Классическая линейная регрессия

1. Понятие о парной и множественной линейной регрессии
2. Сущность метода наименьших квадратов и способы нахождения параметров уравнения

Понятие о парной и множественной линейной регрессии

Регрессия это функциональная зависимость между объясняющими переменными и условным математическим ожиданием (средним значением) зависимой переменой, которая строится с целью предсказания (прогноз) этого среднего значения при фиксированных значениях первых (регрессоров).

Множественная линейная регрессия

$$\tilde{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_k x_{ki} + \varepsilon_i$$

Парная линейная регрессия

$$\tilde{y}_i = a_0 + a_1 x_i + \varepsilon_i$$

Причины возникновения ε_i

1. не включение в уравнение факторов оказывающих существенное влияние на результативный показатель;
2. трудности и ошибки при измерении данных;
3. неверный выбор функциональной формы модели;
4. агрегирование переменных;
5. непредсказуемость человеческого фактора;
6. ограниченность статистических данных.

Очередность «появления» параметров и переменных в регрессионном уравнении:

1. имеем *n* штук пар наблюдений $y_i; x_i$
2. находим параметры уравнения $a_0;\ a_1$
3. находим теоретические значения зависимой переменной $\tilde{y}_i$
4. находим значения случайного члена $\varepsilon_i = y_i - \tilde{y}_i$

Ограниченность парной линейной регрессии:

1. никакая единственная независимая переменная (за редким исключением) не в состоянии «качественно» отразить изменения зависимой переменной;
2. могут существовать несколько переменных оказывающих одинаковое влияние на независимую переменную, но противоречащие друг другу;
3. линейная форма связи очень примитивна.

НО

Нет ни чего лучше по простоте и ясности объяснения чем парная линейная регрессия

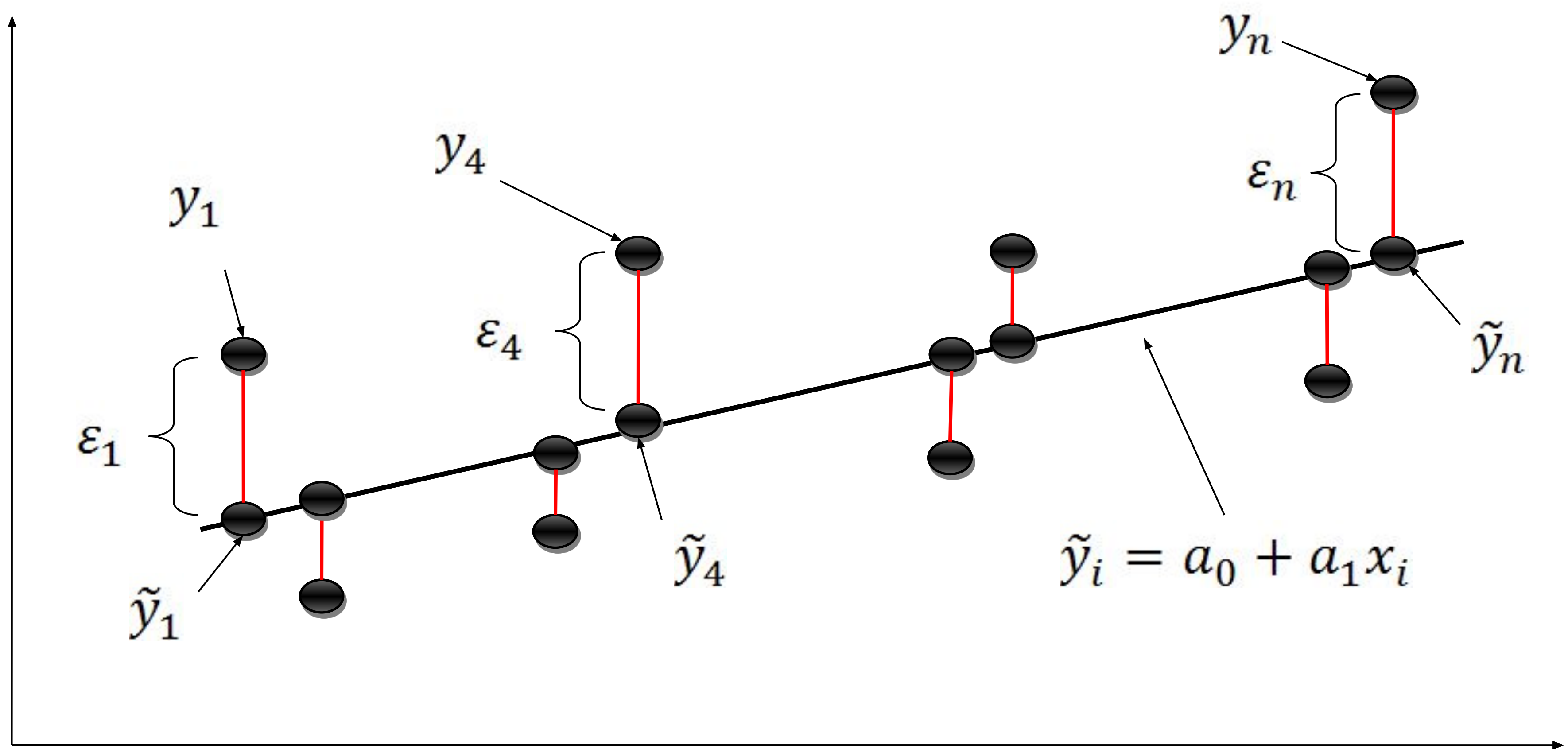
Сущность метода наименьших квадратов и способы нахождения параметров уравнения

Сущность метода наименьших квадратов

состоит в минимизации суммы квадратов отклонений фактических значений от значений, вычисленных по уравнению связи

$$f(a_0; a_1) = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \rightarrow min$$

СУЩНОСТЬ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ



Свойства оценок параметров регрессионного уравнения:

1. Несмещенность оценок параметров регрессии
2. Состоятельность оценок параметров регрессии
3. Эффективность оценок параметров регрессии
4. Достаточность оценки

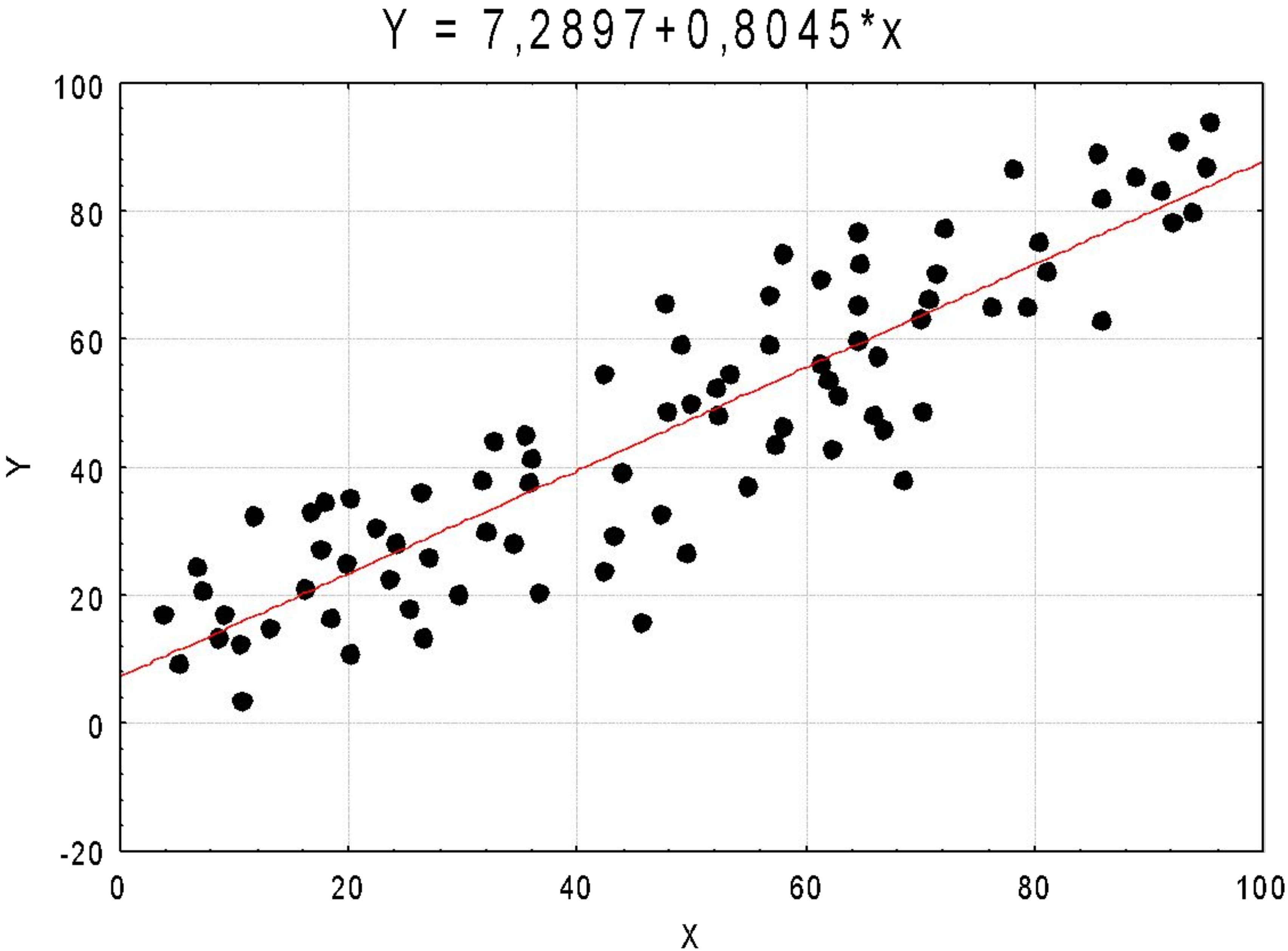
$$\begin{cases} \dfrac{dQ}{da_0} = -2\sum(y_i - a_0 - a_1 x_i) = 0 \\ \dfrac{dQ}{da_1} = -2\sum(y_i - a_0 - a_1 x_i)x_i = 0 \end{cases}$$

$$\Downarrow$$

$$\begin{cases} \sum y_i = na_0 + a_1 \sum x_i \\ \sum y_i x_i = a_0 \sum x_i + a_1 \sum x_i^2 \end{cases}$$

$$\Big\Downarrow \; /n$$

$$\begin{cases} a_0 + a_1 \bar{x} = \bar{y} \\ a_0 \bar{x} + a_1 \overline{x^2} = \overline{xy} \end{cases} \Rightarrow a_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} \Rightarrow a_0 = \bar{y} - a_1 \bar{x}$$

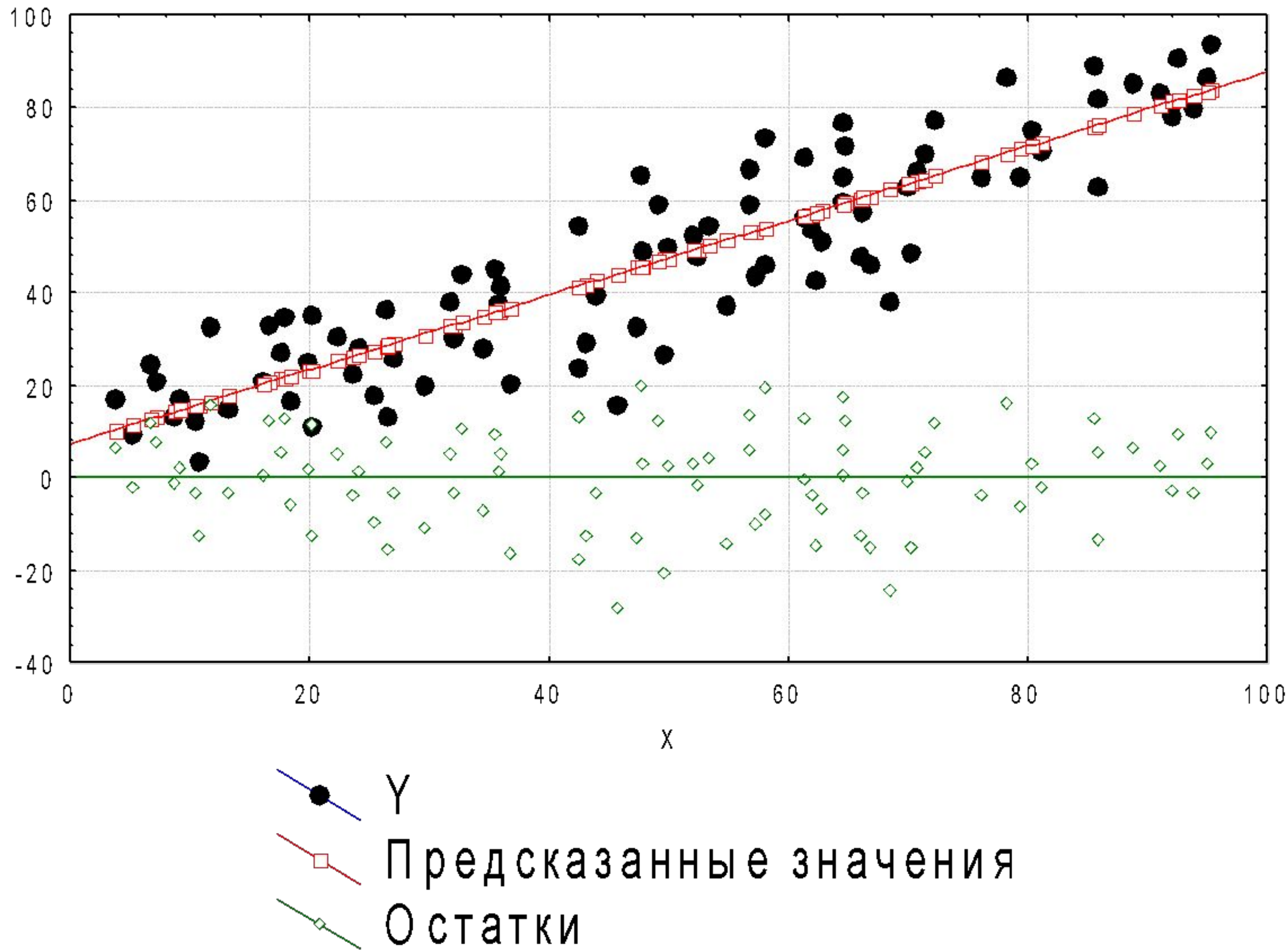

Y = 7,2897+0,8045*x
Y
X
100
80
60
40
20
0
-20
0
20
40
60
80
100

Regression Summary for Dependent Variable: Y (Spreadsheet1) R= ,89444655 R?= ,80003463 Adjusted R?= ,79768209 F(1,85)=340,07 p=0,00

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(85)	p-level
Intercept			7,29	2,39	3,04	0,00
X	0,89	0,05	0,80	0,04	18,44	0,00

1 2 3 4 5 6 7

$$\tilde{y}_i = 7{,}29 + 0{,}80x_i + \varepsilon_i$$


100
80
60
40
20
0
-20
-40
0
20
40
60
80
100
X
Y
Предсказанные значения
Остатки

Система нормальных уравнений регрессии при y и x_1, x_2

$$\begin{cases} \sum y_i = na_0 + a_1 \sum x_{1i} + a_2 \sum x_{2i} \\ \sum y_i x_{1i} = a_0 \sum x_{1i} + a_1 \sum x_i^2 + a_2 \sum x_{1i} x_{2i} \\ \sum y_i x_{2i} = a_0 \sum x_{2i} + a_1 \sum x_{1i} x_{2i} + a_2 \sum x_2^2 \end{cases}$$

ОПРЕДЕЛИТЕЛЬ СИСТЕМЫ

$$\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{vmatrix} =$$

$$= \left(n \sum x_1^2 \sum x_2^2 + \sum x_1 \sum x_1 x_2 \sum x_2 + \sum x_2 \sum x_1 \sum x_1 x_2\right) -$$

$$-\left(\sum x_2 \sum x_1^2 \sum x_2 + \sum x_1 \sum x_1 \sum x_2^2 + n \sum x_1 x_2 \sum x_1 x_2\right)$$

ЧАСТНЫЕ ОПРЕДЕЛИТЕЛИ

$$\Delta a_0 = \begin{vmatrix} \sum y & \sum x_1 & \sum x_2 \\ \sum yx_1 & \sum x_1^2 & \sum x_1x_2 \\ \sum yx_2 & \sum x_1x_2 & \sum x_2^2 \end{vmatrix} \Rightarrow a_0 = \frac{\Delta a_0}{\Delta}$$

$$\Delta a_1 = \begin{vmatrix} n & \sum y & \sum x_2 \\ \sum x_1 & \sum yx_1 & \sum x_1x_2 \\ \sum x_2 & \sum yx_2 & \sum x_2^2 \end{vmatrix} \Rightarrow a_1 = \frac{\Delta a_1}{\Delta}$$

$$\Delta a_2 = \begin{vmatrix} n & \sum x_1 & \sum y \\ \sum x_1 & \sum x_1^2 & \sum yx_1 \\ \sum x_2 & \sum x_1x_2 & \sum yx_2 \end{vmatrix} \Rightarrow a_2 = \frac{\Delta a_2}{\Delta}$$

МНОЖЕСТВЕННОЕ ЛИНЕЙНОЕ УРАВНЕНИЕ РЕГРЕССИИ В МАТРИЧНОЙ ФОРМЕ

$$Y=XA+E$$

где

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{mn} \end{pmatrix} \quad E = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

$$A = (X^T X)^{-1} X^T Y$$

ОЦЕНКИ ПАРАМЕТРОВ МАТРИЦЫ ПАРАМЕТРОВ РЕГРЕССИОННОГО УРАВНЕНИЯ

Параметры множественного уравнения регрессии можно оценить используя:

табличные редакторы

- ✔MS Excel
- ✔Lotus 1-2-3
- ✔Quattro Pro
- ✔StarOffice Calc

эконометрические пакеты программ

- ✔Stata 10.0
- ✔EViews 7

статистические (общего назначения) пакеты программ

- ✔Statistica 6.0
- ✔SPSS 10.0
- ✔StatGraphic
- ✔STADIA6.0

математические пакеты программ

- ✔Mathcad
- ✔Mathematica
- ✔Maple

Стандартизованным коэффициентом регрессии или β - коэффициентом

$$\beta_j = a_j \frac{\sigma_{x_j}}{\sigma_y}$$

Коэффициентов эластичности

$$\theta_j = a_j \frac{\bar{x}_j}{\bar{y}}$$

Регрессионное уравнение в стандартизированной форме

$$\tilde{y}_i = -0{,}56x_{1i} + \underline{0{,}74x_{2i}}$$

$$\tilde{y}_i = -\underline{0{,}96x_{1i}} + 0{,}74x_{2i}$$

$$\tilde{y}_i = -0{,}56x_{1i} + 0{,}74x_{2i} - \underline{0{,}89x_{3i}}$$