Multi-variate Statistics

Extension of Bi-variate Statistics

$(Y, \boldsymbol{X})$~ random variables

where

$\boldsymbol{X}$~ vectors of K random variables

$$\boldsymbol{X} = [X_1, X_2, \ldots, X_K]$$

Y ~ a single random variable

Multi-variate Analyses

- Pair-wise Covariance or Correlation
- Multi-way ANOVA
- <u>Multiple Regression</u>

Multiple Regression Analysis

Focus on the dependency of Y on the $\boldsymbol{X}$ vector, e.g.,

$$\mu_{Y|X} = m(X_1, X_2, \ldots, X_K) = m(X)$$

$$\sigma^2_{Y|X} = v(X_1, X_2, \ldots, X_K) = v(X)$$

X_k - explanatory or independent variable,

$k = 1,\ldots,K$

Y - dependent variable

Multiple Linear Regression

Assumptions

1) linearity $\mu_{Y|X} = X\boldsymbol{\beta}$

   where $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \ldots \quad \beta_K]^T$ are unknown parameters

2) variance-independent or $\sigma^2_{Y|X} = \sigma^2$

3) normality, i.e. $Y \mid X \sim N(X\boldsymbol{\beta}, \sigma^2)$

CLNRM (1)

Classical Linear Normal Regression Model is based upon the assumptions

$$Y_i = X_i\beta + \varepsilon_i$$

where i = index of the observation

ε_i = identical and independent

normal error term

$\varepsilon_i \sim N(0, \sigma^2)$ for all $i=1,\ldots,n$

CLNRM (2)

$\boldsymbol{X}_i$ is pre-selected or non-random but Y_i or ε_i is randomly sampled.

$\boldsymbol{X}_i\boldsymbol{\beta}$ is the non-random component of Y_i

ε_i is the random component of Y_i.

Note that X_1 can be intentionally set to one for all observations so that its coefficient β_1 becomes the y-intercept.

CLNRM
Matrix Representation (1)

Define

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{K1} \\ X_{12} & X_{22} & \cdots & X_{K1} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{Kn} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

OLS Estimation for CLNRM (1)

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} [Y_i - (X_{1i}\beta_1 + X_{2i}\beta_2 + \ldots + X_{Ki}\beta_K)]^2$$

or

$$\min_{\boldsymbol{\beta}} \left[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right]^{\mathrm{T}} \left[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\right]$$

OLS Estimation for CLNRM (2)

First-Order Conditions

$$2\left[-\mathbf{X}\right]^{\mathrm{T}}\left[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right]=\mathbf{0}$$

$$-\mathbf{X}^{\mathrm{T}}\mathbf{Y}+\mathbf{X}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}=\mathbf{0}$$

$$\hat{\boldsymbol{\beta}}=\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$$

OLS Estimation for CLNRM (3)

Estimator for σ^2

$$\hat{\sigma}^2 = \frac{1}{n-K}\left[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right]^{\mathrm{T}}\left[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right]$$

$$= \frac{1}{n-K}\left[\mathbf{Y}^{\mathrm{T}}\mathbf{Y}-\mathbf{Y}^{\mathrm{T}}\hat{\mathbf{Y}}\right]$$

where $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ is called the fitted value of $\mathbf{Y}$

Why *n-K*?

Properties of OLS estimators (1)

<u>Theorem</u> $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2 \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}$$

Does not require normality assumption.

Note that $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$.

Properties of OLS estimators (5)

Variance-Covariance Matrix of $\hat{\boldsymbol{\beta}}$

$$V(\hat{\boldsymbol{\beta}}) = \sigma^2 \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}$$

$$= \begin{bmatrix} V(\hat{\beta}_1) & C(\hat{\beta}_1, \hat{\beta}_2) & \cdots & C(\hat{\beta}_1, \hat{\beta}_K) \\ C(\hat{\beta}_2, \hat{\beta}_1) & V(\hat{\beta}_2) & \cdots & C(\hat{\beta}_2, \hat{\beta}_K) \\ \vdots & \vdots & \ddots & \vdots \\ C(\hat{\beta}_K, \hat{\beta}_1) & C(\hat{\beta}_K, \hat{\beta}_2) & \cdots & V(\hat{\beta}_K) \end{bmatrix}$$

σ^2 is generally unknown.

Properties of OLS estimators (6)

Estimated Variance-Covariance Matrix of $\hat{\boldsymbol{\beta}}$

$$\hat{V}(\hat{\boldsymbol{\beta}}) = \widehat{\sigma^2}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

$$= \begin{bmatrix} \hat{V}(\hat{\beta}_1) & \hat{C}(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \hat{C}(\hat{\beta}_1, \hat{\beta}_K) \\ \hat{C}(\hat{\beta}_2, \hat{\beta}_1) & \hat{V}(\hat{\beta}_2) & \cdots & \hat{C}(\hat{\beta}_2, \hat{\beta}_K) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}(\hat{\beta}_K, \hat{\beta}_1) & \hat{C}(\hat{\beta}_K, \hat{\beta}_2) & \cdots & \hat{V}(\hat{\beta}_K) \end{bmatrix}$$

Properties of OLS estimators (7)

Standard Deviation of $\hat{\beta}_k$

$$sd(\hat{\beta}_k) = \sqrt{V(\hat{\beta}_k)}$$

Standard Error of $\hat{\beta}_k$

$$se(\hat{\beta}_k) = \sqrt{\hat{V}(\hat{\beta}_k)}$$

Properties of OLS estimators (8)

$$t_{cal} = \frac{\dfrac{\hat{\beta}_k - \beta_k}{sd(\hat{\beta}_k)}}{\sqrt{\dfrac{(n-K)\dfrac{\widehat{\sigma^2}}{\sigma^2}}{n-K}}}$$

$$= \frac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)} \sim t(n-K)$$

<<Basis for statistical inference>>

Central Limit Theorem (1)

Similar to that for the Simple Linear Regression Model. Even though the err or terms are <u>not</u> normal, the properties of OLS estimators asymptotically hold when the sample size is very large.

Gauss-Markov Theorem (1)

Similar to that for the Simple Linear Regression Model. Given that $\mathbf{X}$ is non-random, OLS estimator is Best Linear Unbiased Estimator.

Coefficient of Determination (1)

R^2 is a measure for goodness-of-fit. How well does the model fit the observed data ? Low R^2 implies "bad" fit.

<u>Definition</u> $$R^2 \equiv 1 - \frac{SSR}{SST}$$

SSR = Sum of Squared Residuals

SST = Sum of Squared Totals

Coefficient of Determination (2)

where $$SSR = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = [\mathbf{Y} - \hat{\mathbf{Y}}]^{\mathrm{T}} [\mathbf{Y} - \hat{\mathbf{Y}}]$$

$$SST = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

Note that, in general, R^2 cannot be greater than one but could be negative.

Coefficient of Determination (3)

Low R^2 or a bad fit does <u>not</u> mean a bad model. It simply implies a large uncertainty in the nature. It is mainly used as a criterion to select various "candidate" models.

Coefficient of Determination (4)

If an X_i has constant value or a linear combination of X_i 's is equivalent to a c onstant value, then, $0 \le R^2 \le 1$ always

and
$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

Coefficient of Determination (5)

Interpretation if $0 \leq R^2 \leq 1$

$1 - R^2$ or SSR/SST can be interpreted as the fraction of total variation of Y due to the random component (ε).

R^2 is generally regarded as the fraction of total variation of Y explained by the explanatory variables or due to the non-random component.

Adjusted- R^2 (1)

We can cheat on R^2 by adding more irrelevant independent variables on the right-hand side, especially when sampl e is small.

Higher K ==> smaller SSR ==>higher R^2

Adjusted- R^2 (2)

Definition

$$\overline{R}^2 \equiv 1 - \frac{SSR/(n-K)}{SST/(n-1)} = 1 - \frac{\widehat{\sigma^2}}{s_Y^2}$$

Concept

Penalize R^2 by dividing with (n-K)when an irrelevant variable is added.

Adjusted- R^2 (3)

Purpose

For a small sample, it is a better measure for goodness-of-fit than R^2. It is also used as criterion to add or r emove an explanatory variable from the model if it does not contradict th eories.