# LINEAR REGRESSION ANALYSIS

## MODULE – III

## Lecture - 10

# Multiple Linear Regression Analysis

Dr. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

## Maximum likelihood estimation

In the model $y = X\beta + \varepsilon,$ it is assumed that the errors are normally and independently distributed with constant variance $\sigma^2$ i.e.,

$$\varepsilon \sim N(0, \sigma^2 I).$$

The normal density function for the errors is

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[ -\frac{1}{2\sigma^2}\varepsilon_i^2 \right] \quad i = 1, 2, ..., n.$$

The likelihood function is the joint density of $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ given as

$$\begin{aligned} L(\beta, \sigma^{-2}) &= \prod_{i=1}^{n} f(\varepsilon_i) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[ -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[ -\frac{1}{2\sigma^2}\varepsilon'\varepsilon \right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[ -\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta) \right]. \end{aligned}$$

Since the log transformation is monotonic, so we maximize $\ln L(\beta, \sigma^2)$ instead of $L(\beta, \sigma^2)$.

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

The maximum likelihood estimators (m.l.e.) of $\beta$ and $\sigma^2$ are obtained by equating the first order derivatives of $\ln L(\beta,\sigma^2)$ with respect to $\beta$ and $\sigma^2$ to zero as follows:

$$\frac{\partial \ln L(\beta,\sigma^2)}{\partial \beta} = \frac{1}{2\sigma^2} 2X'(y - X\beta) = 0$$

$$\frac{\partial \ln L(\beta,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(y - X\beta)'(y - X\beta).$$

The likelihood equations are given by

$$X'X\beta = X'y$$

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}).$$

Since $\text{rank}(X) = k,$ so that the unique m.l.e. of $\beta$ and $\sigma^2$ are obtained as

$$\tilde{\beta} = (X'X)^{-1}X'y$$

$$\tilde{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}).$$

Next we verify that these values maximize the likelihood function. First we find

$$\frac{\partial^2 \log L(\beta,\sigma^2)}{\partial \beta^2} = -\frac{1}{\sigma^2} X'X$$

$$\frac{\partial^2 \log L(\beta,\sigma^2)}{\partial^2 (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(y - X\beta)'(y - X\beta)$$

$$\frac{\partial^2 \log L(\beta,\sigma^2)}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4} X'(y - X\beta).$$

Thus the Hessian matrix of second order partial derivatives of $\ln L(\beta, \sigma^2)$ with respect to $\beta$ and $\sigma^2$ is

$$\begin{pmatrix} \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial \beta^2} & \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial \beta \partial \sigma^2} \\ \dfrac{\partial^2 \log L(\beta,\sigma^2)}{\partial \sigma^2 \partial \beta} & \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial^2 (\sigma^2)^2} \end{pmatrix}$$

which is negative definite at $\beta = \tilde{\beta}$ and $\sigma^2 = \tilde{\sigma}^2$.

This ensures that the likelihood function is maximized at these values.

Comparing with OLSEs, we find that

i. OLSE and m.l.e. of $\beta$ are same. So m.l.e. of $\beta$ is also an unbiased estimator of $\beta$.

ii. OLSE of $\sigma^2$ is $s^2$ which is related to m.l.e. of $\sigma^2$ as $\tilde{\sigma}^2 = \dfrac{n-k}{n} s^2$. So m.l.e. of $\sigma^2$ is a biased estimator of $\sigma^2$.

# Consistency of estimators

## (i) Consistency of *b*

Under the assumption that $\lim_{n\to\infty}\left(\frac{X'X}{n}\right)=\Delta$ exists as a nonstochastic and nonsingular matrix (with finite elements), we have

$$\begin{aligned}\lim_{n\to\infty} V(b) &= \sigma^2 \lim_{n\to\infty}\frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\\ &= \sigma^2 \lim_{n\to\infty}\frac{1}{n}\Delta^{-1}\\ &= 0.\end{aligned}$$

This implies that OLSE converges to $\beta$ in quadratic mean.

Thus OLSE is a consistent estimator of $\beta$.

This holds true for maximum likelihood estimators also.

Same conclusion can also be proved using the concept of convergence in probability.

An estimator $\hat{\theta}_n$ converges to $\theta$ in probability if

$$\lim_{n\to\infty} P\left[\left|\hat{\theta}_n - \theta\right| \geq \delta\right] = 0 \text{ for any } \delta > 0$$

and is denoted as

$$\text{plim}(\hat{\theta}_n) = \theta.$$

The consistency of OLSE can be obtained under the following weaker assumptions:

(i) $\text{plim}\left(\frac{X'X}{n}\right) = \Delta_*$ exists and is a nonsingular and nonstochastic matrix.

(ii) $\text{plim}\left(\frac{X'\varepsilon}{n}\right) = 0.$

Since

$$\begin{aligned} b - \beta &= (X'X)^{-1} X'\varepsilon \\ &= \left(\frac{X'X}{n}\right)^{-1} \frac{X'\varepsilon}{n}. \end{aligned}$$

So

$$\begin{aligned} \text{plim}(b-\beta) &= \text{plim}\left(\frac{X'X}{n}\right)^{-1} \text{plim}\left(\frac{X'\varepsilon}{n}\right) \\ &= \Delta_*^{-1}.0 \\ &= 0. \end{aligned}$$

Thus *b* is a consistent estimator of $\beta$. The same is true for maximum likelihood estimator also.

## (ii) Consistency of $s^2$

Now we look at the consistency of $s^2$ as an estimate of $\sigma^2$. We have

$$\begin{aligned} s^2 &= \frac{1}{n-k} e'e \\ &= \frac{1}{n-k} \varepsilon' \bar{H} \varepsilon \\ &= \frac{1}{n}\left(1-\frac{k}{n}\right)^{-1}\left[\varepsilon'\varepsilon - \varepsilon' X (X'X)^{-1} X'\varepsilon\right] \\ &= \left(1-\frac{k}{n}\right)^{-1}\left[\frac{\varepsilon'\varepsilon}{n} - \frac{\varepsilon' X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\right]. \end{aligned}$$

Note that $\frac{\varepsilon'\varepsilon}{n}$ consists of terms $\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$ and $\{\varepsilon_i^2, i=1,2,\dots,n\}$ is a sequence of independently and identically distributed random variables with mean $\sigma^2$. Using the law of large numbers

$$\text{plim}\left(\frac{\varepsilon'\varepsilon}{n}\right) = \sigma^2$$

$$\begin{aligned} \text{plim}\left[\frac{\varepsilon' X}{n}\left(\frac{X'X}{n}\right)^{-1}\frac{X'\varepsilon}{n}\right] &= \left(\text{plim}\frac{\varepsilon' X}{n}\right)\left[\text{plim}\left(\frac{X'X}{n}\right)^{-1}\right]\left(\text{plim}\frac{X'\varepsilon}{n}\right) \\ &= 0.\Delta_*^{-1}.0 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \Rightarrow \text{plim}(s^2) &= (1-0)^{-1}\left[\sigma^2 - 0\right] \\ &= \sigma^2. \end{aligned}$$

Thus $s^2$ is a consistent estimator of $\sigma^2$. The same holds true for maximum likelihood estimator also.

## Cramer-Rao lower bound

Let $\theta=(\beta,\sigma^2)'$. Assume that both $\beta$ and $\sigma^2$ are unknown. If $E(\hat{\theta})=\theta$, then the Cramer-Rao lower bound for $\hat{\theta}$ is greater than or equal to the matrix inverse of

$$I(\theta)=-E\left[\frac{\partial^2 \ln L(\theta)}{\partial\theta\partial\theta'}\right]$$

$$=\begin{bmatrix} -E\left[\dfrac{\partial \ln L(\beta,\sigma^2)}{\partial\beta^2}\right] & -E\left[\dfrac{\partial \ln L(\beta,\sigma^2)}{\partial\beta\partial\sigma^2}\right] \\ -E\left[\dfrac{\partial \ln L(\beta,\sigma^2)}{\partial\sigma^2\partial\beta}\right] & -E\left[\dfrac{\partial \ln L(\beta,\sigma^2)}{\partial^2(\sigma^2)^2}\right] \end{bmatrix}$$

$$=\begin{bmatrix} -E\left[-\dfrac{X'X}{\sigma^2}\right] & -E\left[\dfrac{X'(y-X\beta)}{\sigma^4}\right] \\ -E\left[\dfrac{(y-X\beta)'X}{\sigma^4}\right] & -E\left[\dfrac{n}{2\sigma^4}-\dfrac{(y-X\beta)'(y-X\beta)}{\sigma^6}\right] \end{bmatrix}$$

$$=\begin{bmatrix} \dfrac{X'X}{\sigma^2} & 0 \\ 0 & \dfrac{n}{2\sigma^4} \end{bmatrix}.$$

Then

$$[I(\theta)]^{-1}=\begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{bmatrix}$$

is the Cramer-Rao lower bound matrix of $\beta$ and $\sigma^2$.

The covariance matrix of OLSEs of $\beta$ and $\sigma^2$ is

$$\Sigma_{OLS} = \begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n-k} \end{bmatrix}$$

which means that the Cramer-Rao bound is attained for the covariance of *b* but not for $s^2$.