

# ЛЕКЦИЯ 2. Метод наименьших квадратов: как подогнать модель под данные

## Введение

В прошлой лекции вы узнали, что такое простая линейная регрессия: модель, которая позволяет спрогнозировать зависимую переменную  $y$ , основываясь на одной объясняющей переменной  $x$ .

Но возникает логичный вопрос:

**Как именно рисуется линия, которая “наилучшим образом” проходит через точки?**

Почему она — одна? Почему именно эта?

Ответ — в **Методе наименьших квадратов (МНК)**.

## Что делает МНК?

Метод наименьших квадратов подбирает такую прямую, чтобы **суммарное отклонение всех точек от линии было минимальным**.

Но не просто “отклонение” — а **квадраты отклонений**. Почему?

Чтобы **избежать эффекта “плюс-минус”**, когда одни отклонения положительные, а другие — отрицательные.

Квадраты делают все ошибки положительными и “наказывают” за сильные отклонения.



## Формула прямой линии

$$y = b_0 + b_1 * x$$

Где:

$y$  — предсказанное значение (например, оценка сна),

$x$  — значение признака (например, экранное время),

$b_0$  — точка пересечения с осью  $y$  (когда  $x = 0$ ),

$b_1$  — коэффициент наклона (показывает, насколько  $y$  меняется при изменении  $x$ ).

## Что минимизирует МНК?

МНК минимизирует сумму квадратов ошибок:

$$\sum (y_i - \hat{y}_i)^2$$

Где:

$y_i$  — фактическое значение

$\hat{y}_i$  — предсказанное значение по модели

$(y_i - \hat{y}_i)$  — остаток (ошибка предсказания)

## Пример гипотезы

Гипотеза:

Чем чаще студент использует экран перед сном, тем ниже качество его сна.

### Пример данных

Респондент	Экран (x)	Качество сна (y)
1	2	7
2	4	6
3	6	5
4	8	3
5	10	2

## Как рассчитать МНК вручную (понимание, не для вычисления)

Формулы коэффициентов:

$$b_1 = (\sum(x - \bar{x})(y - \bar{y})) / \sum(x - \bar{x})^2$$

$$b_0 = \bar{y} - b_1 * \bar{x}$$

Где:

$\bar{x}$  — среднее значение  $x$

$\bar{y}$  — среднее значение  $y$

## Как применить МНК в Excel и Google Sheets

**Google Sheets (английский язык):**

```
=LINEST(B2:B6, A2:A6, TRUE, TRUE)
```

**Excel (русский язык):**

```
=ЛИНЕЙН(B2:B6; A2:A6; ИСТИНА; ИСТИНА)
```

Где:

A2:A6 — значения  $x$  (время у экрана)

B2:B6 — значения  $y$  (качество сна)

## Что вы получаете

Например:

$$b_1 = -0.6$$

$$b_0 = 8.5$$

$$R^2 = 0.92$$

**Интерпретация:**

$b_1 = -0.6 \rightarrow$  каждое дополнительное "единица экрана" снижает качество сна на 0.6 балла

$b_0 = 8.5 \rightarrow$  если экранного времени = 0, качество сна  $\approx 8.5$

$R^2 = 0.92 \rightarrow$  модель объясняет 92% изменений в качестве сна

## Как читать результат

Показатель	Что он говорит
$b_1 < 0$	Связь отрицательная (больше X — меньше Y)
$b_1 > 0$	Связь положительная
$b_1 \approx 0$	Связь слабая или отсутствует
$R^2 \rightarrow 1$	Модель хорошо объясняет данные
$R^2 \rightarrow 0$	Модель почти ничего не объясняет

## Визуализация: как выглядит линия МНК

Вы можете построить график с точками (x, y) и добавить **трендовую линию** (Trendline):

### В Google Sheets:

Вставка → Диаграмма → “Линейная диаграмма”

Настройки → “Трендовая линия” → Тип: линейная

Показать уравнение и  $R^2$

### В Excel:

Постройте график → Щелчок по линии → “Добавить трендовую линию”

Отметьте “Показать уравнение на диаграмме” и “Показать  $R^2$ ”

## Лайфхаки: как использовать ИИ

Инструмент	Что делает
Excel AI / Copilot	Сам рассчитает регрессию
ChatGPT (Data Analyst)	Объяснит коэффициенты и подскажет шаги
Google Sheets Explore	Найдет зависимости и покажет тренды

### Категорически запрещено:

Вставлять формулу без понимания

Игнорировать смысл коэффициентов

“Придумывать” значения для красивой линии

Не интерпретировать результат в контексте гипотезы

## Вывод

Метод наименьших квадратов — это не просто техника, это **сердце линейной регрессии**.

Именно он превращает данные в модель, а модель — в анализ.

Понимая его, вы больше не “строите график” — вы **контролируете поведение модели**.