Chapter 2 Organizing and Summarizing Data

Ch2.1 Organizing Qualitative Data

Objective A: Interpretation of a Basic Statistical Graph

Example 1:

Identity Theft  Identity fraud occurs someone else’s personal information is used to open credit card accounts, apply for a job, receive benefits, and so on. The following relative frequency bar graph represents the various types of identity theft based on a study conducted by the Federal of Trade Commission.

(a) Approximate what percentage of identity theft was loan fraud (such as applying for a loan in someone else’s name)?

(b) If there were 10 million cases of credit card fraud in 2008, how many were credit card fraud (someone uses someone else’s credit card to make a purchase)?

Objective B: Construct a Frequency / Relative Frequency Distribution, Bar Graph, Pareto Chart and Pie Chart

B1. Frequency / Relative Frequency Distribution

- A frequency distribution lists each category of data and the frequency which is the number of occurrences for each category data.

- A relative frequency distribution lists each category of data and the relative frequency which is the proportion of observation within a category.

Relative frequency = \( \frac{\text{frequency}}{\text{sum of all frequency}} \)
Example 1: In a national survey conducted by the Centers for Disease Control to determine health-risk behaviors among college students, college students were asked, “How often do you wear a seat belt when driving a car?” The frequencies were as follows:

<table>
<thead>
<tr>
<th>Response</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>I do not drive a car</td>
<td>249</td>
</tr>
<tr>
<td>Never</td>
<td>118</td>
</tr>
<tr>
<td>Rarely</td>
<td>249</td>
</tr>
<tr>
<td>Sometimes</td>
<td>345</td>
</tr>
<tr>
<td>Most of the time</td>
<td>716</td>
</tr>
<tr>
<td>Always</td>
<td>3093</td>
</tr>
</tbody>
</table>

(a) Construct a relative frequency distribution.

<table>
<thead>
<tr>
<th>Response</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>I do not drive a car</td>
<td>249</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>Rarely</td>
<td>249</td>
<td></td>
</tr>
<tr>
<td>Sometimes</td>
<td>345</td>
<td></td>
</tr>
<tr>
<td>Most of the time</td>
<td>716</td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>3093</td>
<td></td>
</tr>
</tbody>
</table>

(b) What percentage of respondents answered “Always”?

(c) What percentage of respondents answered “Never” or “Rarely”?

(d) Suppose that a representative from the Centers for Disease Control says, “2.5% of the college students in this survey responded that they never wear a seat belt.” Is this a descriptive or inferential statement?
B2. Construct a Bar Graph, a Pareto Chart, or a Pie Chart

- A bar graph is constructed by labeling each category of data on either the horizontal or vertical axis and the frequency or relative frequency of the category on the other axis. Rectangles of equal width are drawn for each category. The height of each rectangle represents the category’s frequency or relative frequency.
- A Pareto chart is a bar graph whose bars are drawn in decreasing order of frequency or relative frequency.
- A pie chart is a circle divided into sectors. Each sector represents a category of data. The area of each sector is proportional to the frequency of the category.

Example 2: A sample of 40 randomly selected registered voters in Sylmar was asked their Political affiliation: Democrat (D), Republican (R), or Independent (I). The results of the survey are as follows:

R D R D R R D D D D R D R D D I D D R R D D D I D
R D D I R D R D D R

(a) Construct a frequency distribution of the data.

(b) Construct a relative frequency distribution of the data.

(c) Construct a frequency bar graph.
(d) Construct a relative frequency bar graph.

(e) Construct a Pareto chart.

(f) Construct a pie chart.

(g) Use StatCrunch to construct a pie chart.

Follow the steps from the PowerPoint presentation.
Objective A : Histogram
- A histogram is constructed by drawing rectangles for each class of data. If the discrete data set is small, each number is a class. If the discrete data set is large or the data are continuous, the classes must be created using interval of numbers. 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.

Construct Frequency Distribution and Histogram for Discrete Data
Example 1: The following data represent the number of customers waiting for a table at 6:00 P.M. for 40 consecutive Saturdays at Bobak’s Restaurant:
11  5  11  3  6  8  6  7  4  5  13  9  6  4  14  11  13  10  9  6  8
10  9  5  10  8  7  3  8  8  7  8  7  9  10  4  8  6  11  8

(a) Are these data discrete or continuous? Explain.

(b) Construct a frequency distribution of the data.

(c) Construct a relative frequency distribution of the data.
(d) What percentage of the Saturdays had 10 or more customers waiting for a table at 6:00 p.m.?

(e) Construct a frequency histogram of the data.

Identify the shape of each distribution.
Objective B: Constructing a Stem-and-Leaf Plot

The stem of a data value will consist of the digits to the left of the rightmost digit.
The leaf of a data value will be the rightmost digit.

Example 1: The following data represent the number of miles per gallon achieved on the highway for small cars for the model year 2008.

27  31  28  30  52  25  33  33  29  23  27  37  30  45  24  32
34  35  31  44  42  26  43  35  36  36  54  33  32  35  34  37

(a) Construct a stem-and-leaf plot.

(b) Describe the shape of the distribution.

Objective C: Construct Frequency Distributions and Histogram for Continuous Data

- Classes are categories into which data are grouped.
- The lowest class limit is the smallest value within a class.
- The upper class limit is the largest value within a class.
- The class width is the difference between consecutive lower class limits.

- The class width is computed by the following formula.

\[
\text{Class width} \approx \frac{\text{largest data value} - \text{smallest data value}}{\text{number of classes}}
\]

------> Round this value up to the same decimal place as the raw data.
Example 1: The following data represent the fall 2006 student headcount enrollments for all public community colleges in the state of Illinois.

<table>
<thead>
<tr>
<th>Number of Students Enrolled</th>
<th>Number of Community Colleges*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–4,999</td>
<td>15</td>
</tr>
<tr>
<td>5,000–9,999</td>
<td>16</td>
</tr>
<tr>
<td>10,000–14,999</td>
<td>9</td>
</tr>
<tr>
<td>15,000–19,999</td>
<td>4</td>
</tr>
<tr>
<td>20,000–24,999</td>
<td>0</td>
</tr>
<tr>
<td>25,000–29,999</td>
<td>1</td>
</tr>
</tbody>
</table>

*Treats the City Colleges of Chicago as seven distinct institutions, but all other multicampus districts as single institutions.

(a) Find the number of classes.

(b) Find the class limits.

(c) Find the class width.

Example 2: **Uninsured Rates** The following data represent the percentage of people without health insurance for the 50 states and the District of Columbia in 2009. (Ch 2.2 Q36 p. 94)

<table>
<thead>
<tr>
<th>4.2</th>
<th>10.6</th>
<th>12.6</th>
<th>14.8</th>
<th>17.8</th>
<th>18.9</th>
<th>21.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.6</td>
<td>10.6</td>
<td>13.0</td>
<td>15.5</td>
<td>18.1</td>
<td>19.4</td>
<td>22.2</td>
</tr>
<tr>
<td>9.2</td>
<td>10.9</td>
<td>13.3</td>
<td>15.9</td>
<td>18.3</td>
<td>19.6</td>
<td>25.0</td>
</tr>
<tr>
<td>9.6</td>
<td>10.9</td>
<td>13.4</td>
<td>15.9</td>
<td>18.3</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>9.6</td>
<td>11.3</td>
<td>13.9</td>
<td>16.1</td>
<td>18.4</td>
<td>20.6</td>
<td></td>
</tr>
<tr>
<td>9.7</td>
<td>11.4</td>
<td>14.0</td>
<td>16.1</td>
<td>18.4</td>
<td>21.1</td>
<td></td>
</tr>
<tr>
<td>10.2</td>
<td>11.6</td>
<td>14.3</td>
<td>16.1</td>
<td>18.6</td>
<td>21.2</td>
<td></td>
</tr>
<tr>
<td>10.5</td>
<td>12.3</td>
<td>14.7</td>
<td>16.2</td>
<td>18.7</td>
<td>21.3</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Gallup*
With the first class having a lower class limit of 4 and a class width of 2:

(a) Construct a frequency distribution.

Put the data in ascending order first:

4.2  8.6  9.2  9.6  9.6  9.7  10.2  10.5  10.6  10.6  10.9  10.9  11.3  11.4  11.6  
12.3  12.6  13.0  13.3  13.4  13.9  14.0  14.3  14.7  14.8  15.5  15.9  15.9  16.1  16.1  
16.1  16.2  17.8  18.1  18.3  18.3  18.4  18.4  18.6  18.7  18.9  19.4  19.6  19.7  20.6  
21.1  21.2  21.3  21.4  22.2  25.0

(b) Construct a relative frequency distribution.
(c) Construct a frequency histogram of the data.

(d) Construct a relative frequency histogram of the data.

(e) Describe the shape of the distribution.

(f) Use StatCrunch to repeat parts (a) to (e) with the first class having a lower class of 4 and a class width of 4.
   Note: For qualitative data, we can use StatCrunch to construct a frequency histogram first, then from the histogram we can obtain the frequency distribution.
Follow the PowerPoint presentation for the part (f).

Use StatCrunch to construct a frequency histogram.

From the frequency histogram obtained above, create the frequency distribution.

Example 3: The largest value of a data set is 125 and the smallest value of the data set is 27. If six classes are to be formed, calculate an appropriate class width.
Objective D : Time Series Graphs
- A time series graph represents the values of a variable that have been collected over a specified period of time. The horizontal axis is the time and the vertical axis is the value of the variable. Line segments are drawn by connective consecutive points of time and corresponding value of the variable.

Example 1: The following time-series graph shows the annual U.S. motor vehicle production from 1990 through 2008.

![Annual U.S. Motor Vehicle Production Graph](image)

*Source: Bureau of Transportation Statistics*

(a) Estimate the number of motor vehicles produced in the United States in 1991.

(b) Estimate the number of motor vehicles produced in the United States in 1999.
(c) Use the results from (a) and (b) to estimate the percent increase in the number of motor vehicles produced from 1991 to 1999.

(d) Estimate the percent decrease in the number of motor vehicles produced from 1999 to 2008.

Ch 2.4 Graphical Misrepresentations of Data
- The most common graphical misinterpretation of data is accomplished through manipulation of the scale of the graph.

Example 1:
Union Membership  The following relative frequency histogram represents the proportion of employed people aged 25 to 64 years old who were members of a union.

Source: U.S. Bureau of Labor Statistics

(a) Describe how this graph is misleading. What might a reader conclude from the graph?
(b) Redraw the histogram so that it is not misleading

Example 2:

**Inauguration Cost**  The following is a *USA Today*-type graph. Explain how it is misleading.