W6 Formulae

Wednesday, 16 November 2022 20:10

- SLR has one explanatory variable.
- MLR has multiple explanatory variables.

$k \rightarrow$ number of explanatory variables in MLR

Multiple Linear Regression Model

- Observed values of Y are linearly related to the k explanatory variables as:

$$\boldsymbol{y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \epsilon},$$

where, $\boldsymbol{\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)}$

- We makes 3 assumptions about the error term:

1. Independent
2. Equal variance, $Var(\epsilon) = \sigma_\epsilon^2$
3. Normal. $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$

- If you add explanatory variables to the model:
  - $\bar{R}^2 \uparrow$ and $s_e \downarrow$

- R
  - In SLR, R represents the correlation between X and Y.
  - In MLR, R represents the correlation between observed value(y) predicted value($\hat{y}$).

- Calibration plot: Scatterplot between y and $\hat{y}$.

Adjusted R Squared Formula

$$\boldsymbol{\bar{R}^2 = 1 - \left[\frac{(1 - R^2) \times (n - 1)}{(n - k - 1)}\right]}$$

Slopes

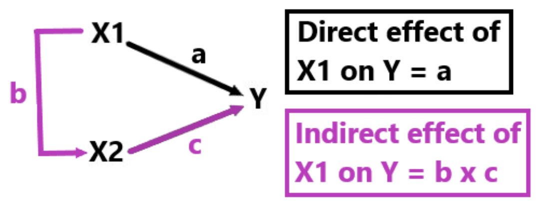
- In SLR: $\beta_1 \rightarrow$ Marginal Slope.
  - Change in Y variable with one unit change in X variable.

- In MLR: $\beta_1 \rightarrow$ Partial Slope.
  - Change in Y variable with one unit change in X variable keeping all the other X variables constant.

- In MLR, if variables are independent of each other, *i.e.*, correlation $= 0$, then
  Marginal slope = Partial slope

Path Diagrams

Schematic drawing of the relationships among various X's and Y.



Total effect = Direct effect + Indirect effect

$Y = \beta_0 + aX_1 + cX_2$
$X_2 = \tilde{\beta}_0 + bX_1$

Total effect of $X1$ on $Y = a + b \times c$

Total effect of X_i on Y is represented in the Marginal Slope.

VIF (Variance Inflation Factor)

quantifies the amount of unique variation in each explanatory variable and measures the effect of collinearity.

$$\boldsymbol{VIF(X_j) = \frac{1}{1 - R_j^2}}$$

- $R_j^2 \uparrow \quad \Rightarrow \; VIF \uparrow$
- R_j^2 is the coefficient of determination on a regression where that particular j^{th} variable is the response variable and all the other explanatory variables are the explanatory variables.
- If explanatory variables are uncorrelated: $R_j^2 = 0$ and $VIF = 1$
- If they are correlated: $VIF > 1$
- Larger the VIF larger the collinearity.

Typically,

$$se(b_1) = \frac{s_e}{\sqrt{n}} \times \frac{1}{s_{X_1}}$$

With VIF,

$$\boldsymbol{se(b_1) = \frac{s_e}{\sqrt{n}} \times \frac{1}{s_{X_1}} \times \sqrt{VIF(X_1)}}$$

- $s_e \rightarrow$ Standard error, estimate of σ_ε
- $s_{X_1} \rightarrow$ Standard deviation in X_1
- As $VIF \uparrow \Rightarrow s_e(b_1) \uparrow$

Collinearity

Occurs when explanatory variables are highly correlated.

- Signs:
  - R^2 does not increase as much on adding explanatory variables.
  - Value of Marginal slopes $>$ Partial slopes
  - Standard errors for Partial slopes $>$ Marginal slopes
  - VIF increases
- Remedies
  - Remove redundant explanatory variables
  - Re-express explanatory variables