Chapter 1
Introduction

1.1 Regression and Model Building

Regression analysis (RA) is a statistical technique for investigating and modeling the relationship between variables. RA has a lot of applications in almost every filed, including engineering, economics, physical and chemical sciences, management, biological and social sciences.

As an example, suppose that an industrial engineer employed by a soft drink beverage bottler is analyzing the product delivery and service operations for vending machines. He suspects that the time required by a route delivery man to load and service a machine is related to the number of cases of product delivered. The engineer visits 25 randomly selected retail outlets having vending machines and the in-outlet delivery time (in minutes) and the volume of product delivered (in cases) are observed for each. The 25 data points are plotted in Figure 1.1 a. This graph is called scatter plot.

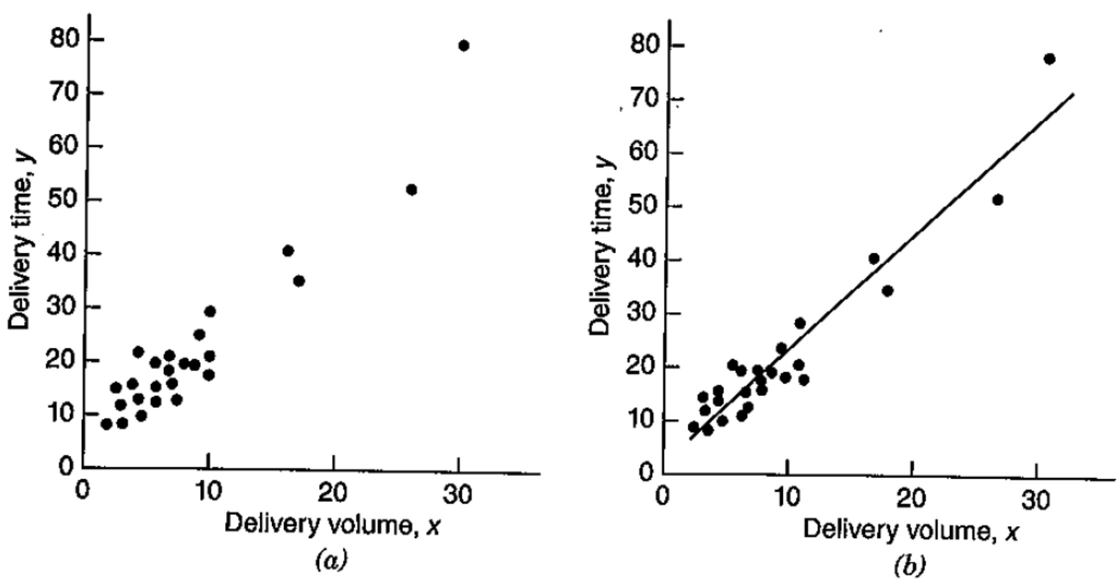

Figure 1.1 (*a*) Scatter diagram for delivery volume. (*b*) Straight-line relationship between delivery time and delivery volume.

Let y represent delivery time and x represent delivery volume, then the relationship between y and x can be expressed as

$$y = \beta_0 + \beta_1 x \qquad (1.1)$$

where β_0 is the intercept and β_1 is the slope. Let the difference between the observed value of y and the straight line ($\beta_0 + \beta_1 x$) be an error, ε. It is a random variable that accounts for the failure of the model to fit the data exactly. Then a plausible model for the delivery time would be

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (1.2)$$

Equation (1.2) is called a linear regression model. Here y is called dependent or response variable and x is called independent or predator or regressor variable. Suppose the mean and variance of ε are 0 and σ^2 respectively. Then the mean and variance of y would be respectively

$$E(y \mid x) = \beta_0 + \beta_1 x \quad \text{and} \quad \mathrm{V}(y \mid x) = \sigma^2$$

The slope can be interpreted as the change in the mean of y for a unit change in x.

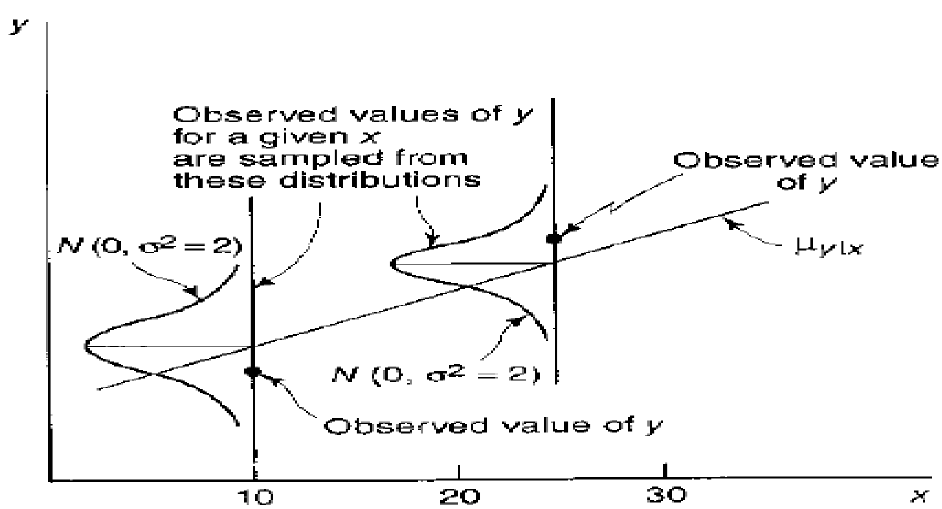

Figure 1.2 How observations are generated in linear regression.

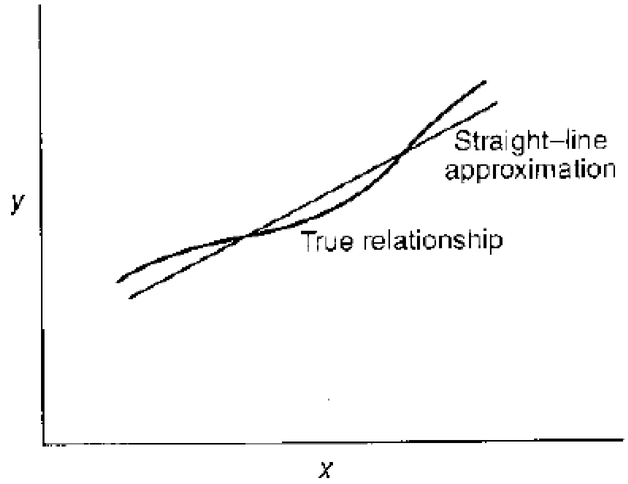

Figure 1.3 Linear regression approximation of a complex relationship.

In general, the response variable y may be related to k regressors, x_1, x_2, ..,x_k, so that

$$y = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + ... + \beta_1 x_k + \varepsilon$$

This is called multiple linear regression model, as it has more than one regressors or independent variables.

1.2 Data Collection

An essential aspect of regression analysis is data collection. Any regression analysis is only as good as the data on which it is based. Three basic methods (among others) for collecting data are as follows:

- A retrospective study based on historical data
- An observational study
- A designed experiment

1.3 Uses of Regression

Regression models are used for several purposes, including the followings

1. Data description
2. Parameter estimation
3. Predation and estimation
4. Control

Engineers and scientists frequently use equations to summarize or describe a set of data. Regression analysis is helpful in developing such equations. For example, we may collect a considerable amount of delivery time and delivery volume data and a regression model probably be a much more convenient and useful summary of these data than a table or even a graph.

1.4 Role of the Computer

Fitting or building a regression model is an iterative process. The model-building process is illustrated in Figure 1.8.

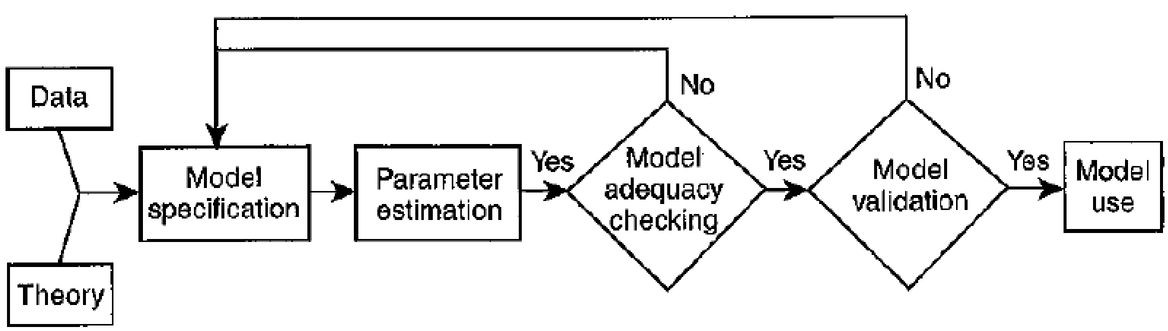

Figure 1.8 Regression model-building process.

We will use mostly R (or Splus, SPSS, Minitab) software to analyze the data of this subject.