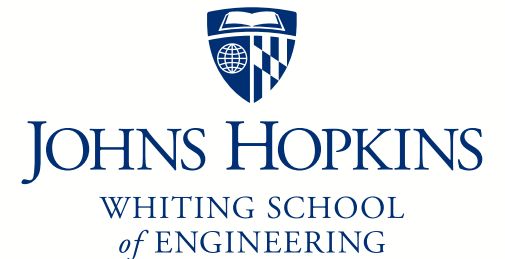


Multiple Linear Regression Model – Part III

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 3 – Lecture 3D



Properties of Least-Squares Estimators

$\hat{\beta}$ is an unbiased estimator of β

$$\begin{aligned} E(\hat{\beta}) &= E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}] = E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\beta + \epsilon)] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\epsilon] = \beta \end{aligned}$$

$$MS_{\text{Res}} = \frac{SS_{\text{Res}}}{n - p}$$

$$\hat{\sigma}^2 = MS_{\text{Res}} : \text{unbiased for } \sigma^2$$

Variance-Covariance of Least-Squares Estimators

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \text{Var}(\hat{\boldsymbol{\beta}}) = \text{Var}[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}] \quad \begin{array}{l} \text{Var}(\mathbf{A}\mathbf{y}) = \mathbf{A}\text{Var}(\mathbf{y})\mathbf{A}' \\ \text{Var}(\mathbf{y}) = \sigma^2 \mathbf{I}_n \end{array}$$

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}) &= \text{Var}[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \text{Var}(\mathbf{y}) [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}']' \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Let $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$.

The variance of $\hat{\beta}_j$ is $\sigma^2 C_{jj}$

The covariance between $\hat{\beta}_i$ and $\hat{\beta}_j$ is $\sigma^2 C_{ij}$.

Estimate Variance-Covariance of Least-Squares Estimators

To estimate the variances of the least-square estimators, just replace σ^2 with the unbiased estimator $\hat{\sigma}^2$ in the variance-covariance formulas.

Gauss-Markov Theorem

For the linear regression with $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2$, and uncorrelated errors, the least-square estimators are unbiased and have the minimum variances when compared to any estimator that is a linear combination of y_i .

Test on Individual Regression Coefficient

$$H_0: \beta_j = 0, \quad H_1: \beta_j \neq 0$$

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)}$$

C_{jj} is the diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$ corresponding to $\hat{\beta}_j$.

$H_0: \beta_j = 0$ is rejected if $|t_0| > t_{\alpha/2, n-k-1}$

The Delivery Time Data (Ex 3.4, page 88 of Textbook)

Suppose that we wish to assess the value of the x_2 (distance) given x_1 (cases) is fixed in the model.

$$H_0: \beta_2 = 0, \quad H_1: \beta_2 \neq 0$$

$$t_0 = \frac{\hat{\beta}_2}{\sqrt{\hat{\sigma}^2 C_{22}}} = \frac{0.01438}{\sqrt{(10.6239)(0.00000123)}} = 3.98$$

which is larger than $t_{0.025,22} = 2.074$, we reject H_0

Maximum Likelihood Estimation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad , \quad \boldsymbol{\varepsilon} \sim \mathbf{NIND}(\mathbf{0}, \sigma^2 \mathbf{I})$$

Loglikelihood:

$$\ln L(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

MLEs:

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

$$\tilde{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}$$



JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING