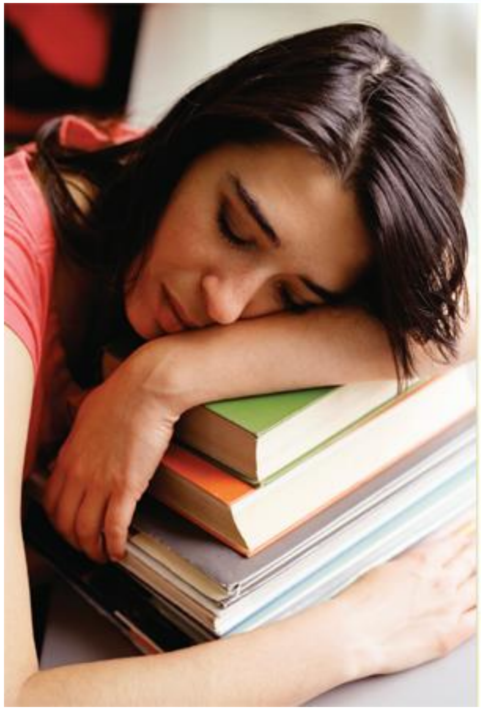
2 Descriptive Analysis and Presentation of Single-Variable Data



2.6 Interpreting and Understanding Standard Deviation

Interpreting and Understanding Standard Deviation

Standard deviation is a measure of variation (dispersion) in the data. It has been defined as a value calculated with the use of formulas.

Even so, you may be wondering what it really is and how it relates to the data. It is a kind of yardstick by which we can compare the variability of one set of data with that of another.

This particular “measure” can be understood further by examining two statements that tell us how the standard deviation relates to the data: the *empirical rule* and *Chebyshev’s theorem*.

The Empirical Rule and Testing for Normality

The Empirical Rule and Testing for Normality

Empirical rule If a variable is normally distributed, then
(1) within 1 standard deviation of the mean, there will be
approximately 68% of the data;

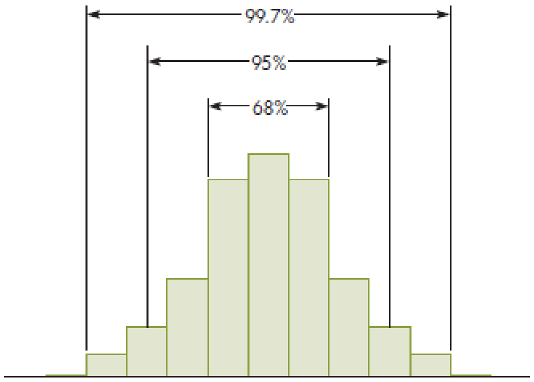
(2) within 2 standard deviations of the mean, there will be
approximately 95% of the data; and

(3) within 3 standard deviations of the mean, there will be
approximately 99.7% of the data. (This rule applies specifically to a normal **[bell-shaped] distribution**, but
it is frequently applied as an interpretive guide to

The Empirical Rule and Testing for Normality

Figure 2.31 shows the intervals of 1, 2, and 3 standard deviations about the mean of an approximately normal distribution.

Usually these proportions do not occur exactly in a sample, but your observed values will be close when a large sample is drawn from a normally distributed population.

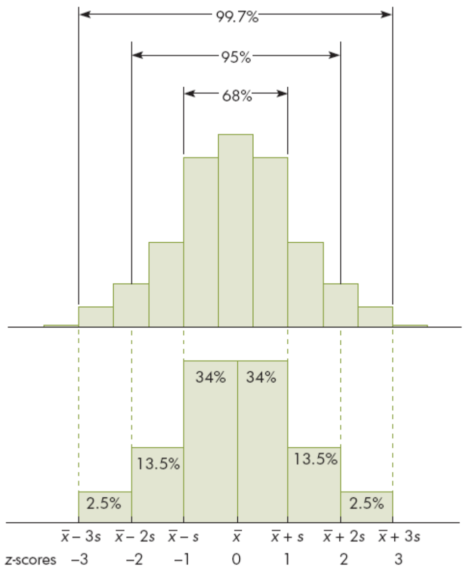


Empirical Rule

Figure 2.31

The Empirical Rule and Testing for Normality

If a distribution is approximately normal, it will be nearly symmetrical and the mean will divide the distribution in half (the mean and the median are the same in a symmetrical distribution). This allows us to refine the empirical rule, as shown in Figure 2.32.



Refinement of Empirical Rule

Figure 2.32

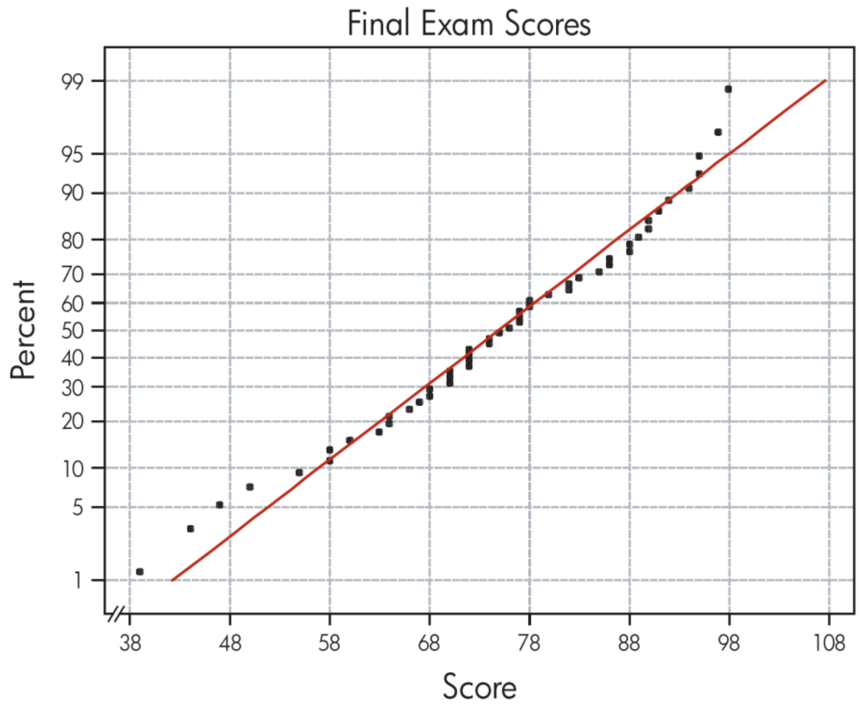
The Empirical Rule and Testing for Normality

The empirical rule can be used to determine whether a set of data is approximately normally distributed.

There is another way to test for normality—by drawing a probability plot (an ogive drawn on probability paper) using a computer or graphing calculator.

The Empirical Rule and Testing for Normality

For our illustration, a probability plot of the statistics final exam scores is shown on Figure 2.33.



Probability Plot of Statistics Exam Scores

Figure 2.33

The Empirical Rule and Testing for Normality

The test for normality, at this point in our study of statistics, is simply to compare the graph of the data (the ogive) with the straight line drawn from the lower left corner to the upper right corner of the graph.

If the ogive lies close to this straight line, the distribution is said to be approximately normal. The vertical scale used to construct the probability plot is adjusted so that the ogive for an exactly normal distribution will trace the straight line.

The ogive of the exam scores follows the straight line quite closely, suggesting that the distribution of exam scores is approximately normal.

Chebyshev’s Theorem

Chebyshev's Theorem

In the event that the data do not display an approximately normal distribution, Chebyshev's theorem gives us
information about how much of the data will fall within intervals centered at the mean for all distributions.

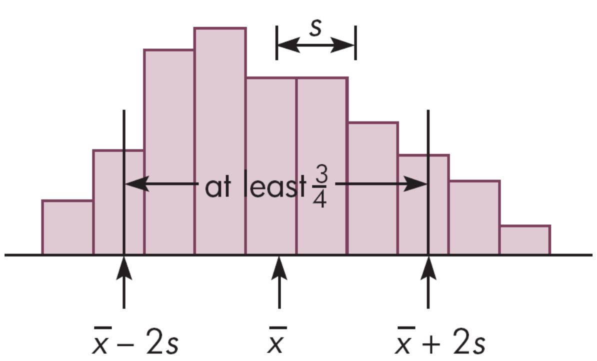
Chebyshev's theorem The proportion of any distribution that lies within *k* standard deviations of the $1 - \frac{1}{k^2}$n is at least,
, where *k* is any positive number greater than 1. This theorem applies to all distributions of data.

Chebyshev's Theorem

This theorem says that within 2 standard deviations of the mean (k = 2), you will always find at least 75% (that is, 75% or more) of the data:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75, \textbf{ at least 75}\,\%$$

Figure 2.34 shows a mounded distribution that illustrates at least 75%.



Chebyshev's Theorem with k = 2

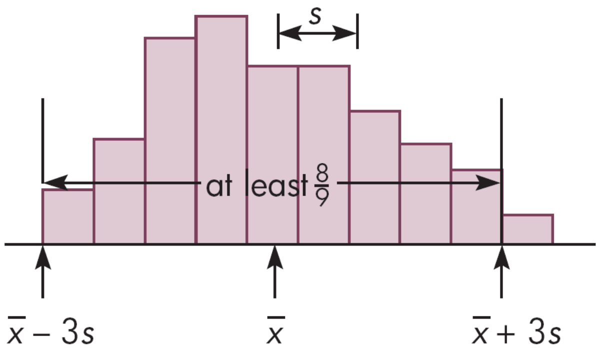
Figure 2.34

Chebyshev's Theorem

If we consider the interval enclosed by 3 standard deviations on either side of the mean ($k = 3$), the theorem says that we will always find at least 89% (that is, 89% or more) of the data:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} = 0.89, \textbf{ at least 89}\%$$

Figure 2.35 shows a mounded distribution that illustrates at least 89%.



Chebyshev's Theorem with $k = 3$

Figure 2.35