Simple Regression Model

$$y = \beta_0 + \beta_1 x + u.$$

Slope variable

Error term / disturbance.

intercept

Independent / explanatory / regressor } variable

Dependent / explained / response } variable

$$\Delta y = \beta_1 \Delta x + \Delta u.$$

if $\Delta u = 0$ Then $\Delta y = \beta_1 \Delta x$.

$$\Delta x = 1 \Longrightarrow \Delta y = \beta_1$$

(Zero conditional mean assumption).

$E(u \mid \underline{x}) = E(u) = 0$ (if $E(u) = \alpha \neq 0$,

$(x_1 = x_1^* \quad x_2 = x_2^* \cdots x_n = x_n^*)$

then $\beta_0 \to \beta_0' = \beta_0 + \alpha$, $u \to u' = u - \alpha$. $E(u') = 0$.)

$\Longrightarrow Cov(u, x) = 0.$

$$E(Y \mid x) = E(\beta_0 + \beta_1 x + u \mid x)$$
$$= \beta_0 + \beta_1 x.$$

PRF: population regression function.

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

$\hat{u}_i = Y_i - \hat{Y}_i$ residuals.

$\sum_{i=1}^{n} \hat{u}_i^2$ sum of squared residuals (SSR).

The method of ordinary least squares (OLS).

$$\min SSR = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = 0. \quad \frac{\partial SSR}{\partial \hat{\beta}_1} = 0.$$

$$SSR = \sum_{i=1}^{n} \left[Y_i^2 + (\hat{\beta}_0 + \hat{\beta}_1 X_i)^2 - 2Y_i(\hat{\beta}_0 + \hat{\beta}_1 X_i) \right]$$

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} \left(2(\hat{\beta}_0 + \hat{\beta}_1 X_i) - 2Y_i \right) = 0$$

$$\sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 X_i - Y_i) = 0.$$

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i = 0$$

$$\hat{\beta}_0 + \hat{\beta}_1 \bar{X} - \bar{Y} = 0.$$

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} \left(2X_i(\hat{\beta}_0 + \hat{\beta}_1 X_i) - 2Y_i X_i \right) = 0.$$

$$\sum_{i=1}^{n} X_i (\hat{\beta}_0 + \hat{\beta}_1 X_i - Y_i) = 0.$$

$$\hat{\beta}_0 n\bar{X} + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} X_i Y_i = 0$$

$$(\bar{y} - \hat{\beta}_1 \bar{x})n\bar{x} + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i Y_i = 0.$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i Y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$ form 1.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Algebraic Properties of OLS.

(1) $\sum_{i=1}^{n} \hat{u}_i = 0$ $\quad \hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

(2)

$$\sum_{i=1}^{n} \hat{u}_i X_i = 0 \Longrightarrow r_{u,x} = 0$$

$\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0 \Rightarrow Cov(u, \hat{y}) = 0.$

$$r_{u,x} = \frac{\overbrace{Cov(u,x)}^{=0}}{\sqrt{Var(u)\,Var(x)}} = \frac{E(uX) - E(U)E(X)}{\sqrt{\sum(\hat{u}_i - \underbrace{\bar{u}}_{=0})^2 \sum(x_i - \bar{x})^2}}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i X_i - 0}{\sqrt{Var\,u \quad Var\,X}}$$

$$\sum_{i=1}^{n} \hat{u}_i X_i = 0 = n\bar{u}\bar{x} \Rightarrow r_{u,x} = 0.$$

total sum of squares (SST) $\equiv \sum_{i=1}^{n} (y_i - \bar{y})^2$

explained sum of squares (SSE) $\equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

residual sum of squares (SSR) $\equiv \sum_{i=1}^{n} \hat{u}_i^2$

$$= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} \left[(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}) \right]^2$$

$$= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2$$

$$+ 2 \sum_{i=1}^{n} \underbrace{(y_i - \hat{y}_i)}_{\hat{u}_i - \bar{u}} (\hat{y}_i - \bar{y})$$

$\sum \hat{u}_i x_i = 0$

$= \sum \hat{u}_i (\hat{\beta}, x_i)$

$= Cov(u, \hat{y}) = 0$

SST = SSR + SSE

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Coefficent of determination.

$100 \cdot R^2$ is the percent of varation in Y can be explained by X.