модели линейной регрессии:	
парная линейная регрессия	$\tilde{y} = a + bx$
многофакторная линейная регрессия	$\tilde{y} = a + bx_1 + cx_2$

модели нелинейной регрессии:	
гипербола	$\tilde{y} = a + \frac{b}{x}$
степенная	$\tilde{y} = a \cdot x^b$
экспоненциальная	$\tilde{y} = e^{a + b \cdot x}$
показательная	$\tilde{y} = a \cdot b^x$
полином разных степеней	$\tilde{y} = a + b_1 x + b_2 x^2 + b_3 x^3$

Этапы корреляционно-регрессионного анализа

- Найти характеристики исходного статистического ряда (среднее значение переменных, дисперсия, среднее квадратическое отклонение, ковариация).
- Рассчитать коэффициент корреляции.
- Выбрать вид исходного уравнения регрессии.
- Определить теоретические значения переменной *«y(x)».*
- Определить качество полученных оценок параметров уравнения регрессии путем расчета коэффициента детерминации, стандартных ошибок и критерия Стьюдента.
- Определить пригодность полученного уравнения для прогнозирования при помощи критерия Фишера.
- Прогнозирование значений переменной «*y*».

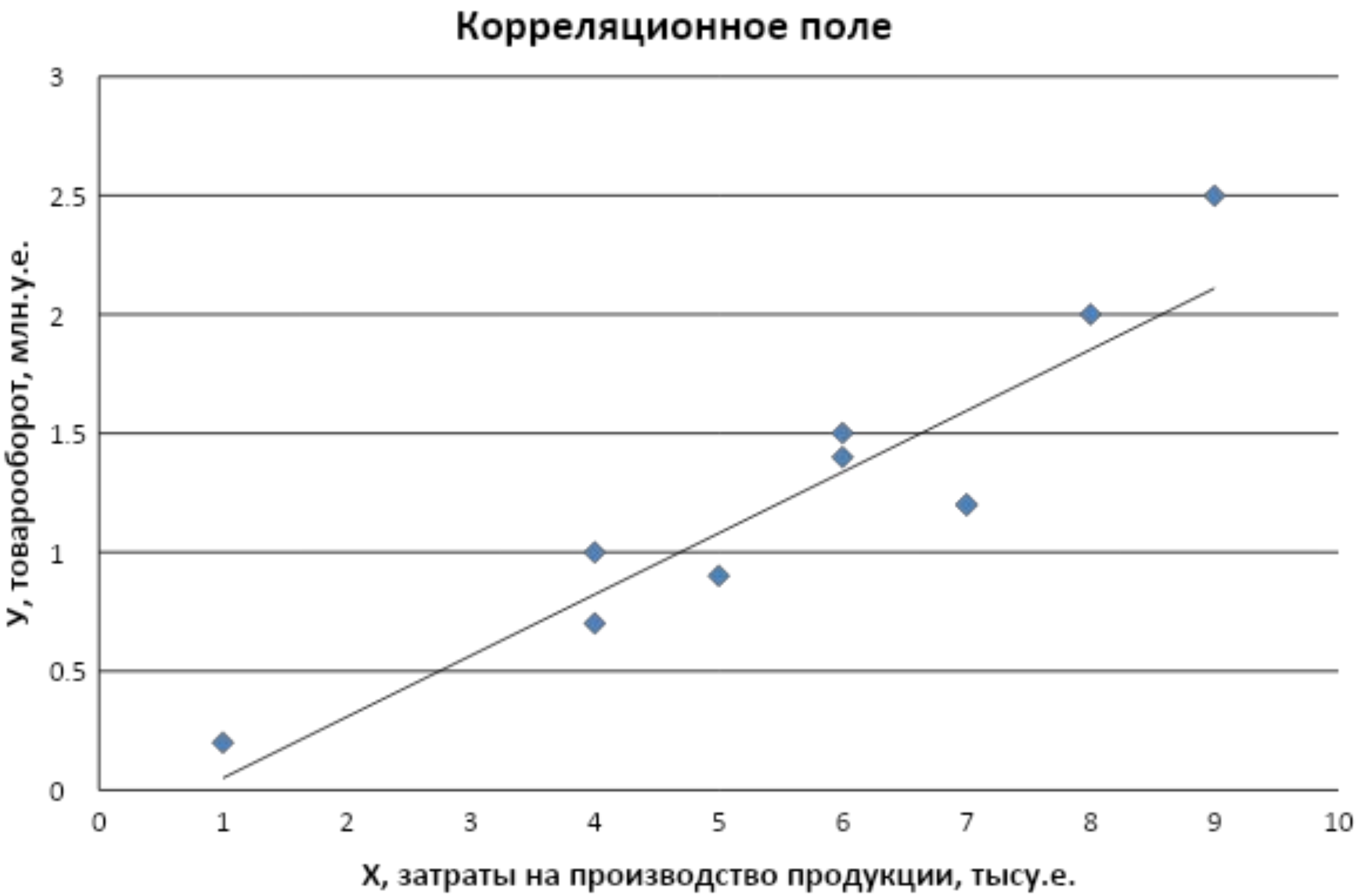
Решение типовой задачи

Изучить зависимость товарооборота от объема затрат на производство продукции по 10-ти предприятиям:

№ предприятия	**затраты на производство продукции, тыс. у.е., *X***	**товарооборот, млн. у.е., *У***
1	6	1,5
2	8	2
3	6	1,4
4	9	2,5
5	7	1,2
6	1	0,2
7	4	0,7
8	7	1,2
9	4	1
10	5	0,9

Решение:

- **1. Построим корреляционное поле зависимости** между объемом затрат на производство продукции и товарооборотом:



Зависимость между затратами на производство *X* и товарооборотом описывается **уравнением прямой**:

$$y = a + b \cdot x$$

Данное уравнение решается методом наименьших квадратов.

$$\begin{cases} a \cdot n + b \cdot \Sigma x_i = \Sigma y_i; \\ a \cdot \Sigma x_i + b \cdot \Sigma x_i^2 = \Sigma xy \end{cases}$$

Расчет параметров *a* и *b*

$$a = \overline{y} - b \cdot \overline{x}, \qquad b = \frac{\overline{y \cdot x} - \overline{y} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2}$$

Выполним вспомогательные расчеты:

товарооборот, млн. у.е., *У*	**затраты на произ-водство продукции, тыс. у.е., *X***	**X*Y**	**X^2**
1,5	6	9	36
2	8	16	64
1,4	6	8,4	36
2,5	9	22,5	81
1,2	7	8,4	49
0,2	1	0,2	1
0,7	4	2,8	16
1,2	7	8,4	49
1	4	4	16
0,9	5	4,5	25
$\Sigma Y=12{,}6$	**$\Sigma X=57$**	**$\Sigma(X*Y)=84{,}2$**	**$\Sigma(X^2)=373$**

Определили параметры уравнения регрессии:

$$a = \bar{y} - b \cdot \bar{x}, \qquad \xcancel{b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}}$$

$$b = \frac{(84{,}2/10) - (12{,}6/10) \cdot (57/10)}{(373/10) - (57/10)\hat{}2} = 0{,}257$$

$$a = (12{,}6/10) - 0{,}257 \cdot (57/10) = -0{,}207$$

$$a = -\,0,\,207; \qquad b = 0,\,257$$

Тогда **уравнение регрессии** примет вид:

$$y = -0{,}207 + 0{,}257 \cdot x$$

2. Определим тесноту связи между показателями *Х* и *У*:

– ***рассчитаем линейный коэффициент корреляции:***

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y} = 0{,}915,$$

Значение коэффициента корреляции | ***r*** | >0,9 и со знаком «+», значит **связь прямая и очень сильная.**

− ***рассчитаем коэффициент детерминации:***

$$R^2 = r^2 = 0{,}915^2 = 0{,}837$$

Коэффициент детерминации показывает,

что **на 83,7%** товарооборот предприятия зависит от затрат на производство продукции,

а **остальные 16,3%** обусловлены неучтенными факторами.

– ***рассчитаем ранговый коэффициент Спирмена:***

$$R^2 = r^2 = 0{,}915^2 = 0{,}837$$

где d – разность рангов (Rx – Ry);

n – число пар рангов.

$$p = 1 - \frac{6 \cdot \sum d^2}{n(n^2 - 1)}$$

Затраты на производство продукции, тыс. у.е., **Х**	**Rx**	Товарооборот, млн. у.е., **У**	**Ry**	**$(Rx – Ry)^2$**
6	5,5	1,5	8	6,25
8	9	2	9	0
6	5,5	1,4	7	2,25
9	10	2,5	10	0
7	7,5	1,2	5,5	4
1	1	0,2	1	0
4	2,5	0,7	2	0,25
7	7,5	1,2	5,5	4
4	2,5	1	4	2,25
5	4	0,9	3	1
Всего	**-**	**-**	**-**	**20**

$$p = 1 - \frac{6 \cdot 20}{10\,(100 - 1)} = 0,88$$

Значение **коэффициента Спирмена** стремится к единице и величина положительная, значит, **связь** между товарооборотом и затратами на производство продукции **сильная и прямая.**

3. Проверим значимость параметров уравнения регрессии.

- Для оценки статистической значимости коэффициентов регрессии применяется

t-критерий Стьюдента.

- Коэффициенты значимы при выполнении условия

$$t_{расч} > t_{таб}$$

- $t_{таб}$ определяется по таблицам t-критерия Стьюдента по числу степеней свободы $\boldsymbol{v_1 = n-k-1}$ (k - число независимых переменных в уравнении регрессии) и заданному уровню значимости

$\alpha = 0{,}05$ или $0{,}1$.

- Расчетные значения **t-критерия Стьюдента** определяются отношениями:

$$t_{pa} = \frac{a}{CO(a)} \qquad t_{pb} = \frac{b}{CO(b)}$$

- *где* ***CO_a*** *и* ***CO_b*** *-* ***«стандартные ошибки»*** *коэффициентов регрессии, которые представляют собой оценку стандартного отклонения функции от плотности вероятности коэффициента.*

Стандартные ошибки коэффициентов регрессии (Co_a; CO_b) определяются соотношениями:

$$COa = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widetilde{y}_i)^2}{(n-2)} \cdot \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = \sigma_{ocm} \frac{\sqrt{\sum_{i=1}^{n} x_i^2}}{n\sigma_x},$$

$$COb = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widetilde{y}_i)^2 / (n-2)}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = \frac{\sigma_{ocm}}{\sigma_x \sqrt{n}}$$

где σ^2_{ocm} *представляет собой несмещенную оценку остаточной дисперсии*

$$\sigma^2_{ocm} = \frac{\sum_{i=1}^{n}(y_i - \widetilde{y}_i)^2}{(n-2)}.$$

4. Проверим качество уравнения регрессии по F-критерию Фишера.

- Согласно *F*-критерию Фишера уравнение считается адекватным при выполнении условия:

$$F_{расч} > F_{таб},$$

где $F_{таб}$ *определяется по таблицам F-критерия Фишера,*

при заданном уровне значимости **α = 0,05 или 0,1**

и числе степеней свободы $v_1 = k$; $v_2 = n - k - 1$.

При $\alpha = 0{,}05;\ v_1 = 1\ ;\ v_2 = 8$

$$F_{таб} = 5{,}317$$

- Величину $F_{расч}$ можно выразить через коэффициент детерминации R^2

$$F_{расч} = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k} = \frac{0{,}837}{1-0{,}837} \cdot \frac{10-1-1}{1} = 41{,}08$$

Получили $F_{расч} > F_{таб}$,
значит полученное уравнение регрессии адекватно описывает зависимость между переменными ***У*** *и* ***Х***.

5. Рассчитаем возможные значения товарооборота при увеличении затрат на производство продукции до 12 тыс. у.е.

Для этого используем построенное уравнение регрессии:

$$\boxed{y_{(x)} = -0{,}207 + 0{,}257 \cdot x}$$

$$y_{(12)} = -0{,}207 + 0{,}257 \cdot 12 = 2{,}877$$

Таким образом, при затратах на производство в 12 тыс.у.е.

товарооборот составит 2,877 млн.у.е.