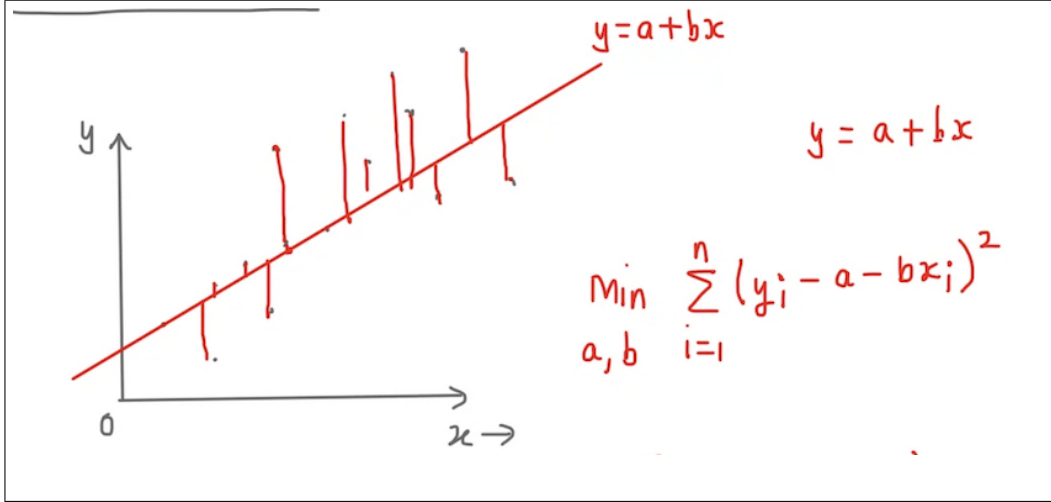


## Linear Regression



**Figure 1:** The Best fit line

$$\min_{a,b} \sum_{i=1}^n (y_i - a - bx_i)^2$$

Differentiating w.r.t  $a$  we get,

$$-2 \sum_{i=1}^n (y_i - a - bx_i) = 0$$

Differentiating w.r.t  $b$  we get,

$$-2 \sum_{i=1}^n (y_i - a - bx_i)x_i = 0$$

Rewriting the above equations;

$$\begin{aligned} na + b \sum x_i &= \sum y_i \\ a \sum x_i + b \sum x_i^2 &= \sum x_i y_i \end{aligned}$$

or,

$$\frac{\sum x_i}{n} \times [na + b \sum x_i = \sum y_i] \quad \dots (1)$$

$$a \sum x_i + b \sum x_i^2 = \sum x_i y_i \quad \dots (2)$$

From 1 we get;

$$a \sum x_i + \frac{b (\sum x_i)^2}{n} = \frac{\sum x_i \sum y_i}{n} \quad \dots (3)$$

Subtracting 3 from 2 we get;

$$b \left( \sum x_i^2 - \frac{(\sum x_i)^2}{n} \right) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$\Rightarrow \text{Slope of the best fit line, } b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

So,

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

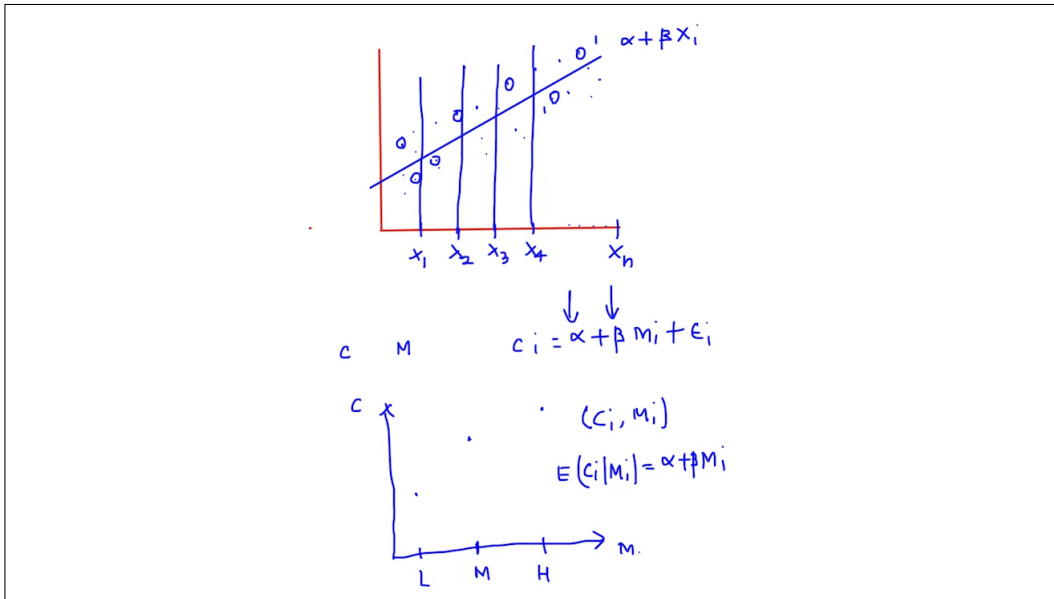
$$= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum (X_i - \bar{X})Y_i}{\sum (X_i - \bar{X})^2}, \quad \text{where } \bar{X} = \frac{\sum x_i}{n} \text{ and } \bar{Y} = \frac{\sum y_i}{n}$$

and since,  $a + b\bar{X} = \bar{Y}$  we get,

$$a = \bar{Y} - b\bar{X}$$

## Linear Regression Model



**Figure 2:** Linear Regression Model

There are some unknown parameters that we want to estimate just like in estimation.

$$E(Y_i|X_i) = \alpha + \beta X_i$$

Now we have the following assumptions about our model;

- $Y_i = \alpha + \beta X_i + \epsilon_i \quad \dots (A_1)$

where  $\epsilon_i$  is a random variable.

So we are trying to estimate  $\alpha$  and  $\beta$  for which we will draw some points from around the line (population)  $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$  and then using those points we will try to estimate the line.

- Now if we fix  $X_i$ 's and then for each  $X_i$  we draw a value of  $Y_i$  then  $E(\epsilon_i) = 0 \quad \forall i \quad \dots (A_2)$

and if we do not fix  $X_i$ 's then  $E(\epsilon_i|X_i) = 0 \quad \forall X_i$

- $E(\epsilon_i^2) = \sigma^2 \quad \forall i. \quad \dots (A_3)$

- $E(\epsilon_i \epsilon_j) = 0 \quad \forall i \neq j \quad \dots (A_4)$

or we can have a much stronger assumption rather than all the last three above, i.e.,

$$\epsilon_i \text{'s} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2). \quad \dots (A_5)$$

So now,

$$\hat{\beta}_{OLS} = \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

and

$$\hat{\alpha}_{OLS} = \bar{Y} - \hat{\beta}_{OLS} \bar{X}$$

Now under assumptions  $A_1, A_2, A_3, A_4$  and specifically  $A_1, A_2$ ;

$$\begin{aligned} E(\hat{\beta}_{OLS}) &= E\left(\frac{\sum (X_i - \bar{X})(\alpha + \beta X_i + \epsilon_i)}{\sum (X_i - \bar{X})^2}\right) \\ &= E\left(\frac{\alpha \sum (X_i - \bar{X})}{\sum (X_i - \bar{X})^2} + \frac{\beta \sum (X_i - \bar{X}) X_i}{\sum (X_i - \bar{X})^2} + \frac{\sum (X_i - \bar{X}) \epsilon_i}{\sum (X_i - \bar{X})^2}\right) \\ &= E\left(\beta + \frac{\sum (X_i - \bar{X}) \epsilon_i}{\sum (X_i - \bar{X})^2}\right) = \beta \quad \text{since } \frac{\sum (X_i - \bar{X}) X_i}{\sum (X_i - \bar{X})^2} = 1 \text{ and } \frac{\alpha \sum (X_i - \bar{X})}{\sum (X_i - \bar{X})^2} = 0 \end{aligned}$$

Now under assumptions  $A_1, A_2, A_3, A_4$ ;

$$\begin{aligned}
 \text{Var}(\hat{\beta}_{\text{OLS}}) &= E(\hat{\beta}_{\text{OLS}} - \beta)^2 \\
 &= E\left(\left(\frac{\sum(X_i - \bar{X})\epsilon_i}{\sum(X_i - \bar{X})^2}\right)^2\right) \\
 &= \frac{1}{\left(\sum(X_i - \bar{X})^2\right)^2} E\left(\left(\sum(X_i - \bar{X})\epsilon_i\right)^2\right) \\
 &= \frac{1}{\left(\sum(X_i - \bar{X})^2\right)^2} E\left(\sum_{i=1}^n (X_i - \bar{X})^2 \epsilon_i^2 + 2 \sum_{i < j} (X_i - \bar{X})(X_j - \bar{X}) \epsilon_i \epsilon_j\right) \\
 &= \frac{\sum(X_i - \bar{X})^2 \sigma^2}{\left(\sum(X_i - \bar{X})^2\right)^2} = \frac{\sigma^2}{\sum(X_i - \bar{X})^2}
 \end{aligned}$$

Now if we assume assumption  $A_5$  we can even find the distribution of  $\hat{\beta}_{\text{OLS}}$ ;  
The Distribution of

$$\hat{\beta}_{\text{OLS}} = \beta + \frac{\sum(X_i - \bar{X})\epsilon_i}{\sum(X_i - \bar{X})^2} \quad \text{is } \mathcal{N}\left(\beta, \frac{\sigma^2}{\sum(X_i - \bar{X})^2}\right)$$

## Multiple Linear Regression

Suppose we have  $n$  observations; and Assume that  $x$  is full rank;

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \\ \vdots & & \\ x_{m1} & \cdots & x_{nm} \end{bmatrix}_{n \times m} \begin{bmatrix} \beta_1 \\ \vdots \\ \vdots \\ \beta_m \end{bmatrix}_1 + \epsilon_i$$

$$\begin{aligned} & (y - x\beta)_{1 \times n}^T (y - x\beta)_{n \times 1} \\ &= (y^T - \beta^T x^T)(y - x\beta) \\ &= y^T y - y^T x\beta - \beta^T x^T y + \beta^T x^T x\beta \\ &= y^T y - 2\beta^T x^T y + \beta^T x^T x\beta \\ & \text{differentiating w.r.t } \beta^T \text{ we get,} \\ &= -2x^T y + 2x^T x\beta = 0 \\ &= \hat{\beta}_{\text{OLS}} = (x^T x)^{-1} (x^T y) \\ &E(\hat{\beta}_{\text{OLS}}) = E(\beta + (x^T x)^{-1} x^T \epsilon) = \beta \end{aligned}$$

and

$$\begin{aligned} \text{Var}(\hat{\beta}_{\text{OLS}}) &= E \left( \hat{\beta}_{\text{OLS}} - \beta \right)^2 \\ &= E \left( (x^T x)^{-1} x^T \epsilon \epsilon^T x (x^T x)^{-1} \right) \\ &= \sigma^2 E \left( (x^T x)^{-1} x^T x (x^T x)^{-1} \right) \quad \text{since } E(\epsilon_{n \times 1} \epsilon_{1 \times n}^T) = \sigma^2 I_{n \times n} \\ &= \sigma^2 (x^T x)^{-1} \end{aligned}$$