Multiple Linear Regression

General Linear Regression Model

In general, the variables $X_1, \ldots, X_{p-1}$ in a regression model do not need to represent different predictor variables, as we shall shortly see. We therefore define the general linear

regression model, with normal error terms, simply in terms of X variables:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

where:

$\beta_0, \beta_1, \ldots, \beta_{p-1}$ are parameters
$X_{i1}, \ldots, X_{i,p-1}$ are known constants
ε_i are independent $N(0, \sigma^2)$
$i = 1, \ldots, n$

To express general linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

in matrix terms, we need to define the following matrices:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

$$\underset{p\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Note that the $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ vectors are the same as for simple linear regression. The $\boldsymbol{\beta}$ vector contains additional regression parameters, and the $\mathbf{X}$ matrix contains a column of 1s as well as a column of the n observations for each of the $p-1$ X variables in the regression model. The row subscript for each element X_{ik} in the $\mathbf{X}$ matrix identifies the trial or case, and the column subscript identifies the $\mathbf{X}$ variable.

In matrix terms, the general linear regression model

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}} \underset{n\times p}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}$$

where:

$\mathbf{Y}$ is a vector of responses

$\boldsymbol{\beta}$ is a vector of parameters

$\mathbf{X}$ is a matrix of constants

$\boldsymbol{\varepsilon}$ is a vector of independent normal random variables with expectation

$\mathbf{E}\{\boldsymbol{\varepsilon}\} = \mathbf{0}$ and variance-covariance matrix:

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\mathbf{I}$$

Consequently, the random vector $\mathbf{Y}$ has expectation:

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \mathbf{X}\boldsymbol{\beta}$$

and the variance-covariance matrix of $\mathbf{Y}$ is the same as that of $\boldsymbol{\varepsilon}$:

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\mathbf{Y}\}} = \sigma^2\mathbf{I}$$

Estimation of Regression Coefficients

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{bmatrix} = \hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{Y} = Xb = X\hat{\beta} = HY, \quad H = X(X'X)^{-1}X'$$

Statistical inference for the multiple linear model

Lemma:

Lemma

$$E(\hat{\beta}) = \beta$$

$$Var(\hat{\beta}) = MSE(X'X)^{-1}$$

Proof

$$E(\widehat{\beta}) = E(X'X)^{-1}X'Y = (X'X)^{-1}X'E(Y)$$
$$= (X'X)^{-1}X'XB$$
$$= I\beta$$
$$= \beta.$$

This show that the Least square estimate of β is an unbiased estimator.

$$Var(\widehat{\beta}) = Var[(X'X)^{-1}X'Y]$$
$$= (X'X)^{-1}X'Var(Y)[(X'X)^{-1}X']'$$
$$= (X'X)^{-1}X'\sigma^2[(X'X)^{-1}X']'$$
$$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$
$$= \sigma^2(X'X)^{-1}I$$
$$= \sigma^2(X'X)^{-1}$$
$$= MSE(X'X)^{-1}.$$

Example: (Dwaine Studios)

Form the estimated model, we get

$$\hat{\mathbf{Y}} = \mathbf{Xb}$$

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_{21} \end{bmatrix} = \begin{bmatrix} 1 & 68.5 & 16.7 \\ 1 & 45.2 & 16.8 \\ \vdots & \vdots & \vdots \\ 1 & 52.3 & 16.0 \end{bmatrix} \begin{bmatrix} -68.857 \\ 1.455 \\ 9.366 \end{bmatrix} = \begin{bmatrix} 187.2 \\ 154.2 \\ \vdots \\ 157.1 \end{bmatrix}$$

we find:

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$$

$$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{21} \end{bmatrix} = \begin{bmatrix} 174.4 \\ 164.4 \\ \vdots \\ 166.5 \end{bmatrix} - \begin{bmatrix} 187.2 \\ 154.2 \\ \vdots \\ 157.1 \end{bmatrix} = \begin{bmatrix} -12.8 \\ 10.2 \\ \vdots \\ 9.4 \end{bmatrix}$$

Hence,

$$MSE = \frac{SSE}{n-p} = \frac{e'e}{21-3} = \frac{2180.927}{18} = 121.1626$$

and

$$Var(\widehat{\beta}) = MSE\,(X'X)^{-1}$$

$$= 121.1626 \begin{bmatrix} 29.7289 & .0722 & -1.9926 \\ .0722 & .00037 & -.0056 \\ -1.9926 & -.0056 & .1363 \end{bmatrix}$$

$$= \begin{bmatrix} 3,602.0 & 8.748 & -241.43 \\ 8.748 & .0448 & -.679 \\ -241.43 & -.679 & 16.514 \end{bmatrix}$$

Take the square root of the diagonal we get

$$SE(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = \sqrt{3602} = 60.017$$

$$SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{0.0448} = 0.212$$

$$SE(\hat{\beta}_2) = \sqrt{Var(\hat{\beta}_2)} = \sqrt{16.514} = 4.06$$

Calculate the variance in Dwaine Studios, Inc. data.

```
# How to read txt file in R

mat <- scan('DSD.txt')

mat <- matrix(mat, ncol = 3, byrow = TRUE)

X1=mat[,1]

X2=mat[,2]
```

```
Y=mat[,3]
n=length(mat[,1])
one=as.vector(rep(1, n))
X=cbind(one,X1,X2)
b=solve(t(X)%*%X)%*%t(X)%*%Y
model=lm(Y~X1+X2)

e=model$residual
SSE=t(e)%*%e
MSE=SSE[1,1]/18
varb=MSE*solve(t(X)%*%X)
SEb0=sqrt(varb[1,1])
SEb1=sqrt(varb[2,2])
SEb2=sqrt(varb[3,3])
summary(model)
```

Confidence Interval of the multiple linear regression model coefficients

The $100(1-\alpha)\%$ confidence interval for the model coefficients can be obtained by

$$\widehat{\beta}_i \pm t_{1-\alpha/2,n-p} SE(\widehat{\beta}_i), \quad i = 0,1,2,\dots,p$$

Example: (Dwaine Studios)

Form the estimated model, calculate 95% C.Is for the model coefficients

$$Y = -68.9 + 1.46X\,1 + 9.37X\,2.$$

$$\begin{aligned}\widehat{\beta}_0 \pm t_{1-\alpha/2,n-p} SE(\widehat{\beta}_0) &= \widehat{\beta}_0 \pm t_{0.975,18} SE(\widehat{\beta}_0)\\ &= -68.9 \pm 2.1(60.017)\\ &= (-195.94,\ 56.14)\end{aligned}$$

$$\begin{aligned}\widehat{\beta}_1 \pm t_{1-\alpha/2,n-p} SE(\widehat{\beta}_1) &= \widehat{\beta}_1 \pm t_{0.975,18} SE(\widehat{\beta}_1)\\ &= 1.46 \pm 2.1(0.212)\\ &= (1.01,\ 1.91)\end{aligned}$$

$$\begin{aligned}\widehat{\beta}_2 \pm t_{1-\alpha/2,n-p} SE(\widehat{\beta}_2) &= \widehat{\beta}_2 \pm t_{0.975,18} SE(\widehat{\beta}_2)\\ &= 9.37 \pm 2.1(4.06)\\ &= (0.84,\ 17.90)\end{aligned}$$

```
confint(model,level=0.95) #CIs for all parameters
```