

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**. Parameters 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.

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

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

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

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

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

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

with response function given by

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

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_i s. 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

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

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 containing higher-order terms of predictor variable, making the response function curilinear. 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 regression 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 curilinear 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

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

The model is given by

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

where

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

So the expectation of response given by

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

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

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

yielding least squared estimator

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

3. Fitted Values and Residuals

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

4. Sum of Squares and Mean Squares

Same as before

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

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

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

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, \dots, X_{p-1}

$$\begin{cases} H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0 \\ H_\alpha : \text{not all } \beta_k \text{ equal zero } k = 1, \dots, p-1 \end{cases}$$

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 $k = 1, \dots, p-1$ 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 , fitted 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

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

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

$p \times p$

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

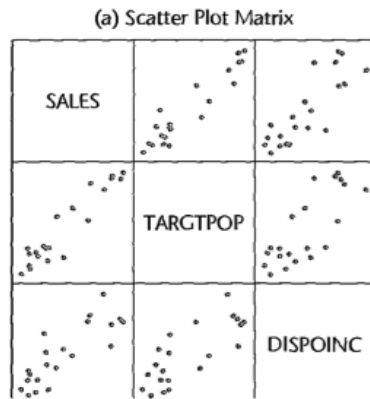
$p \times p$

6.8 Diagnostics and Remedial Measures

1. Scatter Plot Matrix

Y variable is found in its row, and X variable is name found in the column

FIGURE 6.4
SYGRAPH
Scatter Plot
Matrix and
Correlation
Matrix—
Dwaine Studios
Example.



(b) Correlation Matrix

| | SALES | TARGTOP | DISPOINC |
|----------|-------|---------|----------|
| SALES | 1.000 | .945 | .836 |
| TARGTOP | | 1.000 | .781 |
| DISPOINC | | | 1.000 |

First column we can compare plots of different predictor variables X_i with Y .

2. Correlation Matrix

contains coefficients of simple correlation $r_{Y,1}, r_{Y,2}, \dots, r_{Y,p-1}$ between Y and each of the predictor variables, and coefficients of simple correlation among predictor variables $r_{1,2}$ between X_1 and X_2

$$\begin{bmatrix} 1 & r_{Y,1} & r_{Y,2} & \cdots & r_{Y,p-1} \\ r_{Y,1} & 1 & r_{1,2} & \cdots & r_{1,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{Y,p-1} & r_{1,p-2} & r_{2,p-1} & \cdots & 1 \end{bmatrix}$$

which is symmetric with main diagonals containing 1s. Usually the first column/row is the response variable Y

3. Residual Plots

- (a) Good for investigating constancy of variance of error terms, as well as for providing information about outliers.
- (b) Residuals should be plotted against each of the predictor variables, which provide information about adequacy of the regression function w.r.t. that predictor variable.
- (c) Should try plotting against important predictor variables omitted from the model.

4. Remedial Measures

Add a variable to account a curvature effect X^2 or interaction between X_1 and X_2 as X_1X_2 . Box-Cox transformation used to determine an appropriate power transformation on Y for MLR.

10.1 Added-Variable Plots

1. **Motivation** Normal residual plot may not properly show nature of marginal effect of a predictor variable, given other predictor variables in the model. The added variable plot provides graphical information about the marginal importance of a predictor variable X_k , given the other predictor variables already in the model.
2. **Idea** Both response variable Y and predictor variable X_k under consideration are regressed against other predictor variables in the regression model and the residuals are obtained for each.

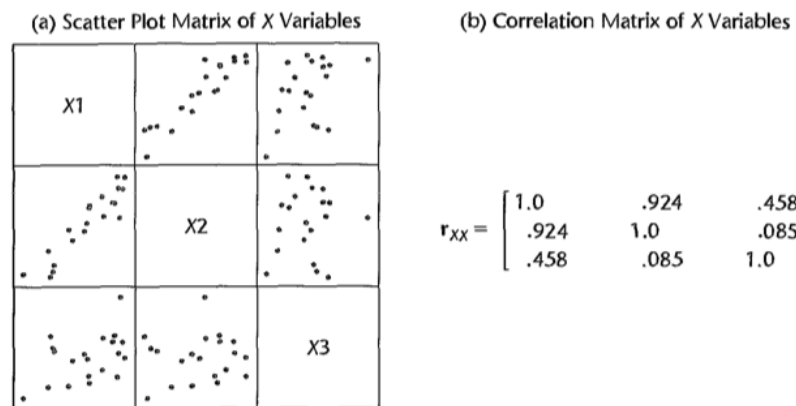
22 Multicollinearity and Its Effects

1. **Motivation** In nonexperimental situations, Predictor variables tend to be correlated among themselves and with other variables that are related to the response variable but are not included in the model
2. **Multicollinearity** occurs when predictor variables are correlated among themselves.
3. **Uncorrelated Predictor Variables** predictor variable X_1 and X_2 are uncorrelated if coefficient of simple determination between the two are small, i.e. $r_{12}^2 \sim 0$. When predictors are uncorrelated, their regression coefficients stay approximately the same regardless of any other uncorrelated predictors are in the model or not. Also, the marginal contribution of one predictor variable in reducing the error sum of squares when the other predictor variables are in the model is exactly the same as when this predictor variable is in the model alone.
4. **Problem with perfectly correlated predictor variables** If X_1 and X_2 are perfectly related, many different response function will lead to perfectly fitted values for the observation, and to the fitted values for any other (X_1, X_2) combinations following the relation between X_1 and X_2 . The response functions are not the same and will lead to different fitted values for (X_1, X_2) combinations that do not follow the relation between X_1 and X_2

- (a) perfect correlation does not inhibit our ability to find a good fit to data, or inferences about the mean responses or predictions of new observations
- (b) However, since many different response functions provide the same good fit, we cannot interpret any one set of regression coefficients as reflecting the effects of different predictor variables. This results in estimated regression coefficients having large sampling variability. We get different estimates of regression coefficients from one sample to the next.
- (c) We can no longer interpret a regression coefficients as measuring the change in the expected value of the response variable when the given predictor variable is increased by one unit while all other predictors are held constant.

5. **Effects of Multicollinearity** Given scatter plot matrix and the correlation matrix of

FIGURE 7.3
Scatter Plot
Matrix and
Correlation
Matrix of the
Predictor
Variables—
Body Fat
Example.



Effects on Regression Coefficients. Note from Table 7.2³ that the regression for X_1 , triceps skinfold thickness, varies markedly depending on which other included in the model:

| Variables in Model | b_1 | b_2 |
|--------------------|-------|--------|
| X_1 | .8572 | — |
| X_2 | — | .8565 |
| X_1, X_2 | .2224 | .6594 |
| X_1, X_2, X_3 | 4.334 | −2.857 |

the predictor variables.

Note X_1 and X_2 are highly correlated $r_{12} = 0.924$, X_3 is not highly correlated with X_1 and X_2 ($r_{13} = 0.458$, $r_{23} = 0.085$) but X_3 is highly correlated with X_1 and X_2 together by looking at coefficients of multiple determination. Note $\hat{\beta}_2$ changes sign when X_3 is added to the model that includes X_1 and X_2 .

- (a) **regression coefficients** Idea is when predictors are correlated, the regression coefficients of any one variable depends on which other predictors are included in the model and which one are left out. A regression coefficients does not reflect

any inherent effect of the particular predictor variable on the response variable but only a marginal or partial effect.

- (b) **sum of squares** marginal contribution of any one predictor variable in reducing error sum of squares varies, depending on which other variables are in the model.

22 Interaction Regression Models

22 Analysis of Covariance

1. **ANCOVA** A special type of regression model, used for observational/designed studies. Idea is to augment ANOVA model. It is intended reduce variance of error terms in the model
2. **How ANCOVA reduce error variability** ANCOVA utilizes relationship between response variable and one or more quantitative variable for which the observations are available in order to reduce error term variability and make the study a more powerful one for comparing treatment effects.
3. **Concomitant (covariates) variable** Quantitative variable in covariance analysis is called concomitant variable. Concomitant variable usually has relationship with response variable. For example, age, status, aptitude in studying human subjects, etc. A concomitant variable should be observed before the study, or not be influenced by treatment in any way, if observed during the study, i.e. age.
4. When qualitative concomitant variable added, analysis of variance model where some factors are of primary interest and others represent concomitant variables are included for purpose of error variance reduction

Chapter 5 Multiple Linear Regression

Chapter 5.2 Estimation and Inference in MLR

1. Estimation and Inference in MLR

$$\mathbb{E}\{Y|X_1 = x_1, X_2 = x_2, \dots, X_p = x_p\} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

so

$$Y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + e_i$$

where $\mathbb{E}\{e_i|X\} = 0$, in which case response variable Y is predicted from p predictor variables X_1, X_2, \dots, X_p

2. Least Squared matrix formation Regression model

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

with $\text{var}(e) = \sigma^2 \mathbf{I}$

$$RSS(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Least square estimate given by

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

Fitted value given by

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

3. Least Squared Estimates properties

$$\mathbb{E}\{\hat{\boldsymbol{\beta}}|\mathbf{X}\} = \boldsymbol{\beta} \quad \text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

4. Residual Sum of Squares

$$RSS = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \hat{\mathbf{e}}'\hat{\mathbf{e}} = \sum_i \hat{e}_i^2$$

Error variance estimated by

$$s^2 = \frac{RSS}{n - p - 1} = \frac{1}{n - p - 1} \sum_i \hat{e}_i^2$$

is an unbiased estimator of σ^2

5. Confidence interval and tests of significance Assume normally distributed error with constant variance, it can be shown that for $i = 0, 1, \dots, p$,

$$T_i = \frac{\hat{\beta}_i - \beta_i}{se(\hat{\beta})} \sim t_{n-p-1}$$

degree of freedom is difference between sample size and number of parameters estimated, i.e. $p + 1$ for parameters $\beta_0, \beta_1, \dots, \beta_p$

6. Adjusted R^2

$$R_{adj}^2 = 1 - \frac{RSS/(n-p-1)}{SST/(n-1)}$$

7. ANOVA to test linear association

$$\begin{cases} \mathcal{H}_0 : \beta_1 = \dots = \beta_p = 0 \\ \mathcal{H}_\alpha : \text{at least some of } \beta_i \neq 0 \end{cases}$$

$$F = \frac{SS_{reg}/p}{RSS/(n-p-1)}$$

F has F distribution with p and $n-p-1$ degrees of freedom when H_0 is true. F-test used to test for existence of a linear association between Y and **ANY** of the p x variable. If F-test is indeed significant then a natural question is to ask which of the p x-variables is there evidence of a linear association with Y . We perform p separate t-test $\mathcal{H}_0 : \beta_1 = 0$. However, there are problems interpreting results when predictors are highly correlated

8. Partial F-test

Test if a subset of predictors have regression coefficients equal to 0.

$$\begin{cases} \mathcal{H}_0 : \beta_1 = \dots = \beta_k = 0 \text{ where } k < p & \text{reduced model } Y = \beta_0 + \beta_{k+1}x_{k+1} + \dots + \beta_px_p + e \\ \mathcal{H}_\alpha : \mathcal{H}_0 \text{ not true} & \text{full model } Y = \beta_0 + \beta_1x_1 + \dots + \beta_px_p + e \end{cases}$$

$$F = \frac{(RSS_{reduced} - RSS_{full})/(df_{reduced} - df_{full})}{RSS_{full}/df_{full}} = \frac{(RSS_{reduced} - RSS_{full})/k}{RSS_{full}/(n-p-1)}$$

Note, the reduced model includes all variable currently of interest except whose statistical significance we wish to test. Intuitively, we look at the average difference in RSS between 2 models and compare this to average predictor error in the best model we have (the full model). If error are roughly the same for two models, then the added regressors in the full model are not contributing a lot. But if average decrease in prediction error due to added regressor is quite large, then the regressors are contributing a good deal of explanatory power and thus should be retained.

Chapter 5.3 Analysis of Covariance

1. **Analysis of Covariance** Naively, ANOVA model has predictor variables that is quantitative (x) and qualitative (d).
2. **Coincident regression lines** The dummy variable x has no effect on Y

$$Y = \beta_0 + \beta_1x + e$$

3. **Parallel regression lines** Dummy variable produces an additive change in Y

$$Y = \beta_0 + \beta_1 x + \beta_2 d + e = \begin{cases} \beta_0 + \beta_1 x + e & d = 0 \\ \beta_0 + \beta_2 + \beta_1 x + e & d = 1 \end{cases}$$

4. **Regression line with equal intercepts but different slopes** Dummy variable only changes the size of the effect of x on Y , that is,

$$Y = \beta_0 + \beta_1 x + \beta_3 d \times x + e = \begin{cases} \beta_0 + \beta_1 x + e & d = 0 \\ \beta_0 + (\beta_1 + \beta_3)x + e & d = 1 \end{cases}$$

5. **Unrelated regression lines** Dummy variable produces an additive change in Y and also changes the size of effect of x on Y

$$Y = \beta_0 + \beta_1 x + \beta_2 d + \beta_3 d \times x + e = \begin{cases} \beta_0 + \beta_1 x + e & d = 0 \\ \beta_0 + \beta_2 + (\beta_1 + \beta_3)x + e & d = 1 \end{cases}$$

Chapter 6 Diagnostics and Transformation for MLR

Chapter 6.1 Regression Diagnostics for Multiple Regression

1. Determine if proposed regression model is valid by looking at **standardized residuals** plot
2. Determine if any data points are **leverage point** (with large effect on model)
3. Determine if any data points are **outliers**
4. Assess effect of each predictor variable on the response variable, having adjusted for the effect of other predictors using **added variable plots**
5. Assess extent of **collinearity** among predictor variables using **variance inflation factors**
6. Examine if assumption of constant error variance is valid,
7. For time series data, examine if data are correlated over time

1. Leverage Points

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

To classify i -th point as point of high leverage, in MLR model with p predictors if

$$h_{ii} > 2 \times \text{avg}(h_{ii}) = 2 \times \frac{(p+1)}{n}$$

2. Standardized residuals

$$\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$\mathbb{E}\{\hat{\mathbf{e}}|\mathbf{X}\} = 0$$

$$\text{Var}\{\hat{\mathbf{e}}|\mathbf{X}\} = \sigma^2(\mathbf{I} - \mathbf{H})$$

The i -th residual has variance given by

$$\text{Var}(\hat{e}_i) = \sigma^2 [1 - h_{ii}]$$

The i -th standardized residual is given by

$$r_i = \frac{\hat{e}_i}{s\sqrt{1 - h_{ii}}} \quad s = \sqrt{\frac{\sum_j \hat{e}_j^2}{n - (p + 1)}} \text{ estimates } \sigma$$

3. **Outlier** Points are outliers if the point falls outside interval from -2 to 2 for small datasets.
4. **Validity of Model** A MLR model is a valid model for the data if the conditional mean of Y given X is a linear function of X and the conditional variance of Y given X is constant

$$\mathbb{E}\{Y|X = x\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad \text{Var}\{Y|X = x\} = \sigma^2$$

When a valid model has been fit, a plot of standardized residuals, r_i against any predictor or any linear combination of predictors (i.e. response variable) will have properties

- (a) random scatter of points around horizontal axis, since mean function of e_i is zero when a corrected model has been fit
- (b) Constant variability as we look along the horizontal axis

So any pattern in a plot of standardized residual is indicative that an invalid model has been fit to the data.

5. **Added variable plot** used to visually assess the effect of each predictor, having adjusted for the effects of the other predictors. It is simply a plot showing residual of response except X_i against all predictors except X_i vs. residual of X_i on all other predictors. This is mainly to overcome the shortcoming of plotting scatter plots of response variable against each of the predictor variable, which does not take into account the effect of the other predictor variable in the model where predictor variables are correlated (other variables are not held constant).
 - (a) Compute residuals of regressing response against predictor variables omitting X_i

- (b) Compute residuals of regressing X_i against the remaining independent variables
- (c) Plot residuals from 1 against residuals from 2

It is used to identify outliers and heteroscedasticity. The slope in the plot is the partial regression coefficients. A linear pattern indicates that X_i is useful in the model over and above other predictor variables, i.e. plot shows strengths of linear relationship Y and X_i over and above other variables. It is also useful for detecting nonlinear relationships, nonconstant variance, and influential points

6. **Partial correlation** measures degree of association between 2 random variables, with effect of the set of controlling random variables removed. Formally, partial correlation between X and Y given a set of n controlling variables $\mathbf{Z} = \{Z_1, \dots, Z_n\}$ is the correlation between residual e_X and e_Y resulting from the linear regression of X with \mathbf{Z} and of Y with \mathbf{Z}

Chapter 6.4 Multicollinearity

1. **Scenario** highly correlated data yields highly statistically significant F-test but very few of the regression coefficients may be statistically significant
2. **Variance Inflation Factors** quantifies severity of multicollinearity. It provides an index that measures how much the variance (the square of the estimate's standard deviation) of an estimated regression coefficient is increased because of collinearity. It can be shown that

$$Var\{\hat{\beta}_j\} = \frac{1}{1 - R_j^2} \times \frac{\sigma^2}{(n-1)s_{s_j}^2} \quad j = 1, \dots, p$$

variance inflation factor is given by

$$VIF = \frac{1}{1 - R_j^2}$$

where R_j^2 denote value of coefficients of multiple correlation obtained from regression of x_j on the other x 's, i.e. amount of variability explained by this regression.

3. **Ridged Regression**

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

4. **Principle Component analysis**

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

where \mathbf{Z} is a lower dimension version of X