MA425/652: Applied Regression Analysis - Fall 2021

Linear Regression with One Predictor Variable

Emmanuel Thompson

7-17-2021

Contents

1.1 Relations Between Variables

1.2 Regression Models & Their Uses

1.3 Simple Linear Regression Model with Distribution of Error Terms Unspecified

1.3.1 Formal Statement of Model

- The Basic Model (one predictor with linear regression function) is written as:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \text{ for } i = 1, 2, \cdots, n. \tag{1}$$

- where:
  - Y_i = value of the response variable in the ith trial (Unknown).
  - X_i = value of the predictor variable in the ith trial (Constant).
  - β_0 & β_1 are unknown parameters.
  - ϵ_i = random error term.
    - **Mean:** $\mathrm{E}\{\epsilon_i\} = 0$.
    - **Variance:** $\sigma^2\{\epsilon_i\} = \sigma^2$.
    - **Covariance:** $\sigma\{\epsilon_i, \epsilon_j\} = 0$ for all $i, j : i \neq j$.

1.3.2 Important Features of Model (1)

1. Y_i comprises of two components:

- Constant Term $\implies \beta_0 + \beta_1 X_i$.
- Error Term: $\implies \epsilon_i$.
- Therefore Y_i is a random variable.

2. Since $\mathrm{E}\{\epsilon_i\} = 0$, it follows that $\mathrm{E}\{Y_i\} = \beta_0 + \beta_1 X_i$.
3.

1.4 Data for Regression Analysis

- Observational Data
- Experimental Data
- Completely Randomized Design

1.4.1 Observational Data

- Data obtained not through experimental studies.
- Observational studies do not control the explanatory or predictor variable(s) of interest.

Example:

- *Company officials wished to study the relation between age of employee (X) and number of days of illness last year (Y).*
- **The needed data for use in the regression analysis were obtained from personnel records. Such data are observational data since the explanatory variable, age, is not controlled.**

1.4.2 Experimental Data

1.4.3 Completely Randomized Design