

Тема 1

Простая линейная эконометрическая модель

Оценивание параметров эконометрической модели методом наименьших квадратов

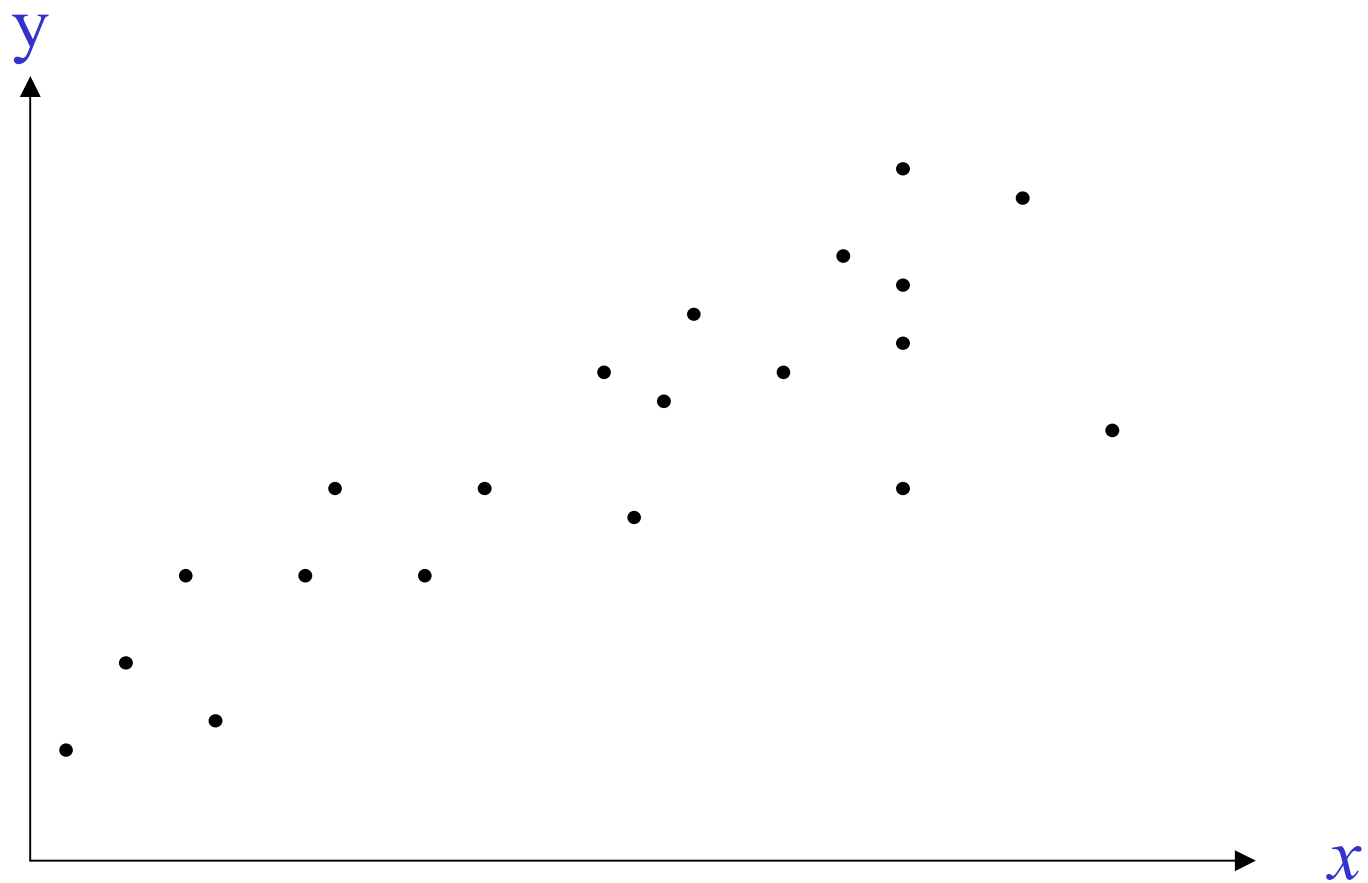
Уравнение парной регрессии

$$y = a_0 + a_1 x + u$$

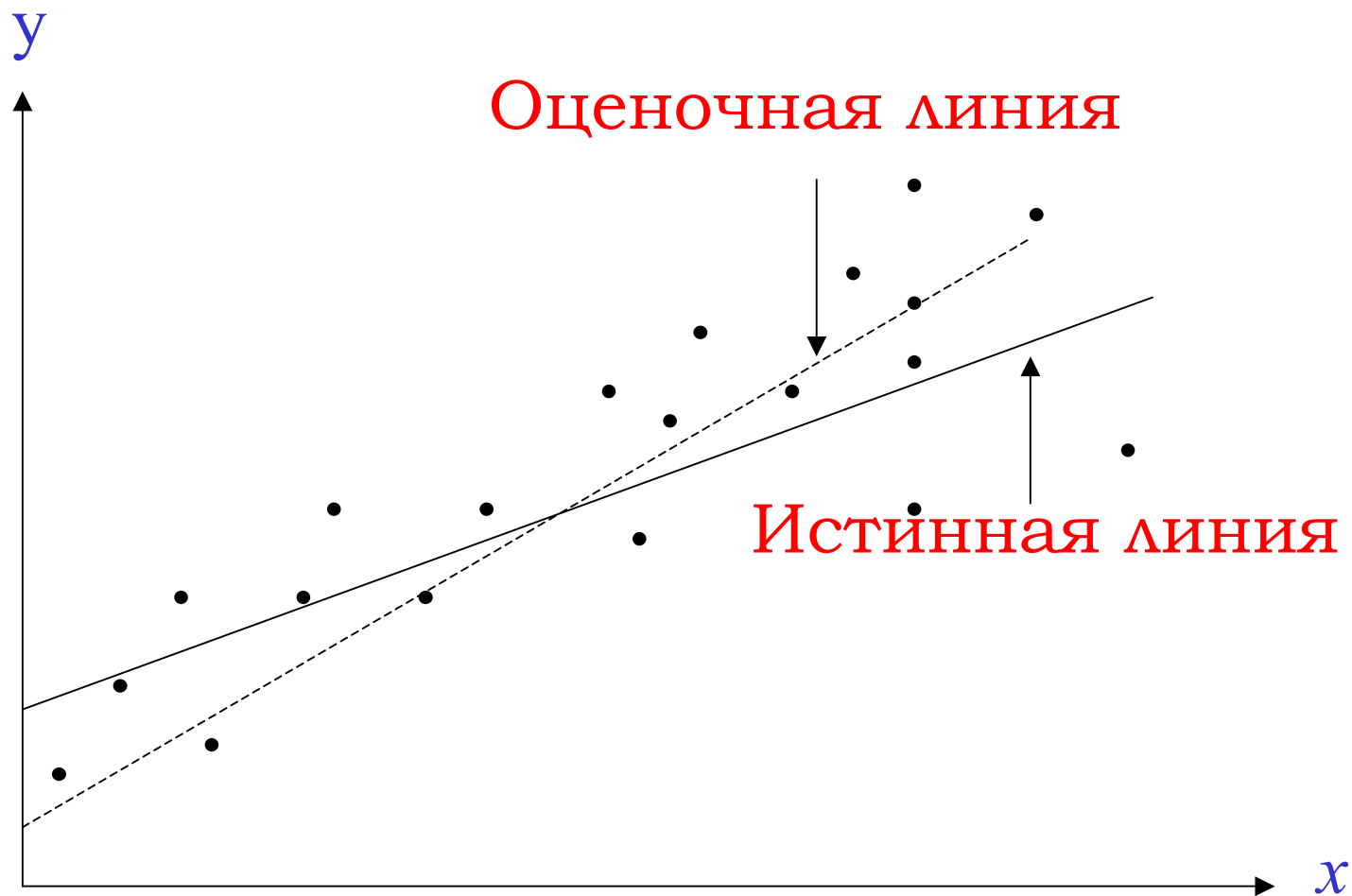
Основная задача – оценить

a_0

a_1



Корреляционное поле



Корреляционное поле и линии регрессии

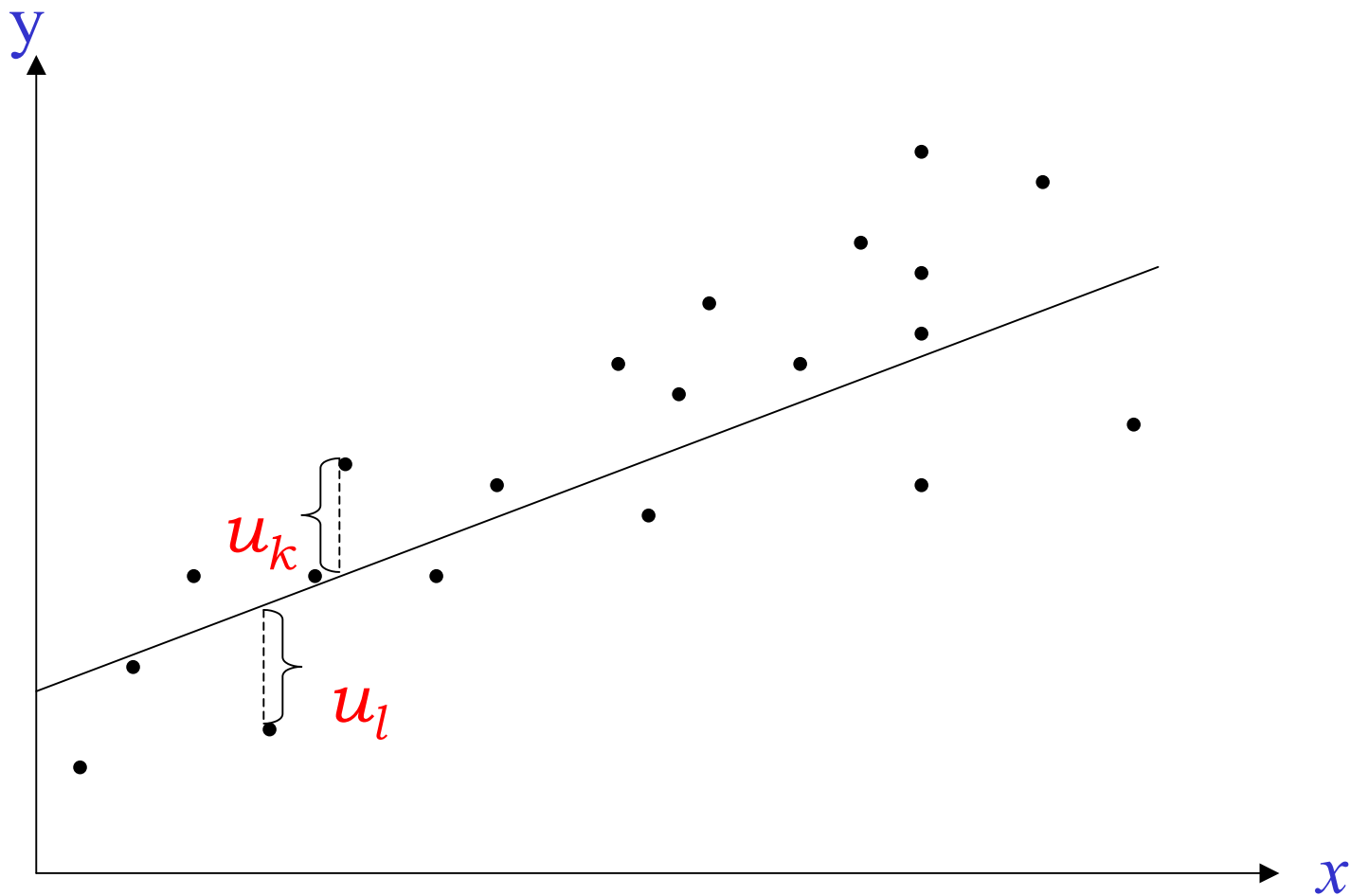


Иллюстрация отклонений

Критерии оценки:

- Сумма отклонений
- Сумма модулей отклонений
- Сумма квадратов отклонений
- И др.

Метод наименьших квадратов (МНК)

Пусть

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x$$

$\hat{y}, \hat{a}_0, \hat{a}_1$ - оценки величин y, a_0 и a_1

Основная задача:

подобрать такие $\hat{a}_0, \hat{a}_1$, которые
минимизируют сумму квадратов
отклонений $\hat{y}$ от y

$$\min \left\{ \sum_{i=1}^n u_i^2 \right\} = f(\hat{a}_0, \hat{a}_1)$$

Очень важное замечание

- Применение метода 1МНК возможно при условии, когда отклонения распределены по нормальному закону, т.е. когда их среднее значение равно нулю, а дисперсия – постоянная величина

Найдем

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i)^2$$

$$\left\{ \begin{array}{l} \frac{\partial \left(\sum_{i=1}^n u_i^2 \right)}{\partial \hat{a}_0} = -2 \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i) = 0 \\ \frac{\partial \left(\sum_{i=1}^n u_i^2 \right)}{\partial \hat{a}_1} = -2 \sum_{i=1}^n x_i (y_i - \hat{a}_0 - \hat{a}_1 x_i) = 0 \end{array} \right.$$



$$\begin{cases} \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i) = 0 \\ \sum_{i=1}^n x_i (y_i - \hat{a}_0 - \hat{a}_1 x_i) = 0 \end{cases}$$

Система нормальных уравнений

$$\left\{ \begin{array}{l} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ \hat{a}_0 \sum_{i=1}^n x_i + \hat{a}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{array} \right.$$

Найдем $\widehat{a}_0$

$$\widehat{a}_0 = \frac{\sum_{i=1}^n y_i - \widehat{a}_1 \sum_{i=1}^n x_i}{n}$$

Подставим во второе уравнение

$$\frac{\sum_{i=1}^n y_i - \hat{a}_1 \sum_{i=1}^n x_i}{n} \sum_{i=1}^n x_i + \hat{a}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i$$

Находим $\widehat{a}_1$

1.
$$\widehat{a}_1 = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2}$$

2.

$$\widehat{a}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

3.

$$\hat{a}_1 = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}$$

4.

$$\hat{a}_1 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

Наконец еще одна формула

$$\widehat{a}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$\hat{a}_1$ — коэффициент регрессии

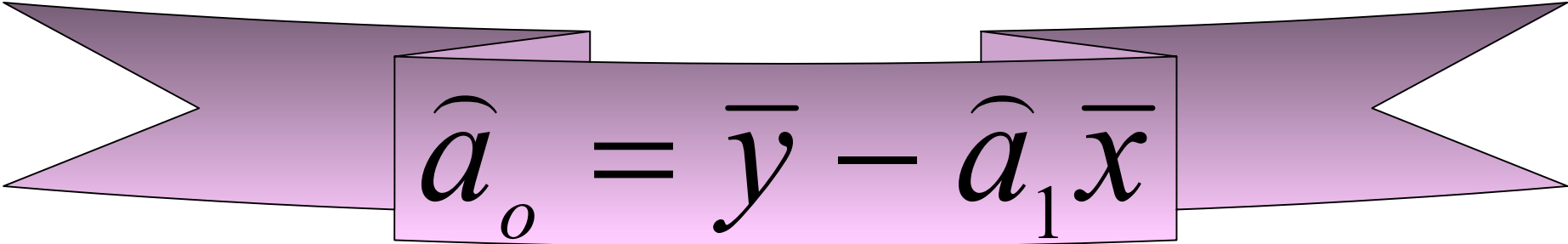
Определение

- Коэффициент регрессии в линейной модели показывает, на сколько единиц собственного измерения изменится в среднем зависимая переменная при изменении i -й независимой переменной в среднем на единицу собственного измерения при прочих неизменных условиях.

Найдем $\hat{a}_0$

$$\hat{a}_0 = \frac{\sum_{i=1}^n y_i - \hat{a}_1 \sum_{i=1}^n x_i}{n}$$



A purple ribbon banner with a 3D effect, featuring a central rectangular section and two flared ends. It contains the simplified regression equation.
$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$$

Очень важное заключение

Линия регрессии проходит через среднюю точку!

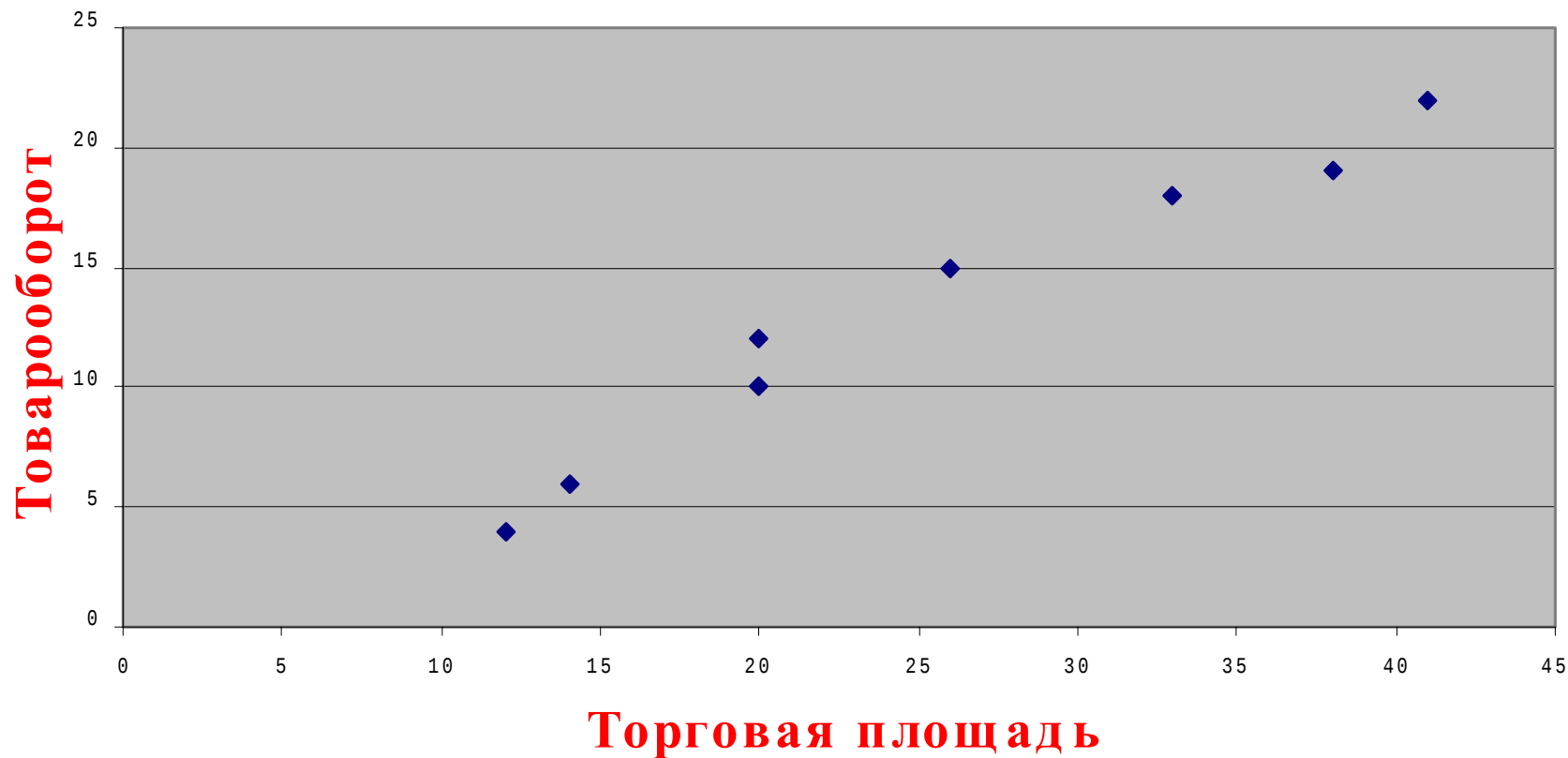
$$\bar{y} = \hat{a}_0 + \hat{a}_1 \bar{x}$$

Пример: имеется выборка

Наблюдение	Торговая площадь, м ² (x)	Товарооборот, тыс. грн. (y)
1	20	10
2	14	6
3	12	4
4	20	12
5	33	18
6	38	19
7	41	22
8	26	15

Необходимо построить модель,
отражающую зависимость суточного
товарооборота от размера торговой
площади, в предположении о линейной
связи между рассматриваемыми
показателями

Решение



Разброс точек, характеризующих зависимость между рассматриваемыми показателями

	x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	x_i^2
	20	10	-5,5	-3,25	17,875	30,25	400
	14	6	-11,5	-7,25	83,375	132,25	196
	12	4	-13,5	-9,25	124,875	182,25	144
	20	12	-5,5	-1,25	6,875	30,25	400
	33	18	7,5	4,75	35,625	56,25	1089
	38	19	12,5	5,75	71,875	156,25	1444
	41	22	15,5	8,75	135,625	240,25	1681
	26	15	0,5	1,75	0,875	0,25	676
Сумма	204	106			477	828	6030
Среднее	25,5	13,25					

$$\hat{a}_1 = \frac{477}{828} = 0,576$$

$$\hat{a}_0 = 13,25 - 0,576 \cdot 25,2 = -1,44$$

Уравнение регрессии

$$\hat{y} = -1,44 + 0,576 \cdot x$$

Коэффициент регрессии $\hat{a}_1 = 0,576$

показывает, что при увеличении торговой

площади в среднем на 1 м^2 суточный

товарооборот возрастет в среднем

на $0,576$ тыс. грн. при прочих неизменных

условиях

При интерпретации оценок необходимо ПОМНИТЬ

- Прочие неизменные условия
- Что каждая числовая величина представляет собой оценку, которая, как правило, несет в себе ошибку
- Единицы измерения переменных, которые рассматриваются

Некоторые применения уравнения регрессии

1. Прогнозирование

Для $x = 30 \text{ м}^2$

$$\hat{y} = -1,44 + 0,576 \cdot x = -1,44 + 0,576 \cdot 30 = 15,84 \text{ тыс.грн.}$$

2. Определение площади торговой точки для достижения запланированного товарооборота

$\hat{y} = 20$ тыс. грн.

$$x = \frac{\hat{y} + 1,44}{0,576} = \frac{20 + 1,44}{0,576} = 37,2 \text{ м}^2$$

Уравнение регрессии обладает тем свойством,
что оно справедливо для всех без исключения
наблюдений переменных

$$\hat{y}_1 = -1,44 + 0,576 \cdot x_1$$

$$\hat{y}_2 = -1,44 + 0,576 \cdot x_2$$

$\vdots$

$$\hat{y}_8 = -1,44 + 0,576 \cdot x_8$$

Числовые характеристики коэффициентов регрессии, полученные с помощью 1МНК, описывают в среднем зависимость между зависимой и независимыми переменными



$$u_i = y_i - \hat{y}_i$$

Коэффициенты эластичности,
парной корреляции,
детерминации

Коэффициент эластичности

- Показывает, на сколько процентов в среднем изменится зависимая переменная при изменении независимой переменной в среднем на 1 процент при прочих неизменных условиях.

$$E_{y/x} = \frac{\partial y}{\partial x} \cdot \frac{\bar{x}}{\bar{y}}$$

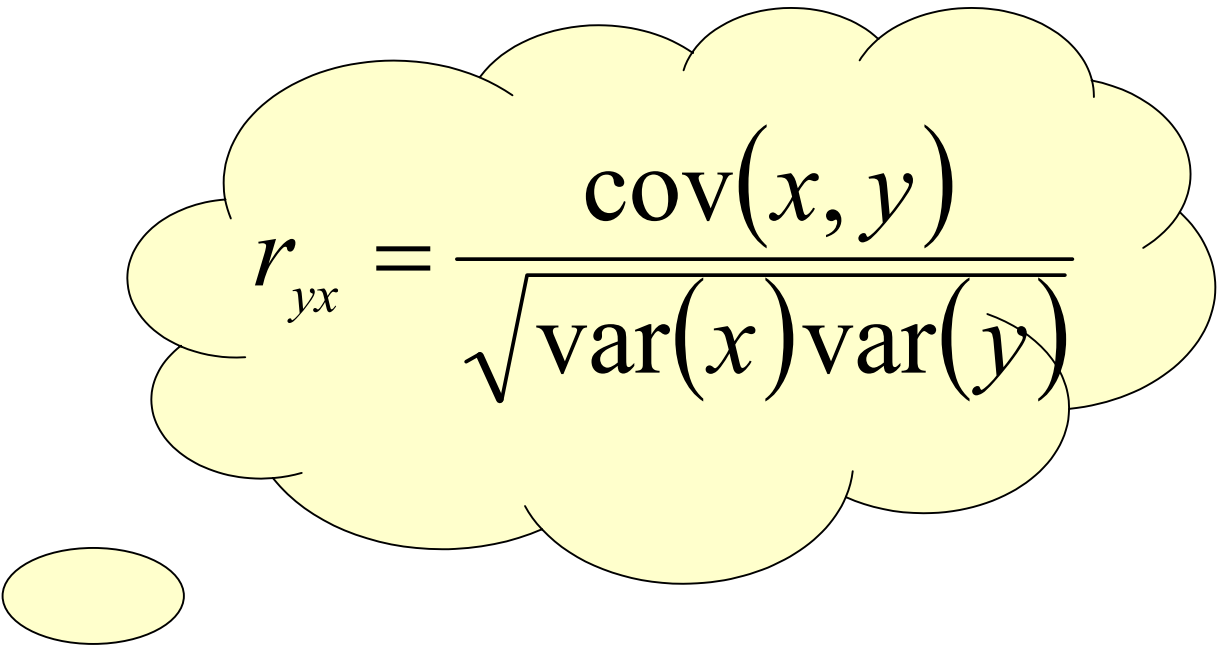
Пример расчета

$$\frac{\partial y}{\partial x} = \hat{a}_1 = 0,576$$

$$E_{y/x} = \frac{\partial y}{\partial x} \cdot \frac{\bar{x}}{\bar{y}} = 0,576 \cdot \frac{25,5}{13,25} = 1,1$$

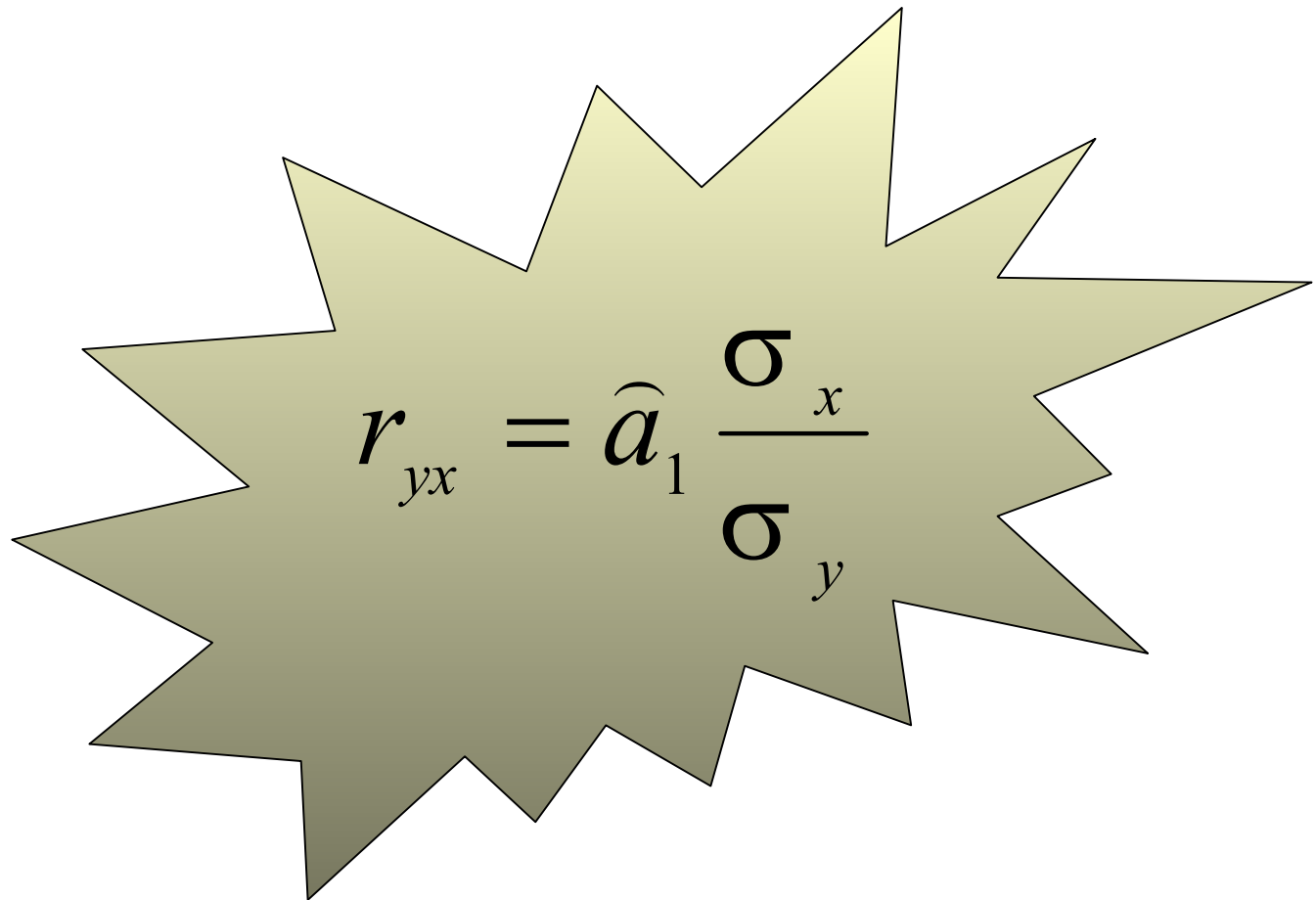
Коэффициент парной корреляции

- характеризует тесноту и направление связи между переменными


$$r_{yx} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)\text{var}(y)}}$$

$$-1 \leq r_{yx} \leq 1$$

Связь между коэффициентом корреляции и
коэффициентом регрессии $\hat{a}_1$


$$r_{yx} = \hat{a}_1 \frac{\sigma_x}{\sigma_y}$$

Рассчитаем коэффициент корреляции

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 103,5 \quad \sigma_x = 10,17349$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = 35,6875 \quad \sigma_y = 5,9739$$

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = 59,625$$

$$r_{yx} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{59,625}{10,17349 \cdot 5,973902} = 0,981$$

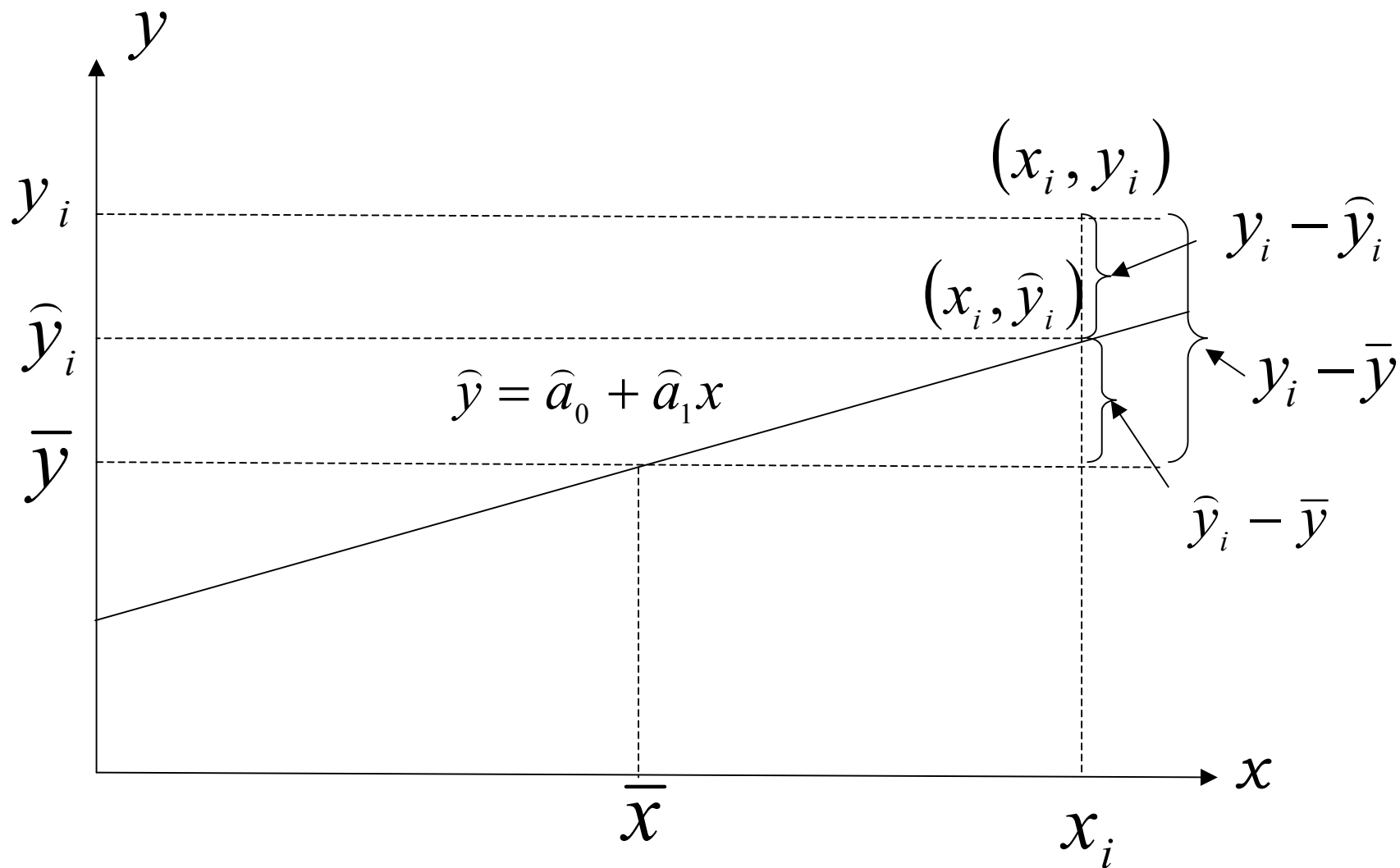


Иллюстрация отклонений значений
зависимой переменной

$(y_i - \hat{y}_i)$ - отклонение от точки на линии регрессии

$(y_i - \bar{y})$ - общее отклонение

$(\hat{y}_i - \bar{y})$ - отклонение, связанное с линией регрессии

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$



$$\sum_{i=1}^n (y_i - \bar{y})^2 =$$

$$= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$



$$\begin{aligned} & \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \\ & = \sum_{i=1}^n (\hat{a}_0 + \hat{a}_1 x_i - \hat{a}_0 - \hat{a}_1 \bar{x}) u_i = \hat{a}_1 \sum_{i=1}^n (x_i - \bar{x}) u_i \end{aligned}$$

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \quad \longrightarrow \quad \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$$



$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = SST$$

- общая сумма квадратов отклонений

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = SSE$$

- сумма квадратов необъяснимых отклонений (ошибок)

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = SSR$$

- сумма квадратов отклонений, объясняемая регрессией

$$SST = SSE + SSR$$



$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} + \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n}$$



$$\sigma_{общ.}^2 = \sigma_{ош.}^2 + \sigma_{рег.}^2$$

$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \sigma_{\text{общ.}}^2$$

– общая дисперсия

$$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} = \sigma_{\text{ош.}}^2$$

– дисперсия ошибок

$$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n} = \sigma_{\text{регр.}}^2$$

– дисперсия, объясняемая регрессией

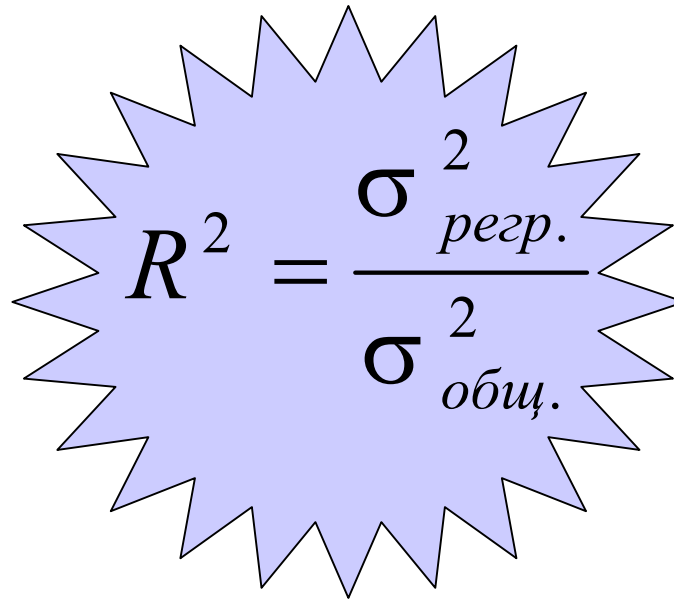
$$1 = \frac{\sigma_{ош.}^2}{\sigma_{общ.}^2} + \frac{\sigma_{регр.}^2}{\sigma_{общ.}^2}$$

$$\frac{\sigma_{ош.}^2}{\sigma_{общ.}^2}$$

– доля ошибок в общей дисперсии

$$\frac{\sigma_{регр.}^2}{\sigma_{общ.}^2}$$

– часть дисперсии, которая объясняется регрессией


$$R^2 = \frac{\sigma_{\text{пер.}}^2}{\sigma_{\text{общ.}}^2}$$

$$R^2 = \frac{SSR}{SST}$$

$$0 \leq R^2 \leq 1$$

Определение

- Часть дисперсии, которая объясняется регрессией, называется **коэффициентом детерминации**

- **Связь между коэффициентом корреляции и коэффициентом детерминации**

$$R^2 = \frac{\sigma_{\text{пер.}}^2}{\sigma_{\text{общ.}}^2} = \frac{SSR}{SST}$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (\hat{a}_0 + \hat{a}_1 x_i - (\hat{a}_0 + \hat{a}_1 \bar{x}))^2 = \hat{a}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

$$R^2 = \frac{SSR}{SST} = \frac{\hat{a}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 \cdot \left(\frac{1}{n}\right)}{\sum_{i=1}^n (y_i - \bar{y})^2 \cdot \left(\frac{1}{n}\right)} = \hat{a}_1^2 \frac{\sigma_x^2}{\sigma_y^2}$$

$$r_{yx} = \widehat{a}_1 \frac{\sigma_x}{\sigma_y}$$



$$R^2 = r_{yx}^2$$

Пример расчета коэффициента детерминации

y_i	$\hat{y}_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$	$u_i = y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	u_i / y_i
10	10,1	-3,25	10,563	-3,168	10,039	-0,0815	0,0066	-0,008
6	6,62	-7,25	52,563	-6,625	43,891	-0,625	0,3906	-0,104
4	5,47	-9,25	85,563	-7,777	60,484	-1,4728	2,1692	-0,368
12	10,1	-1,25	1,5625	-3,168	10,039	1,91848	3,6806	0,1599
18	17,6	4,75	22,563	4,321	18,668	0,42935	0,1843	0,0239
19	20,5	5,75	33,063	7,201	51,856	-1,4511	2,1056	-0,076
22	22,2	8,75	76,563	8,929	79,733	-0,1794	0,0322	-0,008
15	13,5	1,75	3,0625	0,288	0,0829	1,46196	2,1373	0,0975
Сумма			285,5		274,79		10,707	-0,284

$$\sigma_{\text{обш.}}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{285,5}{8} = 35,6875$$

$$\sigma_{\text{пер.}}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n} = \frac{274,7935}{8} = 34,34919$$

$$R^2 = \frac{\sigma_{пер.}^2}{\sigma_{общ.}^2} = \frac{34,34919}{35,6875} = 0,9625$$

$$R^2 = r_{yx}^2 = (0,981)^2 = 0,9625$$

Коэффициент остаточной детерминации

- Рассчитывается для определения доли вариации y за счет неучтенных в модели факторов

$$1 - R^2 = 1 - 0,9625 = 0,0375$$