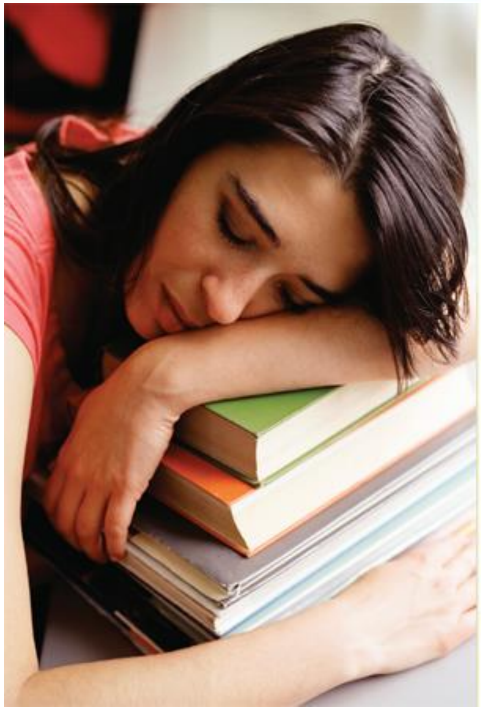
2 Descriptive Analysis and Presentation of Single-Variable Data



2.3 Measures of Central Tendency

Measures of Central Tendency

Measures of central tendency are numerical values that locate, in some sense, the center of a set of data. The term *average* is often associated with all measures of central tendency.

Mean (arithmetic mean) The average with which you are probably most familiar. The sample mean is represented by $\bar{x}$ (read "*x*-bar" or "sample mean").

The mean is found by adding all the values of the variable *x* (this sum of *x* values is symbolized Σx) and dividing the sum by the number of these values, *n* (the "sample size").

Measures of Central Tendency

We express this in formula form as

$$\textit{Sample mean:} \quad x\text{-}bar = \frac{\textit{sum of all } x}{\textit{number of } x}$$

$$\bar{x} = \frac{\sum x}{n} \qquad \textbf{(2.1)}$$

Median The value of the data that occupies the middle position when the data are ranked in order according to size. The sample median is represented by $\tilde{x}$ (read "*x*-tilde" or "sample median").

Procedure for Finding the Median

Procedure for Finding the Median

Step 1: Rank the data.

Step 2: **Determine the depth of the median.** The **depth,** or position (number of positions from either end), of the median is determined by the formula

depth of median: $$\textit{depth of median} = \frac{\textit{sample size} + 1}{2}$$

$$d(\tilde{x}) = \frac{n+1}{2} \qquad \textbf{(2.2)}$$

Procedure for Finding the Median

The median's depth (or position) is found by adding the position numbers of the smallest data (1) and the largest data (*n*) and dividing the sum by 2 (*n* is the number of pieces of data).

Step 3: **Determine the value of the median.** Count the ranked data, locating the data in the $d(\tilde{x})$th position. The median will be the same regardless of which end of the ranked data (high or low) you count from. In fact, counting from both ends will serve as an excellent check.

Procedure for Finding the Median

Mode The mode is the value of x that occurs most frequently.

If two or more values in a sample are tied for the highest frequency (number of occurrences), we say there is **no mode**.

For example, in the sample 3, 3, 4, 5, 5, 7, the 3 and the 5 appear an equal number of times. There is no one value that appears most often; thus, this sample has no mode.

Procedure for Finding the Median

Midrange The number exactly midway between a lowest-valued data, L, and a highest-valued data, H. It is found by averaging the low and the high values:

$$\text{midrange} = \frac{\textit{low value} + \textit{high value}}{2}$$

$$\text{midrange} = \frac{L + H}{2} \qquad \mathbf{(2.3)}$$

Applied Example 11 – *"Average" Means Different Things*

When it comes to convenience, few things can match that wonderful mathematical device called averaging. With an average, you can take a fistful of figures on any subject and compute one figure that will represent the whole fistful.

But there is one thing to remember. There are several kinds of measures ordinarily known as averages, and each gives a different picture of the figures it is called on to represent.

Applied Example 11 – *“Average” Means Different Things* cont’d

Take an example. Table 2.11 shows the annual incomes of 10 families.

\$54,000	\$39,000	\$37,500	\$36,750	\$35,250	\$31,500	\$31,500	\$31,500	\$31,500	\$25,500

Annual Incomes of 10 Families **[TA02-11]**

Table 2.11

What would this group’s “typical” income be? Averaging would provide the answer, so let’s compute the typical income by the simpler and more frequently used kinds of averaging.

Applied Example 11 – *“Average” Means Different Things* cont’d

- *The arithmetic mean.* This is the most common form of average, obtained by adding items in the data set, then dividing by the number of items; for these data, the arithmetic mean is $35,400.

  The mean is representative of the data set in the sense that the sum of the amounts by which the higher figures exceed the mean is exactly the same as the sum of the amounts by which the lower figures fall short of the mean.

  The higher incomes exceed the mean by a total of $25,650. The lower incomes fall short of the mean by a total of $25,650.

Applied Example 11 – *“Average” Means Different Things* cont’d

- *The median*. As you may have observed, six families earn less than the mean and four families earn more.

  You might wish to represent this varied group by the income of the family that is smack dab in the middle of the whole bunch. The median works out to $33,375.

- *The midrange.* Another number that might be used to represent the average is the midrange, computed by calculating the figure that lies halfway between the highest and lowest incomes: $39,750.

Applied Example 11 – *“Average” Means Different Things* cont’d

- *The mode.* So, three kinds of averages, and not one family actually has an income matching any of them. Say you want to represent the group by stating the income that occurs most frequently. That is called a mode. The modal income would be $31,500.

Four different averages are available, each valid, correct, and informative in its way. But how they differ!

arithmetic mean	*median*	*midrange*	*mode*
$35,400	$33,375	$39,750	$31,500

Applied Example 11 – *"Average" Means Different Things* cont'd

And they would differ still more if just one family in the group were millionaires—or one were jobless!

The large value of $54,000 (extremely different from the other values) is skewing the data out toward larger data values.

This skewing causes the mean and midrange to become much larger in value.

Procedure for Finding the Median

Round-off rule When rounding off an answer, let’s agree to keep one more decimal place in our answer than was present in the original information.

To avoid round-off buildup, round off only the final answer, not the intermediate steps. That is, avoid using a rounded value to do further calculations.