2.1 Organizing Qualitative Data

Definitions:

**Frequency Distribution** – lists the *number* of occurrences of each category of data.

**Relative Frequency Distribution** – lists the *proportion* of occurrences of each category of data. All relative frequencies sum to one.

*Example*: A survey of a random sample of 25 men revealed the following eye color:

<table>
<thead>
<tr>
<th>Eye Color</th>
<th>Tally</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td></td>
<td>13</td>
<td>0.52</td>
</tr>
<tr>
<td>Blue</td>
<td></td>
<td>8</td>
<td>0.32</td>
</tr>
<tr>
<td>Hazel</td>
<td></td>
<td>3</td>
<td>0.12</td>
</tr>
<tr>
<td>Green</td>
<td></td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>25</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Bar Graph** – a graph of the frequency or relative frequency distribution of qualitative data, consisting of vertical or horizontal rectangles, called bars, for each class of data.

*Example*: Construct a bar graph for the eye-color data above.

*Solution*: We construct a vertical bar graph in descending order of frequency, called a *Pareto chart*. 
Side-by-side bar graphs are used when two or more data sets are graphed. Horizontal bar graphs make it easier to read the category names.

**Example:** A random sample of 25 women revealed the distribution of eye color shown below. Construct a bar graph of this distribution alongside the one for men given above.

<table>
<thead>
<tr>
<th>Eye Color</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown</td>
<td>14</td>
</tr>
<tr>
<td>Blue</td>
<td>7</td>
</tr>
<tr>
<td>Hazel</td>
<td>2</td>
</tr>
<tr>
<td>Green</td>
<td>2</td>
</tr>
<tr>
<td>Sum</td>
<td>25</td>
</tr>
</tbody>
</table>

**Solution:** To display these two data sets, we will construct a side-by-side, horizontal bar graph.
**Pie Chart** – a graph of the relative frequencies of qualitative data, consisting of a circle divided into sectors for each category of data. The sectors sum to 100% or 360°.

*Example:* Construct a pie chart of the eye-color data for men shown above.

*Solution:* We use the relative frequency to express each section as a percentage, and color coordinate the areas with the eye colors.

![Pie Chart Example](image)

*Exercises:*

2. Interpret a pie chart (5).
3. Interpret a bar graph (6 – 12).
4. Construct a bar graph or pie chart (13 – 26).
2.2 Organizing Quantitative Data

Definitions:

**Classes** – categories by which data are grouped. Data sets with discrete values may have one class per value. Continuous data have classes that range from the *lower class limit* to the *upper class limit*. The class width is the difference between consecutive lower class limits. The first and last classes may be open ended with only one class limit.

*Example:* The following data represent the volume of Atria Group stock in millions traded in a random sample of 35 days in 2007. (a) If six classes are to be formed, choose an appropriate lower class limit for the first class and a class width. (b) Construct frequency and relative frequency distributions.

\[
\begin{align*}
6.42 & \quad 23.59 & \quad 18.91 & \quad 7.85 & \quad 7.76 \\
8.51 & \quad 9.05 & \quad 14.83 & \quad 14.43 & \quad 8.55 \\
6.37 & \quad 10.30 & \quad 10.16 & \quad 10.90 & \quad 11.20 \\
13.57 & \quad 9.13 & \quad 7.83 & \quad 15.32 & \quad 14.05 \\
7.84 & \quad 7.88 & \quad 17.10 & \quad 16.58 & \quad 7.68 \\
7.69 & \quad 10.22 & \quad 10.49 & \quad 8.41 & \quad 7.85 \\
10.94 & \quad 20.15 & \quad 8.97 & \quad 15.39 & \quad 8.32
\end{align*}
\]

*Solution:* Since \( x_{\text{min}} = 6.37 \) and \( x_{\text{max}} = 23.59 \), the class width is

\[
\text{class width} = \frac{23.59 - 6.37}{6} = 2.87 \approx 3
\]

We choose a lower class limit of the first class of 6.00 and a class width of 3.00.

<table>
<thead>
<tr>
<th>Volume</th>
<th>Tally</th>
<th>Freq.</th>
<th>Rel. Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.00-8.99</td>
<td>II</td>
<td>15</td>
<td>0.4286</td>
</tr>
<tr>
<td>9.00-11.99</td>
<td>III</td>
<td>9</td>
<td>0.2571</td>
</tr>
<tr>
<td>12.00-14.99</td>
<td>IIII</td>
<td>4</td>
<td>0.1143</td>
</tr>
<tr>
<td>15.00-17.99</td>
<td>IIII</td>
<td>4</td>
<td>0.1143</td>
</tr>
<tr>
<td>18.00-20.99</td>
<td>II</td>
<td>2</td>
<td>0.0571</td>
</tr>
<tr>
<td>21.00-23.99</td>
<td>I</td>
<td>1</td>
<td>0.0286</td>
</tr>
</tbody>
</table>
**Histogram** – a graph for quantitative data similar to the bar graph for qualitative data, where non-overlapping classes with equal class widths (except possibly the first and last classes) are used to plot frequencies or relative frequencies with bars that touch (unless one class has no frequency).

*Example:* Construct a frequency histogram of the data for the Atria Group stock price.

*Solution:*

![Histogram Example](image)

**Stem-and-Leaf Plot** – a graph used to plot raw data, where the data is grouped into classes that are the stems and the last digits are the leaves. It is best for small data sets and a small range (the difference between the maximum and minimum values). If there are too many leaves on the stem, we use *split stems*. [See Figs. 13 & 14] If two sets of data are being compared, we use back-to-back stem-and-leaf plots. [Work problem 2.2.48]

**Dot Plot** – a graph similar to a bar graph, except dots are used to show the frequency. [See Fig. 15]

**Time-Series Plot** – a line graph of the value of a variable (such as a company’s stock price) versus time. [See Fig. 18]

**Bell-Shaped Distribution** – a symmetrical distribution, also called a normal distribution, that is shaped like a bell. [See Fig. 16]

**Uniform Distribution** – a symmetrical distribution where all classes have the same frequency.

**Skewed-Right Distribution** – a nonsymmetrical distribution where the tail to the right extends out further than the left.

**Skewed-Left Distribution** – a nonsymmetrical distribution where the tail to the left extends out further than the right.
Exercises:

2. Read a histogram (9 – 14).
3. Read a time-series graph (15 – 18).
4. Read a frequency distribution (19 – 20).
8. Construct a dot plot (53 – 54).
2.3 Graphical Misrepresentations of Data

Definitions:

**Deceptive Graph** – a graph that intentionally creates a false impression of the data. These are sometimes used by marketing firms, public policy advocates, and political groups.

**Misleading Graph** – a graph that unintentionally creates a false impression of the data. These are sometimes used by newspapers or other media outlet.

A graph can be deceptive or misleading by

1. using a bar with different intervals, a histogram that skips regions, or a pictorial graph that is not proportional.

2. using an axis scale that is not labeled.

3. using an axis scale that does not start with zero or does not show a symbol indicating the axis has been truncated or broken.

4. using a graph that is too cluttered.

5. using a graph that stands out more than the data.

Exercises:

1. Describe what can make a graph misleading or deceptive (1 – 11, 18 - 20).

2. Construct misleading and non-misleading graphs (12 – 17).