Sections 3-1 and 3-2

Overview and Measures of Center

TWO TYPES OF STATISTICS

• Descriptive statistics summarize or describe the important characteristics of data.
• Inferential statistics use sample data to make inferences (or generalizations) about a population.

MEASURE OF CENTER

A measure of center is a value at the center or middle of a data set.

There are four measures of center that we will discuss in this class
  • (Arithmetic) Mean
  • Median
  • Mode
  • Midrange
THE MEAN

The arithmetic mean, or the mean, of a set of data is the measure of center found by adding all the data values and dividing by the total number of data values. This is what is commonly referred to as the average.

PROPERTIES OF THE MEAN

• Sample means drawn from the same population tend to vary less than other measures of center.
• The mean of the data set uses every data value.
• A disadvantage of the mean is that just one extreme value (outlier) can change the value of the mean substantially. (Since the mean cannot resist substantial changes caused by extreme values, we say that the mean is not a resistant measure of center.

NOTATION

• Σ – denotes the summation (addition) of a set of values.
• x – is the variable usually used to represent the individual data values.
• n – represents the number of values in a sample.
• N – represents the number of values in a population.
MEAN OF A SAMPLE AND MEAN OF A POPULATION

• \( \bar{x} = \frac{\sum x}{n} \) is the mean of sample values.
• \( \mu = \frac{\sum x}{N} \) is the mean of population values.

EXAMPLE

The given values are the numbers of Dutchess County car crashes for each month in a recent year. Find the mean.

27 8 17 11 15 25 16 14 14 14 13 18

FINDING THE MEAN ON THE TI-83/84

1. Press STAT; select 1:Edit…
2. Enter your data values in L1. (You may enter the values in any of the lists.)
3. Press 2ND, MODE (for QUIT).
4. Press STAT; arrow over to CALC. Select 1:1-Var Stats.
5. Enter L1 by pressing 2ND, 1.
6. Press ENTER.
FINDING THE MEAN ON THE TI-84 WITH THE NEW OS

1. Press STAT; select 1:Edit…
2. Enter your data values in L1. (You may enter the values in any of the lists.)
3. Press 2ND, MODE (for QUIT).
4. Press STAT; arrow over to CALC. Select 1:1-Var Stats.
5. For "List" enter L1 by pressing 2ND, 1.
7. Highlight “Calculate” and press ENTER.

MEDIAN

The median, denoted by \( \bar{x} \), of a data set is the middle value when the original data values are arranged in order of increasing (or decreasing) magnitude.

PROPERTIES OF THE MEDIAN

- The median does not change by large amounts when we include just a few extreme values (so the median is a resistant measure of center).
- The median does not use every data value.
EXAMPLE

The given values are the numbers of Dutchess County car crashes for each month in a recent year. Find the median.

27 8 17 11 15 25 16 14 14 14 13 18

FINDING THE MEDIAN ON THE TI-83/84

This is done in exactly the same way as finding the mean. After you have finished use the down arrow to scroll down and you will see Med=.
MODE

- The mode, denoted by \(M\), of a data set is the number that occurs most frequently.
- When two values occur with the same greatest frequency, each one is a mode and the data set is said to be **bimodal**.
- When more than two values occur with the same greatest frequency, each is a mode and the data set is said to be **multimodal**.
- When no data value is repeated, we say that there is **no mode**.
- This is the **only** measure of center that can be used with **nominal** data.

EXAMPLE

| a. 5 5 5 3 1 5 1 4 3 5 |
| b. 1 2 2 2 3 4 5 6 6 7 9 |
| c. 1 2 3 6 7 8 9 10 |

\(\Rightarrow\) Mode is 5
\(\Rightarrow\) Bimodal - 2 and 6
\(\Rightarrow\) No Mode

EXAMPLE

The given values are the numbers of Dutchess County car crashes for each month in a recent year. Find the mode.

27 8 17 11 15 25 16 14 14 13 18

\(\Rightarrow\) Mode is 14
FINDING THE MODE ON THE TI-83/84

The TI-83/84 will **not** calculate the mode of a data set. However, the data in a list can be easily sorted to help in finding the mode.

To sort L1 in *ascending* order: STAT, 2:SortA, L1, ) and ENTER.

To sort in *descending* order, use 3:SortD.

MIDRANGE

The **midrange** of a data set is the measure of center that is the value midway between the highest and lowest values of the original data set. It is found by adding the highest data value and the lowest data value and then dividing by 2; that is,

\[
\text{midrange} = \frac{\text{maximum data value} + \text{minimum data value}}{2}
\]

PROPERTIES OF THE MIDRANGE

• Because the midrange uses only the maximum and minimum values, it is very sensitive to those extremes.
• In practice, the midrange is rarely used, but it has three redeeming features:
  1. The midrange is easy to compute.
  2. The midrange helps reinforce the very important point that there are several different ways to define the center of a data set.
  3. The value of the midrange is sometimes used incorrectly for the median, so confusion can be reduced by clearly defining the midrange along with the median.
EXAMPLE

The given values are the numbers of Dutchess County car crashes for each month in a recent year. Find the midrange.

27 8 17 11 15 25 16 14 14 14 13 18

ROUND-OFF RULES FOR MEAN, MEDIAN, AND MIDRANGE

For the mean, median, and midrange, carry one more decimal place than is present in the original set of values.

For the mode, leave the value without rounding (because values of the mode are the same as some of the original data values).

MEAN FROM A FREQUENCY DISTRIBUTION

To compute the mean from a frequency distribution, we assume that all sample values are equal to the class midpoint.

\[ \bar{x} = \frac{\Sigma(f \cdot x)}{\Sigma f} \]

\( x = \) class midpoint

\( f = \) frequency

\( \Sigma f = \) sum of frequencies = \( n \)
EXAMPLE

The following data represent the number of people aged 25–64 covered by health insurance in 2007. Approximate the mean age.

<table>
<thead>
<tr>
<th>Age</th>
<th>Number (in billions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25–34</td>
<td>39.8</td>
</tr>
<tr>
<td>35–44</td>
<td>41.9</td>
</tr>
<tr>
<td>45–54</td>
<td>43.8</td>
</tr>
<tr>
<td>55–64</td>
<td>33.3</td>
</tr>
</tbody>
</table>

Source: ebri.org Notes • August 2009 • Vol. 30, No. 8

FINDING THE MEAN FROM A FREQUENCY TABLE ON TI-83/84

1. Enter the class midpoints in L1.
2. Enter the frequencies in L2.
3. Press STAT, arrow over to CALC, and select 1:1-Var Stats.
4. Press L1,L2 followed by ENTER.
5. The mean will be the first item.

FINDING THE MEAN FROM A FREQUENCY TABLE ON TI-84 WITH NEW OS

1. Enter the class midpoints in L1.
2. Enter the frequencies in L2.
3. Press STAT, arrow over to CALC, and select 1:1-Var Stats.
4. For "List", enter L1
5. For "FreqList", enter L2.
6. Highlight "Calculate" and press ENTER.
7. The mean will be the first item.
WEIGHTED MEAN

In some cases, the values vary in their degree of importance, so we may want to weight them accordingly. We can compute the weighted mean for such values.

\[
\bar{x} = \frac{\sum (w \cdot x)}{\sum w}
\]

- \( x \) = value
- \( w \) = weight of value
- \( \sum w \) = sum of weights

EXAMPLE

Marissa just completed her first semester in college. She earned an "A" in her four-hour statistics course, a "B" in her three-hour sociology course, an "A" in her three-hour psychology course, a "C" in her five-hour computer programming course, and an "A" in her one-hour drama course. Determine Marissa's grade point average.

BEST MEASURE OF CENTER

- **Mean**: Add all data, then divide by number of data points.
- **Median**: The middle value when the data is ordered.
- **Mode**: The most frequent value in the data.
- **Midrange**: The average of the highest and lowest values.

The mean is sensitive to extreme values, often used more than the other measures of center.

The median is resistant to extreme values, often used when the data is skewed.

The mode is the most frequently occurring value in the data.

The midrange is the average of the highest and lowest values.

The best measure of center depends on the distribution of the data.