# Chapter 3 – Numerically Summarizing Data

## OUTLINE

1. Measures of Central Tendency
2. Measures of Dispersion
3. Measures of Central Tendency and Dispersion from Grouped Data
4. Measures of Position
5. The Five-Number Summary and Boxplots

## Putting It Together

When we look at a distribution of data, we should consider three characteristics of the distribution: shape, center, and spread. In the last chapter, we discussed methods for organizing raw data into tables and graphs. These graphs (such as the histogram) allow us to identify the shape of the distribution: symmetric (in particular, bell shaped or uniform), skewed right, or skewed left.

The center and spread are numerical summaries of the data. The center of a data set is commonly called the average. There are many ways to describe the average value of a distribution. In addition, there are many ways to measure the spread of a distribution. The most appropriate measure of center and spread depends on the distribution’s shape.

Once these three characteristics of the distribution are known, we can analyze the data for interesting features, including unusual data values, called outliers.

## Section 3.1 Measures of Central Tendency

### Objectives

1. Determine the Arithmetic Mean of a Variable from Raw Data
2. Determine the Median of a Variable from Raw Data
3. Explain What It Means for a Statistic to be Resistant
4. Determine the Mode of a Variable from Raw Data

#### Objective 1: Determine the Arithmetic Mean of a Variable from Raw Data

Introduction, Page 1

*Answer the following after watching the video.*

1. What does a measure of central tendency describe?

Objective 1, Page 1

1. Explain how to compute the arithmetic mean of a variable.
2. What symbols are used to represent the population mean and the sample mean?

Objective 1, Page 2

1. List the formulas used to compute the population mean and the sample mean.

**Note:** Throughout this course, we agree to round the mean to one more decimal place than that in the raw data.

Objective 1, Page 3

**Example 1 *Computing a Population Mean and a Sample Mean***

Table 1 shows the first exam scores of the ten students enrolled in Introductory Statistics.

**Table 1**

| **Student** | **Score** |
| --- | --- |
| 1.Michelle | 82 |
| 2. Ryanne | 77 |
| 3. Bilal | 90 |
| 4. Pam | 71 |
| 5. Jennifer | 62 |
| 6. Dave | 68 |
| 7. Joel | 74 |
| 8. Sam | 84 |
| 9. Justine | 94 |
| 10. Juan | 88 |

1. Compute the population mean, .
2. Find a simple random sample of size *n* = 4 students.
3. Compute the sample mean, , of the sample found in part (B).

Objective 1, Page 5

*Answer the following after experimenting with the fulcrum animation*.

1. What is the mean of the data?
2. Explain why it is helpful to think of the mean as the center of gravity.

#### Objective 2: Determine the Median of a Variable from Raw Data

Objective 2, Page 1

1. Define the median of a variable.

Objective 2, Page 2

1. List the three steps in finding the median of a data set.

Objective 2, Page 3

**Example 2 *Determining the Median of a Data Set (Odd Number of Observations)***

Table 2 shows the length (in seconds) of a random sample of songs released in the 1970s. Find the median length of the songs.

**Table 2**

| **Song Name** | **Length** |
| --- | --- |
| "Sister Golden Hair" | 201 |
| "Black Water" | 257 |
| "Free Bird" | 284 |
| "The Hustle" | 208 |
| "Southern Nights" | 179 |
| "Stayin' Alive" | 222 |
| "We Are Family" | 217 |
| "Heart of Glass" | 206 |
| "My Sharona" | 240 |

Objective 2, Page 5

**Example 3 *Determining the Median of a Data Set (Even Number of Observations)***

Find the median score of the data in Table 1.

**Table 1**

| **Student** | **Score** |
| --- | --- |
| 1. Michelle | 82 |
| 2. Ryanne | 77 |
| 3. Bilal | 90 |
| 4. Pam | 71 |
| 5. Jennifer | 62 |
| 6. Dave | 68 |
| 7. Joel | 74 |
| 8. Sam | 84 |
| 9. Justine | 94 |
| 10. Juan | 88 |

#### Objective 3: Explain What It Means for a Statistic to be Resistant

Objective 3, Page 1

*Answer the following as you work through the Mean versus Median Applet.*

1. When the mean and median are approximately 2, how does adding a single observation near 9 affect the mean? How does it affect the median?
2. When the mean and median are approximately 2, how does adding a single observation near 24 affect the mean? The median?

Objective 3, Page 1 (continued)

1. When the mean and median are approximately 40, how does dragging the new observation from 35 toward 0 affect the mean? How does it affect the median?

Objective 3, Page 2

*Answer the following as you watch the video.*

1. Which measure, the mean or the median, is least affected by extreme observations?
2. Define what it means for a numerical summary of data to be resistant.
3. Which measure, the mean or the median, is resistant?

Objective 3, Page 3

1. State the reason that we compute the mean.

Objective 3, Page 7

*Answer the following as you work through Activity 2: Relation among the Mean, Median, and Distribution Shape.*

1. If a distribution is skewed left, what is the relation between the mean and median?
2. If a distribution is skewed right, what is the relation between the mean and median?

Objective 3, Page 7 (continued)

1. If a distribution is symmetric, what is the relation between the mean and median?

Objective 3, Page 11

1. Sketch three graphs showing the relation between the mean and median for distributions that are skewed left, symmetric, and skewed right.

Objective 3, Page 12

**Example 4 *Describing the Shape of a Distribution***

The data in Table 4 represent the birth weights (in pounds) of 50 randomly sampled babies.

1. Find the mean and median birth weight.
2. Describe the shape of the distribution.
3. Which measure of central tendency best describes the average birth weight?

**Table 4**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 5.8 | 7.4 | 9.2 | 7 | 8.5 | 7.6 |
| 7.9 | 7.8 | 7.9 | 7.7 | 9 | 7.1 |
| 8.7 | 7.2 | 6.1 | 7.2 | 7.1 | 7.2 |
| 7.9 | 5.9 | 7 | 7.8 | 7.2 | 7.5 |
| 7.3 | 6.4 | 7.4 | 8.2 | 9.1 | 7.3 |
| 9.4 | 6.8 | 7 | 8.1 | 8 | 7.5 |
| 7.3 | 6.9 | 6.9 | 6.4 | 7.8 | 8.7 |
| 7.1 | 7 | 7 | 7.4 | 8.2 | 7.2 |
| 7.6 | 6.7 |  |  |  |  |

#### Objective 4: Determine the Mode of a Variable from Raw Data

Objective 4, Page 1

1. Define the mode of a variable.
2. Under what conditions will a set of data have no mode?
3. Under what conditions will a set of data have two modes?

Objective 4, Page 2

**Example 5 *Finding the Mode of Quantitative Data***

The following data represent the number of O-ring failures on the shuttle Columbia for the 17 flights prior to its fatal flight:

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 2, 3

Find the mode number of O-ring failures.

Objective 4, Page 3

**Example 6 *Finding the Mode of Quantitative Data***

Find the mode of the exam score data listed in Table 1.

**Table 1**

| **Student** | **Score** |
| --- | --- |
| 1. Michelle | 82 |
| 2. Ryanne | 77 |
| 3. Bilal | 90 |
| 4. Pam | 71 |
| 5. Jennifer | 62 |
| 6. Dave | 68 |
| 7. Joel | 74 |
| 8. Sam | 84 |
| 9. Justine | 94 |
| 10. Juan | 88 |

Objective 4, Page 5

1. What does it mean when we say that a data set is bimodal? Multimodal?

Objective 4, Page 6

**Example 7 *Finding the Mode of Qualitative Data***

The data in Table 5 represent the location of injuries that required rehabilitation by a physical therapist. Determine the mode location of injury.

**Table 5**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Back | Back | Hand | Neck | Knee | Knee |
| Wrist | Back | Groin | Shoulder | Shoulder | Back |
| Elbow | Back | Back | Back | Back | Back |
| Back | Shoulder | Shoulder | Knee | Knee | Back |
| Hip | Knee | Hip | Hand | Back | Wrist |

Data from Krystal Catton, student at Joliet Junior College

Objective 4, Page 8

#### ***Summa***ry

1. List the conditions for determining when to use the following measures of central tendency.
2. Mean
3. Median
4. Mode

## Section 3.2 Measures of Dispersion

### Objectives

1. Determine the Range of a Variable from Raw Data
2. Determine the Standard Deviation of a Variable from Raw Data
3. Determine the Variance of a Variable from Raw Data
4. Use the Empirical Rule to Describe Data That Are Bell-Shaped

Introduction, Page 1

Measures of central tendency describe the typical value of a variable. We also want to know the amount of dispersion (or spread) in the variable. Dispersion is the degree to which the data are spread out.

Introduction, Page 2

**Example 1 *Comparing Two Sets of Data***

The data tables represent the IQ scores of a random sample of 100 students from two different universities.

For each university, compute the mean IQ score and draw a histogram, using a lower class limit of 55 for the first class and a class width of 15. Comment on the results.

#### Objective 1: Determine the Range of a Variable from Raw Data

Objective 1, Page 1

1. What is the range of a variable?

Objective 1, Page 2

**Example 2 *Computing the Range of a Set of Data***

The data in the table represent the first exam scores of 10 students enrolled in Introductory Statistics. Compute the range.

| **Student** | **Score** |
| --- | --- |
| 1. Michelle | 82 |
| 2. Ryanne | 77 |
| 3. Bilal | 90 |
| 4. Pam | 71 |
| 5. Jennifer | 62 |
| 6. Dave | 68 |
| 7. Joel | 74 |
| 8. Sam | 84 |
| 9. Justine | 94 |
| 10. Juan | 88 |

#### Objective 2: Determine the Standard Deviation of a Variable from Raw Data

Objective 2, Page 1

1. Explain how to compute the population standard deviation  and list its formula.

Objective 2, Page 2

**Example 3 *Computing a Population Standard Deviation***

Compute the population standard deviation of the test scores in Table 6.

**Table 6**

| **Student** | **Score** |
| --- | --- |
| 1. Michelle | 82 |
| 2. Ryanne | 77 |
| 3. Bilal | 90 |
| 4. Pam | 71 |
| 5. Jennifer | 62 |
| 6. Dave | 68 |
| 7. Joel | 74 |
| 8. Sam | 84 |
| 9. Justine | 94 |
| 10. Juan | 88 |

Objective 2, Page 5

1. If a data set has many values that are “far” from the mean, how is the standard deviation affected?

Objective 2, Page 6

1. Explain how to compute the sample standard deviation *s* and list its formula.

Objective 2, Page 7

1. What do we call the expression ?

Objective 2, Page 8

**Example 4 *Computing a Sample Standard Deviation***

In a previous lesson we obtained a simple random sample of exam scores and computed a sample mean of 73.75. Compute the sample standard deviation of the sample of test scores for that data.

Objective 2, Page 10

*Answer the following after you watch the video.*

1. Is standard deviation resistant? Why or why not?

Objective 2, Page 11

1. When comparing two populations, what does a larger standard deviation imply about dispersion?

Objective 2, Page 14

**Example 5 *Comparing the Standard Deviations of Two Sets of Data***

The data tables represent the IQ scores of a random sample of 100 students from two different universities.

Use the standard deviation to determine whether University A or University B has more dispersion in the IQ scores of its students.

Objective 2, Page 17

*Answer the following after using the applet in Activity 1: Standard Deviation as a Measure of Spread.*

1. Compare the dispersion of the observations in Part A with the observations in Part B. Which set of data is more spread out?
2. In Part D, how does adding a point near 10 affect the standard deviation? How is the standard deviation affected when that point is moved near 25? What does this suggest?

Objective 2, Page 18

*Watch the video to reinforce the ideas from Activity 1: Standard Deviation as a Measure of Spread.*

#### Objective 3: Determine the Variance of a Variable from Raw Data

Objective 3, Page 1

1. Define variance.

Objective 3, Page 2

**Example 6 *Determining the Variance of a Variable for a Population and a Sample***

In previous examples, we considered population data of exam scores in a statistics class. For this data, we computed a population mean of  points and a population standard deviation of  points. Then, we obtained a simple random sample of exam scores. For this data, we computed a sample mean of points and a sample standard deviation of *s* = 11.3points. Use the population standard deviation exam score and the sample standard deviation exam score to determine the population and sample variance of scores on the statistics exam.

Objective 3, Page 3

*Answer the following after you watch the video.*

1. Using a rounded value of the standard deviation to obtain the variance results in a round-off error. How should you deal with this issue?

Objective 3, Page 5

Whenever a statistic consistently underestimates a parameter, it is said to be **biased**. To obtain an unbiased estimate of the population variance, divide the sum of the squared deviations about the sample mean by .

#### Objective 4: Use the Empirical Rule to Describe Data That Are Bell-Shaped

Objective 4, Page 1

1. According to the Empirical Rule, if a distribution is roughly bell shaped, then approximately what percent of the data will lie within 1 standard deviation of the mean? What percent of the data will lie within 2 standard deviations of the mean? What percent of the data will lie within 3 standard deviations of the mean?

Objective 4, Page 2

1. Sketch the third part of Figure 5.

Objective 4, Page 3

**Example 7 *Using the Empirical Rule***

Table 9 represents the IQs of a random sample of 100 students at a university.

1. Determine the percentage of students who have IQ scores within 3 standard deviations of the mean according to the Empirical Rule.
2. Determine the percentage of students who have IQ scores between 67.8 and 132.2 according to the Empirical Rule.
3. Determine the actual percentage of students who have IQ scores between 67.8 and 132.2.
4. According to the Empirical Rule, what percentage of students have IQ scores between 116.1 and 148.3?

**Table 9**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 73 | 103 | 91 | 93 | 136 | 108 | 92 | 104 | 90 | 78 |
| 108 | 93 | 91 | 78 | 81 | 130 | 82 | 86 | 111 | 93 |
| 102 | 111 | 125 | 107 | 80 | 90 | 122 | 101 | 82 | 115 |
| 103 | 110 | 84 | 115 | 85 | 83 | 131 | 90 | 103 | 106 |
| 71 | 69 | 97 | 130 | 91 | 62 | 85 | 94 | 110 | 85 |
| 102 | 109 | 105 | 97 | 104 | 94 | 92 | 83 | 94 | 114 |
| 107 | 94 | 1121 | 113 | 115 | 106 | 97 | 106 | 85 | 99 |
| 102 | 109 | 76 | 94 | 103 | 112 | 107 | 101 | 91 | 107 |
| 107 | 110 | 106 | 103 | 93 | 110 | 125 | 101 | 91 | 119 |
| 118 | 85 | 127 | 141 | 129 | 60 | 115 | 80 | 111 | 79 |

## Section 3.3 Measures of Central Tendency and Dispersion from Grouped Data

### Objectives

1. Approximate the Mean of a Variable from Grouped Data
2. Compute the Weighted Mean
3. Approximate the Standard Deviation from a Frequency Distribution

#### Objective 1: Approximate the Mean of a Variable from Grouped Data

Objective 1, Page 1

1. Explain how to find the class midpoint.
2. List the formulas for approximating the population mean and sample mean from a frequency distribution.

Objective 1, Page 2

**Example 1 *Approximating the Mean for Continuous Quantitative Data from a Frequency Distribution***

The frequency distribution in Table 10 represents the five-year rate of return of a random sample of 40 large-blend mutual funds. Approximate the mean five-year rate of return.

**Table 10**

| **Class (5-year rate of return)** | **Frequency** |
| --- | --- |
| 8-8.99 | 2 |
| 9-9.99 | 2 |
| 10-10.99 | 4 |
| 11-11.99 | 1 |
| 12-12.99 | 6 |
| 13-13.99 | 13 |
| 14-14.99 | 7 |
| 15-15.99 | 3 |
| 16-16.99 | 1 |
| 17-17.99 | 0 |
| 18-18.99 | 0 |
| 19-19.99 | 1 |

Objective 1, Page 2 (Continued)

#### Objective 2: Compute the Weighted Mean

Objective 2, Page 1

1. When data values have different importance, or weights, associated with them, we compute the weighted mean. Explain how to compute the weighted mean and list its formula.

Objective 2, Page 2

**Example 2 *Computing the Weighted Mean***

Marissa just completed her first semester in college. She earned an A in her 4-hour statistics course, a B in her 3-hour sociology course, an A in her 3-hour psychology course, a C in her 5-hour computer programming course, and an A in her 1-hour drama course. Determine Marissa’s grade point average.

#### Objective 3: Approximate the Standard Deviation from a Frequency Distribution

Objective 3, Page 1

1. List the formulas for approximating the population standard deviation and sample standard deviation of a variable from a frequency distribution.

Objective 3, Page 2

**Example 3 *Approximating the Standard Deviation from a Frequency Distribution***

The frequency distribution in Table 11 represents the five-year rate of return of a random sample of 40 large-blend mutual funds. Approximate the standard deviation five-year rate of return.

**Table 11**

| **Class (5-year rate of return)** | **Frequency** |
| --- | --- |
| 8-8.99 | 2 |
| 9-9.99 | 2 |
| 10-10.99 | 4 |
| 11-11.99 | 1 |
| 12-12.99 | 6 |
| 13-13.99 | 13 |
| 14-14.99 | 7 |
| 15-15.99 | 3 |
| 16-16.99 | 1 |
| 17-17.99 | 0 |
| 18-18.99 | 0 |
| 19-19.99 | 1 |

## Section 3.4 Measures of Position

### Objective

1. Determine and Interpret *z*-Scores
2. Interpret Percentiles
3. Determine and Interpret Quartiles
4. Determine and Interpret the Interquartile Range
5. Check a Set of Data for Outliers

#### Objective 1: Determine and Interpret z-Scores

Objective 1, Page 1

1. What does a *z*-score represent?
2. Explain how to find a *z*-score and list the formulas for computing a population *z*-score and a sample *z*-score.
3. What does a positive *z*-score for a data value indicate? What does a negative z-score indicate?
4. What does a *z*-score measure?

Objective 1, Page 1 (continued)

1. How are *z*-scores rounded?

Objective 1, Page 2

**Example 1 *Determine and Interpret z-Scores***

Determine whether the Boston Red Sox or the Colorado Rockies had a relatively better run-producing season. The Red Sox scored 878 runs and play in the American League, where the mean number of runs scored was  and the standard deviation was  runs. The Rockies scored 845 runs and play in the National League, where the mean number of runs scored was  and the standard deviation was  runs.

Objective 1, Page 5

With negative *z*-scores, we need to be careful when deciding the better outcome. For example, when comparing finishing times for a marathon the lower score is better because it is more standard deviations below the mean.

#### Objective 2: Interpret Percentiles

Objective 2, Page 1

1. What does the *k*th percentile represent?

Objective 2, Page 2

**Example 2 *Interpreting a Percentile***

Jennifer just received the results of her SAT exam. Her math score of 600 is at the 74th percentile. Interpret this result.

#### Objective 3: Determine and Interpret Quartiles

Objective 3, Page 1

1. Define the first, second, and third quartiles.

Objective 3, Page 2

1. List the three steps for finding quartiles.

Objective 3, Page 3

**Example 3 *Finding and Interpreting Quartiles***

The Highway Loss Data Institute routinely collects data on collision coverage claims. Collision coverage insures against physical damage to an insured individual’s vehicle. Table 12 represents a random sample of 18 collision coverage claims based on data obtained from the Highway Loss Data Institute for 2007 models. Find and interpret the first, second, and third quartiles for collision coverage claims.

**Table 12**

|  |  |  |
| --- | --- | --- |
| $6751 | $9908 | $3461 |
| $2336 | $21,147 | $2,332 |
| $189 | $1185 | $370 |
| $1414 | $4668 | $1953 |
| $10,034 | $735 | $802 |
| $618 | $180 | $1657 |

#### Objective 4: Determine and Interpret the Interquartile Range

Objective 4, Page 1

1. Which measure of dispersion is resistant?
2. Define the interquartile range, IQR.

Objective 4, Page 2

**Example 4 *Finding and Interpreting the Interquartile Range***

Determine and interpret the interquartile range of the collision claim data from Table 12 in Example 3.

**Table 12**

|  |  |  |
| --- | --- | --- |
| $6751 | $9908 | $3461 |
| $2336 | $21,147 | $2332 |
| $189 | $1185 | $370 |
| $1414 | $4668 | $1953 |
| $10,034 | $735 | $802 |
| $618 | $180 | $1657 |

Objective 4, Page 4

1. If the shape of a distribution is symmetric, which measure of central tendency and which measure of dispersion should be reported?
2. If the shape of a distribution is skewed left or skewed right, which measure of central tendency and which measure of dispersion should be reported? Why?

#### Objective 5: Check a Set of Data for Outliers

Objective 5, Page 1

1. What is an outlier?

Objective 5, Page 2

1. List the four steps for checking for outliers by using quartiles.

Objective 5, Page 3

**Example 5 *Checking for Outliers***

Check the data in Table 12 on collision coverage claims for outliers.

**Table 12**

|  |  |  |
| --- | --- | --- |
| $6751 | $9,908 | $3461 |
| $2336 | $21,147 | $2332 |
| $189 | $1185 | $370 |
| $1414 | $4668 | $1953 |
| $10034 | $735 | $802 |
| $618 | $180 | $1657 |

## Section 3.5 The Five-Number Summary and Boxplots

### Objectives

1. Determine the Five-Number Summary
2. Draw and Interpret Boxplots

#### Objective 1: Determine the Five-Number Summary

Objective 1, Page 1

1. What values does the five-number summary consist of?

Objective 1, Page 2

**Example 1 *Obtaining the Five-Number Summary***

Table 13 shows the finishing times (in minutes) of the men in the 60- to 64-year-old age group in a 5-kilometer race. Determine the five-number summary of the data.

**Table 13**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 19.95 | 23.25 | 23.32 | 25.55 | 25.83 | 26.28 | 42.47 |
| 28.58 | 28.72 | 30.18 | 30.35 | 30.95 | 32.13 | 49.17 |
| 33.23 | 33.53 | 36.68 | 37.05 | 37.43 | 41.42 | 54.63 |

Data from Laura Gillogly, student at Joliet Junior College

#### Objective 2: Draw and Interpret Boxplots

Objective 2, Page 1

1. List the five steps for drawing a boxplot.

Objective 2, Page 2

**Example 2 *Constructing a Boxplot***

Use the results of Example 1 to construct a boxplot of the finishing times of the men in the 60- to 64-year-old age group.

(The five-number summary is: 19.95, 26.06, 30.95, 37.24, 54.63.)

Objective 2, Page 4

1. If the right whisker of a boxplot is longer than the left whisker and the median is left of the center of the box, what is the most likely shape of the distribution?

Objective 2, Page 5

When describing the shape of a distribution from a boxplot, be sure to justify your conclusion. Possible areas to discuss:

* Compare the length of the left whisker to the length of the right whisker
* The position of the median in the box
* Compare the distance between the median and the first quartile to the distance between the median and the third quartile
* Compare the distance between the median and the minimum value to the distance between the median and the maximum value

Objective 2, Page 10

**Example 3 *Comparing Two Distributions Using Boxplots***

Table 14 shows the red blood cell mass (in millimeters) for 14 rats sent into space (flight group) and for 14 rats that were not sent into space (control group). Construct side-by-side boxplots for red blood cell mass for the flight group and control group. Does it appear that space flight affects the rats' red blood cell mass?

**Table 14**

| **Flight** | **Control** |
| --- | --- |
| 7.43 | 8.65 |
| 7.21 | 6.99 |
| 8.59 | 8.40 |
| 8.64 | 9.66 |
| 9.79 | 7.62 |
| 6.85 | 7.44 |
| 6.87 | 8.55 |
| 7.89 | 8.70 |
| 9.30 | 7.33 |
| 8.03 | 8.58 |
| 7.00 | 9.88 |
| 8.80 | 9.94 |
| 6.39 | 7.14 |
| 7.54 | 9.14 |

Data from NASA Life Sciences Data Archive