

Тема: Линейная регрессия, коэффициент корреляции

ЗАДАНИЕ. Имеются данные средней выработки на одного рабочего y (тыс. руб.) и товарооборота x (тыс. руб.) в 20 магазинах за квартал. На основе указанных данных требуется:

- 1) определить зависимость (коэффициент корреляции) средней выработки на одного рабочего от товарооборота,
- 2) составить уравнение прямой регрессии этой зависимости.

		Магазины											
		1	2	3	4	5	6	7	8	9	10	11	12
x		10	14	21	23	27	32	39	45	55	61	62	68
y		3,8	4,8	5,9	6,1	6,2	6,3	6,6	7,4	8,5	9,7	10,5	12,4

РЕШЕНИЕ. Связь между изучаемыми признаками может быть выражена уравнением прямой линии регрессии \bar{Y} на X : $\bar{y}_x = ax + b$. Для вычисления параметров a , b и коэффициента корреляции r составим расчетную таблицу.

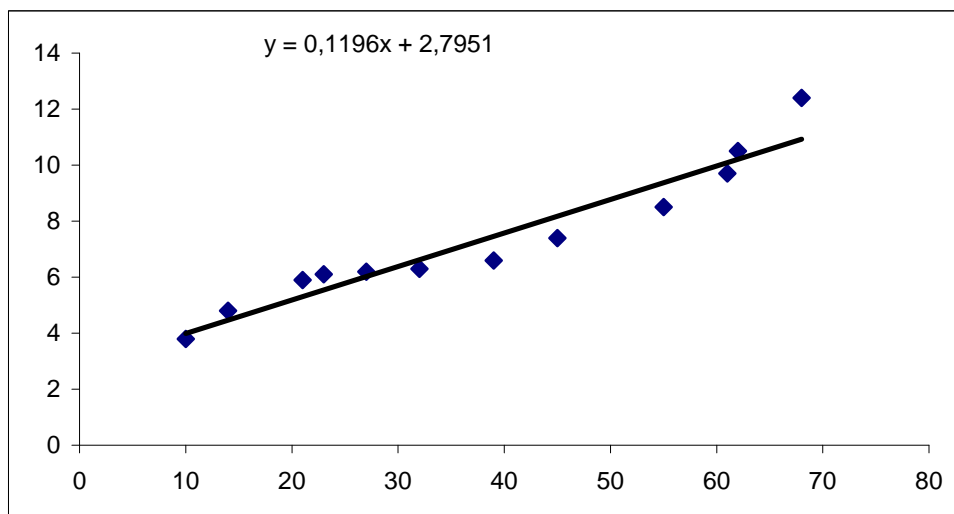
		Магазины												
		1	2	3	4	5	6	7	8	9	10	11	12	Сумма
X		10	14	21	23	27	32	39	45	55	61	62	68	457
Y		3,8	4,8	5,9	6,1	6,2	6,3	6,6	7,4	8,5	9,7	10,5	12,4	88,2
X^2		100	196	441	529	729	1024	1521	2025	3025	3721	3844	4624	21779
Y^2		14,44	23,04	34,81	37,21	38,44	39,69	43,56	54,76	72,25	94,09	110,3	153,8	716,3
XY		38	67,2	123,9	140,3	167,4	201,6	257,4	333	467,5	591,7	651	843,2	3882,2

Параметры a , b найдем из системы уравнений

$$\begin{cases} 21779a + 457b = 3882,2 \\ 457a + 12b = 88,2. \end{cases}$$

Получаем $a = 0,12$, $b = 2,8$, тогда $\bar{y}_x = 0,12x + 2,8$

Геометрическая иллюстрация:



Найдем выборочный коэффициент корреляции:

$$r_B = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n \sum y_i^2 - (\sum y_i)^2}} =$$
$$= \frac{12 \cdot 3882,2 - 457 \cdot 88,2}{\sqrt{12 \cdot 21779 - (457)^2} \cdot \sqrt{12 \cdot 716,3 - (88,2)^2}} \approx 0,959.$$

Связь очень сильная, прямая.