PROBABILITY & STATISTICS

BS 1402

Contents

- Simple Linear Regression
- Least Squares Method for Fitted Linear Regression Model

Linear Relationship

Linear Equation:

$$Y = \beta_0 + \beta_1 x,$$

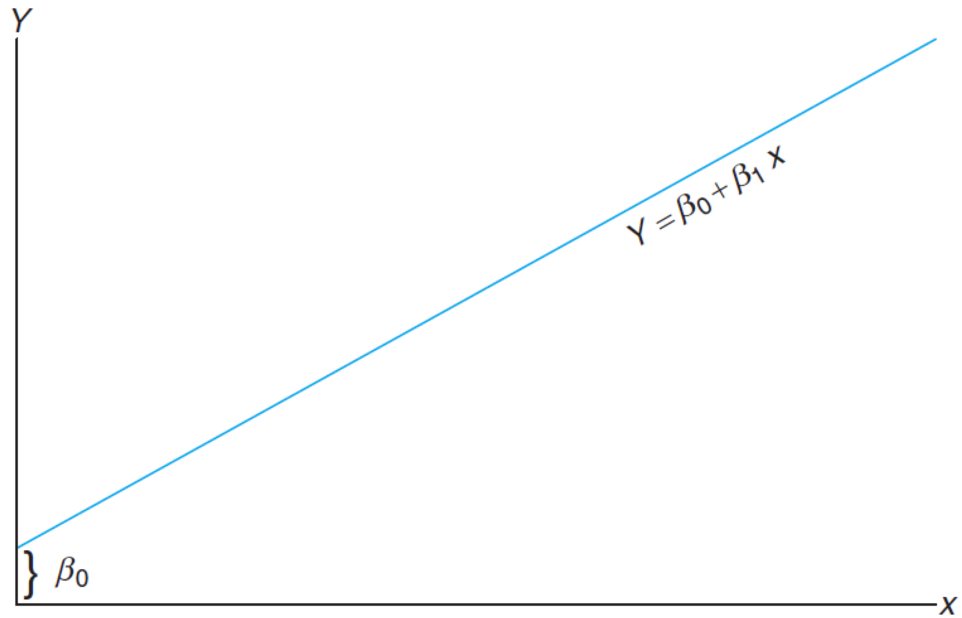


Figure 11.1: A linear relationship; β_0: intercept; β_1: slope.

The Simple Linear Regression (SLR) Model

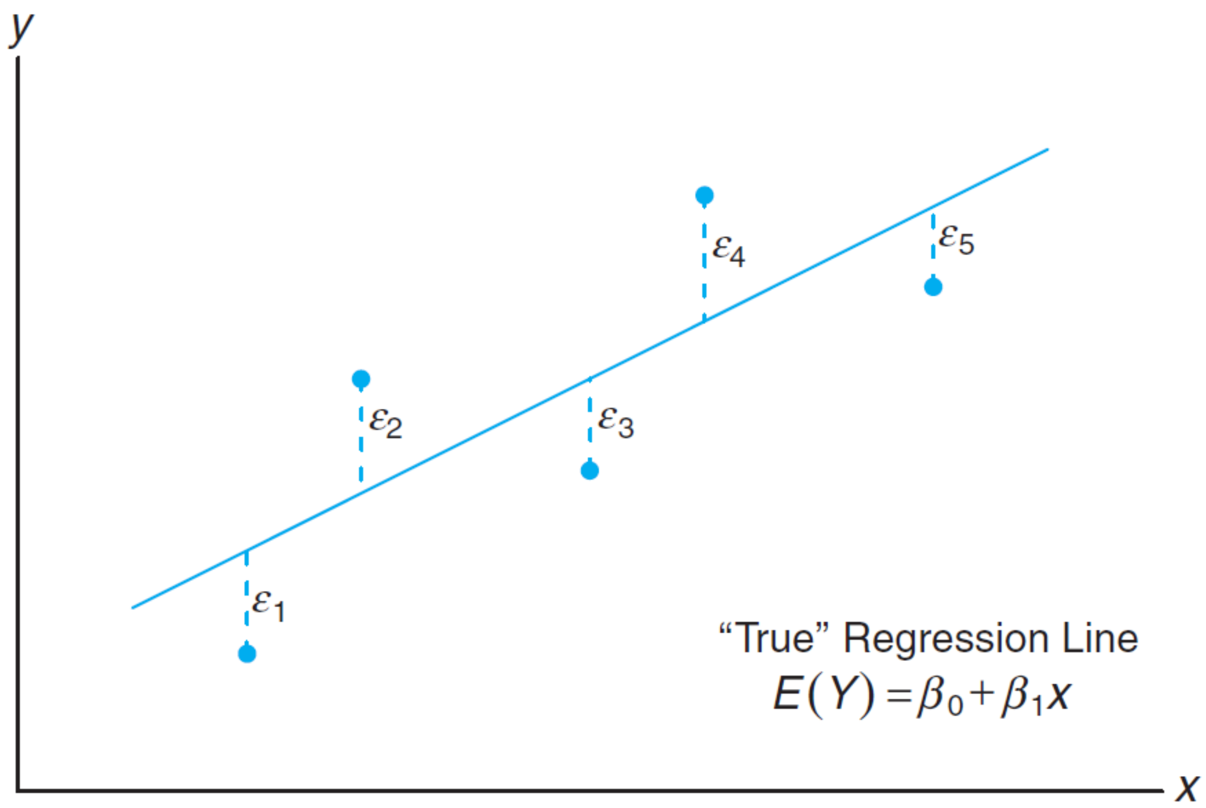
The response Y is related to the independent variable x through the equation

$$Y = \beta_0 + \beta_1 x + \epsilon.$$

β_0 and β_1 are unknown intercept and slope parameters, respectively,

and ϵ is a random variable that is assumed to be distributed with $E(\epsilon) = 0$

The Simple Linear Regression (SLR) Model



The Fitted Regression Model

An important aspect of regression analysis is, very simply, to estimate the parameters β_0 and β_1 (i.e., estimate the so-called regression coefficients).

Suppose we denote the estimates b_0 for β_0 and b_1 for β_1. Then the estimated or fitted regression line is given by

$$\hat{y} = b_0 + b_1 x,$$

Least Squares and the Fitted Model

A residual is an error in the fit of the model:

$$\hat{y} = b_0 + b_1 x,$$

Given a set of regression data $\{(x_i, y_i); i = 1, 2, \ldots, n\}$ and a fitted model, $\hat{y}_i = b_0 + b_1 x_i$, the ith residual e_i is given by

$$e_i = y_i - \hat{y}_i, \quad i = 1, 2, \ldots, n.$$

If a set of n residuals is large, then the fit of the model is not good.

Small residuals are a sign of a good fit.

The Method of Least Squares

$$b_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right) \left(\sum_{i=1}^{n} y_i \right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2}$$

$$b_0 = \frac{\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} x_i}{n}$$

Example 1:

Estimate the regression line for the pollution data of Table 11.1.

$$\sum_{i=1}^{33} x_i = 1104, \ \sum_{i=1}^{33} y_i = 1124, \ \sum_{i=1}^{33} x_i y_i = 41{,}355, \ \sum_{i=1}^{33} x_i^2 = 41{,}086$$

Therefore,

$$b_1 = \frac{(33)(41{,}355) - (1104)(1124)}{(33)(41{,}086) - (1104)^2} = 0.903643 \text{ and}$$

$$b_0 = \frac{1124 - (0.903643)(1104)}{33} = 3.829633.$$

Thus, the estimated regression line is given by

$$\hat{y} = 3.8296 + 0.9036x.$$

Example 1 Data:

Table 11.1: Measures of Reduction in Solids and Oxygen Demand

Solids Reduction, x (%)	**Oxygen Demand Reduction, y (%)**	**Solids Reduction, x (%)**	**Oxygen Demand Reduction, y (%)**
3	5	36	34
7	11	37	36
11	21	38	38
15	16	39	37
18	16	39	36
27	28	39	45
29	27	40	39
30	25	41	41
30	35	42	40
31	30	42	44
31	40	43	37
32	32	44	44
33	34	45	46
33	32	46	46
34	34	47	49
36	37	50	51
36	38		

Thank You.