Simple Linear Regression (Univariate)

SLR is a model with single regressor X that has a linear relationship with a response Y. The SLR is,

$$Y = \beta_0 + \beta_1 x + \epsilon .$$

$Y \rightarrow$ is a response (a R.V)

$x \rightarrow$ is a regressor (not a R.V)

$\epsilon \rightarrow$ random error.

$\beta_0 \rightarrow$ intercept

$\beta_1 \rightarrow$ slope

Basic assumptions,

$$Y_i = \beta_0 + \beta_1 x + \epsilon_i , \quad i = 1 : n$$

1) ϵ_i is a RV with zero mean & variance σ^2 (unknown)

i.e $E[\epsilon_i] = 0$ & $V[\epsilon_i] = \sigma^2$

2) ϵ_i & ϵ_j are uncorrelated, $i \neq j$ So, $cov(\epsilon_i, \epsilon_j) = 0$

3) ϵ_i is a normally distributed R.V with mean zero & variance σ^2, $\epsilon \sim N(0, \sigma^2)$

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$E[Y_i] = E[\beta_0 + \beta_1 x_i + \epsilon_i] = \beta_0 + \beta_1 x_i$$

$$V[Y_i] = Var[\beta_0 + \beta_1 x_i + \epsilon_i] = V[\epsilon_i] = \sigma^2$$

$\epsilon_i \sim N(0, \sigma^2)$ and independent

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

Least Squares Estimation of the parameters.

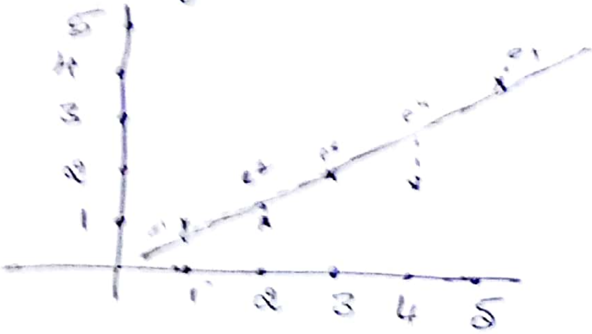
The parameters β_0, β_1 are unknown & must be estimated (fitting a linear model)

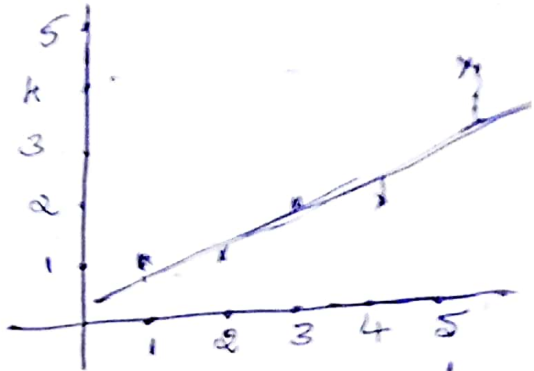
sample data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

Example

Advertising (x)	Sales (y)
1	1
2	1
3	2
4	2
5	4

Scatterplot





Error is the difference between the actual value of "y" for a given x & the predicted value of y using the estimated linear model. $(e_i) \Rightarrow y_i - \hat{y}_i$

The line fitted by least square is the one that makes the sum of square of all vertical discrepancy as small as possible

SS_{Res} (Sum of square residual)

$= \sum_{i=1}^{n} e_i^2$ is minimum

So $e_i = Y_i - \hat{Y}_i$

$Y_i \Rightarrow$ is the actual response

$\hat{Y}_i \rightarrow$ predicted response.

Hence

We need to estimate β_0 & β_1 such that the SS_{res} is as minimum as possible

$$SS_{res} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \sum_{1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_0 x_i)^2 \text{ is min}$$

To find out the E s.

the least square estimators β_0 & β_1 are referred as, $\hat{\beta}_0$ & $\hat{\beta}_1$

~~is note~~

We make the differentiation of the as 0 w.r.t β_0 & β_1 both.

$\sum x^2$

$$\frac{\partial S}{\partial \beta_0}\Big|_{\hat{\beta}_0, \hat{\beta}_1} = 0$$

$$-2\sum_{1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \text{—— (1)}$$

$$\sum_{1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial S}{\partial \beta_1}\Big|_{\hat{\beta}_0, \hat{\beta}_1} = -2\sum x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \text{—— (2)}$$

Eqn (1). $\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.$

$$\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0.$$

$$n\hat{\beta}_0 = \sum_{i=1}^{n} y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i$$

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{\hat{\beta}_1}{n}\sum_{i=1}^{n} x_i = \bar{y} - \hat{\beta}_1 \bar{x}$$

Eqn (2)

$$\sum x_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

Substitute, $\hat{\beta}_0 = \bar{Y}_i - \hat{\beta}_1 \bar{x}$

$$\sum x_i (y_i - \bar{Y}_i + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) = 0$$

$$\sum x_i (y_i - \bar{Y}_i) + \sum (\hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) x_i = 0.$$

$$\beta_1 = \frac{\sum x_i (y_i - \bar{Y}_i)}{\sum (x_i - \bar{x}) x_i}$$

By proving that, $\sum (y_i - \bar{y}) \bar{x} = 0$ & $\sum (x_i - \bar{x}) \bar{x} = 0$, add these terms to numerator & denominator, we get

$$\beta_1 = \frac{\sum (y_i - \hat{Y}_i)(x_i - \bar{x})}{\sum (x_i - \bar{x})(x_i - \bar{x})}$$

Multivariate Regression

In Multivariate regression, the numeric output r is assumed to be written as a linear function, that is a weighted sum of several input variables, $x_1, \dots, x_p$ & noise.

(Actually in statistical literature, this is called multiple regression & statisticians use the term multivariate when there are multiple outputs).

TV adver	Newspaper	Radio	Sales
1	1	2	1
2	1	3	2
4	1	5	4
6	1	2	5

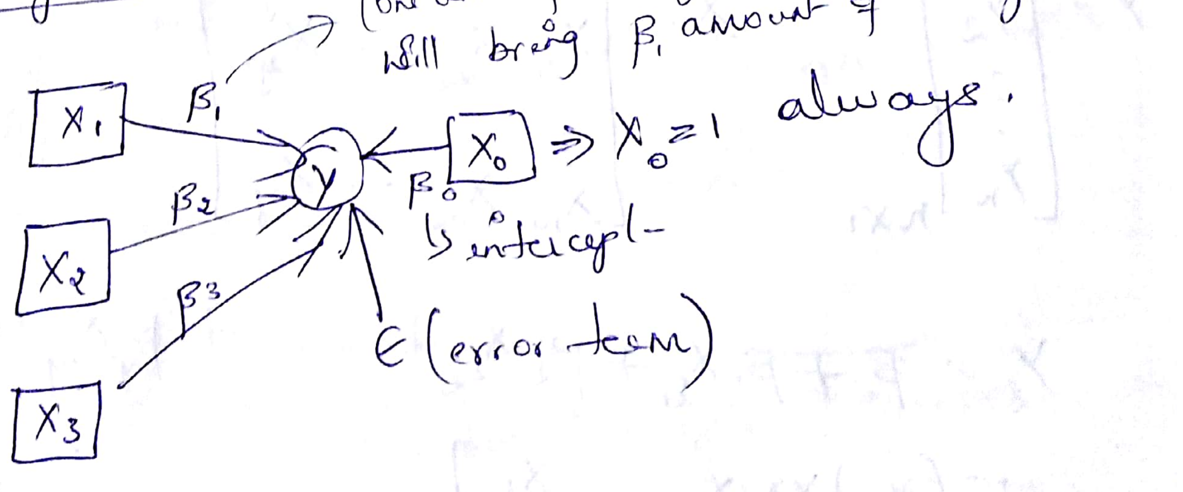
More than one regressor variable, say 'p'

Sales ⇒ Y $\quad X_1 \rightarrow$ TV advertising

$X_2 \rightarrow$ Newspaper.

$X_3 \rightarrow$ Radio

Conceptual Model.



$$Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$ is the equation for the above pictorial model.

Since $X_0 = 1$, we can write it

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

If we ~~generalize~~ generalize this ~~to~~ where we have 'p' variable $(X_1, X_2, \ldots, X_p)$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}_{p \times 1}$$

$\Rightarrow$ adding intercept parameter X_0 it becomes

$$\Downarrow$$

$$X = \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_p \end{bmatrix}_{(p+1) \times 1}$$ $\rightarrow$ one data point.

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1) \times 1}$$

So, $Y = \beta_0 X_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon$

'n' data points (sample dataset) is represented as,

$$\underline{DV} \qquad\qquad \underline{\underline{IVs}}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1} \qquad X = \begin{bmatrix} X_{11} & X_{12} & \ldots & X_{1p} \\ X_{21} & X_{22} & \ldots & X_{2p} \\ \vdots & & & \\ X_{n1} & X_{n2} & \ldots & X_{np} \end{bmatrix}$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \epsilon_i$$

$$E(Y_i \mid X_{i1}, X_{i2}, \ldots X_{ip}]$$

$$Y_i = E(Y_i \mid X_{i1}, X_{i2}, \ldots, X_{ip}) + \epsilon$$

$$Y_i = \hat{Y}_i + \epsilon$$

$$\epsilon = Y_i - \hat{Y}_i$$

Summarized,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i$$

For 'n' data points,

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_p X_{1p} + \epsilon_1$$

$$Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_p X_{2p} + \epsilon_2$$

$$\vdots$$

$$Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_p X_{np} + \epsilon_n$$

$$\underset{n \times 1}{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}} = \underset{n \times (p+1)}{\begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & & X_{np} \end{bmatrix}} \underset{(p+1) \times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}} + \underset{n \times 1}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

$$n \times (p+1) \times (p+1) \times 1$$

So we can write the above equation in vector form as,

$$\boxed{\underset{n \times 1}{Y} = \underset{n \times (p+1)}{X} \underset{(p+1) \times 1}{\beta} + \underset{(n \times 1)}{\epsilon}}$$

$X \Rightarrow$ is also called as design matrix

$\beta \Rightarrow$ Regression coefficients.

↳ Eq^n for Multiple Regression in matrix form.

Assumptions ⇒ (1) Linearity.

(2) Equal Y variance across values of X

(3) ϵ is uncorrelated terms

(4) Normality of the error terms

$$\epsilon_i \sim N(0, \sigma_i^2)$$

$$\sigma_i^2 = \sigma$$

$$\epsilon_i, \epsilon_j \Rightarrow \underline{cov(\epsilon_i, \epsilon_j) = 0}.$$

Estimation of Model parameters (β).

$$Y = X\beta + \epsilon$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \epsilon_i$$

$$\epsilon_i = Y_i - \sum_{j=0}^{p} \beta_j X_{ij}$$

$$\epsilon_i^2 = \left[Y_i - \sum_{j=0}^{p} \beta_j X_{ij}\right]^2 \quad \text{for (} i^{th} \text{ observation)}$$

For 'n' observations, we need to take sum over $i=1$ to n. (SSE)

$$SSE = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left[Y_i - \sum_{j=0}^{p} \beta_j X_{ij}\right]^2$$

Solve for β & chose β in such a way that $SSE=0$ (objective function).

$$\frac{\partial SSE}{\partial \beta_j} = 0 \quad \text{~~subject to~~} \quad \cancel{\frac{\partial^2 SSE}{\partial \beta_j \partial p}}$$

...rivation

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n \times 1} \qquad SSE = \underset{(1\times n)}{\epsilon^T} \underset{(n\times 1)}{\epsilon}$$

$$= (y - X\beta)^T (y - X\beta) \text{ is minimum}$$

$$(y - X\beta)^2$$

Derivation

$$\frac{\partial SSE}{\partial \beta} = -2X^T(y - X\beta) = 0 \qquad 2(-1)X^T(y - X\beta)$$

$$\Rightarrow -X^Ty + X^TX\beta = 0$$

$$X^TX\beta = X^Ty$$

Multiply with inverse of (X^TX) i.e $(X^TX)^{-1}$

$$(X^TX^{-1})(X^TX)\beta = (X^TX)^{-1}X^Ty \qquad A \times A^{-1} = 1$$

$$= 1 \qquad \beta = (X^TX)^{-1}X^Ty$$

Problem Ex

$$y = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 5 \\ 1 & 7 \\ 1 & 10 \\ 1 & 12 \\ 1 & 20 \end{bmatrix}$$

$$\hat{\beta} = (X^TX)^{-1}X^Ty$$

$$X^TX = \underset{(2\times 5)}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix}} \underset{(5\times 2)}{\begin{bmatrix} 1 & 5 \\ 1 & 7 \\ 1 & 10 \\ 1 & 12 \\ 1 & 20 \end{bmatrix}} = \begin{bmatrix} 5 & 54 \\ 54 & 718 \end{bmatrix}$$

↳ step (1)

$$(X^TX)^{-1} =$$

Step 2 $\Rightarrow (X^TX)^{-1} = \frac{1}{|X^TX|} adj(X^TX)$

$$|X^TX| = \begin{vmatrix} 5 & 54 \\ 54 & 718 \end{vmatrix} = 5 \times 718 - 54^2 = 3590 - 2916 = 674$$

$$adj(X^TX) = \begin{bmatrix} 718 & -54 \\ -54 & 5 \end{bmatrix}$$

$$(X^TX)^{-1} = \frac{1}{674} \begin{bmatrix} 718 & -54 \\ -54 & 5 \end{bmatrix} = \begin{bmatrix} 1.07 & -0.08 \\ -0.08 & 0.007 \end{bmatrix}$$

Next, (Step 3)

$$\cancel{(X^TX)} \quad X^Ty = \underset{(2\times5)}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 7 & 10 & 12 & 20 \end{bmatrix}} \underset{(5\times1)}{\begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}}$$

$$= \begin{bmatrix} 150 \\ 1970 \end{bmatrix}$$

(Step 4)

$$\hat{\beta} = (X^TX)^{-1} X^Ty$$

$$= \begin{bmatrix} 1.07 & -0.08 \\ -0.08 & 0.007 \end{bmatrix} \begin{bmatrix} 150 \\ 1970 \end{bmatrix} = \begin{bmatrix} 2.70 \\ 1.775 \end{bmatrix}$$

Regression Eq^n

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

$$y = 2.70 + 1.775 X_1 + \epsilon$$

$$\hat{y} = 2.70 + 1.775 X_1$$

$$\epsilon = y - \hat{y}$$

$$= \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix} - \begin{bmatrix} 2.70 + 1.775 \times 5 \\ 2.70 + 1.775 \times 7 \\ 2.70 + 1.775 \\ \vdots \\ \vdots \end{bmatrix}$$

Problem 2

$$\boxed{Y = 1 + 2X}$$

$$Y = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{\det |X^T X|} \operatorname{adj}(X^T X)$$

$$= \frac{1}{150 - 100} \begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix} = \frac{1}{50} \begin{bmatrix} 30 & -10 \\ -10 & 5 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix}$$

$$\boxed{\beta = (X^T X)^{-1} X^T Y}$$

$$\cancel{X^T Y = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 4 & 3 & 8 & 9 \end{bmatrix} =}$$

1×5

5×2

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 25 \\ 70 \end{bmatrix}$$

$(2 \times 5) \qquad (5 \times 1)$

$$(X^T X)^{-1} X^T Y = \begin{bmatrix} \cancel{0.6} & \cancel{-0.2} \\ & \\ \cancel{10} & \end{bmatrix} \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix} \begin{bmatrix} 25 \\ 70 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\hat{Y} = X\hat{\beta} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix}$$

$$\hat{\epsilon} = Y - \hat{Y} = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} - \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

Sum of square $\hat{\epsilon}^2$

$$\hat{\epsilon}'\epsilon = \begin{bmatrix} 0 & 1 & -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} = \underline{\underline{6}}$$