Simple Linear Regression

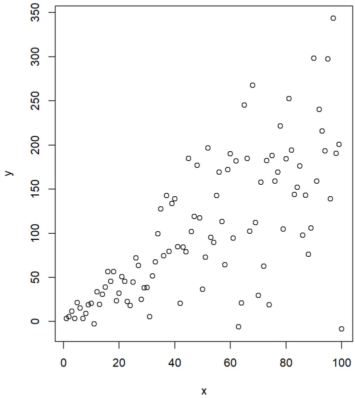
Dr. Kiah Wah Ong

Introduction

Given 100 data points

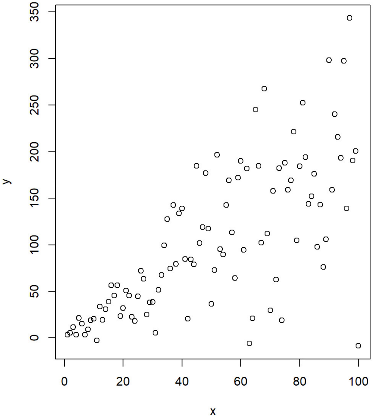
$$(x_1, y_1), (x_2, y_2), \cdots, (x_{100}, y_{100})$$

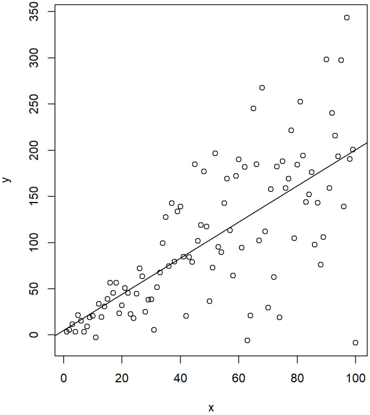
we want to fit a line (or more generally an equation) to describe the relationship between x and y.



Introduction

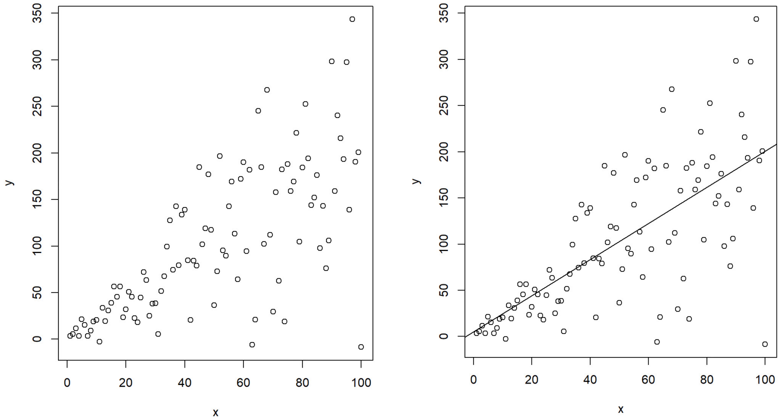
A **regression model** is a function that describes the relationship between the outcome (response) y, and the predictor (regressor) x.





Simple Linear Regression (SLR)

The **simple linear regression model** is a model with a single predictor x that has a relationship with a outcome y that is a straight line.

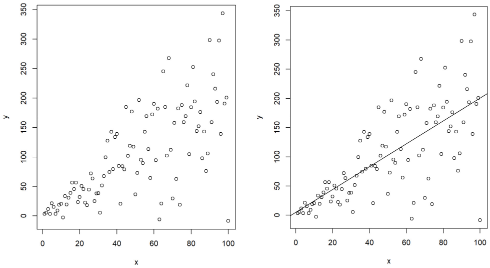


Simple Linear Regression (SLR)

In SLR, the relationship between a response variable y and a predictor variable x is postulated as

$$y = \beta_0 + \beta_1 x + \epsilon$$

where β_0 and β_1 are constants called the *model regression coefficients*, and ϵ is a random disturbance or error.



Simple Linear Regression (SLR)

For data points (x_i, y_i), $i = 1, 2, \cdots, n$, we write

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

(i) y_i is the value of the response variable in the i-th trial

(ii) β_0 and β_1 are parameters

(iii) x_i is a **known constant**, namely, the value of the predictor variable in the i-th trial

(iv) ϵ_i is a random error with

$$E(\epsilon_i) = 0 \quad \text{and} \quad \sigma^2(\epsilon_i) = \sigma^2$$

and ϵ_i, ϵ_j are uncorrelated so that their covariance is zero, i.e. $\sigma(\epsilon_i, \epsilon_j) = 0$ for all i, j, $i \neq j$, $i = 1, 2, \cdots, n$.

Estimation of β_0 and β_1

The parameters β_0 and β_1 are unknown and must be estimated using sample data

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

The **least squares method** is used to estimate β_0 and β_1. To do this, from

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

we write

$$\epsilon_i = y_i - (\beta_0 + \beta_1 x_i)$$

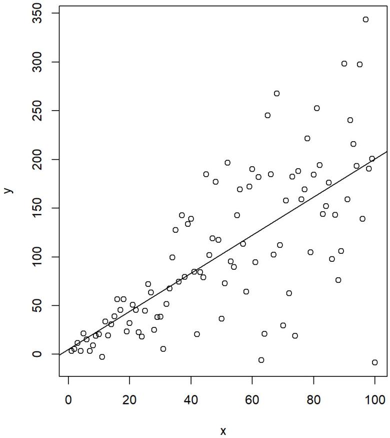
and call the following as

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

the residual sum of squares.

Estimation of β_0 and β_1

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$



Estimation of β_0 and β_1

In the next video we will show how to estimate β_0 and β_1 by minimizing the residual sum of squares.