Contents

Last update: 2025 January 23

Chapter 1

Statistics for the Working Economist

§ 1.1. Measurements

Definition 1.1.1.

(1) The large body of data that is the target of our interest is called the *population*.

(2) A subset selected from a given population is called a *sample*.

Definition 1.1.2. The *mean* of a sample of n measured responses $y_1, ..., y_n$ is given by:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

The corresponding *population mean* is denoted μ.

Remark. We usually cannot measure the value of the population mean, μ; rather, μ is an unknown constant that we may want to estimate using sample information.

Definition 1.1.3. The *variance* of a sample of measurements $y_1, ..., y_n$ is given by:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

The corresponding *population variance* is denoted by σ^2. The larger the variance of a set of measurements, the greater will be the amount of variation within the set

Definition 1.1.4. The *standard deviation* of a sample of measurements is given by:

$$s = \sqrt{s^2}.$$

The corresponding *population standard deviation* is denoted by $\sigma = \sqrt{\sigma^2}$

Definition 1.1.5. An *estimator* is a formula that tells how to calculate the value of an estimate based on the measurements contained in a sample.

Example 1.1.1. The sample mean $\overline{y}$ is one possible point estimator of a population mean μ.

§ 1.2. Linear Models

In this chapter, we undertake a study of inferential procedures that can be used when a random variable Y, called the *dependent variable*, has a mean that is a function of one or more non-random variables $x_1, ..., x_k$ called *independent variables*.

Definition 1.2.1. A *linear statistical model* relating a random response Y to a set of independent variables $x_1, ..., x_k$ is of the form:

$$Y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + \epsilon,$$

where $\beta_0, ..., \beta_k$ are unknown parameters, $epsilon$ is a random variable (typically an error of some sort), and the variables $x_1, ..., x_k$ are known values.

Remark. Although unreasonable, we will assume that $E(\epsilon) = 0$, hence that:

$$E(Y) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k.$$