



CHAPTER 1

Simple Linear Regression

Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:

- One variable, denoted x , is regarded as the predictor, explanatory, or independent variable.
- The other variable, denoted y , is regarded as the response, outcome, or dependent variable.

Because the other terms are used less frequently today, we'll use the "predictor" and "response" terms to refer to the variables encountered in this course. The other terms are mentioned only to make you aware of them should you encounter them in other contexts.

Simple linear regression gets its adjective "simple," because it concerns the study of only one predictor variable. In contrast, multiple linear regression, a topic that will be covered later, gets its adjective "multiple," because it concerns the study of two or more predictor variables.

1.0.1 The model behind simple linear regression

Given a scatterplot of the response variable y versus the predictor variable x , we fit the line

$$y = \beta_0 + \beta_1 x + \epsilon \quad (1.1)$$

that minimizes the distances from the observed points to the line!

- y = dependent variable (output/outcome/prediction/estimation)
- β_0 = y-intercept (constant term)
- β_1 = slope of the regression line (the effect that X has on Y)
- x = independent variable (input variable used in the prediction of Y)
- ϵ = error (the difference between the actual and predicted/estimated value)

This line can be used to predict future values of y given new data values of x .



APPENDIX A

Answers to Exercises

