6-6 Assessing Normality

Key Concept The following chapters include several important statistical methods requiring that sample data are a simple random sample from a population having a *normal* distribution. In this section we present criteria for determining whether the requirement of a normal distribution is satisfied. The criteria involve (1) visual inspection of a histogram to see if it is roughly bell-shaped; (2) identifying any outliers; and (3) constructing a graph called a *normal quantile plot*.

Part 1: Basic Concepts of Assessing Normality

When trying to determine whether a collection of data has a distribution that is approximately normal, we can visually inspect a histogram to see if it is approximately bell-shaped (as discussed in Section 2-3), we can identify outliers, and we can also use a *normal quantile plot* (discussed briefly in Section 2-3).

DEFINITION A **normal quantile plot** (or **normal probability plot**) is a graph of points (x, y) where each x value is from the original set of sample data, and each y value is the corresponding z score that is a quantile value expected from the standard normal distribution.

Procedure for Determining Whether It Is Reasonable to Assume That Sample Data Are from a Population Having a Normal Distribution

1. *Histogram:* Construct a histogram. If the histogram departs dramatically from a bell shape, conclude that the data do not have a normal distribution.
2. *Outliers:* Identify outliers. If there is more than one outlier present, conclude that the data do not have a normal distribution. (Just one outlier could be an error or the result of chance variation, but be careful, because even a single outlier can have a dramatic effect on results.)
3. *Normal quantile plot:* If the histogram is basically symmetric and the number of outliers is 0 or 1, use technology to generate a *normal quantile plot*. Apply the following criteria to determine whether or not the distribution is normal. (These criteria can be used loosely for small samples, but they should be used more strictly for large samples.)

   Normal Distribution: The population distribution is normal if the pattern of the points is reasonably close to a straight line and the points do not show some systematic pattern that is not a straight-line pattern.

   Not a Normal Distribution: The population distribution is *not* normal if either or both of these two conditions applies:

   - The points do not lie reasonably close to a straight line.
   - The points show some *systematic pattern* that is not a straight-line pattern.

Later in this section we will describe the actual process of constructing a normal quantile plot, but for now we focus on interpreting such a plot.

Example 1 Determining Normality

The accompanying displays show histograms of data along with the corresponding normal quantile plots.

continued