2.3 Organizing Quantitative Data

This section will focus on ways to organize quantitative data into tables, charts, and graphs.

Quantitative data is organized by dividing the observations into classes (bins or categories). Each class is treated as distinct values for the quantitative data.

We can group quantitative data into three different kinds of groups and the first type of group to be discussed is single value grouping.

In single value grouping each class has one distinct value. We can organize quantitative data into a frequency distribution like the qualitative data was organized.

Example 1

2.30 Household Size. The U.S. Census Bureau conducts nationwide surveys on characteristics of U.S. households and publishes the results in Current Population Reports. Following are data on the number of people per household for a sample of 40 households.

| 2 5 2 1 1 2 3 4 |
| 1 4 4 2 1 4 3 3 |
| 7 1 2 2 3 4 2 2 |
| 6 5 2 5 1 3 2 5 |
| 2 1 3 3 2 2 3 3 |

Construct a frequency distribution for single value grouping. Add in a last column for the relative frequency data.

There are two more methods for grouping quantitative data into classes. The next method to be discussed is called limit grouping.
Limit grouping uses class limits. Each class has a range of values.

**Terms used in Limit Grouping**
Lower class limit: The smallest value that can go into a class.
Upper class limit: The largest value that can go into a class.
Class width: The difference between the smallest value in a class and the smallest value in the next higher class.
Class mark: The average of the two class limits of a class.

**Example 2**

2.28 Super Bowl Ratings. Super Bowl television broadcasts rate among the highest in all television broadcasts. The Television Bureau of Advertising published a report titled *TV Basics*. Part of the report listed the television ratings for 32 Super Bowls, as shown in the following table.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Rating</th>
<th>Rating</th>
<th>Rating</th>
<th>Rating</th>
<th>Rating</th>
<th>Rating</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>49.1</td>
<td>44.4</td>
<td>42.7</td>
<td>42.3</td>
<td>46.4</td>
<td>45.8</td>
<td>44.5</td>
<td>41.9</td>
</tr>
<tr>
<td>40.4</td>
<td>41.3</td>
<td>45.5</td>
<td>47.2</td>
<td>46.3</td>
<td>43.3</td>
<td>41.6</td>
<td>40.2</td>
</tr>
<tr>
<td>43.3</td>
<td>45.1</td>
<td>48.6</td>
<td>44.4</td>
<td>42.4</td>
<td>40.7</td>
<td>40.3</td>
<td>44.2</td>
</tr>
<tr>
<td>46.0</td>
<td>48.3</td>
<td>43.5</td>
<td>47.1</td>
<td>41.9</td>
<td>40.4</td>
<td>41.4</td>
<td>46.4</td>
</tr>
</tbody>
</table>

Use a class width of 2 and construct a frequency distribution using limit grouping. Add in a last column for the relative frequency data.

The last method of grouping quantitative data into classes is called cutpoint grouping.

**Terms used in Cutpoint Grouping**
Lower class cutpoint: The smallest value that can go into a class.
Upper class cutpoint: The smallest value that can go into the next higher class.
Class width: The difference between the cutpoints in a class.
Class midpoint: The average of the two cutpoints in a class.
Construct a frequency distribution using cutpoint grouping. Add in a last column for the relative frequency data.

In general how do we chose the classes width? Decide on the (approximate) number of classes. Normally between 5 and 20 classes (pg 69 in the book gives guidelines on how to determine the number of classes)

1. Calculate an approximate class width as

\[ Class \ Width = \frac{Maximum \ Observation \ - \ Minimum \ Observation}{Number \ of \ Classes} \]

- If the answer is a decimal, then round up to the nearest whole number and that is your class width.
- If the answer is a whole number, then that is your class width.

3. Choose a number for the lower limit (or cutpoint) of the first class, noting that it must be less than or equal to the minimum observation.

4. Obtain the other lower class limits (or cutpoints) by successively adding the class width chosen in step 2.

5. Use the results obtained from Step 4 to specify all of the classes.
Histograms: Displays the classes of the quantitative data on the horizontal axis and the frequencies (relative frequencies or percentages) of those values are on the vertical axis. The frequency (relative frequency) of each distinct value is represented by a vertical bar whose height is equal to the frequency (relative frequency) of that class. The bars should be positioned so that they touch each other (unless there are not any values for the frequency in that class).

- For single-value grouping, we use the distinct values of the observations to label the bars and each value is centered under its bar.
- For limit grouping or cutpoint grouping, we use the lower class limits (or equivalently the lower cutpoint) to label the bars. (Note that some statistical programs use the class marks or class midpoints centered under the bars.)

Procedure to Construct a Histogram
1. Obtain a frequency (relative frequency, percent) distribution of the data.
2. Draw a horizontal axis on which to place the bars and a vertical axis to display the frequencies (relative frequencies, percents)
3. For each distinct value, construct a vertical bar whose height equals the frequency (relative frequency, percent) of that class.
4. Label the bars with the distinct values, the horizontal axis with the name of the variable, and the vertical axis with “Frequency” (“Relative Frequency,” “Percent”).

Example 4
Use the data and frequency distribution from the Household Size data in Example 1 to create a histogram for the single value data.

Example 5
Use the data and frequency distribution from Example 2 on Superbowl Ratings to create a histogram for the limit grouping data.
Example 6
Use the data and relative frequency distribution from Example 3 on the Cheetah speeds to create a histogram for the cutpoint grouping data.

Dotplots are a graphical display of quantitative data that are useful in comparing the relative positions of data in a data set or to compare two data sets to each other. (Normally used for small data sets)

Procedure to Construct a Dotplot
1. Draw a horizontal axis to display the possible values of the quantitative data.
2. Record each observation by placing a dot over the appropriate value on the horizontal axis.
3. Label the horizontal axis with the name of the variable.

Example 7
Go back to the Household size data in example 2 and construct a dotplot.

Example 8

2.60 Stressed-Out Bus Drivers. Frustrated passengers, congested streets, time schedules, and air and noise pollution are just some of the physical and social pressures that lead many urban bus drivers to retire prematurely with disabilities such as coronary heart disease and stomach disorders. An intervention program designed by the Stockholm Transit District was implemented to improve the work conditions of the city’s bus drivers. Improvements were evaluated by Evans et al. who collected physiological and psychological data for bus drivers who drove on the improved routes (intervention) and for drivers who were assigned the normal routes (control). Their findings were published in the article “Hassles on the Job: A Study of a Job Intervention With Urban Bus Drivers” (Journal of Organizational Behavior, Vol. 20, pp. 199–208). Following are data, based on the results of the study, for the heart rates, in beats per minute, of the intervention and control drivers.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>68 66</td>
<td>74 52</td>
</tr>
<tr>
<td>74 58</td>
<td>77 53</td>
</tr>
<tr>
<td>69 63</td>
<td>60 77</td>
</tr>
<tr>
<td>68 73</td>
<td>66 71</td>
</tr>
<tr>
<td>64 76</td>
<td>63 73</td>
</tr>
</tbody>
</table>

a. Obtain dotplots for each of the two data sets, using the same scales.

b. Use your result from part (a) to compare the two data sets.
Stem and Leaf Diagram
A stem and leaf diagram is a visual display of the “raw” data. In this type of diagram, each observation except for the rightmost digit is thought of as the stem. The right most digit is thought of as the leaf.

Procedure: Construct a Stem and Leaf Diagram
1. Think of each observation as a stem-consisting of all, but the right most digit and a leaf, the right most digit
2. Write the stems from smallest to largest in a vertical column to the left of a vertical rule.
3. Write each leaf to the right of the vertical rule in the row that contains the appropriate stem.
4. Arrange the leave in each row in ascending order

Example 9

2.65 Ages of Baseball Players. From MLB Roster Analysis on the ESPN Web site, we found the average age of the players on each of the 30 major league baseball teams, as of May 2, 2005, to be as follows.

<table>
<thead>
<tr>
<th>Ages</th>
<th>26.6</th>
<th>27.9</th>
<th>27.9</th>
<th>29.9</th>
<th>29.3</th>
<th>28.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28.4</td>
<td>28.9</td>
<td>27.7</td>
<td>28.7</td>
<td>30.5</td>
<td>29.8</td>
</tr>
<tr>
<td></td>
<td>28.5</td>
<td>27.9</td>
<td>30.9</td>
<td>29.3</td>
<td>28.8</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>29.1</td>
<td>31.0</td>
<td>30.7</td>
<td>30.3</td>
<td>29.7</td>
<td>31.0</td>
</tr>
<tr>
<td></td>
<td>29.4</td>
<td>29.8</td>
<td>29.4</td>
<td>32.7</td>
<td>34.0</td>
<td>31.8</td>
</tr>
</tbody>
</table>

Construct a stem-and-leaf diagram for these data using
a. one line per stem.  
b. two lines per stem.
c. Which stem-and-leaf diagram do you find more useful? Why?