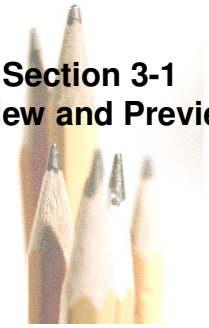
Chapter 3
Statistics for Describing, Exploring, and Comparing Data

Section 3-1
Review and Preview



Created by Tom Wegleitner, Centreville, Virginia

Review

- **Chapter 1**
  Distinguish between population and sample, parameter and statistic
  Good sampling methods: ***simple random sample***, collect in appropriate ways

- **Chapter 2**
  Frequency distribution: summarizing data
  Graphs designed to help understand data
  Center, variation, distribution, outliers, changing characteristics over time

Preview

- **Important Statistics**

  Mean, median, standard deviation, variance

- **Understanding and Interpreting**

  these important statistics

Preview

- **Descriptive Statistics**

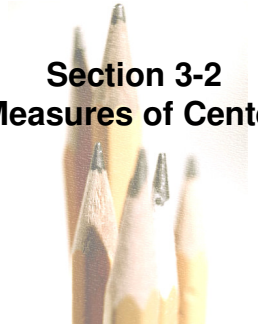
  In this chapter we'll learn to summarize or describe the important characteristics of a known set of data

- **Inferential Statistics**

  In later chapters we'll learn to use sample data to make inferences or generalizations about a population

2

Section 3-2
Measures of Center



Key Concept

Characteristics of center. Measures of center, including mean and median, as tools for analyzing data. Not only determine the value of each measure of center, but also interpret those values.

Part 1

Basics Concepts of Measures of Center

3

Measure of Center

- **Measure of Center**

  the value at the center or middle of a data set

Arithmetic Mean

- **Arithmetic Mean (Mean)**
  the measure of center obtained by adding the values and dividing the total by the number of values

  What most people call an *average*.

Notation

- Σ denotes the sum of a set of values.
- x is the variable usually used to represent the individual data values.
- n represents the number of data values in a sample.
- N represents the number of data values in a population.

4

Notation

$\overline{x}$ is pronounced 'x-bar' and denotes the mean of a set of sample values

$$\overline{x} = \frac{\Sigma x}{n}$$

μ is pronounced 'mu' and denotes the mean of all values in a population

$$\mu = \frac{\Sigma x}{N}$$

Mean

- **Advantages**
  Is relatively reliable, means of samples drawn from the same population don't vary as much as other measures of center
  Takes every data value into account

- **Disadvantage**
  Is sensitive to every data value, one extreme value can affect it dramatically; is not a *resistant* measure of center

Median

- **Median**
  the middle value when the original data values are arranged in order of increasing (or decreasing) magnitude

- often denoted by $\tilde{x}$ (pronounced 'x-tilde')

- is not affected by an extreme value - is a resistant measure of the center

Finding the Median

First *sort* the values (arrange them in order), the follow one of these

1. If the number of data values is odd, the median is the number located in the exact middle of the list.

2. If the number of data values is even, the median is found by computing the mean of the two middle numbers.

5.40 1.10 0.42 0.73 0.48 1.10

0.42 0.48 0.73 1.10 1.10 5.40

(in order - even number of values – no exact middle shared by two numbers)

$\frac{0.73 + 1.10}{2}$

MEDIAN is 0.915

5.40 1.10 0.42 0.73 0.48 1.10 0.66

0.42 0.48 0.66 0.73 1.10 1.10 5.40

(in order - odd number of values)

exact middle

MEDIAN is 0.73

Mode

- **Mode**
  the value that occurs with the greatest frequency
- Data set can have one, more than one, or no mode

Bimodal two data values occur with the same greatest frequency

Multimodal more than two data values occur with the same greatest frequency

No Mode no data value is repeated

Mode is the only measure of central tendency that can be used with nominal data

Mode - Examples

a. 5.40 1.10 0.42 0.73 0.48 1.10 ⇦ Mode is 1.10

b. 27 27 27 55 55 55 88 88 99 ⇦ Bimodal - 27 & 55

c. 1 2 3 6 7 8 9 10 ⇦ No Mode

Definition

- **Midrange**
  the value midway between the maximum and minimum values in the original data set

$$\text{Midrange} = \frac{\text{maximum value} + \text{minimum value}}{2}$$

Midrange

- **Sensitive to extremes**
  because it uses only the maximum and minimum values, so rarely used
- **Redeeming Features**
  (1) very easy to compute
  (2) reinforces that there are several ways to define the center
  (3) Avoids confusion with median

Round-off Rule for Measures of Center

Carry one more decimal place than is present in the original set of values.

Critical Thinking

Think about whether the results are reasonable.

Think about the method used to collect the sample data.

Part 2

Beyond the Basics of Measures of Center

Mean from a Frequency Distribution

Assume that all sample values in each class are equal to the class midpoint.

Mean from a Frequency Distribution

use class midpoint of classes for variable x

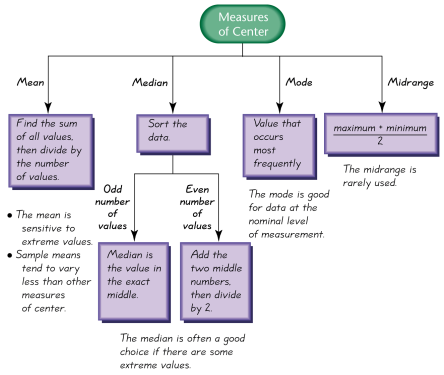
$$\bar{x} = \frac{\Sigma(f \cdot x)}{\Sigma f}$$

Weighted Mean

When data values are assigned different weights, we can compute a weighted mean.

$$\bar{x} = \frac{\Sigma(w \cdot x)}{\Sigma w}$$

Best Measure of Center



Skewed and Symmetric

- **Symmetric**

  distribution of data is symmetric if the left half of its histogram is roughly a mirror image of its right half

- **Skewed**

  distribution of data is skewed if it is not symmetric and extends more to one side than the other

Skewed Left or Right

- **Skewed to the left**

  (also called negatively skewed) have a longer left tail, mean and median are to the left of the mode

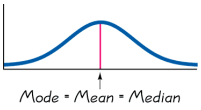
- **Skewed to the right**

  (also called positively skewed) have a longer right tail, mean and median are to the right of the mode

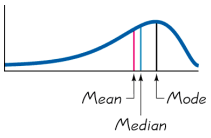
Shape of the Distribution

The mean and median cannot always be used to identify the shape of the distribution.

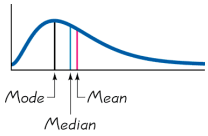
Skewness



(b) Symmetric



(a) Skewed to the Left (Negatively)



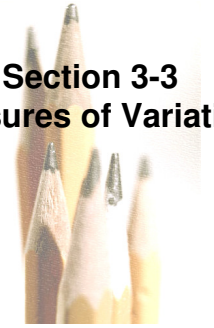
(c) Skewed to the Right (Positively)

Recap

In this section we have discussed:

- **Types of measures of center**
  - **Mean**
  - **Median**
  - **Mode**
- **Mean from a frequency distribution**
- **Weighted means**
- **Best measures of center**
- **Skewness**

Section 3-3
Measures of Variation



Key Concept

Discuss characteristics of variation, in particular, measures of variation, such as standard deviation, for analyzing data.

Make understanding and interpreting the standard deviation a priority.

Part 1

Basics Concepts of Measures of Variation

Definition

The range of a set of data values is the difference between the maximum data value and the minimum data value.

Range = (maximum value) – (minimum value)

It is very sensitive to extreme values; therefore not as useful as other measures of variation.

Round-Off Rule for Measures of Variation

When rounding the value of a measure of variation, carry one more decimal place than is present in the original set of data.

Round only the final answer, not values in the middle of a calculation.

Definition

The **standard deviation** of a set of sample values, denoted by *s*, is a measure of variation of values about the mean.

Sample Standard Deviation Formula

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}}$$

Sample Standard Deviation (Shortcut Formula)

$$s = \sqrt{\frac{n\Sigma(x^2) - (\Sigma x)^2}{n\,(n-1)}}$$