1. Use the data in exercise 3.72 on page 111 in the textbook. Obtain the
   a. range of the data.
   b. sample standard deviation of the data from the defining formula.
   c. sample standard deviation of the data from the computing formula

2. Do Exercise 3.93 on page 113 in the textbook.

   Hint:
   \[
   \begin{array}{ccc}
   & \sum x_i & \sum x_i^2 \\
   Data set I & 189 & 3157 \\
   Data set II & 142 & 2052 \\
   \end{array}
   \]

3. Consider the following sample of exam scores, arranged in increasing order.

   28  57  58  64  69  74  79  80  83  85  85  87  87  89  89  90  92  93  94  94  95  96  96  97  97  97  98  100  100

   (Hint: the sample mean and sample standard deviation of these exam scores are, respectively, 85 and 16.1)

   a. Determine the interquartile range (IQR).
   b. Obtain the five number summary.
   c. Identify potential outliers if any.
   d. Construct and interpret a boxplot or if appropriate, a modified boxplot.
Exercises for Recitations (09/28-10/03)

1. Consider the following four data sets:

<table>
<thead>
<tr>
<th></th>
<th>Data Set I</th>
<th>Data Set II</th>
<th>Data Set III</th>
<th>Data Set IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>9</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

\[ \sum x_i = 50 \quad 50 \quad 50 \quad 50 \]
\[ \sum x_i^2 = 350 \quad 410 \quad 250 \quad 316 \]

\[ s = \begin{array}{c} 3.33 \\ 2.71 \end{array} \]

a. Compute the mean of each data set.
b. Although the four data sets have the same means, they are quite different in another respect. How are they different?
c. In looking at the data, which data set appears to have the least variation? The greatest variation?
d. Compute the range of each data set.
e. Compute the sample SD of each data set.
f. From your answers to parts (d) and (e), which measure of variation better distinguishes the spread in the four data sets - the range or the SD?
g. Are your answers in parts (c) and (e) consistent?

2. The U.S. National Center for Health Statistics compiles data on the length of stay by patients in short-term hospitals and publishes its findings in *Vital and Health Statistics*. A random sample of 21 patients, arranged in ascending order, yielded the following data on length of stay, in days.

1 1 3 3 4 4 5 6 6 7 7 9 9 10 12 12 13 15 18 23 55

a. Obtain the five number summary.
b. Determine the IQR.
c. Identify potential outliers, if any.
d. Construct and interpret a boxplot and, if appropriate a modified boxplot.
### Descriptive Statistics: Length Stay

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length S</td>
<td>21</td>
<td>10.62</td>
<td>7.00</td>
<td>8.79</td>
<td>11.62</td>
<td>2.54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length S</td>
<td>1.00</td>
<td>55.00</td>
<td>4.00</td>
<td>12.50</td>
</tr>
</tbody>
</table>