# LINEAR REGRESSION ANALYSIS

## MODULE – III

## Lecture - 12

# Multiple Linear Regression Analysis

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

# Testing of hypothesis

There are several important questions which can be answered through the test of hypothesis concerning the regression coefficients.

For example

1. What is the overall adequacy of the model?
2. Which specific explanatory variables seems to be important?

etc.

In order to the answer such questions, we first develop the test of hypothesis for a general framework, viz., general linear hypothesis.

Then several tests of hypothesis can be derived as its special cases.

So first we discuss the test of a general linear hypothesis.

## Test of hypothesis for $H_0 : R\beta = r$

We consider a general linear hypothesis that the parameters in $\beta$ are contained in a subspace of parameter space for which $R\beta = r,$ where *R* is a (*J* x *k*) matrix of known elements and *r* is a (*J* x 1) vector of known elements. Note that the matrix *X'X* is of full rank. In general, the null hypothesis

$$H_0 : R\beta = r$$

is termed as general linear hypothesis and

$$H_1 : R\beta \neq r$$

is the alternative hypothesis.

We assume that rank (*R*) = *J*, i.e., *R* is of full column rank, so that there is no linear dependence in the hypothesis.

Some special cases and interesting example of $H_0 : R\beta = r$ are as follows:

(i) $H_0 : \beta_i = 0$

Choose $J = 1,\ r = 0,\ R = [0, 0, ..., 0, 1, 0, ..., 0]$ where 1 occurs at the $i^{th}$ position is *R*.

This particular hypothesis explains whether $X_i$ has any effect on the linear model or not.

(ii) $H_0 : \beta_3 = \beta_4$ or $H_0 : \beta_3 - \beta_4 = 0$

Choose $J = 1,\ \ r = 0,\ \ R = [0, 0, 1, -1, 0, ..., 0]$

(iii) $H_0 : \beta_3 = \beta_4 = \beta_5$

or $H_0 : \beta_3 - \beta_4 = 0,\ \beta_3 - \beta_5 = 0$

Choose $J = 2,\ \ r = (0,\ \ 0)',\ \ R = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & \dots & 0 \end{bmatrix}.$

(iv) $H_0 : \beta_3 + 5\beta_4 = 2$

Choose $J = 1,\ \ r = 2,\ \ R = \left[0,\ \ 0,\ \ 1,\ \ 5,\ \ 0, \dots, 0\right]$

(v) $H_0 : \beta_2 = \beta_3 = \dots = \beta_k = 0$

Choose

$J = k - 1$

$r = (0,\ \ 0, \dots, 0)'$

$$R = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{(k-1)\times k} = \begin{bmatrix} 0 & \\ 0 & I_{k-1} \\ \vdots & \\ 0 & \end{bmatrix}.$$

This particular hypothesis explains the goodness of fit. It tells whether $\beta_i$ has linear effect or not and are they of any importance. It also tests whether $X_2, X_3, \dots, X_k$ have any influence in the determination of *y* or not. Here $\beta_1 = 0$ is excluded because this involves additional implication that the mean level of *y* is zero. Our main concern is to know whether the explanatory variables help in explaining the variation in *y* around its mean value or not.

We develop the likelihood ratio test for $H_0 : R\beta = r.$

## Likelihood ratio test

The likelihood ratio test statistic is

$$\lambda = \frac{\max L(\beta,\sigma^2 \mid y, X)}{\max L(\beta,\sigma^2 \mid y, X, R\beta = r)} = \frac{\hat{L}(\Omega)}{\hat{L}(\omega)}$$

where $\Omega$ denotes the whole parametric space and $\omega$ denotes the sample space.

If both the likelihoods are maximized, one constrained and the other unconstrained, then the value of the unconstrained will not be smaller than the value of the constrained. Hence $\lambda \geq 1.$

First we discus the likelihood ratio test for a simpler case when $R = I_k$ and $r = \beta_0,\ i.e.,\ \beta = \beta_0.$

This will give us better and detailed understanding for the minor details and then we generalize it for $R\beta = r$ , in general.

## Likelihood ratio test for $H_0 : \beta = \beta_0$

Let the null hypothesis related to $k \times 1$ vector $\beta$ is $H_0 : \beta = \beta_0$

where $\beta_0$ is specified by the investigator. The elements of $\beta_0$ can take on any value, including zero.

The concerned alternative hypothesis is $H_1 : \beta \neq \beta_0.$

Since $\varepsilon \sim N(0,\sigma^2 I)$ in $y = X\beta + \varepsilon,$ so $y \sim N(X\beta,\sigma^2 I).$ Thus the whole parametric space and sample space are $\Omega$ and $\omega$ respectively given by

$$\Omega : \left\{(\beta,\sigma^2) : -\infty < \beta_i < \infty,\ \sigma^2 > 0,\ i = 1, 2, ..., k\right\}$$

$$\omega : \left\{(\beta,\sigma^2) : \beta = \beta_0, \sigma^2 > 0\right\}.$$

The unconstrained likelihood under $\Omega$ is

$$L(\beta,\sigma^2 \mid y,X) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right].$$

This is maximized over $\Omega$ when

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

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

where $\tilde{\beta}$ and $\tilde{\sigma}^2$ are the maximum likelihood estimates of $\beta$ and $\sigma^2$ which are the values obtained by maximizing the likelihood function. Thus

$$\begin{aligned}
\hat{L}(\Omega) &= \max L\left(\beta,\sigma^2 \mid y,X\right) \\
&= \frac{1}{\left[\dfrac{2\pi}{n}(y - X\tilde{\beta})'(y - X\tilde{\beta})\right]^{\frac{n}{2}}} \exp\left[-\frac{(y - X\tilde{\beta})'(y - X\tilde{\beta})}{\left(\dfrac{2(y - X\tilde{\beta})'(y - X\tilde{\beta})}{n}\right)}\right] \\
&= \frac{n^{n/2}\exp\left(-\dfrac{n}{2}\right)}{(2\pi)^{n/2}\left[(y - X\tilde{\beta})'(y - X\tilde{\beta})\right]^{\frac{n}{2}}}.
\end{aligned}$$

The constrained likelihood under $\omega$ is

$$\hat{L}(\omega) = Max\, L(\beta, \sigma^2 \mid y, X, \beta = \beta_0)$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[ -\frac{1}{2\sigma^2}(y - X\beta_0)'(y - X\beta_0) \right].$$

Since $\beta_0$ is known, so the constrained likelihood function has an optimum variance estimator

$$\tilde{\sigma}_\omega^2 = \frac{1}{n}(y - X\beta_0)'(y - X\beta_0)$$

$$\hat{L}(\omega) = \frac{n^{n/2} \exp\left(-\frac{n}{2}\right)}{(2\pi)^{n/2}\left[(y - X\beta_0)'(y - X\beta_0)\right]^{n/2}}.$$

The likelihood ratio is

$$\frac{\hat{L}(\Omega)}{\hat{L}(\omega)} = \frac{\left( \dfrac{n^{n/2}\exp(-n/2)}{(2\pi)^{n/2}\left[(y - X\tilde{\beta})'(y - X\tilde{\beta})\right]^{n/2}} \right)}{\left( \dfrac{n^{n/2}\exp(-n/2)}{(2\pi)^{n/2}\left[(y - X\tilde{\beta}_0)'(y - X\tilde{\beta}_0)\right]^{n/2}} \right)}$$

$$= \left[ \frac{(y - X\beta_0)'(y - X\beta_0)}{(y - X\tilde{\beta})'(y - X\tilde{\beta})} \right]^{n/2}$$

$$= \left( \frac{\tilde{\sigma}_\omega^2}{\tilde{\sigma}^2} \right)^{n/2}$$

$$= (\lambda)^{n/2}$$

where

$$\lambda = \frac{(y - X\beta_0)'(y - X\beta_0)}{(y - X\tilde{\beta})'(y - X\tilde{\beta})}$$

is the ratio of the quadratic forms. Now we simplify the numerator of $\lambda$ as follows:

$$\begin{aligned}(y - X\beta_0)'(y - X\beta_0) &= \left[(y - X\tilde{\beta}) + X(\tilde{\beta} - \beta_0)\right]'\left[(y - X\tilde{\beta}) + X(\tilde{\beta} - \beta_0)\right] \\ &= (y - X\tilde{\beta})'(y - X\tilde{\beta}) + 2y'\left[I - X(X'X)^{-1}X'\right]X(\tilde{\beta} - \beta_0) + (\tilde{\beta} - \beta_0)'X'X(\tilde{\beta} - \beta_0) \\ &= (y - X\tilde{\beta})'(y - X\tilde{\beta}) + (\tilde{\beta} - \beta_0)'X'X(\tilde{\beta} - \beta_0).\end{aligned}$$

Thus

$$\begin{aligned}\lambda &= \frac{(y - X\tilde{\beta})'(y - X\tilde{\beta}) + (\tilde{\beta} - \beta_0)'X'X(\tilde{\beta} - \beta_0)}{(y - X\tilde{\beta})'(y - X\tilde{\beta})} \\ &= 1 + \frac{(\tilde{\beta} - \beta_0)'X'X(\tilde{\beta} - \beta_0)}{(y - X\tilde{\beta})'(y - X\tilde{\beta})}\end{aligned}$$

$$\text{or } \lambda - 1 = \lambda_0 = \frac{(\tilde{\beta} - \beta_0)'X'X(\tilde{\beta} - \beta_0)}{(y - X\tilde{\beta})'(y - X\tilde{\beta})}$$

where

$$0 \le \lambda_0 < \infty.$$

## Distribution of ratio of quadratic forms

Now we find the distribution of the quadratic forms involved in $\lambda_0$ to find the distribution of $\lambda_0$ as follows:

$$\begin{aligned}(y - X\tilde{\beta})'(y - X\tilde{\beta}) &= \tilde{e}'\tilde{e} \\ &= y'\left[I - X(X'X)^{-1}X'\right]y \\ &= y'\bar{H}y \\ &= (X\beta + \varepsilon)'\bar{H}(X\beta + \varepsilon) \\ &= \varepsilon'\bar{H}\varepsilon \qquad \text{(using } \bar{H}X = 0\text{)} \\ &= (n-k)\hat{\sigma}^2.\end{aligned}$$

**Result**: Let $Z$ is a $n$ x 1 random vector that is distributed as $N(0, \sigma^2 I_n)$ and $A$ is any symmetric idempotent $n$ x $n$ matrix of rank $p$ then $\frac{Z'AZ}{\sigma^2} \sim \chi^2(p)$. Let $B$ is another $n$ x $n$ symmetric idempotent matrix of rank $q$, then $\frac{Z'BZ}{\sigma^2} \sim \chi^2(q)$.

If $AB = 0$ then $Z'AZ$ is distributed independently of $Z'BZ$.

So using this result, we have

$$\frac{y'\bar{H}y}{\sigma^2} = \frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-k).$$

Further, if $H_0$ is true, then $\beta = \beta_0$. Substituting $\beta = \beta_0$ in this expression, we have the quantity which is same as the numerator of $\lambda_0$. The numerator of $\lambda_0$ can be rewritten in a general form for any $\beta$ as

$$\begin{aligned}(\tilde{\beta}-\beta)'X'X(\tilde{\beta}-\beta) &= \varepsilon'X(X'X)^{-1}X'X(X'X)^{-1}X'\varepsilon \\ &= \varepsilon'X(X'X)^{-1}X'\varepsilon \\ &= \varepsilon'H\varepsilon\end{aligned}$$

where *H* is an idempotent matrix with rank *k*.

Thus using this result, we have

$$\frac{\varepsilon'H\varepsilon}{\sigma^2} = \frac{\varepsilon'X'(X'X)^{-1}X'\varepsilon}{\sigma^2} \sim \chi^2(k).$$

Furthermore, the product of the quadratic form matrices in the numerator $(\varepsilon'\bar{H}\varepsilon)$ and in the denominator $(\varepsilon'H\varepsilon)$ of $\lambda_0$ is

$$\left[I - X(X'X)^{-1}X'\right]X(X'X)^{-1}X' = X(X'X)^{-1}X' - X(X'X)^{-1}X'X(X'X)^{-1}X' = 0$$

and hence the $\chi^2$ random variables in numerator and denominator of $\lambda_0$ are independent. Dividing each of the $\chi^2$ random variable by their respective degrees of freedom, we get

$$\lambda_1 = \left( \frac{\dfrac{\dfrac{(\tilde{\beta}-\beta_0)'X'X(\tilde{\beta}-\beta_0)}{\sigma^2}}{k}}{\left( \dfrac{\dfrac{(n-k)\hat{\sigma}^2}{\sigma^2}}{n-k} \right)} \right)$$

$$= \frac{(\tilde{\beta}-\beta_0)'X'X(\tilde{\beta}-\beta_0)}{k\hat{\sigma}^2}$$

$$= \frac{(y-X\beta_0)'(y-X\beta_0)-(y-X\tilde{\beta})'(y-X\tilde{\beta})}{k\hat{\sigma}^2}$$

$\sim F(k, n-k)$ under $H_0$.

Note that

$(y-X\beta_0)'(y-X\beta_0)$ : Restricted error sum of squares

$(y-X\tilde{\beta})'(y-X\tilde{\beta})$ : Unrestricted error sum of squares

Numerator of $\lambda_1$ : Difference between the restricted and unrestricted error sum of squares.

The decision rule is to reject $H_0 : \beta = \beta_0$ at $\alpha$ level of significance whenever

$$\lambda_1 \geq F_\alpha(k, n-k)$$

where $F_\alpha(k, n-k)$ is the upper critical points on the central *F*-distribution with *k* and *n - k* degrees of freedom.