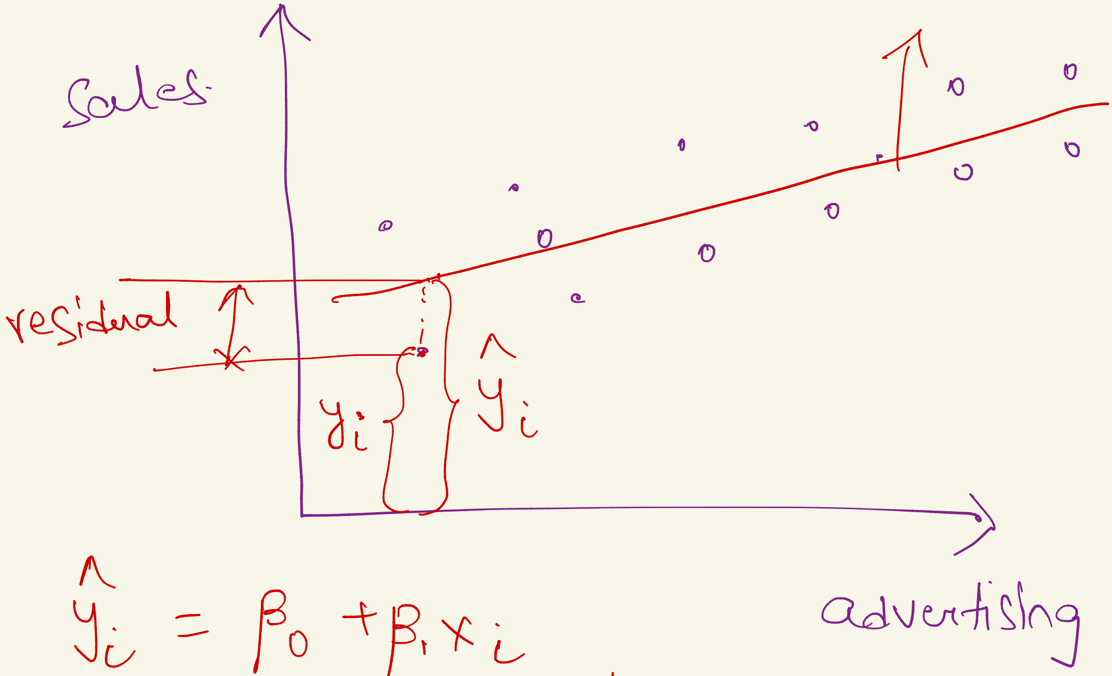


$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

error $e_i = y_i - \hat{y}_i$

$$= y_i - (\beta_0 + \beta_1 x_i)$$

Find β_0 and β_1 such that the square of error is minimum

$$\min \sum e_i^2 = \min \sum \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2$$

$$\frac{\partial}{\partial \beta_0} = 0 \qquad \frac{\partial}{\partial \beta_1} = 0$$

$\hat{\beta}_1$ (or) $b_1 = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$ - slope

$\bar{x}$ = Mean of x

$\bar{y}$ = Mean of y

$\hat{\beta}_0 = b_0 = \bar{y} - b_1 \bar{x} \rightarrow$ intercept.

(OR)

$\bar{y} - \hat{\beta}_1 \bar{x}$

$\hat{\beta}_1 = \dfrac{\Delta \hat{y}}{\Delta x}$ = change in y for one unit change in x

Example

Advertising and Sales.

X (advertising)	Y (Sales)
5	50
2	15
3	25
1	10
4	40

$$\bar{X} = \frac{5+2+3+1+4}{5} = \frac{15}{5} = 3$$

$$\bar{Y} = \frac{50+15+25+10+40}{5} = \cancel{28}$$

x	y	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$
5	50	$5-3=2$	$50-28=22$	$2\times22=44$	4
2	15	$2-3=-1$	$15-28=-13$	$-1\times-13=13$	1
3	25	$3-3=0$	$25-28=-3$	$0\times-3=0$	0
1	10	$1-3=-2$	$10-28=-18$	$-2\times-18=36$	4
4	40	$4-3=1$	$40-28=12$	$1\times12=12$	1
		0	0	105	10

$\bar{x}=3$

$\bar{y}=28$

$$\hat{\beta}_1 = b_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{105}{10}$$

$$\hat{\beta}_1 = 10.5 = \text{Slope.}$$

$$b_0 = \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$= 28 - 10.5(3)$$

$$= -3.5$$

If you do not advertise you expect to have -3.5 in sales.

Best fit linear regression line is

$$\hat{y} = -3.5 + 10.5(x)$$

OLS estimation

↳ Ordinary Least Squares.

Algebraic Properties of OLS estimates.

1. Sum of residuals = 0

Residuals = observed − fitted.

$$= \sum_{i=1}^{n} y - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

$$= 0$$

2. Covariance between x and residuals is zero

$$\sum_{i=1}^{n} x_i e_i = 0$$

3. $(\bar{x}, \bar{y})$ always fall on the estimated regression line

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

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

$$y = -3.5 + 10.5(x)$$

$$\bar{x} = 3$$

$$y = -3.5 + 10.5(3)$$

$$= -3.5 + 31.5$$

$$= \underline{28}$$

$$= \bar{y}$$

4. Covariance between fitted values and residuals is zero.

$$e_i = y_i - \hat{y}_i$$

Sum of Squares total

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

total Variation from the mean to be explained. This is a measure of total sample variation.

Sum of Squares Residual

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

total variation explained by X

Sum of Squared Error

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y})^2$$

total variation left unexplained

$$SST = SSR + SSE$$

Goodness-of-fit

Measures how well x explain y

R^2 = Fraction of sample variation in y explained by x

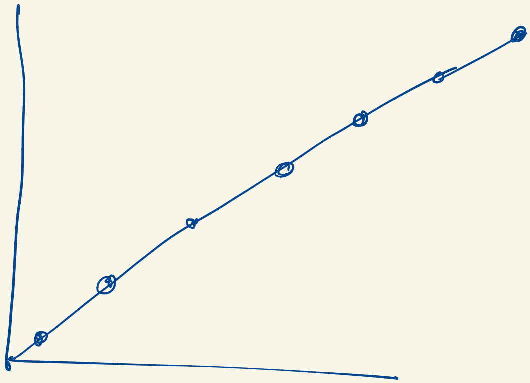
(OR)

Ratio of variation explained to the total variation.

$$R^2 = \frac{SSE}{SST}$$

$$= 1 - \frac{SSR}{SST}$$

R^2 is also called Coefficient of determination



$R^2 = 1$

$R^2 = 0$ poor fit