

Дано:

Дана таблица распределения 100 автомашин по затратам на перевозки X (ден. ед.) и по протяженности маршрутов перевозок Y (км). Известно, что между X и Y существует линейная корреляционная зависимость. Требуется:

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

$X \backslash Y$	20	40	60	80	100	120	140	160	m_x
1000	2	7	3	–	–	–	–	–	12
2000	–	6	4	5	–	–	–	–	15
3000	–	–	8	9	7	–	–	–	24
4000	–	–	–	7	14	5	–	–	26
5000	–	–	–	–	5	7	4	–	16
6000	–	–	–	–	–	–	4	3	7
m_y	2	13	15	21	26	12	8	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:

$$\begin{aligned}\sum_{i=1}^6 m_{x_i} &= \sum_{j=1}^8 m_{y_j} = n = 100 \\ \sum_{i=1}^6 \sum_{j=1}^8 m_{ij} x_i &= \sum_{i=1}^6 m_{x_i} x_i = 340000 \\ \sum_{i=1}^6 \sum_{j=1}^8 m_{ij} y_j &= \sum_{j=1}^8 m_{y_j} y_j = 8780 \\ \sum_{i=1}^6 \left(x_i \sum_{j=1}^8 m_{ij} y_j \right) &= \sum_{j=1}^8 \left(y_j \sum_{i=1}^6 m_{ij} x_i \right) = 33940000\end{aligned}$$

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

$$\begin{aligned}\bar{x} &= \frac{\sum \sum m_{ij} x_i}{n} = \frac{\sum m_{x_i} x_i}{n} = \frac{340000}{100} = 3400 \\ \bar{y} &= \frac{\sum m_{y_j} y_j}{n} = \frac{8780}{100} = 87,8\end{aligned}$$

Выборочные дисперсии находим по формулам:

$$\begin{aligned}s_x^2 &= \frac{1}{n-1} \left(\sum m_{x_i} x_i^2 - \frac{1}{n} (\sum m_{x_i} x_i)^2 \right) = \frac{1}{99} \left(1356000000 - \frac{1}{100} (340000)^2 \right) = 2020202,02 \\ s_y^2 &= \frac{1}{n-1} \left(\sum m_{y_j} y_j^2 - \frac{1}{n} (\sum m_{y_j} y_j)^2 \right) = \frac{1}{99} \left(876400 - \frac{1}{100} (8780)^2 \right) = 1065,82\end{aligned}$$

	j	1	2	3	4	5	6	7	8	9	10	11	12	13
i	$X \backslash Y$	20	40	60	80	100	120	140	160	m_x	$m_x x_i$	$\sum_{j=1}^k m_{ij} y_j$	$m_{x_i} x_i^2$	$x_i \sum_{j=1}^k m_{ij} y_j$
1	1000	2	7	3	-	-	-	-	-	12	12000	500	12000000	500000
2	2000	-	6	4	5	-	-	-	-	15	30000	880	60000000	1760000
3	3000	-	-	8	9	7	-	-	-	24	72000	1900	216000000	5700000
4	4000	-	-	-	7	14	5	-	-	26	104000	2560	416000000	10240000
5	5000	-	-	-	-	5	7	4	-	16	80000	1900	400000000	9500000
6	6000	-	-	-	-	-	-	4	3	7	42000	1040	252000000	6240000
7	m_y	2	13	15	21	26	12	8	3	100	340000	8780	1356000000	33940000
8	$m_y y_j$	40	520	900	1680	2600	1440	1120	480	8780	-	-	-	-
9	$\sum_{i=1}^m m_{x_i} x_i$	2000	19000	35000	65000	102000	55000	44000	18000	340000	-	-	-	-
10	$m_{ij} y_j^2$	800	20800	54000	134400	260000	172800	156800	76800	876400	-	-	-	-
11	$y_j \sum_{i=1}^m m_{x_i} x_i$	40000	760000	2100000	5200000	10200000	6600000	6160000	2880000	33940000	-	-	-	-

Корреляционный момент вычисляем по формуле:

$$S_{xy} = \frac{1}{n-1} \left(\sum \sum m_{ij} x_i y_j - \frac{1}{n} (\sum m_{xi} x_i) (\sum m_{yj} y_j) \right) = \frac{1}{99} \left(33940000 - \frac{1}{100} (340000 * 8780) \right) \approx 41292,93$$

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 $S_x = \sqrt{2020202,02} \approx 1421,33811$; $S_y = \sqrt{1065,82} \approx 32,6469$;

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{41292,93}{1421,33811 * 32,6469} \approx 0,88989$$

Составляем уравнение эмпирической линии регрессии y на x :

$$y = 87,8 + 0,88989 * \frac{32,6469}{1421,33811} (x - 3400)$$

$$y = 0,02044x + 18,304$$

