

Задача. Найти выборочное уравнение прямой линии регрессии Y на X по данным, приведенным в корреляционной таблице.

	2	6	10	14	18	22	26	30	34
11	1	-	-	-	-	-	-	-	-
13	1	2	1	-	-	-	-	-	-
15	1	3	4	2	1	-	-	-	-
17	-	1	2	5	3	2	-	-	-
19	-	-	1	3	7	5	1	1	-
21	-	-	-	1	4	5	4	3	1
23	-	-	-	-	1	1	2	2	2
25	-	-	-	-	-	-	1	1	3
27	-	-	-	-	-	-	-	-	1

Решение.

Вносим данные для расчетов, подсчитываем частоты по строкам и столбцам. Так как не указано, где переменная X , а где Y , выбираем это произвольно.

$Y \setminus X$	2	6	10	14	18	22	26	30	34	n_y
11	1									1
13	1	2	1							4
15	1	3	4	2	1					11
17		1	2	5	3	2				13
19			1	3	7	5	1	1		18
21				1	4	5	4	3	1	18
23					1	1	2	2	2	8
25							1	1	3	5
27									1	1
n_x	3	6	8	11	16	13	8	7	7	79

Составим корреляционную таблицу в условных вариантах, выбрав в качестве ложных нулей $C_1 = 18$ и $C_2 = 19$ (находятся в середине ряда).

Сделаем преобразования $u_i = \frac{x_i - C_1}{h_x}$, $v_j = \frac{y_j - C_2}{h_y}$, где $h_x = 4$ и $h_y = 2$. Получим таблицу:

$V \setminus U$	-4	-3	-2	-1	0	1	2	3	4	n_v
-4	1	0	0	0	0	0	0	0	0	1
-3	1	2	1	0	0	0	0	0	0	4
-2	1	3	4	2	1	0	0	0	0	11
-1	0	1	2	5	3	2	0	0	0	13
0	0	0	1	3	7	5	1	1	0	18
1	0	0	0	1	4	5	4	3	1	18
2	0	0	0	0	1	1	2	2	2	8
3	0	0	0	0	0	0	1	1	3	5
4	0	0	0	0	0	0	0	0	1	1
n_u	3	6	8	11	16	13	8	7	7	79

Найдем \bar{u} и \bar{v} :

$$\bar{u} = \frac{1}{n} \sum n_u \cdot u = \frac{21}{79} = 0,2658.$$

$$\bar{v} = \frac{1}{n} \sum n_v \cdot v = \frac{2}{79} = 0,0253.$$

Найдем вспомогательные величины $\overline{u^2}$ и $\overline{v^2}$:

$$\overline{u^2} = \frac{1}{n} \sum n_u \cdot u^2 = \frac{365}{79} = 4,6203.$$

$$\overline{v^2} = \frac{1}{n} \sum n_v \cdot v^2 = \frac{220}{79} = 2,7848.$$

Найдем:

$$\sigma_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{4,6203 - 0,2658^2} \approx 2,133,$$

$$\sigma_v = \sqrt{\overline{v^2} - (\bar{v})^2} = \sqrt{2,7848 - 0,0253^2} \approx 1,6686.$$

Расчетные таблицы:

u_i	-4	-3	-2	-1	0	1	2	3	4	Сумма
n_{ui}	3	6	8	11	16	13	8	7	7	79
$u_i n_{ui}$	-12	-18	-16	-11	0	13	16	21	28	21
$u_i^2 n_{ui}$	48	54	32	11	0	13	32	63	112	365

v_j	-4	-3	-2	-1	0	1	2	3	4	Сумма
n_{vj}	1	4	11	13	18	18	8	5	1	79
$v_j n_{vj}$	-4	-12	-22	-13	0	18	16	15	4	2
$v_j^2 n_{vj}$	16	36	44	13	0	18	32	45	16	220

Найдем $\sum u \cdot v \cdot n_{uv} = 238$, для чего составим расчетную таблицу:

$v \backslash u$	-4	-3	-2	-1	0	1	2	3	4	n_v	$u \cdot n_{uv}$	$u \cdot v \cdot n_{uv}$
-4	1	0	0	0	0	0	0	0	0	1	-4	16
-3	1	2	1	0	0	0	0	0	0	4	-12	36
-2	1	3	4	2	1	0	0	0	0	11	-23	46
-1	0	1	2	5	3	2	0	0	0	13	-10	10
0	0	0	1	3	7	5	1	1	0	18	5	0
1	0	0	0	1	4	5	4	3	1	18	25	25
2	0	0	0	0	1	1	2	2	2	8	19	38
3	0	0	0	0	0	0	1	1	3	5	17	51
4	0	0	0	0	0	0	0	0	1	1	4	16
n_u	3	6	8	11	16	13	8	7	7	79		
												238

Найдем выборочный коэффициент корреляции:

$$r_{\hat{a}} = \frac{\sum u \cdot v \cdot n_{uv} - n \cdot \bar{u} \cdot \bar{v}}{n \cdot \sigma_u \cdot \sigma_v} = \frac{238 - 79 \cdot 0,2658 \cdot 0,0253}{79 \cdot 2,133 \cdot 1,6686} \approx 0,845.$$

Найдем характеристики исходных случайных величин X, Y .

$$\bar{x} = \bar{u} \cdot h_x + C_1 = 0,2658 \cdot 4 + 18 = 19,063,$$

$$\bar{y} = \bar{v} \cdot h_y + C_2 = 0,0253 \cdot 2 + 19 = 19,051.$$

$$\sigma_x = h_x \sigma_u = 4 \cdot 2,133 = 8,532,$$

$$\sigma_y = h_y \sigma_v = 2 \cdot 1,6686 = 3,337.$$

Уравнение регрессии Y на X имеет вид $\bar{y}_x - \bar{y} = r_{\hat{a}} \frac{\sigma_y}{\sigma_x} (x - \bar{x})$. Подставляем все величины:

$$\bar{y}_x - 19,051 = 0,845 \frac{3,337}{8,532} (x - 19,063),$$

$$\bar{y}_x = 0,33x + 12,753.$$