Q13

Rahul Atre

2023-10-11

Q13 [Regression Modeling]

It is claimed that in the case of simple linear regression of Y onto X, the R^2 statistic is equal to the square of the correlation between X and Y. Prove that this relationship holds. For simplicity, assume that $\bar{x} = \bar{y} = 0$.

First, let us define the equations that we know, and see if we can simplify them using the above information. From regression, we know that the beta estimate parameters are as follows:

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0$
- $\hat{\beta}_1 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{(x_i - \bar{x})^2} = \sum_{i=1}^{n} \frac{x_i y_i}{x_i^2}$

For R^2:

- Also, $SST = SSE + SSR \rightarrow SSR = SST - SSE$
- So, $R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$
- $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2$

(For the next few equations we will drop the formal sum notation for ease of understanding)

Also, let's try simplifying the correlation of X and Y:

$Cor(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$

Now, we can expand and simplify SSE:

$SSE = \sum (y - \hat{y})^2$

$= \sum (y^2 - 2y\hat{y} + \hat{y}^2)$

$= \sum (y^2 - 2y(\hat{\beta}_0 + \hat{\beta}_1 x) + (\hat{\beta}_0 + \hat{\beta}_1 x)^2)$

$= \sum (y^2 - 2y(\hat{\beta}_1 x) + (\hat{\beta}_1 x)^2)$

$= \sum (y^2 - 2y(\hat{\beta}_1 x) + \hat{\beta}_1^2 x^2)$

$= \sum (y^2 - 2y(\hat{\beta}_1 x) + \hat{\beta}_1^2 x^2)$

$= \sum y^2 - 2\hat{\beta}_1 \sum xy + \hat{\beta}_1^2 \sum x^2$

$= \sum y^2 - 2(\sum \frac{xy}{x^2}) \sum xy + (\sum \frac{xy}{x^2})^2 \sum x^2$

$= \sum y^2 - 2\frac{(\sum xy)^2}{\sum x^2} + \frac{(\sum xy)^2}{\sum x^2}$

$= \frac{\sum y^2 \sum x^2 - \sum (xy)^2}{\sum x^2}$

Thus, $R^2 = \frac{SST - SSE}{SST}$

$= \frac{\sum y^2 - (\frac{\sum y^2 \sum x^2 - \sum (xy)^2}{\sum x^2})}{\sum y^2}$

$= \frac{\sum y^2 \sum x^2 - \sum y^2 \sum x^2 + (\sum xy)^2}{\sum y^2 \sum x^2}$

$= \frac{(\sum xy)^2}{\sum y^2 \sum x^2}$

$= (\frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}})^2$

$= Cor(x, y)^2$

Hence, we have proved that $R^2 = Cor(x, y)^2$.