STAT2331-801 Spring 2011
Practical Exam 1

Disclaimer: These questions are intended as a guide to the kind of questions and topics you may encounter on the real exam. The real exam may contain questions on topics that are not presented in this practice exam. In addition it may contain questions on the same topics, but exploring different aspects. If you do well in this practice exam, then it is likely that you will do well on the real one, but there are no guarantees. Notice that there are two types of problems on the test, short answer and multiple choice. For the multiple choice problems, only one answer is correct, although parts of some of the wrong answers may be correct. The actual exam 1 will probably contain 2 short answer problems and 9 multiple-choice problems. In the short answer problems, please keep at least 4 decimal places in the calculation and put your answers in at least 2 decimal places.

Topics include:
- Chapter 1: Displaying data (stemplots etc. including ideas of symmetry, skewness, outliers, modes)
- Chapter 2: Measures of center (mean, median), ideas of robustness vs. efficiency
- Measures of variation (IQR, s)
- Chapter 3: Density curves including the normal curve, the 68-95-99.7 rule, the z-score and use of Table A.
- Chapter 4: Scatter plots and correlation

Multiple Choice

1. Suppose the correlation between length in inches and weight in pounds for raccoons is 0.82. The correlation between length in centimeters and weight in kilograms would be exactly the same because
   a. both length and weight are converted to metric units
   b. correlation is a resistant measure
   c. correlation is unit-free
   d. No! Of course the correlation would change!
   e. None of the above

2. In general, which of the following statements is FALSE?
   a. The sample mean is more sensitive to extreme values than the median.
   b. The sample standard deviation is more sensitive to extreme values than the interquartile range (IQR).
   c. The sample standard deviation is a measure of spread around the sample mean.
   d. The sample standard deviation is a measure of central tendency around the median.
   e. If a distribution is symmetric, then the mean will be equal to the median.
3. In the below boxplot, the first quartile of the heights of the Oak Trees is closest to which height?
   a. 67 feet
   b. 60 feet
   c. 53 feet
   d. 48 feet
   e. 40 feet

4. Boxplots are most often used to ____________. The central box spans from the _______ up to the ____________. A line across the box indicates the location of the ___________.

   Fill in the blanks above in order:

   a. display single distributions, minimum, maximum, mean
   b. display single distributions, lower quartile, upper quartile, median
   c. compare distributions, lower quartile, upper quartile, mean
   d. compare distributions, lower quartile, upper quartile, median
   e. compare distributions, minimum, maximum, median

5. If the correlation between body weight and annual income were high and positive, we could conclude from this fact alone that:
   a. high incomes cause people to eat more food.
   b. low incomes cause people to eat less food.
   c. high income people tend to spend a greater proportion of their income on food than low income people, on average.
   d. high income people tend to be heavier than low income people, on average.
   e. high incomes cause people to gain weight.
6. A histogram of the heights of 39 plants is as follows:

The Q₃ of the height distribution is approximately:

a. 7.7  
b. 7.5  
c. 9.4  
d. 10.0  
e. 11.0

7. From tax records, it is relatively easy to determine the average amount of liquor consumed per person (X) and the average number of cigarettes consumed per person (Y) for each of the 50 states of the U.S. These are plotted on a scatter plot and a high positive correlation is found. Which of the following is a reasonable conclusion from this information?

a. This implies that heavy smoking causes people to drink more.  
b. This implies that heavy drinking causes people to smoke more.  
c. States which have high tax rates on liquor will also have high tax rates on tobacco.  
d. This could be an example of a correlation caused by a hidden variable because both activities could be highly correlated with average family income and average income might vary widely among the states.  
e. The number of cigarettes a person smokes is positively correlated with the amount of liquor they consume.

8. Which of the following is NOT a correct statement about Normal Distributions?

a. All of the normal distributions can be transformed into a standard normal distribution  
b. The normal distribution is a bimodal distribution  
c. The normal distribution is a symmetric distribution  
d. The central 68% of the distribution is 2 standard deviations wide  
e. None of the above
Short Answer / Calculation

1. The height (in feet) and volume of usable lumber (in cubic feet) of 6 cherry trees are measured by a researcher.

<table>
<thead>
<tr>
<th>X: Height (in feet)</th>
<th>61</th>
<th>65</th>
<th>75</th>
<th>80</th>
<th>85</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y: Volume of usable lumber (in cubic feet)</td>
<td>7</td>
<td>15</td>
<td>27</td>
<td>38</td>
<td>38</td>
<td>55</td>
</tr>
</tbody>
</table>

(a) Use scatter plot to decide whether it would be appropriate to use correlation coefficient to describe the relation between the height and the volume of usable lumber.
(b) Calculate the correlation $r$ for height and the volume of usable lumber. Show all your calculation.

2. Suppose that exam scores in a large class average 80, with a standard deviation of 5. Suppose further that the scores are normally distributed.
   (a) What percentage of students will have exam scores between 80 and 90? (Use 68-95-99.7 rule)
   (b) What percentage of students will have exam scores above 90? (Use 68-95-99.7 rule)
   (c) If I pick a person at random from the class, what’s the chance their exam score is bigger than 85? (Use 68-95-99.7 rule)
   (d) What percentage of students will have exam scores above 74? (Use Table A)
   (e) Suppose I decide to give prizes to students who score in the top 0.6% of the class. About what score does a student need to get a prize? (Use Table A)

3. The following set of data represents scores of 6 varieties of wine in a wine tasting competition.

   3 7 4 8 4 4

(a) Find the mean and standard deviation of these scores.
(b) Find the five-number summary and interquartile range.