

Outline Stat 332

Regression Analysis

Instructor: Dr. Mahmoud Ibrahim

Textbook:

Applied Linear Regression Models, Fifth Edition by Kutner, Nachtsheim and Neter

كتاب مترجم للطبعة الرابعة

نماذج إحصائية خطية تطبيقية (الجزء الأول)

المؤلف: نيتير واخرون. ترجمة: د. انيس كنجو – د. عبد الحميد الزيد – د. الحسيني عبد البر

Course Scope and Contents:

This course is an introduction to applied data analysis. We will explore data sets, examine various models for the data, assess the validity of their assumptions, and determine which conclusions we can make (if any). Data analysis is a bit of an art; there may be several valid approaches. We will strongly emphasize the importance of critical thinking about the data and the question of interest. Our overall goal is to use a basic set of modeling tools to explore and analyze data and to present the results in a scientific report. We then consider simple linear regression, a model that uses only one predictor. After briefly reviewing some linear algebra, we turn to multiple linear regression, a model that uses multiple variables to predict the response of interest. For all models, we will examine the underlying assumptions. More specifically, do the data support the assumptions? Do they contradict them? What are the consequences for inference? Also, we will explore some nonlinear models and data transformations. Finally, we discuss Linear regression based on the categorical with some applications

Course Calendar

Week	Date	Topics Covered
1	30/8/2020	Introduction and some basic concepts of probability and statistics
2	6/9/2020	Definition of the Simple linear regression model with some applications
3	13/9/2020	Estimation of the unknown parameters of the simple linear regression model
4	20/9/2020	Properties of the least square method
5	27/9/2020	Confidence estimation of the least square estimated of the coefficient of simple linear regression model.
6	4/10/2020	Hypotheses Testing of the simple linear regression model
7	11/10/2020	The efficiency of the simple linear regression model by using ANOVA
8	18/10/2020	Predication and residual analysis of the simple linear regression model
9	25/10/2020	Like-of-fit test
10	1/11/2020	Multiple linear regression model
11	8/11/2020	Estimation of the unknown parameters of the multiple linear regression model.
12	15/11/2020	Hypothesis testing of the multiple linear regression model
13	22/11/2020	Prediction and residual analysis of the multiple linear regression model
14	29/11/2020	Stepwise regression
15	6/12/2020	Linear regression based on the categorical with some application

Stat 332
Regression Analysis

Chapter 1
**Linear Regression with One
Predictor Variable**

- Regression analysis is a statistical methodology that utilizes the relation between two or more quantitative variables so that a response or outcome variable can be predicted from the other, or others.
- This methodology is widely used in business, the social and behavioral sciences, the biological sciences, and many other disciplines.

A few examples of applications are:

- 1- Sales of a product can be predicted by utilizing the relationship between sales and amount of advertising expenditures.
- 2- The performance of an employee on a job can be predicted by utilizing the relationship between performance and a battery of aptitude tests.
- 3- The size of the vocabulary of a child can be predicted by utilizing the relationship between size of vocabulary and age of the child and amount of education of the parents.
- 4- The length of hospital stay of a surgical patient can be predicted by utilizing the relationship between the time in the hospital and the severity of the operation.

➤ Relations between Variables

Functional Relation between Two Variables (mathematical relations)

Example (1-1)

$$Y=2X$$

<i>X</i> Units	<i>Y (Dollars)</i> Sales
75	150
25	50
130	260

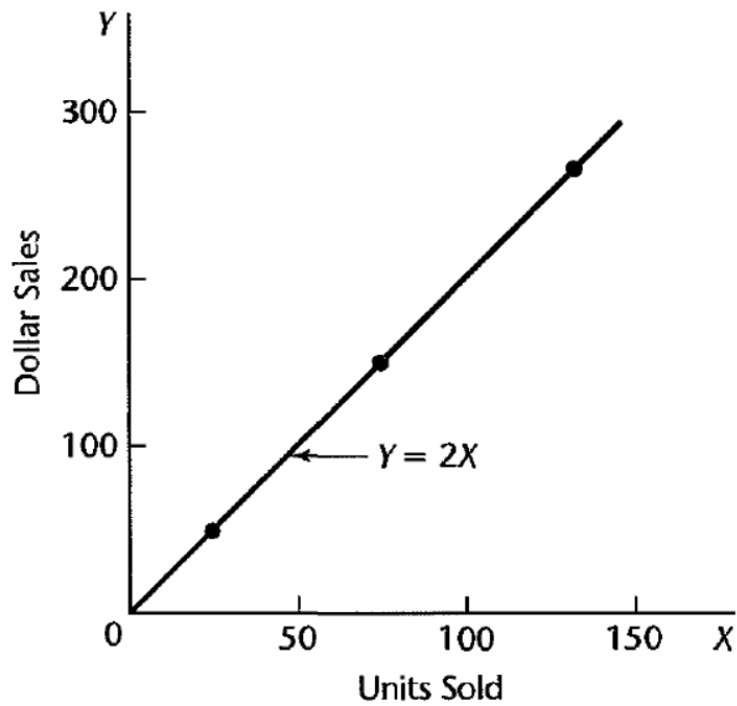


Fig.(1.1)

➤ Statistical Relation between Two Variables

Example (1-2)

Performance evaluations for 10 employees were obtained at midyear and at year-end. These data are plotted in Figure 1.2. Year-end evaluations are taken as the *dependent* or *response variable* Y , and midyear evaluations as the *independent*, *explanatory*, or *predictor*

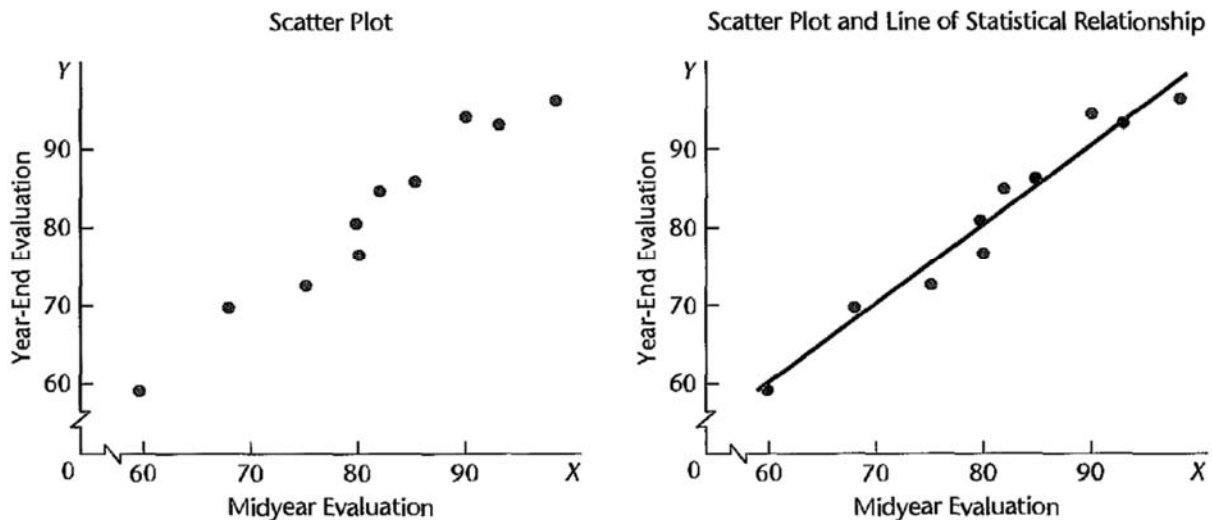
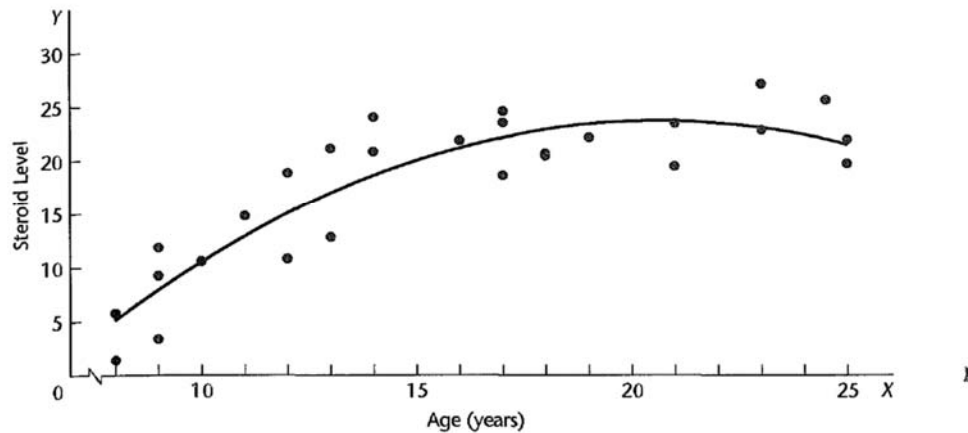


Figure (1.2)

Example (1-3)

Figure 1.3 presents data on age and level of a steroid in plasma for 27 healthy females between 8 and 25 years old. The data strongly suggest that the statistical relationship is *curvilinear* (not linear). The curve of relationship has also been drawn in Figure 1.3. It implies that, as age increases, steroid level increases up to a point and then begins to level off. Note again the scattering of points around the curve of statistical relationship, typical of all statistical relations

FIGURE 1.3 Curvilinear Statistical Relation between Age and Steroid Level in Healthy Females Aged 8 to 25.

What is the functional relations to describe Fig 1.2 and Fig 1.3?

It is a regression relation or regression model!

➤ Uses of Regression Analysis

Regression analysis serves three major purposes:

- (1) description
- (2) control
- (3) prediction

➤ **Regression and Causality**

The existence of a statistical relation between the **response variable Y** and the **explanatory or predictor variable X** does not imply in any way that Y depends causally on X .

No matter how strong is the statistical relation between X and Y , no cause-and-effect pattern is necessarily implied by the regression model.

For example, data on size of vocabulary (X) and writing speed (Y) for a sample of young children aged 5-10 will show a positive regression relation. This relation does not imply, however, that an increase in vocabulary causes a faster writing speed.

Here, other explanatory variables, such as age of the child and amount of education, affect both the vocabulary (X) and the writing speed (Y). Older children have a larger vocabulary and a faster writing speed.

Simple Linear Regression Model with Distribution of Error Terms Unspecified

Let

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (1.1)$$

where:

Y_i is the value of the response variable in the i th trial

β_0 and β_1 are parameters

X_i is a known constant, namely, the value of the predictor variable in the i th trial

X_i is a known constant, namely, the value of the predictor variable in the i th trial

ε_i is a random error term with mean $E\{\varepsilon_i\} = 0$ and variance $\sigma^2\{\varepsilon_i\} = \sigma^2$; ε_i and ε_j are uncorrelated so that their covariance is zero (i.e., $\sigma\{\varepsilon_i, \varepsilon_j\} = 0$ for all $i, j; i \neq j$)

$i = 1, \dots, n$

Remark: we will use the symbol $Var(\varepsilon_i) = \sigma^2\{\varepsilon_i\} = \sigma^2$

Important Features of the Model are:

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

$$Var(Y_i) = \sigma^2\{Y_i\} = \sigma^2$$

Example (2-4)

A consultant for an electrical distributor is studying the relationship between the number of bids requested by construction contractors for basic lighting equipment during a week and the time required to prepare the bids. Suppose that regression model (1.1) is applicable and is as follows:

$$Y_i = 9.5 + 2.1X_i + \varepsilon_i$$

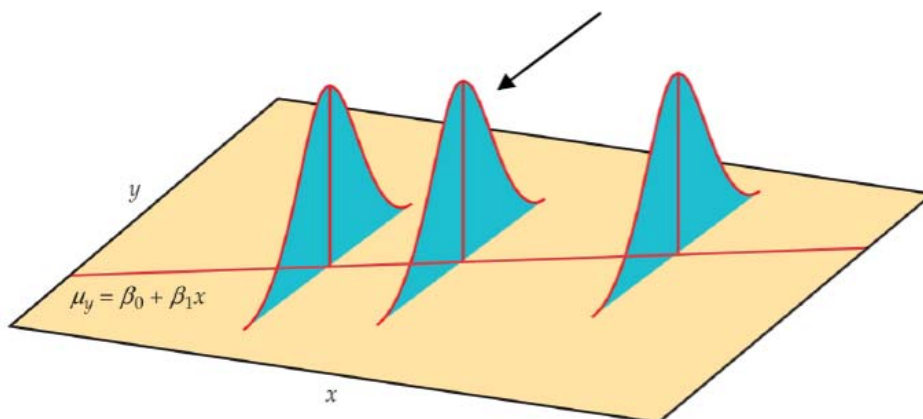
where X is the number of bids prepared in a week and Y is the number of hours required to prepare the bids. Figure 1.6 contains a presentation of the regression function:

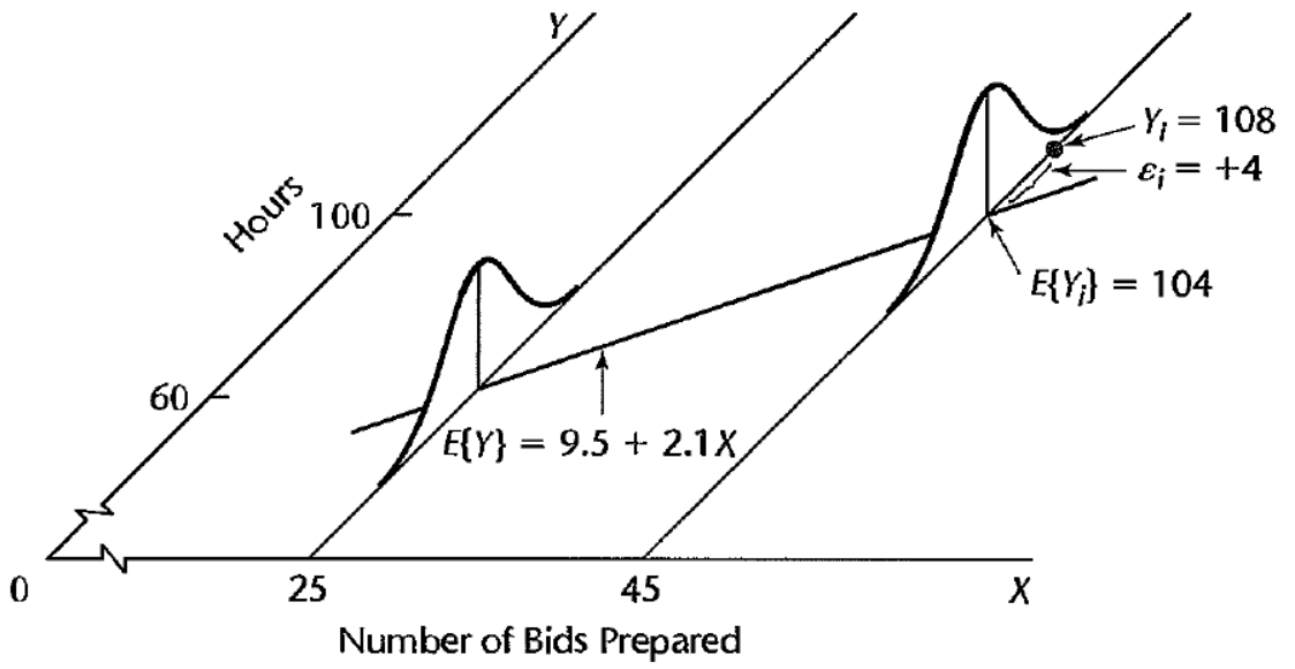
$$E\{Y\} = 9.5 + 2.1X$$

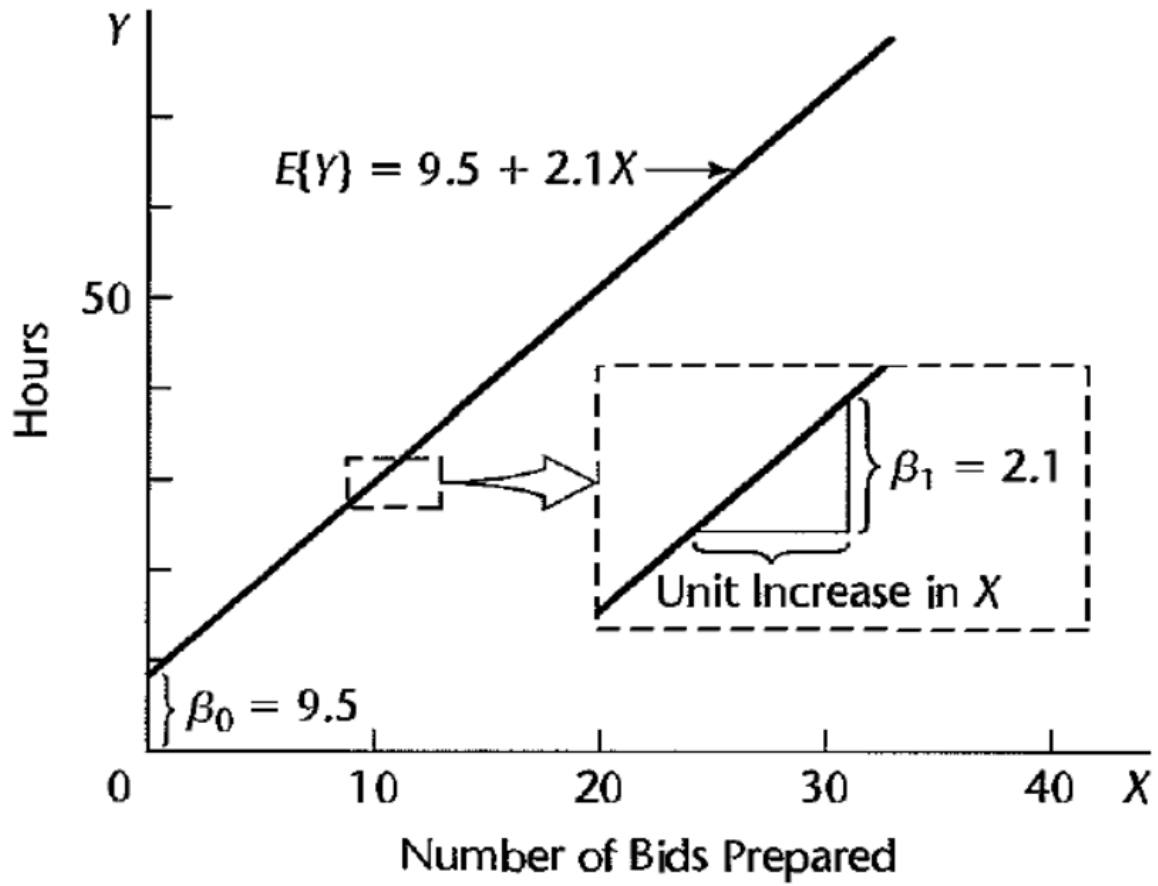
Suppose that in the i th week, $X_i = 45$ bids are prepared and the actual number of hours required is $Y_i = 108$. In that case, the error term value is $\varepsilon_i = 4$, for we have

$$E\{Y_i\} = 9.5 + 2.1(45) = 104$$

For any fixed 'x', the response 'y' follows a normal distribution with standard deviation σ .







Alternative Versions of Regression Model

The simple linear regression model is

$$Y_i = \beta_0 X_0 + \beta_1 X_i + \varepsilon_i \quad \text{where } X_0 \equiv 1$$

Alternative Versions of Regression Model is

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 (X_i - \bar{X}) + \beta_1 \bar{X} + \varepsilon_i \\ &= (\beta_0 + \beta_1 \bar{X}) + \beta_1 (X_i - \bar{X}) + \varepsilon_i \\ &= \beta_0^* + \beta_1 (X_i - \bar{X}) + \varepsilon_i \end{aligned}$$

Estimation of β_0 and β_1

The least square method for estimating the unknown parameters of the simple linear regression model can be explained as follows:

To find “good” estimators of the regression parameters β_0 and β_1 , we employ the method of least squares. For the observations (X_i, Y_i) for each case, the method of least squares considers the deviation of Y_i from its expected value:

$$Y_i - (\beta_0 + \beta_1 X_i)$$

In particular, the method of least squares requires that we consider the sum of the n squared deviations. This criterion is denoted by Q :

$$Q = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2, \quad (1.1)$$

Differentiating (1.1) with respect to β_0 and β_1 and equating to zero, we get

$$\begin{aligned} \frac{\partial Q}{\partial \beta_0} &= -2 \sum (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum X_i (Y_i - \beta_0 - \beta_1 X_i) \end{aligned}$$

Hence

$$\begin{aligned} -2 \sum (Y_i - b_0 - b_1 X_i) &= 0 \\ -2 \sum X_i (Y_i - b_0 - b_1 X_i) &= 0 \end{aligned}$$

$$\sum_{i=1}^n (Y_i - b_0 - b_1 X_i) = 0$$

$$\sum_{i=1}^n X_i (Y_i - b_0 - b_1 X_i) = 0$$

$$\sum Y_i - nb_0 - b_1 \sum X_i = 0$$

$$\sum X_i Y_i - b_0 \sum X_i - b_1 \sum X_i^2 = 0$$

Solving with respect b_0 and b_1 , we get

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$b_0 = \frac{1}{n} \left(\sum Y_i - b_1 \sum X_i \right) = \bar{Y} - b_1 \bar{X}$$

Which can be written as

$$\hat{\beta}_1 = b_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - \left(\sum x \right)^2}$$

$$\hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

Example 1 (Toluca Company Data)

The Toluca company manufactures refrigeration equipment as well as many replacement parts. In the past, one of the replacement parts has been produced periodically in lots of varying sizes. When a cost improvement program was undertaken, company officials wished to determine the optimum lot size for producing this part. The production of this part involves setting up the production process (which must be done no matter what is the lot size) and machining and assembly operations. One key input for the model to ascertain the optimum lot size was the relationship between lot size and labor hours required to produce the lot.

To determine this relationship, data on lot size and work hours for 25 recent production runs were utilized. The production conditions were stable during the six-month period in which the 25 runs were made and were expected to continue to be the same during the next three years, the planning period for which the cost improvement program was being conducted.

The data is given in the book data website:

<http://www.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/Chapter%20%20Data%20Sets/CH01TA01.txt>

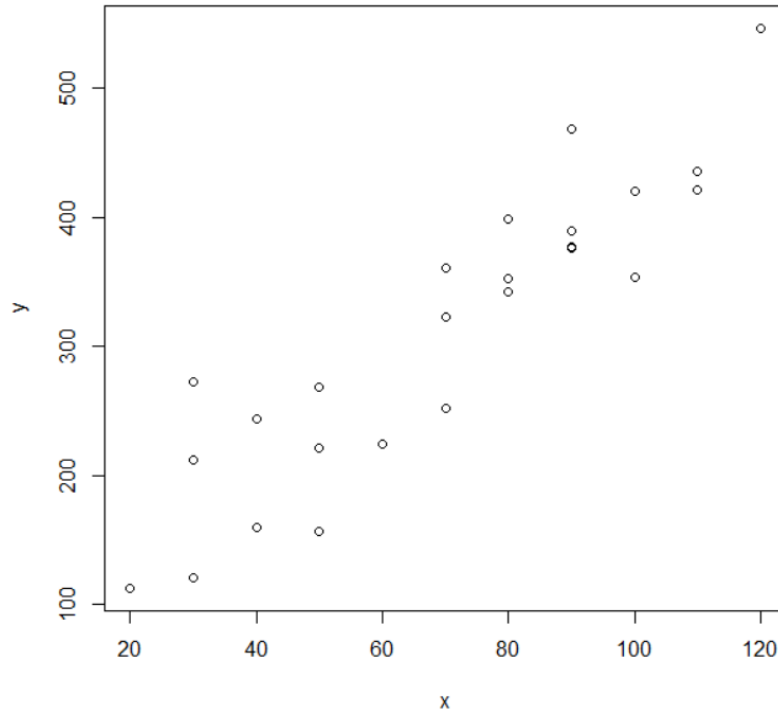
(1) Lot Size X_i	(2) Work Hours Y_i
80	399
30	121
50	221
...	...
40	244
80	342
70	323
<hr/> 1,750	<hr/> 7,807

Find the estimation of the simple linear regression model

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Solution

The scatter plot for the data shows that the simple linear model represents a good fit for the data as follows:



Then it is easy to calculate

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 70,690$$

$$\sum (X_i - \bar{X})^2 = 19,800$$

$$\bar{X} = 70.0$$

$$\bar{Y} = 312.28$$

Then

$$b_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{70,690}{19,800} = 3.5702$$

$$b_0 = \bar{Y} - b_1\bar{X} = 312.28 - 3.5702(70.0) = 62.37$$

One can use R as follows:

```
mydata = read.table("TCD.txt", header=TRUE)
# How to separate some variables the data file
x=mydata$X
y=mydata$Y
x.bar=mean(x)
y.bar=mean(y)
print(c(x.bar,y.bar))
t1=sum((x-x.bar)*(y-y.bar))
t2=sum((x-x.bar)^2)
b1=t1/t2
```

Or one can use the direct R command for regression as: `lm(y~x)` to get the same results.

```
Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)      x
    62.37      3.57
```

As we can see from the results the estimated linear regression model is

$$Y = 62.4 + 3.57 X \quad (2)$$

➤ Interpretation of the results

- 1- When the lot size (X) increases by one units, the work hours (Y) increase by 3.57 hours.
- 2- There is 62.4 hour of the work hours (Y) do not depend on the lot size (X).

➤ The estimated simple linear regression model in equation (2) can be used to predict the work hours required for a certain lot size. For example, if the lot size is 85 units, then

$$Y = 62.4 + 3.57 * 85 = 365.85 \text{ hours.}$$

➤ We can use the alternative Model as:

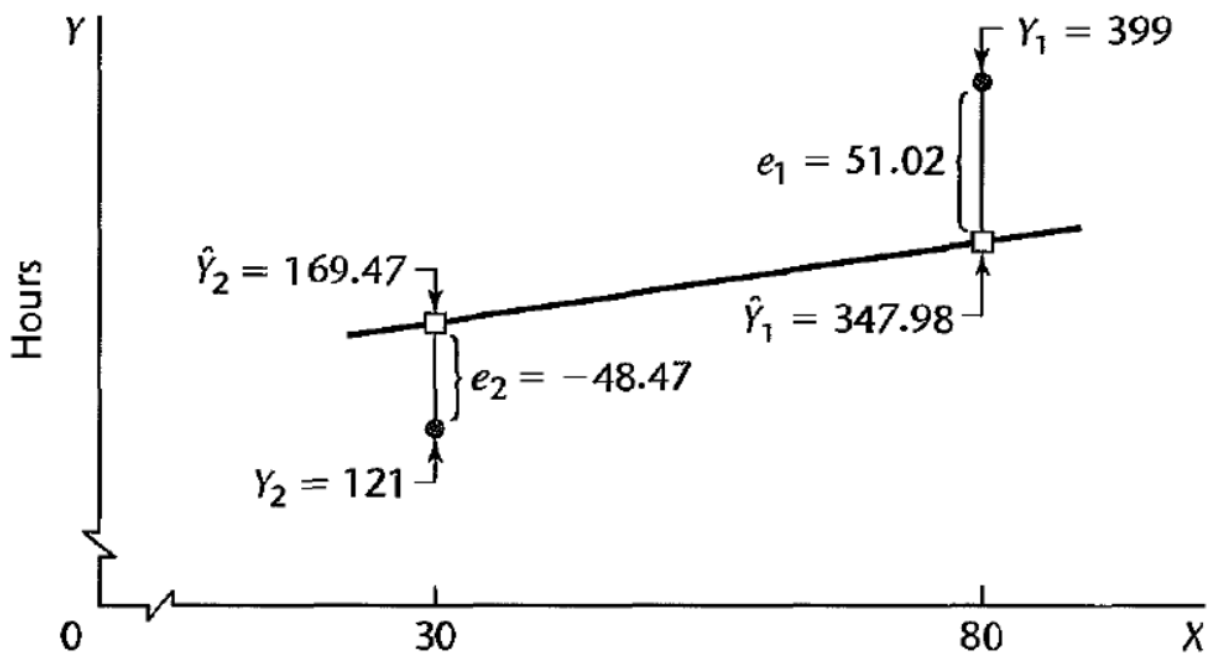
$$\hat{Y} = 312.28 + 3.5702(X - 70.0)$$

➤ The residual can be calculated at each point of the intendent variable x as

This residual is denoted by e_i and is defined in general as follows:

$$e_i = Y_i - \hat{Y}_i$$

For example, when $X=30$ and $X=80$, we calculate the residuals as:



From the Figure, we see that

$$e_i = Y_i - (b_0 + b_1 X_i) = Y_i - b_0 - b_1 X_i$$

$$e_1 = Y_1 - \hat{Y}_1 = 399 - 347.98 = 51.02$$

Similarly, we can calculate the residuals at the all point of x to get

<i>i</i>	<i>e</i>	<i>i</i>	<i>e</i>	<i>i</i>	<i>e</i>
1	51.02	11	-45.17	21	103.53
2	-48.47	12	-60.28	22	84.32
3	-19.88	13	5.32	23	38.83
4	-7.68	14	-20.77	24	-5.98
5	48.72	15	-20.09	25	10.72
6	-52.58	16	0.61		
7	55.21	17	42.53		
8	4.02	18	27.12		
9	-66.39	19	-6.68		
10	-83.88	20	-34.09		

These residuals can be calculated directly from R providing as:

`summary(model)$res`

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2.$$

Let's introduce some more notations:

$$S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$$

$$S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$$

$$S_{xy} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}$$

The least square estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

Properties of the fitted regression line

The residuals $e_i = Y_i - \hat{Y}_i$, $i = 1, 2, \dots$ of the simple linear regression model satisfies the following properties

1- The sum of the residuals is zero i.e.

$$\sum_{i=1}^n e_i = 0$$

Proof.

$$\begin{aligned} \sum e_i &= \sum (Y_i - \hat{Y}_i) = \sum (Y_i - (b_0 + b_1 X_i)) \\ &= \sum (Y_i - b_0 - b_1 X_i) \\ &= n\bar{Y} - nb_0 - nb_1 \bar{X} \\ &= n\bar{Y} - n(\bar{Y} - b_1 \bar{X}) - nb_1 \bar{X} \\ &= n\bar{Y} - n\bar{Y} + nb_1 \bar{X} - nb_1 \bar{X} \\ &= 0 \end{aligned}$$

2- The regression line always goes through the point (\bar{X}, \bar{Y}) .

Proof.

$$Y_i = b_0 + b_1 X_i = b_0 + b_1 \bar{X} = \bar{Y} - b_1 \bar{X} + b_1 \bar{X} = \bar{Y}$$

3- The sum of the observed values Y_i ; equals the sum of the fitted values \hat{Y}_i .

Proof.

$$\begin{aligned} \sum \hat{Y}_i &= \sum (b_0 + b_1 X_i) = nb_0 + nb_1 \bar{X} \\ &= n(\bar{Y} - b_1 \bar{X}) + nb_1 \bar{X} = n\bar{Y} - nb_1 \bar{X} + nb_1 \bar{X} \\ &= n\bar{Y} = \sum Y_i \end{aligned}$$

4- The sum of the weighted residuals is zero when the residual in the i th trial is

weighted by the level of the predictor variable in the i th trial $\sum_{i=1}^n e_i X_i = 0$

Proof.

$$\begin{aligned}
 \sum_{i=1}^n e_i X_i &= \sum_{i=1}^n (Y_i - \hat{Y}_i) X_i = \sum_{i=1}^n (Y_i - b_0 - b_1 X_i) X_i \\
 &= \sum_{i=1}^n (Y_i - (\bar{Y} - b_1 \bar{X}) - b_1 X_i) X_i \\
 &= \sum_{i=1}^n Y_i X_i - \bar{Y} \sum_{i=1}^n X_i - b_1 \sum_{i=1}^n X_i (X_i - \bar{X}) \\
 &= \sum_{i=1}^n Y_i X_i - \bar{Y} \sum_{i=1}^n X_i - b_1 (\sum_{i=1}^n X_i^2 - n\bar{X}^2) \\
 &= \sum_{i=1}^n Y_i X_i - \bar{Y} \sum_{i=1}^n X_i - \sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X}) \\
 &= \sum_{i=1}^n Y_i X_i - n\bar{X}\bar{Y} - \sum_{i=1}^n Y_i X_i + n\bar{X}\bar{Y} \\
 &= 0.
 \end{aligned}$$

5- A consequence of properties (1) and (4) is that the sum of the weighted residuals is zero when the residual in the i th trial is weighted by the fitted

value of the response variable for the i th trial $\sum_{i=1}^n e_i \hat{Y}_i$.

Proof.

$$\sum_{i=1}^n e_i \hat{Y}_i = \sum_{i=1}^n e_i (b_0 + b_1 X_i) = b_0 \sum_{i=1}^n e_i + b_1 \sum_{i=1}^n e_i X_i = 0 + 0 = 0$$

Point Estimator of σ^2

Under the assumption that the residuals $\varepsilon_i \sim N(0, \sigma^2)$, the maximum likelihood methods can be used to derive the MLE of σ^2 as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} = \frac{SSE}{n} \quad \text{which is biased estimate for } \sigma^2$$

The unbiased estimate of σ^2 can be obtained as

$$S^2 = MSE = \frac{n}{n-2} \hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} = \frac{SSE}{n-2} \quad \text{which is called the residual mean square}$$

MSE

Then

$$MSE = \frac{SSE}{n-2}$$

Example

We will calculate SSE for the Toluca Company example. The residuals were obtained earlier. From these results, we obtain:

$$SSE = 54825$$

Then

$$MSE = S^2 = \frac{54825}{25-2} = 2384 \quad \text{and} \quad S = \sqrt{MSE} = \sqrt{2384} = 48.8 \text{ hours}$$

The sum of squared errors can be calculated from R results simply as:

```
model=lm(y~x)
summary(model)
summary(model)$coef
e=summary(model)$res
sum(e^2)
```

Chapter 2

Inferences in Regression and Correlation Analysis

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2.$$

Let's introduce some more notations:

$$S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$$

$$S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$$

$$S_{xy} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}} \tag{2.1}$$

$$\hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

Properties of Point estimation of β_1, β_0

The point estimation of the coefficients of the simple linear regression model in (2.1) can be written in linear combination forms of Y_i as follows:

$$\begin{aligned}
 \hat{\beta}_1 = b_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i - \bar{Y} \sum_{i=1}^n (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \\
 &= \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2}, \\
 &= \sum_{i=1}^n \left(\frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) Y_i \\
 &= \sum_{i=1}^n K_i Y_i
 \end{aligned}$$

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i \tag{2.2}$$

where

$$K_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (2.3)$$

As we can see, the form of the point estimation of slope the simple liner model is given in a liner combination form of Y_i .

Similarly,

$$\begin{aligned} \hat{\beta}_0 = b_0 &= \bar{Y} - b_1 \bar{X} = \bar{Y} - \bar{X} \sum_{i=1}^n k_i Y_i \\ &= \sum_{i=1}^n \frac{Y_i}{n} - \bar{X} \sum_{i=1}^n k_i Y_i \\ &= \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} k_i \right) Y_i \\ &= \sum_{i=1}^n L_i Y_i \end{aligned}$$

$$\hat{\beta}_0 = b_0 = \sum_{i=1}^n L_i Y_i \quad (2.4)$$

where

$$L_i = \frac{1}{n} - \bar{X} K_i \quad (2.5)$$

and K_i is given in (2.3).

The coefficients K_i and L_i satisfy the following properties:

Lemma

The coefficients K_i given in (2.3) satisfies the following properties

$$\begin{aligned} \sum k_i &= 0 \\ \sum k_i X_i &= 1 \\ \sum k_i^2 &= \frac{1}{\sum (X_i - \bar{X})^2} \end{aligned}$$

Proof.

$$\sum k_i = \sum \left[\frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \right] = \frac{1}{\sum (X_i - \bar{X})^2} \sum (X_i - \bar{X}) = \frac{0}{\sum (X_i - \bar{X})^2} = 0$$

$$K_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} \implies K_i X_i = \frac{(X_i^2 - \bar{X} X_i)}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\implies \sum_{i=1}^n K_i X_i = \frac{\sum_{i=1}^n (X_i^2 - \bar{X} X_i)}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{\sum_{i=1}^n X_i^2 - n\bar{X}^2} = 1$$

$$\sum k_i^2 = \sum \left[\frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2} \right]^2 = \frac{1}{[\sum (X_i - \bar{X})^2]^2} \sum (X_i - \bar{X})^2 = \frac{1}{\sum (X_i - \bar{X})^2}$$

Lemma

The coefficients L_i given in (2.5) satisfies the following properties

$$\begin{aligned}\sum_{i=1}^n L_i &= 1 \\ \sum_{i=1}^n L_i X_i &= 0 \\ \sum_{i=1}^n L_i^2 &= \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\end{aligned}$$

Proof.

$$\sum_{i=1}^n L_i = \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} K_i \right) = \frac{n}{n} - \bar{X} \sum_{i=1}^n K_i = 1 - 0 = 1$$

$$\begin{aligned}\sum_{i=1}^n L_i X_i &= \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} K_i \right) X_i = \frac{1}{n} \sum_{i=1}^n X_i - \bar{X} \sum_{i=1}^n K_i X_i \\ &= \bar{X} - \bar{X} = 0\end{aligned}$$

$$\begin{aligned}
\sum_{i=1}^n L_i^2 &= \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} K_i \right)^2 = \sum_{i=1}^n \left(\frac{1}{n^2} - \frac{2K_i \bar{X}}{n} + K_i^2 \bar{X}^2 \right) \\
&= \sum_{i=1}^n \frac{1}{n^2} - \frac{2\bar{X} \sum_{i=1}^n K_i}{n} + \bar{X}^2 \sum_{i=1}^n K_i^2 \\
&= \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2}
\end{aligned}$$

1- Unbiasedness of Point estimation of β_1, β_0

Lemma

The point estimators of β_1 and β_0 are unbiased

Proof.

From (2.2), we have

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i \text{ then}$$

$$\begin{aligned}
E(\hat{\beta}_1) &= E(b_1) = \sum_{i=1}^n K_i E(Y_i) = \sum_{i=1}^n K_i (\beta_0 + \beta_1 X_i) \\
&= \sum_{i=1}^n K_i (\beta_0 + \beta_1 X_i) \\
&= \beta_0 \sum_{i=1}^n K_i + \beta_1 \sum_{i=1}^n K_i X_i \\
&= \beta_1.
\end{aligned}$$

Similarly, from (2.4), we have

$$\begin{aligned}
E(\hat{\beta}_0) &= E(b_0) = \sum_{i=1}^n L_i E(Y_i) = \sum_{i=1}^n L_i (\beta_0 + \beta_1 X_i) \\
&= \sum_{i=1}^n L_i (\beta_0 + \beta_1 X_i) \\
&= \beta_0 \sum_{i=1}^n L_i + \beta_1 \sum_{i=1}^n L_i X_i \\
&= \beta_0.
\end{aligned}$$

2- Variances

Lemma

The point estimators of β_1 and β_0 have the following variances, respectively

$$\text{Var}(\hat{\beta}_1) = \text{Var}(b_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}.$$

and

$$\text{Var}(\hat{\beta}_0) = \text{Var}(b_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right].$$

Proof.

From (2.2), we have

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i \text{ then}$$

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \text{Var}(b_1) = \sum_{i=1}^n K_i^2 \text{Var}(Y_i) = \sum_{i=1}^n K_i^2 \sigma^2 \\ &= \frac{\sigma^2}{\sum (X_i - \bar{X})^2}. \end{aligned}$$

Similarly, from (2.4), we have

$$\begin{aligned} \text{Var}(\hat{\beta}_0) &= \text{Var}(b_0) = \sum_{i=1}^n L_i^2 \text{Var}(Y_i) = \sum_{i=1}^n L_i^2 \sigma^2 \\ &= \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right]. \end{aligned}$$

Example

Consider the Toluca Company example, the variance of β_1, β_0 are:

$$\begin{aligned} \text{Var}(\hat{\beta}_1) = \text{Var}(b_1) &= \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \\ &= \frac{MSE}{\sum (X_i - \bar{X})^2} = \frac{2384}{19800} = .12040 \end{aligned}$$

Hence the squared error of $\hat{\beta}_1$ is

$$S.E.(\hat{\beta}_1) = \sqrt{\text{Var}(\hat{\beta}_1)} = \sqrt{.12040} = .3470$$

Similarly,

$$\begin{aligned} \text{Var}(\hat{\beta}_0) &= \text{Var}(b_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right] \\ &= \text{MSE} \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right] \\ &= 2,384 \left[\frac{1}{25} + \frac{70^2}{19800} \right] \\ &= 685.34 \end{aligned}$$

Hence the squared error of $\hat{\beta}_0$ is

$$S.E(\hat{\beta}_0) = \sqrt{\text{Var}(\hat{\beta}_0)} = \sqrt{685.34} = 26.18$$

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{for all } i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2.$$

Let's introduce some more notations:

$$S_{xx} = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$$

$$S_{yy} = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$$

$$S_{xy} = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y}$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i, \quad K_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \sum_{i=1}^n L_i Y_i, \quad L_i = \frac{1}{n} - \bar{X} K_i$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}, \quad \text{S.E}(\hat{\beta}_1) = \sqrt{\text{Var}(\hat{\beta}_1)}$$

and

$$\text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right], \quad \text{S.E}(\hat{\beta}_0) = \sqrt{\text{Var}(\hat{\beta}_0)}$$

The unbiased estimate of σ^2 is

$$s^2 = \text{MSE} = \hat{\sigma}^2 = \frac{\text{SSE}}{n-2}, \quad \text{SSE} = \sum_{i=1}^n e_i^2, \quad e_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \dots$$

3- Sampling distribution

The normal error regression model is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where:

Y_i is the observed response in the i th trial

X_i is a known constant, the level of the predictor variable in the i th trial

β_0 and β_1 are parameters

ε_i are independent $N(0, \sigma^2)$

$i = 1, \dots, n$

3-1 Sampling distribution of $\frac{\hat{\beta}_1 - \beta_1}{S.E(\hat{\beta}_1)} = \frac{b_1 - \beta_1}{S.E(b_1)}$ and $\frac{\hat{\beta}_0 - \beta_0}{S.E(\hat{\beta}_0)} = \frac{b_0 - \beta_0}{S.E(b_0)}$

Lemma

Let b_1 is the estimator of the slop in the simple linear regression model, then

$\frac{\hat{\beta}_1 - \beta_1}{S.E(\hat{\beta}_1)} = \frac{b_1 - \beta_1}{S.E(b_1)}$ has t distribution with (n-2) degrees of freedom.

Similarly,

Let b_0 is the estimator of the intercept in the simple linear regression model, then

$\frac{\hat{\beta}_0 - \beta_0}{S.E(\hat{\beta}_0)} = \frac{b_0 - \beta_0}{S.E(b_0)}$ has t distribution with (n-2) degrees of freedom.

Interval Estimation

This distribution of $\frac{\hat{\beta}_1 - \beta_1}{S.E(\hat{\beta}_1)} = \frac{b_1 - \beta_1}{S.E(b_1)}$ can be used to construct $100(1-\alpha)\%$

confidence interval for β_1 as follows

$$\hat{\beta}_1 \pm t_{1-\alpha/2, n-2} S.E(\hat{\beta}_1)$$

Similarly, the distribution of $\frac{\hat{\beta}_0 - \beta_0}{S.E(\hat{\beta}_0)} = \frac{b_0 - \beta_0}{S.E(b_0)}$ can be used to construct $100(1-\alpha)\%$

confidence interval for β_0 as follows

$$\hat{\beta}_0 \pm t_{1-\alpha/2, n-2} S.E(\hat{\beta}_0)$$

Example

Consider the Toluca Company example, find 95% confidence intervals for both of β_1 and β_0 .

For such data, we have calculated

$$Var(\hat{\beta}_1) = Var(b_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2} = \frac{MSE}{\sum (X_i - \bar{X})^2} = \frac{2384}{19800} = .12040$$

Hence the squared error of $\hat{\beta}_1$ is

$$S.E(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{.12040} = .3470$$

Similarly,

$$S.E(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = \sqrt{685.34} = 26.18$$

$$t_{1-\alpha/2, n-2} = t_{0.975, 23} = 2.068658 = 2.069$$

and

$$\hat{\beta}_1 = 3.5702, \hat{\beta}_0 = 62.4$$

Then 95% confidence interval for β_1 as follows

$$\hat{\beta}_1 \pm t_{1-\alpha/2, n-2} S.E(\hat{\beta}_1)$$

$$3.5702 \pm 2.069(0.3470)$$

Hence

$$2.85 < \beta_1 < 4.29$$

Similarly, then 95% confidence interval for β_0 as follows

$$\widehat{\beta}_0 \pm t_{1-\alpha/2, n-2} S.E(\widehat{\beta}_0)$$

$$62.4 \pm 2.069(26.18)$$

Hence

$$8.25 < \beta_0 < 116.57$$

This can be done easily using R as:

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

Hypotheses Testing

This distribution of $\frac{\widehat{\beta}_1 - \beta_1}{S.E(\widehat{\beta}_1)} = \frac{b_1 - \beta_1}{S.E(b_1)}$ and $\frac{\widehat{\beta}_0 - \beta_0}{S.E(\widehat{\beta}_0)} = \frac{b_0 - \beta_0}{S.E(b_0)}$ can be used to test

some hypotheses concerning the coefficients of the simple linear regression model as follows:

Steps for testing β_1

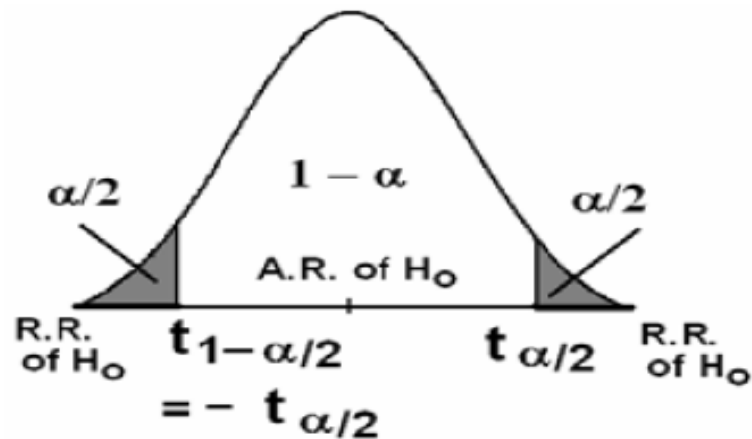
i) Setup the hypotheses

$$H_0 : \beta_1 = \beta_1^{(0)} \quad vs \quad H_1 : \beta_1 \neq \beta_1^{(0)}$$

ii) Test statistic under H_0

$$T_0 = \frac{b_1 - \beta_1^{(0)}}{S.E(b_1)} \quad \text{this statistic has t distribution with } (n-2) \text{ d.f}$$

iii) Critical regions



R.R: Rejection Region and A.R: Acceptance Region

iv) Decision

When the calculated T_0 belongs to the shaded areas, we reject the null hypothesis H_0 , otherwise Accept H_0 .

Steps for testing β_0

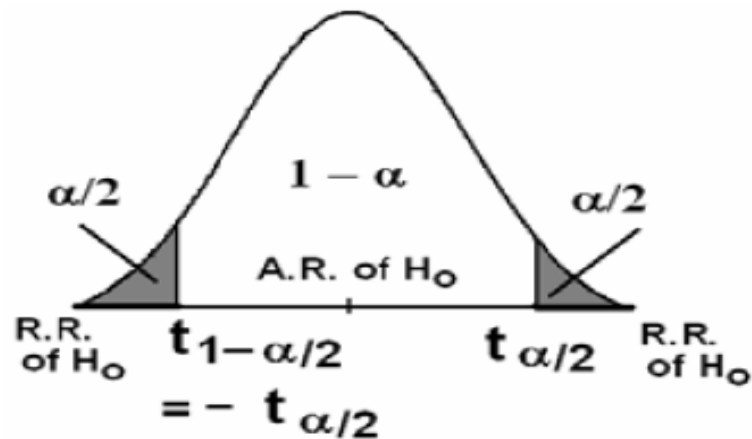
- i) Setup the hypotheses

$$H_0 : \beta_0 = \beta_0^{(0)} \quad vs \quad H_1 : \beta_0 \neq \beta_0^{(0)}$$

- ii) Test statistic under H_0

$$T_0 = \frac{b_0 - \beta_0^{(0)}}{S.E(b_0)} \quad \text{this statistic has t distribution with } (n-2) \text{ d.f}$$

- iii) Critical regions



R.R: Rejection Region and A.R: Acceptance Region

- iv) Decision

When the calculated T_0 belongs to the shaded areas, we reject the null hypothesis H_0 , otherwise Accept H_0 .

Remarks:

(1) In some applications, we may need to test

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

and

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

In these cases, you need to replace both of $\beta_1^{(0)}$ and $\beta_0^{(0)}$ by zeros in the test steps. This equivalent to testing the significance of the linear term (β_1) or the intercept term β_0 . R output are designed for these cases.

(2) In some applications, we may need to test

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 > (<) 0$$

and

$$H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 > (<) 0$$

In these cases, you need to replace the critical regions to one-sided critical regions.

(3) One may use the two-sided p-value approach: $p\text{-value} = 2P(T > T_0)$, then reject H_0 when $p\text{-value} \leq \alpha$, otherwise accept H_0 . The one-sided p-value is $p\text{-value} = P(T > T_0)$, then reject H_0 when $p\text{-value} \leq \alpha$, otherwise accept H_0 .

Example:

Consider the Toluca Company example, test the hypotheses

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

and

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

Testing β_1

- i) Setup the hypotheses

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

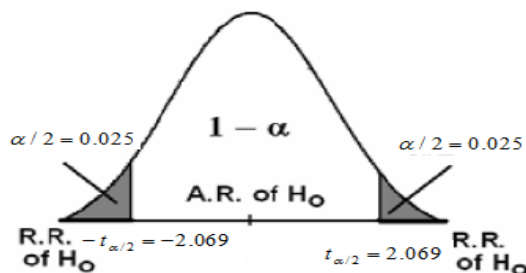
- ii) Test statistics under H_0

$$T_0 = \frac{b_1 - \beta_1^{(0)}}{S.E.(b_1)} = \frac{3.5702 - 0}{0.3470} = 10.29$$

- iii) Critical regions

$$\alpha = 0.05 \quad \rightarrow \quad \alpha / 2 = 0.025$$

$$t_{\alpha/2} = 2.069 \quad \text{and} \quad -t_{\alpha/2} = -2.069$$



R.R: Rejection Region and A.R: Acceptance Region

- iv) Decision: The calculated $T_0 = 10.29$ belongs to the shaded areas, then we reject the null hypothesis H_0

These calculations can be conducted using the R output as:

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	62.366	26.177	2.382	0.0259 *
x	3.570	0.347	10.290	4.45e-10 ***

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

As we can see from the results that, $T=10.29$. Also, the $p\text{-value}=0.000 < 0.05$, then reject H_0 .

Testing β_0

- i) Setup the hypotheses

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

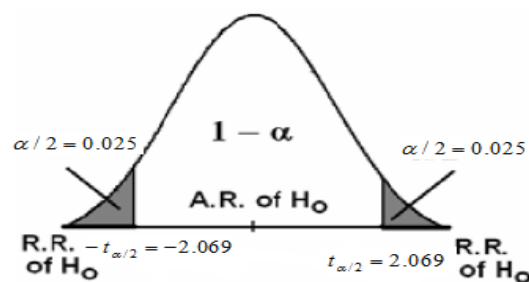
- ii) Test statistics under H_0

$$T_0 = \frac{b_0 - \beta_0^{(0)}}{S.E(b_0)} = \frac{62.4 - 0}{26.18} = 2.38$$

- iii) Critical regions

$$\alpha = 0.05 \rightarrow \alpha / 2 = 0.025$$

$$t_{\alpha/2} = 2.069 \quad \text{and} \quad -t_{\alpha/2} = -2.069$$



R.R: Rejection Region and A.R: Acceptance Region

- iv) Decision: The calculated $T_0 = 2.38$ belongs to the shaded areas, then we reject the null hypothesis H_0 .

As we can see from the results that, $T=10.29$. Also, the $p\text{-value}=0.026 < 0.05$, then reject H_0 .

Recall:

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2.$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i, \quad K_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \sum_{i=1}^n L_i Y_i, \quad L_i = \frac{1}{n} - \bar{X} K_i$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right],$$

The unbiased estimate of σ^2 is

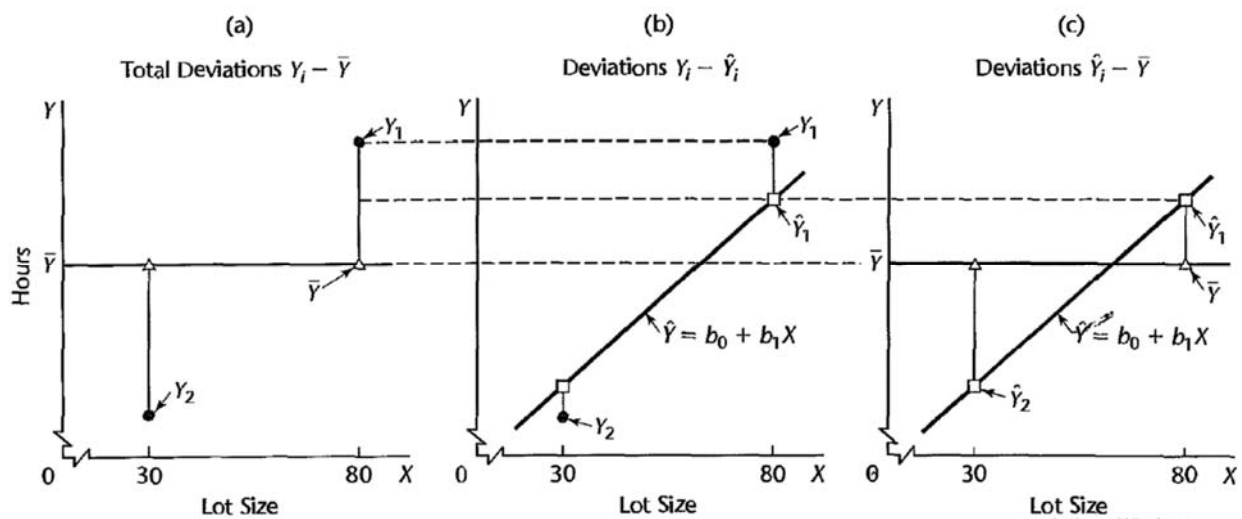
$$s^2 = \text{MSE} = \hat{\sigma}^2 = \frac{\text{SSE}}{n-2}, \quad \text{SSE} = \sum_{i=1}^n e_i^2, \quad e_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \dots$$

Analysis of Variance Approach to Regression Analysis

We now have developed the basic regression model and demonstrated its major uses. At this point, we consider the regression analysis from the perspective of analysis of variance. This new perspective will not enable us to do anything new, but the analysis of variance approach will come into its own when we take up multiple regression models and other types of linear statistical models.

Types of sum of squared Errors

Use the data in Toluca Company example, we show three types of sum of the squared errors as:



1- *SSTO* stands for *total sum of squares*

$$SSTO = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

2- *SSE* stands for *error sum of squares*

$$SSE = \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

3- *SSR* stands for *regression sum of squares*

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

Lemma:

$$SSTO = SSR + SSE$$

Proof.

$$\underbrace{Y_i - \bar{Y}}_{\text{Total deviation}} = \underbrace{\hat{Y}_i - \bar{Y}}_{\substack{\text{Deviation of fitted} \\ \text{regression value} \\ \text{around mean}}} + \underbrace{Y_i - \hat{Y}_i}_{\substack{\text{Deviation} \\ \text{around} \\ \text{fitted} \\ \text{regression} \\ \text{line}}}$$

$$\begin{aligned}
 \sum (Y_i - \bar{Y})^2 &= \sum [(\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)]^2 \\
 &= \sum [(\hat{Y}_i - \bar{Y})^2 + (Y_i - \hat{Y}_i)^2 + 2(\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i)] \\
 &= \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2 + 2 \sum (\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i)
 \end{aligned}$$

Activate Windows
Go to Settings to

But

$$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})(Y_i - \hat{Y}_i) = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})e_i = \sum_{i=1}^n \hat{Y}_i e_i - \bar{Y} \sum_{i=1}^n e_i = 0 - 0 = 0$$

Then

$$\begin{aligned}
 SSTO &= \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \\
 &= SSR + SSE
 \end{aligned}$$

Types of sum of squared Errors

4- *SSTO* stands for *total sum of squares*

$$SSTO = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

5- *SSE* stands for *error sum of squares*

$$SSE = \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

6- *SSR* stands for *regression sum of squares*

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

Lemma:

$$SSTO = SSR + SSE$$

$$SSR = b_1^2 \sum (X_i - \bar{X})^2$$

ANOVA TABLE

Source of Variation	SS	df	MS
Regression	$SSR = \sum(\hat{Y}_i - \bar{Y})^2$	1	$MSR = \frac{SSR}{1}$
Error	$SSE = \sum(Y_i - \hat{Y}_i)^2$	$n - 2$	$MSE = \frac{SSE}{n - 2}$
Total	$SSTO = \sum(Y_i - \bar{Y})^2$	$n - 1$	

$F_0 = \frac{MSR}{MSE}$

ANOVA tables are widely used; we shall usually utilize the basic type of the tables.

F-test

This test is basically depending on the ANOVA table and it works like t-test (in the simple linear regression model). We summarize the test as given below:

i) Setup the hypotheses

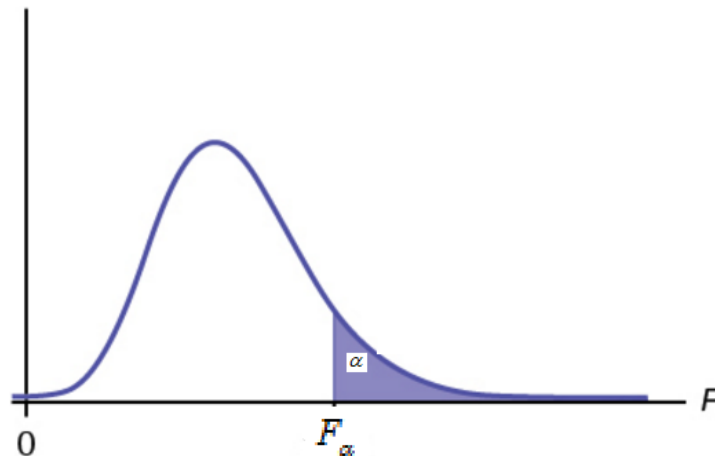
$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$$

ii) Test statistic under H0

$$F_0 = \frac{MSR}{MSE} \quad (\text{From ANOVA table})$$

this statistic has F_α distribution with (1,n-2) d.f

iii) Critical region



The shaded area is the rejection Region and the unshaded area is acceptance Region

iv) Decision

When the calculated F_0 belongs to the shaded area, we reject the null hypothesis H0, otherwise Accept H0.

P-value approach

- i) Setup the hypotheses

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

- ii) Calculate p-value

$$P \text{ -value} = P(F \geq F_0)$$

Reject H_0 , otherwise, Accept H_0

Remarks:

Since Both of F-test and t-test do the same job for testing

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

So, there is a relation between F_0 and T_0 as ($F_0 = T_0^2$)

Example

Consider the Toluca Company example, Use F test (ANOVA) for testing the significance of the linear term in the simple linear regression model.

From the data, we have

$$SSTO = \sum_{i=1}^n (Y_i - \bar{Y})^2 = 307203$$

$$SSE = \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 = 54825$$

$$\begin{aligned} SSR &= \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = b_1^2 S_{xx} = SSTO - SSE \\ &= 307203 - 54825 = 252378 \end{aligned}$$

Then The ANOVA Table is

Source of Variation	SS	df	MS	F0
Regression	SSR=252378	1	MSR=252378	F ₀ =105.9
Error	SSE=54825	23	MSE=2384	
SSTO	SSTo=307203	24		

i) Setup the hypotheses

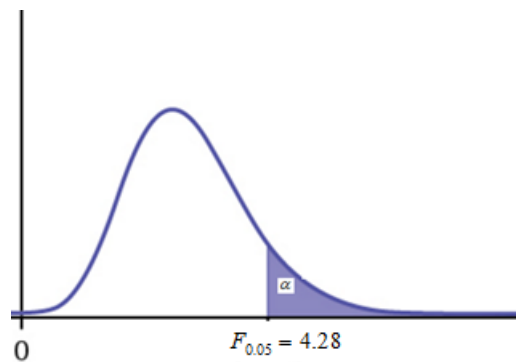
$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$$

ii) Test statistic under H0

$$F_0 = 105.9 \quad (\text{From ANOVA table})$$

this statistic has F_α distribution with (1,n-2) d.f

iii) Critical regions



The shaded area is the rejection Region and the unshaded shaded area is acceptance Region

iv) Decision

since the calculated $F_0 = 105.9 > 4.28$ belongs to the shaded area, we reject the null

P-value approach

i) Setup the hypotheses

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

ii) Calculate p-value

$$P\text{-value} = P(F \geq F_0) \approx 0.0000 = .0000 < 0.05$$

Then Reject H0

Remarks:

Since Both of F-test and t-test do the same job for testing

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

So, there is a relation between F_0 and T_0 as $(105.9 = (10.29)^2)$

```
> anova(model)
Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)
x         1 252378  252378  105.88 4.449e-10 ***
Residuals 23  54825    2384
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Coefficient of Determination

We saw earlier that $SSTO$ measures the variation in the observations Y_i , or the uncertainty in predicting Y , when no account of the predictor variable X is taken. Thus, $SSTO$ is a measure of the uncertainty in predicting Y when X is not considered. Similarly, SSE measures the variation in the Y_i when a regression model utilizing the predictor variable X is employed. A natural measure of the effect of X in reducing the variation in Y , i.e., in reducing the uncertainty in predicting Y , is to express the reduction in variation ($SSTO - SSE = SSR$) as a proportion of the total variation

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

The measure R^2 is called the *coefficient of determination*.

For the Toluca Company example, we obtained $SSTO = 307,203$ and $SSR = 252,378$. Hence:

$$R^2 = \frac{252,378}{307,203} = .822$$

Thus, the variation in work hours is reduced by 82.2 percent when lot size is considered.

Coefficient of Correlation

A measure of linear association between Y and X when both Y and X are random is the *coefficient of correlation*. This measure is the signed square root of R^2 :

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

A plus or minus sign is attached to this measure according to whether the slope of the fitted regression line is positive or negative. Thus, the range of r is: $-1 \leq r \leq 1$.

Example

For the Toluca Company example, we obtained $R^2 = .822$. Treating X as a random variable, the correlation coefficient here is:

$$r = +\sqrt{.822} = .907$$

The plus sign is affixed since b_1 is positive.

Remark

The Correlation Coefficient between two variables Y_1 and Y_2 is given by

$$r_{12} = \frac{\sum(Y_{i1} - \bar{Y}_1)(Y_{i2} - \bar{Y}_2)}{[\sum(Y_{i1} - \bar{Y}_1)^2 \sum(Y_{i2} - \bar{Y}_2)^2]^{1/2}}$$

```
> summary(model)$r.squared
[1] 0.8215335
```

```
cor(x,y)
```

```
[1] 0.9063848
```

```
tt=summary(model)$r.squared
```

```
> sqrt(tt)
```

```
[1] 0.9063848
```

Recall:

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{for all } i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2.$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i, \quad K_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \sum_{i=1}^n L_i Y_i, \quad L_i = \frac{1}{n} - \bar{X} K_i$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right],$$

The unbiased estimate of σ^2 is

$$s^2 = \text{MSE} = \hat{\sigma}^2 = \frac{\text{SSE}}{n-2}, \quad \text{SSE} = \sum_{i=1}^n e_i^2, \quad e_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \dots$$

Interval Estimation of the mean response when $X = X_h$

- The mean response when $X = X_h$ is denoted $E(Y_h)$
- The point estimation of the response mean is given by

$$\hat{Y}_h = b_0 + b_1 X_h$$

- Sampling distribution of \hat{Y}_h :

Since $\hat{Y}_h = b_0 + b_1 X_h$ and

$b_0 \sim N(\beta_0, \text{Var}(b_0))$, $b_1 \sim N(\beta_1, \text{Var}(b_1))$, then \hat{Y}_h is normally distributed with mean

$$E(\hat{Y}_h) = E(b_0) + E(b_1 X_h) = E(Y_h) = \beta_0 + \beta_1 X_h$$

and

$$\begin{aligned} \text{Var}(\hat{Y}_h) &= \text{Var}(b_0 + b_1 X_h) = \text{Var}(\bar{Y} + b_1(X_h - \bar{X})) \\ &= \text{Var}(\bar{Y}) + (X_h - \bar{X})^2 \text{Var}(b_1) + 2(X_h - \bar{X}) \text{Cov}(\bar{Y}, b_1) \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2 (X_h - \bar{X})^2}{S_{xx}} + 0 \end{aligned}$$

because

$$\begin{aligned}
Cov(\bar{Y}, b_1) &= Cov\left(\frac{1}{n} \sum_{i=1}^n y_i, \sum_{i=1}^n k_i y_i\right) \\
&= \frac{1}{n} \sum_{i=1}^n k_i Cov(y_i, y_i) \\
&= \frac{1}{n} \sum_{i=1}^n k_i Var(y_i) \\
&= \frac{1}{n} \sum_{i=1}^n k_i \sigma^2 = \frac{\sigma^2}{n} \sum_{i=1}^n k_i = 0
\end{aligned}$$

Then

$$\begin{aligned}
Var(\hat{Y}_h) &= \sigma^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{xx}} \right) = MSE \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{xx}} \right) \\
S.E.(\hat{Y}_h) &= \sqrt{Var(\hat{Y}_h)}
\end{aligned}$$

Lemma

The statistics $\frac{\hat{Y}_h - E(Y_h)}{S.E.(\hat{Y}_h)}$ had t distribution with (n-2) degrees of freedom. For more details, see the book (page 54).

This lemma enables us to construct $100(1-\alpha)\%$ confidence interval about the response mean Y_h as follows:

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

Example

Consider the Toluca Company example, find 90% confidence interval for the response mean $E(Y_h)$ when $X_h = 65$.

From the data, we have

$$\hat{Y}_h = b_0 + b_1 X_h = 62.37 + 3.5702(65) = 294.4$$

$$Var(\hat{Y}_h) = MSE \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{xx}} \right) = 2384 \left[\frac{1}{25} + \frac{(65 - 70.0)^2}{19800} \right] = 98.37$$

$$S.E(\hat{Y}_h) = \sqrt{Var(\hat{Y}_h)} = 9.918$$

For a 90 percent confidence coefficient, we require $t(.95; 23) = 1.714$.

Hence, our confidence interval with confidence coefficient .90 is by

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

$$294.4 - 1.714(9.918) < E(Y_h) < 294.4 + 1.714(9.918)$$

$$277.4 < E(Y_h) < 311.4$$

We conclude with confidence coefficient .90 that the mean number of work hours required when lots of 65 units are produced is somewhere between 277.4 and 311.4 hours. We see that our estimate of the mean number of work hours is moderately precise.

Example 2

Suppose the Toluca Company wishes to estimate $E\{Y_h\}$ for lots with $X_h = 100$ units with a 90 percent confidence interval. We require:

$$\begin{aligned}\hat{Y}_h &= 62.37 + 3.5702(100) = 419.4 \\ s^2\{\hat{Y}_h\} &= 2,384 \left[\frac{1}{25} + \frac{(100 - 70.00)^2}{19,800} \right] = 203.72 \\ s\{\hat{Y}_h\} &= 14.27 \\ t(.95; 23) &= 1.714\end{aligned}$$

Hence, the 90 percent confidence interval is:

$$\begin{aligned}419.4 - 1.714(14.27) &\leq E\{Y_h\} \leq 419.4 + 1.714(14.27) \\ 394.9 &\leq E\{Y_h\} \leq 443.9\end{aligned}$$

Note that this confidence interval is somewhat wider than that for Example 1, since the X_h level here ($X_h = 100$) is substantially farther from the mean $\bar{X} = 70.0$ than the X_h level for Example 1 ($X_h = 65$).

One may use R commands:

```
newx = data.frame(x=100)
```

```
predict(model, newx, level=0.90,interval="confidence")
```

Predicting a Future Observation When X is Known

We consider now the prediction of a new observation Y corresponding to a given level X of the predictor variable.

If β_0, β_1, σ were known, we'd know that the distribution of responses when $X=X_h$ is normal with mean $\beta_0 + \beta_1 X_h$ and standard deviation σ . Thus, making use of the normal distribution (and equivalently, the empirical rule) we know that if we took a sample item from this distribution, it is very likely that the value fall within 2 standard deviations of the mean. That is, we would know that the probability that the sampled item lies within the range $(\beta_0 + \beta_1 X_h - \sigma, \beta_0 + \beta_1 X_h + \sigma)$ is approximately 0.95.

In practice, we don't know the mean $\beta_0 + \beta_1 X_h$ or the standard deviation σ . However, we just constructed a $(1-\alpha)100\%$ Confidence Interval for $E\{Y_h\}$,

The prediction error is for the new observation is the difference between the observed value and its predicted value: $Y_{h(new)} - \hat{Y}_h$. Since the data are assumed to be independent, the new (future) value is independent of its predicted value, since it wasn't used in the regression analysis. The variance of the prediction error can be obtained as follows:

$$\begin{aligned} \text{Var}\{pred\} &= \text{Var}\{Y_{h(new)} - \hat{Y}_h\} = \text{Var}\{Y_{h(new)}\} + \text{Var}\{\hat{Y}_h\} = \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \\ &= \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \end{aligned}$$

and an unbiased estimator is:

$$\text{Var}\{pred\} = \text{MSE} \left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right]$$

$$S.E\{pred\} = \sqrt{\text{Var}(pred)}$$

A $(1-\alpha)100\%$ Prediction Interval for New Observation When $X=X_h$

$$\hat{Y}_h \pm t(\alpha/2; n-2) \sqrt{\text{MSE} \left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right]}$$

Example (book: page.59)

The Toluca Company studied the relationship between lot size and work hours primarily to obtain information on the mean work hours required for different lot sizes for use in determining the optimum lot size. The company was also interested, however, to see whether the regression relationship is useful for predicting the required work hours for individual lots. Suppose that the next lot to be produced consists of $X_h = 100$ units and that a 90 percent prediction interval is desired. We require $t(.95; 23) = 1.714$. From earlier work, we have:

In this example

$$\hat{Y}_h = b_0 + b_1 X_h = 62.37 + 3.5702(100) = 419.4$$

$$Var\{pred\} = MSE \left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] = 2587.72$$

$$S.E\{pred\} = \sqrt{Var(pred)} = 50.87$$

For a 90 percent confidence coefficient, we require $t(.95; 23) = 1.714$.

Hence, our confidence interval with confidence coefficient .90 is by

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

$$419.4 - 1.714(50.87) < Y_{h(\text{new})} < 419.4 + 1.714(50.87)$$

$$332.2 < Y_{h(\text{new})} < 506.6$$

With confidence coefficient .90, we predict that the number of work hours for the next production run of 100 units will be somewhere between 332 and 507 hours

R commands:

```
newx = data.frame(x=100)
```

```
predict(model, newx, level=0.90, interval="predict")
```

Diagnostics and Remedial Measures

Recall:

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{for all } i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2.$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i, \quad K_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \sum_{i=1}^n L_i Y_i, \quad L_i = \frac{1}{n} - \bar{X} K_i$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right],$$

The unbiased estimate of σ^2 is

$$s^2 = \text{MSE} = \hat{\sigma}^2 = \frac{\text{SSE}}{n-2}, \quad \text{SSE} = \sum_{i=1}^n e_i^2, \quad e_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \dots$$

When a regression model, such as the simple linear regression model is considered for an application, we can usually not be certain in advance that the model is appropriate for that application. Anyone, or several, of the features of the model (conditions), such as linearity of the regression function or normality of the error terms, may not be appropriate for the particular data at hand. Hence, it is important to examine the aptness of the model for the data before inferences based on that model are undertaken. In this part, we discuss some simple graphic methods for studying the appropriateness of a model.

We also consider some remedial techniques that can be helpful when the data are not in accordance with the conditions of regression model. One of these techniques is the transformations.




Transformations for Nonlinear Relation Only

We first consider transformations for linearizing a nonlinear regression relation when the distribution of the error terms is reasonably close to a normal distribution and the error terms have approximately constant variance. In this situation, transformations on X should be attempted. The reason why transformations on Y may not be desirable here is that a transformation on Y, such as $Y' = \sqrt{y}$, may materially change the shape of the distribution

of the -error terms from the normal distribution and may also lead to substantially differing error term variances.

The following Figure contains some prototype nonlinear regression relations with constant error variance and also presents some simple transformations on X that may be helpful to linearize the regression relationship without affecting the distributions of Y. Several alternative transformations may be tried. Scatter plots and residual plots based on each transformation should then be prepared and analyzed, to decide which transformation is most effective

Prototype Nonlinear Regression Patterns with Constant Error Variance and Simple Transformations of X.

	Prototype Regression Pattern	Transformations of X
(a)		$X' = \log_{10} X$ $X' = \sqrt{X}$
(b)		$X' = X^2$ $X' = \exp(X)$
(c)		$X' = 1/X$ $X' = \exp(-X)$

Example (page 129)

Data from an experiment on the effect of number of days of training received (X) on performance (Y) in a battery of simulated sales situations are presented in the following table, columns 1 and 2, for the 10 participants in the study. A scatter plot of these data is shown. Clearly the regression relation appears to be curvilinear, so the simple linear regression model does not be appropriate. Since the variability at the different X levels appears to be fairly constant, we shall consider a transformation on X. Based on the prototype plot in (a), we shall consider initially the square root transformation

$X' = \sqrt{x}$ The transformed values are shown below:

Sales Trainee	(1) Days of Training	(2) Performance Score	(3)
<i>i</i>	X_i	Y_i	$X'_i = \sqrt{X_i}$
1	.5	42.5	.70711
2	.5	50.6	.70711
3	1.0	68.5	1.00000
4	1.0	80.7	1.00000
5	1.5	89.0	1.22474
6	1.5	99.6	1.22474
7	2.0	105.3	1.41421
8	2.0	111.8	1.41421
9	2.5	112.3	1.58114
10	2.5	125.7	1.58114

To examine further whether the simple linear regression model (2.1) is appropriate now, we fit it to the transformed X data. The regression calculations with the transformed X data are carried out in the usual fashion, except that the predictor variable now is X' . We obtain the following fitted regression function:

$$Y = -10.33 + 83.45X'$$

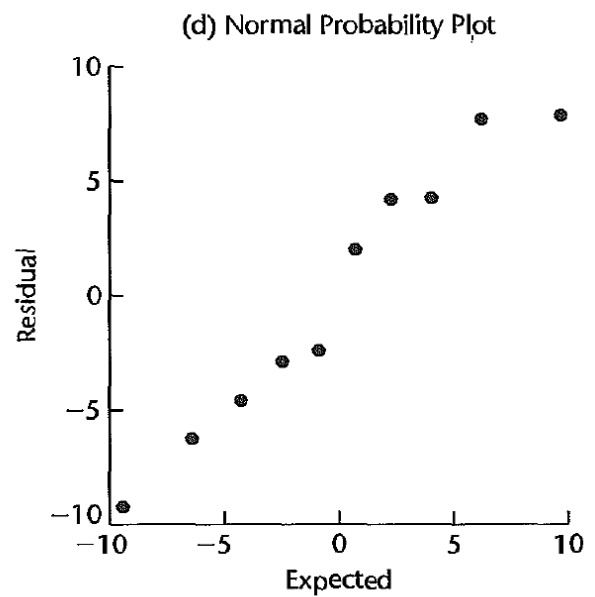
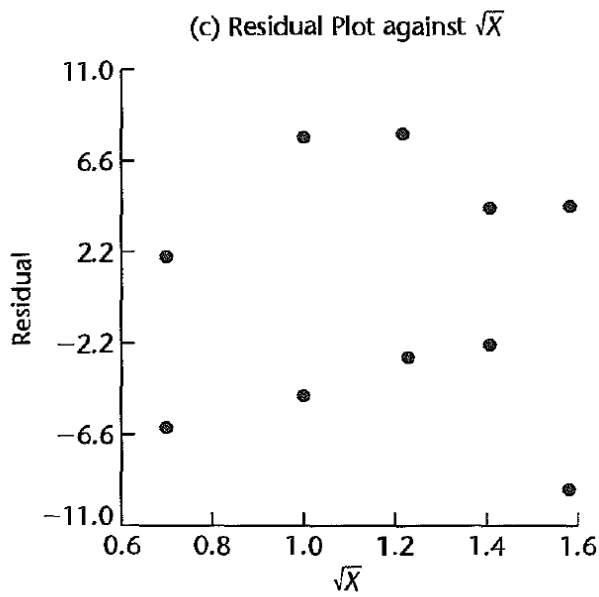
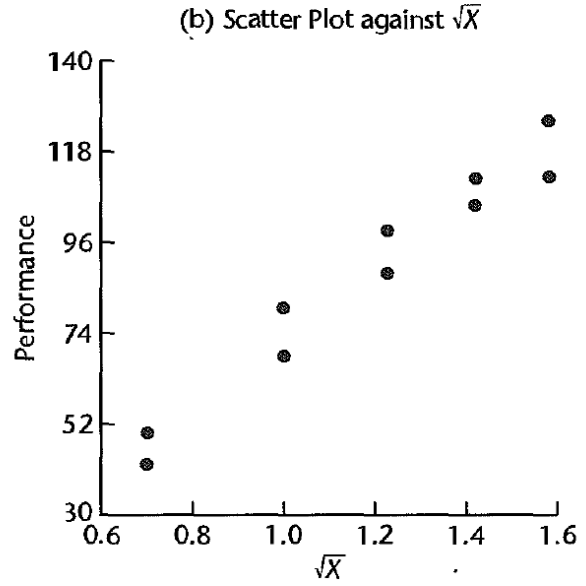
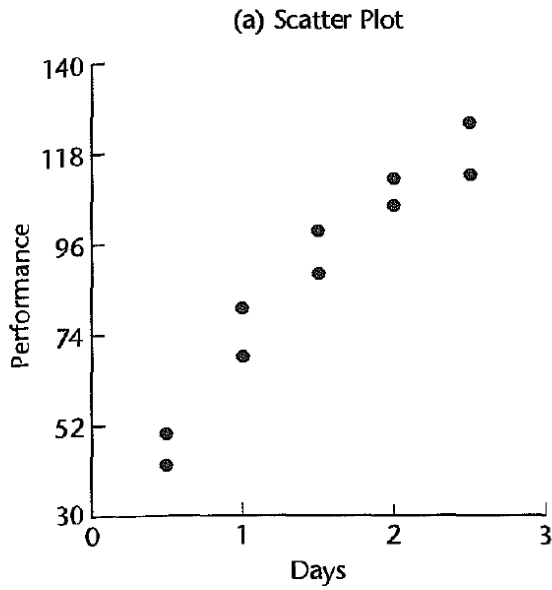
Form this model, we see that

- 1- No strong indications of substantial departures from normality are indicated by this plot.
- 2- The coefficient of correlation between the ordered residuals and their expected values under normality, .979. is substantially larger and supports the reasonableness of normal error terms. Thus, the new simple linear regression model appears to be appropriate here for the transformed data.

The fitted regression function in the original units of X can easily be obtained, if desired:

$$Y = -10.33 + 83.4S \sqrt{x}$$

Scatter Plots and Residual Plots—Sales Training Example.



Diagnostics and Remedial Measures: Continue.

Recall:

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2.$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i, \quad K_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \sum_{i=1}^n L_i Y_i, \quad L_i = \frac{1}{n} - \bar{X} K_i$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right],$$

The unbiased estimate of σ^2 is

$$s^2 = MSE = \hat{\sigma}^2 = \frac{SSE}{n-2}, \quad SSE = \sum_{i=1}^n e_i^2, \quad e_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \dots$$

When a regression model, such as the simple linear regression model is considered for an application, we can usually not be certain in advance that the model is appropriate for that application. Anyone, or several, of the features of the model (conditions), such as linearity of the regression function or normality of the error terms, may not be appropriate for the particular data at hand. Hence, it is important to examine the aptness of the model for the data before inferences based on that model are undertaken. In this part, we discuss some simple graphic methods for studying the appropriateness of a model.

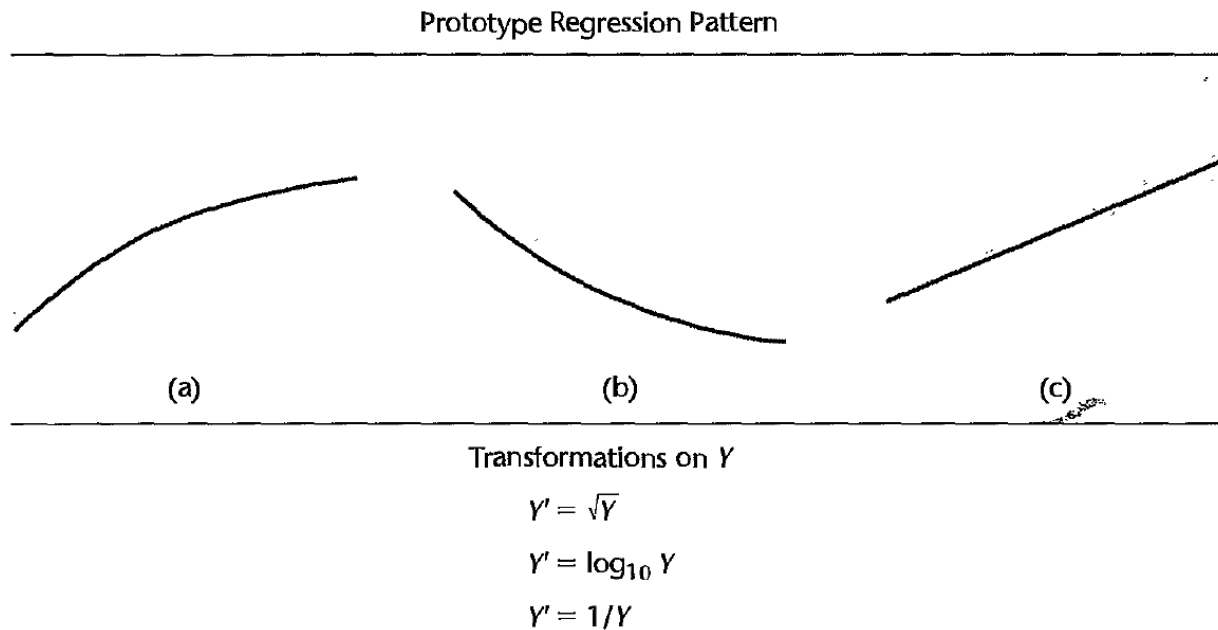
We also consider some remedial techniques that can be helpful when the data are not in accordance with the conditions of regression model. One of these techniques is the transformations.

Transformations for Nonnormality and Unequal Error Variances

Unequal error variances and nonnormality of the error terms frequently appear together. To remedy these departures from the simple linear regression model, we need a transformation on Y , since the shapes and spreads of the distributions of Y need to be changed. Such a transformation on Y may also at the same time help to linearize a curvilinear regression relation. At other times, a simultaneous transformation on X may be needed to obtain or maintain a linear regression relation.

Frequently, the nonnormality and unequal variances departures from regression model take the form of increasing skewness and increasing variability of the distributions of the error terms as the mean response $E\{Y\}$ increases. For example, in a regression of yearly household expenditures for vacations (Y) on household income (X), there will tend to be more variation and greater positive skewness (i.e., some very high yearly vacation expenditures) for high-income households than for low-income households, who tend to consistently spend much less for vacations.

The figure below also presents some simple transformations on Y that may be helpful for these cases. Several alternative transformations on Y may be tried, as well as some simultaneous transformations on X . Scatter plots and residual plots should be prepared to determine the most effective transformation(s).



Example (page 132)

Data on age (X) and plasma level of a polyamine (Y) for a portion of the 25 healthy children in a study are presented in columns 1 and 2 of Table 3.8. These data are plotted in Figure 3.16a as a scatter plot. Note the distinct curvilinear regression relationship, as well as the greater variability for younger children than for older ones.

Child <i>i</i>	(1) Age X_i	(2) Plasma Level Y_i	(3) $Y'_i = \log_{10} Y_i$
1	0 (newborn)	13.44	1.1284
2	0 (newborn)	12.84	1.1086
3	0 (newborn)	11.91	1.0759
4	0 (newborn)	20.09	1.3030
5	0 (newborn)	15.60	1.1931
6	1.0	10.11	1.0048
7	1.0	11.38	1.0561
...
19	3.0	6.90	.8388
20	3.0	6.77	.8306
21	4.0	4.86	.6866
22	4.0	5.10	.7076
23	4.0	5.67	.7536
24	4.0	5.75	.7597
25	4.0	6.23	.7945

On the basis of the prototype regression pattern in Figure 3.15b, we shall first try the logarithmic transformation $Y' = \log_{10}(Y)$. The transformed Y values are shown in column 3 of Table 3.8. Figure 3.16b contains the scatter plot with this transformation. Note that the transformation not only has led to a reasonably linear regression relation, but the variability at the different levels of X also has become reasonably constant. To further examine the reasonableness of the transformation $Y' = \log_{10}(Y)$, we fitted the

simple linear regression model (2.1) to the transformed Y data and obtained:

$$Y' = 1.135 - .1023X$$

A plot of the residuals against X is shown in Figure 3.16c, and a normal probability plot of the residuals is shown in Figure 3.16d. The coefficient of correlation between the ordered residuals and their expected values under normality is .981. For $\alpha = .05$, Table B.6 indicates that the critical value is .959 so that the observed coefficient supports the assumption of normality of the error terms. All of this evidence supports the appropriateness of regression model (2.1) for the transformed Y data.

F- Test for Lack of Fit

In this part, we take up a formal test for determining whether a specific type of regression function adequately fits the data.

We illustrate this test for ascertaining whether a linear regression function is a good fit for the data under some assumptions.

Assumptions

The lack of fit test assumes that the observations Y for given X are:

- (1) independent of errors
- (2) the error is normally distributed
- (3) The distributions of Y have the same variance σ^2

The lack of fit test requires repeat, observations at one or more X levels. These data can be obtained as:

- 1- In non-experimental data, these may occur by chance, as when in a productivity study relating workers' output and

age, several workers of the same age happen to be included in the study.

2- In an experiment, one can assure by design that there are repeat observations. Repeat trials for the same level of the predictor variable, of the type described, are called replications. The resulting observations are called replicates.

Example [see text book page.120]

In an experiment involving 12 similar but scattered suburban branch offices of a commercial bank, holders of checking accounts at the offices were offered gifts for setting up money market accounts. Minimum initial deposits in the new money market account were specified to qualify for the gift. The value of the gift was directly proportional to the specified minimum deposit. Various levels of minimum deposit and related gift values were used in the experiment in order to ascertain the relation between the specified minimum deposit and gift value, on the one hand, and number of accounts opened at the office, on the other. Altogether, six levels of minimum deposit and proportional gift value were used, with two of the branch offices assigned at random to each level. One branch office had a fire during the period and was dropped from the study. The data are given below:

Branch	Size of Minimum Deposit (dollars)	Number of New Accounts	Branch	Size of Minimum Deposit (dollars)	Number of New Accounts
i	X_i	Y_i	i	X_i	Y_i
1	125	160	7	75	42
2	100	112	8	175	124
3	200	124	9	125	150
4	75	28	10	200	104
5	150	152	11	100	136
6	175	156			

The simple regression model is estimated as

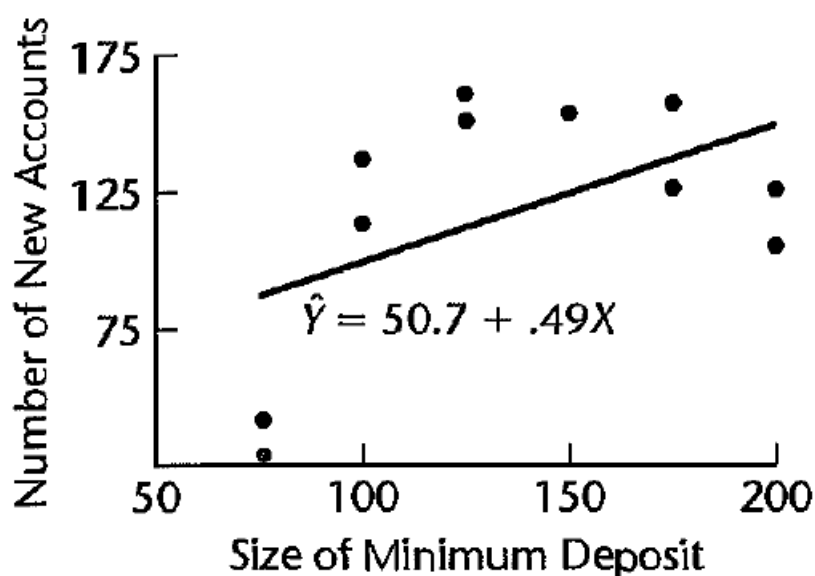
$$Y = 50.72251 + 0.48670X$$

R-results

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	50.7225	39.3979	1.287	0.23
x	0.4867	0.2747	1.772	0.11
Residual standard error: 40.47 on 9 degrees of freedom				
Multiple R-squared: 0.2586, Adjusted R-squared: 0.1762				
F-statistic: 3.139 on 1 and 9 DF, p-value: 0.1102				

> summary(aov(model))

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x	1	5141	5141	3.139	0.11
Residuals	9	14742	1638		



A scatter plot, together with the fitted regression line, and the other results indications that a linear regression function is inappropriate. Here we can use the **for Lack of Fit** to test this.

F-Test for lake of fit

Let the simple regression model (the reduced model) in the form

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, \dots, n,$$

While the true model (Full model) in the form

$$Y_i = \mu_i + \varepsilon_i, i = 1, \dots, n,$$

$Y_i = \mu_i + \varepsilon_i, i = 1, \dots, n \Rightarrow$ the true model and μ_i might not be

$$\beta_0 + \beta_1 x_i .$$

Now if the used model is not the true model ($\mu_i \neq \beta_0 + \beta_1 x_i$), then

$$\hat{y}_i = b_0 + b_1 x_i \text{ based on the simple linear regression model cannot be an}$$

accurate predicted. value of u_i .

$$\text{Thus, } e_i = y_i - \hat{y}_i = \mu_i - b_0 - b_1 x_i + \varepsilon_i = s_i \neq \varepsilon_i .$$

Then, the mean residual sum of squares

$$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} = \frac{\sum_{i=1}^n s_i^2}{n-2} \neq \frac{\sum_{i=1}^n \varepsilon_i^2}{n-2}$$

is no longer a sensible estimate of σ^2 .

To resolve this problem, we could try to obtain repeat observations with respect to the same covariate. Let

$$y_{11}, y_{12}, \dots, y_{1n_1} \Rightarrow n_1 \text{ repeated observation at } x_1;$$

$$y_{21}, y_{22}, \dots, y_{2n_2} \Rightarrow n_2 \text{ repeated observation at } x_2;$$

⋮

$$y_{c1}, y_{c2}, \dots, y_{cn_c} \Rightarrow n_m \text{ repeated observation at } x_c;$$

Note: $\sum_{j=1}^c n_j = n .$

In view of the repeated observation the full model can be formatted as:

$$Y_{ij} = \mu_j + \varepsilon_{ij}$$

where:

μ_j are parameters $j = 1, \dots, c$

ε_{ij} are independent $N(0, \sigma^2)$

Since the error terms have expectation zero, it follows that:

$$E\{Y_{ij}\} = \mu_j$$

Thus, the parameter μ_j ($j = 1, \dots, c$) is the mean response when $X = X_j$.

To fit the full model to the data, we require the least squares or maximum likelihood estimators for the parameters μ_j . It can be shown that these

estimators of μ_j are simply the sample means \bar{Y}_j as follows

$$Q = \sum_{i=1}^{n_j} \sum_{j=1}^c \varepsilon_{ij}^2 = \sum_{i=1}^{n_j} \sum_{j=1}^c (Y_{ij} - \mu_j)^2$$

$$\frac{\partial}{\partial \mu_j} Q = 0 \Rightarrow -2 \sum_{i=1}^{n_j} (Y_{ij} - \mu_j) = 0 \Rightarrow n_j \mu_j = \sum_{i=1}^{n_j} (Y_{ij})$$

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (Y_{ij}) = \bar{Y}_j$$

Thus, the estimated expected value for observation Y_{ij} is \bar{Y}_j , and the error sum of squares for the full model therefore is:

$$SSE(F) = \sum_j \sum_i (Y_{ij} - \bar{Y}_j)^2 = SSPE$$

In the context of the test for lack of fit, the full model error sum of squares is called the *pure error sum of squares* and is denoted by *SSPE*.

Note that $SSPE$ is made up of the sums of squared deviations at each X level. At level $X = X_j$, this sum of squared deviations is:

$$\sum_i (Y_{ij} - \bar{Y}_j)^2$$

The degrees of freedom of the $SSPE$ is

$$df_F = \sum_j (n_j - 1) = \sum_j n_j - c = n - c$$

The reduced model

The general linear test approach next requires consideration of the reduced model under H_0 . For testing the appropriateness of a linear regression relation, the alternatives are:

$$H_0: E\{Y\} = \beta_0 + \beta_1 X$$

$$H_a: E\{Y\} \neq \beta_0 + \beta_1 X$$

Thus, H_0 postulates that μ_j in the full model is linearly related to X_j :

$$\mu_j = \beta_0 + \beta_1 X_j$$

The reduced model under H_0 therefore is:

$$Y_{ij} = \beta_0 + \beta_1 X_j + \varepsilon_{ij} \quad \text{Reduced model}$$

Hence, the error sum of squares for the reduced model is the usual error sum of squares SSE :

$$\begin{aligned} SSE(R) &= \sum \sum [Y_{ij} - (b_0 + b_1 X_j)]^2 \\ &= \sum \sum (Y_{ij} - \hat{Y}_{ij})^2 = SSE \end{aligned}$$

We also know that the degrees of freedom associated with $SSE(R)$ are:

$$df_R = n - 2$$

Test Statistic

The general linear test statistic

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

here becomes:

$$F^* = \frac{SSE - SSPE}{(n - 2) - (n - c)} \div \frac{SSPE}{n - c}$$

The difference between the two error sums of squares is called the *lack of fit sum of squares* here and is denoted by *SSLF*:

$$SSLF = SSE - SSPE$$

We can then express the test statistic as follows:

$$\begin{aligned} F^* &= \frac{SSLF}{c - 2} \div \frac{SSPE}{n - c} \\ &= \frac{MSLF}{MSPE} \end{aligned}$$

where *MSLF* denotes the *lack of fit mean square* and *MSPE* denotes the *pure error mean square*.

We know that large values of F^* lead to conclusion H_a in the general linear test. Decision rule

If $F^* \leq F(1 - \alpha; c - 2, n - c)$, conclude H_0

If $F^* > F(1 - \alpha; c - 2, n - c)$, conclude H_a

Then, we summarize the test steps as follows:

Let the given data as follows

X_1	X_1	X_c
-------	-------	-------	-------

Y_{11}	Y_{21}		Y_{c1}
Y_{12}	Y_{22}		Y_{c2}
\vdots	\vdots		\vdots
Y_{1n_1}	Y_{2n_2}		Y_{cn_c}
\bar{Y}_1	\bar{Y}_2	\bar{Y}_c

From the data we calculate the following quantities

$$\underbrace{Y_{ij} - \hat{Y}_{ij}}_{\text{Error deviation}} = \underbrace{Y_{ij} - \bar{Y}_j}_{\text{Pure error deviation}} + \underbrace{\bar{Y}_j - \hat{Y}_{ij}}_{\text{Lack of fit deviation}}$$

hence

$$\sum \sum (Y_{ij} - \hat{Y}_{ij})^2 = \sum \sum (Y_{ij} - \bar{Y}_j)^2 + \sum \sum (\bar{Y}_j - \hat{Y}_{ij})^2$$

$$SSE = SSPE + SSLF$$

$$\hat{Y}_{ij} = b_0 + b_1 X_j$$

$$\bar{y} = \frac{\sum_{j=1}^c \sum_{i=1}^{n_j} y_{ji}}{n}, \quad \bar{Y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}, \quad SSR = \sum_{j=1}^c n_j (\hat{y}_j - \bar{y})^2$$

we can define the lack of fit sum of squares directly as follows:

$$SSLF = \sum \sum (\bar{Y}_j - \hat{Y}_{ij})^2$$

Since all Y_{ij} observations at the level X_j have the same fitted value, which we can denote by \hat{Y}_j , we can express

$$SSLF = \sum_j n_j (\bar{Y}_j - \hat{Y}_j)^2$$

Step 1:

$$H_0 : E(y_{ji}) = \beta_0 + \beta_1 x_j \quad \text{vs} \quad H_1 : E(y_{ji}) \neq \beta_0 + \beta_1 x_j$$

Step 2:

Calculate the test statistic F^* from the ANOVA table

Step 3:

Calculate the critical region $F(1-\alpha, c-2, n-c)$, where $c-2$ and $n-c$ are the degrees of freedom (from ANOVA table)

Step 4:**Decision:**

We know that large values of F^* lead to conclusion H_a in the general linear test. Decision rule

$$\text{If } F^* \leq F(1 - \alpha; c - 2, n - c), \text{ conclude } H_0$$

$$\text{If } F^* > F(1 - \alpha; c - 2, n - c), \text{ conclude } H_a$$

All what we need now is to construct the ANOVA Table as follows:

The ANOVA table

Source of Variation	SS	df	MS
Regression	$SSR = \sum \sum (\hat{Y}_{ij} - \bar{Y})^2$	1	$MSR = \frac{SSR}{1}$
Error	$SSE = \sum \sum (Y_{ij} - \hat{Y}_{ij})^2$	$n - 2$	$MSE = \frac{SSE}{n - 2}$
Lack of fit	$SSLF = \sum \sum (\bar{Y}_j - \hat{Y}_{ij})^2$	$c - 2$	$MSLF = \frac{SSLF}{c - 2}$
Pure error	$SSPE = \sum \sum (Y_{ij} - \bar{Y}_j)^2$	$n - c$	$MSPE = \frac{SSPE}{n - c}$
Total	$SSTO = \sum \sum (Y_{ij} - \bar{Y})^2$	$n - 1$	

Example [see text book page.120]

In this example, we see

Replicate	Size of Minimum Deposit (dollars)					
	$j = 1$ $X_1 = 75$	$j = 2$ $X_2 = 100$	$j = 3$ $X_3 = 125$	$j = 4$ $X_4 = 150$	$j = 5$ $X_5 = 175$	$j = 6$ $X_6 = 200$
$j = 1$	28	112	160	152	156	124
$j = 2$	42	136	150		124	104
Mean \bar{Y}_j	35	124	155	152	140	114

These sums of squares are then added over all of the X levels ($j = 1, \dots, c$).

For the bank example, we have:

$$SSPE = (28 - 35)^2 + (42 - 35)^2 + (112 - 124)^2 + (136 - 124)^2 + (160 - 155)^2 + (150 - 155)^2 + (152 - 152)^2 + (156 - 140)^2$$

$$+ (124 - 140)^2 + (124 - 114)^2 + (104 - 114)^2$$

$$= 1148$$

$$df_f = n - c = 11 - 6 = 5$$

$$SSE(R) = SSE = 14741.6, \quad df_R = n - 2 = 9$$

$$SSLF = SSE - SSPE = 14741.6 - 1148 = 13593.6$$

Then ANOVA Table is

Source	df	SS	MS	F
SSR	1	5141	5141	$F_1 = 5141/1637.956 = 3.138668$
SSE	$n - 2 = 9$	14741.6	1637.956	
Lack of fit	$c - 2 = 6 - 4 = 4$	13593.6	3398.4	$F^* = 3398.4/229.6 = 14.80139$
Pure error	$n - c = 11 - 6 = 5$	1148	229.6	
SSTOT	10	19882.6		

(b) Bank Example

Source of Variation	SS	df	MS
Regression	5,141.3	1	5,141.3
Error	14,741.6	9	1,638.0
Lack of fit	13,593.6	4	3,398.4
Pure error	1,148.0	5	229.6
Total	19,882.9	10	

If the level of significance is to be $\alpha = .01$, we require $F(.99; 4, 5) = 11.4$.

Since

$F^* = MSLF/MSPE = 13593.6/1148 = 14.80 > 11.4$, we conclude H_0 , that the regression function is not linear.

This, of course, accords with our visual impression from scatter plot. The P-value for the test is .006 \rightarrow reject \rightarrow simple linear model is not good fit for the given data.


```
install.packages("olsrr")
library(olsrr)
x=c(125,100,200,75,150,175,75,175,125,200,100)
y=c(160,112,124,28,152,156,42,124,150,104,136)
fit=lm(y~x)
ols_pure_error_anova(fit)
```

```
> library(olsrr)
> x=c(125,100,200,75,150,175,75,175,125,200,100)
> y=c(160,112,124,28,152,156,42,124,150,104,136)
> fit=lm(y~x)
> ols_pure_error_anova(fit)
Lack of Fit F Test
-----
Response : y
Predictor: x
```

Analysis of Variance Table					
	DF	Sum Sq	Mean Sq	F Value	Pr(>F)
x	1	5141.338	5141.338	22.39259	0.001070765
Residual	9	14741.57	1637.952		
Lack of fit	4	13593.57	3398.393	14.80136	0.005593812
Pure Error	5	1148.00	229.60		

$$F = \frac{(RSS(\text{model 2}) - RSS(\text{model m})) / (m - 2)}{RSS(\text{model m}) / (n - m)} = \frac{\sum_{j=1}^m n_j (\bar{y}_j - \hat{y}_j)^2}{\sum_{j=1}^m \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2}$$

In general, we use the following procedure to fit simple regression model when the data contain repeated observations.

1. Fit the model, write down the usual analysis of variance table. Do **not** perform an F-test for regression ($H_0: \beta_1 = 0$).
2. Perform the F-test for lack of fit. There are two possibilities.
 - (a) If significant lack of fit, stop the analysis of the model fitting

and seek ways to improve the model by examining residuals.

(b) If lack of fit test is not significant, carry out an F-test for regression, obtain confidence interval and so on. The residuals should still be plotted and examined for peculiarities.

Example:

X	Y
90	81,83
79	75
66	68,60,62
51	60,64
35	51,53

$$\sum_{i=1}^{10} x_i = 629, \sum_{i=1}^{10} y_i = 657, \sum_{i=1}^{10} x_i^2 = 43161, \sum_{i=1}^{10} y_i^2 = 44249, \sum_{i=1}^{10} x_i y_i = 43189.$$

Thus, total sum of squares:

$$\sum_{i=1}^{10} (y_i - \bar{y})^2 = \sum_{i=1}^{10} y_i^2 - 10\bar{y}^2 = 44249 - 10(65.7)^2 = 1084.1.$$

$$\Rightarrow b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^{10} x_i y_i - 10\bar{x}\bar{y}}{\sum_{i=1}^{10} x_i^2 - 10\bar{x}^2} = \frac{43189 - 10 * 62.9 * 65.7}{43161 - 10 * (62.9)^2} = 0.51814$$

$$\Rightarrow \text{regression sum of squares} = b_1^2 S_{XX} = 965.65636$$

$$\Rightarrow \text{residual sum of squares (reduced model)} = 1084.1 - 965.66 = 118.44$$

Pure error sum of squares:

X

$$90: \bar{Y}_1 = \frac{81+83}{2} = 82, \sum_{i=1}^2 (Y_{1i} - \bar{Y}_1)^2 = (81-82)^2 + (83-82)^2 = 2.$$

$$79: (75-75)^2 = 0$$

$$66: \bar{Y}_3 = \frac{68+60+62}{3} = 63.33, \sum_{i=1}^3 (Y_{3i} - \bar{Y}_3)^2 = (68-63.33)^2 + (60-63.33)^2$$

$$+ (62-63.33)^2 = 34.67$$

$$51: \bar{Y}_4 = \frac{60+64}{2} = 62, \sum_{i=1}^2 (Y_{4i} - \bar{Y}_4)^2 = (60-62)^2 + (64-62)^2 = 8.$$

$$35: \bar{Y}_5 = \frac{51+53}{2} = 52, \sum_{i=1}^2 (Y_{5i} - \bar{Y}_5)^2 = (51-52)^2 + (53-52)^2 = 2$$

Then, pure error sum of squares = 2+0+34.67+8+2=46.67

Lack of fit sum of squares = 118.44-46.67=71.77

Source	df	SS	MS
Regression	1	965.66	965.66
Lack of fit	3	71.77	23.92
Pure error	5	46.67	9.33
Total	9	1084.1	

$$\Rightarrow F = \frac{23.92}{9.33} = 2.56 < f_{3,5,0.05} = 5.41$$

\Rightarrow Not significant!! That is, the simple linear regression is adequate. The standard F-test for regression can be carried out.

Matrix Approach to Simple Linear Regression Analysis

Recall:

In the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

$$E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i \neq j.$$

Then

$$E(Y_i) = \beta_0 + \beta_1 X_i \quad \text{and} \quad \text{Var}(Y_i) = \sigma^2.$$

The point estimates of β_0, β_1 are

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = b_0 = \bar{Y} - b_1 \bar{X}$$

$$\hat{\beta}_1 = b_1 = \sum_{i=1}^n K_i Y_i, \quad K_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \sum_{i=1}^n L_i Y_i, \quad L_i = \frac{1}{n} - \bar{X} K_i$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}, \quad \text{Var}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} \right],$$

The unbiased estimate of σ^2 is

$$s^2 = \text{MSE} = \hat{\sigma}^2 = \frac{\text{SSE}}{n-2}, \quad \text{SSE} = \sum_{i=1}^n e_i^2, \quad e_i = Y_i - \hat{Y}_i, \quad i = 1, 2, \dots$$

Basic Definitions of the matrices and operations

The matrix A of with r rows and c columns will be represented either in full

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{bmatrix}$$

or in abbreviated form:

$$\mathbf{A} = [a_{ij}] \quad i = 1, \dots, r; j = 1, \dots, c$$

or simply by a boldface symbol, such as \mathbf{A} .

Transpose

The transpose of a matrix \mathbf{A} is another matrix, denoted by \mathbf{A}' , that is obtained by interchanging corresponding columns and rows of the matrix \mathbf{A} .

For example, if:

$$\mathbf{A}_{3 \times 2} = \begin{bmatrix} 2 & 5 \\ 7 & 10 \\ 3 & 4 \end{bmatrix}$$

then the transpose \mathbf{A}' is:

$$\mathbf{A}'_{2 \times 3} = \begin{bmatrix} 2 & 7 & 3 \\ 5 & 10 & 4 \end{bmatrix}$$

Equality of Matrices

Two matrices **A** and **B** are said to be equal if they have the same dimension and if all corresponding elements are equal. Conversely, if two matrices are equal, their corresponding elements are equal. For example, if:

$$\mathbf{A} = \begin{matrix} 3 \times 1 \\ \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \end{matrix} \quad \mathbf{B} = \begin{matrix} 3 \times 1 \\ \begin{bmatrix} 4 \\ 7 \\ 3 \end{bmatrix} \end{matrix}$$

then $\mathbf{A} = \mathbf{B}$ implies:

$$a_1 = 4 \quad a_2 = 7 \quad a_3 = 3$$

Similarly, if:

$$\mathbf{A} = \begin{matrix} 3 \times 2 \\ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \end{matrix} \quad \mathbf{B} = \begin{matrix} 3 \times 2 \\ \begin{bmatrix} 17 & 2 \\ 14 & 5 \\ 13 & 9 \end{bmatrix} \end{matrix}$$

then $\mathbf{A} = \mathbf{B}$ implies:

$$\begin{aligned} a_{11} &= 17 & a_{12} &= 2 \\ a_{21} &= 14 & a_{22} &= 5 \\ a_{31} &= 13 & a_{32} &= 9 \end{aligned}$$

Regression Examples

In regression analysis, one basic matrix is the vector **Y**, consisting of the n observations on the response variable:

$$\mathbf{Y} = \begin{matrix} n \times 1 \\ \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \end{matrix}$$

Note that the transpose \mathbf{Y}' is the row vector:

$$\mathbf{Y}' = \begin{matrix} 1 \times n \\ [Y_1 \quad Y_2 \quad \cdots \quad Y_n] \end{matrix}$$

Another basic matrix in regression analysis is the **X** matrix, which is defined as follows for simple linear regression analysis:

Another basic matrix in regression analysis is the \mathbf{X} matrix, which is defined as follows for simple linear regression analysis:

$$\mathbf{X}_{n \times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$$

The matrix \mathbf{X} consists of a column of 1s and a column containing the n observations on the predictor variable X . Note that the transpose of \mathbf{X} is:

$$\mathbf{X}'_{2 \times n} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix}$$

The \mathbf{X} matrix is often referred to as the *design matrix*.

Matrix Addition and Subtraction

Adding or subtracting two matrices requires that they have the same dimension. The sum, or difference, of two matrices is another matrix whose elements each consist of the sum, or difference, of the corresponding elements of the two matrices. Suppose:

$$\mathbf{A}_{3 \times 2} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \quad \mathbf{B}_{3 \times 2} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$$

then:

$$\mathbf{A} + \mathbf{B}_{3 \times 2} = \begin{bmatrix} 1+1 & 4+2 \\ 2+2 & 5+3 \\ 3+3 & 6+4 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 4 & 8 \\ 6 & 10 \end{bmatrix}$$

Similarly:

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} 1-1 & 4-2 \\ 2-2 & 5-3 \\ 3-3 & 6-4 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 0 & 2 \end{bmatrix}$$

In general, if:

$$\mathbf{A} = [a_{ij}]_{r \times c} \quad \mathbf{B} = [b_{ij}]_{r \times c} \quad i = 1, \dots, r; j = 1, \dots, c$$

then:

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]_{r \times c} \quad \text{and} \quad \mathbf{A} - \mathbf{B} = [a_{ij} - b_{ij}]_{r \times c}$$

Formula (1) generalizes in an obvious way to addition and subtraction of more than two matrices. Note also that $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, as in ordinary algebra. \square

Regression Examples

The regression model:

$$Y_i = E\{Y_i\} + \varepsilon_i \quad i = 1, \dots, n$$

can be written compactly in matrix notation. First, let us define the vector of the mean responses:

$$\mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \vdots \\ E\{Y_n\} \end{bmatrix}$$

and the vector of the error terms:

$$\mathbf{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Recalling the definition of the observations vector \mathbf{Y} , we can write the regression model as follows:

$$\mathbf{Y} = \mathbf{E}\{\mathbf{Y}\} + \boldsymbol{\varepsilon}$$

$n \times 1 \quad n \times 1 \quad n \times 1$

because:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \vdots \\ E\{Y_n\} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\} + \varepsilon_1 \\ E\{Y_2\} + \varepsilon_2 \\ \vdots \\ E\{Y_n\} + \varepsilon_n \end{bmatrix}$$

Thus, the observations vector \mathbf{Y} equals the sum of two vectors, a vector containing the expected values and another containing the error terms.

Multiplication of a Matrix by a Scalar

A *scalar* is an ordinary number or a symbol representing a number. In multiplication of a matrix by a scalar, every element of the matrix is multiplied by the scalar. For example, suppose the matrix \mathbf{A} is given by:

$$\mathbf{A} = \begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix}$$

Then $4\mathbf{A}$, where 4 is the scalar, equals:

$$4\mathbf{A} = 4 \begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 8 & 28 \\ 36 & 12 \end{bmatrix}$$

Similarly, $k\mathbf{A}$ equals:

$$k\mathbf{A} = k \begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 2k & 7k \\ 9k & 3k \end{bmatrix}$$

where k denotes a scalar.

Matrix multiplications

Here is an example of matrix multiplication:

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix} \end{aligned}$$

In general, if \mathbf{A} has dimension $r \times c$ and \mathbf{B} has dimension $c \times s$, the product \mathbf{AB} is a matrix of dimension $r \times s$ whose element in the i th row and j th column is:

$$\sum_{k=1}^c a_{ik} b_{kj}$$

so that:

$$\mathbf{AB}_{r \times s} = \left[\sum_{k=1}^c a_{ik} b_{kj} \right] \quad i = 1, \dots, r; j = 1, \dots, s$$

Thus, in the foregoing example, the element in the first row and second column of the product \mathbf{AB} is:

$$\sum_{k=1}^3 a_{1k} b_{k2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}$$

as indeed we found by taking the cross products of the elements in the first row of \mathbf{A} and second column of \mathbf{B} and summing.

Regression Examples

A product frequently needed is $\mathbf{Y}'\mathbf{Y}$, where \mathbf{Y} is the vector of observations on the response variable as defined in

$$\mathbf{Y}'\mathbf{Y} = \underset{1 \times 1}{[Y_1 \quad Y_2 \quad \cdots \quad Y_n]} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = [Y_1^2 + Y_2^2 + \cdots + Y_n^2] = \left[\sum Y_i^2 \right]$$

Note that $\mathbf{Y}'\mathbf{Y}$ is a 1×1 matrix, or a scalar. We thus have a compact way of writing a sum of squared terms: $\mathbf{Y}'\mathbf{Y} = \sum Y_i^2$.

We also will need $\mathbf{X}'\mathbf{X}$, which is a 2×2 matrix, \mathbf{v}

$$\mathbf{X}'\mathbf{X} = \underset{2 \times 2}{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix}} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

and $\mathbf{X}'\mathbf{Y}$, which is a 2×1 matrix:

$$\mathbf{X}'\mathbf{Y} = \underset{2 \times 1}{\begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix}} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Identity Matrix. The identity matrix or unit matrix is denoted by \mathbf{I} . It is a diagonal matrix whose elements on the main diagonal are all 1s. Premultiplying or postmultiplying any $r \times r$ matrix \mathbf{A} by the $r \times r$ identity matrix \mathbf{I} leaves \mathbf{A} unchanged. For example:

$$\mathbf{IA} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Similarly, we have:

$$\mathbf{AI} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Note that the identity matrix \mathbf{I} therefore corresponds to the number 1 in ordinary algebra, since we have there that $1 \cdot x = x \cdot 1 = x$.

In general, we have for any $r \times r$ matrix \mathbf{A} :

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

Vector and Matrix with All Elements Unity

A column vector with all elements 1 will be denoted by $\mathbf{1}$:

$$\mathbf{1}_{r \times 1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

and a square matrix with all elements 1 will be denoted by \mathbf{J} :

$$\mathbf{J}_{r \times r} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

For instance, we have:

$$\mathbf{1}_{3 \times 1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{J}_{3 \times 3} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Note that for an $n \times 1$ vector $\mathbf{1}$ we obtain:

$$\mathbf{1}'\mathbf{1} = [1 \ \cdots \ 1] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = [n] = n$$

and:

$$\mathbf{1}\mathbf{1}' = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} [1 \ \cdots \ 1] = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \mathbf{J}_{n \times n}$$

Zero Vector

A zero vector is a vector containing only zeros. The zero column vector will be denoted by $\mathbf{0}$:

$$\mathbf{0}_{r \times 1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Finding the Inverse

Up to this point, the inverse of a matrix \mathbf{A} has been given, and we have only checked to make sure it is the inverse by seeing whether or not $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. But how does one find the inverse, and when does it exist?

An inverse of a square $r \times r$ matrix exists if the rank of the matrix is r . Such a matrix is said to be *nonsingular* or of *full rank*. An $r \times r$ matrix with rank less than r is said to be *singular* or *not of full rank*, and does not have an inverse. The inverse of an $r \times r$ matrix of full rank also has rank r .

Finding the inverse of a matrix can often require a large amount of computing. We shall take the approach in this book that the inverse of a 2×2 matrix and a 3×3 matrix can be calculated by hand. For any larger matrix, one ordinarily uses a computer to find the inverse, unless the matrix is of a special form such as a diagonal matrix. It can be shown that the inverses for 2×2 and 3×3 matrices are as follows:

1. If:

$$\mathbf{A}_{2 \times 2} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then:

$$\mathbf{A}_{2 \times 2}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{d}{D} & \frac{-b}{D} \\ \frac{-c}{D} & \frac{a}{D} \end{bmatrix}$$

where:

$$D = ad - bc$$

D is called the *determinant* of the matrix \mathbf{A} . If \mathbf{A} were singular, its determinant would equal zero and no inverse of \mathbf{A} would exist.

2. If:

$$\mathbf{B}_{3 \times 3} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$$

then:

$$\mathbf{B}_{3 \times 3}^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}^{-1} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix}$$

Example

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}$$

We have:

$$\begin{aligned} a &= 2 & b &= 4 \\ c &= 3 & d &= 1 \end{aligned}$$

$$D = ad - bc = 2(1) - 4(3) = -10$$

Hence:

$$\mathbf{A}^{-1} = \begin{bmatrix} \frac{1}{-10} & \frac{-4}{-10} \\ \frac{-3}{-10} & \frac{2}{-10} \end{bmatrix} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix}$$

Regression Example

The principal inverse matrix encountered in regression analysis is the inverse of the matrix $\mathbf{X}'\mathbf{X}$

$$\mathbf{X}'\mathbf{X}_{2 \times 2} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

Using rule , we have:

$$\begin{aligned} a &= n & b &= \sum X_i \\ c &= \sum X_i & d &= \sum X_i^2 \end{aligned}$$

so that:

$$D = n \sum X_i^2 - \left(\sum X_i \right) \left(\sum X_i \right) = n \left[\sum X_i^2 - \frac{(\sum X_i)^2}{n} \right] = n \sum (X_i - \bar{X})^2$$

Hence:

$$\left(\mathbf{X}'\mathbf{X} \right)_{2 \times 2}^{-1} = \begin{bmatrix} \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} & \frac{-\sum X_i}{n \sum (X_i - \bar{X})^2} \\ \frac{-\sum X_i}{n \sum (X_i - \bar{X})^2} & \frac{n}{n \sum (X_i - \bar{X})^2} \end{bmatrix}$$

Since $\sum X_i = n\bar{X}$ and $\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$, we can simplify

$$\left(\mathbf{X}'\mathbf{X} \right)_{2 \times 2}^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\sum (X_i - \bar{X})^2} & \frac{1}{\sum (X_i - \bar{X})^2} \end{bmatrix}$$

Some Basic Results for Matrices

We list here, without proof, some basic results for matrices which we will utilize in later work.

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

$$\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB}$$

$$k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$$

$$(\mathbf{A}')' = \mathbf{A}$$

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$$

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$$

$$(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$$

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

Random Vectors and Matrices

Expectation of Random Vector or Matrix

Suppose we have $n = 3$ observations in the observations vector \mathbf{Y} .

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

The expected value of \mathbf{Y} is a vector, denoted by $\mathbf{E}\{\mathbf{Y}\}$, that is defined as follows:

$$\mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ E\{Y_3\} \end{bmatrix}$$

Regression Example

Suppose the number of cases in a regression application is $n = 3$. The three error terms ε_1 , ε_2 , ε_3 each have expectation zero. For the error terms vector:

$$\mathbf{\varepsilon}_{3 \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

we have:

$$\mathbf{E}\{\mathbf{\varepsilon}\}_{3 \times 1} = \mathbf{0}_{3 \times 1}$$

since:

$$\begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ E\{\varepsilon_3\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Variance-Covariance Matrix of Random Vector

If we have p random variables we can put them into a random vector as

$Y = [Y_1, \dots, Y_n]'$, then

$$\text{Var}(Y) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}, \text{ where}$$

$$\sigma_{ij} = \sigma_{ji} = \text{cov}(Y_i, Y_j) \quad \text{and} \quad \sigma_{ii} = \text{var}(Y_i).$$

Regression Example

$$\text{Var}(\varepsilon) = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_{pp} \end{bmatrix} = \sigma^2 I.$$

Some Basic Results

Frequently, we shall encounter a random vector \mathbf{W} that is obtained by premultiplying the random vector \mathbf{Y} by a constant matrix \mathbf{A} (a matrix whose elements are fixed):

$$\mathbf{W} = \mathbf{A}\mathbf{Y}$$

Some basic results for this case are:

$$\mathbf{E}\{\mathbf{A}\} = \mathbf{A}$$

$$\mathbf{E}\{\mathbf{W}\} = \mathbf{E}\{\mathbf{A}\mathbf{Y}\} = \mathbf{A}\mathbf{E}\{\mathbf{Y}\}$$

$$\text{Var}(\mathbf{A}\mathbf{W}) = \mathbf{A} \text{var}(\mathbf{W}) \mathbf{A}'.$$

Simple Linear Regression Model in Matrix Terms

We are now ready to develop simple linear regression in matrix terms. Remember again that we will not present any new results, but shall only state in matrix terms the results obtained earlier. We begin with the normal error regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i = 1, \dots, n$$

This implies:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned}$$

Let us repeat these definitions and also define the β vector of the regression

$$\mathbf{Y}_{n \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X}_{n \times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \boldsymbol{\beta}_{2 \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \boldsymbol{\varepsilon}_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Then

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

$n \times 1 \quad n \times 2 \quad 2 \times 1 \quad n \times 1$

since:

$$\begin{aligned} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} &= \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \\ &= \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 X_n + \varepsilon_n \end{bmatrix} \end{aligned}$$

and the conditions are

$$E(\boldsymbol{\varepsilon}) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad \text{and} \quad \text{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$$

Then

$$E(\mathbf{Y}) = \mathbf{X} \boldsymbol{\beta} \quad \text{and} \quad \text{Var}(\mathbf{Y}) = \sigma^2 \mathbf{I}$$

Matrix Approach to Simple Linear Regression Analysis

Least squared estimation

The simple linear regression model can be formed in the matrix form as

$$Y = X \beta + \varepsilon, \quad E(\varepsilon) = 0, \text{Var}(\varepsilon) = \sigma^2 I,$$

where

$$\mathbf{Y}_{n \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X}_{n \times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \boldsymbol{\beta}_{2 \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \boldsymbol{\varepsilon}_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Now, we apply the least squared method to

to find the estimation of the vector β as follows:

$$Q = \sum_{i=1}^n e_i^2 = e'e = (Y - X\beta)'(Y - X\beta)$$

Expanding, we obtain:

$$Q = Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta$$

since $(X\beta)' = \beta'X'$ Note now that $Y'X\beta$ is 1×1 , hence is equal to its transpose, is $\beta'X'Y$. Thus, we find:

$$Q = Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

To find the value of β that minimizes Q , we differentiate with respect to β_0 and β_1 . Let:

$$\frac{\partial}{\partial \beta}(Q) = \begin{bmatrix} \frac{\partial Q}{\partial \beta_0} \\ \frac{\partial Q}{\partial \beta_1} \end{bmatrix}$$

Then it follows that:

$$\frac{\partial}{\partial \beta}(Q) = -2X'Y + 2X'X\beta$$

Equating to the zero vector, we get

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

where

$$(\mathbf{X}'\mathbf{X})_{2 \times 2}^{-1} = \begin{bmatrix} \frac{\sum X_i^2}{n \sum (X_i - \bar{X})^2} & \frac{-\sum X_i}{n \sum (X_i - \bar{X})^2} \\ \frac{-\sum X_i}{n \sum (X_i - \bar{X})^2} & \frac{n}{n \sum (X_i - \bar{X})^2} \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y}_{2 \times 1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

Example

We shall use matrix methods to obtain the estimated regression coefficients for the Toluca Company example. The data On the Y and X Variables. Using these data, we define the Y observations vector and the X matrix as follows:

$$\mathbf{Y} = \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix}$$

We now require the following matrix products:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 80 & 30 & \dots & 70 \end{bmatrix} \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix} = \begin{bmatrix} 25 & 1,750 \\ 1,750 & 142,300 \end{bmatrix}$$

we find the inverse of $\mathbf{X}'\mathbf{X}$:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix}$$

In subsequent matrix calculations utilizing this inverse matrix and other matrix results, we shall actually utilize more digits for the matrix elements than are shown.

$$\begin{aligned} \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} = \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix} \begin{bmatrix} 7,807 \\ 617,180 \end{bmatrix} \\ &= \begin{bmatrix} 62.37 \\ 3.5702 \end{bmatrix} \end{aligned}$$

or $b_0 = 62.37$ and $b_1 = 3.5702$. These results agree with the ones in Chapter 1. Any differences would have been due to rounding effects.

One can use R code direct to get the estimates as:

```
mat <- scan('a.txt')
```

```
mat <- matrix(mat, ncol = 2, byrow = TRUE)
```

```
mat[,1]
```

```
mat[,2]
```

```
length(mat[,1])
```

```
one=as.vector(rep(1, 25))
```

```
x=cbind(one,mat[,1])
```

```
y=mat[,2]
```

```
b=solve(t(x)%*%x)%*%t(x)%*%y
```

```
b
```

Recall:

The simple linear regression model can be formed in the matrix form as

$$\begin{aligned}
 Y &= X \beta + \varepsilon, \\
 E(\varepsilon) &= 0, \text{Var}(\varepsilon) = \sigma^2 I, \\
 E(Y) &= X \beta, \text{Var}(Y) = \sigma^2 I,
 \end{aligned}$$

where

$$\mathbf{Y}_{n \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X}_{n \times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \boldsymbol{\beta}_{2 \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \boldsymbol{\varepsilon}_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

The Fitted Values and Residuals

Let the vector of the fitted values \hat{Y}_i be denoted by $\hat{\mathbf{Y}}$:

$$\hat{\mathbf{Y}}_{n \times 1} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$$

In matrix notation, we then have:

$$\hat{\mathbf{Y}}_{n \times 1} = \mathbf{X}_{n \times 2} \mathbf{b}_{2 \times 1}$$

because:

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \hat{b}_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix}$$

Example

For the Toluca Company example, we obtain the vector of fitted values using the matrices

$$\hat{\mathbf{Y}} = \mathbf{Xb} = \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix} \begin{bmatrix} 62.37 \\ 3.5702 \end{bmatrix} = \begin{bmatrix} 347.98 \\ 169.47 \\ \vdots \\ 312.28 \end{bmatrix}$$

It can be easily calculated using R as: `fits=x0%*%b`

Hat Matrix. We can express the matrix result for \hat{Y}

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

or, equivalently:

$$\hat{Y}_{n \times 1} = H_{n \times n} Y_{n \times 1}$$

where:

$$H_{n \times n} = X(X'X)^{-1}X'$$

The matrix H is symmetric and has the special property (called idempotency):

$$HH=H$$

In general, a matrix M is said to be idempotent if $MM = M$.

Residuals

Let the vector of the residuals $e_i = Y_i - \hat{Y}_i$ be denoted by \mathbf{e} :

$$\mathbf{e}_{n \times 1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

In matrix notation, we then have:

$$\mathbf{e}_{n \times 1} = \mathbf{Y}_{n \times 1} - \hat{\mathbf{Y}}_{n \times 1} = \mathbf{Y}_{n \times 1} - \mathbf{X}_{n \times 1} \mathbf{b}_{n \times 1}$$

For the Toluca Company example, we obtain the vector of the residuals by using the results

$$\mathbf{e} = \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} - \begin{bmatrix} 347.98 \\ 169.47 \\ \vdots \\ 312.28 \end{bmatrix} = \begin{bmatrix} 51.02 \\ -48.47 \\ \vdots \\ 10.72 \end{bmatrix}$$

It can be easily calculated using R as: Res= y-fits

Variance-Covariance Matrix of Residuals. The residuals e_i , like the fitted values \hat{Y}_i , can be expressed as linear combinations of the response variable observations Y_i , using the result for $\hat{\mathbf{Y}}$:

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

We thus have the important result:

$$\mathbf{e}_{n \times 1} = (\mathbf{I}_{n \times n} - \mathbf{H}_{n \times n}) \mathbf{Y}_{n \times 1}$$

Now

$$\begin{aligned}
 \text{Var}(e) &= \text{Var}((I - H)Y) = (I - H) \text{var}(Y)(I - H)' \\
 &= (I - H)\sigma^2 I (I - H)' \\
 &= \sigma^2 (I - H)I (I - H)' \\
 &= \sigma^2 (I - H) = \text{MSE} (I - H),
 \end{aligned}$$

where

$$\text{MSE} = \text{SSE} / (n - 2) = e'e / (n - 2).$$

$$\text{SSE} = e'e.$$

Analysis of Variance Results

Sums of Squares

To see how the sums of squares are expressed in matrix notation, we begin with the total sum of squares $SSTO$, It will be convenient to use an algebraically equivalent expression:

$$SSTO = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

We know from (5.13) that:

$$\mathbf{Y}'\mathbf{Y} = \sum Y_i^2$$

The subtraction term $(\sum Y_i)^2/n$ in matrix form uses \mathbf{J} , the matrix of 1s

$$\frac{(\sum Y_i)^2}{n} = \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y}$$

For instance, if $n = 2$, we have:

$$\left(\frac{1}{2}\right) [Y_1 \quad Y_2] \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \frac{(Y_1 + Y_2)(Y_1 + Y_2)}{2}$$

Hence, it follows that:

$$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y}$$

Just as $\sum Y_i^2$ is represented by $\mathbf{Y}'\mathbf{Y}$ in matrix terms, so $SSE = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$ can be represented as follows:

$$SSE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$

hence

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}.$$

Finally, it can be shown that:

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right) \mathbf{Y}'\mathbf{J}\mathbf{Y}$$

Remark: $SSTO = SSE + SSR$

Example

For Toluca Company example, find the SSE, SSTO and SSR

$$\mathbf{Y}'\mathbf{Y} = [399 \quad 121 \quad \dots \quad 323] \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} = 2,745,173$$

$$\mathbf{b}'\mathbf{X}'\mathbf{Y} = [62.37 \quad 3.5702] \begin{bmatrix} 7,807 \\ 617,180 \end{bmatrix} = 2,690,348$$

Hence:

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} = 2,745,173 - 2,690,348 = 54,825$$

$$\begin{aligned} SSTO &= \mathbf{Y}'\mathbf{Y} - \frac{1}{n}(\mathbf{Y}'\mathbf{J}\mathbf{Y}) = 2745173 - 2437970 \\ &= 307203 \end{aligned}$$

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \frac{1}{n}(\mathbf{Y}'\mathbf{J}\mathbf{Y}) = 2690348 - 243790 = 252378$$

Remark:

$$SSTO = t(y)^2 - t(y)^2 J^2 / 25$$

$$SSR = t(b)^2 t(x)^2 - t(y)^2 J^2 / 25$$

Sums of Squares as Quadratic Forms

The ANOVA sums of squares can be shown to be *quadratic forms*. An example of a quadratic form of the observations Y_i when $n = 2$ is:

$$5Y_1^2 + 6Y_1Y_2 + 4Y_2^2$$

Note that this expression is a second-degree polynomial containing terms involving the squares of the observations and the cross product. We can express this in matrix terms as follows:

$$[Y_1 \quad Y_2]' \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{Y}'\mathbf{A}\mathbf{Y}$$

where \mathbf{A} is a symmetric matrix.

In general, a quadratic form is defined as:

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} Y_i Y_j \quad \text{where } a_{ij} = a_{ji}$$

\mathbf{A} is a symmetric $n \times n$ matrix and is called the *matrix of the quadratic form*.

The ANOVA sums of squares $SSTO$, SSE , and SSR are all quadratic forms, as can be seen by reexpressing $\mathbf{b}'\mathbf{X}'$. From (5.71), we know,

$$\mathbf{b}'\mathbf{X}' = (\mathbf{X}\mathbf{b})' = \hat{\mathbf{Y}}'$$

We now use the result in (5.73) to obtain:

$$\mathbf{b}'\mathbf{X}' = (\mathbf{H}\mathbf{Y})'$$

Since \mathbf{H} is a symmetric matrix so that $\mathbf{H}' = \mathbf{H}$, we finally obtain

$$\mathbf{b}'\mathbf{X}' = \mathbf{Y}'\mathbf{H}$$

This result enables us to express the ANOVA sums of squares as follows:

$$SSTO = \mathbf{Y}' \left[\mathbf{I} - \left(\frac{1}{n} \right) \mathbf{J} \right] \mathbf{Y}$$

$$SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$SSR = \mathbf{Y}' \left[\mathbf{H} - \left(\frac{1}{n} \right) \mathbf{J} \right] \mathbf{Y}$$

Each of these sums of squares can now be seen to be of the form $\mathbf{Y}'\mathbf{A}\mathbf{Y}$, where the three \mathbf{A} matrices are:

$$\mathbf{I} - \left(\frac{1}{n} \right) \mathbf{J}$$

$$\mathbf{I} - \mathbf{H}$$

$$\mathbf{H} - \left(\frac{1}{n} \right) \mathbf{J}$$

Recall:

The simple linear regression model can be formed in the matrix form as

$$\begin{aligned}
 Y &= X \beta + \varepsilon, \\
 E(\varepsilon) &= 0, \text{Var}(\varepsilon) = \sigma^2 I, \\
 E(Y) &= X \beta, \text{Var}(Y) = \sigma^2 I,
 \end{aligned}$$

where

$$\mathbf{Y}_{n \times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X}_{n \times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \boldsymbol{\beta}_{2 \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \boldsymbol{\varepsilon}_{n \times 1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

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

$$\begin{aligned}
 \text{Var}(e) &= \text{Var}((\mathbf{I} - \mathbf{H})\mathbf{Y}) = (\mathbf{I} - \mathbf{H})\text{var}(\mathbf{Y})(\mathbf{I} - \mathbf{H})' \\
 &= (\mathbf{I} - \mathbf{H})\sigma^2\mathbf{I}(\mathbf{I} - \mathbf{H})' \\
 &= \sigma^2(\mathbf{I} - \mathbf{H})\mathbf{I}(\mathbf{I} - \mathbf{H})' \\
 &= \sigma^2(\mathbf{I} - \mathbf{H}) = \text{MSE}(\mathbf{I} - \mathbf{H}),
 \end{aligned}$$

$$MSE = SSE / (n - 2) = e'e / (n - 2).$$

$$SSE = e'e$$

$$SSE = Y'Y - b'X'Y.$$

$$SSTO = Y'Y - \frac{1}{n}(Y'JY)$$

$$SSR = b'X'Y - \frac{1}{n}(Y'JY)$$

$$SSTO = Y' \left[\mathbf{I} - \left(\frac{1}{n} \right) \mathbf{J} \right] Y$$

$$SSE = Y'(\mathbf{I} - \mathbf{H})Y$$

$$SSR = Y' \left[\mathbf{H} - \left(\frac{1}{n} \right) \mathbf{J} \right] Y$$

Each of these sums of squares can now be seen to be of the form $Y'AY$, where the three A matrices are:

$$\mathbf{I} - \left(\frac{1}{n} \right) \mathbf{J}$$

$$\mathbf{I} - \mathbf{H}$$

$$\mathbf{H} - \left(\frac{1}{n} \right) \mathbf{J}$$

Inferences in Regression Analysis

Lemma

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

Proof

$$\begin{aligned} E(\hat{\beta}) &= E(X'X)^{-1}X'Y = (X'X)^{-1}X'E(Y) \\ &= (X'X)^{-1}X'XB \\ &= I\beta \\ &= \beta. \end{aligned}$$

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

$$\begin{aligned}
\text{Var}(\hat{\beta}) &= \text{Var}[(X'X)^{-1}X'Y] \\
&= (X'X)^{-1}X'\text{Var}(Y)[(X'X)^{-1}X']' \\
&= (X'X)^{-1}X'\sigma^2[(X'X)^{-1}X']' \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}I \\
&= \sigma^2(X'X)^{-1} \\
&= \text{MSE}(X'X)^{-1}.
\end{aligned}$$

This can be written as

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

$$= \begin{bmatrix} \frac{\sigma^2}{n} + \frac{\sigma^2\bar{X}^2}{\sum(X_i - \bar{X})^2} & \frac{-\bar{X}\sigma^2}{\sum(X_i - \bar{X})^2} \\ \frac{-\bar{X}\sigma^2}{\sum(X_i - \bar{X})^2} & \frac{\sigma^2}{\sum(X_i - \bar{X})^2} \end{bmatrix}$$

Or

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{MSE} (X'X)^{-1} \\ &= \begin{bmatrix} \frac{\text{MSE}}{n} + \frac{\bar{X}^2 \text{MSE}}{\sum (X_i - \bar{X})^2} & \frac{-\bar{X} \text{MSE}}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X} \text{MSE}}{\sum (X_i - \bar{X})^2} & \frac{\text{MSE}}{\sum (X_i - \bar{X})^2} \end{bmatrix} \end{aligned}$$

Example:

For Toluca Company example by matrix methods, we get

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \sigma^2 (X'X)^{-1} \\ &= \text{MSE}(\mathbf{X}'\mathbf{X})^{-1} = 2,384 \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix} \\ &= \begin{bmatrix} 685.34 & -8.428 \\ -8.428 & .12040 \end{bmatrix} \end{aligned}$$

Mean Response

To estimate the mean response at X_h , let us define the vector:

$$\mathbf{X}_h = \begin{bmatrix} 1 \\ X_h \end{bmatrix} \quad \text{or} \quad \mathbf{X}'_h = [1 \quad X_h]$$

The fitted value in matrix notation then is:

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

since:

$$\mathbf{X}'_h \mathbf{b} = [1 \quad X_h] \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [b_0 + b_1 X_h] = [\hat{Y}_h] = \hat{Y}_h$$

Note that $\mathbf{X}'_h \mathbf{b}$ is a 1×1 matrix; hence, we can write the final result as a scalar.

$$\begin{aligned} \text{Var}(\hat{Y}_h) &= \text{Var}(X'_h \mathbf{b}) = \text{Var}[X'_h (X'X)^{-1} X'Y] \\ &= X'_h (X'X)^{-1} X' \text{Var}(Y) [X'_h (X'X)^{-1} X']' \\ &= X'_h (X'X)^{-1} X' \sigma^2 [X'_h (X'X)^{-1} X']' \\ &= \sigma^2 X'_h (X'X)^{-1} X' X (X'X)^{-1} X_h \\ &= \sigma^2 X'_h (X'X)^{-1} X_h \\ &= \text{MSE}(X'_h (X'X)^{-1} X_h). \end{aligned}$$

Example

For Toluca Company example, the variance of the mean of the response when $X=65$ can be calculated using the matrix form as

$$\begin{aligned} \text{Var}(\hat{Y}_h) &= X'_h \text{Var}(b) X_h \\ &= [1 \quad 65] \begin{bmatrix} 685.34 & -8.428 \\ -8.428 & .12040 \end{bmatrix} \begin{bmatrix} 1 \\ 65 \end{bmatrix} = 98.37 \end{aligned}$$

Which is the same result as that obtained before.

Prediction of New Observation

Proceeding similarly, we get

$$\text{Var}(\hat{Y}_{new}) = MSE (1 + X'_h (X'X)^{-1} X_h).$$

Multiple Linear Regression

In This chapter, we generalized the simple linear regression model as

General Linear Regression Model

In general, the variables X_1, \dots, 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:

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

where:

$\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters

$X_{i1}, \dots, X_{i,p-1}$ are known constants

ε_i are independent $N(0, \sigma^2)$

$i = 1, \dots, n$

If we let $X_{i0} \equiv 1$, regression model can be written as follows:

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

where $X_{i0} \equiv 1$, or:

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \quad \text{where } X_{i0} \equiv 1$$

The response function for regression model is, since $E\{\varepsilon_i\} = 0$:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1}$$

This model can be specialized for different cases as follows

1- Simple linear model when $p=2$

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon,$$

2- Model with some Qualitative Predictor Variables

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon,$$

This model can be used in different applications such as:

The first-order regression model then is as follows:

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

where:

$$\begin{aligned} X_{i1} &= \text{patient's age} \\ X_{i2} &= \begin{cases} 1 & \text{if patient female} \\ 0 & \text{if patient male} \end{cases} \end{aligned}$$

3- Polynomial regression

Polynomial Regression. Polynomial regression models are special cases of the general linear regression model. They contain squared and higher-order terms of the predictor variable(s), making the response function curvilinear. The following is a polynomial regression model with one predictor variable:

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

4- Transformed Variables

Transformed Variables. Models with transformed variables involve complex, curvilinear response functions, yet still are special cases of the general linear regression model. Consider the following model with a transformed Y variable:

$$\log Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

Here, the response surface is complex, yet model can still be treated as a general linear regression model. If we let $Y'_i = \log Y_i$, we can write regression model

$$Y'_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

which is in the form of general linear regression model. The response variable just happens to be the logarithm of Y .

Many models can be transformed into the general linear regression model. For instance, the model:

$$Y_i = \frac{1}{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i}$$

can be transformed to the general linear regression model by letting $Y'_i = 1/Y_i$. We then have:

$$Y'_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

5- Interaction Effects

Interaction Effects. When the effects of the predictor variables on the response variable are not additive, the effect of one predictor variable depends on the levels of the other predictor variables. The general linear regression model encompasses regression models with nonadditive or interacting effects. An example of a nonadditive regression model with two predictor variables X_1 and X_2 is the following:

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

Here, the response function is complex because of the interaction term $\beta_3 X_{i1} X_{i2}$. Yet regression model is a special case of the general linear regression model. Let $X_{i3} = X_{i1} X_{i2}$ and then write

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

6-Combination of Cases

Combination of Cases. A regression model may combine several of the elements we have just noted and still be treated as a general linear regression model. Consider the following regression model containing linear and quadratic terms for each of two predictor variables and an interaction term represented by the cross-product term:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \varepsilon_i$$

Let us define:

$$Z_{i1} = X_{i1} \quad Z_{i2} = X_{i1}^2 \quad Z_{i3} = X_{i2} \quad Z_{i4} = X_{i2}^2 \quad Z_{i5} = X_{i1} X_{i2}$$

We can then write regression model (6.16) as follows:

$$Y_i = \beta_0 + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \beta_4 Z_{i4} + \beta_5 Z_{i5} + \varepsilon_i$$

General Linear Regression Model in Matrix Terms

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:

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

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

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:

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

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

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

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

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

Estimation of Regression Coefficients

The least squares criterion is generalized as follows for general linear regression model

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

The least squares estimators are those values of $\beta_0, \beta_1, \dots, \beta_{p-1}$ that minimize Q . Let us denote the vector of the least squares estimated regression coefficients b_0, b_1, \dots, b_{p-1} as \mathbf{b} :

$$\mathbf{b}_{p \times 1} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}$$

The least squares normal equations for the general linear regression model

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

Hence

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

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

The prove is similar to the simple linear model in the matrix form.

Example: Multiple Regression with Two Predictor Variables (Dwaine Studios)

Dwaine Studios, Inc., operates portrait studios in 21 cities of medium size. These studios specialize in portraits of children. The company is considering an expansion into other cities of medium size and wishes to investigate whether sales (Y)- thousands- in a community can be predicted from the number of persons aged 16 or younger in the community (X_1)- thousands- and the per capita disposable personal income in the community (X_2)- *thousands*. Data on these variables for the most recent year for the 21 cities in which Dwaine Studios is now operating are shown below:

city	X1	X2	Y
1	68.5	16.7	174.4
2	45.2	16.8	164.4
3	91.3	18.2	244.2
4	47.8	16.3	154.6
5	46.9	17.3	181.6
6	66.1	18.2	207.5
7	49.5	15.9	152.8
8	52	17.2	163.2
9	48.9	16.6	145.4
10	38.4	16	137.2
11	87.9	18.3	241.9
12	72.8	17.1	191.1
13	88.4	17.4	232
14	42.9	15.8	145.3
15	52.5	17.8	161.1
16	85.7	18.4	209.7
17	41.3	16.5	146.4
18	51.7	16.3	144
19	89.6	18.1	232.6
20	82.7	19.1	224.1
21	52.3	16	166.5

```
# 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)
summary(model)
```

then we obtain

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

The model coefficients can be interpreted as:

- 1- There is -68.9 of the sales when X1 and X2 are zeros
- 2- When in population size increases by one unit (thousand), the sales increases by 1.46 thousand with fixed income.
- 3- When in income increases by one unit (thousand), the sales increases by 9.37 thousand with fixed population size.

Multiple Linear Regression

General Linear Regression Model

In general, the variables X_1, \dots, 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:

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

where:

$\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters

$X_{i1}, \dots, X_{i,p-1}$ are known constants

ε_i are independent $N(0, \sigma^2)$

$i = 1, \dots, n$

To express general linear regression model

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
 E(\hat{\beta}) &= E(X'X)^{-1}X'Y = (X'X)^{-1}X'E(Y) \\
 &= (X'X)^{-1}X'XB \\
 &= I\beta \\
 &= \beta.
 \end{aligned}$$

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

$$\begin{aligned}
 Var(\hat{\beta}) &= Var[(X'X)^{-1}X'Y] \\
 &= (X'X)^{-1}X'Var(Y)[(X'X)^{-1}X']' \\
 &= (X'X)^{-1}X'\sigma^2[(X'X)^{-1}X']' \\
 &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
 &= \sigma^2(X'X)^{-1}I \\
 &= \sigma^2(X'X)^{-1} \\
 &= MSE(X'X)^{-1}.
 \end{aligned}$$

Example: (Dwayne 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

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

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

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

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

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

Example: (Dwayne Studios)

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

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

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

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

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

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

Multiple Linear Regression

General Linear Regression Model

In general, the variables X_1, \dots, 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:

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

where:

$\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters

$X_{i1}, \dots, X_{i,p-1}$ are known constants

ε_i are independent $N(0, \sigma^2)$

$i = 1, \dots, n$

To express general linear regression model

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

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

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

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

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

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

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

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

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

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

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

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

$$H_0 : \beta_i = \beta_i^{(0)} , i = 0,1,2,\dots,p$$

$$H_1 : \beta_i \neq (> \text{ or } <)\beta_i^{(0)}$$

Step 2: The test statistic

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

Step 3: The Critical regions

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

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

Step 4: The decision: Reject H_0 , 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: (Dwayne Studios)

Test the significance of coefficients in the Dwayne Studio data (use $\alpha = 5\%$ 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

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

Step 4:

The test statistic belongs to the acceptance region, then accept H_0 .

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

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

Step 4:

The test statistic belongs to the rejection region, then reject H_0 .

Testing β_2

Step 1: The hypotheses

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

Step 2:

The test statistic

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

Step 3:

The Critical regions

The critical region in this case is

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

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

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$

$$H_1 : \beta_i \neq \beta_j \text{ for } i \neq j$$

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 H_0 .

Example: (Dwayne Studios)

Test the significance of model in the Dwayne Studio data (use $\alpha = 5\%$ 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

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

and.

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

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 H_0 . 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

Dwayne 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

$$\begin{aligned} s^2\{\hat{Y}_h\} &= \mathbf{X}'_h \mathbf{s}^2\{\mathbf{b}\} \mathbf{X}_h \\ &= [1 \quad 65.4 \quad 17.6] \begin{bmatrix} 3,602.0 & 8.748 & -241.43 \\ 8.748 & .0448 & -.679 \\ -241.43 & -.679 & 16.514 \end{bmatrix} \begin{bmatrix} 1 \\ 65.4 \\ 17.6 \end{bmatrix} \\ &= 7.656 \end{aligned}$$

Then

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

90% CI for the mean of Y is

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

$$191.10 \pm 2.101 (2.77)$$

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

Prediction Limits for New Observations

Dwayne 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}

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

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

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

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

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

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

Similarly, For City B, we have

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

R code

```
mydata=read.table("Dwayne 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)
```

Multiple Regression II

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:

$\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters

$X_{i1}, \dots, X_{i,p-1}$ are known constants

ε_i are independent $N(0, \sigma^2)$

$i = 1, \dots, n$

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:

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

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

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

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

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

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

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

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

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

Estimation of Regression Coefficients

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

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

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

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

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

The Extra Sum of Squares

An extra sum of squares measures the marginal reduction in the error sum of squares when one or several predictor variables are added to the regression model, given that other predictor variables are already in the model. Equivalently, one can view an extra sum of squares as measuring the marginal increase in the regression sum of squares when one or several predictor variables are added to the regression model. We first utilize an example to illustrate these ideas, and then we present definitions of extra sums of squares and discuss a variety of uses of extra sums of squares in tests about regression coefficients.

Example (Book: page 256) Body fat example,

From the example, we define

$$SSR(X_1|X_2) = SSE(X_2) - SSE(X_1, X_2)$$

or, equivalently:

$$SSR(X_1|X_2) = SSR(X_1, X_2) - SSR(X_2)$$

If X_2 is the extra variable, we define:

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2)$$

or, equivalently:

$$SSR(X_2|X_1) = SSR(X_1, X_2) - SSR(X_1)$$

Extensions for three or more variables are straightforward. For example, we define:

$$SSR(X_3|X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3)$$

or:

$$SSR(X_3|X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2)$$

and:

$$SSR(X_2, X_3|X_1) = SSE(X_1) - SSE(X_1, X_2, X_3)$$

or:

$$SSR(X_2, X_3|X_1) = SSR(X_1, X_2, X_3) - SSR(X_1)$$

and:

$$SSR(X_2, X_3|X_1) = SSE(X_1) - SSE(X_1, X_2, X_3)$$

Decomposition of SSR into Extra Sums of Squares

In multiple regression, unlike simple linear regression, we can obtain a variety of decompositions of the regression sum of squares SSR into extra sums of squares.

$$SSTO = SSR(X_1) + SSE(X_1)$$

where the notation now shows explicitly that X_1 is the X variable in the model. Replacing $SSE(X_1)$ by its equivalent

$$SSTO = SSR(X_1) + SSR(X_2|X_1) + SSE(X_1, X_2)$$

We now make use of the same identity for multiple regression with two X variables as for a single X variable, namely:

$$SSTO = SSR(X_1, X_2) + SSE(X_1, X_2)$$

Solving (7.7) for $SSE(X_1, X_2)$:

$$SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)$$

Of course, the order of the X variables is arbitrary. Here, we can also obtain the decomposition:

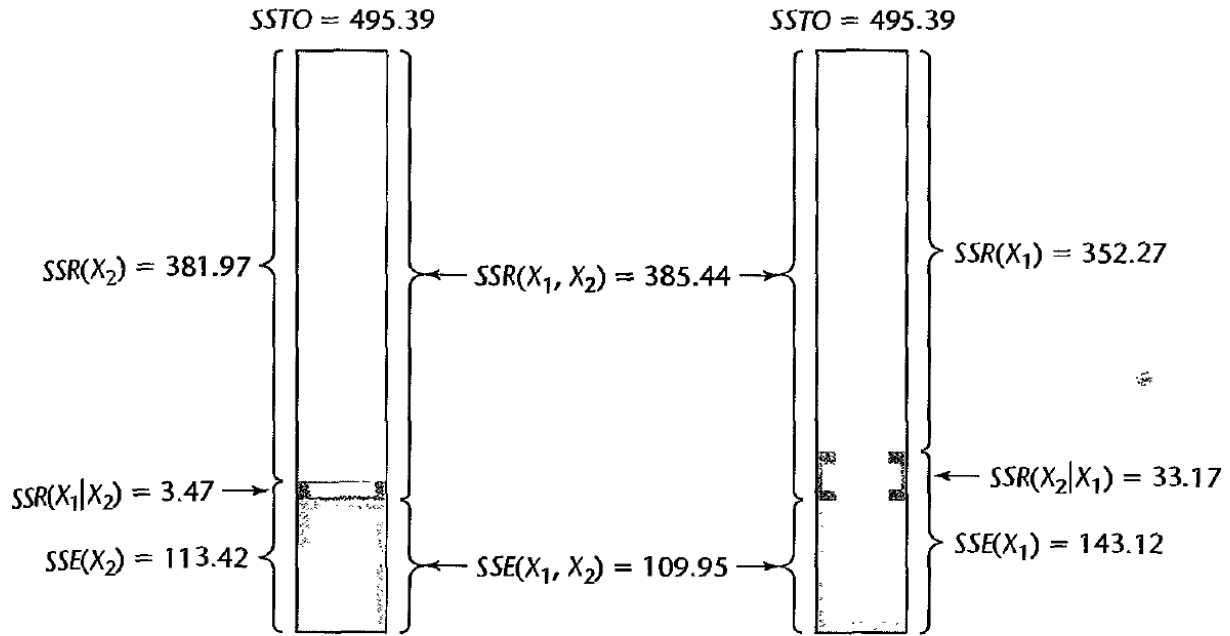
$$SSR(X_1, X_2) = SSR(X_2) + SSR(X_1|X_2)$$

When the regression model contains three X variables, a variety of decompositions of $SSR(X_1, X_2, X_3)$ can be obtained. We illustrate three of these:

$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2)$$

$$SSR(X_1, X_2, X_3) = SSR(X_2) + SSR(X_3|X_2) + SSR(X_1|X_2, X_3)$$

$$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2, X_3|X_1)$$



Example of ANOVA Table with Decomposition of SSR for Three X Variables.

Source of Variation	SS	df	MS
Regression	$SSR(X_1, X_2, X_3)$	3	$MSR(X_1, X_2, X_3)$
X_1	$SSR(X_1)$	1	$MSR(X_1)$
$X_2 X_1$	$SSR(X_2 X_1)$	1	$MSR(X_2 X_1)$
$X_3 X_1, X_2$	$SSR(X_3 X_1, X_2)$	1	$MSR(X_3 X_1, X_2)$
Error	$SSE(X_1, X_2, X_3)$	$n - 4$	$MSE(X_1, X_2, X_3)$
Total	$SSTO$	$n - 1$	

Uses of Extra Sums of Squares in Tests for Regression Coefficients

Test whether a Single $\beta_k = 0$

When we wish to test whether the term $\beta_k X_k$ can be dropped from a multiple regression model, we are interested in the alternatives:

$$H_0: \beta_k = 0$$

$$H_a: \beta_k \neq 0$$

*

We already know that test statistic

$$t^* = \frac{b_k}{s\{b_k\}}$$

is appropriate for this test.

We, now show that this can also be done using the extra sum of squares. Let us consider the first-order regression model with three predictor variables:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad \text{Full model}$$

To test the alternatives:

$$H_0: \beta_3 = 0$$

$$H_a: \beta_3 \neq 0$$

we fit the full model and obtain the error sum of squares $SSE(F)$. We now explicitly show the variables in the full model, as follows:

$$SSE(F) = SSE(X_1, X_2, X_3)$$

The degrees of freedom associated with $SSE(F)$ are $df_F = n - 4$ since there are four parameters in the regression function for the full model

The reduced model when H_0 holds is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad \text{Reduced model}$$

We next fit this reduced model and obtain:

$$SSE(R) = SSE(X_1, X_2)$$

There are $df_R = n - 3$ degrees of freedom associated with the reduced model.

The general linear test statistic

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

here becomes:

$$F^* = \frac{SSE(X_1, X_2) - SSE(X_1, X_2, X_3)}{(n - 3) - (n - 4)} \div \frac{SSE(X_1, X_2, X_3)}{n - 4}$$

Note that the difference between the two error sums of squares in the numerator term is the extra sum of squares

$$SSE(X_1, X_2) - SSE(X_1, X_2, X_3) = SSR(X_3|X_1, X_2)$$

Hence the general linear test statistic here is:

$$F^* = \frac{SSR(X_3|X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_3)}{n-4} = \frac{MSR(X_3|X_1, X_2)}{MSE(X_1, X_2, X_3)}$$

And this can be compared with the critical region $F(1, n-4)$ to have the decision.

Remark: F-statistic in this case also equal to (t-statistic)²

Example:

In the body fat example, can we remove the X3 from the model?

Solution

1- $H_0: B_3=0$ vs $H_1: B_3 \neq 0$

2-

$$\begin{aligned} F^* &= \frac{SSR(X_3|X_1, X_2)}{1} \div \frac{SSE(X_1, X_2, X_3)}{n-4} \\ &= \frac{11.54}{1} \div \frac{98.41}{16} = 1.88 \end{aligned}$$

3-

For $\alpha = .01$, we require $F(.99; 1, 16) = 8.53$. Since $F^* = 1.88 \leq 8.53$, we conclude H_0 , that X_3 can be dropped from the regression model that already contains X_1 and X_2 .

Remark: if we use t-test we see that

Since $(t^*)^2 = (-1.37)^2 = 1.88 = F^*$, we see that the two test statistics are equivalent, just as for simple linear regression.

Test whether Several Coefficients

In multiple regression we are frequently interested in whether several terms in the regression model can be dropped. For example, we may wish to know whether both $\beta_2 X_2$ and $\beta_3 X_3$ can be dropped from the full model. The alternatives here are:

$$H_0: \beta_2 = \beta_3 = 0$$

$$H_a: \text{not both } \beta_2 \text{ and } \beta_3 \text{ equal zero}$$

With the general linear test approach, the reduced model under H_0 is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \quad \text{Reduced model}$$

and the error sum of squares for the reduced model is:

$$SSE(R) = SSE(X_1)$$

This error sum of squares has $df_R = n - 2$ degrees of freedom associated with it.

The general linear test statistic (2.70) thus becomes here:

$$F^* = \frac{SSE(X_1) - SSE(X_1, X_2, X_3)}{(n - 2) - (n - 4)} \div \frac{SSE(X_1, X_2, X_3)}{n - 4}$$

Again the difference between the two error sums of squares in the numerator term is an extra sum of squares, namely:

$$SSE(X_1) - SSE(X_1, X_2, X_3) = SSR(X_2, X_3|X_1)$$

Hence, the test statistic becomes:

$$F^* = \frac{SSR(X_2, X_3|X_1)}{2} \div \frac{SSE(X_1, X_2, X_3)}{n - 4} = \frac{MSR(X_2, X_3|X_1)}{MSE(X_1, X_2, X_3)}$$

Note that $SSR(X_2, X_3|X_1)$ has two degrees of freedom associated with it, as we pointed out earlier.

Example:

In the body fat example, can we remove the X_2 and X_3 from the model?

Multiple Regression II

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:

$\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters

$X_{i1}, \dots, X_{i,p-1}$ are known constants

ε_i are independent $N(0, \sigma^2)$

$i = 1, \dots, n$

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:

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

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

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

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

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

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

Summary of Tests Concerning Regression Coefficients

Test whether All $\beta_k = 0$

This is the *overall F test* of whether or not there is a regression relation between the response variable Y and the set of X variables. The alternatives are:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$

$$H_a: \text{not all } \beta_k \text{ (} k = 1, \dots, p-1 \text{) equal zero}$$

and the test statistic is:

$$\begin{aligned} F^* &= \frac{SSR(X_1, \dots, X_{p-1})}{p-1} \div \frac{SSE(X_1, \dots, X_{p-1})}{n-p} \\ &= \frac{MSR}{MSE} \end{aligned}$$

If H_0 holds, $F^* \sim F(p-1, n-p)$. Large values of F^* lead to conclusion H_a .

Test whether a Single $B_k = 0$

This is a *partial F test* of whether a particular regression coefficient β_k equals zero. The alternatives are:

$$H_0: \beta_k = 0$$

$$H_a: \beta_k \neq 0$$

and the test statistic is:

$$\begin{aligned} F^* &= \frac{SSR(X_k | X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_{p-1})}{1} \div \frac{SSE(X_1, \dots, X_{p-1})}{n-p} \\ &= \frac{MSR(X_k | X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_{p-1})}{MSE} \end{aligned}$$

If H_0 holds, $F^* \sim F(1, n - p)$. Large values of F^* lead to conclusion H_a . Statistics packages that provide extra sums of squares permit use of this test without having to fit the reduced model.

An equivalent test statistic is

$$t^* = \frac{b_k}{s\{b_k\}}$$

If H_0 holds, $t^* \sim t(n - p)$. Large values of $|t^*|$ lead to conclusion H_a .

Test whether Some $B_k = 0$

This is another *partial F test*. Here, the alternatives are:

$$H_0: \beta_q = \beta_{q+1} = \dots = \beta_{p-1} = 0$$

$$H_a: \text{not all of the } \beta_k \text{ in } H_0 \text{ equal zero}$$

where for convenience, we arrange the model so that the last $p - q$ coefficients are the ones to be tested. The test statistic is:

$$\begin{aligned} F^* &= \frac{SSR(X_q, \dots, X_{p-1} | X_1, \dots, X_{q-1})}{p - q} \div \frac{SSE(X_1, \dots, X_{p-1})}{n - p} \\ &= \frac{MSR(X_q, \dots, X_{p-1} | X_1, \dots, X_{q-1})}{MSE} \end{aligned}$$

If H_0 holds, $F^* \sim F(p - q, n - p)$. Large values of F^* lead to conclusion H_a .

Remark:

The partial F^* for several $B_k=0$ can be formed in terms of R^2 as

$$\begin{aligned}
 F^* &= \frac{MSR(X_q, \dots, X_{p-1} | X_1, \dots, X_{q-1})}{MSE(X_1, \dots, X_{p-1})} = \frac{\frac{SSR(X_q, \dots, X_{p-1} | X_1, \dots, X_{q-1})}{p-q}}{\frac{SSE(X_1, \dots, X_{p-1})}{n-p}} \\
 &= \frac{\frac{SSR(X_1, \dots, X_{p-1}) - SSR(X_1, \dots, X_{q-1})}{p-q}}{\frac{SSE(X_1, \dots, X_{p-1})}{n-p}} \\
 &= \frac{\frac{SSR(X_1, \dots, X_{p-1}) - SSR(X_1, \dots, X_{q-1})}{(p-q)SST}}{\frac{SSE(X_1, \dots, X_{p-1})}{(n-p)SST}} \\
 &= \frac{\frac{R^2(X_1, \dots, X_{p-1}) - R^2(X_1, \dots, X_{q-1})}{(p-q)}}{\frac{1 - R^2(X_1, \dots, X_{p-1})}{(n-p)}} = \frac{\frac{R_F^2 - R_R^2}{df_R - df_F}}{\frac{1 - R_F^2}{df_F}}
 \end{aligned}$$

Coefficients of Partial determination

Two Predictor Variables

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

$SSE(X_2)$ measures the variation in Y when X_2 is included in the model. $SSE(X_1, X_2)$ measures the variation in Y when both X_1 and X_2 are included in the model. Hence, the relative marginal reduction in the variation in Y associated with X_1 when X_2 is already in the model is:

$$\frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSR(X_1|X_2)}{SSE(X_2)}$$

This measure is the coefficient of partial determination between Y and X_1 , given that X_2 is in the model. We denote this measure by $R_{Y1|2}^2$:

$$R_{Y1|2}^2 = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{SSR(X_1|X_2)}{SSE(X_2)}$$

Thus, $R_{Y1|2}^2$ measures the proportionate reduction in the variation in Y remaining after X_2 is included in the model that is gained by also including X_1 in the model.

The coefficient of partial determination between Y and X_2 , given that X_1 is in the model, is defined correspondingly:

Thus, $R_{Y2|1}^2$ measures the proportionate reduction in the variation in Y remaining after X_1 is included in the model that is gained by also including X_2 in the model.

The coefficient of partial determination between Y and X_2 , given that X_1 is in the model, is defined correspondingly:

$$R_{Y2|1}^2 = \frac{SSR(X_2|X_1)}{SSE(X_1)}$$

General Case

The generalization of coefficients of partial determination to three or more X variables in the model is immediate. For instance:

$$R_{Y1|23}^2 = \frac{SSR(X_1|X_2, X_3)}{SSE(X_2, X_3)}$$

$$R_{Y2|13}^2 = \frac{SSR(X_2|X_1, X_3)}{SSE(X_1, X_3)}$$

$$R_{Y3|12}^2 = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)}$$

$$R_{Y4|123}^2 = \frac{SSR(X_4|X_1, X_2, X_3)}{SSE(X_1, X_2, X_3)}$$

Note that in the subscripts to R^2 , the entries to the left of the vertical bar show in turn the variable taken as the response and the X variable being added. The entries to the right of the vertical bar show the X variables already in the model.

Example

For the body fat example, we can obtain a variety of coefficients of partial determination. Here are three (Tables 7.2 and 7.4):

$$R_{Y2|1}^2 = \frac{SSR(X_2|X_1)}{SSE(X_1)} = \frac{33.17}{143.12} = .232$$

$$R_{Y3|12}^2 = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)} = \frac{11.54}{109.95} = .105$$

$$R_{Y1|2}^2 = \frac{SSR(X_1|X_2)}{SSE(X_2)} = \frac{3.47}{113.42} = .031$$

We see that when X_2 is added to the regression model containing X_1 here, the error sum of squares $SSE(X_1)$ is reduced by 23.2 percent. The error sum of squares for the model containing both X_1 and X_2 is only reduced by another 10.5 percent when X_3 is added to the model. Finally, if the regression model already contains X_2 , adding X_1 reduces $SSE(X_2)$ by only 3.1 percent.

Coefficients of Partial Correlation

Example

For the body fat example, we have:

$$r_{Y2|1} = \sqrt{.232} = .482$$

$$r_{Y3|12} = -\sqrt{.105} = -.324$$

$$r_{Y1|2} = \sqrt{.031} = .176$$

Note that the coefficients $r_{Y2|1}$ and $r_{Y1|2}$ are positive because we see from $b_2 = .6594$ and $b_1 = .2224$ are positive. Similarly, $r_{Y3|12}$ is negative because we see from Table 7.2d that $b_3 = -2.186$ is negative.

Comment

Coefficients of partial determination can be expressed in terms of simple or other partial correlation coefficients. For example:

$$R_{Y2|1}^2 = [r_{Y2|1}]^2 = \frac{(r_{Y2} - r_{12}r_{Y1})^2}{(1 - r_{12}^2)(1 - r_{Y1}^2)}$$

$$R_{Y2|13}^2 = [r_{Y2|13}]^2 = \frac{(r_{Y2|3} - r_{12|3}r_{Y1|3})^2}{(1 - r_{12|3}^2)(1 - r_{Y1|3}^2)}$$

where r_{Y1} denotes the coefficient of simple correlation between Y and X_1 , r_{12} denotes the coefficient of simple correlation between X_1 and X_2 , and so on. Extensions are straightforward.

```
install.packages("asbio")
```

```
library(asbio)
```

```
lm1=lm(y~x1)
```

```
lm12=lm(y~x1+x2)
```

```
partial.R2(lm1, lm12)
```

Standardized Multiple Regression Model

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:

$\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters

$X_{i1}, \dots, X_{i,p-1}$ are known constants

ε_i are independent $N(0, \sigma^2)$

$i = 1, \dots, n$

The standardized regression model is as follows:

$$Y_i^* = \beta_1^* X_{i1}^* + \cdots + \beta_{p-1}^* X_{i,p-1}^* + \varepsilon_i^*$$

where the response variable Y^* and the independent

variables X_i^* are given by

$$Y^* = \frac{Y - \bar{Y}}{S_Y} \quad , \quad X_i^* = \frac{X_i - \bar{X}_i}{S_{X_i}} \quad , \quad i = 1, 2, \dots, p-1,$$

where \bar{Y} and \bar{X}_k are the respective means of the Y and the X_k observations, and s_Y and s_k are the respective standard deviations defined as follows:

$$s_Y = \sqrt{\frac{\sum_i (Y_i - \bar{Y})^2}{n-1}}$$

$$s_k = \sqrt{\frac{\sum_i (X_{ik} - \bar{X}_k)^2}{n-1}} \quad (k = 1, \dots, p-1)$$

The correlation transformation is a simple function of the standardized variables

$$Y_i^* = \frac{1}{\sqrt{n-1}} \left(\frac{Y_i - \bar{Y}}{s_Y} \right)$$

$$X_{ik}^* = \frac{1}{\sqrt{n-1}} \left(\frac{X_{ik} - \bar{X}_k}{s_k} \right) \quad (k = 1, \dots, p-1)$$

The relation between the coefficients of the original model and standardized model are

$$\beta_k = \left(\frac{s_Y}{s_k} \right) \beta_k^* \quad (k = 1, \dots, p-1)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \dots - \beta_{p-1} \bar{X}_{p-1}$$

We see that the standardized regression coefficients β_k^* and the original regression coefficients β_k ($k = 1, \dots, p-1$) are related by simple scaling factors involving ratios of standard deviations.

Estimated Standardized Regression Coefficients

Let

$$\mathbf{X}_{n \times (p-1)} = \begin{bmatrix} X_{11}^* & \cdots & X_{1,p-1}^* \\ X_{21}^* & \cdots & X_{2,p-1}^* \\ \vdots & & \vdots \\ X_{n1}^* & \cdots & X_{n,p-1}^* \end{bmatrix}$$

and

$$\mathbf{Y} = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{bmatrix}$$

Then

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

It can be shown that for the transformed variables, $X'Y$ and $X'X$ become

$$X'X = r_{XX} \quad \text{and} \quad X'Y = r_{yX}$$

$$r_{XX} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1,p-1} \\ r_{12} & 1 & r_{23} & r_{2,p-1} \\ & & \ddots & \\ r_{1,p-1} & r_{2,p-1} & & r_{(1-p)(1-p)} \end{bmatrix}, \quad r_{yX} = \begin{bmatrix} r_{y1} \\ r_{y2} \\ \vdots \\ r_{y,p-1} \end{bmatrix},$$

$$r_{ij} = \text{corr}(X_i, X_j), \quad r_{yi} = \text{corr}(Y, X_i)$$

and hence

$$\mathbf{b} = \mathbf{r}_{XX}^{-1} \mathbf{r}_{yX}$$

$$\mathbf{b}_{(p-1) \times 1} = \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_{p-1}^* \end{bmatrix}$$

The regression coefficients b_1^*, \dots, b_{p-1}^* are often called *standardized regression coefficients*.

The return to the estimated regression coefficients for regression model original variables is accomplished by employing the relations:

$$b_k = \left(\frac{s_Y}{s_k} \right) b_k^* \quad (k = 1, \dots, p-1)$$

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - \dots - b_{p-1} \bar{X}_{p-1}$$

Example: In Dwaine Studios example data

$$\begin{aligned} Y_1^* &= \frac{1}{\sqrt{n-1}} \left(\frac{Y_1 - \bar{Y}}{s_Y} \right) & X_{11}^* &= \frac{1}{\sqrt{n-1}} \left(\frac{X_{11} - \bar{X}_1}{s_1} \right) \\ &= \frac{1}{\sqrt{21-1}} \left(\frac{174.4 - 181.90}{36.191} \right) & &= \frac{1}{\sqrt{21-1}} \left(\frac{68.5 - 62.019}{18.620} \right) \\ &= -.04634 & &= .07783 \end{aligned}$$

$$X_{12}^* = \frac{1}{\sqrt{n-1}} \left(\frac{X_{12} - \bar{X}_2}{s_2} \right) = \frac{1}{\sqrt{21-1}} \left(\frac{16.7 - 17.143}{.97035} \right) = -.10208$$

$$\hat{Y}^* = .7484X_1^* + .2511X_2^*$$

and

$$b_1 = \left(\frac{s_Y}{s_1} \right) b_1^* = \frac{36.191}{18.620} (.7484) = 1.4546$$

$$b_2 = \left(\frac{s_Y}{s_2} \right) b_2^* = \frac{36.191}{.97035} (.2511) = 9.3652$$

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 = 181.90 - 1.4546(62.019) - 9.3652(17.143) = -68.860$$

R:Code

```
mydata=read.table("Dwayne Studios.txt",header=TRUE)
Y=mydata$Y
X1=mydata$X1
X2=mydata$X2
n=length(X1)
one=as.vector(rep(1, n))
X=cbind(one,X1,X2)
Model=lm(Y~X1+X2)
library(QuantPsyc)
lm.beta(Model)
```

Regression Model Selection

Likelihood function

In the multiple linear regression model, we have

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

where:

$\beta_0, \beta_1, \dots, \beta_{p-1}$ are parameters

$X_{i1}, \dots, X_{i,p-1}$ are known constants

ε_i are independent $N(0, \sigma^2)$

$i = 1, \dots, n$

As we can see, the error term is follow normal distribution with mean 0 and variance σ^2 , then we can write

$$Likelihood = (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^n \frac{(Y_i - \beta_0 - \beta_1 X_{i1} - \dots - \beta_{p-1} X_{i,p-1})^2}{2\sigma^2}\right)$$

When replace σ^2 by its estimate ($s_n^2 = SSE/n$) and the model coefficients by their estimates, we get

$$Likelihood = L = (2\pi SSE/n)^{-n/2} \exp\left(-\sum_{i=1}^n \frac{(Y_i - \hat{Y}_i)^2}{2SSE/n}\right)$$

then

$$-2\text{Log } L = n[\log(2\pi) + \log(SSE/n) + 1]$$

Akaike's An Information Criterion

Generic function calculating Akaike's 'An Information Criterion' for one or several fitted model objects.

Akaike's Information Criterion is usually calculated with software. The basic formula is defined as:

$$\mathbf{AIC} = -2(\mathbf{\log\text{-likelihood}}) + 2\mathbf{K}$$

Where:

- K is the number of model parameters (the number of variables in the model plus the intercept).
- Log-likelihood is a measure of model fit. The higher the number, the better the fit. This is usually obtained from statistical output.

or small sample sizes ($n/K < \approx 40$), use the second-order AIC:

$$\mathbf{AICc} = -2(\mathbf{\log\text{-likelihood}}) + 2\mathbf{K} + (2\mathbf{K(K+1)/(n-K-1)})$$

Where:

n = sample size,

K= number of model parameters,

Log-likelihood is a measure of model fit.

formula $-2 \cdot \log\text{-likelihood} + k \cdot n \cdot p$,

where n represents the number of parameters in the fitted model,

$k = 2$ for the usual AIC,

or $k = \log(n)$ (n being the number of observations). This can be used when $n/p < 40$.

i.e

$$\text{AIC} = -2 \log L + 2(p+1),$$

Where

L is the likelihood function

p is the number of parameters in linear model and we add one because we have σ^2 to be estimated.

How to know if the model is best fit for your data?

The most common metrics to look at while selecting the model are:

STATISTIC	CRITERION
R-Squared	Higher the better (> 0.70)
Adj R-Squared	Higher the better
F-Statistic	Higher the better
Std. Error	Closer to zero the better
t-statistic	Should be greater 1.96 for p-value to be less than 0.05
AIC	Lower the better
BIC	Lower the better
Mallows cp	Should be close to the number of predictors in model
MAPE (Mean absolute percentage error)	Lower the better
MSE (Mean squared error)	Lower the better
Min_Max Accuracy => mean(min(actual,	Higher the better

In this part we use AIC as a criterion to the model selection.

Example: Use the mtcars data to select the best model

<https://gist.github.com/seankross/a412dfbd88b3db70b74b>

Using

Stepwise: Backward selection

Stepwise: Forward selection

Stepwise: Combination of Forward and Backward selection

R code:

```
data(mtcars)
```

```
d=mtcars
```

```
head(d)
```

```
FitAll=lm(mpg~., data=d)
```

```
summary(FitAll)
```

```
p=length(FitAll$coef)
```

```
n=length(d$mpg)
```

```
SSE=sum((FitAll$res)^2)
```

```
AIC=n*(log(2*pi)+1+log(SSE/n))+2*(p+1)
```

```
logL=as.numeric(logLik(FitAll))
```

```
AIC=-2*logL+2*(p+1)
```

```
AIC(FitAll)
```

```
#===== Backward=====
step(FitAll, direction="backward")
```

```
#===== Forward =====
Fitstart=lm(mpg~1,data=d)
summary(Fitstart)
step(Fitstart, direction="forward", scope=formula(FitAll))
#===== Both =====
```

```
Fitstart=lm(mpg~1,data=d)
summary(Fitstart)
step(Fitstart, direction="both", scope=formula(FitAll))
#=====
```

R:Code

```
mydata=read.table("Dwaine Studios.txt",header=TRUE)
Y=mydata$Y
X1=mydata$X1
X2=mydata$X2
n=length(X1)
one=as.vector(rep(1, n))
X=cbind(one,X1,X2)
Model=lm(Y~X1+X2)
library(QuantPsyc)
lm.beta(Model)
```