Chapter 4 – Summarizing Data Numerically
Section 1 – Measures of Center

Objectives
1. Compute the median of some data.
2. Describe the meaning of sigma notation.
3. Compute the arithmetic mean of some data.
4. Compare the means of two groups of data.
5. Compare the mean and the median of data.
6. Find the means and medians of bimodal and multimodal distributions.
7. Find the mode of some data.

Vocabulary
1. median
2. summation notation
3. arithmetic mean/mean
4. formula
5. sensitive/resistant to outliers
6. mode

Lesson/Activity
OBJECTIVE 1 – Compute the median of some data.

Definition: Median
The **median** of some data is the 50th percentile.

Finding the Median of a Distribution
To find the median of some data values, first list the observations from smallest to largest.
- If the number of observations is odd, then the median is the middle observation.
- If the number of observations is even, then the median is the average of the two middle observations.

1. The percentages of adults who exercise frequently are shown in the following table for Northeastern states and Mountain states.

Northeastern State	Percent	Mountain State	Percent
Connecticut	54	Arizona	53
Maine	55	Colorado	60
Massachusetts	53	Idaho	58
New Hampshire	54	Montana	60
New Jersey	48	Nevada	55
New York	49	New Mexico	57
Pennsylvania	50	Utah	54
Rhode Island	48	Wyoming	54
Vermont	65		

Source: The Gallup Organization

a. Find the median percent for the Northeastern states. What does it mean in this situation?
b. Find the median percent for the Mountain states. What does it mean in this situation?
c. Compare the results you found in Parts (a) and (b). What does this mean in this situation?
d. A student says that the result you found in Part (a) should be greater than the result you found in Part (b) because there are more Northeastern states than Mountain states. What would you tell the student?

OBJECTIVE 2 – Describe the meaning of sigma notation.

Definition: Summation Notation
Let $x_1, x_2, x_3, \ldots, x_n$ be some data values. The **summation notation** Σx_i stands for the sum of the data values: $\Sigma x_i = x_1 + x_2 + x_3 + \cdots + x_n$

2. The TV series with the top five most-watched finales and the finale viewerships are shown in the following table. Let x_i be the finale viewership (in millions) listed in the ith row of the table.
   a. Find the values of $x_1, x_2, x_3, x_4,$ and x_5.
   b. Find Σx_i. What does it mean in this situation?

TV Series	Finale Viewership (in millions)
M*A*S*H	105.9
Cheers	80.4
Seinfeld	76.3
Friends	52.5
Magnum, P.I.	50.7

Source: USA Today, Reuters, TV Guide, ABC

OBJECTIVE 3 – Compute the arithmetic mean of some data.

Definition: Mean
The **arithmetic mean** (or **mean**) of n data values $x_1, x_2, x_3, \ldots, x_n$ is given by $\bar{x} = \frac{\Sigma x_i}{n}$

- **Remember to always include the units of the mean.**
- A **formula** is an equation that contains two or more variables.

3. The numbers of teaspoons of sugar in 12 fluid ounces of various beverages are shown in the following table. Find the mean number of teaspoons of sugar in the beverages. Then construct a dotplot and verify that the mean is a reasonable measure of the center.

Beverage	Number of Teaspoons of Sugar
Mountain Dew®	11
Pepsi®	9
Ame Orange and Grape®	7
Dr. Pepper®	9
Fanta Grape Flavored Drink®	11
Coca-Cola®	9
Club Orange®	12
Tesco Original Cola®	8

Source: The companies

Mean is a Measure of the Center
If a distribution is unimodal and approximately symmetric, the mean is a reasonable measure of the center. In this case, we say the mean is a typical value.

4. The students in one of the author's prestatistics classes estimated his age. Their estimates are shown in the following table.

48	45	45	41	40	44	47	44
50	35	45	54	38	40	46	50
38	42	47	43	49	45	43	43
49							

Source: J. Lehmann

a. Determine whether the mean should be a reasonable measure of the center of the distribution.
b. Compute the mean. What does it tell us in this situation?
c. A student calculates the mean estimate to be 44.4 years and concludes that a typical estimate by all students at the college where the survey was performed is 44.4 years. What would you tell the student?

**When computing the mean or other measures of data,
we will round to one more decimal place than the data.**

OBJECTIVE 4 – Compare the means of two groups of data.

5. The prices of hot dogs at each of the stadiums for the baseball teams in the National League Central (NLC) and the National League West (NLW) are shown in the following table.

NLC Team	Hot Dog Price (dollars)	NLW Team	Hot Dog Price (dollars)
Chicago Cubs	5.50	Arizona Diamondbacks	2.75
Cincinnati Reds	1.00	Colorado Rockies	4.75
Milwaukee Brewers	3.50	Los Angeles Dodgers	5.50
Pittsburgh Pirates	3.25	San Diego Padres	4.00
St. Louis Cardinals	4.25	San Francisco Giants	5.25

Source: Team Marketing Report

a. Determine whether the mean hot dog price of the NLC distribution should be a reasonable measure of the center of the distribution. Do the same for the NLW distribution.
b. Find the mean hot dog price for the NLC.
c. Find the mean hot dog price for the NLW.
d. What do your results in Parts (b) and (c) mean in this situation?

6a. Compute the mean for the numbers 3, 4, and 5.
b. Compute the mean for the numbers 3, 4, 5, and 4.
c. A student says that the result you found in Part (b) should be larger than the result you found in Part (a) because there are more observations in Part (b). What would tell the student? Include in your explanation a discussion about division as well as a comparison of dotplots of the two data sets.

OBJECTIVE 5 – Compare the mean and the median of data.

7. The following data are the savings (in thousands of dollars) of five adults: 15, 9, 11, 6, and 3.
a. Find the mean savings.
b. Find the median savings.
c. Suppose that the adult who had $15 thousand in savings wins $800 thousand (after taxes) from the lottery. So, the adult's new savings is $815 thousand. Find the mean savings of the five adults after the lottery win.
d. Find the median savings of the five adults after the lottery win.
e. Describe how much the outlier $815 thousand affected the mean and the median.

The Effect of Outliers on the Mean and the Median
The **mean** is **sensitive** to outliers.
The **median** is **resistant** to outliers.

8. The top-paid female CEOs and their compensations are shown in the following table.

CEO	Company	Compensation (in billions of dollars)
Irene Rosenfeld	Mondelez	21.0
Virginia Rometty	IBM	19.3
Safra Ada Catz	Oracle	37.7
Phebe Novakovic	General Dynamics	19.3
Carol Meyrowitz	TJK	28.7
Indra Nooyi	PepsiCo	22.5
Meg Whitman	Hewlett-Packard	19.6
Debra Reed	Sempra	16.9
Marissa Mayer	Yahoo	42.1
Ursula Burns	Xerox	22.2
Marillyn Hewson	Lockheed Martin	33.7
Mary Barra	GM	16.2

Source: S&P Capital IQ, USA Today, Bespoke Investment Group

a. Find the mean compensation.
b. Find the median compensation.
c. Construct a frequency histogram and indicate the mean and the median on it. Describe the shape.
d. Which measures the center better, the mean or the median? Explain.
e. Explain why it makes sense that the mean is larger than the median.

How the Shape of a Distribution Affects the Mean and the Median
- If a distribution is skewed left, the mean is usually less than the median and the median is usually a better measure of the center.
- If a distribution is symmetric, the mean is approximately equal to the median and both are reasonable measures of the center.
- If a distribution is skewed right, the mean is usually greater than the median and the median is usually a better measure of the center.

OBJECTIVE 6 – Find the means and medians of bimodal and multimodal distributions.

9a. Consider how many days each president of the United States has served in office. What shape
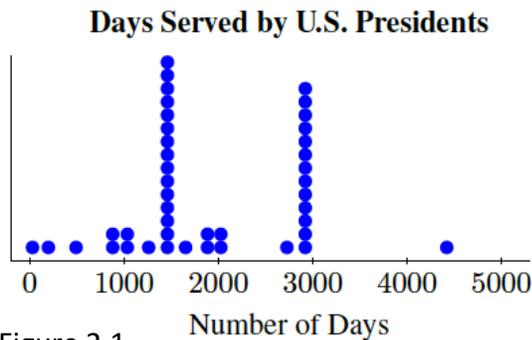would the distribution have?

Days Served by U.S. Presidents

Figure 2.1

Summary statistics:

Column	Mean	Median
Length (days)	1910.5581	1461

Figure 2.2

b. The numbers of days served by the presidents are described by the dotplot in Fig. 2.1. The mean
and median are shown in Fig. 2.2. How do the mean and the median relate to the distribution?
How well do they describe a typical observation?

c. What would be a more useful way to analyze this situation?

OBJECTIVE 7 – Find the mode of some data.

Definition: Mode
The **mode** of some data is an observation with the greatest frequency. There can be
more than one mode, but if all the observations have frequency 1, then there is no
mode.

10. For each data set, find the mode.
a. The author surveyed the students in one of his prestatistics classes about the number of states they
have lived in. Some of their responses are shown in the following table.

1	10	1	2	1	1
2	1	1	1	4	1

Source: J. Lehmann

b. The author surveyed the students in one of his prestatistics classes about what superpower they
wished they had. Some of their responses are shown in the following table.

| fly	teleport	mind read	fly	fly
invisible	fly	teleport	invisible	clone myself

Source: J. Lehmann

c. The tuitions of the five most expensive colleges are shown in the following table.

College	Tuition, Fees, Room, and Board (dollars)
Sarah Lawrence College	61,236
New York University	59,837
Harvey Mudd College	58,913
Columbia University	58,742
Wesleyan University	58,202

Source: CampusGrotto.com

11. The author surveyed students in one of his statistics classes about the how many times per week they eat at fast-food restaurants. A dotplot and the mean, median, and mode are shown in the following figure. Explain why the median and the mean (rounded to the ones place) are better measures of the center than the mode.

Fast-Food Restaurants

Number of visits per week
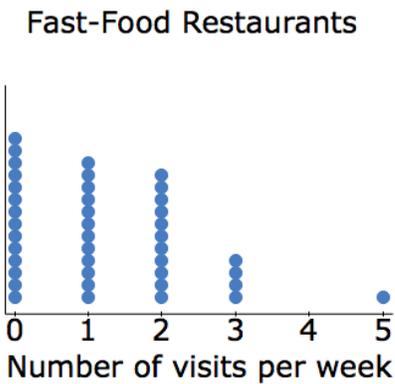
Summary statistics:

Column	n	Mean	Median	Mode
Visits	42	1.2142857	1	0

Homework/Assessment
1, 3, 5, 7, 15, 17, 19, 23, 27, 29, 35, 37, 43, 49, 53, 59, 63