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

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

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

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

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

$$MSE = \frac{SSE}{n-p}$$

Hypothesis Testing

To test the coefficients of the multiple linear regression model, we follow the standard steps as: follows:

Step 1: The hypotheses

$$H_0: \beta_i = \beta_i^{(0)} \quad , \quad i = 0,1,2,....,p$$
$$H_1: \beta_i \neq (> or <)\beta_i^{(0)}$$

Step 2: The test statistic

$$T_i = \frac{\hat{\beta}_i - \beta_i^{(0)}}{S.E(\hat{\beta}_i)} \quad , \quad i = 0,1,2,....,p$$

Step 3: The Critical regions

Use the quantiles of t distribution to find the critical regions corresponding the null hypnosis $H_1: \beta_i \neq ,> or < \beta_i^{(0)}$, respectively, as

$(-\infty, -t_{1-\alpha/2,n-p}) \cup (, t_{1-\alpha/2,n-p}, \infty)$, $(t_{1-\alpha,n-p}, \infty)$ *or* $(-\infty, t_{1-\alpha,n-p})$

Step 4: The decision: Reject H0, if the calculate test statistic in step 2 belongs to the corresponding critical region.

p-value approach:
one can use p-value approach testing the hypotheses.

Remark:
Testing the significance of any of the coefficient is equivalent testing whether that coefficient is zero.

Example: (Dwaine Studios)

Test the significance of coefficients in the Dwaine Studio data (use α = 5% if it is not given).

$$Y = -68.9 + 1.46X_1 + 9.37X_2.$$

In this model, we run the test as follows:

Testing β_0

Step 1: The hypotheses

$$H_0: \beta_0 = 0$$
$$H_1: \beta_0 \neq 0$$

Step 2:

The test statistic

$$T_0 = \frac{\hat{\beta}_0 - 0}{S.E(\hat{\beta}_0)} = \frac{-68.9 - 0}{S.E(\hat{\beta}_0)} = \frac{-68.9}{60.017} = -1.147$$

Step 3:

The Critical regions

The critical region in this case is

$$(-\infty, -t_{1-\alpha/2,n-p}) \cup (, t_{1-\alpha/2,n-p}, \infty) = (-\infty, -t_{0.975,18}) \cup (t_{0.975,18}, \infty)$$
$$= (-\infty, -2.101) \cup (2.101, \infty)$$

Step 4:

The test statistic belongs to the acceptance region, then accept H0.

Testing β_1

Step 1: The hypotheses

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

Step 2:

The test statistic

$$T_1 = \frac{\hat{\beta}_1 - 0}{S.E(\hat{\beta}_1)} = \frac{1.46 - 0}{S.E(\hat{\beta}_1)} = \frac{1.46}{0.212} = 6.88$$

Step 3:

The Critical regions

The critical region in this case is

$$(-\infty, -t_{1-\alpha/2,n-p}) \cup (, t_{1-\alpha/2,n-p}, \infty) = (-\infty, -t_{0.975,18}) \cup (t_{0.975,18}, \infty)$$
$$= (-\infty, -2.101) \cup (2.101, \infty)$$

Step 4:

The test statistic belongs to the rejection region, then reject H0.

Testing β_2

Step 1: The hypotheses

$$H_0: \beta_2 = 0$$
$$H_1: \beta_2 \neq 0$$

Step 2:

The test statistic

$$T_2 = \frac{\widehat{\beta}_2 - 0}{S.E(\widehat{\beta}_2)} = \frac{9.37 - 0}{4.06} = \frac{9.37}{4.06} = 2.31$$

Step 3:

The Critical regions

The critical region in this case is

$$(-\infty, -t_{1-\alpha/2,n-p}) \cup (, t_{1-\alpha/2,n-p}, \infty) = (-\infty, -t_{0.975,18}) \cup (\ t_{0.975,18}, \infty)$$
$$= (-\infty, -2.101) \cup (\ 2.101, \infty)$$

Step 4:

The test statistic belongs to the rejection region, then reject H0.

Also, one can use p-value approach

The R-results in this example as:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-68.8571	60.0170	-1.147	0.2663	
X1	1.4546	0.2118	6.868	2e-06	***
X2	9.3655	4.0640	2.305	0.0333	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

ANOVA TEST (F-test)

Step 1: The hypotheses

$$H_0: \ \beta_1 = \beta_2 = ... = \beta_{p-1} = 0$$
$$H_1: \ \beta_i \neq \beta_j \ \ for \ i \neq j$$

Step 2:

The test statistic

$$F = \frac{MSR}{MSE}$$

Step 3:

The Critical regions The critical region in this case is

$(F_{1-\alpha,p-1,n-p}, \infty)$

Step 4:

If the test statistic belongs to the rejection region, then reject H0.

Example: (Dwaine Studios)

Test the significance of model in the Dwaine Studio data (use α = 5% if it is not given).

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

In this model, we run the test as follows:

Step 1: The hypotheses

$$H_0: \beta_1 = \beta_2 = 0$$
$$H_1: \beta_1 \neq \beta_2$$

Step 2:

The test statistic

$$SSTO = \mathbf{Y'Y} - \left(\frac{1}{n}\right)\mathbf{Y'JY} = 721{,}072.40 - 694{,}876.19 = 26{,}196.21$$

and

$$SSE = \mathbf{Y'Y} - \mathbf{b'X'Y}$$
$$= 721{,}072.40 - \begin{bmatrix} -68.857 & 1.455 & 9.366 \end{bmatrix} \begin{bmatrix} 3{,}820 \\ 249{,}643 \\ 66{,}073 \end{bmatrix}$$
$$= 721{,}072.40 - 718{,}891.47 = 2{,}180.93$$

Finally, we obtain by subtraction:

$$SSR = SSTO - SSE = 26{,}196.21 - 2{,}180.93 = 24{,}015.28$$

$$F = \frac{MSR}{MSE} = \frac{24015.28/2}{21180.93/18} = 99.1$$

Step 3:

The Critical Region: The critical region in this case is

$$(F_{1-\alpha,p-1,n-p}, \infty) = (F_{0.95,2,18}, \infty) = (3.55, \infty)$$

Step 4:

If the test statistic belongs to the rejection region, then reject H0. The model is significant.

Also, one can use p-value in such test.

R-results are:

F-statistic: 99.1 on 2 and 18 DF, p-value: 1.921e-10

Coefficient of Multiple Determination. For our example, we have

$$R^2 = \frac{SSR}{SSTO} = \frac{24{,}015.28}{26{,}196.21} = .917$$

Thus, when the two predictor variables, target population and per capita disposable income, are considered, the variation in sales is reduced by 91.7 percent.

coefficient of multiple correlation

The coefficient of multiple correlation is given by

$$R = \sqrt{R^2} = \sqrt{0.917} = 0.96$$

Estimation of Mean Response

Dwaine Studios would like to estimate expected (mean) sales in cities with target population $X_{h1} = 65.4$ thousand persons aged 16 years or younger and per capita disposable income

$X_{h2} = 17.6$ thousand dollars with a 95 percent confidence interval. We define:

$$\mathbf{X}_h = \begin{bmatrix} 1 \\ 65.4 \\ 17.6 \end{bmatrix}$$

The point estimate of mean sales is by (6.55):

$$\hat{Y}_h = \mathbf{X}_h'\mathbf{b} = [1 \quad 65.4 \quad 17.6] \begin{bmatrix} -68.857 \\ 1.455 \\ 9.366 \end{bmatrix} = 191.10$$

The estimated variance

$$s^2\{\hat{Y}_h\} = \mathbf{X}_h' s^2\{\mathbf{b}\}\mathbf{X}_h$$

$$= [1 \quad 65.4 \quad 17.6] \begin{bmatrix} 3{,}602.0 & 8.748 & -241.43 \\ 8.748 & .0448 & -.679 \\ -241.43 & -.679 & 16.514 \end{bmatrix} \begin{bmatrix} 1 \\ 65.4 \\ 17.6 \end{bmatrix}$$

$$= 7.656$$

Then

$$S.E(\widehat{Y}_h) = \sqrt{7.656} = 2.77$$

90% CI for the mean of Y is

$$\widehat{Y}_h \pm t_{(1-\alpha/2,n-p)} S.E(\widehat{Y}_h)$$

$$191.10 \pm 2.101\ (2.77)$$

$$185.3 \leq E\{Y_h\} \leq 196.9$$

Prediction Limits for New Observations

Dwaine Studios as part of a possible expansion program would like to predict sales for two new cities, with the following characteristics:

	City A	City B
X_{h1}	65.4	53.1
X_{h2}	17.6	17.7

Prediction intervals with a 90 percent family confidence coefficient are desired. Note that the two new cities have characteristics that fall well within the pattern of the 21 cities on which the regression analysis is based.

For City A, we have

A $100(1-\alpha)\%$ CI for Y_{new}

$$\widehat{Y}_{new} \pm t_{(1-\alpha/2,n-p)} S.E(\widehat{Y}_{new})$$

$$S.E(\widehat{Y}_{new}) = \sqrt{MSE + \text{var}(\widehat{Y}_h)} = \sqrt{121.626 + 7.656} = 11.35$$

In similar fashion, we obtain for city B (calculations not shown):

$$\hat{Y}_h = 174.15 \qquad s\{\text{pred}\} = 11.93$$

$$\hat{Y}_{new} \pm t_{1-\alpha/2}, n-p\, S.E(\hat{Y}_{new})$$

$$\text{City A: } 167.3 \leq Y_{h(\text{new})} \leq 214.9$$

Similarly, For City B, we have

$$\text{City B: } 149.1 \leq Y_{h(\text{new})} \leq 199.2$$

R code

```
mydata=read.table("Dwaine Studios.txt",header=TRUE)
Y=mydata$Y
X1=mydata$X1
X2=mydata$X2
n=length(X1)
model=lm(Y~X1+X2)
summary(model)
one=matrix(1,n)
X=cbind(one,X1,X2)
b=solve(t(X)%*%X)%*%t(X)%*%Y
p=3
```

```
J=matrix(1,21,21)
#  SSTO=Y'Y - 1/n(Y'JY)
SSTO=t(Y)%*%Y-1/n*(t(Y)%*%J%*%Y)
SSE=t(Y)%*%Y-t(b)%*%t(X)%*%Y
SSR=SSTO-SSE
MSR=SSR/(p-1)
MSE=SSE/(n-p)
F=MSR/MSE
RS=SSR/SSTO
R=sqrt(RS)
vb=vcov(Model)
MSE[1,1]*solve(t(X)%*%X)
Xh=c(1, 65.4, 17.6)
Yh=t(Xh)%*%b
vYh=t(Xh)%*%vb%*%Xh
vYh[1,1]
seYh=sqrt(vYh[1,1])
vYnew=MSE+vYh[1,1]
seYhnew=sqrt(vYnew)
newx = data.frame(X1=65.4, X2=17.6)
predict(Model, newx, level=0.95,interval="confidence")
predict(Model, newx, level=0.95,interval="predict")
```

```
model=lm(Y~X1+X2)
summary(model)
anova(model)
r=model$res
sum(r^2)
```