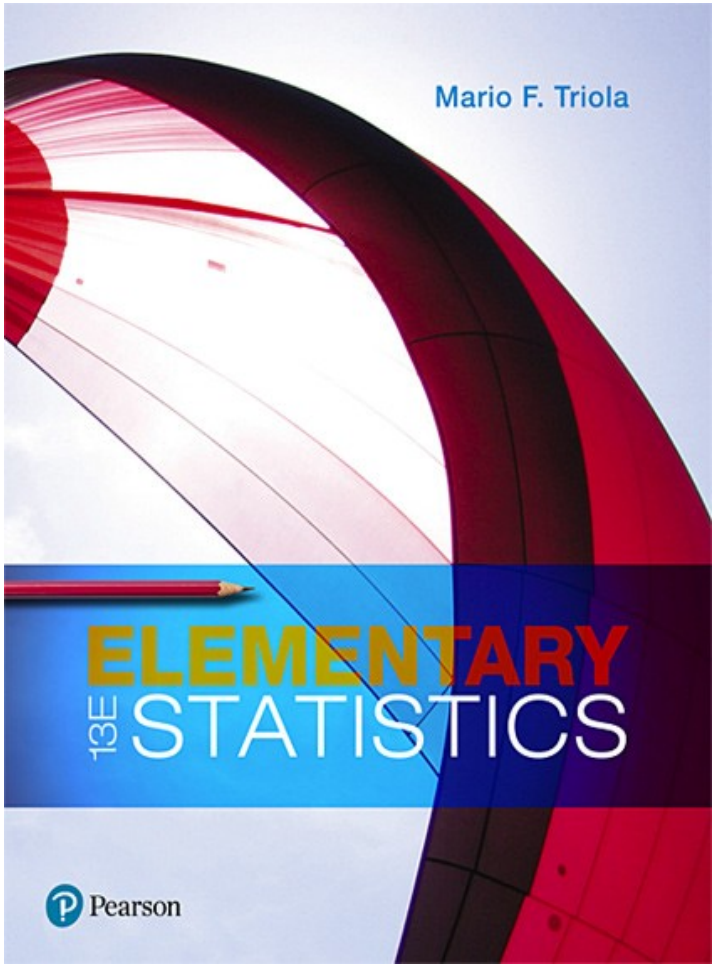
Elementary Statistics

Thirteenth Edition



Chapter 6

Normal Probability Distributions

Normal Probability Distributions

6-1 The Standard Normal Distribution

6-2 Real Applications of Normal Distributions

6-3 Sampling Distributions and Estimators

6-4 The Central Limit Theorem

6-5 Assessing Normality

6-6 Normal as Approximation to Binomial

Key Concept

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 **normal quantile plot.**

Normal Quantile Plot

- Normal Quantile Plot
  - 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 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 of 2)

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 might not have a normal distribution.
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.**

Procedure for Determining Whether It Is Reasonable to Assume That Sample Data Are from a Population Having a Normal Distribution (2 of 2)

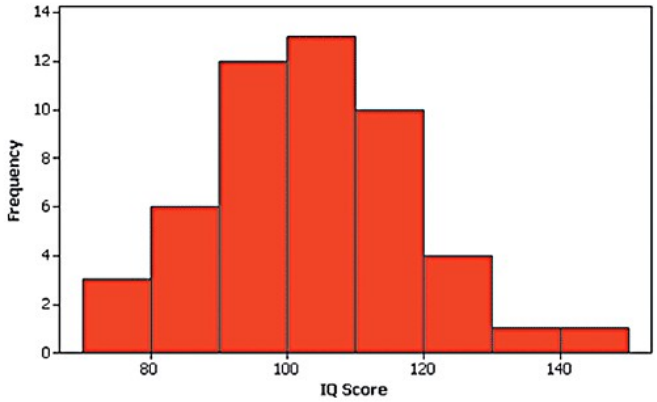
Apply the following criteria to determine whether the distribution is normal. (These criteria can be used loosely for small samples, but they should be used more strictly for large samples.)

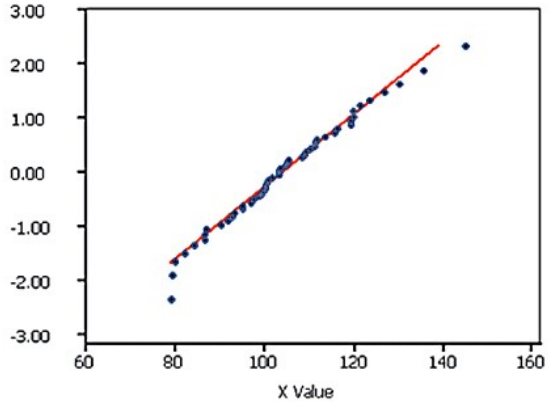
Normal Distribution: The population distribution is normal if the pattern of the points is reasonably close to a straight line and the points do not show some systematic pattern that is not a straight-line pattern.

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.

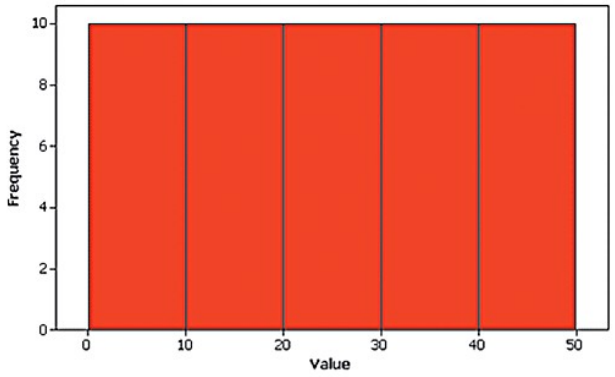
Normal Example

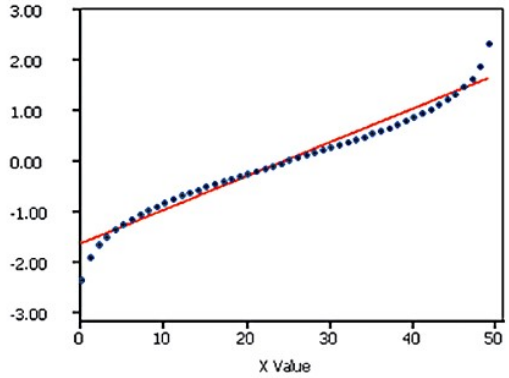




Normal: The first case shows a histogram of IQ scores that is close to being bell-shaped suggesting a normal distribution. The corresponding normal quantile plot shows points that are reasonably close to a straight-line pattern, and the points do not show any other systematic pattern that is not a straight line. It is safe to assume that these IQ scores are from a population that has a normal distribution.

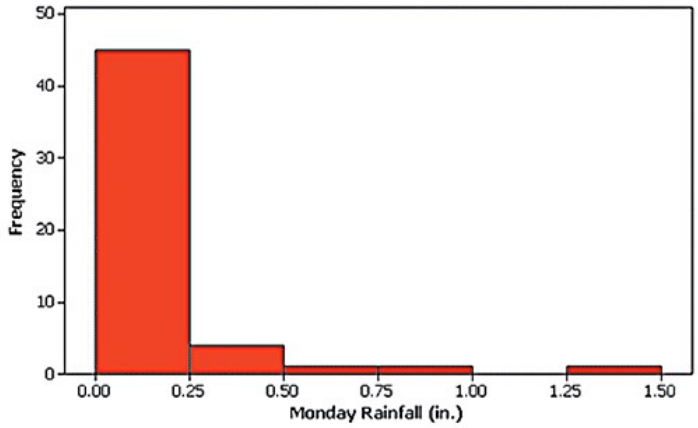
Uniform Example

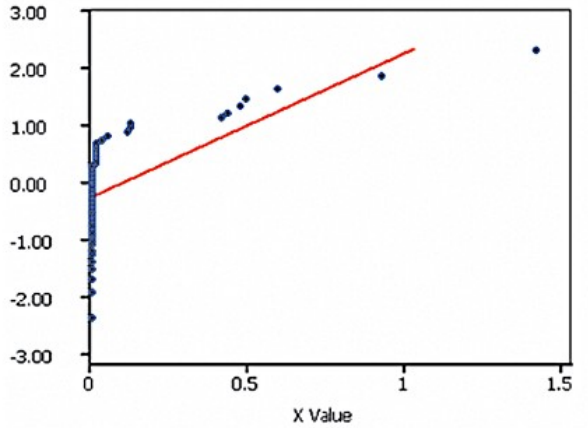




Uniform: The second case shows a histogram of data having a uniform (flat) distribution. The corresponding normal quantile plot suggests that the points are not normally distributed. Although the pattern of points is reasonably close to a straight-line pattern, **there is another systematic pattern that is not a straight-line pattern.** We conclude that these sample values are from a population having a distribution that is not normal.

Skewed Example





Skewed: The third case shows a histogram of the amounts of rainfall (in inches) in Boston for every Monday during one year. The shape of the histogram is skewed to the right, not bell-shaped. The corresponding normal quantile plot shows points that are not at all close to a straight-line pattern. These rainfall amounts are from a population having a distribution that is not normal.

Tools for Determining Normality (1 of 2)

- **Histogram / Outliers:** If the requirement of a normal distribution is not too strict, simply look at a histogram and find the number of outliers. If the histogram is roughly bell-shaped and the number of outliers is 0 or 1, treat the population as if it has a normal distribution.
- Normal Quantile Plot: Normal quantile plots can be difficult to construct on your own, but they can be generated with a TI-83/84 Plus calculator or suitable software, such as Statdisk, Minitab, Excel, or StatCrunch.

Tools for Determining Normality (2 of 2)

- Advanced Methods: In addition to the procedures discussed in this section, there are other more advanced procedures for assessing normality, such as the chi-square goodness-of-fit test, the Kolmogorov-Smirnov test, the Lilliefors test, the Anderson-Darling test, the Jarque-Bera test, and the Ryan-Joiner test (discussed briefly in Part 2).

Manual Construction of Normal Quantile Plots (1 of 3)

Step 1: First sort the data by arranging the values in order from lowest to highest.

Step 2: With a sample of size n, each value represents a proportion of $\frac{1}{n}$ of the sample. Using the known sample size n, find the values of $\frac{1}{2n}, \frac{3}{2n}, \frac{5}{2n}$, and so on, until you get n values.

These values are the cumulative areas to the left of the corresponding sample values.

Manual Construction of Normal Quantile Plots (2 of 3)

Step 3: Use the standard normal distribution (software or a calculator or Table A-2) to find the *z* scores corresponding to the cumulative left areas found in Step 2. (These are the *z* scores that are expected from a normally distributed sample.)

Step 4: Match the original sorted data values with their corresponding *z* scores found in Step 3, then plot the points (*x*, *y*), where each *x* is an original sample value and *y* is the corresponding *z* score.

Manual Construction of Normal Quantile Plots (3 of 3)

Step 5: Examine the normal quantile plot and use the criteria given in Part 1. Conclude that the population has a normal distribution 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.

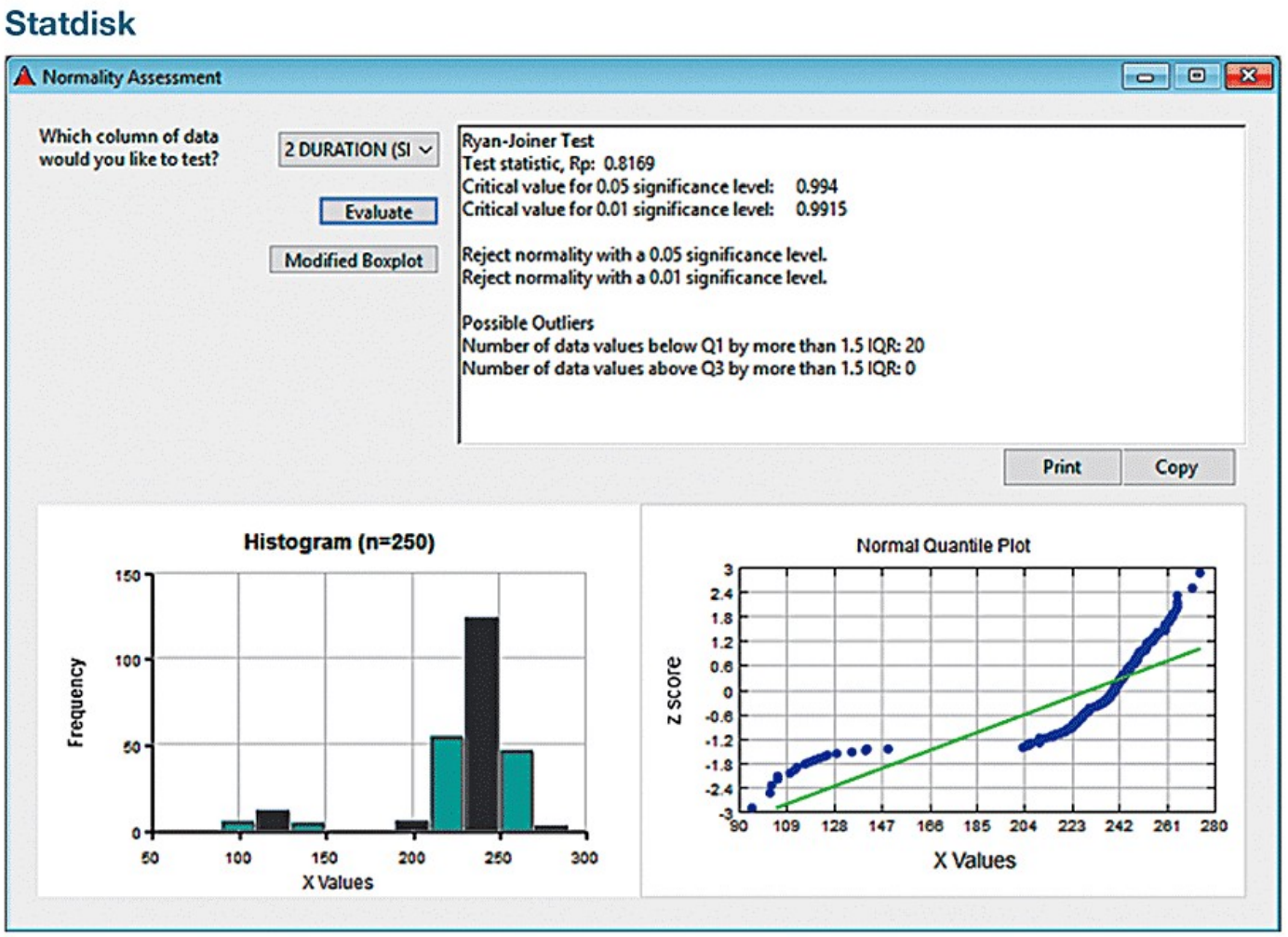
Ryan-Joiner Test

- Ryan-Joiner Test
  - The **Ryan-Joiner test** is one of several formal tests of normality, each having its own advantages and disadvantages.
  - Statdisk has a feature of **Normality Assessment** that displays a histogram, normal quantile plot, the number of potential outliers, and results from the Ryan-Joiner test.

Example: Old Faithful Eruption Times (1 of 3)

We can use the **Normality Assessment** feature of Statdisk with all 250 eruption times listed in the “Old Faithful” data set to get the accompanying display.

Example: Old Faithful Eruption Times (2 of 3)



Example: Old Faithful Eruption Times (3 of 3)

Let's use the display with the three criteria for assessing normality.

1. **Histogram:** We can see that the histogram is **skewed** to the left and is far from being bell-shaped.
2. **Outliers:** The display shows that there are 20 possible outliers. If we examine a sorted list of the 250 eruption times, the 20 lowest times do appear to be outliers.
3. **Normal quantile plot:** Whoa! The points in the normal quantile plot are very far from a straight-line pattern. We conclude that the 250 eruption times do **not** appear to be from a population with a normal distribution.

Data Transformations

Many data sets have a distribution that is not normal, but we can **transform** the data so that the modified values have a normal distribution.

One common transformation is to transform each value of x by taking its logarithm. (You can use natural logarithms or logarithms with base 10. If any original values are 0, take logarithms of values of $x + 1$).

If the distribution of the logarithms of the values is a normal distribution, the distribution of the original values is called a **lognormal distribution**.