CHAPTER TWO

Frequency Distributions

NOTE TO INSTRUCTORS

In this chapter, instructors should emphasize the importance of visually representing data. The chapter describes the different ways of organizing data in terms of a frequency distribution. The various shapes of distributions are also presented. Students often forget the importance of visually representing the data that they work with. As a result, it would be useful to show students how valuable visual representation can be by demonstrating how frequency distributions can be used to aid in getting a quick "snapshot" of data that are collected.

OUTLINE OF RESOURCES

I. Frequency Distributions
- Discussion Question 2-1
- Discussion Question 2-2
- Discussion Question 2-3
- Discussion Question 2-4
- Classroom Activity 2-1: Comparing Visualization Methods
- LaunchPad Statistical Applets: One-Variable Statistical Calculator

II. Shapes of Distributions
- Discussion Question 2-5
- Classroom Activity 2-2: Exploring Shapes of Distribution
- LaunchPad Video Resources
- Additional Reading
- Online Resources

III. Handouts
- Handout 2-1: Exploring Shapes of Distribution

CHAPTER GUIDE

I. Frequency Distributions

1. When we organize our data that is composed of **raw scores**, or data that has not yet been analyzed, it is useful to look at the

distribution of scores. The distribution allows us to examine the pattern of our data.

2. We organize our raw scores into a **frequency distribution**, which describes the pattern of a set of numbers by displaying a count or proportion for each possible value of a variable.
3. The best and simplest way to arrange our data is to use a **frequency table**, which visually displays the data so that we can see how often each value occurs.
4. To create a frequency table, we determine the range of our scores. Then, we create two columns. In the first column, add the highest value to the top of the column and the lowest value to the bottom. In the second column, mark the number of times each of these values has occurred in our data set.
5. Sometimes it is better to use a **grouped frequency table** that displays the frequencies for an interval rather than a specific value. A grouped frequency table is a better choice than a frequency table when the data are composed of continuous interval variables, cover a huge range, or are very large.

> Discussion Question 2-1

What is the difference between a frequency table and a grouped frequency table? When would you want to use one type rather than the other?

Your students' answers should include:

- A frequency table reports every value in a given data set, whereas a grouped frequency table reports intervals or ranges of values.
- A frequency table is used to depict data showing how often certain values occurred and how many scores were at each value. A grouped frequency table is used when the values are either:
  a. vast in number (such as when reporting hundreds of values); or
  b. several decimal places; or
  c. both vast in number and several decimal places long.

6. To create a grouped frequency table, find the highest and lowest scores in the distribution. If the highest and lowest values are decimals, round down. Subtract the lowest score from the highest score and add one. Next, determine the number of intervals and best interval size. List the intervals from lowest to highest in a column. Then, in the other column, count the number of values in each interval.

> Discussion Question 2-2

What steps are involved in creating a frequency table? A grouped frequency table?

Your students' answers should include:

- To create a frequency table:
  a. examine the data;

b. create two columns; in the first column record the values, putting highest at the top and lowest at the bottom;
c. tally the occurrence of each value; and
d. record the tallies in the second column.

- To create a grouped frequency table:
  a. find the highest and lowest scores;
  b. use the full range of data, but round scores down to whole numbers;
  c. determine the number of intervals and best interval size;
  d. determine which number will be the bottom of the lowest interval; and
  e. list the intervals from highest to lowest and then count the numbers of scores in each.

7. Another way to organize the data is to use a **histogram**. Histograms typically depict just one variable, usually based on scale data, with the values of the variable on the *x*-axis and the frequencies on the *y*-axis. Each bar represents the frequencies for each value or interval.

8. To create a histogram, start with a frequency table. Draw your *x*- and *y*-axis and label them with your variable of interest. Draw a bar for each value, centering the bar over that value on the *x*-axis. The bar should be as high as the frequency for that value.

9. Histograms can also be created from a grouped frequency table. Instead of values, the midpoints of the intervals are listed on the *x*-axis. The remaining steps are the same as those that you used when constructing a histogram from a frequency table.

10. **Frequency polygons** are another way of visually representing our data using a line graph, where the *x*-axis represents the value (or interval midpoint) and the *y*-axis represents the frequency. Frequency polygons are similar to histograms except that dots are used instead of bars and a line is used to connect the dots.

> Discussion Question 2-3

What is the difference between a histogram and a frequency polygon?

Your students' answers should include:

- A histogram looks like a bar graph and often depicts interval data, with the values of the variables represented on the *x*-axis and the frequencies represented on the *y*-axis.
- A frequency polygon is a line graph depicting interval data. It also represents values on the *x*-axis and frequencies on the *y*-axis.

> Discussion Question 2-4

What steps are involved in creating a histogram? A frequency polygon?

Your students' answers should include:

- To create a histogram:
  a. determine the midpoint for each interval, if needed;

b. draw and label the *x*-axis and the *y*-axis of a graph; and
c. draw a bar for each value.

- To create a frequency polygon:
  a. determine the midpoint for each interval, if needed;
  b. draw and label the *x*-axis and the *y*-axis;
  c. mark a dot above each value and connect the dots with a line; and
  d. add hypothetical values at both ends of the *x*-axis and mark dots for the frequency of 0 for each value to create a grounded shape rather than a floating line.

Classroom Activity 2-1
Comparing Visualization Methods

In this exercise, you will work with one set of data and compare how the different visualization methods capture the data. You could collect any sort of information from your students, perhaps using the data that you collected earlier to demonstrate different measurement scales, (e.g., height, year in school, age, etc.) and then demonstrate how frequency table, grouped frequency table, histogram, and so forth., might be used to visualize the data.

LaunchPad Statistical Applets
One-Variable Statistical Calculator

This applet calculates standard numerical statistics (e.g., mean, standard deviation, quartiles) and shows graphical displays (a histogram and a stemplot) of one-variable data sets. You can choose to view data sets from the textbook, or enter your own set of data.

II. Shapes of Distributions

1. A **normal distribution** refers to a bell-shaped, symmetrical, and unimodal frequency distribution.
2. We could also have a **skewed distribution**. Skewed distributions are distributions where one of the tails of the distribution is pulled away from the center.
3. When the tail of our distribution extends to the right, we say that our data are **positively skewed**. We typically observe positively skewed data when there is a **floor effect—**when a variable is prevented from taking values below a certain point.
4. Data can also be **negatively skewed**, meaning that the tail of our distribution extends to the left. We may observe negatively skewed data in the case of a **ceiling effect**—when a variable is prevented from taking values above a certain point.

> Discussion Question 2-5
What is skewness? What is the difference between the two different types of skewness?

Your students' answers should include:

- Skewness is the amount that a tail of a distribution is pulled away from the center.
  - **a.** Positively skewed data: The tail of the distribution extends to the right.
  - **b.** Negatively skewed data: The tail of the distribution extends to the left.

Classroom Activity 2-2
Exploring Shapes of Distribution

In this exercise, students will generate examples of two variables.

- Have the students predict whether the variables will be positively skewed or negatively skewed.
- Students can then develop questionnaires in groups to measure these two variables.
- Have them hand out versions of their questionnaires in class to see if they were correct in their predictions.

See Handout 2-1.

LaunchPad Video Resources

Snapshots: Data and Distributions
StatClips Examples: Exploratory Pictures for Quantitative Data, Example A
StatClips Examples: Summaries and Pictures for Categorical Data, Examples A and B
StatClips: Summaries and Pictures for Categorical Data

Additional Reading

Moore, Thomas L., Ed. (2001). *Teaching Statistics: Resources for Undergraduates.* Mathematical Association of America.

This book is an instructor's manual for teaching undergraduate statistics that advocates a hands-on approach.

Online Resources

The following Web site provides a wealth of information on statistics: http://www.math.yorku.ca/SCS/StatResource.html.

For information on good and bad visual graphic presentations, see: http://www.datavis.ca/gallery/index.php.

HANDOUT 2-1: EXPLORING SHAPES OF DISTRIBUTION

Directions: Answer the following questions regarding skewness.

1. What two variables do you think would generate data from your class that would be: positively skewed or negatively skewed? Why do you think the data would follow this pattern?

2. Develop short questionnaires to measure these two variables.

3. Distribute these questionnaires to your classmates and draw histograms of the data collected. Were you correct in any of your predictions? Why or why not?