\sum f_j}$$

$$\bar{\sigma}^2 = \frac{92.24 \cdot 286.67 + 97.37 \cdot 79.53}{286.67 + 79.53} = 93.35$$

4. Межгрупповая дисперсия характеризует изменение (вариацию) изучаемого (результативного) признака под действием на него фактора (факторного признака), положенного в основание группировки.

Межгрупповую дисперсию определим как:

$$\delta^2 = \frac{\sum(\bar{x}_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$

где

$$\bar{x} = \frac{\sum \bar{x}_j \cdot f_j}{\sum f_j}$$

$$\bar{x} = \frac{43.35 \cdot 286.67 + 45.18 \cdot 79.53}{286.67 + 79.53} = 43.75$$

Тогда

$$\sigma^2 = \frac{(43.35 - 43.75)^2 286.67 + (45.18 - 43.75)^2 79.53}{286.67 + 79.53} = 0.57$$

Общая дисперсия характеризует изменение (вариацию) изучаемого (результативного) признака под действием на него всех без исключения факторов (факторных признаков).

Общая дисперсия будет равна: $\sigma = 93.35 + 0.57 = 93.92$

Общую дисперсию также можно рассчитать и по формуле:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$

Эмпирическое корреляционное отношение измеряет, какую часть общей колеблемости результативного признака вызывает изучаемый фактор. Это отношение факторной дисперсии к общей дисперсии:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}$$

Определяем эмпирическое корреляционное отношение:

$$\eta = \sqrt{\frac{0.57}{93.92}} = 0.0779$$

Связи между признаками могут быть слабыми и сильными (тесными). Их критерии оцениваются по шкале Чеддока:

0.1 < η < 0.3: слабая;

0.3 < η < 0.5: умеренная;

0.5 < η < 0.7: заметная;

0.7 < η < 0.9: высокая;

0.9 < η < 1: весьма высокая;

В нашем примере связь между признаком Y фактором X слабая

Коэффициент детерминации.

$$\eta^2 = \frac{\delta^2}{\sigma^2}$$

Определим коэффициент детерминации:

$$\eta^2 = \frac{0.57}{93.92} = 0.00607$$

Таким образом, на 0.61% вариация обусловлена различиями между признаками, а на 99.39% – другими факторами.