STATISTICAL COMPUTATION

WEEK 13 – LINEAR REGRESSION (2/2)

Annisa Auliya
I Melda Puspita

GET TO KNOW US

ANNISA AULIYA R.

082334174749

I MELDA PUSPITA L.

085257113961

https://intip.in/KomstatC2023

MATERIALS

Multiple Linear Regression

- Parameter Estimation
- Significance Test (Simultan & Partial

01

PARAMETER ESTIMATION

ESTIMATION

Observation, i	Response, y	Regressors x_1	x_2	$\ldots$	x_k
1	y_1	x_{11}	x_{12}	$\ldots$	x_{1k}
2	y_2	x_{21}	x_{22}	$\ldots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
n	y_n	x_{n1}	x_{n2}	$\ldots$	x_{nk}

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ESTIMATION

We wish to find the vector of least-squares estimators, $\boldsymbol{\beta}$, that minimizes

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Note that $S(\boldsymbol{\beta})$ may be expressed as

$$\begin{aligned} S(\boldsymbol{\beta}) &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \end{aligned}$$

since $\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ is a 1×1 matrix, or a scalar, and its transpose$(\boldsymbol{\beta}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\boldsymbol{\beta}$ is the same scalar. The least-squares estimators must satisfy

ESTIMATION

since $\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}$ is a 1×1 matrix, or a scalar, and its transpose$(\boldsymbol{\beta}'\mathbf{X}'\mathbf{y})' = \mathbf{y}'\mathbf{X}\boldsymbol{\beta}$ is the same scalar. The least-squares estimators must satisfy

$$\left.\frac{\partial S}{\partial \boldsymbol{\beta}}\right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$$

which simplifies to

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{3.12}$$

To solve the normal equations, multiply both sides of (3.12) by the inverse of $\mathbf{X}'\mathbf{X}$. Thus, the **least-squares estimator** of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3.13}$$

02

SIGNIFICANCE TEST

SIMULTAN

$$H_0: \beta_1 = \beta_1 = \cdots = \beta_k = 0$$

$$H_1: \beta_j \neq 0 \quad \text{for at least one } j$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	SS_R	k	MS_R	MS_R/MS_{Res}
Residual	SS_{Res}	$n-k-1$	MS_{Res}	
Total	SS_T	$n-1$		

$$SS_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$SS_{Res} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} - \left[\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right]$$

$$SS_T = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$SS_T = SS_R + SS_{Res}$$

$$\hat{\sigma}^2 = MS_{Res}$$

Therefore, to test the hypothesis H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$, compute the test statistic F_0 and reject H_0 if $F_0 > F_{\alpha,k,n-k-1}$

PARTIAL

The hypotheses for testing the significance of any individual regression coefficient, such as β_j, are

$$H_0\colon \beta_j = 0, \quad H_1\colon \beta_j \neq 0 \tag{3.28}$$

If $H_0\colon \beta_j = 0$ is not rejected, then this indicates that the regressor x_j can be deleted from the model. The **test statistic** for this hypothesis is

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{\mathrm{se}\left(\hat{\beta}_j\right)} \tag{3.29}$$

where C_{jj} is the diagonal element of $(\mathbf{X'X})^{-1}$ corresponding to $\hat{\beta}_j$. The null hypothesis $H_0\colon \beta_j = 0$ is rejected if $|t_0| > t_{\alpha/2,n-k-1}$. Note that this is really a **partial** or **marginal test** because the regression coefficient $\hat{\beta}_j$ depends on all of the other regressor variables $x_i (i \neq j)$ that are in the model. Thus, this is a test of the **contribution** of x_j **given the other regressors in the model**.

THANKS

https://intip.in/KomstatC2023