Теория Вероятностей. Практическая работа №6

Марьин Григорий 466650, Вариант 9

Дано:
Дана таблица распределения 100 заводов по производственным средствам X (тыс. ден. ед.) и по суточной выработке Y(т). Известно, что между X и Y существует линейная корреляционная зависимость. Требуется:

а) Найти уравнение прямой регрессии y на x;
б) Построить уравнение эмпирической линии регрессии и случайные точки выборки (X, Y)

X\Y	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	m_x
250	3	4	5	0	0	0	0	0	12
450	0	6	2	8	0	0	0	0	16
650	0	0	0	5	14	9	0	0	28
850	0	0	0	6	8	6	0	0	20
1050	0	0	0	0	5	7	4	0	16
1250	0	0	0	0	0	0	5	3	8
m_y	3	10	7	19	27	22	9	3	100

Решение:
Для подсчета числовых характеристик (выборочных средних $\bar{x}$ и $\bar{y}$, выборочных средних квадратичных отклонений s_x и s_y и выборочного корреляционного момента s_{xy}) составляем расчетную таблицу. При заполнении таблицы осуществляем контроль по строкам и столбцам:

$$\sum_{i=1}^{6} m_{x_i} = \sum_{j=1}^{8} m_{y_j} = n = 100$$

$$\sum_{i=1}^{6}\sum_{j=1}^{8} m_{ij}x_i = \sum_{i=1}^{6} m_{x_i}x_i = 72200$$

$$\sum_{i=1}^{6}\sum_{j=1}^{8} m_{ij}y_i = \sum_{j=1}^{8} m_{y_j}y_j = 94{,}8$$

$$\sum_{i=1}^{6}\left(x_i \sum_{j=1}^{8} m_{ij}y_j\right) = 76220$$

Вычисляем выборочные средние $\bar{x}$ и $\bar{y}$, $i = \overline{1,6}; j = \overline{1,8}$:

$$\bar{x} = \frac{\sum\sum m_{ij}x_i}{n} = \frac{\sum m_{x_i}x_i}{n} = \frac{72200}{100} = 722$$

$$\bar{y} = \frac{\sum m_{y_j}y_j}{n} = \frac{94{,}8}{100} = 0{,}948$$

Выборочные дисперсии находим по формулам:

$$s_x^2 = \frac{1}{n-1}\left(\sum m_{x_i}x_i^2 - \frac{1}{n}\left(\sum m_{x_i}x_i\right)^2\right) = \frac{1}{99}\left(60410000 - \frac{1}{100}(72200)^2\right) = 83652{,}52525$$

$$s_y^2 = \frac{1}{n-1}\left(\sum m_{y_j}y_j^2 - \frac{1}{n}\left(\sum m_{y_j}y_j\right)^2\right) = \frac{1}{99}\left(100{,}4 - \frac{1}{100}(94{,}8)^2\right) = 0{,}10635959$$

	j	1	2	3	4	5	6	7	8	9	10	11	12	13
i	X\Y	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	m_{x_i}	$m_{x_i} * x_i$	$\sum_{j=1}^{k} m_{y_j} y_j$	$x_i^2 m_{x_i}$	$x_i \sum_{j=1}^{k} m_{ij} y_j$
1	250	3	4	5	0	0	0	0	0	12	3000	5,2	750000	1300
2	450	0	6	2	8	0	0	0	0	16	7200	10	3240000	4500
3	650	0	0	0	5	14	9	0	0	28	18200	28,8	11830000	18720
4	850	0	0	0	6	8	6	0	0	20	17000	20	14450000	17000
5	1050	0	0	0	0	5	7	4	0	16	16800	19	17640000	19950
6	1250	0	0	0	0	0	0	5	3	8	10000	11,8	12500000	14750
7	m_{y_j}	3	10	7	19	27	22	9	3	100	72200	94,8	60410000	76220
8	$m_{y_j} * y_j$	0,6	4	4,2	15,2	27	26,4	12,6	4,8	94,8	-	-	-	-
9	$\sum_{i=1}^{m} m_{ij} x_i$	750	3700	2150	11950	21150	18300	10450	3750	72200	-	-	-	-
10	$y_j^2 m_{ij}$	0,12	1,6	2,52	12,16	27	31,68	17,64	7,68	100,4	-	-	-	-
11	$y_j \sum_{i=1}^{m} m_{ij} x_i$	150	1480	1290	9560	21150	21960	14630	6000	76220	-	-	-	-

Корреляционный момент вычисляем по формуле:

$$S_{xy} = \frac{1}{n-1}\left(\sum\sum m_{ij}x_i y_j - \frac{1}{n}\left(\sum m_{x_i}x_i\right)\left(\sum m_{y_j}y_j\right)\right) = \frac{1}{99}\left(76220 - \frac{1}{100}(94{,}8 * 72200)\right) \approx 78{,}53$$

Оценкой теоретической линии регрессии является эмпирическая линия регрессии, уравнение которой имеет вид

$$y = \bar{y} + r_{xy}\frac{S_y}{S_x}(x - \bar{x}),$$

где $S_x = \sqrt{83652{,}53} \approx 289{,}23; S_y = \sqrt{0{,}1064} \approx 0{,}3261;$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{78{,}53}{289{,}23 * 0{,}3261} = \approx 0{,}833$$

Составляем уравнение эмпирической линии регрессии у на x:

$$y = 0{,}948 + 0{,}833 * \frac{0{,}3261}{289{,}23}(x - 722)$$

$$y = 0{,}000939188x + 0{,}269906$$