ЭКОНОМЕТРИКА

Лекция 6

Метод наименьших квадратов

Уравнение парной регрессии

Метод наименьших квадратов

В математической статистике методы получения наилучшего приближения к исходным данным в виде аппроксимирующей функции получили название регрессионного анализа

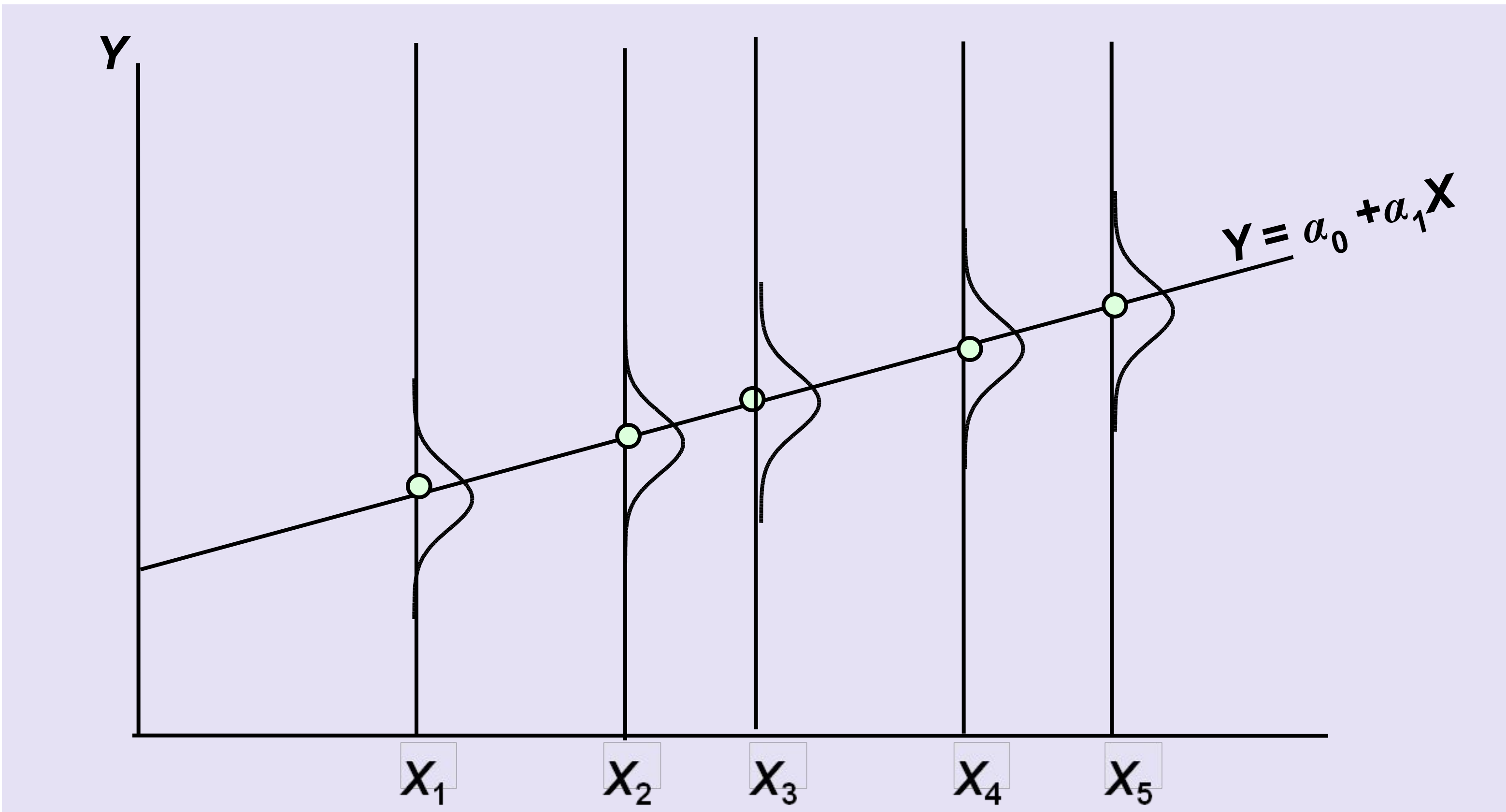
Основными задачами регрессионного анализа являются установление зависимости между переменными и оценка (прогноз) значений зависимой переменной

В экономических исследованиях часто заданному значению одной переменной может соответствовать множество значений другой переменной

Другими словами, каждому значению одной переменной соответствует условное распределение другой переменной

Метод наименьших квадратов

Графическая иллюстрация сказанного:



Зависимость, при которой каждому значению одной переменной соответствует условное математическое ожидание другой называется **регрессионной:**

Метод наименьших квадратов

Начнем с построения модели в виде линейного уравнения парной регрессии

$$y_t = a_0 + a_1 x_t + u_t \qquad (6.1)$$

Постановка задачи

Дано:
Выборка наблюдений за поведением переменных y_t и x_t

Найти:
1. Оценки значений параметров $\mathbf{a_0}$ и $\mathbf{a_1}$
2. Оценки точности $\sigma(a_0)$ и $\sigma(a_1)$.
3. Оценка рассеяния случайного возмущения σ_u
4. Оценку точности прогнозирования $\sigma(y(x_0))$

Метод наименьших квадратов

Введем следующие обозначения и определения

1. Выборка

$$\begin{pmatrix} y_1 & x_1 \\ y_2 & x_2 \\ \dots & \dots \\ y_n & x_n \end{pmatrix}$$

2. Система уравнений наблюдений

$$\begin{cases} y_1 = a_0 + a_1 x_1 + u_1 \\ y_2 = a_0 + a_1 x_2 + u_2 \\ \dots\dots\dots\dots\dots\dots\dots \\ y_n = a_0 + a_1 x_n + u_n \end{cases} \qquad (6.2)$$

3. В е к т о р а

$$\vec{Y} = \begin{Bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{Bmatrix} \quad \vec{X} = \begin{Bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{Bmatrix} \quad \vec{U} = \begin{Bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{Bmatrix} \quad \vec{a} = \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix}$$

4. Матрица коэффициентов при параметрах

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}$$

Метод наименьших квадратов

Идея метода.

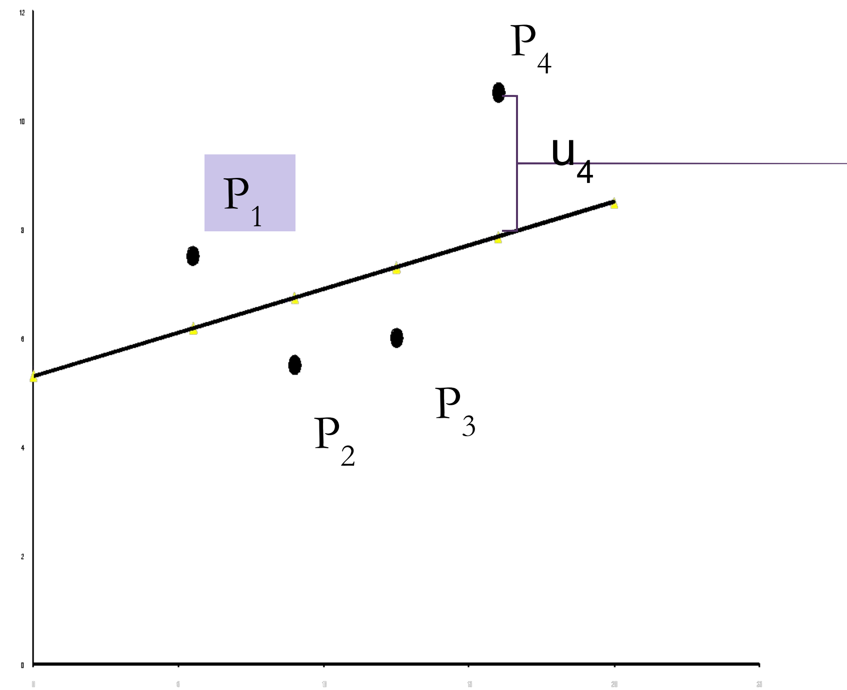
Пусть имеем выборку из 4-х точек (n=4):

$P_1 = (x_1, y_1)$

$P_2 = (x_2, y_2)$

$P_3 = (x_3, y_3)$

$P_4 = (x_4, y_4)$



На практике мы имеем возможность наблюдать только исходные точки

Предполагаем, что существует теоретическая прямая, которая наилучшим образом проходит через них

<u>Задача:</u> оценить с некоторой точностью, как может проходить эта прямая

Метод наименьших квадратов

Итак, оценки параметров модели парной регрессии согласно МНК будем искать из условия:

$$S = \sum_{i=1}^{n} u_i^2 = \sum \left(y_i - \tilde{a}_0 - \tilde{a}_1 x_i \right)^2 \Rightarrow \min \qquad (6.2)$$

Условиями минимума функции являются равенство нулю первых производных и положительность вторых производных по $\tilde{a}_0$ и $\tilde{a}_1$

$$\begin{cases} \dfrac{\partial S}{\partial \tilde{a}_0} = \sum 2\left(y_i - \tilde{a}_0 - \tilde{a}_1 x_i\right)\left(-1\right) = 0 \\ \dfrac{\partial S}{\partial \tilde{a}_1} = \sum 2\left(y_i - \tilde{a}_0 - \tilde{a}_1 x_i\right)\left(-x_i\right) = 0 \end{cases} \qquad (6.3)$$

Система (6.3) называется **системой нормальных уравнений** для вычисления оценок параметров уравнения парной регрессии (6.1)

Метод наименьших квадратов

Упростим систему нормальных уравнений (6.3)

$$\begin{cases} n\tilde{a}_0 + \tilde{a}_1 \sum x_i = \sum y_i \\ \sum x_i \tilde{a}_0 + \tilde{a}_1 \sum x_i^2 = \sum x_i y_i \end{cases} \qquad (6.4)$$

Убеждаемся, что решение системы уравнений (6.4) будет соответствовать минимуму функции (6.1)

Для этого вычисляем значения вторых частных производных функции (6.1)

$$\frac{\partial^2 S}{\partial \tilde{a}_0^2} = n > 0$$

$$\frac{\partial^2 S}{\partial \tilde{a}_1^2} = \sum x_i^2 > 0$$

Вторые производные больше нуля – функция (6.1) принимает минимальное значение в точке $\tilde{a}_0$, $\tilde{a}_1$

Метод наименьших квадратов

$$\begin{cases} n\tilde{a}_0 + \tilde{a}_1 \sum x_i = \sum y_i \\ \sum x_i \tilde{a}_0 + \tilde{a}_1 \sum x_i^2 = \sum x_i y_i \end{cases} \qquad (6.4)$$

Для решения системы (6.4) выразим из первого уравнения $\tilde{a}_0$, подставим его во второе уравнение

$$\begin{cases} \tilde{a}_0 = \dfrac{1}{n}\left(\sum y_i - \tilde{a}_1 \sum x_i\right) \\ \sum x_i \dfrac{1}{n}\left(\sum y_i - \tilde{a}_1 \sum x_i\right) + \sum x_i^2 a_1 = \sum x_i y_i \end{cases} \qquad (6.5)$$

Решив второе уравнение системы (6.5) получим:

$$\tilde{a}_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\left(n\sum x_i^2 - \sum x_i \sum x_i\right)} \qquad (6.6)$$

Метод наименьших квадратов

$$\tilde{a}_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \sum x_i \sum x_i} \qquad (6.6)$$

Проанализируем выражение (6.6)
Для этого вычислим COV(x,y) и σ^2(x)

$$\mathrm{Cov}(x,y) = \frac{1}{(n-1)}\sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{(n-1)}\sum \left(x_i y_i - \bar{x} y_i - x_i \bar{y} + \overline{xy}\right) =$$

$$= \frac{1}{n(n-1)}\left(n\sum x_i y_i - \sum x_i \sum y_i\right)$$

$$\sigma^2(x) = \frac{1}{(n-1)}\sum (x_i - \bar{x})^2 = \frac{1}{(n-1)}\sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) == \frac{1}{n(n-1)}\left(n\sum x_i^2 - \sum x_i \sum x_i\right)$$

$$\tilde{a}_1 = \frac{\mathrm{Cov}(x,y)}{\sigma^2(x)} \qquad (6.7)$$

Метод наименьших квадратов

Проверим выполнение условия несмещенности для оценки (6.7)
Для этого вычислим числитель выражения (6.7)

$$Cov(x,y) = Cov(x, a_0 + a_1 x + u) =$$

$$= Cov(x, a_0) + Cov(x, a_1 x) + Cov(x, u)$$

Подставив в (6.7) полученное выражение получим:

$$\tilde{a}_1 = a_1 + \frac{Cov(x,u)}{\sigma^2(x)} \qquad (6.8)$$

Математическое ожидание выражения (6.7) имеет вид:

$$M(\tilde{a}_1) = a_1 + \frac{1}{\sigma^2(x)} Cov(x,u) \qquad (6.9)$$

Метод наименьших квадратов

Вычислим дисперсии параметров уравнения регрессии и дисперсию прогнозирования эндогенной переменной

1. Дисперсия параметра ã$_1$

$$\sigma^2(\tilde{a}_1) = \sigma^2\left(\frac{Cov(x,y)}{\sigma^2(x)}\right) = \sigma^2\left(\frac{\frac{1}{(n-1)}\sum(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{(n-1)}\sum(x_i-\bar{x})^2}\right) =$$

$$= \sigma^2\left(\frac{\sum(x_i-\bar{x})(y_i-a_0-a_1x)}{\sum(x_i-\bar{x})^2}\right) = \sigma^2\left(\frac{\sum(x_i-\bar{x})u_i}{\sum(x_i-\bar{x})^2}\right) =$$

$$= \sigma^2(\tilde{a}_1) = \sum\frac{(x_i-\bar{x})^2\sigma^2(u_i)}{\left[(x_i-\bar{x})^2\right]^2} = \sigma_u^2\frac{1}{n\sigma^2(x)} \qquad (6.10)$$

Метод наименьших квадратов

2. Дисперсия параметра $\tilde{a}_0$

$$\tilde{a}_0 = \frac{1}{n}\left(\sum y_i - \tilde{a}_1 \sum x_i\right) = \bar{y} - \tilde{a}_1 \bar{x}$$

$$\sigma^2\left(\tilde{a}_0\right) = \sigma^2\left(\bar{y} - \tilde{a}_1 \bar{x}\right) = \sigma^2(\bar{y}) + \bar{x}^2 \sigma^2(\tilde{a}_1) - 2\bar{x}\mathrm{Cov}\left(\bar{y}, \tilde{a}_1\right)$$

$$\sigma^2\left(\tilde{a}_0\right) = \sigma^2(\bar{y}) + \bar{x}^2 \sigma^2(\tilde{a}_1)$$

$\sigma^2(y)$ Определяется с помощью (6.10)

$$\sigma^2(\bar{y}) = \sigma^2\left(\frac{1}{n}\sum(y_i)\right) = \frac{1}{n^2}\sum \sigma^2\left(a_0 + a_1 x_i + u_i\right) = \frac{\sigma^2(u)}{n}$$

В результате получаем:

$$\sigma^2\left(\tilde{a}_0\right) = \sigma^2(u)\left(\frac{1}{n} + \frac{\bar{x}^2}{n\sigma^2(x)}\right)$$

Оценка параметров уравнения парной регрессии с помощью ММП

Исходные предположения

1. Уравнение имеет вид: $y_t = a_0 + a_1 x_t + u_t$
2. Случайное возмущение имеет нормальное распределение с параметрами 0 и σ_u
3. Для получения ММП-оценок имеем выборку из n наблюдений

Тогда:

$$u_t = y_t - a_0 - a_1 x_t$$

Закон распределения для случайного возмущения принимает вид:

$$p_u(u_t a_0, a_1) = \frac{1}{(2\pi)^{1/2}\sigma_u} \exp\left(- \frac{\left(\left(y_t - a_0 - a_1 x_t\right) - 0\right)^2}{2\sigma_u^2} \right)$$

Оценка параметров уравнения парной регрессии с помощью ММП

1. Функция правдоподобия получит вид:

$$L(u_t a_0, a_1) = \prod_{t=1}^{t=n} \left[\frac{1}{(2\pi)^{\frac{1}{2}} \sigma_u} e^{-\frac{((y_t - a_0 - a_1 x_t) - 0)^2}{2\sigma_u^2}} \right] = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_u^n} e^{-\frac{\sum (y_t - a_0 - a_1 x_t)^2}{2\sigma_u^2}}$$

2. Логарифм функции правдоподобия

$$\ln(L(u_t a_0, a_1)) = \ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}} \right) + \ln\left(\frac{1}{\sigma_u^n} \right) + \frac{\sum (y_t - a_0 - a_1 x_t)^2}{2\sigma_u^2}$$

Оценка параметров уравнения парной регрессии с помощью ММП

3. Составляем уравнения для вычисления оценок a_0 и a_1

$$\begin{cases} \dfrac{\partial \ln(L)}{\partial a_0} = \sum y_i - n a_0 - a_1 \sum x_i = 0 \\ \dfrac{\partial \ln(L)}{\partial a_1} = \sum x_t y_t - \sum x_t a_0 - a_1 \sum x_t^2 = 0 \end{cases}$$

Получили систему уравнений совпадающую с (6.3)

Следовательно, и решения совпадут

Метод наименьших квадратов

Вывод

С помощью метода наименьших квадратов получили

1. Оценки параметров уравнения регрессии, по крайней мере, состоятельными

2. Если случайное возмущение подчиняется нормальному закону распределения, то оценки параметров модели несмещенные и эффективные

3. Нет необходимости в знании закона распределения случайных возмущений

Пример применения МНЛ

X	Y	ỹ	U	U^2	σ(y)
1,0	1,91	2,17	-0,26	0,07	1,20
2,0	2,76	2,72	0,04	0,00	1,19
3,0	2,67	3,26	-0,59	0,35	1,17
4,0	4,03	3,80	0,23	0,05	1,16
5,0	4,12	4,34	-0,22	0,05	1,15
6,0	2,81	4,88	-2,07	4,30	1,15
7,0	6,53	5,42	1,11	1,22	1,14
8,0	6,24	5,97	0,27	0,07	1,14
9,0	9,03	6,51	2,52	6,36	1,13
10,0	6,87	7,05	-0,18	0,03	1,13
11,0	9,09	7,59	1,50	2,24	1,13
12,0	7,08	8,13	-1,05	1,11	1,13
13,0	7,79	8,68	-0,89	0,78	1,14
14,0	8,75	9,22	-0,47	0,22	1,14
15,0	11,19	9,76	1,43	2,05	1,15
16,0	10,15	10,30	-0,15	0,02	1,15
17,0	10,52	10,84	-0,32	0,10	1,16
18,0	10,89	11,38	-0,49	0,24	1,17
19,0	10,59	11,93	-1,34	1,78	1,19
20,0	13,40	12,47	0,93	0,87	1,20
			ΣU^2	21,93	

X-стаж работы сотрудника

Y- часовая оплата труда

Модель: $Y=a_0+aX_t+U_t$

$\Sigma x_i=210$; $\Sigma y_i=146.42$;
$\Sigma x_i^2=2870$; $\Sigma x_i y_i=1897.66$

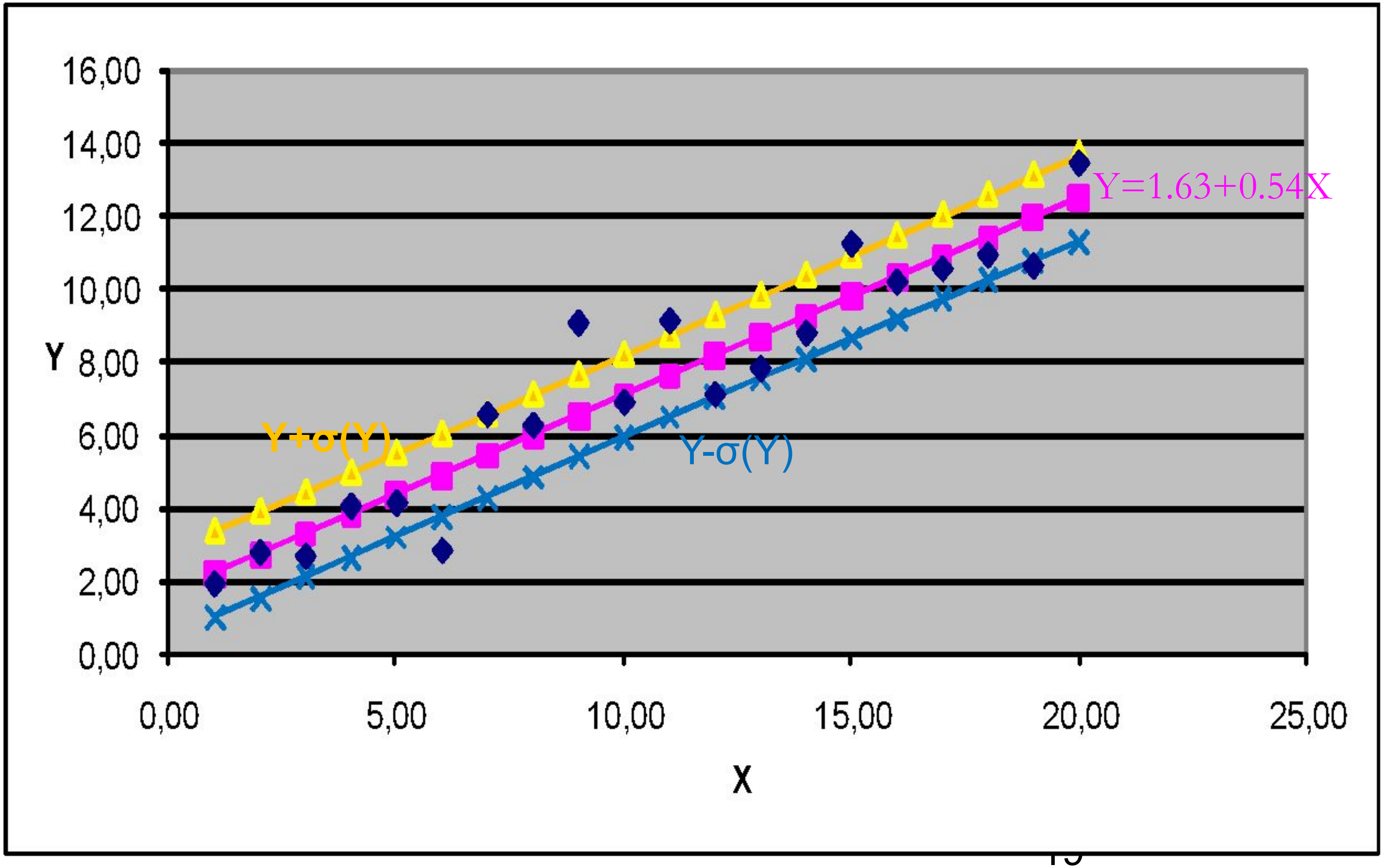
$$a_1 = \frac{20\cdot 1897.66 - 210.14\cdot 6.42}{20\cdot 2870 - 44100} = 1.63$$

$$a_0 = \frac{1}{20}\left(155.42 - 1.63\cdot 210\right) = 0.54$$

$$\sigma_u^2 = \frac{21.93}{20-2} = 1.1$$

Пример применения МНЛ

Графическое отображение результатов



Метод наименьших квадратов

Выводы:

1. Метод наименьших квадратов имеет следующие преимущества:

- не требуется знания закона распределения случайного возмущения
- дает оценки по крайней мере состоятельные
- в случае нормального распределения случайного возмущения оценки параметров линейной модели несмещенные и эффективные

2. Для получения несмещенных и эффективных оценок параметров в случае, если случайное возмущение имеет закон распределения отличный от нормального, необходимо наложить на него дополнительные требования