Week 6 Wednesday

Valentino Aceves

May 8, 2024

Measures of Fit for MLR

- SSR, ESS, TSS

$$SSR = \sum_{i=1}^{n}(y_i - \hat{y_i})^2 = \sum_{i=1}^{n}(y_i - (\hat{b_0} + \hat{b_1}x_{i1} + ... + \hat{b_p}x_{ip}))^2$$
$$ESS = \sum_{i=1}^{n}(\hat{y_i} - \bar{y})^2 = \sum_{i=1}^{n}(\hat{b_0} + \hat{b_1}x_{i1} + ... + \hat{b_p}x_{ip} - \bar{y})^2$$
$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

- R^2

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} \in [0, 1]$$

Notice that there is no longer a relationship between the correlation r_{XY} and R^2. This is only true for simple linear regression. Also notice, R^2 is **strictly increasing** as we introduce more variables (as $p \uparrow$), no matter how irrelevant they are. This is because as we introduce more variables, SSR decreases. We will introduce an R^2 that punishes too many variables in the model.

- $\bar{R}^2$ (Adjusted R^2)

$$\bar{R}^2 = 1 - \frac{n-1}{n-p-1}\frac{SSR}{TSS} < R^2$$

Now as $p \uparrow \implies SSR \downarrow \frac{n-1}{n-p-1} \uparrow$. This creates a counter effect to the measure.

Side note: $n - p - 1 =$ degrees of freedom of the model.

- Residual Standard Error (RSE)

$$RSE = \sqrt{\frac{SSR}{n-p-1}} = \hat{\sigma}$$

$E(\hat{\sigma}) = \sigma$ however $E(\hat{\sigma}^2) \neq \sigma^2$ where σ^2 is the true variance of the errors.