APPLIED STATISTICS

Multiple Linear Regression and Its Estimation

Dr Tao Zou

Research School of Finance, Actuarial Studies & Statistics
The Australian National University

Last Updated: Wed Aug 16 08:59:21 2017

Overview

- Introduction to Multiple Linear Regression (MLR)
- MLR Model Assumptions
- Estimation of MLR Model

References

1. **F.L. Ramsey and D.W. Schafer** (2012)
   Chapter 9 of *The Statistical Sleuth*
2. The slides are made by **R Markdown**.
   `http://rmarkdown.rstudio.com`

Multiple Linear Regression

Multiple linear regression (MLR) models the mean of the response variable as a function of several explanatory variables, namely

$$\boxed{\mu\{Y|X\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k,}\ \text{where } X = (X_1, \cdots, X_k).$$

$\beta_1\ X_1$

- $\mu\{Y|X\}$, still represents **the regression of** Y **on** $X = (X_1, \cdots, X_k)$ or equivalently **the mean of** Y **as a function of** $X = (X_1, \cdots, X_k)$.
- $\sigma\{Y|X\}$, represents the standard deviation of Y as a function of X.

The term "linear" means "linear in the regression coefficients $\beta_0, \cdots, \beta_k$".
Examples of MLR models include

$X = (X_1, X_2)$

$X_1 = Z, \quad X_2 = Z^2$

$$\mu\{Y|Z\} = \beta_0 + \beta_1 Z + \beta_2 Z^2, \text{ and } \Rightarrow \mu\{Y|X\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$\mu\{Y|Z\} = \beta_0 + \beta_1\sqrt{Z_1} + \beta_2 Z_1^2 + \beta_3 Z_2, \text{ where } Z = (Z_1, Z_2).$$

The following is not a MLR model

$$\mu\{Y|Z\} = \beta_0 + \beta_1 Z_1^{\beta_2} + \beta_3 Z_2^{\beta_4}. \qquad \times$$

MLR and Interpretation

Consider the following MLR model with two explanatory variables:

$$\mu\{Y|X_1, X_2\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$$

Marginal effect of X_1 is

$$\mu\{Y|X_1 = x_1 + 1, X_2 = x_2\} - \mu\{Y|X_1 = x_1, X_2 = x_2\} = \beta_1.$$

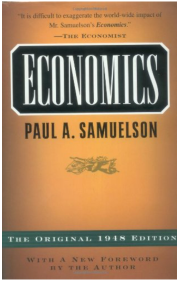
Marginal effect of X_2 is

$$\mu\{Y|X_1 = x_1, X_2 = x_2 + 1\} - \mu\{Y|X_1 = x_1, X_2 = x_2\} = \beta_2.$$

Hence, β_1 givens the increase in the mean of response for a unit increase in X_1, with X_2 held constant.

MLR and Interpretation (Con'd)

Hold other things constant! – from



In practice, we **only care about** how X_1 affects Y. However, Y can also be influenced by X_2. In this case, we should control X_2 and hold it constant (X_2 is called a control variable).

The interpretation of β_1 in MLR

$$\mu\{Y|X_1, X_2\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

realises the above idea.

MLR Model Assumptions

1. **Linearity**: The means of response fall on a linear function of the explanatory variables ($\mu\{Y|X\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$).

2. **Normality**: There is a normally distributed (sub)population of responses for given values of the explanatory variables ($X_1 = x_1, \cdots, X_k = x_k$).

3. **Constant variance**: The (sub)population standard deviations are all equal: $\sigma\{Y|X\} = \sigma$.

4. **Independence**: Observations

$$\text{Estimation} \Leftarrow \left\{ \begin{array}{ll} (X_{1,1}, \cdots X_{k,1}, Y_1), & \text{1-st of } X_1 \cdots X_k \\ \cdots & \\ (X_{1,n}, \cdots X_{k,n}, Y_n), & \text{n-th of } X_1 \cdots X_k \end{array} \right.$$

   are independent, where n is the sample size.

Remark: 2 & 3 can be described by $Y = \mu\{Y|X\} + \mathcal{E}$, where $\mathcal{E} \sim N(0, \sigma^2)$. It follows $Y \sim N(\mu\{Y|X\}, \sigma^2)$.

Estimation of SLR Parameters (Review)

Given the observations

$$(X_1, Y_1),$$
$$\vdots$$
$$(X_n, Y_n),$$

the LS estimates of β_1 and β_0 are chosen to minimise:

$$Q(b_1, b_0) = \sum_{i=1}^{n} (\underbrace{Y_i - b_0 - b_1 X_i}_{\text{Error}})^2 \Rightarrow \hat{\beta}_1, \hat{\beta}_0$$

Estimation of MLR Parameters

Given the observations

$$(X_{1,1}, \cdots X_{k,1}, Y_1),$$
$$\vdots$$
$$(X_{1,n}, \cdots X_{k,n}, Y_n),$$

$\mu(Y|X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$

the LS estimates of $\beta_0, \cdots, \beta_k$ are chosen to minimise:

Error

$$Q(b_0, \cdots, b_k) = \sum_{i=1}^{n} \{Y_i - (b_0 + b_1 X_{1,i} + \cdots + b_k X_{k,i})\}^2 .$$

Estimation of MLR Parameters

The values of $b_0, \cdots, b_k$ that minimise $Q(b_0, \cdots, b_k)$ are given by:

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{Y},$$

k×1 vector

$X^\top X$ does not have inverse ?

where the $n \times (k+1)$ matrix

$$\mathbb{X} = \begin{pmatrix} 1 & X_{1,1} & \cdots & X_{k,1} \\ 1 & X_{1,2} & \cdots & X_{k,2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1,n} & \cdots & X_{k,n} \end{pmatrix} \text{ is called design matrix, and } \mathbb{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}.$$

1st obs of $X_1 \cdots X_k$

X_1 X_k

intercept β_0

$\mu(Y|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n$

$\beta_0 \times 1$ $n \times 1$

Review of Matrix Algebra

$$A = \begin{array}{r} \\ \\ \text{Row } i \\ \\ \\ \end{array} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,q-1} & a_{1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{i,q-1} & a_{iq} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{p,q-1} & a_{pq} \end{pmatrix}$$

, is a $p \times q$ matrix.

$$B = \begin{array}{c} \begin{array}{ccccc} & & \text{Column } j & & \end{array} \\ \begin{pmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1r} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{q-1,1} & \cdots & b_{q-1,j} & \cdots & b_{q-1,r} \\ b_{q1} & \cdots & b_{qj} & \cdots & b_{qr} \end{pmatrix} \end{array}$$

, is a $q \times r$ matrix.

Matrix Multiplication

$$C = AB = \begin{array}{c} \\ \\ \text{Row } i \\ \\ \\ \end{array} \begin{array}{c} \text{Column } j \\ \begin{pmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{i1} & \cdots & c_{ij} & \cdots & c_{ir} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{p1} & \cdots & c_{pj} & \cdots & c_{pr} \end{pmatrix} \end{array}, \text{is a } p \times r \text{ matrix,}$$

$p \times \cancel{q}\ \cancel{q} \times r$

$\Downarrow$

$p \times r$

where the (i, j)-th element of matrix C is

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{i,q-1}b_{q-1,j} + a_{iq}b_{qj}.$$

Matrix Transpose and Inverse

$$
A^\top = \begin{array}{c}
\color{red}{\text{Column } i} \\
\begin{pmatrix}
a_{11} & \cdots & \color{red}{a_{i1}} & \cdots & a_{p1} \\
a_{12} & \cdots & \color{red}{a_{i2}} & \cdots & a_{p2} \\
\vdots & \vdots & \color{red}{\vdots} & \vdots & \vdots \\
a_{1,q-1} & \cdots & \color{red}{a_{i,q-1}} & \cdots & a_{p,q-1} \\
a_{1q} & \cdots & \color{red}{a_{iq}} & \cdots & a_{pq}
\end{pmatrix}
\end{array}, \text{is a } q \times p \text{ matrix.}
$$

Suppose $p = q$ and A is a $p \times p$ matrix. Then A^{-1} is defined as a $p \times p$ matrix such that

$$
A^{-1}A = AA^{-1} = \begin{pmatrix}
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 \\
0 & 0 & \cdots & 0 & 1
\end{pmatrix},
$$

which is a $p \times p$ matrix with diagonals being 1 and off-diagonals being 0 (called identity matrix).

Estimation of MLR Parameters (Con'd)

$k=1$

SLR is a special case of MLR. By employing the formula above, the LS estimates of β_0 and β_1 of SLR are

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \mathbb{Y},$$

where the $n \times \overset{(k+1)}{2}$ design matrix

$$\mathbb{X} = \begin{array}{c} \begin{array}{c} \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \end{array} \begin{array}{l} \leftarrow \text{1st obs} \\ \\ \\ \\ \end{array} \\ \begin{array}{cc} \text{intercept} & X \end{array} \end{array}, \text{ and } \mathbb{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}.$$

Recall that

$$\triangle \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \text{ and } \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X},$$

where $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$ and $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$.

One can verify that this is equal to the matrix expression.

Example: Corn Yield vs Rainfall

(example from "The Statistical Sleuth")

Six U.S. corn-producing states (Iowa, Nebraska, Illinois, Indiana, Missouri, and Ohio, 1890 - 1927)

Data from M. Ezekiel and K. A. Fox, Methods of Correlation and Regression Analysis, New York: John Wiley & Sons, 1959; originally from E. G. Misner, "Studies of the Relationship of Weather to the Production and Price of Farm Products, I. Corn" [mimeographed publication, Cornell University, March 1928].

Year	Yield	Rainfall
1890	24.5	9.6
1891	33.7	12.9
1892	27.9	9.9
1893	27.5	8.7
1894	21.7	6.8
...		
1927	32.6	10.4

Yield: the yield of corn per unit area of land cultivation (bushels per acre).

Rainfall: precipitation (inches).

Example: Corn Yield vs Rainfall (Con'd)

/L/0

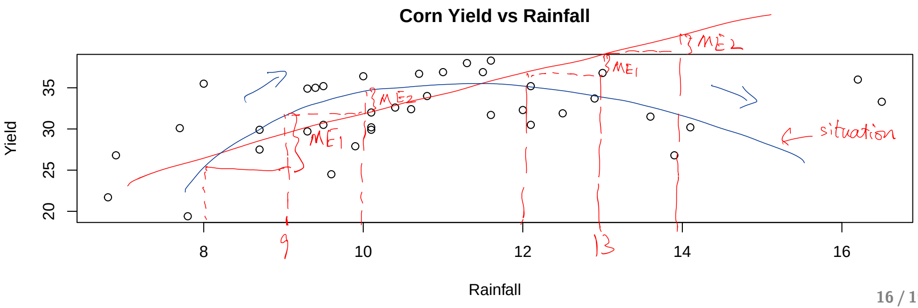
R Code

```
rm(list=ls())
setwd('~/Desktop/Research/AppliedStat2017/L4')
#install.packages('Sleuth3')
library(Sleuth3)
head(ex0915)
```

data.frame

```
##   Year Yield Rainfall
## 1 1890  24.5      9.6
## 2 1891  33.7     12.9
## 3 1892  27.9      9.9
## 4 1893  27.5      8.7
## 5 1894  21.7      6.8
## 6 1895  31.9     12.5
```

```
y=ex0915$Yield
z=ex0915$Rainfall
plot(z,y,ylab="Yield",xlab="Rainfall",main="Corn Yield vs Rainfall")
```



Example: Corn Yield vs Rainfall (Con'd)

In this example a straight line regression model is inappropriate. See Lecture Notes 3.

One model for incorporating curvature is

$$\mu\{\text{Yield}|\text{Rainfall}\} = \beta_0 + \beta_1 \underset{z}{\text{Rainfall}} + \beta_2 \underset{z^2}{\text{Rainfall}^2}.$$

The model incorporates curvature by allowing the (marginal) effect of rainfall to be different at different levels of rainfall.

Marginal effect of Rainfall is

$$\mu\{\text{Yield}|\text{Rainfall}+1\} - \mu\{\text{Yield}|\text{Rainfall}\} = \beta_1 + \beta_2(2\text{Rainfall}+1),$$

which depends on the rainfall level.

vs.

β_1

Example: Corn Yield vs Rainfall (Con'd)

```
fitm<-lm(y~z+I(z^2))
summary(fitm)
```

Response - y   Rainfall - z

z^2

```
##
## Call:
## lm(formula = y ~ z + I(z^2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.4642 -2.3236 -0.1265  3.5151  7.1597
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.01467   11.44158  -0.438  0.66387
## z            6.00428    2.03895   2.945  0.00571 **
## I(z^2)      -0.22936    0.08864  -2.588  0.01397 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.763 on 35 degrees of freedom
## Multiple R-squared:  0.2967, Adjusted R-squared:  0.2565
## F-statistic: 7.382 on 2 and 35 DF,  p-value: 0.002115
```

$\hat\beta_0$, $\hat\beta_1$, $\hat\beta_2$

$\rightarrow H_0: \beta_2 = 0 \leftrightarrow H_a: \beta_2 \neq 0$

$\beta_2 \neq 0$

design matrix

```
X=cbind(1,z,z^2)
Y=y
solve(t(X)%*%X)%*%t(X)%*%Y
```

$(X^T X)^{-1} X^T Y = \begin{pmatrix} \hat\beta_0 \\ \hat\beta_1 \\ \hat\beta_2 \end{pmatrix}$

```
##          [,1]
##    -5.0146670
## z   6.0042835
##    -0.2293639
```

(1) solve (A) returns A^{-1} — inverse

(2) A %*% B $\rightarrow$ AB

(3) t(A) $\rightarrow$ A^T

Example: Corn Yield vs Rainfall (Con'd)

```
plot(z,y,ylab="Yield",xlab="Rainfall",main="Corn Yield vs Rainfall")
points(z,fitm$fitted,pch=3,col='red')
```

x-axis