Contents

Last update: 2025 February 18

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.*

It is important to note that we cannot make any measurements based off of a given population —our only resource is making inferences based off of data gathered from a sample. For example, suppose we make N observations $Y_1, ..., Y_N$ from a given population and compute its mean:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{N} Y_i.$$

The value $\overline{Y}$ is merely an *approximation* or *estimation* of what the true value of the population mean is. The population mean, typically denoted μ, is an unknown constant which we can only estimate using information from a given sample. This leaves us with the following definition:

Definition 1.1.2. An *estimator* is a formula that tells how to calculate the value of an estimate based on the measurements contained in a sample.

Typically, if θ is a fixed parameter from a population, we denote its estimator as $\widehat{\theta}$.

§ 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, ϵ is a random variable, and the variables $X_1, ..., X_k$ assume known values. We assume that $E[\epsilon] = 0$ and hence that:

$$E[Y] = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k.$$

The notation of our linear statistical model needs to be extended to include reference to the number of observations. Suppose that from our random response Y we make n independent observations $Y_1, ..., Y_n$. We can write the observation Y_i as:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + ... + \beta_N X_{i,n} + \epsilon_i,$$

where $X_{i,j}$ is the i^{th} observation of the j^{th} independent variable and ϵ_i is the i^{th} observation of the random variable. This is essentially a system of n linear equations. Let:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}.$$

We can now express our linear statistical model as:

$$\mathbf{Y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\epsilon}.$$

Note that if $\widehat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$, then $\widehat{\mathbf{Y}} = \widehat{\boldsymbol{\beta}}\mathbf{X} + \boldsymbol{\epsilon}$ is an estimator for $E[\mathbf{Y}]$.

Definition 1.2.2. The *sum of squares for errors* is:

$$\text{SSE} = \mathbf{Y}^t\mathbf{Y} - \boldsymbol{\beta}^t\mathbf{X}^t\mathbf{Y},$$

Exercise 1.2.1. Show that $\mathbf{Y}^t\mathbf{Y} - \boldsymbol{\beta}^t\mathbf{X}^t\mathbf{Y} = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2$.

§ 1.3. Method of Least Squares

The least-squares procedure for fitting a line through a set of n data points is similar to the method that we might use if we fit a line by eye; that is, we want the differences between the observed values and corresponding points on the fitted line to be "small" in some overall sense. A convenient way to accomplish this, and one that yields estimators with good properties, is to minimize the sum of squares of the vertical deviations from the fitted line:

$$\frac{\partial \text{SSE}}{\partial \widehat{\boldsymbol{\beta}}} = \begin{pmatrix} \frac{\partial \text{SSE}}{\partial \widehat{\beta_1}} \\ \vdots \\ \frac{\partial \text{SSE}}{\partial \widehat{\beta_n}} \end{pmatrix} = \mathbf{0}.$$

We will not show the full derivation of finding $\widehat{\beta}$, as this is not used in practice often. Typically, one will use computational software such as Stata to compute linear regressions using the method of least-squares. Regardless, consider the following proposition for *simple linear regression models*, which has merely one independent variable X.

Proposition 1.3.1. *The least-squares estimators for simple linear regression models is given by:*

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$
$$\widehat{\beta_0} = \overline{Y} - \widehat{\beta_1}\,\overline{X}.$$

Exercise 1.3.1. For a simple linear regression model, show that:

$$\widehat{\beta} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^{-1}\mathbf{Y} = \begin{pmatrix} \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2} \\ \\ \overline{Y} - \widehat{\beta_1}\,\overline{X} \end{pmatrix}.$$