$TSS = ESS + RSS$

$\downarrow$

$\underline{\underline{R^2}}$

\# $y_i = \hat{y}_i + e_i$

or, $y_i - \bar{y} = (\hat{y}_i - \bar{y}) + e_i$

or, $\Sigma(y_i - \bar{y})^2 = \Sigma\left[(\hat{y}_i - \bar{y}) + e_i\right]^2$

⋮

or, $\boxed{\Sigma(y_i - \bar{y})^2 = \hat{\beta}^2 \Sigma(x_i - \bar{x})^2 + \Sigma e_i^2}$

$\Sigma y_i^2 = \hat{\beta}^2 \Sigma x_i^2 + \Sigma e_i^2$

Total sum of square (TSS) = Explain ss (ESS) + unexplained or residual (RSS)

$\therefore R^2 = \dfrac{\text{Variation explained}}{\text{Variation required to be explained}} = \dfrac{\hat{\beta} \Sigma x_i^2}{\Sigma y_i^2}$

or, $R^2 = \dfrac{\Sigma y_i^2 - \Sigma e_i^2}{\Sigma y_i^2} = 1 - \dfrac{\Sigma e_i^2}{\Sigma y_i^2}$

$\therefore \boxed{R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}}$

...em

... in

CLRM

if $R^2 = 0.8$ → 80% of the variation in dependent variable (Y) is explained by the independent variable (X) in the model.

increase in x also increases β (parameters or coefficients of variable)

∴ R^2 will give misleading result if degree of freedom is not considered.

To correct this defect we (adjust R^2) by taking degrees of freedom into account which, as we know, get decreased with inclusion of additional explanatory variable in the model.

We define $\bar{R}^2$ as adjusted R^2 by dividing both sides of R^2 with degree of freedom.

$$\Rightarrow R^2 \Sigma y_i^2 = \Sigma y_i^2 - \Sigma e_i^2$$

$$\Sigma e_i^2 = \Sigma y_i^2 - \Sigma y_i^2 R^2$$

n (obs)

K (~~parameters~~ regressor)

$$R^2 = 1 - \frac{\Sigma e_i^2}{\Sigma y_i^2}$$

(n-1) → d.f → (1)

$$\Sigma e_i^2 = \Sigma y_i^2 (1 - R^2)$$

d.f ∃e

$$RSS = TSS(1 - R^2) \quad \text{— (1)}$$

$(n-1)$

$(n-K) \rightarrow$ d.f 3e

$$RSS = TSS(1-R^2) \quad \text{— (1)}$$

Dividing both sides by degree of freedom

$$\frac{RSS}{(n-K)} = \frac{TSS}{(n-1)}(1-\bar{R}^2)$$

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

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-K}\right)\left(\frac{RSS}{TSS}\right)$$

$$\boxed{\bar{R}^2 = 1 - \frac{RSS/(n-K)}{TSS/(n-1)}}$$

$$\boxed{\bar{R}^2 = 1 - \frac{\Sigma e_i^2/(n-K)}{\Sigma y_i^2/(n-1)}}$$

adjusted R^2 $(\bar{R}^2) < R^2$.

Relation between Simple and multiple Regression coefficients:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \Rightarrow \text{Multiple Regrn}$$

We break three variables into simple regression.

(1) Y on X_2: $\quad Y_i = \beta_{1\cdot2} + \beta_{12} X_{2i} + U_{(1\cdot2)i}$

[Here $U_{1\cdot2}$ include $= \beta_3 X_3 + u_i$

(1) Y on X_2: $Y_i = \underbrace{\beta_{1.2}} + \underbrace{\beta_{12} X_{2i}} + U_{(1.2)i}$

[$U_{1.2}$ include $= \beta X_3 + u_i$]

Here. $\beta_{12} = \dfrac{\sum x_2 y}{\sum x_2^2}$ (regression coeff)

(2) Y on X_3: $Y_i = \beta_{1.3} + \beta_{13} X_{3i} + U_{1.3}$

[$U_{1.3} = U + \beta X_2$]

Here $\beta_{13} = \dfrac{\sum x_3 y}{\sum x_3^2}$ (regression coeff)

(3) X_2 on X_3: $X_{2i} = \beta_{2.3} + \beta_{23} X_{3i} + U_{2.3}$

Here $\beta_{23} = \dfrac{\sum x_2 x_3}{\sum x_3^2}$

actual model
multiple

Y X_2 X_3

β_{12} — Y on X_2

β_{13} — Y on X_3

β_{23} — X_2 on X_3

$Y = \beta_{1.2} + \beta_{12} X_2 + U_{1.2}$

$Y = \beta_{1.3} + \beta_{13} X_3 + U_{1.3}$

$X_2 = \beta_{2.3} + \beta_{23} X_3 + U_{2.3}$

X_3 X_2

1 → Y
2 → X_2
3 → X_3

$\ldots + \beta\, x \ldots + u_i$

3 → x_3

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

or, $$Y_i = \beta_{1.23} + \beta_{12.3} X_{2i} + \beta_{13.2} X_{3i} + U_{1.23}$$

In simple regression β_{12} and β_{13} indicates the effect of X_2 and X_3 on Y respectively

i.e, $\boxed{\beta_{12} = \frac{\partial Y}{\partial x_2}}$ and $\boxed{\beta_{13} = \frac{\partial Y}{\partial x_3}}$

In multiple regression $\beta_{12.3}$ indicates existence of variable X_3 but represents only the effect of X_2 on Y. Similarly $\beta_{13.2}$

∴ Multiple regression coefficients like $\beta_{12.3}$, $\beta_{13.2}$ are known as partial regression coefficient.

How can we relate partial (multiple) regression coefficient with simple regression coefficient.

$$\hat{\beta}_{12.3} = \hat{\beta}_2 = \frac{\Sigma x_2 y \left(\Sigma x_3^2\right) - \Sigma x_2 x_3 \, \Sigma x_3 y}{\Sigma x_2^2 \, \Sigma x_3^2 - \left(\Sigma x_2 x_3\right)^2}$$

... the denominator by $\Sigma x_2^2 (\Sigma x_3^2$

Divide numerator & denominator by $\sum x_2^2 \sum x_3^2$

$$\hat{\beta}_{12\cdot 3} = \frac{\dfrac{\sum x_2 y}{\sum x_2^2} - \dfrac{\sum x_2 x_3}{(\sum x_2)^2} \cdot \dfrac{\sum x_3 y}{\sum x_3^2}}{1 - \left(\dfrac{\sum x_2 x_3}{\sum x_2}\right)\left(\dfrac{\sum x_2 x_3}{\sum x_3}\right)}$$

$$\boxed{\hat{\beta}_{12\cdot 3} = \frac{\hat{\beta}_{12} - \hat{\beta}_{23}\hat{\beta}_{13}}{1 - \hat{\beta}_{32} \cdot \hat{\beta}_{23}}}$$

(Proved)

$\hat{\beta}_{13\cdot 2} = \hat{\beta}_3 = \dfrac{\sum x_3 y \sum x_3^2 - \sum x_2 x_3 \sum x_2 y}{\sum x_2^2 \sum x_3^2 - (\sum x_2 x_3)^2}$

$$= \frac{\dfrac{\sum x_3 y}{\sum x_2^2} - \dfrac{\sum x_2 x_3}{\sum x_3^2} \cdot \dfrac{\sum x_2 y}{\sum x_2^2}}{1 - \dfrac{\sum x_2 x_3}{\sum x_2^2} \cdot \dfrac{\sum x_2 x_3}{\sum x_3^2}}$$

or, $\hat{\beta}_{13\cdot 2} = \dfrac{\hat{\beta}_{13} - \hat{\beta}_{23}\hat{\beta}_{13}}{1 - \hat{\beta}_{32}\hat{\beta}_{23}}$

Regression m y on x: $Y = \bar{Y} + b_{yx}\bar{X}$ $\quad b_{yx} = \dfrac{\mathrm{cov}(x,y)}{V(x)}$

Consider $\hat{\beta}_{32} \cdot \hat{\beta} = \sum x_2 x_3 \quad \sum x_2 x_3$

Consider $\hat{\beta}_{23}\cdot\hat{\beta}_{32} = \frac{\Sigma x_2 x_3}{\Sigma x_3^2}\cdot\frac{\Sigma x_2 x_3}{\Sigma x_2^2}$

$\sigma_3^2 \ \sigma_2^2$

$\frac{1}{n}\Sigma x_3^2 = \frac{1}{n}\Sigma(x_3-\bar{x}_3)^2 = n\sigma_3^2$

$r_{xy} = \frac{cov}{\sqrt{v_x}\sqrt{v_y}}$

$$= \frac{(\Sigma x_2\cdot x_3)^2}{\Sigma x_3^2\,\Sigma x_2^2}$$

$$= \frac{(\Sigma x_2\cdot x_3)^2}{n\sigma_3^2\cdot n\sigma_2^2}$$

$$= \left(\frac{\Sigma x_2\cdot x_3}{n\,\sigma_3\cdot\sigma_2}\right)^2$$

$$= r_{23}^2$$

$$\hat{\beta}_{12.3} = \frac{\hat{\beta}_{12}-\hat{\beta}_{32}\,\hat{\beta}_{13}}{1-r_{23}^2} \quad \text{and} \quad \hat{\beta}_{13.2} = \frac{\hat{\beta}_{13}-\hat{\beta}_{22}\,\hat{\beta}_{13}}{1-r_{23}^2}$$

a) $\mu \Rightarrow \sigma$ known $\Rightarrow Z/\gamma$ stat.

b) $\mu \Rightarrow \sigma$ unknown (sample s) $\longrightarrow$ t statistics.

δ e^2 unknown

(No relation b/w x and y)

$H_0: \hat{\beta}_1 = 0$ (No relation b/w x and y)

vs $H_1: \beta_1 \neq 0$ (Relation x and y)

$$SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\Sigma x_i^2}}$$

Test statistics: $t = \dfrac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)}$

$\hat{Y} = \hat{\alpha} + \hat{\beta} x$

Under H_0: $t = \dfrac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \dfrac{\hat{\beta}_1 \sqrt{\Sigma x_i^2}}{\hat{\sigma}}$

$SE(\ )$

$\hat{Y} = 0.1 + 0.3X$
$(?)$
3

SE: 2

$t = \dfrac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$

$\dfrac{0.3}{3}$

$t = ans$