R^2 = SSR/SST = The proportion of the variation in y being explained by the variation in x. R square range from 0 to 1.

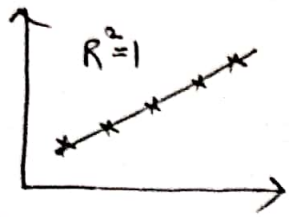
SSR + SSE = SST

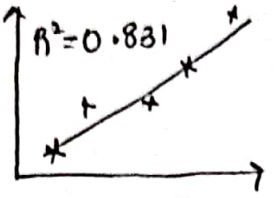
Sum of square due to regression | Sum of Square due to error.

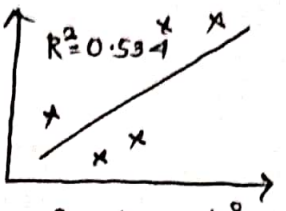
Also known as,

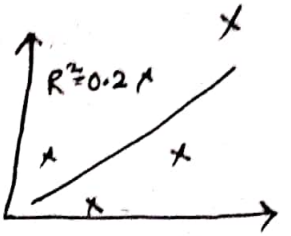
ESS + RSS = SST/TSS

Explained Sum of square | Residual Sum of square | Total Sum of Square.







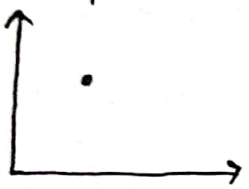


Q) What is the minimum number of observations require to estimate the regression?

$Y_i = B_0 + B_1 X_i + \varepsilon_i$.

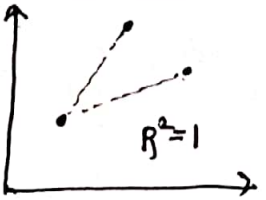
(Y_i = Height, X_i = weight)

If one point is there



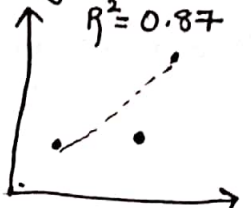
we cannot make a line, so we need two point

2 observation



So wherever second point is, line will join. But we cannot check strength of / test of relationship.

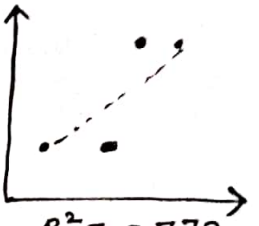
3rd observation (In order to check strength of line)



So, 1 point will check strength of line.

So, df = 1 (degree of freedom)

4 observations



$R^2 = 0.772$

df = 2 //.

df = 4 − 1 − 1

df = 2

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$. How many minimum number of observation require to construct regression line?

(X_{1i} = height, X_{2i} = Mothers height)

df = 5 − 2 − 1 = 5 − 3 = 2.

- We need minimum 3 points to construct a plane.

N = 3, R^2 = 1. N = 4, R^2 = 0.80, df = 1. N = 5, R^2 = 0.73, df = 2.

- So, degree of freedom, $\boxed{df = n - k - 1}$

n = number of observation

k = number of explanatory (x) variables (Independent)

Q) How does degree of freedom related to R square?

→ As degree of freedom (df) decreases (i.e. more variable added to given model) R square will only increase. So, even if we add useless variable, R square will only increase. So we use Adjusted R^2.

Adjusted R^2

Adjusted $R^2 = (1) - (1 - R^2)\frac{n-1}{n-k-1}$ or $1 - \frac{SSE}{SST}\left(\frac{n-1}{n-k-1}\right)$

as k increases, Adjusted R^2 will tend to decrease, reflecting the reduced power in the model. Only if we add useful variable to the model, Adjusted R^2 will only increase.

Number of observation (n)	number of variable (k)	R^2	Adj-R^2	
25	4	0.71	0.65	
25	5	0.76	0.69	
25	6	0.78	0.71	→ Choose this one.
25	7	0.79	0.70	