Math 106 Spring 2010
Review Questions for Test 1

1. Suppose that 60% of the calls to a customer service number for an airline company are resolved within 5 minutes. A manager assumes that the calls from customers to a particular customer service representative are independent.

   (a) What is the probability that all of the next 20 calls handled by this customer service representative will be resolved within 5 minutes?

   (b) What is the probability that none of the next 20 calls handled by this customer service representative will be resolved within 5 minutes?

   (c) What is the probability that at least one of the next 20 calls handled by this customer service representative will not be resolved within 5 minutes?

   (d) If the manager has 15 customer service representatives working a particular shift, do you think the probabilities in parts (a), (b), and (c) would be the same for each representative? Explain.

2. Researchers examined the possible association between the use of cellular cellphones and the development of a tumor in the inner ear (acoustic neuroma) by comparing 106 cases of individuals with acoustic neuroma to 212 cases of individuals who did not have acoustic neuroma. The data obtained included information on the use of cell phones from personal interviews, data from medical records, and the results of radiologic examinations. Did the researchers conduct an observational or an experimental study? Explain.

3. On many websites there is a poll question that appears on the screen, and you simply click buttons to vote Yes, No, Not sure, or Don't care. One of the poll questions I have read is about the commercials they have showed during Super Bowl, 2009, "Did you like the commercials?". In all, 4316 (29%) said Yes, another 8986 (62%) said No, and the remaining 1182 were not sure or didn't care.

   (a) What is the sample size for this poll?

   (b) That is a much larger sample than standard sample surveys. Do you think this large sample gives good information about any clearly defined population? Explain.

4. The distribution of heights of woman aged 20 to 29 is approximately normal with mean 64 inches and standard deviation 2.7 inches. Use the 68-95-99.7 rule to answer the following questions.

   (a) Between what heights do the middle 95% of young woman fall?

   (b) What percent of young women are taller than 61.3 inches?

5. For the following data set with 5 entries, you know that the mean is 2.3 pounds, but one of the data has become corrupt. Determine the missing entry.

   Weight of yams: 3.1 2.4 1.5 ??? 1.7

   Because of the hormones some yams are gigantic. Suppose in the data above the 1.7 pounds yam is replaced by a 5 pounds yam. What is the new mean of the data set? This change explains that the mean is ..... measure of the center.

6. The relationship between the depth of flooding and the amount of flood damage was examined. The data on P-drive under exercises 5.75 represents x=depth of flooding and y=flood damage.

   (a) Obtain the equation of the least-squares line

   (b) Construct a scatterplot and draw the least-squares line on the plot. Does it look as though a straight line provides an adequate description on the relationship between y and x? Explain, support your answer numerically.

   (c) Predict flood damage for a structure subjected to 6.5 ft of flooding.

   (d) Would you use the least-squares line to predict flood damage when depth of flooding is 18 ft?

   (e) Find and interpret the z-score of flood damage of 26?

(f) Plot the residuals versus the depth of flooding?

(g) Consider the flood damage as a univariate data. Find the mean and standard deviation of this data. Find the five number summary for this data set.

(h) What measures of center and spread would you use for the data of flood damage?

7. Open the data on P-drive under Exercises 4.65. The data is on franchise cost as a percentage of total room revenue for chains of three different types: Budget, Midrange, First-class. Construct a side by side box plot for each type of hotel , and comment on interesting features, similarities and differences.

8. The percentage of juice lost after thawing for 19 different strawberry varieties are given. (Exercise 4.58)

46 51 44 50 33 46 60 41 55 46 53 53 42 44 50 54 46 41 48

(a) Are there any mild outliers, extreme outliers?

(b) Construct a stem and leaf graph for this data.

(c) Construct a histogram for this data?

(d) How would you describe the shape of the distribution?

9. Explain in 2-3 sentences why the median is resistant, that is, why the addition or removal of outliers does not strongly affect the median.

10. Identify and briefly explain the statistical blunders committed.

(a) A newspaper reports a high correlation between gender and personal income, with a correlation coefficient of over $1100.

(b) There is a high correlation between the gender of American workers and their income.

11. The usual way to study the brain's response to sounds is to have subjects listen to "pure tones." The response to recognizable sounds may differ. To compare responses, researchers anesthetized macaque monkeys. They fed pure tones and also monkey calls directly to their brains by inserting electrodes. Response to the stimulus was measured by firing rate (electrical spikes per second) of neurons in various areas of the brain. The file monkey.mtw on P-drive contains the responses for 37 neurons.

(a) One notable finding is that responses to monkey calls are generally stronger than responses to pure tones. Give a numerical measure that supports this finding.

(b) Find the least squares line for predicting monkey call response from pure tone response.

(c) Identify and interpret the value of r^2.

(d) Identify the point with the largest residual.

(e) One point is an outlier in the x direction. Identify this point. How influential is this point on the correlation coefficient?

(f) Would you be willing to use your least squares regression line from part b to predict the monkey call response when the pure tone response is 550? Explain.

12. Different varieties of the tropical flower Heliconia are fertilized by different species of hummingbirds. Over time, the lengths of the flowers and the form of the hummingbirds' beaks have evolved to match each other. The data on the lengths in millimeters of three varieties of these flowers on the island of Dominca are in the file heliconia.mtw under P-drive. Use visual displays and descriptive statistics to compare the three distributions. Your comparison should address center, spread, and shape of the three distributions. What are the most important differences among the three varieties of flowers?

13. You are assigned to direct a study at Kenyon College to discover factors that are associated with strong academic performance. You decide to identify 20 students who have perfect GPAs of 4.0, and then measure explanatory variables for them that you think may be important, such as high school GPA and average amount of time spent studying per day.

(a) Explain what is wrong with this study design.

(b) Describe a study design that would provide more useful information.

14. A study was done on the effect of moving the three-point line in college basketball. The study explored three-point data on 30 randomly selected U.S. college basketball teams, and one variable that was collected wast the number of three-point shots attempted last season. A histogram of this variable is close to bell shaped. The sample of 30 teams had an average of 610 three-point shots attempted, and all but 2 schools had attempted between 450 and 750 three-point shots.

    (a) Provide an accurate estimate of the standard deviation of the number of three-point shots attempted last season. Show your work.

    (b) Using your standard deviation in part a, estimate the z-score for three-point attempts of University of Arizona's team, which attempted 594 three-point shots last year. How do you interpret this z-score? Explain your work.

15. Bring up the data set Poplar1.mtw on P-drive, which contains the height, diameter, and weight of wood after processing for 14 poplar trees. A company wishes to predict weight of wood after processing using measurable variables on the tree before cutting, such as tree height and diameter.

    (a) Which is a better predictor of wood weight? Support your answer with values of r, r^2, regression line etc.

    (b) Look at now the relationship between weight and the squared diameter. What is the value of r^2 this time. What do you think about this transformation?