I. Review

1. Stem-and-Leaf Plots:
   
   (a) advantage: it can be constructed quickly; we can extract all the data values from plot;
   
   (b) disadvantage: not useful for large data sets; the choice of stem values may affects the distribution pattern of data.

2. Histograms

   (a) advantage: useful for large data sets;

   (b) disadvantage: the choice of class boundaries can affect the appearance of the histogram.

3. Dot Plots:

   (a) advantage: it can be constructed quickly.

   (b) disadvantage: when the number of data is small, it is difficult to identify any pattern of variation.

4. Boxplots: (constructed by: min max median 1stQ 3rdQ—five number summary)
   They are particularly effective for graphically portraying comparisons among sets of data; they have a high visual impact.
5. Measures of Location:

(a) sample mean=$\bar{x}=\frac{\sum_{i=1}^{n} x_i}{n}$
(Sensitive to outlying values.)

(b) sample median: (for ordered data)
when sample size is odd, median=the value for the middle observation;
when sample size is even, median= the average of the middle two.
(Robust to outlying values.)

(c) Finding the $p$th sample quantile(also called the $100p$th percentile) $x_{[p]}$:

i. Put the data in order, from smallest to largest.

ii. Compute $np$, where $n$ is the sample size.

iii. If $np$ is an integer, then $x_{[p]}$ is the average of the $(np)^{th}$ and the $(np+1)^{th}$ numbers in the list.

iv. If $np$ is not an integer, then round up, and use the observation which occurs at that place in the list.

(d) 1st quartile= the 0.25 quantile.

(e) 3rd quartile= the 0.75 quantile.

6. Measures of Spread

(a) range=maximum-minimum

(b) interquartile range(IQR)=3rd quartile-1st quartile

(c) variance=$S^2=\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2=\frac{1}{n-1} [\sum_{i=1}^{n} x_i^2 - n\bar{x}^2].$

(d) standard deviation=$S=\sqrt{S^2}$

(e) coefficient of variation=$cv=\frac{S}{\bar{x}}$
II. Practice Problems

1. Consider the following two sets of data:

\[
\begin{align*}
    x: & \quad 4 \quad 5 \quad 7.8 \quad 6.4 \quad 1.5 \\
    y: & \quad 33 \quad 7.9 \quad 18.7 \quad 6 \quad 55.9
\end{align*}
\]

(a) Evaluate the following:

i. \( \sum_{i=1}^{5} x_i \)

ii. \( \sum_{i=1}^{5} x_i^2 \)

iii. \( (\sum_{i=1}^{5} x_i)^2 \)

iv. \( \sum_{i=1}^{5} y_i \)

v. \( \sum_{i=1}^{5} x_i y_i \)

vi. \( (\sum_{i=1}^{5} x_i)(\sum_{i=1}^{5} y_i) \)

vii. \( \sum_{i=1}^{5} ax_i \) with \( a = 2 \)

viii. \( a \sum_{i=1}^{5} x_i \) with \( a = 2 \)

ix. \( \sum_{i=1}^{5} a \) with \( a = 4 \)

(b) Before making any further calculations, which sample, \( x \) or \( y \), do you think has the larger mean? Calculate \( \bar{x} \) and \( \bar{y} \) and compare.

(c) Before making any further calculations, which sample, \( x \) or \( y \), do you think has the larger variance? Calculate \( s^2 \) for each sample and compare.

(d) Verify numerically that, except for rounding error, the \( n = 5 \) values satisfy the following:

i. \( \sum_{i=1}^{5} (x_i - \bar{x}) = 0 \)

ii. \( \sum_{i=1}^{5} (x_i - \bar{x})^2 = \sum_{i=1}^{5} x_i^2 - n(\bar{x})^2 = \sum_{i=1}^{5} x_i^2 - \frac{(\sum_{i=1}^{5} x_i)^2}{n} \)
2. A company bottles milk in several sizes of container. A random sample of 17 containers is obtained from the “small” container size. The volume of milk (in ounces) is measured for each container. The volumes are:
5.99 5.84 5.95 6.09 5.93 5.88 5.92 6.04 6.00
5.89 5.95 5.97 5.90 5.91 6.03 5.89 5.98

(a) Make a stem and leaf display.

(b) Find the mean, standard deviation, median, 1st quartile, 3rd quartile, range, IQR and 20th percentile of the data.

(c) Construct a box plot for these data.
III Solutions of the Practice problems

1. (a) Evaluate the following:
   
i. \( \sum_{i=1}^{5} x_i = 24.7 \)
   
ii. \( \sum_{i=1}^{5} x_i^2 = 16.00 + 25.00 + 60.84 + 40.96 + 2.25 = 145.05 \)
   
iii. \( (\sum_{i=1}^{5} x_i)^2 = 24.7^2 = 610.09 \)
   
iv. \( \sum_{i=1}^{5} y_i = 121.5 \)
   
v. \( \sum_{i=1}^{5} x_i y_i = 132.00 + 39.50 + 145.86 + 38.40 + 83.85 = 439.61 \)
   
vi. \( (\sum_{i=1}^{5} x_i)(\sum_{i=1}^{5} y_i) = 24.7(121.5) = 3001.05 \)
   
vii. \( \sum_{i=1}^{5} ax_i \) with \( a = 2 \): \( 8.0 + 10.0 + 15.6 + 12.8 + 3.0 = 49.4 \)
   
viii. \( a \sum_{i=1}^{5} x_i \) with \( a = 2 \): \( 2(24.7) = 49.4 \)
   
ix. \( \sum_{i=1}^{5} a \) with \( a = 4 \): \( 4 + 4 + 4 + 4 + 4 = 20 \)

(b) \( \bar{x} = \frac{24.7}{5} = 4.94 \), \( \bar{y} = \frac{121.5}{5} = 24.3 \).

(c) \( s_x^2 = 5.758 \), \( s_y^2 = 427.365 \).

(d) i. \( \sum_{i=1}^{5} (x_i - \bar{x}) = (-0.94) + 0.06 + 2.86 + 1.46 + (-3.44) = 0 \)
   
ii. \( \sum_{i=1}^{5} (x_i - \bar{x})^2 = 0.8836 + 0.0036 + 8.1796 + 2.1316 + 11.8336 = 23.032 \)

\[ \sum_{i=1}^{5} x_i^2 - n(\bar{x})^2 = 145.05 - 5(4.94)^2 = 23.032 \]

\[ \sum_{i=1}^{5} x_i^2 - \frac{(\sum_{i=1}^{5} x_i)^2}{n} = 145.05 - \frac{24.7^2}{5} = 23.032 \]

2. (a)

<table>
<thead>
<tr>
<th>x</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.8</td>
<td>4</td>
</tr>
<tr>
<td>5.8</td>
<td>899</td>
</tr>
<tr>
<td>5.9</td>
<td>0123</td>
</tr>
<tr>
<td>5.9</td>
<td>55789</td>
</tr>
<tr>
<td>6.0</td>
<td>034</td>
</tr>
<tr>
<td>6.0</td>
<td>9</td>
</tr>
</tbody>
</table>

(b) mean=5.9506, standard deviation=0.0657, median=5.95, range=0.25, 
- Q1=x[5]=5.90, Q3=x[13]=5.99, IQR=0.09, 20th percentile \( x_{[4]} = 5.89 \).