4.4 Displaying Distributions—Quantitative Variables

4.4.1 Frequency Plots of quantitative variables

Basic steps for a frequency plot:

- Draw a real line.
- On the real-line axis, mark the minimum and the maximum values of the data.
- Fill in the scale for the numbers on the line in between the minimum and maximum in equal-spaced increments. The scale should cover the whole range of values in the data.
- Mark each data value with an X above the appropriate location on the scale. If there are two or more items with the same value, stack them vertically.

Note:
1. Involve two axes (1) vertical denotes frequency (2) horizontal denotes scale for values.
2. X and • plots: Use if you have a few (n<20) values. Plot x (or •) over each value observed.
3. Terms used to describe values in terms of the frequency plot.

**Outliers:** 1 or two observations which are far removed from the rest of the data.

**Cluster:** group of observations separated from the rest of the data.

**Gap:** a large distance between observations.

Example: The prices (in cents) of a single-scoop ice cream cone at 17 Los Angeles stores:
25, 53, 70, 75, 90, 90, 91, 95, 95, 95, 96, 100, 105, 110, 115, 120

We notice that the values of 25 and 53 are far removed from the rest of the data. They are called outliers. These values stand out due to the gaps between 25 and 52 and between 53 and 70. Starting from 90, we have a cluster.
4.4.2. Shapes of Distributions

Symmetric, bell-shaped
Unimodal
Math Exam Score

Bimodal
Heights of men and
women in a school

Skewed to the right
(skewed positively)
Income

Skewed to the left
(skewed negatively)
Gas price

Uniform
0: number randomly
p: number from a range
Briefly, the terms used to describe the shape of a distribution are:

- **Symmetric** The distribution can be divided into two parts around a central value and each part is the reflection of the other.
- **Unimodal** The distribution has a single peak which shows the most common value(s) in the data.
- **Bimodal** The distribution has two peaks. This often results when two populations are being sampled.
- **Uniform** The possible values appear with equal frequency.
- **Skewed** One side of the distribution is stretched out longer than the other side. The direction of the skewness is the direction of the longer side.
4.4.3. Stem and Leaf Plot

Basic steps for a stem-and-leaf plot:

- Separate each measurement into a stem and a leaf -- generally the leaf consists of exactly one digit (the last one) and the stem consists of 1 or more digits. Sometimes the decimal is left out of the stem but a note is added on how to read each value. For the 2.345 example we would state that $\underline{234} \mid 5$ should be read as 2.345.

- Sometimes, when the observed values have many digits, it may be helpful either to round the numbers (round 2.345 to 2.35, with stem=2.3, leaf=5) or truncate (or dropping) digits (truncate 2.345 to 2.34).

- Write out the stems in order increasing vertically (from top to bottom) and draw a line to the right of the stems.

- Attach each leaf to the appropriate stem.

- Arrange the leaves in increasing order (from left to right).
Example 4.7 Basic Stem-and-Leaf Plot for Age

Consider the ages of the 20 subjects from DATA SET 1 (page 212).
45, 41, 51, 46, 47, 42, 43, 50, 39, 32, 41, 44, 47,
49, 45, 42, 41, 40, 45, 37
Construct the basic stem-and-leaf plot of this data set. Use 3|2 for 32 years old.

Note:
- The stem should be equally spaced.
- If there are too few stems and too many leaves on each stem, one suggestion is to use split stems.
- If there are too many stems and too few leaves on each stem, one suggestion is to use 2-digit leaves and round or truncate the original values.
4.4.4. Histograms

Basic Steps for Creating a Histogram:

- Find the minimum value, maximum value, and overall range of the data.
- Divide the range of the data (smallest to largest) into classes of equal width. The classes should cover the whole range of values, but they should not overlap.
- Count the number of observations that fall into each class. Recall that the counts are also called frequencies.
- Draw a horizontal axis and mark off the classes along this axis.
- The vertical axis can be the count, the proportion, or the percentage.
- Draw a rectangle (a vertical bar) above each class with the height equal to the count, the proportion, or the percentage.
### Example: Real Data

$n = 58$

<table>
<thead>
<tr>
<th>Class</th>
<th>Observations (Count)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>[100,110]</td>
<td>5</td>
<td>8.62%</td>
</tr>
<tr>
<td>(90,100</td>
<td>8</td>
<td>13.79%</td>
</tr>
<tr>
<td>(80,90)</td>
<td>12</td>
<td>20.69%</td>
</tr>
<tr>
<td>(70,80)</td>
<td>11</td>
<td>18.97%</td>
</tr>
<tr>
<td>(60,70)</td>
<td>9</td>
<td>15.52%</td>
</tr>
<tr>
<td>(50,60)</td>
<td>8</td>
<td>13.79%</td>
</tr>
<tr>
<td>(40,50)</td>
<td>3</td>
<td>5.17%</td>
</tr>
<tr>
<td>(30,40)</td>
<td>1</td>
<td>1.72%</td>
</tr>
<tr>
<td>(20,30)</td>
<td>1</td>
<td>1.72%</td>
</tr>
</tbody>
</table>

![Graph of class distribution](image1.png)

![Graph of percentages](image2.png)
- Time plots

- Scatter plots.
  - Positively associated
  - No linear association