**Section 2.2 Organizing Quantitative Data: The Popular Displays**

***Objective 1: Organize Discrete Data in Tables***

Objective 1, Page 1

1. What do we use to create the classes when the number of distinct data values of a discrete variable is small?

Objective 1, Page 2

**Example 1 *Constructing Frequency and Relative Frequency Distributions from Discrete Data***

The manager of a Wendy’s® fast-food restaurant wants to know the typical number of customers who arrive during the lunch hour. The data represent the number of customers who arrive at Wendy’s for 40 randomly selected 15-minute intervals of time during lunch. Construct a frequency and relative frequency distribution.

**Number of Arrivals at Wendy’s**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 7 | 6 | 6 | 6 | 4 | 6 | 2 | 6 |
| 5 | 6 | 6 | 11 | 4 | 5 | 7 | 6 |
| 2 | 7 | 1 | 2 | 4 | 8 | 2 | 6 |
| 6 | 5 | 5 | 3 | 7 | 5 | 4 | 6 |
| 2 | 2 | 9 | 7 | 5 | 9 | 8 | 5 |

**This section deals with summarizing**

**quantitative data.**

When summarizing quantitative data,

**first determine whether the data are discrete or continuous.**

If the data are discrete with relatively few values

of the variable, then the categories of the data,

called classes, will be the observation,

just like it was for qualitative data.

**However, if the data are discrete**

**but with many different values of the variable,**

or if the data are continuous, then the categories

of the data, the classes, must be

created using intervals of numbers, such as 10 to 19,

20 to 29, and so on.

We first present the techniques for organizing

discrete quantitative data when there

are relatively few different values, and then proceed to organize in continuous quantitative data.

OBJECTIVE 1 Organize Discrete Data in Tables

Use the values of the discrete variable to create the classes when the number of distinct data values is small. The approach to summarizing the data is similar to that of constructing frequency or relative frequency distributions from qualitative data where the categories of data are determined by the actual observations.

Problem

The manager of a Wendy's® fast-food restaurant wants to know the typical number of customers who arrive during the lunch hour. The data in Table 8 represent the number of customers who arrive at Wendy's for 40 randomly selected 15-minute intervals of time during lunch. For example, during one 15-minute interval, seven customers arrived. Construct a frequency and relative frequency distribution.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **TABLE 8** | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | --- | | **Number of Arrivals at Wendy's** | | | | | | | | | 7 | 6 | 6 | 6 | 4 | 6 | 2 | 6 | | 5 | 6 | 6 | 11 | 4 | 5 | 7 | 6 | | 2 | 7 | 1 | 2 | 4 | 8 | 2 | 6 | | 6 | 5 | 5 | 3 | 7 | 5 | 4 | 6 | | 2 | 2 | 9 | 7 | 5 | 9 | 8 | 5 | |  |

**Frequency and Relative Frequency Distributions from Discrete Data**

1. Enter the raw data into the spreadsheet. Name the column variable.
2. Select **Stat**, highlight **Tables**, and select **Frequency**.
3. Click on the variable you want to summarize. Click the Type of table you want. If you want both Frequency and Relative Frequency, highlight Frequency; then press Ctrl (or Command on an Apple) and select Relative frequency. Click Compute!.

|  |  |  |
| --- | --- | --- |
| 6 | 11 | 0.275 |
| 5 | 7 | 0.175 |
| 2 | 6 | 0.15 |
| 7 | 5 | 0.125 |
| 4 | 4 | 0.1 |
| 9 | 2 | 0.05 |
| 8 | 2 | 0.05 |
| 11 | 1 | 0.025 |
| 3 | 1 | 0.025 |
| 1 | 1 | 0.025 |

2.2 Organizing Quantitative Data: The Popular Displays

Objective 2, Page 1

Explain how a histogram is constructed?

The *histogram,* a graph used to present quantitative data, is similar to the bar graph.

**DEFINITION**

A **histogram** is constructed by drawing rectangles for each class of data. The height of each rectangle is the frequency or relative frequency of the class. The width of each rectangle is the same, and the rectangles touch each other

Objective 2, Page 2

**Example 2 *Drawing a Histogram of Discrete Data***

Construct a frequency histogram and a relative frequency histogram using the data in Table 9. Recall that this table summarizes the data for the number of customers who arrive at Wendy's for 40 randomly selected 15-minute intervals of time during lunch.

**Table 9**

| **Number of Customers** | **Frequency** | **Relative Frequency** |
| --- | --- | --- |
| 1 | 1 | 0.25 |
| 2 | 6 | 0.15 |
| 3 | 1 | 0.025 |
| 4 | 4 | 0.1 |
| 5 | 7 | 0.175 |
| 6 | 11 | 0.275 |
| 7 | 5 | 0.125 |
| 8 | 2 | 0.05 |
| 9 | 2 | 0.05 |
| 10 | 0 | 0.0 |
| 11 | 1 | 0.025 |

OBJECTIVE 2, PAGE 2 (CONTINUED)

Technology Step-By-Step

**Drawing a Histogram**

1. If necessary, enter the raw data into the spreadsheet. Name the column variable.
2. Select **Graph** and highlight **Histogram**.
3. Click on the variable you want to summarize. Choose the type of histogram (frequency or relative frequency). You have the option of choosing a lower class limit for the first class by entering a value in the cell marked "Bins: Start at:". You have the option of choosing a class width by entering a value in the cell marked "Bins: Width:". Enter labels for the X-axis and Y-axis. Enter a title for the graph. Click Compute!.

OBJECTIVE 3 Organize Continuous Data in Tables

When a data set consists of a large number of different discrete data values or when a data set consists of continuous data, create classes by using intervals of numbers.

Table 10 is a typical frequency distribution created from continuous data. The data represent the number of U.S. residents, ages 25 to 74, who had a bachelor's degree or higher in 2016.

| **TABLE 10** | |
| --- | --- |
| **Age** | **Number (in thousands)** |
| 25–34 | 16,206 |
| 35–44 | 15,102 |
| 45–54 | 14,373 |
| 55–64 | 12,865 |
| 65–74 | 8,775 |
| Data from U.S. Census Bureau | |

**3.lower class limit** (the smallest value within the class)

**4. upper class limit** (the largest value within the class)

5.The **class width** is the difference between consecutive lower class limits

6. Notice that the classes in Table 10 **do not overlap**. This is necessary to avoid confusion as to which class a data value belongs. Notice also that the class widths are equal for all classes

1. Notice that the data are categorized, or grouped, by intervals of numbers. Each interval represents a class. For example, the first class is 25- to 34-year-old U.S. residents who have a bachelor's degree or higher. We read this interval as follows: “The number of U.S. residents, ages 25 to 34, with a bachelor's degree or higher was 16,206,000 in 2016.” There are five classes in the table, each with a **lower class limit** (the smallest value within the class) and an **upper class limit** (the largest value within the class). The lower class limit for the first class in Table 10 is 25; the upper class limit is 34. The **class width** is the difference between consecutive lower class limits. In Table 10, the class width is 35−25=10. The data in Table 10 are continuous. So the class 25−34 actually represents 25−34.999…, or 25 up to every value less than 35.
2. Notice that the classes in Table 10 do not overlap. This is necessary to avoid confusion as to which class a data value belongs. Notice also that the class widths are equal for all classes.

**One exception to the requirement of equal class widths occurs in open-ended tables**. A table is **open-ended** if the **first class has no lower class limit or the last class has no upper class limit**. The data in Table 11 represent the number of births to unmarried mothers in 2015 in the United States. The last class in the table, “40 and older,” is open-ended.

| **TABLE 11** | |
| --- | --- |
| **Age** | **Number of Births (in thousands)** |
| 15–19 | 204 |
| 20–24 | 561 |
| 25–29 | 435 |
| 30–34 | 252 |
| 35–39 | 117 |
| 40 and older | 30 |
| Data from National Vital Statistics Report, Vol. 66, No. 1 | |

 Organize Continuous Data in Tables

2.2 Organizing Quantitative Data: The Popular Displays

OBJECTIVE 3 Organize Continuous Data in Tables

When a data set consists of a large number of different discrete data values or when a data set consists of continuous data, create classes by using intervals of numbers.

Table 10 is a typical frequency distribution created from continuous data. The data represent the number of U.S. residents, ages 25 to 74, who had a bachelor's degree or higher in 2016.

| **TABLE 10** | |
| --- | --- |
| **Age** | **Number (in thousands)** |
| 25–34 | 16,206 |
| 35–44 | 15,102 |
| 45–54 | 14,373 |
| 55–64 | 12,865 |
| 65–74 | 8,775 |
| Data from U.S. Census Bureau | |

Notice that the data are categorized, or grouped, by intervals of numbers. Each interval represents a class. For example, the first class is 25- to 34-year-old U.S. residents who have a bachelor's degree or higher. We read this interval as follows: “The number of U.S. residents, ages 25 to 34, with a bachelor's degree or higher was 16,206,000 in 2016.” There are five classes in the table, each with a **lower class limit** (the smallest value within the class) and an **upper class limit** (the largest value within the class). The lower class limit for the first class in Table 10 is 25; the upper class limit is 34. The **class width** is the difference between consecutive lower class limits. In Table 10, the class width is 35−25=10. The data in Table 10 are continuous. So the class 25−34 actually represents 25−34.999…, or 25 up to every value less than 35.

Notice that the classes in Table 10 do not overlap. This is necessary to avoid confusion as to which class a data value belongs. Notice also that the class widths are equal for all classes.

2.2 Organizing Quantitative Data: The Popular Displays

One exception to the requirement of equal class widths occurs in open-ended tables. A table is **open-ended** if the first class has no lower class limit or the last class has no upper class limit. The data in Table 11 represent the number of births to unmarried mothers in 2015 in the United States. The last class in the table, “40 and older,” is open-ended.

| **TABLE 11** | |
| --- | --- |
| **Age** | **Number of Births (in thousands)** |
| 15–19 | 204 |
| 20–24 | 561 |
| 25–29 | 435 |
| 30–34 | 252 |
| 35–39 | 117 |
| 40 and older | 30 |
| Data from National Vital Statistics Report, Vol. 66, No. 1 | |

2.2 Organizing Quantitative Data: The Popular Displays

EXAMPLE 3 Organizing Continuous Data into a Frequency and Relative Frequency Distribution

Problem

Suppose you are considering investing in a Roth IRA. You collect the data in Table 12, which represent the five-year rate of return (in percent, adjusted for sales charges) for a simple random sample of 40 large-blend mutual funds. Construct a frequency and relative frequency distribution of the data.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **TABLE 12** | | | | | | | | | --- | --- | --- | --- | --- | --- | --- | --- | | **Five-Year Rate of Return of Mutual Funds (in percent)** | | | | | | | | | 10.94 | 14.60 | 12.80 | 16.00 | 11.93 | 15.68 | 9.03 | 13.40 | | 10.53 | 13.98 | 13.86 | 12.36 | 13.54 | 9.94 | 13.93 | 13.63 | | 14.12 | 14.88 | 14.77 | 13.13 | 8.28 | 19.43 | 12.98 | 13.16 | | 12.26 | 14.20 | 14.80 | 13.26 | 13.67 |  |  |  | |  |  |  |  |  |  |  |  | | Data from Morningstar.com | | | | | | | | |

**Frequency and Relative Frequency Tables-Organize Continuous Data**

1. If necessary, enter the raw data into the spreadsheet. Name the column variable.
2. Select **Data** and highlight **Bin.**
3. Click the variable you want to summarize. Click the “Use fixed width bins” radio button. Enter the lower class limit of the first class in the “Bins: Start at:” cell. Enter the class width in the “Bins: Width:” cell. Leave the ”Include left endpoint” radio button selected. Click Compute!.
4. Select **Stat** and highlight **Tables,** then **Frequency.**
5. Click the Bin(column name) variable. Under Type:, select Frequency and Relative Frequency. Click Compute!

### Frequency table results for Bin(Return):

1. Count = 40

| **Bin(Return)** | **Frequency** | **Relative Frequency** |
| --- | --- | --- |
| 8 to 9 | 2 | 0.05 |
| 9 to 10 | 2 | 0.05 |
| 10 to 11 | 4 | 0.1 |
| 11 to 12 | 1 | 0.025 |
| 12 to 13 | 6 | 0.15 |
| 13 to 14 | 13 | 0.325 |
| 14 to 15 | 7 | 0.175 |
| 15 to 16 | 3 | 0.075 |
| 16 to 17 | 1 | 0.025 |
| 19 to 20 | 1 | 0.025 |

***Objective 4: Construct Histograms of Continuous Data***

Objective 4, Page 2

**Example 4 *Drawing a Histogram of Continuous Data***

Construct a frequency and relative frequency histogram of the five-year rate of return data discussed in Example 3.

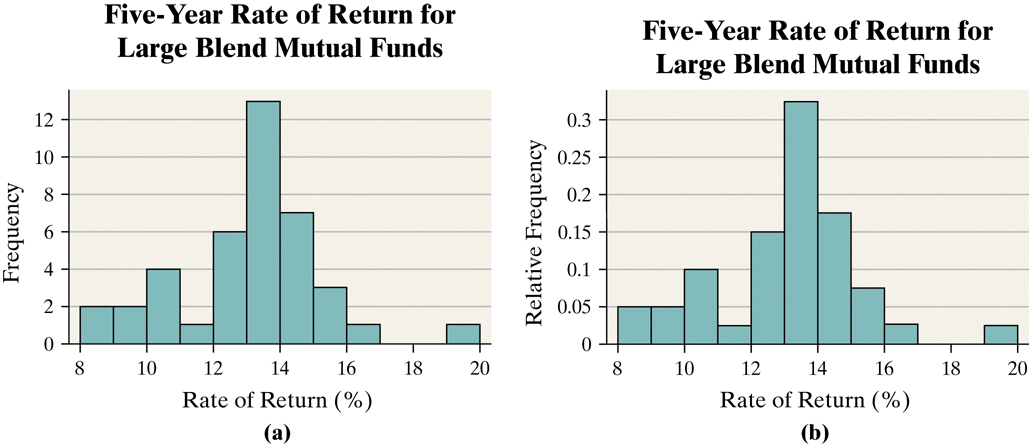
**Table 12 Five-Year Rate of Return of Mutual Funds (in percent)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 10.94 | 14.60 | 12.80 | 16.00 | 11.93 | 15.68 | 09.03 | 13.40 |
| 10.53 | 13.98 | 13.86 | 12.36 | 13.54 | 09.94 | 13.93 | 13.63 |
| 14.12 | 14.88 | 14.77 | 13.13 | 08.28 | 19.43 | 12.98 | 13.16 |
| 12.26 | 14.20 | 14.80 | 13.26 | 13.67 | 10.08 | 14.86 | 8.71 |
| 12.17 | 10.26 | 15.22 | 13.36 | 13.55 | 13.90 | 15.64 | 12.80 |

Data from [Morningstar.com](http://morningstar.com/)

**Draw a Histogram**

1. If necessary, enter the raw data into the spreadsheet. Name the column variable.
2. Select **Graph** and highlight **Histogram**.
3. Click on the variable you want to summarize. Choose the type of histogram (frequency or relative frequency). You have the option of choosing a lower class limit for the first class by entering a value in the cell marked "Bins: Start at:". You have the option of choosing a class width by entering a value in the cell marked "Bins: Width:". Enter labels for the X-axis and Y-axis. Enter a title for the graph. Click Compute!.



Objective 4, Page 7 Objective 4, Page 4

There is no one correct frequency distribution for a particular set of data. However, some frequency distributions better illustrate patterns within the data than others. So constructing frequency distributions is somewhat of an art form. Use the distribution that seems to provide the best overall summary of the data.

Objective 4, Page 5

*Answer the following after using the applet in Activity 1: Choosing Class Width.*

**Constructing Histograms Is Somewhat of an Art Form**

In Examples 3 and 4, the choices of the lower class limit of the first class and the class width were rather arbitrary. Although formulas and procedures exist for creating frequency distributions from raw data, they do not necessarily provide better summaries.

There is no one correct frequency distribution for a particular set of data. However, some frequency distributions better illustrate patterns within the data than others. So constructing frequency distributions is somewhat of an art form. Use the distribution that seems to provide the best overall summary of the data.

Next, you will use an applet to explore how changing the class width and the lower class limit of the first class affects the appearance of a histogram. As you use the applet, remember: The goal is to design a distribution that is best for revealing the patterns within the data

Part A. 6 classes

Part B. 3 classes

Part C. 10 classes

Part D. 3 classes

What happens to the number of classes as the bin width increases? They shrink as the width increases.

1. The number of classes in a frequency distribution is typically between what two numbers?

5 - 20

1. Explain how to choose the lower class limit of the first class in a frequency distribution.

Choose the smallest observation in the data set or a convenient number slightly smaller than the smallest observation in the data set. For example, in [Table 12](https://xlitemprod.pearsoncmg.com/assignment/containerassignmentplayer.aspx#xln-lb-lnk_obj5_7_ca0e9b2c-0d81-43b1-9f0f-b703e2ae3e3a), the smallest observation in 8.28. A convenient lower class limit of the first class is 8.

1. Once you decide on the number of classes, explain how to determine the class width.

**Guidelines for Determining the Lower Class Limit of the First Class and Class Width**

**Choosing the Lower Class Limit of the First Class**

Choose the smallest observation in the data set or a convenient number slightly smaller than the smallest observation in the data set. For example, in [Table 12](https://xlitemprod.pearsoncmg.com/assignment/containerassignmentplayer.aspx#xln-lb-lnk_obj5_7_ca0e9b2c-0d81-43b1-9f0f-b703e2ae3e3a), the smallest observation in 8.28. A convenient lower class limit of the first class is 8.

**Determining the Class Width**

* Decide on the number of classes. Generally, there should be between 5 and 20 classes. The smaller the data set, the fewer the classes. For example, we might think ten classes would be a good choice for the data in Table 12.
* Determine the class width by computing

Class width≈largest data value − smallest data valuenumber of classes

* Round the value to a convenient number. For example, using the data in Table 12, we obtain class width ≈19.43−8.2810=1.115. Round this down to 1 because this is an easy number to work with.

Rounding up may result in fewer classes than were originally intended, while rounding down may result in more class than originally intended.

***Objective 5: Draw Dot Plots***

Objective 5, Page 1

1. Explain how to draw a dot plot. We draw a **dot plot** by placing each observation horizontally in increasing order and placing a dot above the observation each time it is observed

Objective 5, Page 2

**Example 5 *Drawing a Dot Plot***

Draw a dot plot for the data from Table 8.

**Table 8 Number of Arrivals at Wendy’s**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 7 | 6 | 6 | 6 | 4 | 6 | 2 | 6 |
| 5 | 6 | 6 | 11 | 4 | 5 | 7 | 6 |
| 2 | 7 | 1 | 2 | 4 | 8 | 2 | 6 |
| 6 | 5 | 5 | 3 | 7 | 5 | 4 | 6 |
| 2 | 2 | 9 | 7 | 5 | 9 | 8 | 5 |

**Drawing a Dot Plot**

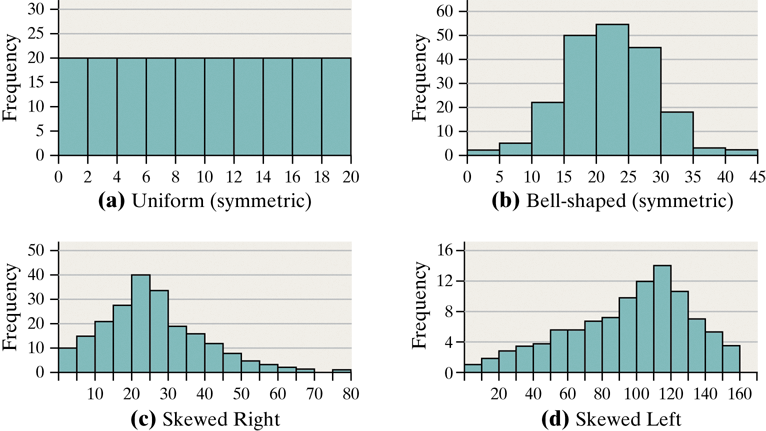
1. If necessary, enter the raw data into the spreadsheet. Name the column variable.
2. Select **Graph** and highlight **Dotplot.**
3. Click on the variable you want to summarize. Enter labels for the X-axis and Y-axis. Enter a title for the graph. Click Compute!.

***Objective 6: Identify the Shape of a Distribution***

One way that a variable is described is through the shape of its distribution. Distribution shapes are typically classified as *symmetric, skewed left,* or *skewed right.* Figure 9 displays various histograms and the shape of the distribution.

Figures 9(a) and (b) show symmetric distributions. They are symmetric because if we split the histogram down the middle, the right and left sides are mirror images. Figure 9(a) is a **uniform distribution** because the frequency of each value of the variable is evenly spread across the values of the variable. Figure 9(b) displays a **bell-shaped distribution** because the highest frequency occurs in the middle and frequencies tail off to the left and right of the middle. The distribution in Figure 9(c) is **skewed right**. Notice that the tail to the right of the peak is longer than the tail to the left of the peak. Finally, Figure 9(d) illustrates a distribution that is **skewed left** because the tail to the left of the peak is longer than the tail to the right of the peak.

It is important to recognize that data will not always exhibit behavior that perfectly matches the any of the shapes in Figure 9. To identify the shape of a distribution, some flexibility is required. In addition, people may disagree on the shape because identifying shape is subjective. Therefore, you should always support the shape of the distribution.



Objective 6, Page 1

1. Draw an example of a uniform distribution.
2. Draw an example of a bell-shaped distribution.
3. Draw an example of a distribution that is skewed right.

Objective 6, Page 1 (continued)

1. Draw an example of a distribution that is skewed left.

Objective 6, Page 2

**Example 6 *Identifying the Shape of a Distribution***

Figure 10 displays the histogram obtained in Example 4 for the five-year rate of return for large-blended mutual funds. Describe the shape of the distribution.