LINEAR REGRESSION ANALYSIS

MODULE – III

Lecture - 9

Multiple Linear Regression Analysis

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

Theorem:

i. Let $\hat{y} = Xb$ be the empirical predictor of *y*. Then $\hat{y}$ has the same value for all solutions *b* of $X'Xb = X'y$.

ii. $S(\beta)$ attains the minimum for any solution of $X'Xb = X'y$.

Proof:

i. Let *b* be any member in $b = (X'X)^{-}X'y + \left[I - (X'X)^{-}X'X\right]\omega$

Since $X(X'X)^{-}X'X = X$, so then

$$
\begin{aligned}
Xb &= X(X'X)^{-}X'y + X\left[I - (X'X)^{-}X'X\right]\omega \\
&= X(X'X)^{-}X'y
\end{aligned}
$$

which is independent of ω. This implies that $\hat{y}$ has same value for all solution *b* of $X'Xb = X'y$.

ii. Note that for any β ,

$$
\begin{aligned}
S(\beta) &= \left[y - Xb + X(b-\beta)\right]'\left[y - Xb + X(b-\beta)\right] \\
&= (y-Xb)'(y-Xb) + (b-\beta)'X'X(b-\beta) + 2(b-\beta)'X'(y-Xb) \\
&= (y-Xb)'(y-Xb) + (b-\beta)'X'X(b-\beta) \qquad \text{(Using } X'Xb = X'y\text{)} \\
&\geq (y-Xb)'(y-Xb) = S(b) \\
&= y'y - 2y'Xb + b'X'Xb \\
&= y'y - b'X'Xb \\
&= y'y - \hat{y}'\hat{y}.
\end{aligned}
$$

Fitted values

Now onwards, we assume that *X* is a full column rank matrix.

If $\hat{\beta}$ is any estimator of β for the model $y = X\beta + \varepsilon,$ then the fitted values are defined as $\hat{y} = X\hat{\beta}$ where $\hat{\beta}$ is any estimator of β .

In case of $\hat{\beta} = b,$

$$\begin{aligned} \hat{y} &= Xb \\ &= X(X'X)^{-1}X'y \\ &= Hy \end{aligned}$$

where $H = X(X'X)^{-1}X'$ is termed as **Hat Matrix** which is

i. symmetric

ii. idempotent (i.e., $HH = H$) and

iii. $tr\, H = tr\; X(X'X)^{-1}X' = tr\, X'X(X'X)^{-1} = tr\, I_k = k.$

Residuals

The difference between the observed and fitted values of study variable is called as residual. It is denoted as

$$
\begin{aligned}
e &= y \sim \hat{y} \\
&= y - \hat{y} \\
&= y - Xb \\
&= y - Hy \\
&= (I - H)y \\
&= \bar{H}y
\end{aligned}
$$

where $\bar{H} = I - H$.

Note that

(i) $\bar{H}$ is a symmetric matrix,

(ii) $\bar{H}$ is an idempotent matrix, i.e., $\bar{H}\bar{H} = (I - H)(I - H) = (I - H) = \bar{H}$ and

(iii) $tr\bar{H} = trI_n - trH = (n - k)$.

Properties of OLSE

(i) Estimation error

The estimation error of *b* is

$$
\begin{aligned}
b-\beta &= (X'X)^{-1}X'y-\beta \\
&= (X'X)^{-1}X'(X\beta+\varepsilon)-\beta \\
&= (X'X)^{-1}X'\varepsilon.
\end{aligned}
$$

(ii) Bias

Since X is assumed to be nonstochastic and $E(\varepsilon)=0$

$$
\begin{aligned}
E(b-\beta) &= (X'X)^{-1}X'E(\varepsilon) \\
&= 0.
\end{aligned}
$$

Thus OLSE is an unbiased estimator of β.

(iii) Covariance matrix

The covariance matrix of *b* is

$$
\begin{aligned}
V(b) &= E(b-\beta)(b-\beta)' \\
&= E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right] \\
&= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'IX(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}.
\end{aligned}
$$

(iv) Variance

The variance of *b* can be obtained as the sum of variances of all $b_1, b_2, \ldots, b_k$ which is the trace of covariance matrix of *b*. Thus

$$
\begin{aligned}
Var(b) &= tr\left[V(b)\right] \\
&= \sum_{i=1}^{k} E(b_i - \beta_i)^2 \\
&= \sum_{i=1}^{k} Var(b_i).
\end{aligned}
$$

Estimation of σ^2

The least squares criterion can not be used to estimate σ^2 because σ^2 does not appear in $S(\beta)$. Since $E(\varepsilon_i^2) = \sigma^2$, so we attempt with residuals e_i to estimate σ^2 as follows:

$$
\begin{aligned}
e &= y - \hat{y} \\
&= y - X(X'X)^{-1}X'y \\
&= [I - X(X'X)^{-1}X']y \\
&= \bar{H}y.
\end{aligned}
$$

Consider the residual sum of squares

$$\begin{aligned} SS_{res} &= \sum_{i=1}^{n} e_i^2 \\ &= e'e \\ &= (y - Xb)'(y - Xb) \\ &= y'(I - H)(I - H)y \\ &= y'(I - H)y \\ &= y'\bar{H}y. \end{aligned}$$

Also

$$\begin{aligned} SS_{res} &= (y - Xb)'(y - Xb) \\ &= y'y - 2b'X'y + b'X'Xb \\ &= y'y - b'X'y \quad \text{(Using } X'Xb = X'y). \end{aligned}$$

$$\begin{aligned} SS_{res} &= y'\bar{H}y \\ &= (X\beta + \varepsilon)'\bar{H}(X\beta + \varepsilon) \\ &= \varepsilon'\bar{H}\varepsilon \quad \text{(Using } \bar{H}X = 0). \end{aligned}$$

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

so $y \sim N(X\beta, \sigma^2 I)$.

Hence $y'\bar{H}y \sim \chi^2(n-k)$.

Thus $E[y'\bar{H}y] = (n-k)\sigma^2$

or $$E\left[\frac{y'\bar{H}y}{n-k}\right] = \sigma^2$$

or $$E\left[MS_{res}\right] = \sigma^2$$

where $MS_{res} = \dfrac{SS_{res}}{n-k}$ is the mean sum of squares due to residual.

Thus an unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = MS_{res} = s^2 \text{ (say),}$$

which is a model dependent estimator.

$\boxed{\text{Covariance matrix of } \hat{y}}$

The covariance Matrix of $\hat{y}$ is

$$\begin{aligned} V(\hat{y}) &= V(Xb) \\ &= XV(b)X' \\ &= \sigma^2 X(X'X)^{-1}X' \\ &= \sigma^2 H. \end{aligned}$$

Gauss-Markov theorem

The ordinary least squares estimator (OLSE) is the best linear unbiased estimator (BLUE) of β .

Proof: The OLSE of β is

$$b = (X'X)^{-1}X'y$$

which is a linear function of *y*. Consider the arbitrary linear estimator $b^* = a'y$ of linear parametric function $\ell'\beta$ where the elements of a are arbitrary constants.

Then for b^*,

$$E(b^*) = E(a'y) = a'X\beta$$

and so b^* is an unbiased estimator of $\ell'\beta$ when

$$E(b^*) = a'X\beta = \ell'\beta$$

$$\Rightarrow a'X = \ell'.$$

Since we wish to consider only those estimators that are linear and unbiased, so we restrict ourselves to those estimators for which $a'X = \ell'$.

Further

$$Var(a'y) = a'Var(y)a = \sigma^2 a'a$$

$$Var(\ell'b) = \ell'Var(b)\ell$$

$$= \sigma^2 a'X(X'X)^{-1}X'a.$$

Consider

$$\begin{aligned} Var(a'y) - Var(\ell' b) &= \sigma^2 \left[a'a - a'X(X'X)^{-1}X'a \right] \\ &= \sigma^2 a' \left[I - X(X'X)^{-1}X' \right] a \\ &= \sigma^2 a'(I - H)a. \end{aligned}$$

Since (*I* - *H*) is a positive semi-definite matrix, so

$$Var(a'y) - Var(\ell' b) \geq 0.$$

.

This reveals that if b^* is any linear unbiased estimator then its variance must be no smaller than that of *b*.

If we consider $\ell = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ (here 1 occurs at i^{th} place), then $\ell' b = b_i$ is best linear unbiased estimator of $\ell' \beta = \beta_i$ for all $i = 1,2,\ldots,k$.

Consequently *b* is the best linear unbiased estimator of β, where ‘best’ refers to the fact that *b* is efficient within the class of linear and unbiased estimators.