Goodness of Fit Measure:

$$\therefore \frac{ESS}{TSS} = \frac{\hat{\beta}^2 \sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2} =$$

$$= \left\{ \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \right\}^2 \cdot \frac{\sum (X_i - \bar{X})^2}{\sum (Y_i - \bar{Y})^2}$$

$$= \frac{\left\{ \sum (X_i - \bar{X})(Y_i - \bar{Y}) \right\}^2}{\left\{ \sum (X_i - \bar{X})^2 \right\}^{\cancel{2}}} \cdot \frac{\cancel{\left\{ \sum (X_i - \bar{X})^2 \right\}}}{\sum (Y_i - \bar{Y})^2}$$

$$= \frac{\left\{ \sum (X_i - \bar{X})(Y_i - \bar{Y}) \right\}^2}{\left\{ \sum (X_i - \bar{X})^2 \right\} \left\{ \sum (Y_i - \bar{Y})^2 \right\}}$$

$$= \left\{ \underbrace{\frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}}}_{\text{corr coeff b/w } X \,\&\, Y} \right\}^2 = (r_{XY})^2 = r^2$$

$$\frac{ESS}{TSS} = R^2$$ [R^2 measure / Goodness of Fit Measure] .

As $-1 \leq r \leq 1 \Rightarrow \boxed{0 \leq R^2 \leq 1} \Rightarrow$ Higher the value of R^2 better fit is the regression model to the given data.

Multiple Regression Analysis :-

True Model : $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_K X_{Ki} + u_i$

No. of explanatory variables $= (K-1)$. $[X_2, X_3, \ldots, X_K]$.

No. of unknown parameters = k.

Eg for $k = 3$:

True Model: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

Estimated Model: $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$

For OLS Estimation: $\underset{\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3}{\text{Min}} \sum e_i^2$

$e_i = Y_i - \hat{Y}_i$

$\therefore \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 0 \Rightarrow (-\cancel{2}) \sum \overbrace{(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})}^{e_i} = 0 \quad \text{---- (i)}$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = 0 \Rightarrow (-\cancel{2}) \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}) X_{2i} = 0 \quad \text{---- (ii)}$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_3} = 0 \Rightarrow (-\cancel{2}) \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}) X_{3i} = 0. \quad \text{---- (iii)}$$

Eqns (i), (ii), (iii) solves for $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$. [HW!].

Properties:

(1) Eqn (i) $\Rightarrow \sum e_i = 0 \Rightarrow \bar{e} = 0$.

Eqn (ii) $\Rightarrow \sum e_i X_{2i} = 0 \Rightarrow cov(X_2, e) = 0 \Rightarrow corr(X_2, e) = 0$

Eqn (iii) $\Rightarrow \sum e_i X_{3i} = 0 \Rightarrow cov(X_3, e) = 0 \Rightarrow corr(X_3, e) = 0$

$\therefore$ The error part is uncorrelated with the entire set of explanatory variables.

(2) $cov(\hat{Y}, e) = \frac{1}{n} \sum (\hat{Y}_i - \bar{\hat{Y}})(e_i - \underbrace{\bar{e}}_{=0})$

$= \frac{1}{n} \sum (\hat{Y}_i - \bar{\hat{Y}}) e_i$

$$= \frac{1}{n}\sum(\hat{Y}_i - \bar{\hat{Y}})e_i$$

$$= \frac{1}{n}\left[\sum \hat{Y}_i e_i - \bar{\hat{Y}}\underbrace{\sum e_i}_{=0}\right].$$

$$= \frac{1}{n}\sum(\hat{Y}_i e_i) = \frac{1}{n}\sum(\hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i})e_i$$

$$= \frac{1}{n}\left[\hat{\beta}_1\underbrace{\sum e_i}_{=0} + \hat{\beta}_2\underbrace{\sum e_i X_{2i}}_{=0} + \hat{\beta}_3\underbrace{\sum e_i X_{3i}}_{=0}\right] = 0.$$

$\Rightarrow corr(\hat{Y}, e) = 0 \Rightarrow$ Error part is uncorrelated with the predicted part of Y.

(3) For ANOVA : $Y = \hat{Y} + e.$

$$Var(Y) = Var(\hat{Y} + e)$$

$$Var(Y) = Var(\hat{Y}) + Var(e) + 2\underbrace{cov(\hat{Y}, e)}_{=0}.$$

$$Var(Y) = Var(\hat{Y}) + Var(e)$$

$$\cancel{\tfrac{1}{n}}\sum(Y_i - \bar{Y})^2 = \cancel{\tfrac{1}{n}}\sum(\hat{Y}_i - \bar{\hat{Y}})^2 + \cancel{\tfrac{1}{n}}\sum e_i^2$$

$$\boxed{\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{\hat{Y}})^2 + \sum e_i^2} \Rightarrow \boxed{TSS = ESS + RSS}$$

Define: $R^2 = \dfrac{ESS}{TSS} \Rightarrow$ Goodness of Fit Measure $[0 \leq R^2 \leq 1]$

Q. Suppose we have data on (Y_i, X_{2i}, X_{3i}), $i = 1, 2, \cdots, n$.
Consider the following models on Y:

$M_1 : Y_i = \beta_1 + \beta_2 X_{2i} + u_i$

$M_2 : Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

How can we assess which model is a better fit to the given data.

Fit $M_1 \Rightarrow$ Obtain R^2_1

Fit $M_2 \Rightarrow$ Obtain R^2_2 $\Big\} \rightarrow$ Compare them,

$\hat{M}_1 : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i}$ ✓ $\Rightarrow \bar{\hat{Y}} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2$

$\hat{M}_2 : \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$ ✓ $\Rightarrow \bar{\hat{Y}} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$

For M_1 : $R_1^2 = \frac{ESS_1}{TSS_1} = \frac{\sum(\hat{Y}_i - \bar{\hat{Y}})^2}{\boxed{\sum(Y_i - \bar{Y})^2}}$

For M_2 : $R_2^2 = \frac{ESS_2}{TSS_2} = \frac{\sum(\hat{Y}_i - \bar{\hat{Y}})^2}{\boxed{\sum(Y_i - \bar{Y})^2}}$ $\rightarrow$ same.

HW.

$ESS_1 =$

$ESS_2 =$