



*Изучайте классиков и решайте трудные задачи.
П.Л. Чебышев*

Парная линейная регрессия

Оценивание по МНК коэффициентов регрессии

Презентация

подготовлена к.э.н., профессором
каф. математической статистики СГЭУ,

Сухановой Е.И.

E-mail: eisukhanova@yandex.ru

План

- 1. Метод наименьших квадратов (МНК).**
- 2. Перечень средств MS Excel.**
- 3. Алгоритм применения функции ЛИНЕЙН.**
- 4. Результаты оценивания регрессии.**

Цели обучения

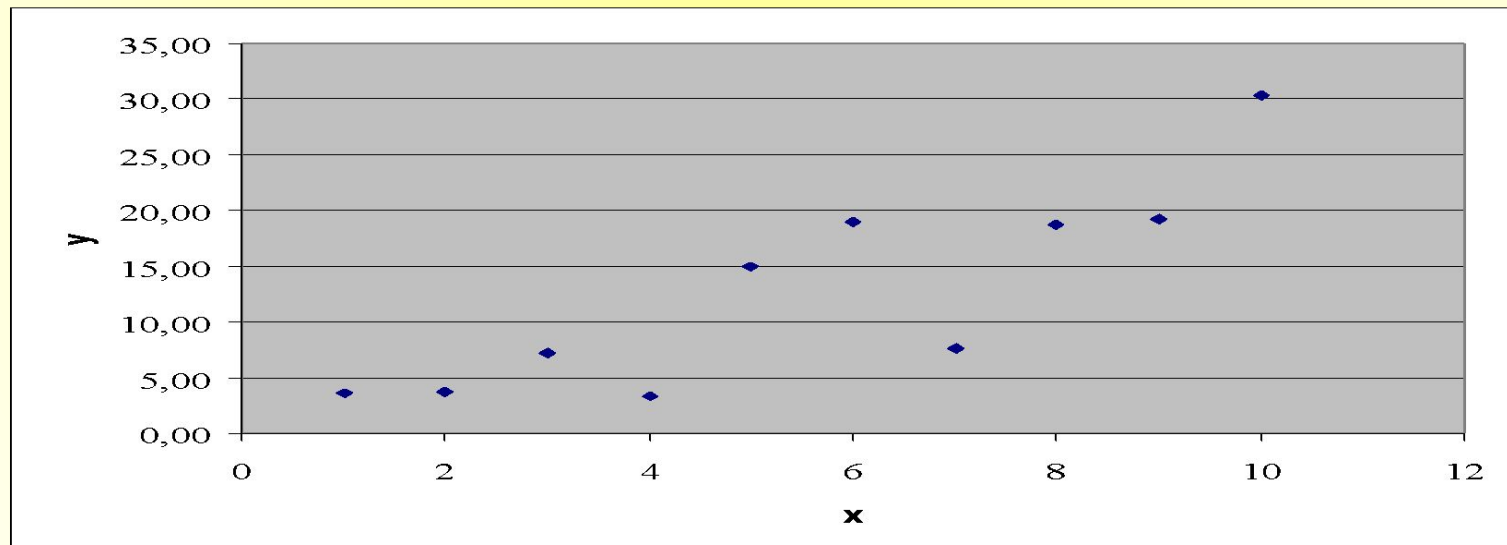
- **научиться** применять МНК для оценивания теоретических коэффициентов уравнения парной линейной регрессии;
- **изучить** структуру дополнительной регрессионной статистики функции ЛИНЕЙН табличного процессора MS Excel.

Метод наименьших квадратов (МНК)

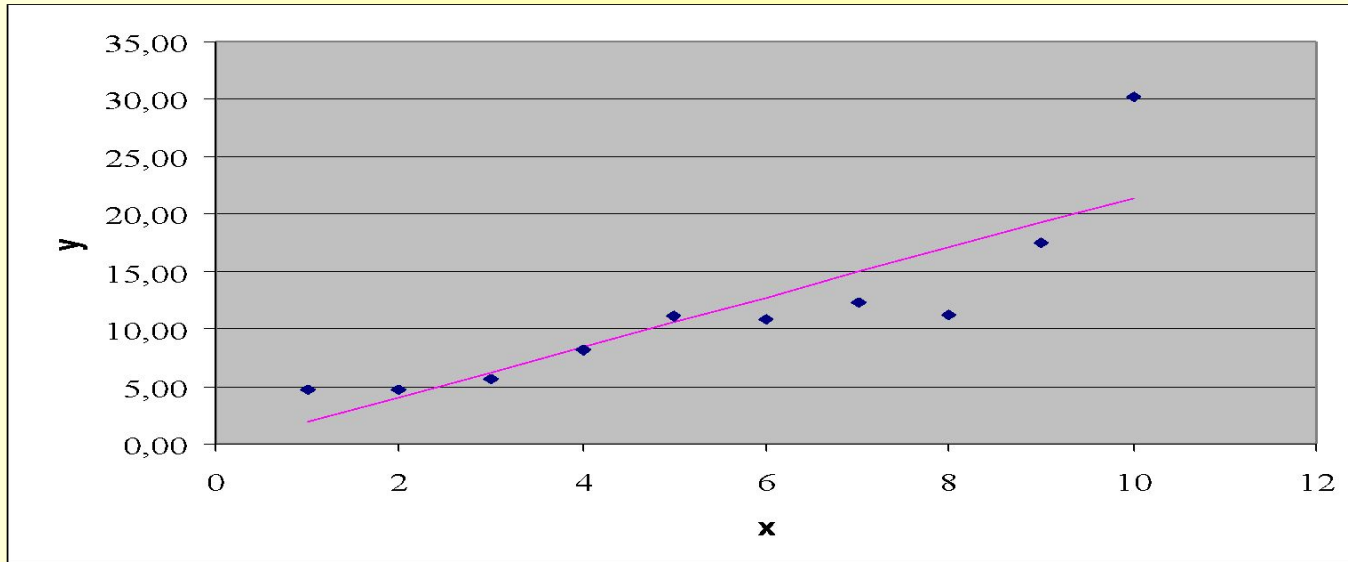
Пусть в генеральной совокупности зависимость между переменными Y и X имеет вид:

$$(1) \quad Y = b_0 + b_1 X + \varepsilon$$

Типичный вид **корреляционного поля** данных наблюдений для выборки значений (X_i, Y_i) , объемом n из генеральной совокупности:



Цель МНК – выполнить наилучшую ” подгонку” прямой под данные наблюдений



Метод наименьших квадратов (МНК) решает задачу «наилучшей» аппроксимации данных наблюдений линейной зависимостью :

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

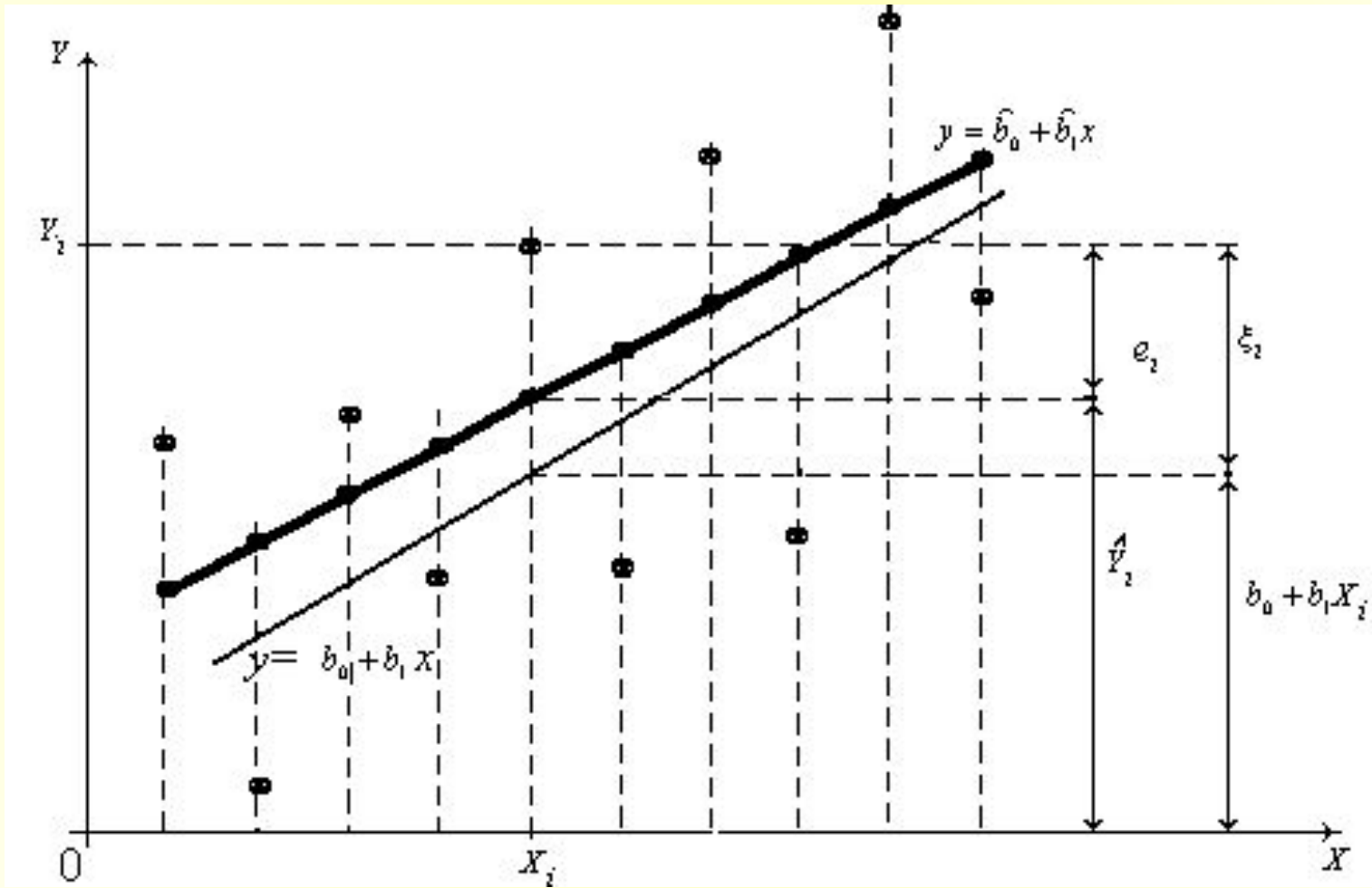
(2) 5

Суть МНК:

следует найти такие коэффициенты уравнения регрессии, чтобы **сумма** квадратов отклонений эмпирических значений результативного признака от расчетных, вычисленных по уравнению, была бы **минимальной**, т.е.

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 =$$
$$= \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \rightarrow \min$$

Корреляционное поле. Истинная зависимость Y от X . МНК-прямая



Формулы для вычисления эмпирических коэффициентов регрессии, полученные по МНК

$$(3) \quad \begin{cases} b_1 = \frac{\widehat{\text{cov}}(x, y)}{\sigma_x^2}; \\ b_0 = \bar{y} - \bar{x} b_1. \end{cases}$$

Пример 1.

Есть данные о количестве внесенных удобрений (Y , кг/га) и урожайности пшеницы (X , ц/га) по десяти фермерским хозяйствам:

x_i	1	2	3	4	5	6	7	8	9	10
y_i	19	16	19	14	23	22	27	33	28	29

Считая форму связи между признаками Y и X линейной, 1) найти по МНК эмпирические коэффициенты регрессии; 2) построить корреляционное поле и эмпирическую линию регрессии; 3) вычислить значение функции $S(\overset{\square}{b}_0, \overset{\square}{b}_1)$

Перечень средств MS Excel

1. Встроенная статистическая функция MS Excel **КОВАР(массив_1;массив_2)**.
2. Встроенная математическая функция MS Excel **СУММКВРАЗН(массив_1;массив_2)**.
3. Встроенная статистическая функция **ЛИНЕЙН** (известные_значения_y ; известные_значения_x; конст; статистика).
4. Мастер диаграмм.

Краткие сведения

1. Функция **СУМКВРАЗН(массив_1;массив_2)** вычисляет сумму квадратов разностей между соответствующими компонентами массивов.
2. Функция **КОВАР(массив_1;массив_2)** находит выборочную ковариацию данных наблюдений, представленных в массивах.
3. Функция **ЛИНЕЙН (известные_значения_y ; известные_значения_x; конст; статистика)** находит по МНК оценки коэффициентов регрессии и дополнительную регрессионную статистику.
4. **Точечная диаграмма** позволяет визуализировать точки из двумерной совокупности.

Алгоритм применения функции ЛИНЕЙН

1. Занести в ячейки с адресами **B1:K1** рабочего листа MS Excel значения X , а в ячейки **B2:K2** – значения Y .
2. Выделить интервал из двух ячеек **A6:B6** .
Вставка -> Функция.
3. Выбрать категорию (вид функции) – **«Статистические»**. Затем в списке с названиями статистических функций, упорядоченными по алфавиту, найти функцию **ЛИНЕЙН**.
4. **ЛИНЕЙН -> ОК.**

Алгоритм применения функции ЛИНЕЙН (продолжение)

5. Задать значения четырех аргументов функции **ЛИНЕЙН**.
Первый аргумент: **известные_значения_u**
-> B2:K2.
Второй аргумент: **известные_значения_x**
-> B1:K1.
6. Задать значения необязательных логических аргументов **конст** и **статистика** по умолчанию, т.е.: **конст -> 1**;
статистика -> 0.
7. **ОК**.

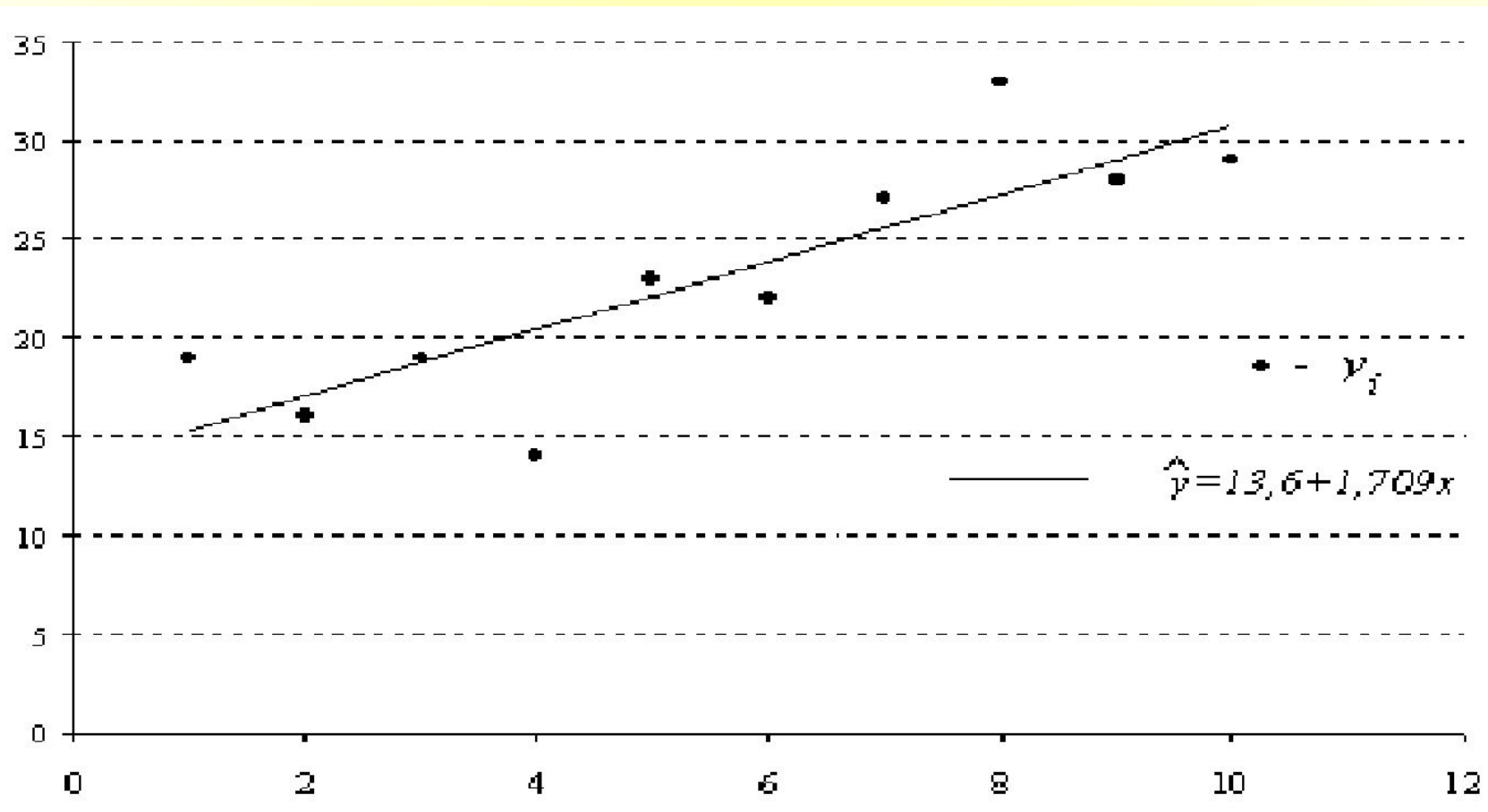
Результаты применения функции ЛИНЕЙН

- В левой из двух выделенных ячеек (А6) появится первый элемент итоговой таблицы – величина коэффициента b_1 .
- Для того, чтобы получить всю таблицу, следует сначала нажать клавишу **F2**, а затем – комбинацию клавиш: **CTRL+SHIFT+ENTER**.
- В ячейке В6 появится значение коэффициента b_0 .

Результаты оценивания регрессии. Рабочий лист MS Excel с исходными данными

B3		fx =\$B\$6+\$A\$6*B1									
	A	B	C	D	E	F	G	H	I	J	K
1	x_i	1	2	3	4	5	6	7	8	9	10
2	y_i	19	16	19	14	23	22	27	33	28	29
3	\hat{y}_i	15,31	17,02	18,73	20,44	22,15	23,85	25,56	27,27	28,98	30,69
4											
5			S(12,46;2,04)								
6	1,709	13,60	99,02								

Визуализация решения, найденного с помощью MS Excel



Основные варианты задания логических аргументов функции ЛИНЕЙН

Варианты вывода результатов функции ЛИНЕЙН для случая парной линейной регрессии

сокращенный

[конст = 1 (или истина),
статистика = 0
(или ложь)]

b_1	b_0
-------	-------

полный

[конст = 1 (или истина),
статистика = 1
(или истина)]

b_1	b_0
с.о. (\hat{b}_1)	с.о. (\hat{b}_0)
R^2	S_e
$F_{\text{набл}}$	k_2
RSS	ESS

Заключение

МНК позволяет получать надежные статистические оценки теоретических коэффициентов регрессии.