A20548395

Homework - 4

4A Given Simple Linear Regression:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots n.$$

To show $R^2 = r^2$, r is correlation between x, y.

$$\Rightarrow r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Lets consider a estimate in linear regression.

$$\left(\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i\right) \text{ — (1)} \quad i = 1, \ldots, n..$$

We know $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$ & $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$R^2 = 1 - \frac{RSS}{TSS}$$

We know, $TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$

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

Also, $R^2 = \dfrac{SS_{Reg}}{TSS}$ where $SS_{Reg} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

Computing R^2:-

$$R^2 = \frac{SS_{Reg}}{TSS} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad \text{substitute (1)}$$

$$= \frac{\sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n} (\cancel{\bar{y}} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i - \cancel{\bar{y}})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$\Rightarrow R^2 = \hat{\beta}_1^2 \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

W.K.T. $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

$$\Rightarrow R^2 = \hat{\beta}_1^2 \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$\Rightarrow R^2 = \left(\frac{\sum_{i=1}^{n} (x_i - x)(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^{\cancel{2}}}\right)^2 \times \frac{\cancel{\sum_{i=1}^{n} (x_i - \bar{x})^2}}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$= \frac{\left(\sum_{i=1}^{n} (x_i - x)(y_i - \bar{y})\right)^2}{\sum_{i=1}^{n} (x_i - x)^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}$$

we have $r = \frac{\sum_{i=1}^{n} (x_i - x)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$

So we proved that $\boxed{R^2 = r^2}$

5A To show, R^2 increases or remains same when moving from simple linear regression model with one variable to multiple linear regression.

We know $R^2 = 1 - \frac{RSS}{TSS}$

Where RSS ~~explains~~ represents the unexplained variance or error.

While TSS represents total variance in target variable (y)

In general

Simple linear Regress:	and Multiple linear regress:
$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$	$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon$
• Here only one independent variable (x_1)	• Here there are two independent variables (x_1, x_2)

Here in Multiple linear regression, by adding second predictor, the model has more flexibility to explain variation in y.

We know $R^2 = 1 - \frac{RSS}{TSS}$ part a

So ~~Thus~~ If RSS decreases, part a ~~increase~~ decreases then R^2 will increase.

So $\boxed{R^2 \propto \frac{1}{RSS}}$

Lets check what happens in RSS when new predictor is added.

→ if x_2 is added, useful information, it will reduce RSS, by explaining more of the time variability in y.

→ But If x_2 adds no useful information, the regression model will effectively ignore it (by diminishing its coefficient) and RSS remains unchanged.

→ Since RSS never increases, and only remain same or decreases we can conclude R^2 either increases or remains same