

2

Frequency Distributions

NOTE TO INSTRUCTORS

This chapter shows students how to summarize data by making frequency distributions. Students do not usually have trouble with the difference between ungrouped and grouped frequency distributions, but they often struggle with how to set up a grouped frequency distribution so it is important to spend more time on the grouped distributions. You need to emphasize that the smaller numbers of the scale need to go at the bottom so that the cumulative frequency and cumulative percentage make sense. Some students have trouble with how the frequency *accumulates*. **Figure 2.1** in the text provides an excellent visual example of how the numbers get added up. **Table 2.5** in the text provides students with a good summary of the decisions that need to be made when setting up a frequency distribution. Students can usually find the real limits of whole numbers, but you should provide examples of how to calculate the real limits of numbers with different levels of decimal places. I have provided you with a handout to give them practice. I have also found that students find it difficult to determine whether a scale is discrete or continuous, so they need to see many examples to get a feel for the difference. It is very important that students learn to see the relationship between the frequency distribution and the graphs. Later in the course when we move to using the normal curve model, students seem to forget that the curve represents a frequency distribution. I have provided handouts that will give the students practice going back and forth between the two representations of data. I have also provided materials that allow the students to practice describing the shapes of distributions and showing they know the shape provides useful information.

OUTLINE OF RESOURCES

1. **Ungrouped and grouped frequency distributions**
 - Discussion question 2-1 (p. 10)
 - Discussion question 2-2 (p. 11)
 - Classroom activity 2-1 (p. 11)
2. **Discrete and continuous numbers**
 - Classroom activity 2-2 (p. 12)
3. **Real limits of numbers**
 - Discussion question 2-3 (p. 12)
 - Classroom activity 2-3 (p. 12)

4. Three types of graphs

- Discussion question 2-4 (p. 13)
- Classroom activity 2-4 (p. 13)

5. Describing the shapes of distributions

- Discussion question 2-5 (p. 13)
- Classroom activity 2-5 (p. 14)
- Classroom activity 2-6 (p. 14)

6. Handouts

- Handout 2-1: Making and comparing frequency distributions (p. 15)
- Handout 2-2: Identifying discrete and continuous numbers (p. 16)
- Handout 2-3: Computing the real limits of numbers (p. 17)
- Handout 2-4: Graphing frequency distributions and graph paper (p. 18)
- Handout 2-5: Interpreting graphs (pp. 19–21)

7. PowerPoint slide template

- Slide 2-1: Kurtosis and variability (p. 22)

CHAPTER GUIDE

1. Ungrouped and grouped frequency distributions

a) Ungrouped frequency distributions can be used when the data take on only a few values. Grouped frequency distributions are used when the variable takes on many values and the full distribution cannot fit on a single page.

b) Frequency distributions need to have a clear title and labeled columns.

c) The first column has the values of the variable, the next column is the frequency, the cumulative frequency, then percentage, and the cumulative percentage. Not all of the columns are required.

Discussion question 2-1

Why is it the number of values of the variable and NOT the number of cases that determines whether you used an ungrouped or a grouped frequency distribution?

Your students' answers should include:

- The number of values determines how many groups there are, but there can be many cases within a single group. Each row in the distribution represents a different group, and you can only view a limited number of rows at a time.

Discussion question 2-2

Why do we start with the lowest values of the variable at the bottom of a frequency distribution?

Your students' answers should include:

- So that the cumulative frequency adds up to the total number of cases at the top of the distribution.
- So that the cumulative percentage equals 100 at the top of the distribution.

Classroom activity 2-1

The goal of this activity is to have students practice making ungrouped and grouped frequency distributions. Additionally, by having different students make different distributions from the same data and then comparing the distributions, students can learn about how their choices about the distribution will influence how well others can understand the distribution and what the distribution communicates about the data.

Instructions

There are five different data sets on **Handout 2-1**. Have students work in teams (pairs or if you have a large class you can use groups of 3 to 4). For each data set, you need five teams of students. Team one makes an ungrouped distribution of the data. Team two makes a grouped distribution with an interval size of 2. Team three makes a grouped distribution with an interval size of 5, team four uses an interval size of 10, and team five uses an interval size of 20. You can expand this to use a greater variety of interval sizes depending on the size of the class. You may also want to eliminate some of the obviously bad choices, for example, use an interval size of 20 when the data go from 1 to 20, or doing an ungrouped distribution when the range of the data is 100 values. Each data set is designed to have a different optimal interval size. Once students have finished creating their distributions, you can have the class compare the five (or more) versions for each data set and discuss which version is best and why. Students should discover that too small an interval does not reduce the data enough and that too large an interval loses too much information.

The following table shows the different groups for this activity:

Data Set	Interval Size				
	Ungrouped	2	5	10	20
A	Team 1	Team 2	Team 3	Team 4	Team 5
B	Team 6	Team 7	Team 8	Team 9	Team 10
C	Team 11	Team 12	Team 13	Team 14	Team 15
D	Team 16	Team 17	Team 18	Team 19	Team 20

2. Discrete and continuous numbers

- Discrete numbers can only take on whole values.
- Continuous numbers can take on decimal values.

Classroom activity 2-2

Give the students copies of **Handout 2-2** on identifying discrete and continuous numbers. The students can work alone or in small groups to identify which variables are going to result in discrete numbers and which in continuous numbers. When most of them seem done you can call on people to share their answers and to explain how they decided which was discrete and which was continuous.

3. Real limits of numbers

In continuous data the real limits of a number are half a unit above and half a unit below the stated value.

Discussion question 2-3

Why do we compute the real limits of continuous numbers but not discrete numbers?

Your students' answers should include:

- Only continuous numbers can take on decimal values.
- We assume that with more sophisticated measuring devices we could get more precise measurements that would include more decimal places.
- The real limits imply that given our current measurement the “real value” of the number could be within a half a unit above or below the measured value.

Classroom activity 2-3

Give students copies of **Handout 2-3** on determining the real limits of data at different levels of precision. They can work alone or in small groups to identify the real limits of each number. When they have had some time to work on it, you can call on people to share their answers. It is important that students can correct any wrong answers so that the handout can be used as a study aid when preparing for an exam.

4. Three different types of graphs

a) Graphs are another way to represent distributions of data. Many people find a graph easier and faster to interpret than a frequency distribution.

b) Bar graphs are used when you have discrete data. In bar graphs the bars do not touch each other to represent that the data are discrete.

c) Histograms can be used to represent continuous data. In a histogram the bars touch each other and start and end at the real limits to indicate that the data is continuous.

d) Frequency polygons can also be used to represent continuous data. The frequency polygon (or line graph) has dots at the frequency of the midpoint of each interval. The dots are then connected by lines.

Discussion question 2-4

Why do the bars in some bar graphs touch but the bars in other bar graphs do not touch? What does this tell us about the data?

Your students' responses should include:

- When the bars do not touch it means the numbers are discrete.
- When the bars touch it means the numbers are continuous.

Classroom activity 2-4

Give each student a copy of **Handout 2-4** and a copy of the graph paper. Students can work in pairs or alone. There are three frequency distributions on the handout. Students should make a bar graph of the first distribution, which has discrete values. They should make a histogram for the second and a frequency polygon (line graph) for the third distribution, both of which have continuous data. Remind students to create careful labeling and to provide titles. When students have finished, they can swap graphs with a neighbor, and they can give each other feedback on the graphs, pointing out any errors or omissions. It would be helpful to get a good version of each graph to show to the class as a whole and to point out what makes a good graph. The bullet points on pages **52** (histogram) and **54** (frequency polygon) of the textbook provide a checklist for the qualities of a good graph.

5. Describing the shapes of distributions

a) The shapes of distributions can vary on three factors: modality, skew, and kurtosis.

b) Modality refers to the number of peaks in distribution, unimodal is one peak, bimodal is two peaks, and three or more peaks is called multimodal.

c) Skew refers to the symmetry of the distribution. If the distribution is symmetric it is said to have no skew. If the peak of the data is toward the right with a shallow tail toward the left, this is called a negative skew (the tail points to the negative numbers). If the peak of the data is toward the left with a shallow tail toward the right, this is called a positive skew (the tail points to the positive numbers).

d) Kurtosis is the degree to which the distribution is peaked. A normal distribution with a rounded high point is called mesokurtic, a sharp peak is called leptokurtic, and a flat distribution is called platykurtic. If you want students to remember these names a useful mnemonic is that platykurtic is flat like a plateau, leptokurtic provides a nice sharp peak that you could leap off of to go gliding, and meso is in the middle.

Discussion question 2-5

If the distribution is skewed to the right, do most of the cases have higher numbers or lower numbers? Does the tail represent most of the data?

Your students' answers should include:

- In a right-skewed distribution, most of the cases have lower numbers because most of the data is to the left.
- The tail represents data values with few cases.

Classroom activity 2-5

*We will see that kurtosis is related to variability. Look at the first graph on **Slide 2-1** (provided on page 22). Do most of the participants have very similar scores, or do lots of people have different scores? Now look at the second graph on **Slide 2-1**. Does this distribution have more or less variability than the first one? Why?*

Your students' answers should include:

- The first graph is leptokurtic so most of the people have the same score.
- Only a few people had different scores.
- The second graph is platykurtic, so the scores are very spread out. There are many people at different values on the graph.

Classroom activity 2-6

Many students have trouble interpreting graphs, so this is a good time to start learning how to interpret graphical representations of data. **Handout 2-5** has descriptions of several experiments and graphs of the data from the experiments. Have students get into small groups to discuss what the graphs tell them about the results of the experiments. When they are done, ask students to share their answers.

Answer key:

- Graph 1 shows the results of having men and women do a mental rotation task.
- The graph is bimodal because there is a gender difference in that men tend to score better than women. There is also a very wide range of scores indicating that some people are very good and some are very bad at this task.
- Graph 2 shows the results of asking psychology majors how many mathematics classes they have taken in college. The graph is leptokurtic and positively skewed because most psychology majors take one or two math classes but a few will take many more.
- Graph 3 shows the results of asking students their favorite color. The graph is platykurtic because there are an almost equal number of cases for each color.
- Graph 4 shows the grades on a first statistics exam. The graph is negatively skewed because most students did fairly well but a few did very poorly.

HANDOUT 2-1

Making and Comparing Frequency Distributions

Use the following data sets to make a frequency distribution. Each team should use a different interval size. Your instructor will tell you which data set to use and which interval size. When you are done, compare your results with the results of other teams who used the same data set. Which interval size worked best for your data set? Why?

DATA SET A

7,1,7,6,8,17,3,4,7,6,17,2,7,8,7,17,15,7,20,16,17,18,3,17,6,7,8,12,17,17,14,12,17,19,20,7,17,3,15,7

DATA SET B

1,12,20,13,14,16,15,17,19,18,10,16,19,15,17,16,14,2,12,15,17,20,14,18,13,16,9,18,15,14,12,19,17,13,3,16,14,15,18,20,12,4,17,19,13,16,15

DATA SET C

99,5,86,55,97,82,13,44,32,91,88,68,72,75,25,35,48,100,94,86,80,84,77,91,88,94,83,87,91,51,82,90,33

DATA SET D

22,10,22,5,22,30,22,41,50,22,1,22,2,17,22,45,22,25,32,22,25,22,21,15,22,9,7,5,22,11,14,22,18,16,22,20,21,22,26,19,18,22,27,22

HANDOUT 2-2

Identifying Discrete and Continuous Numbers

For each variable, identify whether the measurement is most likely to result in discrete or continuous data.

1. annual income in dollars _____
2. top five favorite movie _____
3. marital status _____
4. rating degree of enjoyment on a five-point scale _____
5. zip codes _____
6. socioeconomic status _____
7. IQ scores _____
8. number of parking tickets _____
9. number of Facebook friends _____
10. class rank _____
11. degree of extroversion on a 20-point scale _____

HANDOUT 2-3

Computing the Real Limits of Numbers

The real limits of numbers are one-half unit above and below the stated number.

For the number 10, for example, the real limits are 9.5 and 10.5.

For the number 10.6, the real limits are 10.55 and 10.65.

For the number 10.77, the real limits are 10.765 and 10.775.

Helpful hint: The real limits always have an extra decimal place.

Find the real limits of

1) 27

2) 14.2

3) 135.44

4) 0

5) -2.6

6) 15567.2

7) 1.1

8) 5.65554

HANDOUT 2-4

Graphing Frequency Distributions

Following the guidelines on pages 51 and 52 of your textbook, make a bar graph of the following frequency distribution.

\underline{X}	f
8	9
7	12
6	10
5	8
4	6
3	2
2	1
1	1

Following the guidelines on page 52 of your textbook, make a histogram for the following distribution.

\underline{X}	f
79.5–89.5	5
69.5–79.5	15
59.5–69.5	6
49.5–59.5	7
39.5–49.5	6
29.5–39.5	15
19.5–29.5	4
9.5–19.5	4

Following the guidelines on page 54 of your textbook, make a frequency polygon (line graph) for the following distribution.

\underline{X}	f
35.5–40.5	4
30.5–35.5	5
25.5–30.5	0
20.5–25.5	6
15.5–20.5	7
10.5–15.5	8
5.5–10.5	6
0.5–5.5	4

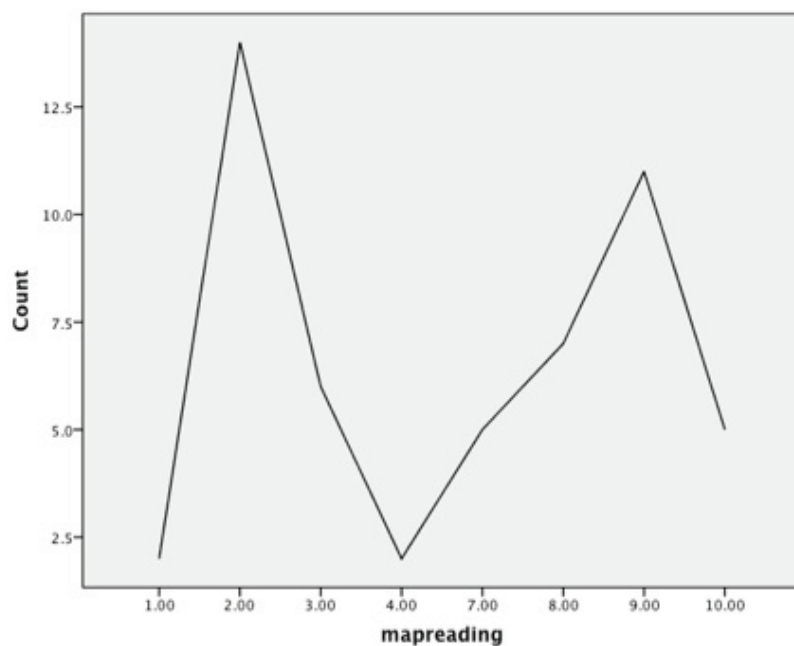
HANDOUT 2-5

Interpreting Graphs

Each of the following examples provides a description of the source of the data and then a graph of the resulting data. Describe the shape of each graph in terms of modality, skew, and kurtosis. Then explain what the graph tells you about the data. Interpret the graph in terms of the source of the data.

Graph A

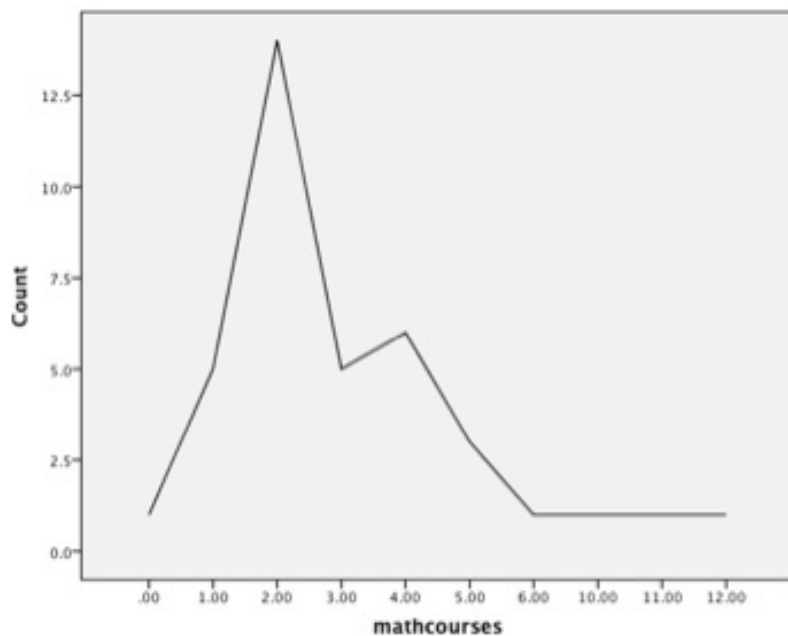
People were asked to do a series of map reading tasks. The scores are graphed below. What does this graph tell you about people's map reading abilities?



[2A]

Graph B

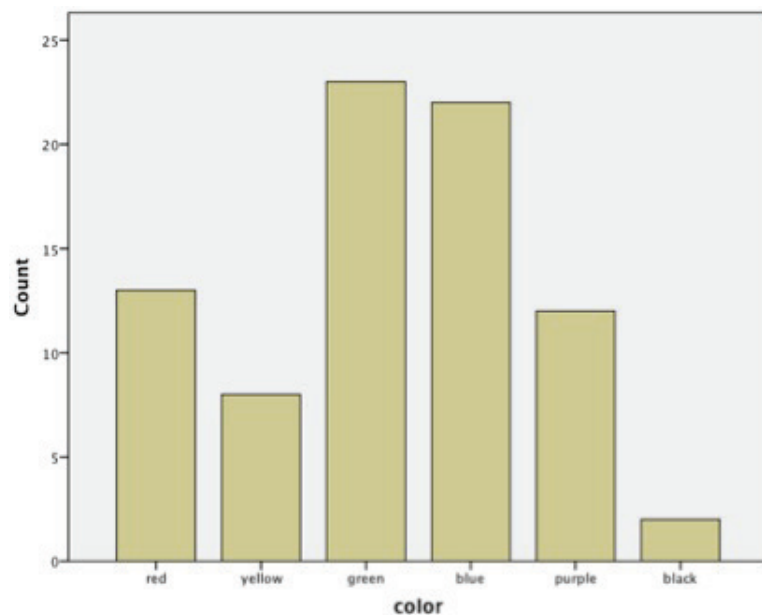
Psychology majors were surveyed to find out how many mathematics classes they have taken in college. The results are graphed below. What does this graph tell you about psychology majors taking math courses?



[2B]

Graph C

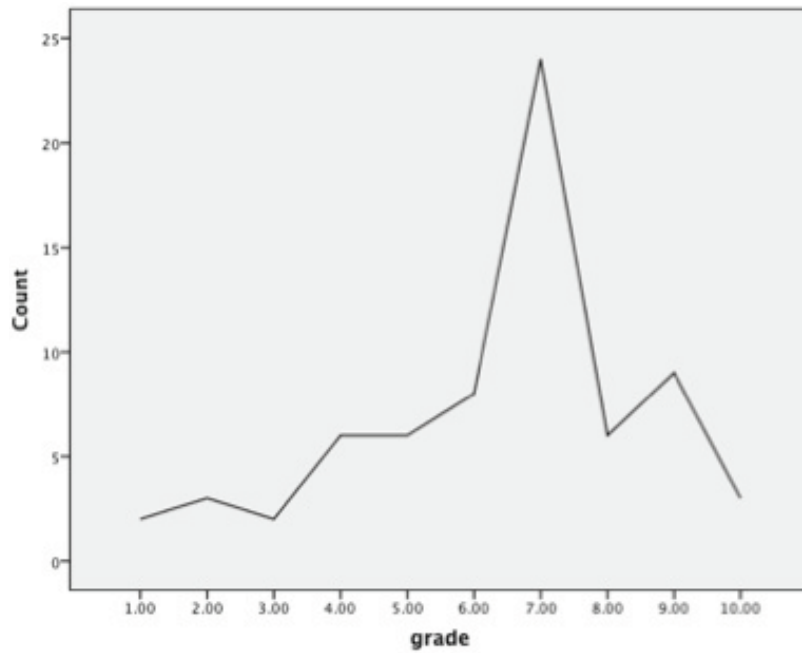
College students were asked to indicate their favorite color. The results are presented in a bar chart below. What does this chart tell you about color preferences?



[2C]

Graph D

A statistics professor made a graph of the grades on the first statistics exam. What does this tell you about the first exam? Why would this graph be useful for students in the class?



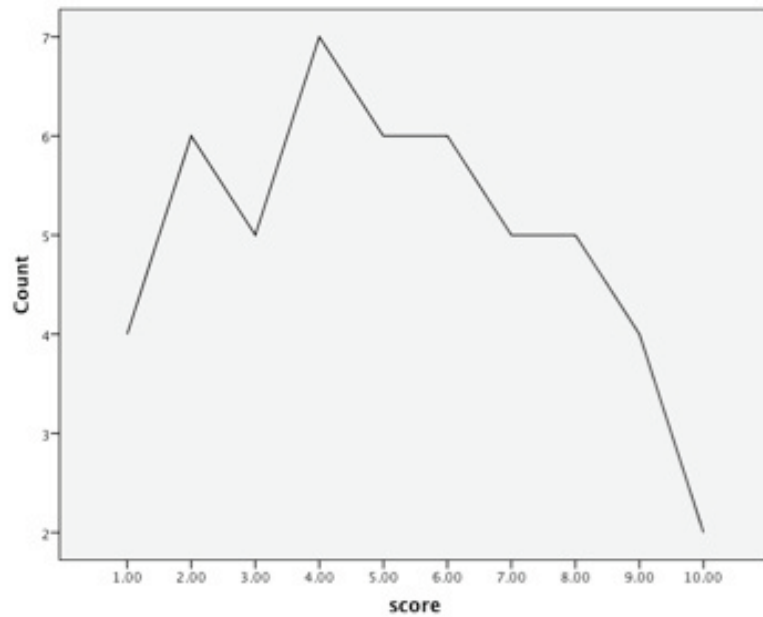
[2D]

POWERPOINT SLIDE 2-1

Kurtosis and Variability

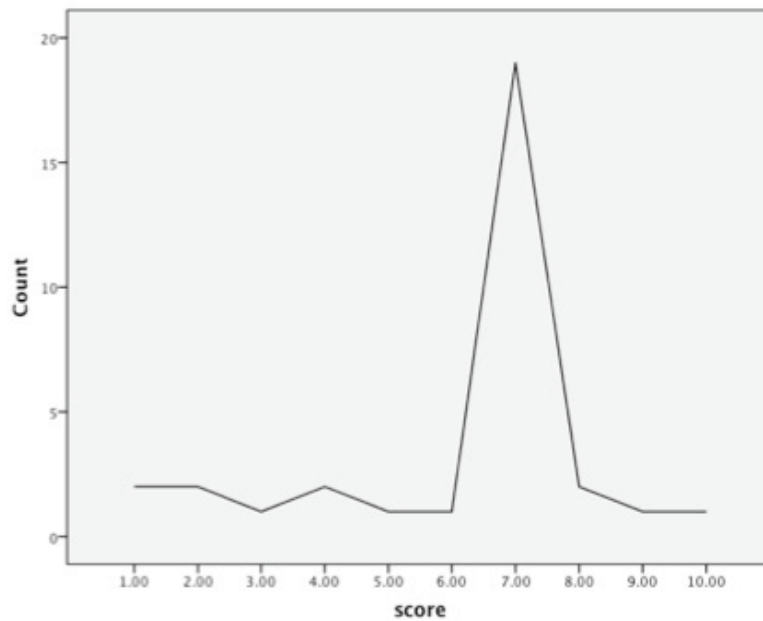
Compare the two graphs. In which graph do the scores vary more?

Graph A



[2E]

Graph B



[2F]