ИДЗ 19.2. Вариант 18. Агаев Хамза Р3234 (поток 1.5).

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 на x;
б) построить уравнение эмпиричесой линии регрессии и случайные точки выборки (X, Y).

$X \backslash Y$	25 200	25 350	25 500	25 650	25 800	25 950	26 100	26 250	m_x
3150	3	4	2	–	–	–	–	–	9
3200	–	5	7	5	–	–	–	–	17
3250	–	–	–	8	14	6	–	–	28
3300	–	–	–	–	8	9	–	–	23
3350	–	–	–	–	–	5	6	3	14
3400	–	–	–	–	–	–	5	4	9
m_y	3	9	9	19	22	20	11	7	100

Для подсчета числовых характеристик (выборочных средних X' и Y', выборочных средних квадратичных отклонений Sx и Sy и выборочного корреляционного момента Sxy) составляем расчетную таблицу. При заполнении таблицы осуществляем контроль по строкам и столбцам:

$$\sum_{i=1}^{6} m_{xi} = \sum_{j=1}^{8} m_{yj} = n = 100,$$

$$\sum_{i=1}^{6}\sum_{j=1}^{8} m_{ij}x_i = \sum_{i=1}^{6} m_{xi}x_i = 327150,$$

$$\sum_{i=1}^{6}\sum_{j=1}^{8} m_{ij}y_j = \sum_{j=1}^{8} m_{yj}y_j = 2578050,$$

$$\sum_{i=1}^{6}\left(x_i\sum_{j=1}^{8} m_{ij}y_j\right) = \sum_{j=1}^{8}\left(y_j\sum_{i=1}^{6} m_{ij}x_i\right) = 8435715000.$$

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

$$\bar{x} = \frac{\sum\sum m_{ij}x_i}{n} = \frac{\sum m_{xi}x_i}{n} = \frac{327150}{100} = 3271,5;$$

$$\bar{y} = \frac{\sum m_{yj}y_j}{n} = \frac{2578050}{100} = 25780,5.$$

Выборочные дисперсии находим по формулам:

$$s_x^2 = \frac{1}{n-1}\left(\sum m_{xi}x_i^2 - \frac{1}{n}\left(\sum m_{xi}x_i\right)^2\right) =$$

$$= \frac{1}{99}\left(1070757500 - \frac{1}{100}(327150)^2\right) = 18232,32;$$

$$s_y^2 = \frac{1}{n-1}\left(\sum m_{yj}y_j^2 - \frac{1}{n}\left(\sum m_{yj}y_j\right)^2\right) =$$

$$= \frac{1}{99}\left(66470377500 - \frac{1}{100}(2578050)^2\right) = 70297,727.$$

	j	1	2	3	4	5	6	7	8	9	10	11	12	13
i	X \ Y	25200	25350	25500	25650	25800	25950	26100	26250	m_{x_i}	$m_{x_i} x_i$	$\sum_{j=1}^{k} m_{ij} y_j$	$x^2_i m_{ij}$	$x_i \sum_{j=1}^{k} m_{ij} y_j$
1	3150	3	4	2						9	28350	228000	89302500	718200000
2	3200		5	7	5					17	54400	433500	174080000	1387200000
3	3250				8	14	6			28	91000	722100	295750000	2346825000
4	3300				6	8	9			23	75900	593850	250470000	1959705000
5	3350						5	6	3	14	46900	365100	157115000	1223085000
6	3400							5	4	9	30600	235500	104040000	800700000
7	m_{y_j}	3	9	9	19	22	20	11	7	100	327150	2578050	1070757500	8435715000
8	$m_{y_j} y_j$	75600	228150	229500	487350	567600	519000	287100	183750	2578050				
9	$\sum_{i=1}^{m} m_{ij} x_i$	9450	28600	28700	61800	71900	65950	37100	23650	327150				
10	$y^2_j m_{ij}$	1905120000	5783602500	5852250000	12500527500	14644080000	13468050000	7493310000	4823437500	66470377500				
11	$y_j \sum_{j=1}^{k} m_{ij} x_i$	238140000	725010000	731850000	1585170000	1855020000	1711402500	968310000	620812500	8435715000				

Корреляционный момент вычисляем по формуле:

$$s_{xy} = \frac{1}{n-1}\left(\sum\sum m_{ij}x_i y_j - \frac{1}{n}\left(\sum m_{xi}x_i\right)\left(\sum m_{yj}y_j\right)\right) =$$
$$= \frac{1}{99}\left(8435715000 - \frac{1}{100}(327150 \cdot 2578050)\right) = 16408,3.$$

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

$$y = \bar{y} + r_{xy}\frac{s_y}{s_x}(x - \bar{x}),$$
$$s_x = \sqrt{18232,32} \approx 135,03; \quad s_y = \sqrt{70297,727} \approx 265,14;$$
$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{16408,3}{135,03 \cdot 265,14} =\approx 0,458.$$

Составляем уравнение эмпирической линии регрессии у на х:

$$y = 25780,5 + 0,458 \cdot \frac{265,14}{135,03}(x - 3271,5) =$$
$$= 22838,4 + 0,899x$$

Строим линию регрессии и случайные точки $(x_i;\ y_j)$.