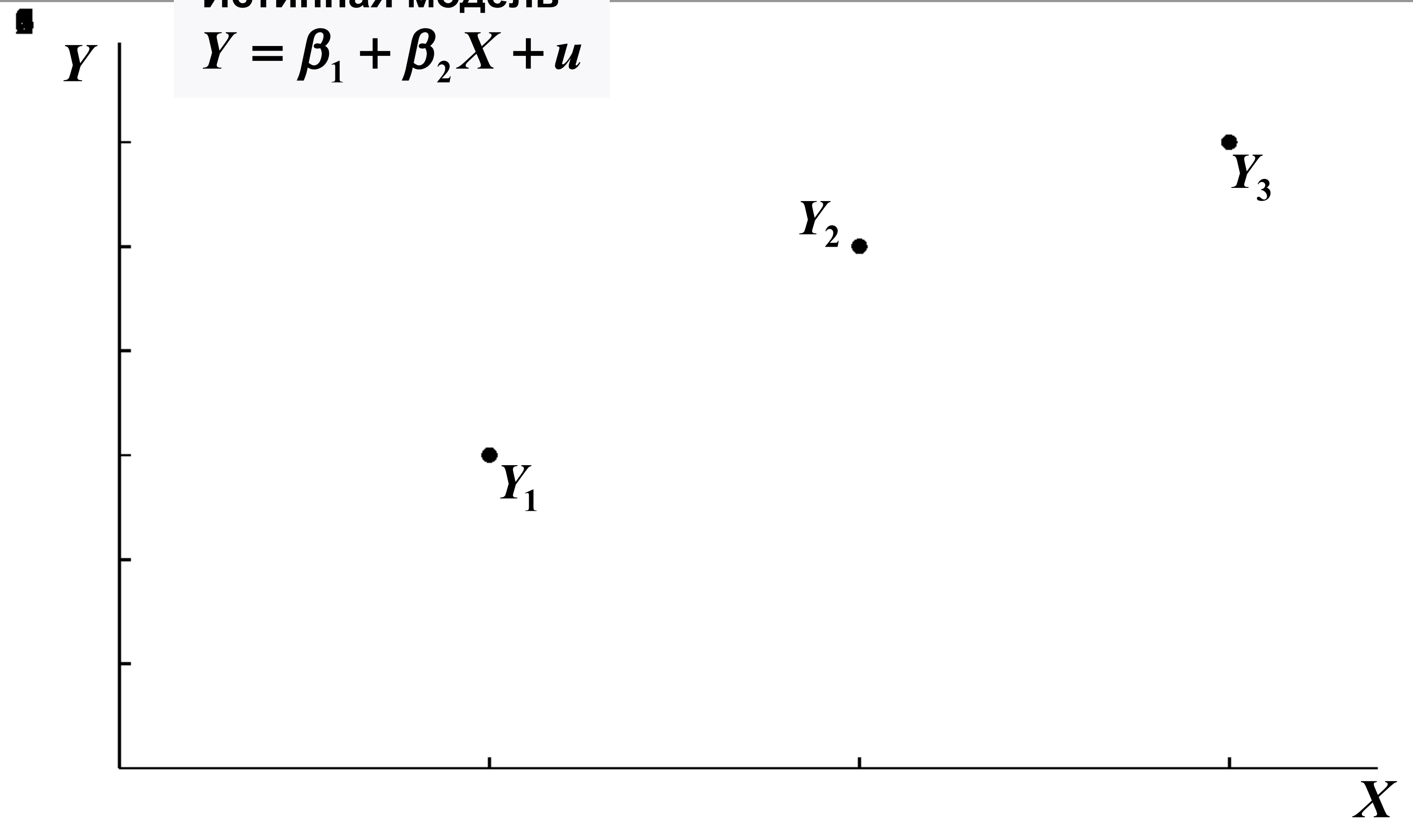
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

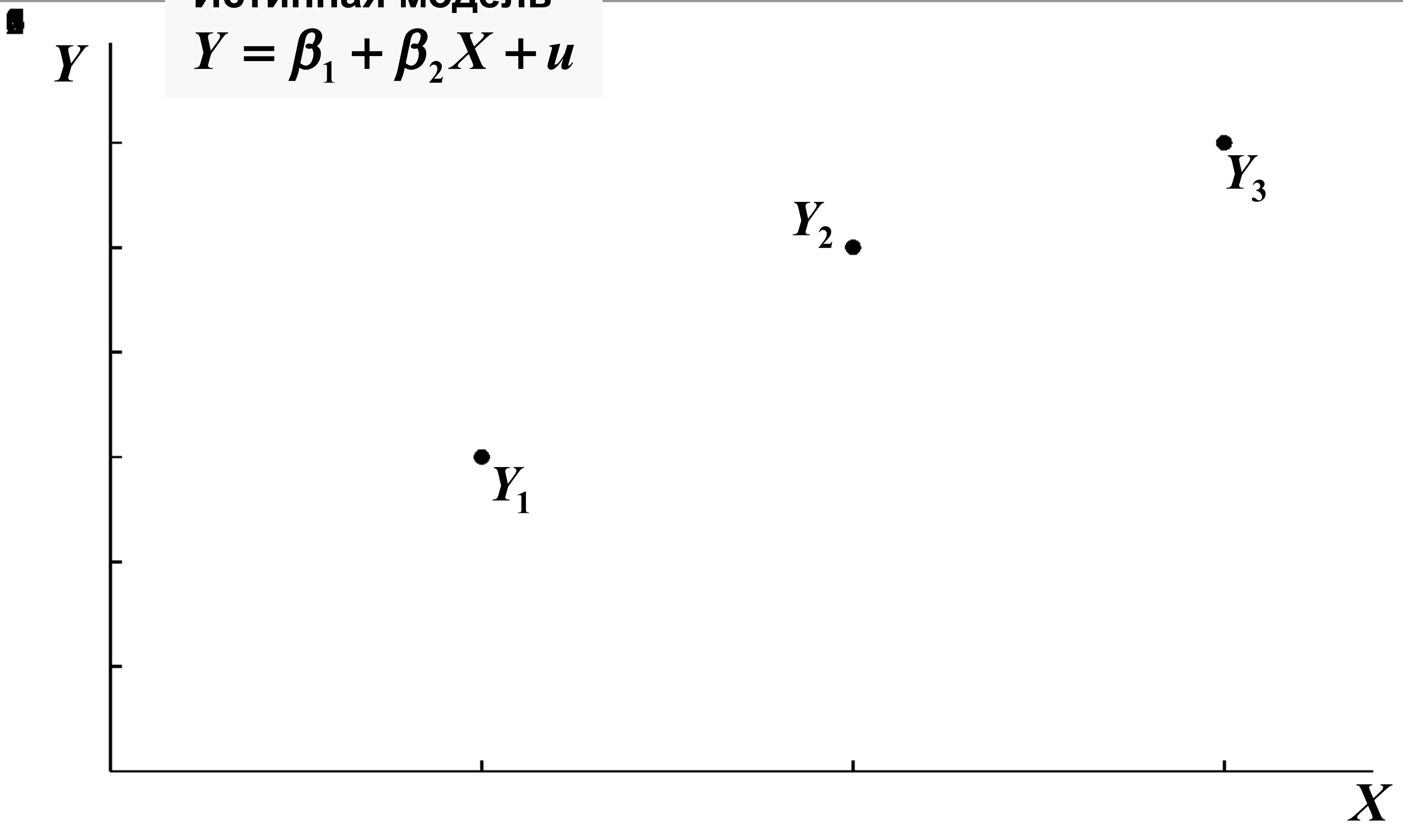


Эта презентация показывает, как рассчитываются коэффициенты регрессии для простой модели регрессии, используя критерий наименьших квадратов (OLS, обычный метод наименьших квадратов)

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$



Начнем с численного примера, имеющего всего три наблюдениями: (1,3), (2,5) и (3,6).

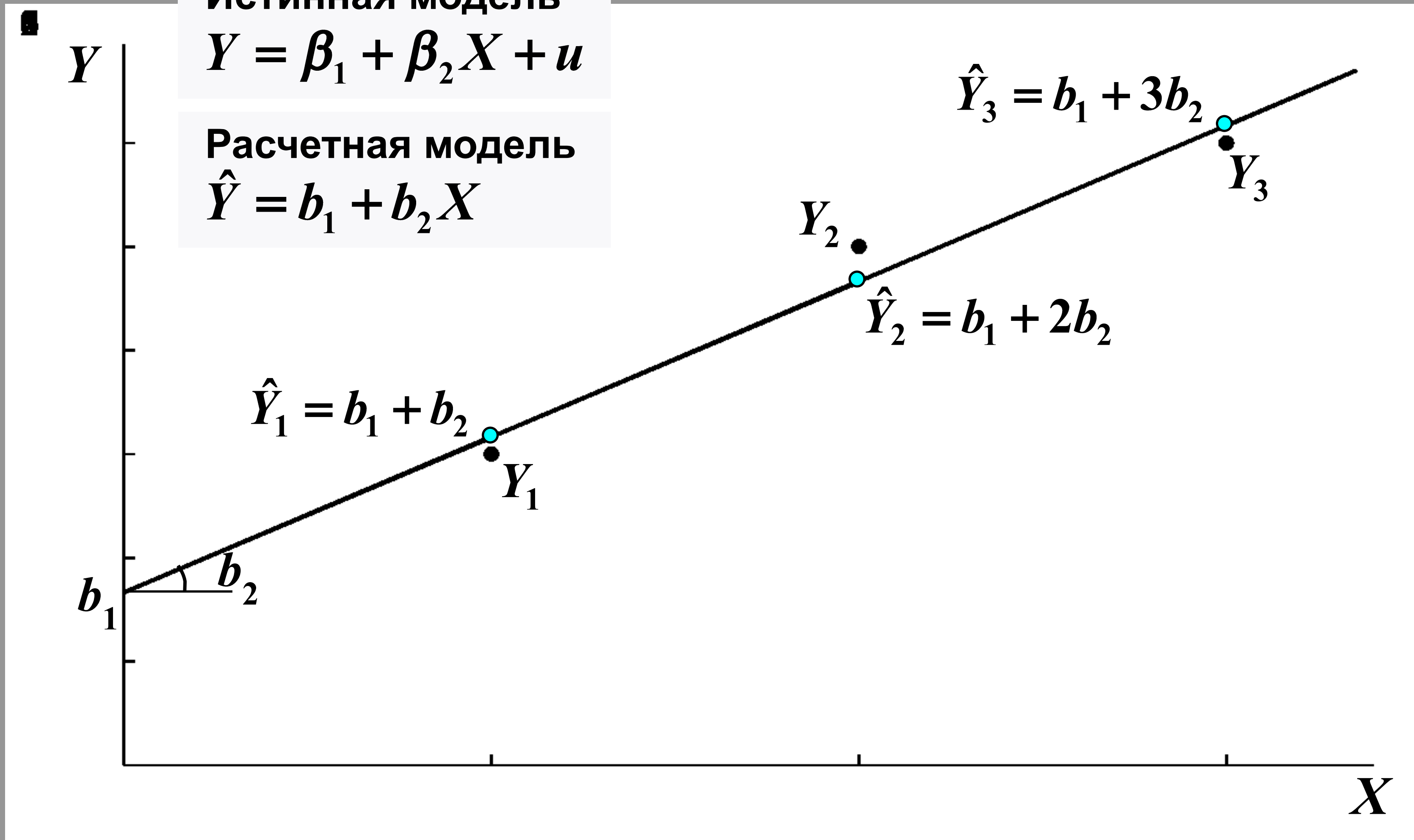
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_1 + b_2 X$$



Написав установленную регрессию как $\hat{Y} = b_1 + b_2X$, мы определим значения b_1 и b_2 так, чтобы минимизировать *RSS*, сумму квадратов остатков.

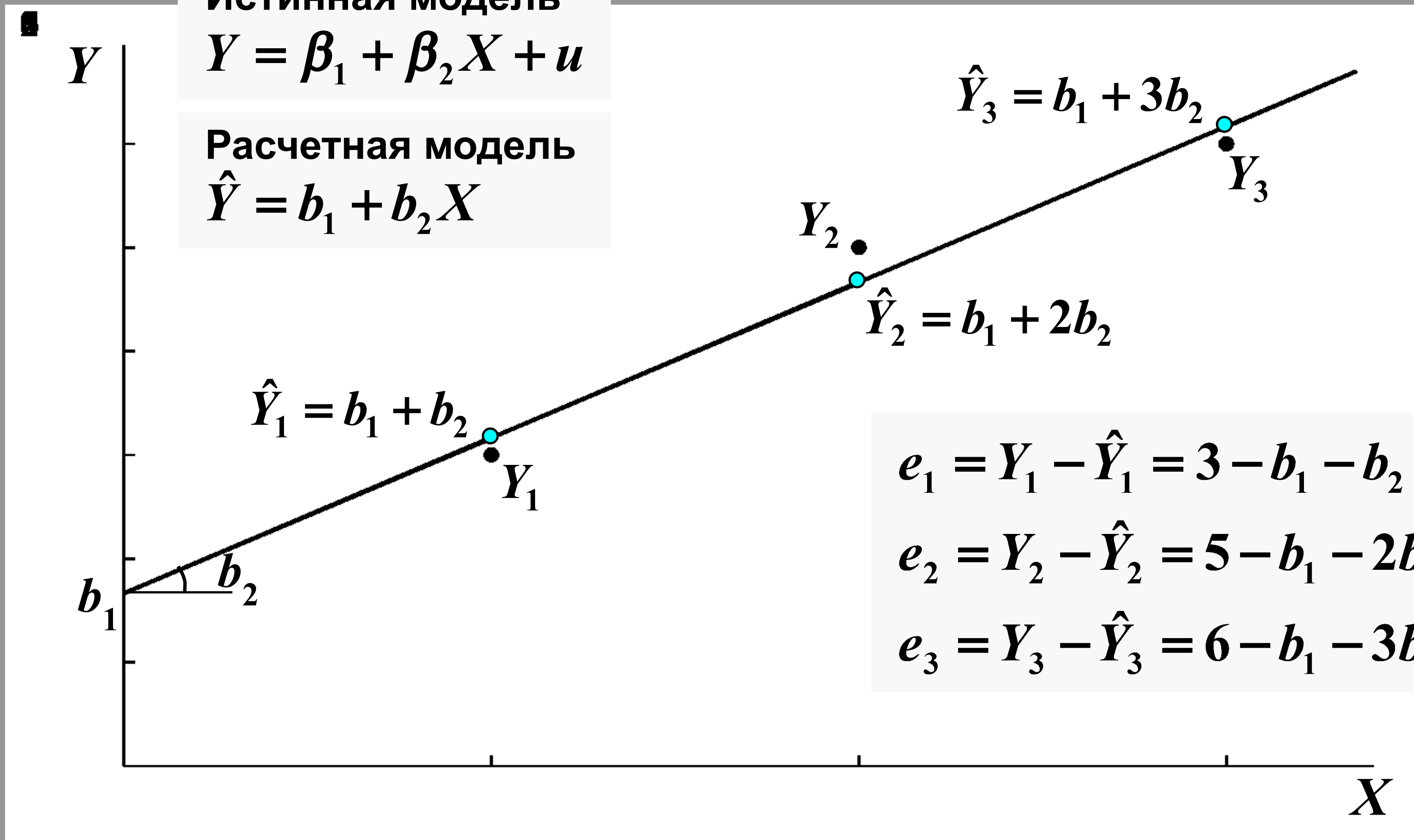
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_1 + b_2 X$$



$$e_1 = Y_1 - \hat{Y}_1 = 3 - b_1 - b_2$$

$$e_2 = Y_2 - \hat{Y}_2 = 5 - b_1 - 2b_2$$

$$e_3 = Y_3 - \hat{Y}_3 = 6 - b_1 - 3b_2$$

Учитывая наш выбор b_1 и b_2, остатки показаны на рисунке.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$$

$$e_1 = Y_1 - \hat{Y}_1 = 3 - b_1 - b_2$$

$$e_2 = Y_2 - \hat{Y}_2 = 5 - b_1 - 2b_2$$

$$e_3 = Y_3 - \hat{Y}_3 = 6 - b_1 - 3b_2$$

Таким образом, показана сумма квадратов остатков.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$$

$$= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2$$

$$+ 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2$$

$$+ 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2$$

Раскрыты скобки.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$
\begin{aligned}
RSS = e_1^2 + e_2^2 + e_3^2 &= (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2 \\
&= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\
&\quad + 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\
&\quad + 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\
&= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2
\end{aligned}
$$

Суммированы подобные члены.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$$

$$\begin{aligned}
&= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\
&+ 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\
&+ 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\
&= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2
\end{aligned}$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0$$

Как минимум, частные производные *RSS* по отношению к b_1 и b_2 должны быть равны нулю. (Мы также должны проверить условие второго порядка.)

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$$

$$\begin{aligned} &= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\ &\quad + 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\ &\quad + 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\ &= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2 \end{aligned}$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0$$

Условия первого порядка дают нам два уравнения с двумя неизвестными.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$$

$$\begin{aligned}
&= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\
&\quad + 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\
&\quad + 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\
&= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2
\end{aligned}$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0$$

$$\therefore b_1 = 1.67, \quad b_2 = 1.50$$

Решая их, мы находим, что RSS минимально, когда b_1 и b_2 равны 1.67 and 1.50 соответственно.

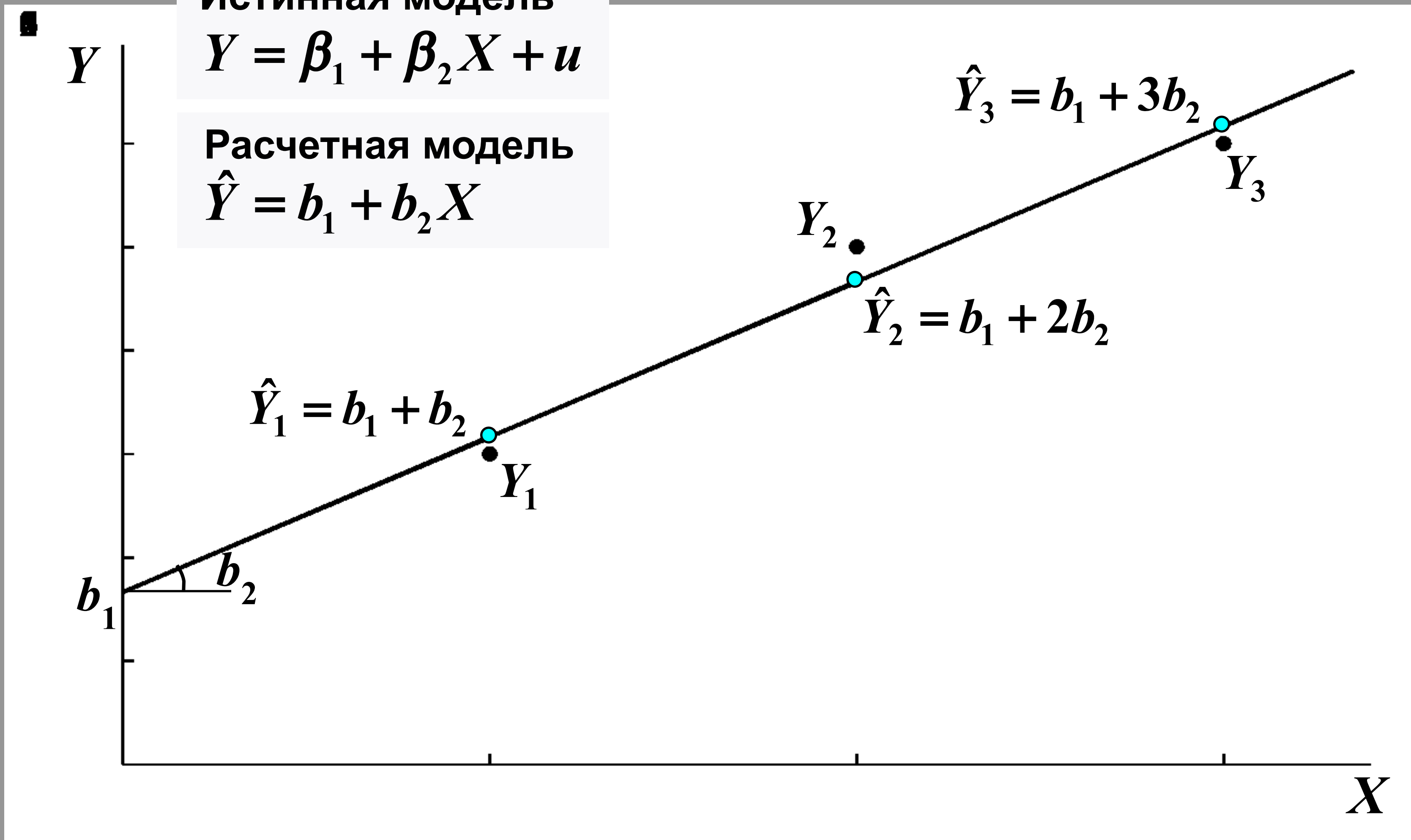
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_1 + b_2 X$$



Это диаграмма дисперсии (рассеяния).

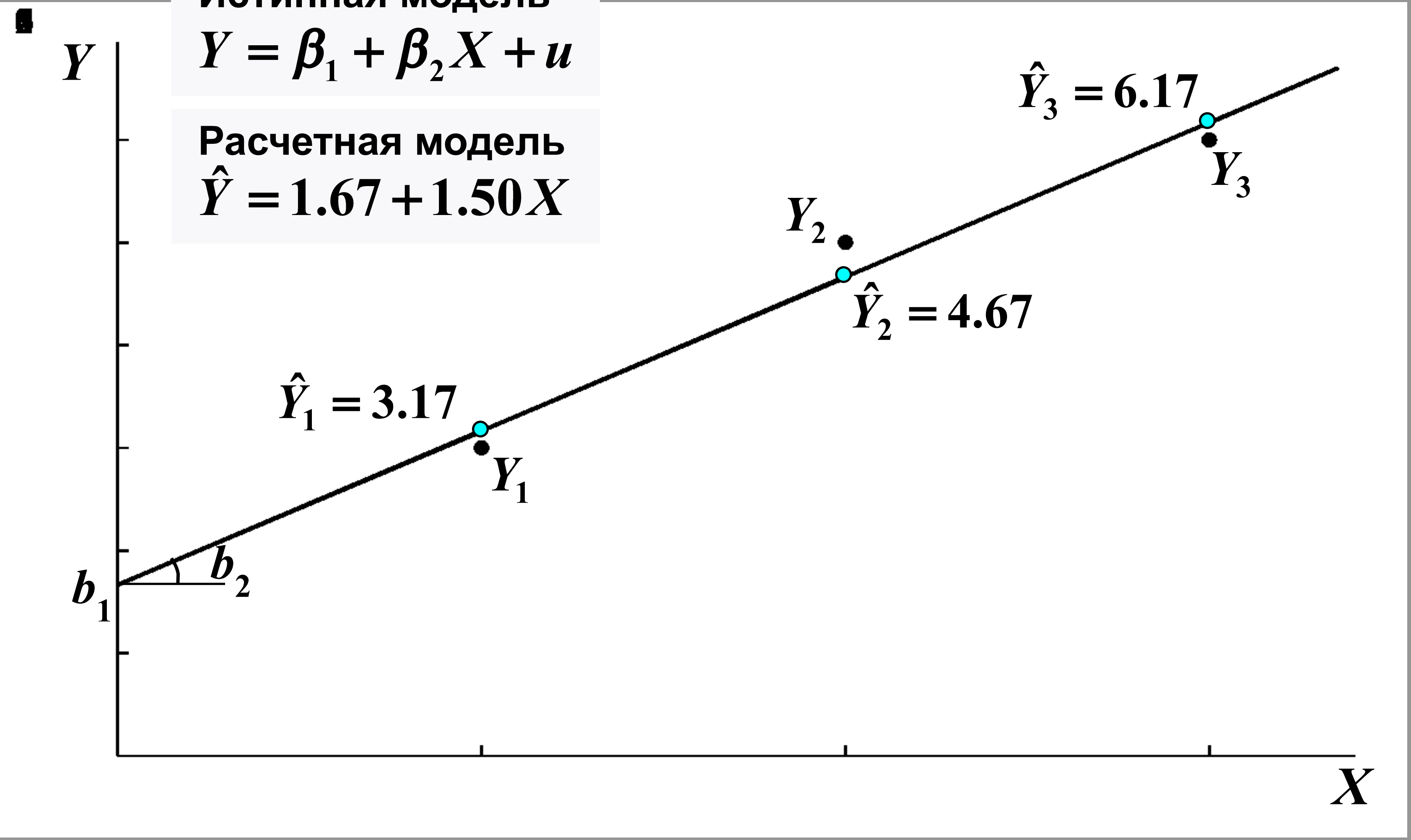
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = 1.67 + 1.50X$$



Показаны расчетная линия и расчетные значения *Y*.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$$

$$= \quad 9 + \quad b_1^2 + \quad b_2^2 - \quad 6b_1 - \quad 6b_2 + \; 2b_1 b_2$$

$$+25 + \quad b_1^2 + \; 4b_2^2 - 10b_1 - 20b_2 + \; 4b_1 b_2$$

$$+36 + \quad b_1^2 + \; 9b_2^2 - 12b_1 - 36b_2 + \; 6b_1 b_2$$

$$= \; 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1 b_2$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0$$

$$\therefore b_1 = 1.67, \quad b_2 = 1.50$$

Прежде чем перейти к общему случаю, нужно сделать небольшое, но важное математическое замечание.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$$

$$\begin{aligned}
&= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\
&+ 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\
&+ 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\
&= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2
\end{aligned}$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0$$

$$\therefore b_1 = 1.67, \quad b_2 = 1.50$$

Когда мы устанавливаем выражение для *RSS*, мы делаем это как функцию от b_1 и b_2. На этом этапе b_1 и b_2 не являются конкретными значениями. Наша задача - определить конкретные значения, которые минимизируют *RSS.*

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$\begin{aligned}
RSS &= e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2 \\
&= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\
&\quad + 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\
&\quad + 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\
&= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2
\end{aligned}$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0$$

$$\therefore b_1 = 1.67, \quad b_2 = 1.50$$

Мы должны дать этим значениям собственные названия, чтобы отличать их от всех остальных.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$
\begin{aligned}
RSS &= e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2 \\
&= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\
&\quad + 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\
&\quad + 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\
&= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2
\end{aligned}
$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1^{\mathrm{OLS}} + 12b_2^{\mathrm{OLS}} - 28 = 0$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1^{\mathrm{OLS}} + 28b_2^{\mathrm{OLS}} - 62 = 0$$

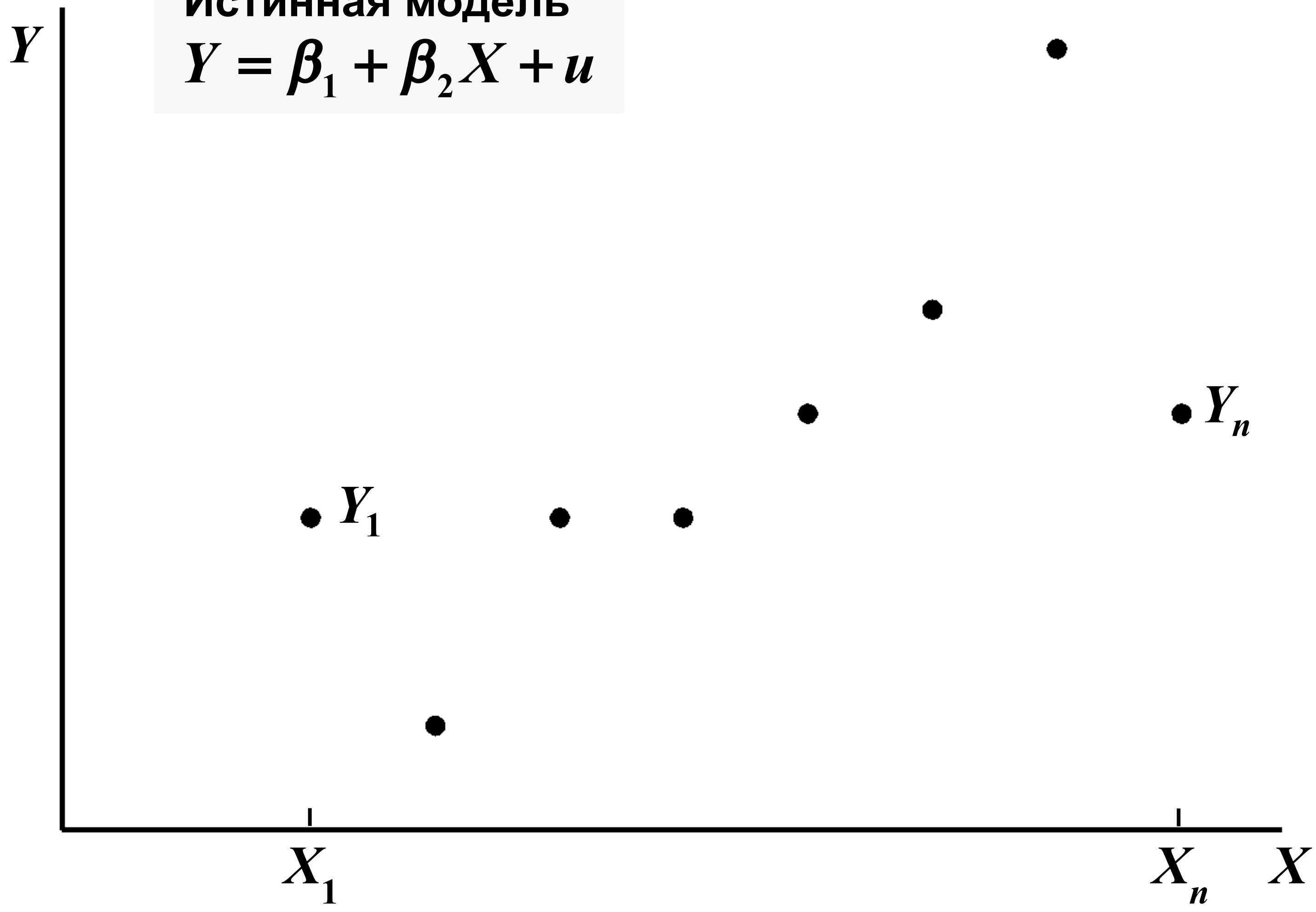
$$\therefore b_1^{\mathrm{OLS}} = 1.67, \quad b_2^{\mathrm{OLS}} = 1.50$$

Обозначим значения, которые минимизируют RSS, следующим образом: b_1^{OLS} and b_2^{OLS}.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

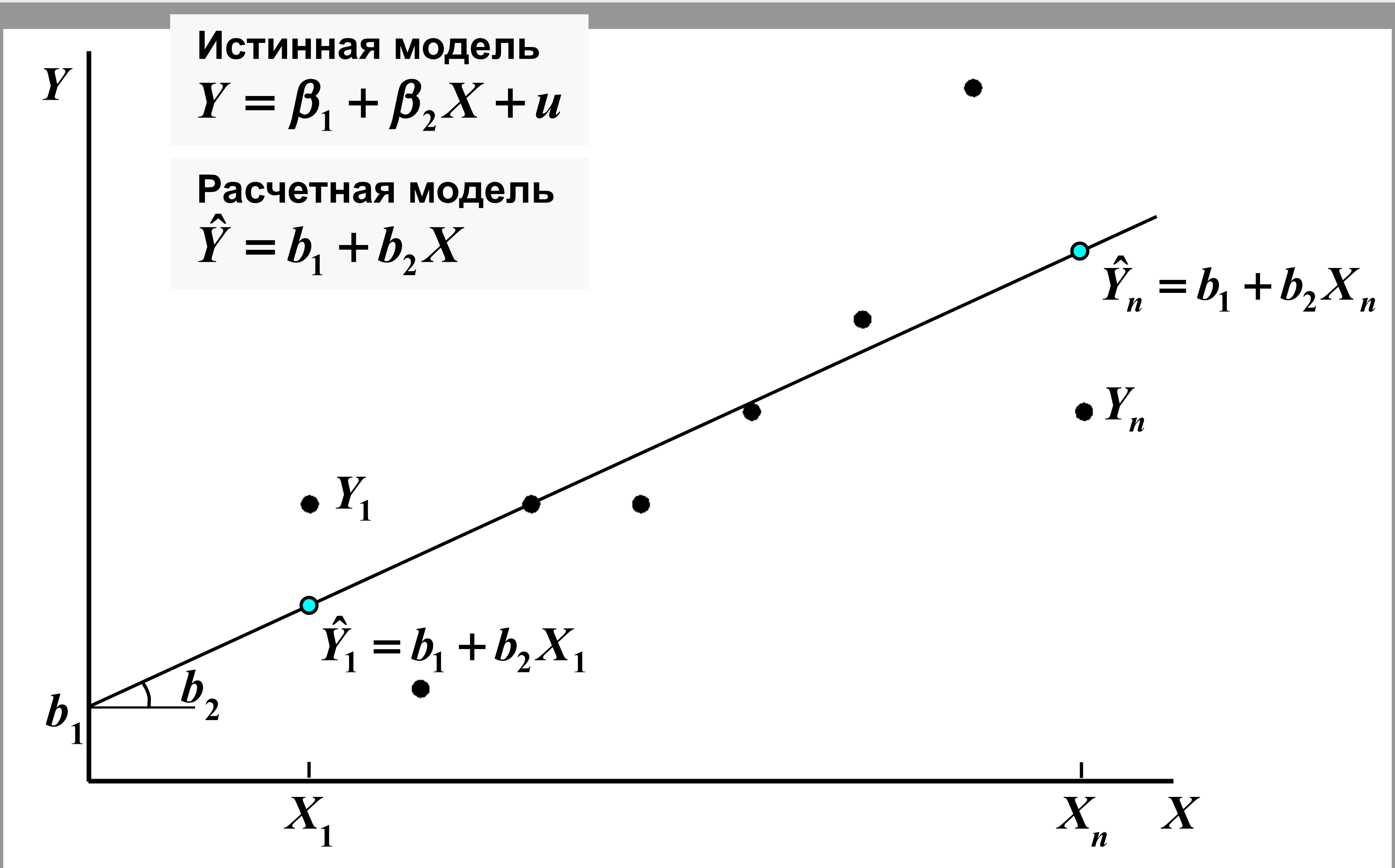
Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$



Теперь мы перейдем к общему случаю с n наблюдениями.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ



Учитывая коэффициенты b_1 и b_2, на слайде мы получили расчетную линию.

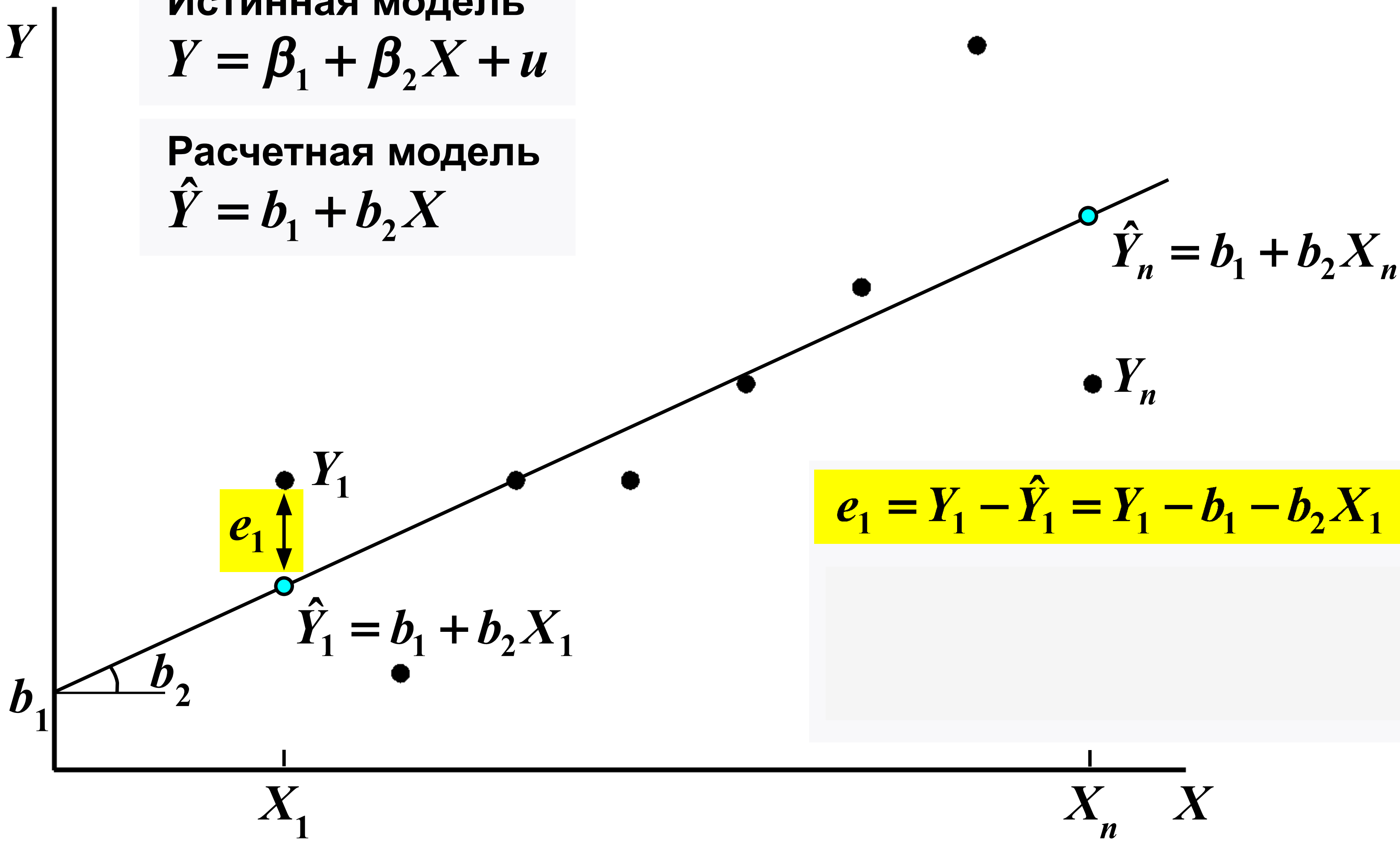
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_1 + b_2 X$$



$$e_1 = Y_1 - \hat{Y}_1 = Y_1 - b_1 - b_2 X_1$$

Получен остаток для первого наблюдения.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Аналогично мы определяем остатки для остальных наблюдений.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$
\begin{aligned}
RSS = e_1^2 + e_2^2 + e_3^2 &= (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2 \\
&= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\
&\quad + 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\
&\quad + 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\
&= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2
\end{aligned}
$$

$$RSS = e_1^2 + \ldots + e_n^2 = (Y_1 - b_1 - b_2X_1)^2 + \ldots + (Y_n - b_1 - b_2X_n)^2$$

RSS, сумма квадратов остатков, определена для общего случая. Для сравнения приведены данные численного примера.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = e_1^2 + e_2^2 + e_3^2 = (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2$$

$$= 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2$$

$$+ 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2$$

$$+ 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2$$

$$= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2$$

$$RSS = e_1^2 + \ldots + e_n^2 = (Y_1 - b_1 - b_2X_1)^2 + \ldots + (Y_n - b_1 - b_2X_n)^2$$

$$= Y_1^2 + b_1^2 + b_2^2X_1^2 - 2b_1Y_1 - 2b_2X_1Y_1 + 2b_1b_2X_1$$

$$+ \ldots$$

$$+ Y_n^2 + b_1^2 + b_2^2X_n^2 - 2b_1Y_n - 2b_2X_nY_n + 2b_1b_2X_n$$

Раскрываем скобки, возведя соответствующие выражения в квадрат.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$
\begin{aligned}
RSS = e_1^2 + e_2^2 + e_3^2 &= (3 - b_1 - b_2)^2 + (5 - b_1 - 2b_2)^2 + (6 - b_1 - 3b_2)^2 \\
&= \ \ 9 + b_1^2 + b_2^2 - 6b_1 - 6b_2 + 2b_1b_2 \\
&\ \ + 25 + b_1^2 + 4b_2^2 - 10b_1 - 20b_2 + 4b_1b_2 \\
&\ \ + 36 + b_1^2 + 9b_2^2 - 12b_1 - 36b_2 + 6b_1b_2 \\
&= 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2
\end{aligned}
$$

$$
\begin{aligned}
RSS = e_1^2 + \ldots + e_n^2 &= (Y_1 - b_1 - b_2X_1)^2 + \ldots + (Y_n - b_1 - b_2X_n)^2 \\
&= Y_1^2 + b_1^2 + b_2^2X_1^2 - 2b_1Y_1 - 2b_2X_1Y_1 + 2b_1b_2X_1 \\
&\ \ + \ldots \\
&\ \ + Y_n^2 + b_1^2 + b_2^2X_n^2 - 2b_1Y_n - 2b_2X_nY_n + 2b_1b_2X_n \\
&= \sum Y_i^2 + nb_1^2 + b_2^2\sum X_i^2 - 2b_1\sum Y_i - 2b_2\sum X_iY_i + 2b_1b_2\sum X_i
\end{aligned}
$$

Приводим подобные члены.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2$$

$$\left.\begin{array}{l} \dfrac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0 \\ \dfrac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0 \end{array}\right\} \therefore b_1 = 1.67, \quad b_2 = 1.50$$

$$RSS = \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i$$

В этом уравнении наблюдения по *X* и *Y* являются просто данными, которые определяют коэффициенты в выражении для *RSS*.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2$$

$$\left.\begin{aligned} \frac{\partial RSS}{\partial b_1} = 0 \quad &\Rightarrow \quad 6b_1 + 12b_2 - 28 = 0 \\ \frac{\partial RSS}{\partial b_2} = 0 \quad &\Rightarrow \quad 12b_1 + 28b_2 - 62 = 0 \end{aligned}\right\} \therefore b_1 = 1.67, \quad b_2 = 1.50$$

$$RSS = \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i$$

Переменными в уравнении для *RSS* являются b_1 и b_2. Обычно b_1 and b_2 являются константами, а *X* и *Y* - переменными.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2$$

$$\left.\begin{aligned} \frac{\partial RSS}{\partial b_1} = 0 \quad &\Rightarrow \quad 6b_1 + 12b_2 - 28 = 0 \\ \frac{\partial RSS}{\partial b_2} = 0 \quad &\Rightarrow \quad 12b_1 + 28b_2 - 62 = 0 \end{aligned}\right\} \therefore b_1 = 1.67, \quad b_2 = 1.50$$

$$RSS = \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i$$

Если есть сомнения, то можно сравнить общий случай с нашим конкретным примером.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2$$

$$\left.\begin{array}{l} \dfrac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0 \\ \dfrac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0 \end{array}\right\} \therefore b_1 = 1.67, \quad b_2 = 1.50$$

$$RSS = \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 2nb_1 - 2\sum Y_i + 2b_2 \sum X_i = 0$$

Первая производная по b_1.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2$$

$$\left.\begin{array}{l} \dfrac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 6b_1 + 12b_2 - 28 = 0 \\[2ex] \dfrac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 12b_1 + 28b_2 - 62 = 0 \end{array}\right\} \therefore b_1 = 1.67, \quad b_2 = 1.50$$

$$RSS = \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 2nb_1 - 2\sum Y_i + 2b_2 \sum X_i = 0$$

$$nb_1 = \sum Y_i - b_2 \sum X_i \qquad b_1 = \overline{Y} - b_2 \overline{X}$$

Получаем окончательное выражение для b_1 .

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$RSS = 70 + 3b_1^2 + 14b_2^2 - 28b_1 - 62b_2 + 12b_1b_2$$

$$\left.\begin{aligned} \frac{\partial RSS}{\partial b_1} = 0 \quad &\Rightarrow \quad 6b_1 + 12b_2 - 28 = 0 \\ \frac{\partial RSS}{\partial b_2} = 0 \quad &\Rightarrow \quad 12b_1 + 28b_2 - 62 = 0 \end{aligned}\right\} \therefore b_1 = 1.67, \quad b_2 = 1.50$$

$$RSS = \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 2nb_1 - 2\sum Y_i + 2b_2 \sum X_i = 0$$

$$nb_1 = \sum Y_i - b_2 \sum X_i \qquad b_1 = \overline{Y} - b_2 \overline{X}$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 2b_2 \sum X_i^2 - 2\sum X_i Y_i + 2b_1 \sum X_i = 0$$

Первая производная по b_2.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 2b_2 \sum X_i^2 - 2\sum X_i Y_i + 2b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + b_1 \sum X_i = 0$$

$$RSS = \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 2nb_1 - 2\sum Y_i + 2b_2 \sum X_i = 0$$

$$nb_1 = \sum Y_i - b_2 \sum X_i \qquad b_1 = \overline{Y} - b_2 \overline{X}$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 2b_2 \sum X_i^2 - 2\sum X_i Y_i + 2b_1 \sum X_i = 0$$

Разделим на 2.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 2b_2 \sum X_i^2 - 2\sum X_i Y_i + 2b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + (\overline{Y} - b_2 \overline{X}) \sum X_i = 0$$

$$RSS = \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_i^2 - 2b_1 \sum Y_i - 2b_2 \sum X_i Y_i + 2b_1 b_2 \sum X_i$$

$$\frac{\partial RSS}{\partial b_1} = 0 \quad \Rightarrow \quad 2nb_1 - 2\sum Y_i + 2b_2 \sum X_i = 0$$

$$nb_1 = \sum Y_i - b_2 \sum X_i \qquad b_1 = \overline{Y} - b_2 \overline{X}$$

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 2b_2 \sum X_i^2 - 2\sum X_i Y_i + 2b_1 \sum X_i = 0$$

Теперь заменим b_1 на выражение, полученное для него, и получим уравнение для b_2.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 2b_2 \sum X_i^2 - 2\sum X_i Y_i + 2b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + (\overline{Y} - b_2 \overline{X}) \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + (\overline{Y} - b_2 \overline{X}) n\overline{X} = 0$$

$$\overline{X} = \frac{\sum X_i}{n}$$

$$\sum X_i = n\overline{X}$$

Используем определение среднего значения выборки.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 2b_2 \sum X_i^2 - 2\sum X_i Y_i + 2b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + (\overline{Y} - b_2 \overline{X}) \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + (\overline{Y} - b_2 \overline{X}) n\overline{X} = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + n\overline{X}\overline{Y} - nb_2 \overline{X}^2 = 0$$

Раскрыты скобки в третьем слагаемом.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$\frac{\partial RSS}{\partial b_2} = 0 \quad \Rightarrow \quad 2b_2 \sum X_i^2 - 2\sum X_i Y_i + 2b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + b_1 \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + (\overline{Y} - b_2 \overline{X}) \sum X_i = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + (\overline{Y} - b_2 \overline{X}) n\overline{X} = 0$$

$$b_2 \sum X_i^2 - \sum X_i Y_i + n\overline{X}\,\overline{Y} - nb_2 \overline{X}^2 = 0$$

$$b_2 \left(\sum X_i^2 - n\overline{X}^2 \right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

В правой части остались слагаемые, не зависимые от b_2.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

Запишем наш результат в новый слайд.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\overline{Y}$$

$$b_2 = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

Получаем выражение для b_2.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

$$b_2 = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$b_2 = \frac{\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum \left(X_i - \overline{X}\right)^2}$$

На практике мы будем использовать альтернативное выражение. Мы продемонстрируем, что они эквивалентны.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

$$b_2 = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$b_2 = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\left(X_i - \overline{X}\right)^2}$$

$$\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \sum X_i Y_i - \sum X_i \overline{Y} - \sum \overline{X} Y_i + \sum \overline{X}\,\overline{Y}$$

Раскрываем скобки в числителе.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

$$b_2 = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$b_2 = \frac{\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum \left(X_i - \overline{X}\right)^2}$$

$$\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \sum X_i Y_i - \sum X_i \overline{Y} - \sum \overline{X} Y_i + \sum \overline{X}\,\overline{Y}$$

$$= \sum X_i Y_i - \overline{Y} \sum X_i - \overline{X} \sum Y_i + n\overline{X}\,\overline{Y}$$

Упрощаем числитель далее.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

$$b_2 = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$b_2 = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum\left(X_i - \overline{X}\right)^2}$$

$$\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \sum X_i Y_i - \sum X_i \overline{Y} - \sum \overline{X} Y_i + \sum \overline{X}\,\overline{Y}$$

$$= \sum X_i Y_i - \overline{Y}\sum X_i - \overline{X}\sum Y_i + n\overline{X}\,\overline{Y}$$

$$= \sum X_i Y_i - \overline{Y}\left(n\overline{X}\right) - \overline{X}\left(n\overline{Y}\right) + n\overline{X}\,\overline{Y}$$

$$\overline{X} = \frac{\sum X_i}{n}$$

$$\sum X_i = n\overline{X}$$

Упрощаем далее.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

$$b_2 = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

$$b_2 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

$$\begin{aligned}\sum (X_i - \overline{X})(Y_i - \overline{Y}) &= \sum X_i Y_i - \sum X_i \overline{Y} - \sum \overline{X} Y_i + \sum \overline{X}\,\overline{Y} \\ &= \sum X_i Y_i - \overline{Y}\sum X_i - \overline{X}\sum Y_i + n\overline{X}\,\overline{Y} \\ &= \sum X_i Y_i - \overline{Y}(n\overline{X}) - \overline{X}(n\overline{Y}) + n\overline{X}\,\overline{Y} \\ &= \sum X_i Y_i - n\overline{X}\,\overline{Y}\end{aligned}$$

Мы показали, что числители двух выражений одинаковы.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

$$b_2\left(\sum X_i^2 - n\overline{X}^2\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

$$b_2 = \frac{\sum X_i Y_i - n\overline{X}\,\overline{Y}}{\sum X_i^2 - n\overline{X}^2}$$

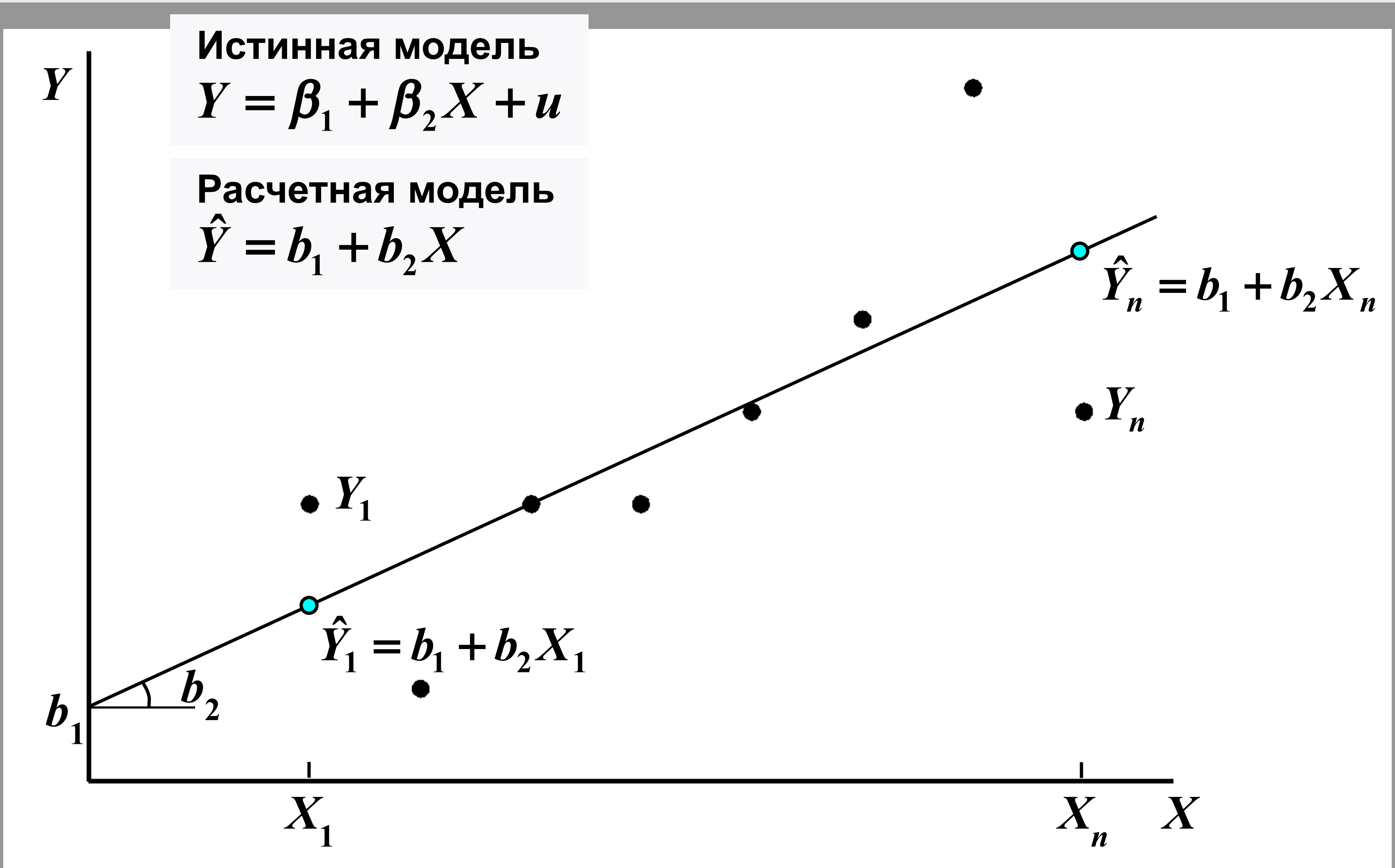
$$b_2 = \frac{\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum \left(X_i - \overline{X}\right)^2}$$

$$\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = \sum X_i Y_i - n\overline{X}\,\overline{Y}$$

$$\sum \left(X_i - \overline{X}\right)^2 = \sum X_i^2 - n\overline{X}^2$$

Знаменатели являются эквивалентными выражениями.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ



Подведем итог тому, что мы сделали.

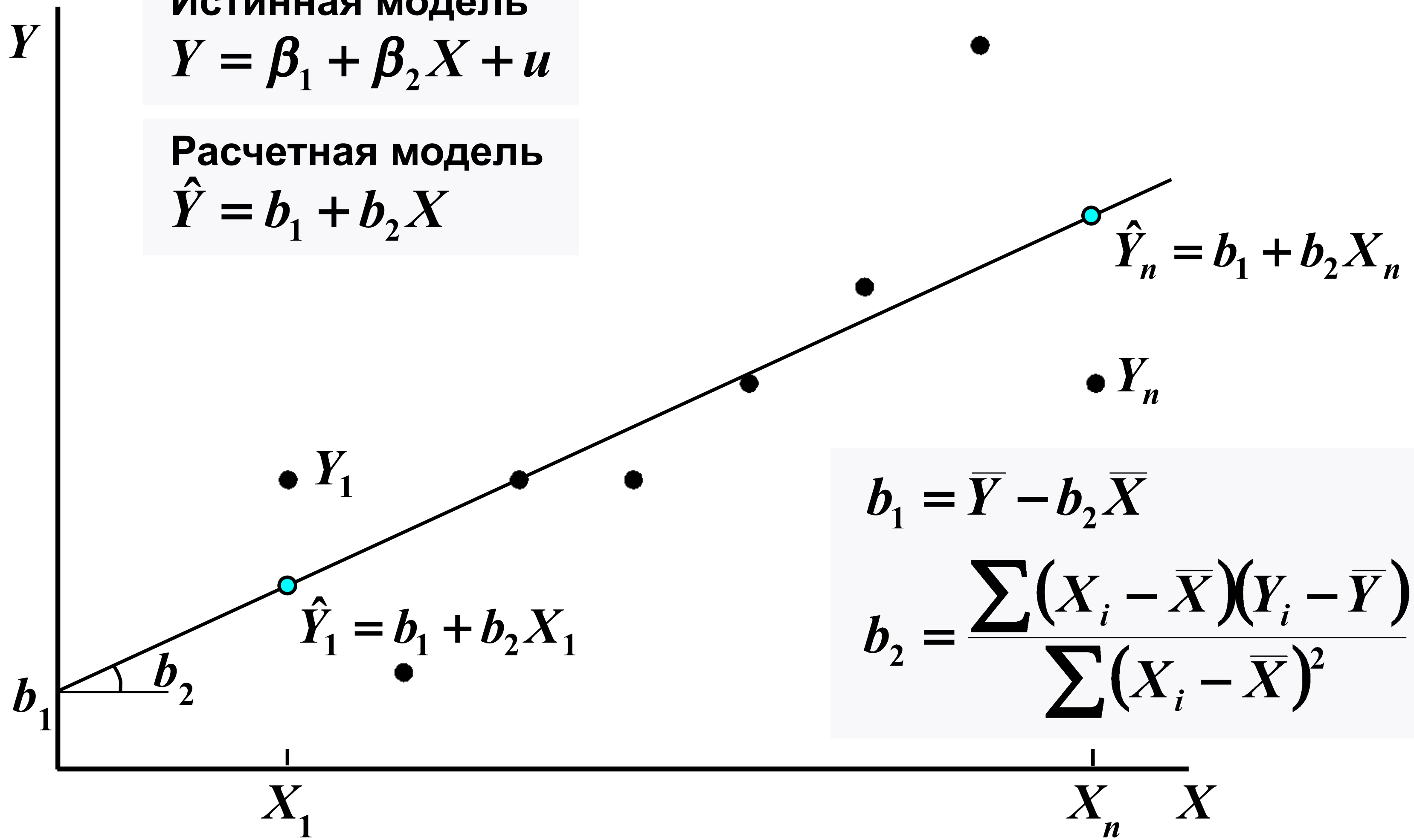
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_1 + b_2 X$$



$$b_1 = \overline{Y} - b_2 \overline{X}$$

$$b_2 = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

Мы нашли параметры расчетной линии, минимизируя сумму квадратов остатков. В результате мы получили выражения для b_1 и b_2.

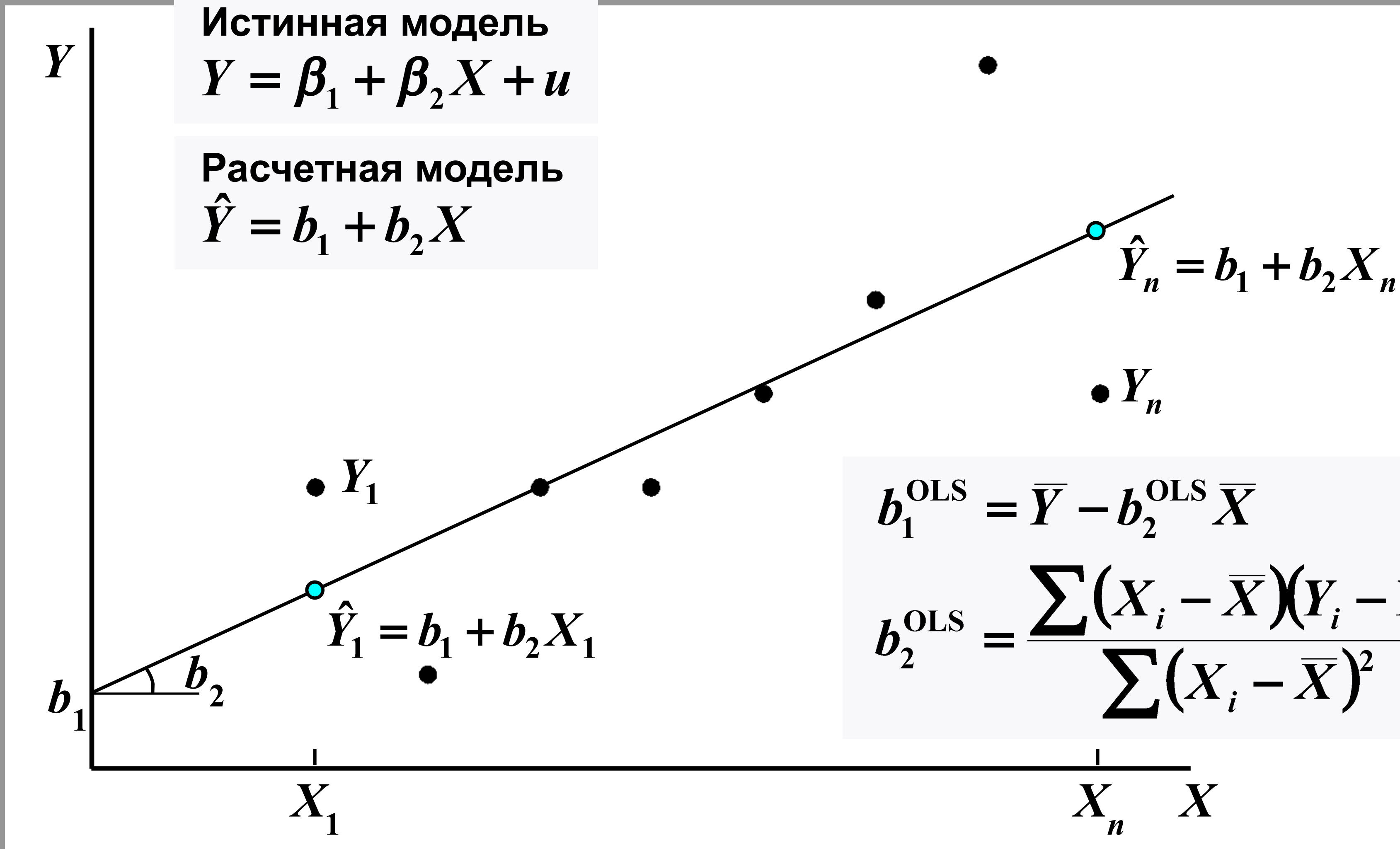
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_1 + b_2 X$$



$$b_1^{\mathrm{OLS}} = \overline{Y} - b_2^{\mathrm{OLS}} \overline{X}$$

$$b_2^{\mathrm{OLS}} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

Обозначаем найденные параметры расчетной модели, как b_1^{OLS} and b_2^{OLS}.

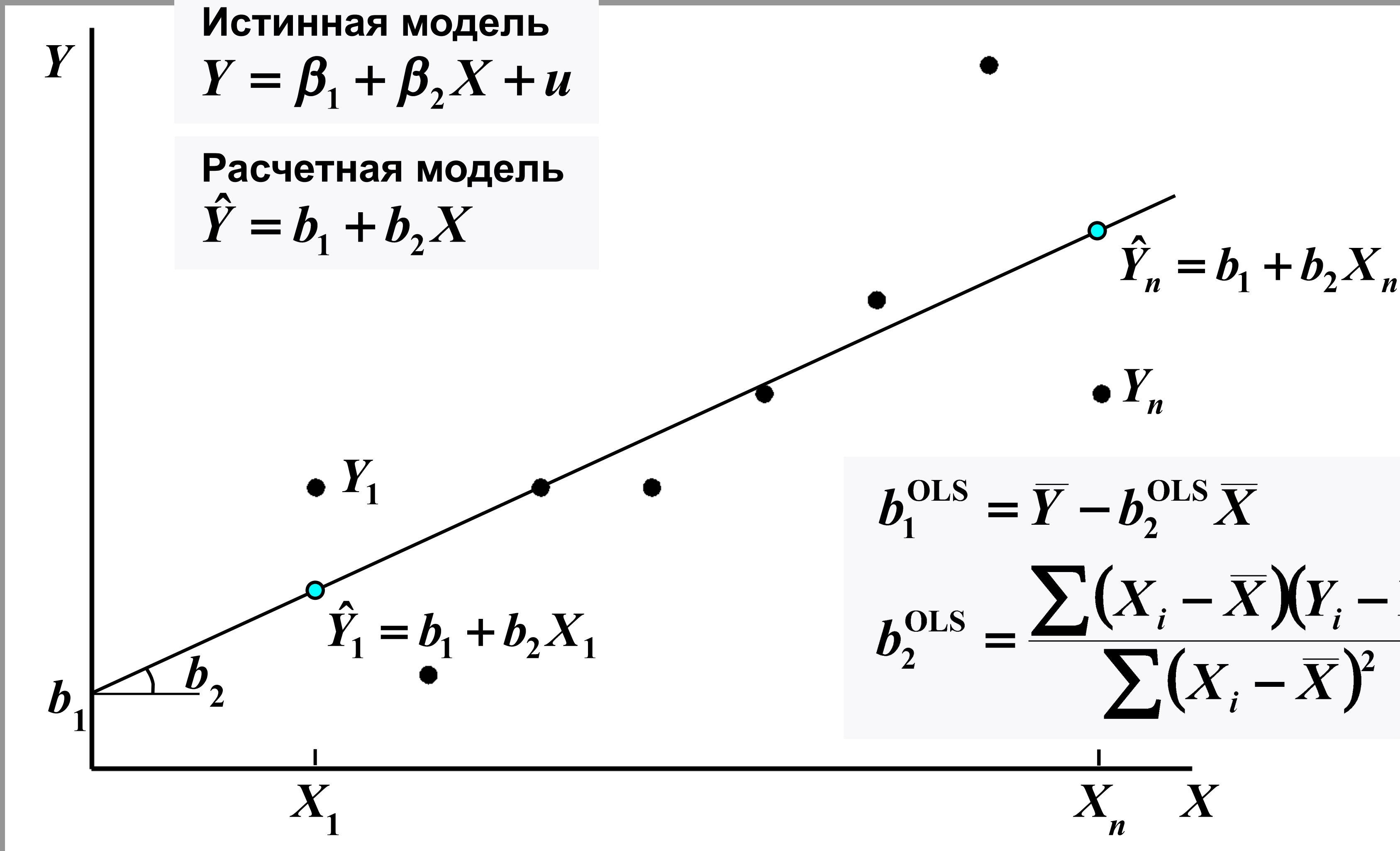
ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_1 + \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_1 + b_2 X$$



$$b_1^{\text{OLS}} = \overline{Y} - b_2^{\text{OLS}}\,\overline{X}$$

$$b_2^{\text{OLS}} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

В дальнейшем мы будем, как правило, пользоваться оценками OLS, и поэтому надстрочную надпись «OLS» мы использовать не будем.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель **Расчетная модель**

$$Y = \beta_2 X + u \qquad \hat{Y} = b_2 X$$

В случае простой регрессионной модели истинная и расчетная модели записываются без свободного члена.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_2 X$$

$$e_i = Y_i - \hat{Y}_i = Y_i - b_2 X_i$$

Мы выведем выражение для b_2, используя критерий наименьших квадратов. Остаток в наблюдении i равен $e_i = Y_i - b_2 X_i$.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель **Расчетная модель**

$$Y = \beta_2 X + u \qquad \hat{Y} = b_2 X$$

$$e_i = Y_i - \hat{Y}_i = Y_i - b_2 X_i$$

$$RSS = \sum (Y_i - b_2 X_i)^2 = \sum Y_i^2 - 2b_2 \sum X_i Y_i + b_2^2 \sum X_i^2$$

Здесь мы получаем выражение для суммы квадратов остатков.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель

$$Y = \beta_2 X + u$$

Расчетная модель

$$\hat{Y} = b_2 X$$

$$e_i = Y_i - \hat{Y}_i = Y_i - b_2 X_i$$

$$RSS = \sum (Y_i - b_2 X_i)^2 = \sum Y_i^2 - 2b_2 \sum X_i Y_i + b_2^2 \sum X_i^2$$

$$\frac{\mathrm{d}RSS}{\mathrm{d}b_2} = 2b_2 \sum X_i^2 - 2\sum X_i Y_i$$

$$2b_2^{\mathrm{OLS}} \sum X_i^2 - 2\sum X_i Y_i = 0$$

Продифференцируем по b_2.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель — **Расчетная модель**

$$Y = \beta_2 X + u \qquad \hat{Y} = b_2 X$$

$$e_i = Y_i - \hat{Y}_i = Y_i - b_2 X_i$$

$$RSS = \sum (Y_i - b_2 X_i)^2 = \sum Y_i^2 - 2b_2 \sum X_i Y_i + b_2^2 \sum X_i^2$$

$$\frac{\mathrm{d}RSS}{\mathrm{d}b_2} = 2b_2 \sum X_i^2 - 2\sum X_i Y_i$$

$$2b_2^{\mathrm{OLS}} \sum X_i^2 - 2\sum X_i Y_i = 0$$

$$b_2^{\mathrm{OLS}} = \frac{\sum X_i Y_i}{\sum X_i^2}$$

Следовательно, мы получаем оценку OLS для b_2 для этой модели.

ВЫЧИСЛЕНИЕ ЛИНЕЙНЫХ КОЭФФИЦИЕНТОВ РЕГРЕССИИ

Истинная модель $Y = \beta_2 X + u$

Расчетная модель $\hat{Y} = b_2 X$

$$e_i = Y_i - \hat{Y}_i = Y_i - b_2 X_i$$

$$RSS = \sum (Y_i - b_2 X_i)^2 = \sum Y_i^2 - 2b_2 \sum X_i Y_i + b_2^2 \sum X_i^2$$

$$\frac{\mathrm{d}RSS}{\mathrm{d}b_2} = 2b_2 \sum X_i^2 - 2\sum X_i Y_i$$

$$2b_2^{\mathrm{OLS}} \sum X_i^2 - 2\sum X_i Y_i = 0$$

$$b_2^{\mathrm{OLS}} = \frac{\sum X_i Y_i}{\sum X_i^2}$$

$$\frac{\mathrm{d}^2 RSS}{\mathrm{d}b_2^2} = 2\sum X_i^2 > 0$$

Вторая производная положительна, подтверждая, что мы нашли минимум.