Multiple Linear Regression

Dr. Kiah Wah Ong

ILLINOIS TECH

Properties of the Least Squares Estimators

Recall from the previous lesson that

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

We now want to investigate the properties of these least squares estimators.

Properties of the Least Squares Estimators

Notice that if A and B are matrices of the same size, then

$$E(A+B) = E(A) + E(B)$$

Also, if $A = (a_{ij})$ is a $r \times p$ constant matrix, while Y is a $p \times c$ random matrix, then

$$\begin{aligned} E(AY) &= E\left(\left[\sum_{k=1}^{p} a_{ik} Y_{kj}\right]\right) \\ &= \left[E\left(\sum_{k=1}^{p} a_{ik} Y_{kj}\right)\right] \\ &= \left[\sum_{k=1}^{p} a_{ik} E(Y_{kj})\right] \\ &= AE(Y) \end{aligned}$$

Similarly, we have $E(AYB) = AE(Y)B$.

Properties of the Least Squares Estimators

From

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

we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \\
&= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}
\end{aligned}$$

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E(\boldsymbol{\epsilon}) = \boldsymbol{\beta}$$

as $E(\boldsymbol{\epsilon}) = \mathbf{0}$.

That is, $E(\hat{\beta}_i) = \beta_i$ for all $i = 0, 1, \cdots, k$. Hence $\hat{\beta}_i$ is an **unbiased estimator** of β_i, $i = 0, 1, \cdots, k$.

Properties of the Least Squares Estimators

Analogous to $\mathrm{Var}(ay) = a^2 \mathrm{Var}(y)$, we have

$$\begin{aligned}
\mathrm{Var}(A\mathbf{y}) &= E[(A\mathbf{y} - A\boldsymbol{\mu})(A\mathbf{y} - A\boldsymbol{\mu})^T] \\
&= E[(A(\mathbf{y} - \boldsymbol{\mu}))(A(\mathbf{y} - \boldsymbol{\mu}))^T] \\
&= E[A(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T A^T] \\
&= AE[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^T]A^T \\
&= A\mathrm{Var}(\mathbf{y})A^T
\end{aligned}$$

by using $E(A\mathbf{y}B) = AE(\mathbf{y})B$ from the previous slide and with $\boldsymbol{\mu} = E(\mathbf{y})$.

Let us now examine the term $\mathrm{Var}(\mathbf{y})$ a little more closely.

Properties of the Least Squares Estimators

Suppose

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \quad \text{and} \quad E(y_i) = \mu_i, \quad i = 1, 2, 3$$

then the variance of $\mathbf{y}$ is given by

$$\begin{aligned}
\mathrm{Var}(\mathbf{y}) &= E\left(\begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \\ y_3 - \mu_3 \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 & y_2 - \mu_2 & y_3 - \mu_3 \end{bmatrix} \right) \\
&= E\left(\begin{bmatrix} (y_1 - \mu_1)^2 & (y_1 - \mu_1)(y_2 - \mu_2) & (y_1 - \mu_1)(y_3 - \mu_3) \\ (y_2 - \mu_2)(y_1 - \mu_1) & (y_2 - \mu_2)^2 & (y_2 - \mu_2)(y_3 - \mu_3) \\ (y_3 - \mu_3)(y_1 - \mu_1) & (y_3 - \mu_3)(y_2 - \mu_2) & (y_3 - \mu_3)^2 \end{bmatrix} \right) \\
&= \begin{bmatrix} E[(y_1 - \mu_1)]^2 & E[(y_1 - \mu_1)(y_2 - \mu_2)] & E[(y_1 - \mu_1)(y_3 - \mu_3)] \\ E[(y_2 - \mu_2)(y_1 - \mu_1)] & E[(y_2 - \mu_2)^2] & E[(y_2 - \mu_2)(y_3 - \mu_3)] \\ E[(y_3 - \mu_3)(y_1 - \mu_1)] & E[(y_3 - \mu_3)(y_2 - \mu_2)] & E[(y_3 - \mu_3)^2] \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{Var}(y_1) & \mathrm{Cov}(y_1, y_2) & \mathrm{Cov}(y_1, y_3) \\ \mathrm{Cov}(y_1, y_2) & \mathrm{Var}(y_2) & \mathrm{Cov}(y_2, y_3) \\ \mathrm{Cov}(y_1, y_3) & \mathrm{Cov}(y_2, y_3) & \mathrm{Var}(y_3) \end{bmatrix}
\end{aligned}$$

ILLINOIS TECH

Properties of the Least Squares Estimators

When y_i are uncorrelated, then $\mathrm{Cov}(y_i, y_j) = 0$ when $i \neq j$ and we have

$$\mathrm{Var}(\mathbf{y}) = \begin{bmatrix} \mathrm{Var}(y_1) & \mathrm{Cov}(y_1, y_2) & \mathrm{Cov}(y_1, y_3) \\ \mathrm{Cov}(y_1, y_2) & \mathrm{Var}(y_2) & \mathrm{Cov}(y_2, y_3) \\ \mathrm{Cov}(y_1, y_3) & \mathrm{Cov}(y_2, y_3) & \mathrm{Var}(y_3) \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{Var}(y_1) & 0 & 0 \\ 0 & \mathrm{Var}(y_2) & 0 \\ 0 & 0 & \mathrm{Var}(y_3) \end{bmatrix}$$

If $\mathrm{Var}(y_i) = \sigma^2$, $i = 1, 2, 3$, then

$$\mathrm{Var}(\mathbf{y}) = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \sigma^2 I$$

Properties of the Least Squares Estimators

Back to the equation

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathrm{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\end{aligned}$$

we can then write

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathrm{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\end{aligned}$$

Properties of the Least Squares Estimators

The **Gauss-Markov** theorem established that the least squares estimator $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator of $\boldsymbol{\beta}$.

Estimation of σ^2

As in simple linear regression, an estimator of σ^2 can be develop from the residual sum of squares

$$SS_R = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e}$$

It can be shown that

$$\frac{SS_R}{\sigma^2} \sim \chi^2_{n-(k+1)}$$

and so

$$E\left(\frac{SS_R}{\sigma^2}\right) = n - k - 1$$

or

$$E\left(\frac{SS_R}{n-k-1}\right) = \sigma^2$$

Estimation of σ^2

From

$$E\left(\frac{SS_R}{n-k-1}\right) = \sigma^2$$

we see that

$$\frac{SS_R}{n-k-1}$$

is an unbiased estimator of σ^2, i.e.

$$\hat{\sigma}^2 = \frac{SS_R}{n-k-1}$$

Also, just like in the case of the simple linear regression, SS_R will be independent of the least squares estimators $\hat{\beta}_0, \hat{\beta}_1, \cdots \hat{\beta}_k$.