Multiple Regression Analysis.

Multiple linear regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u.$$

Example: $cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u.$

$$\frac{\Delta cons}{\Delta inc} \approx \beta_1 + 2\beta_2 inc.$$

$MPC = \beta_1 + 2\beta_2 inc$

OLS estimation of the multiple regression model

Random sample: $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i): i = 1, \ldots, n\}$

Regression residuals: $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}.$

$$\min \sum_{i=1}^{k} \hat{u}_i^2 \longrightarrow \hat{\beta}_0, \ldots, \hat{\beta}_k.$$

Assumption: $E(u \mid x_1, \ldots, x_k) = E(u) = 0.$

$$\Longrightarrow Cov(u, x_i) = 0, \quad i = 1, \ldots, k.$$

$$E(y \mid x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k.$$

parameters regression function (PRF)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots \hat{\beta}_k x_k$$

Sample Regression Line (SRL)

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik})^2$$

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

$$\Longrightarrow \underbrace{\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik})}_{\sum \hat{u}_i = 0} = 0$$

$$\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}_1 - \cdots - \hat{\beta}_k \bar{x}_k = 0$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \cdots - \hat{\beta}_k \bar{x}_k$$

$$\frac{\partial SSR}{\partial \hat{\beta}_j} = \sum_{i=1}^{n} -2x_{i,j} \underbrace{(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik})}_{\hat{u}_i} = 0 \qquad \sum X_{i,j} \hat{u}_i = 0$$

$j = 1, 2, 3, \ldots, k.$

Algebraic Properties of OLS.

- $\sum_{i=1}^{n} \hat{u}_i = 0$
- $\sum_{i=1}^{n} X_{i,j} \hat{u}_i = 0 \qquad j = 1, \ldots, k.$

$$\Longrightarrow r_{\hat{u}, x_j} = 0$$

$$Cov(X_j, \hat{u}) = E(X_j - \bar{X}_j)(\hat{u} - \underbrace{\bar{u}}_{=0})$$

$$= E(X_j \hat{u}) - \bar{X}_j E(\hat{u}) = E(X_j \hat{u}) = 0.$$

Partialling out interpretation of OLS estimators.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u.$$

want to find $\hat{\beta}_1$:

Step 1: $x_1 = \alpha_0 + \alpha_2 x_2 + \cdots + \alpha_k x_k + r.$

$$\hat{x}_1 = \hat{\alpha}_0 + \hat{\alpha}_2 x_2 + \cdots + \hat{\alpha}_k x_k$$

$$\hat{r}_{i,1} = x_{i1} - \hat{x}_{i,1} \implies \sum_{i=1}^{n} \hat{r}_{i,1} = 0, \quad \bar{\hat{r}}_1 = 0$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 (\hat{r}_{i,1} + \hat{x}_{i,1}) + \hat{\beta}_2 x_{i,2} \cdots + \hat{\beta}_k x_{i,k}$$

function about $x_2, \cdots, x_k$

Hence $\hat{r}_{i,1}$ can instead x_i

Step 2: Regress y on $\hat{r}_1$, the estimate of the slope from this simple regression given $\hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{\sum (\hat{r}_{i,1} - \bar{\hat{r}}_1)(y_i - \bar{y})}{\sum (\hat{r}_{i,1} - \bar{\hat{r}}_1)^2}$$

$$= \frac{\sum \hat{r}_{i,1} (y_i - \bar{y})}{\sum \hat{r}_{i,1}^2} = \frac{\sum \hat{r}_{i,1} y_i}{\sum \hat{r}_{i,1}^2}$$

$$SST = \sum (y_i - \bar{y}_i)^2 ,$$

$$SSR = \sum (y_i - \hat{y}_i)^2 ,$$

$$SSE = \sum (\hat{y}_i - \bar{y}_i)^2$$

$$SST = SSR + SSE$$

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

Consider Simple Regression Model.

$$r_{y,x} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$$r^2_{y,x} = \frac{\left(\sum (x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}$$

$$= \frac{\left(\sum\left[(x_i - \bar{x})(y_i - \hat{y}_i) + (x_i - \bar{x})(\hat{y}_i - \bar{y})\right]\right)^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}$$

$$= \frac{\left(\sum (x_i - \bar{x})(\hat{y}_i - \bar{y})\right)^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}$$

$$\frac{\cancel{(x_i - \bar{x})}(\hat{y}_i - \bar{y})(x_j - \bar{x})\cancel{(\hat{y}_j - \bar{y})}}{(x_i - \bar{x})(\hat{y}_j - \bar{y})} = \frac{\hat{\beta}_1 (x_i - \bar{x})(x_j - \bar{x})}{\hat{\beta}_1 (x_i - \bar{x})(x_j - \bar{x})} = 1$$

$$\Longrightarrow \quad = \frac{\sum (x_i - \bar{x})^2 \sum (\hat{y}_i - \bar{y})^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2} = R^2$$

Consider Multiple Regression Model.

$$r^2_{y,\hat{y}} = \frac{\left(\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}$$

$$= \frac{\left(\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{\hat{y}}) + \sum (\hat{y}_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}$$

$$= \frac{\left(\sum (\hat{y}_i - \bar{y})^2\right)^2}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}$$

$$= \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = R^2$$

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u. \Rightarrow \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$$

$$y_1 = \beta_0 + \beta_1 x_1 + v. \Rightarrow \tilde{\beta}_0, \tilde{\beta}_1$$

$$\tilde{\beta}_1 = \hat{\beta}_1 + \tilde{\delta}_1 \hat{\beta}_2$$

$\tilde{\delta}_1$ is the estimator of the slope from the regression of x_2 on x_1. (i.e. $\tilde{\delta}_1 = \frac{\sum (x_{i,1} - \bar{x}_1)(x_{i,2} - \bar{x}_2)}{\sum (x_{i,1} - \bar{x}_1)^2}$)

Proof: $x_2 = \delta_1 x_1 + e. \Rightarrow \bar{x}_2 = \tilde{\delta}_1 \bar{x}_1$

$$\bar{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 \Rightarrow \bar{y}_1 = \hat{\beta}_0 + (\hat{\beta}_1 + \tilde{\delta}_1 \hat{\beta}_2)\bar{x}_1$$

$$\bar{y}_1 = \tilde{\beta}_0 + \tilde{\beta}_1 \bar{x}_1$$

$$\Longrightarrow \begin{cases} \tilde{\beta}_0 = \hat{\beta}_0 \\ \tilde{\beta}_1 = \hat{\beta}_1 + \tilde{\delta}_1 \hat{\beta}_2 \end{cases}$$

when $\tilde{\delta}_1 = 0$ or $\hat{\beta}_2 = 0$, $\tilde{\beta}_1 = \hat{\beta}_1$