Chapter 6 Multiple Regression I

6.1 Multiple Regression Models

1. **Need for Several Predictor Variables** A single predictor is often inadequate since often multiple variables affect the response variable in important and distinctive ways, this is especially true for observational experiments where predictor variables are not controlled. Even in controlled experiments, we often want to investigate a number of predictor variables simultaneously

2. **First-Order Model with Two Predictor Variables** A first order model is linear in predictor variables. A first-order model with 2 predictor variable is given by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

with regression function as

$$\mathbb{E}\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

A regression function in multiple regression is called a **regression surface** or a **response surface**. Parameter β_1 indicates the change in mean response $\mathbb{E}\{Y\}$ per unit increase in X_1 when X_2 is held constant. Hence, effect of X_1 on mean does not depend on levels of X_2, and vice versa, the predictor variables are said to have **additive effects**or **not to interact**. Paramters β_1 and β_2 are called **partial regression coefficients** since they reflect on partial effect of one predictor variable when other predictor variables is included in the model and is held constant.

$$\frac{\partial \mathbb{E}\{Y\}}{\partial X_1} = \beta_1 \qquad \frac{\partial \mathbb{E}\{Y\}}{\partial X_2} = \beta_2$$

3. **First-order model with More Than Two Predictor Variables** Consider $p-1$ predictor variables $X_1, \cdots, X_{p-1}$. A first-order model with $p-1$ predictor variables is given by

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

or equivalently

$$Y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k X_{ik} + \epsilon_i$$

Given $\mathbb{E}\{\epsilon_i\} = 0$, the regression function is given by

$$\mathbb{E}\{Y\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1}$$

which represents a **hyperplane**. Parameter β_k indicates change in the mean response $\mathbb{E}\{Y\}$ with a unit increase in predictor variable X_k, when all other predictor variables in the regression model are held constant

4. **General Linear Regression Model** A general linear model, with normal error terms, is given by

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

where

(a) $\beta_0, \beta_1, \cdots, \beta_{p-1}$ are parameters
(b) $X_{i1}, \cdots, X_{i,p-1}$ are known constants
(c) ϵ_i are independent $\mathcal{N}(0, \sigma^2)$

with response function given by

$$\mathbb{E}\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1} \qquad (\mathbb{E}\{\epsilon_i\} = 0)$$

The general linear model usage cases

(a) $p-1$ **Predictor Variables** as seen with first-order model with $p-1$ predictor variables, which has no interaction effects
(b) **Qualitative Predictor Variables** Indicator variables as X_is. As an example of 1 quantitative predictor X_{i1} and 1 indicator predictor variable X_{i2}

$$\mathbb{E}\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

We have 2 response function

$$\mathbb{E}\{Y\} = \beta_0 + \beta_1 X_1 \qquad \mathbb{E}\{Y\} = (\beta_0 + \beta_2) + \beta_1 X_1$$

which represents parallel straight lines with different intercepts. Generally we represents qualitative variable with c classes by means of $c-1$ indicator variables
(c) **Polynomial Regression** is a special case of generalized linear model contaiing higher-order terms of predictor variable, making the response function curlinear. For example

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

where $X_{i1} = X_i$ and $X_{i2} = X_i^2$
(d) **Transformed Variables**
(e) **Interaction Effects** When effects of predictor variables on response variable are not additive, the effect of one predictor variable depends on levels of other predictor variables. A **nonadditive regressfion model with 2 predictors X_1 and X_2** is given by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i$$

which is a special case of generalized linear model whereby $X_{i3} = X_{i1} X_{i2}$

(f) **Combination of above cases** For example, model that is curlinear with interaction terms.

(g) **Meaning of linear in General Linear Regression Model** Note GLM does is not restricted to linear response surfaces. **Linear Model** refers to the fact that model is linear in the parameters; it does not refer to the shape of the response surface.

$$Y_i = \beta_0 exp(\beta_1 X_i) + \epsilon$$

is a nonlinear regression model

6.2-6.6 General Linear Regression Model in Matrix Terms

1. **GLR Model in Matrix**
Given

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix} \quad \underset{p\times 1}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \quad \underset{n\times 1}{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

The model is given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where

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

So the expectation of response given by

$$\mathbb{E}\{Y\} = \mathbf{X}\boldsymbol{\beta} \qquad \boldsymbol{\sigma^2}\{\mathbf{Y}\} = \sigma^2 \mathbf{I}$$

2. **Estimation of Regression Coefficients**
Idea is to minimize least squared

$$Q = \sum_i^n (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{i,p-1})^2$$

minimize Q to get normal equation

$$\mathbf{X'X}\hat{\boldsymbol{\beta}} = \mathbf{X'Y}$$

yielding least squared estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

3. **Fitted Values and Residuals**

$$
\begin{aligned}
\hat{\mathbf{Y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} \\
\hat{\mathbf{e}} &= (\mathbf{I} - \mathbf{H})\mathbf{Y} && \text{(where } \mathbf{H} = X(X'X)^{-1}X'\text{)} \\
\mathbb{E}\{\hat{\mathbf{e}}\} &= 0 \\
\boldsymbol{\sigma}^{\mathbf{2}}\{\hat{\mathbf{e}}\} &= \sigma^2(\mathbf{I} - \mathbf{H}) \\
\mathbf{s}^2\{\hat{\mathbf{e}}\} &= MSE(\mathbf{I} - \mathbf{H})
\end{aligned}
$$

4. **Sum of Squares and Mean Squares**
Same as before

$$
\begin{aligned}
SST &= \mathbf{Y}'\left[\mathbf{I} - (\frac{1}{n})\mathbf{J}\right]\mathbf{Y} \\
RSS &= \mathbf{Y}'(\mathbf{1} - \mathbf{H})\mathbf{Y} \\
SSReg &= \mathbf{Y}'\left[\mathbf{H} - (\frac{1}{n})\mathbf{J}\right]\mathbf{Y}
\end{aligned}
$$

with

(a) SST has $n - 1$ degrees of freedom
(b) RSS has $n - p$ degrees of freedom
(c) SSReg has $p - 1$ degrees of freedom

So the mean square is given by

$$MSE = \frac{RSS}{n-p} \qquad MSReg = \frac{SSReg}{p-1}$$

TABLE 6.1 ANOVA Table for General Linear Regression Model (6.19).

Source of Variation	SS	df	MS
Regression	$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$p-1$	$MSR = \frac{SSR}{p-1}$
Error	$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$	$n-p$	$MSE = \frac{SSE}{n-p}$
Total	$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$n-1$	

5. **F-test for Regression Relation** tests for if there is a regression relation between the response variable Y and the set of X variable $X_1, \cdots, X_{p-1}$

$$\begin{cases} \mathrm{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0 \\ \mathrm{H}_\alpha : \text{ not all } \beta_k \text{ equal zero } k = 1, \cdots, p-1 \end{cases}$$

With test statistic

$$F* = \frac{MSReg}{MSE}$$

with

$$\begin{cases} \text{if } F^* \leq F_{1-\alpha, p-1, n-p} & \text{conclude } H_0 \\ \text{if } F^* > F_{1-\alpha, p-1, n-p} & \text{conclude } H_\alpha \end{cases}$$

When $p - 1 = 1$, test reduces to F test in SLR for if $\beta_1 = 0$

6. **Coefficients of Multiple Determination**

$$R^2 = \frac{SSReg}{SST} = 1 - \frac{RSS}{SST}$$

which measures the proportionate reduction of total variation in Y associated with the use of the set of X variables.

   (a) R^2 assumes value of 0 when $\hat{\beta}_k = 0$ for $k = 1, \cdots, p-1$ and value 1 when all Y observation fall directly on the fitted regression line
   (b) Adding more X variables can only increase R^2 and never reduce it, because RSS can never become larger with more X variables and SST is always the same for the given set of responses.
   (c) **Adjusted coefficient of multiple determination** R_a^2 adjusts R^2 by dividing each sum of squares by its degree of freedom

$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{RSS}{SST}$$

   R_a^2 may become smaller when another X variable is introduced to the model. Since any decrease in RSS maybe more than offset by the loss of a degree of freedom in the denominator $n - p$
   (d) A large value of R^2 does not imply the fitted model is a useful one. For example, despite high R^2, fittedf model may not be useful if most predictions require extrapolation outside region of observations. Or if MSE may still be too large for inferences to be useful when high precision needed

7. **Coefficient of Multiple Correlation**

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

8. **Inferences about Regression Parameters**
   Least squared and MLE estimators in $\hat{\beta}$ are unbiased

$$\mathbb{E}\{\hat{\boldsymbol{\beta}}\} = \boldsymbol{\beta}$$

$$\boldsymbol{\sigma^2}\{\hat{\boldsymbol{\beta}}\} = \sigma^2 \underset{p \times p}{(\mathbf{X}'\mathbf{X})^{-1}}$$

$$\mathbf{s}^2\{\hat{\boldsymbol{\beta}}\} = MSE \underset{p \times p}{(\mathbf{X}'\mathbf{X})^{-1}}$$

6.2-6.6 Diagnostics and Remedial Measures

1. **Scatter Plot Matrix**