1.1
a. The population of interest is the weight of the shrimp maintained on the specific diet for a period of 6 months.
b. The sample is the 100 shrimp selected from the pond and maintained on the specific diet for a period of 6 months.
c. The weight gain of the shrimp over 6 months.
d. Since the sample is only a small proportion of the whole population, it is necessary to evaluate what the mean weight may be for any other randomly selected 100 shrimps.

1.5
a. The population of interest is the population of those who would vote in the 2004 senatorial campaign.
b. The population from which the sample was selected is registered voters in this state.
c. The sample will adequately represent the population, unless there is a difference between registered voters in the state and those who would vote in the 2004 senatorial campaign.
d. The results from a second random sample of 5,000 registered voters will not be exactly the same as the results from the initial sample. Results vary from sample to sample. With either sample we hope that the results will be close to that of the views of the population of interest.

2.7
The list of registered voters could be used as the sampling frame for selecting the persons to be included in the sample.

2.11
a. Simple random sampling.
b. Stratified sampling.
c. Cluster sampling.

Lab One Problems:

2.
Empirical distribution of gender:

<table>
<thead>
<tr>
<th>gender</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>37.50</td>
</tr>
<tr>
<td>M</td>
<td>62.50</td>
</tr>
</tbody>
</table>

3.
The distribution is quite symmetric (and without obvious outliers), supported by the fact that the mean and median are quite close. Therefore, mean is the better choice of center here. (You may think differently if you draw histograms.)

4.

**Tabulated Statistics: gender, college**

<table>
<thead>
<tr>
<th></th>
<th>Bsns</th>
<th>Eng</th>
<th>Sci</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>33.33</td>
<td>33.33</td>
<td>33.33</td>
<td>100.00</td>
</tr>
<tr>
<td>M</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>40.00</td>
<td>20.00</td>
<td>40.00</td>
<td>100.00</td>
</tr>
<tr>
<td>All</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>37.50</td>
<td>25.00</td>
<td>37.50</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Cell Contents: Count

You can list only counts or add % of column

5.
Here I draw two dotplots into a graph. You can draw side-by-side boxplots. However, since the sample size is very small, dotplots or stemplots are preferred. You can also do back-to-back stemplots in R by installing the package Aplpack: [http://www.inside-r.org/packages/aplpack](http://www.inside-r.org/packages/aplpack). If you are not a R user, I would suggest that you draw the back-to-back stemplots yourself in a word processor or alike. Minitab does not seem to draw such plots.

**Problems on page 26**

Use the two dataset presented on page 125 of the text (in Exercise 3.30) to do the following:

(a) Make a histogram by hand. Explain your choice of intervals. Why might Minitab make a histogram that looks different?

**Answer:**

The range is 13-2=11 and there are 70 observations in total. Since square root of 70 is 8.4, I choose 7 or 8 intervals. After comparing them two, I choose 7 intervals. The width of interval should be at least 11/7 =1.6, rounding up to 2. Therefore, the selected intervals are [2,4), [4,6), [6,8), [8,10), [10,12), [12,14), [14,16).
[6,8), [8,10), [10,12), [12,14), [14,16). Because of the rounding, the last interval is not needed. Here is the resultant histogram:

It might look different to the one drawn by Minitab as it might choose a different set of intervals.

(b) Find the sample mean. On the horizontal axis (bottom scale) of your histogram, mark the position of the sample mean. Does the histogram seem to "balance" there?

Answer:

The sample mean is 7.73. From the histogram you can see 7.73 is pretty much the center. The frequency on its left is about 45, which is much bigger than that on its right (~25), so this histogram is not balanced.

(c) Find the interval that extends two sample standard deviations on either side of the sample mean. Sketch this interval on the bottom scale of your histogram. What proportion of all observations actually falls in this interval? (You may find the sample mean and standard deviation either with software or with a calculator.)

Answer:

The standard deviation is 1.99. The mean +/- 2 standard deviation is (3.75, 11.71). The proportion is 67/70=96%, which matches with the Empirical rule but the histogram shows non-normal.