Касьяненко Вера (P3220, Теор.Вероятн. 5.1)

ИДЗ-19.2 (вариант 5)

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:

a) Найти уравнение прямой регрессии y и x;

b) Построить уравнение эмпирической линии регрессии и случайные точки выборки (X, Y)

$X \backslash Y$	64	72	80	88	96	104	112	120	m_x
1,0	6	2	4	–	–	–	–	–	12
1,3	–	3	8	6	–	–	–	–	17
1,6	–	–	–	8	14	5	–	–	27
1,9	–	–	–	7	8	9	–	–	24
2,2	–	–	–	–	4	5	6	–	15
2,5	–	–	–	–	–	1	1	3	5
m_y	6	5	12	21	26	20	7	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:

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

$$\sum_{i=1}^{6}\sum_{j=1}^{8} m_{ij}x_i = \sum_{i=1}^{6} m_{x_i}x_i = 168{,}4$$

$$\sum_{i=1}^{6}\sum_{j=1}^{8} m_{ij}y_i = \sum_{j=1}^{8} m_{y_j}y_j = 9272$$

$$\sum_{i=1}^{6}\left(x_i\sum_{j=1}^{8} m_{ij}y_j\right) = \sum_{j=1}^{8}\left(y_j\sum_{i=1}^{6} m_{ij}x_i\right) = 16071{,}2$$

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

$$\bar{x} = \frac{\sum\sum m_{ij}x_i}{n} = \frac{\sum m_{x_i}x_i}{n} = \frac{168{,}4}{100} = 1{,}684$$

$$\bar{y} = \frac{\sum m_{y_j}y_j}{n} = \frac{9272}{100} = 92{,}72$$

Выборочные дисперсии находим по формулам:

$$s_x^2 = \frac{1}{n-1}\left(\sum m_{x_i}x_i^2 - \frac{1}{n}\left(\sum m_{x_i}x_i\right)^2\right) = \frac{1}{99}\left(300 - \frac{1}{100}(168{,}4)^2\right) = 0{,}1658$$

$$s_y^2 = \frac{1}{n-1}\left(\sum m_{y_j}y_j^2 - \frac{1}{n}\left(\sum m_{y_j}y_j\right)^2\right) = \frac{1}{99}\left(876864 - \frac{1}{100}(9272)^2\right) = 173{,}3753$$

	j	1	2	3	4	5	6	7	8	9	10	11	12	13
i	X \ Y	64	72	80	88	96	104	112	120	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	1,0	6	2	4	–	–	–	–	–	12	12	848	12	848
2	1,3	–	3	8	6	–	–	–	–	17	22,1	1384	29	1799,2
3	1,6	–	–	–	8	14	5	–	–	27	43,2	2568	69	4108,8
4	1,9	–	–	–	7	8	9	–	–	24	45,6	2320	87	4408
5	2,2	–	–	–	–	4	5	6	–	15	33	1576	73	3467,2
6	2,5	–	–	–	–	–	1	1	3	5	12,5	576	31	1440
7	m_y	6	5	12	21	26	20	7	3	100	168,4	9272	300	16071,2
8	$m_y y_j$	384	360	960	1848	2496	2080	784	360	9272	–	–	–	–
9	$\sum_{i=1}^{m} m_{x_i} x_i$	6	5,9	14,4	33,9	46,4	38,6	15,7	7,5	168,4	–	–	–	–
10	$m_{ij} y_j^2$	24576	25920	76800	162624	239616	216320	87808	43200	876864	–	–	–	–
11	$y_j \sum_{i=1}^{m} m_{x_i} x_i$	384	424,8	1152	2983,2	4454,4	4014,4	1758,4	900	16071,2	–	–	–	–

Корреляционный момент вычисляем по формуле:

$$S_{xy} = \frac{1}{n-1}\left(\sum\sum m_{ij}x_i y_j - \frac{1}{n}\left(\sum m_{x_i}x_i\right)\left(\sum m_{y_j}y_j\right)\right) = \frac{1}{99}\left(16071{,}2 - \frac{1}{100}(168{,}4 * 9272)\right) \approx 4{,}6177$$

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

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

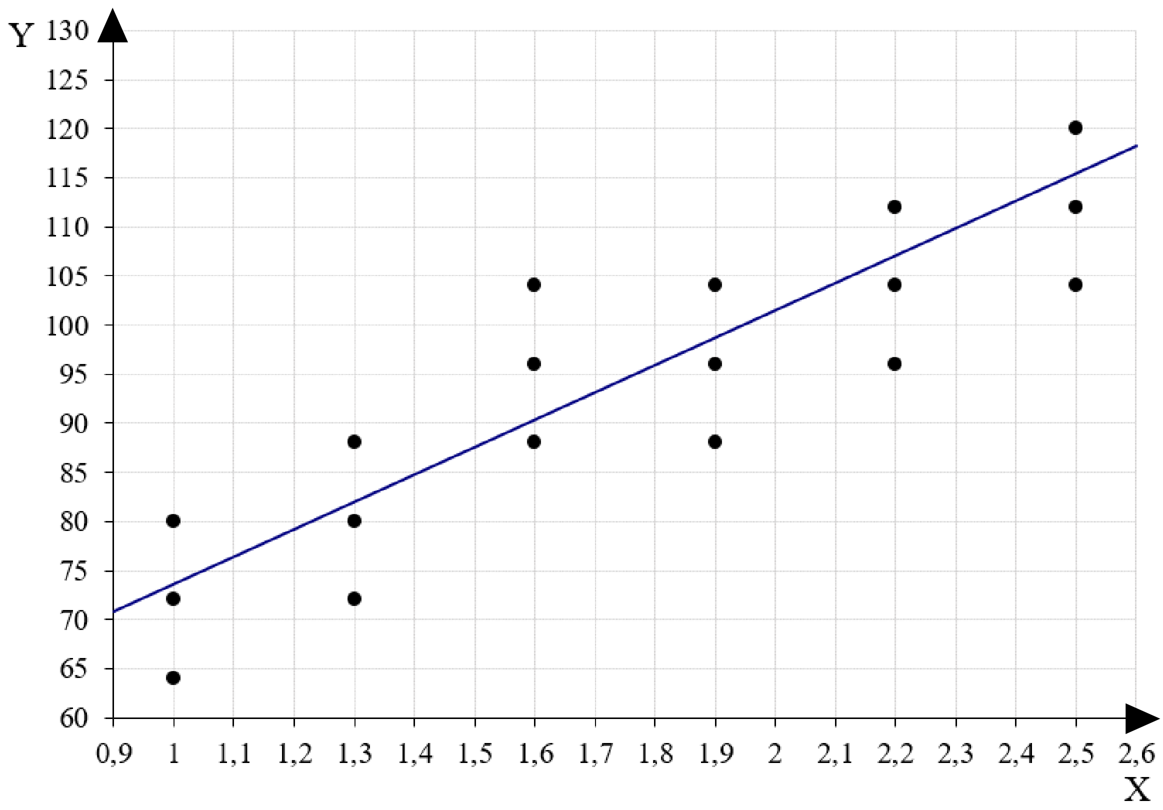
где $S_x = \sqrt{0{,}1658} \approx 0{,}407185; S_y = \sqrt{173{,}3753} \approx 13{,}16721;$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{4{,}6177}{0{,}407185 * 13{,}16721} = \frac{4{,}6177}{5{,}36149} \approx 0{,}861272$$

Составляем уравнение эмпирической линии регрессии y на x:

$$y = 92{,}72 + 0{,}861272 * \frac{13{,}16721}{0{,}407185}(x - 1{,}684)$$

$$y = 27{,}8511x + 45{,}81875$$


Y
130
125
120
115
110
105
100
95
90
85
80
75
70
65
60
0,9
1
1,1
1,2
1,3
1,4
1,5
1,6
1,7
1,8
1,9
2
2,1
2,2
2,3
2,4
2,5
2,6
X