Statistics 1, Fall 2008   Name_____
Final Exam
Calculators and four 8.5" by 11" sheets of handwritten notes allowed. Show all work and answers clearly in the space provided. Partial credit can only be determined based on the work shown on this exam paper. Tables are included at the end of the exam. Have a great Winter Break!

1. Answer the following questions.
    a. The exam scores of five students are: 60, 82, 75, 98 and 88. Calculate the mean, median and sample standard deviation. (OK to use calculator functions)

    b. Circle the dataset having **larger** standard deviation:
        i.   The salaries of professional basketball players or the salaries of elementary school teachers in California
        ii. The weights of 100 dimes or the weights of 100 coins randomly selected from a cash register
        iii. Height of trees sold at a Christmas tree lot or heights of trees on the Sacramento State University campus

    c. A histogram of the amount of time it took 1000 subjects to complete a puzzle is shown below. Does the Empirical Rule apply to this dataset?_____ Does Chebychev's Rule apply?_____

    d. The data have mean 30 and standard deviation 7. Calculate the interval $\mu \pm 2\sigma$. Use both the Empirical Rule and Chebychev's Rule to estimate the proportion of data falling in this interval. Which rule gives a more accurate estimate?_____

Time to complete puzzle (minutes)
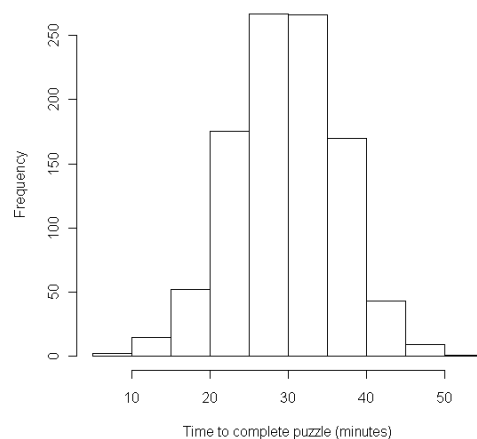
e.   Would Chebychev's Rule apply to the data shown in the histogram below? _____Would the Empirical Rule apply?_____
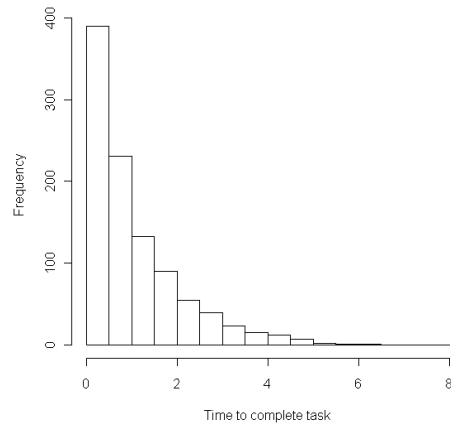
2.   A study is conducted to compare two weight loss diets, Diet 1 and Diet 2. 100 subjects follow Diet 1 for a year and another 100 follow Diet 2 for a year. At the end of the year, the weight loss (in pounds) of each subject is measured. The lower boxplot summarizes the weight loss for subjects on Diet 1. The upper boxplot summarizes weight loss for subjects on Diet 2. Use the boxplots to answer the following questions:
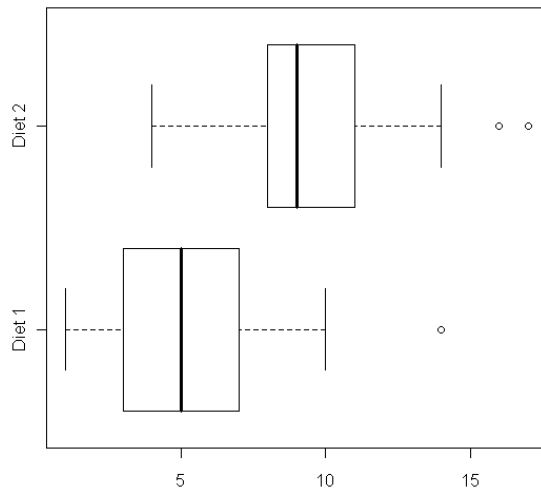

a.   Which diet appears to be more effective? Give reasons for your choice.

b.   Use the boxplot to approximate the first quartile, median and third quartile of weight loss **for Diet 1**. What is the interquartile range?

c.   Give a definition of an outlier. How are outliers denoted in these boxplots?

d. For each subject, two variables are recorded: diet type and weight loss. Classify these as qualitative, quantitative-continuous, or quantitative-discrete variables. Diet type _____ weight loss_____

3. Correlation and Regression
   a. Indicate whether the correlation coefficient in likely to be negative, positive or near zero for each pair of variables.
      i. The number of miles you ride in a taxi cab and the total taxi cab fare
      ii. A stat 1 student's height and his/her score on this final exam
      iii. The number of units a college student has completed toward his/her degree and the number he/she still has to finish
      iv. The number of days left until Christmas and the number of shoppers in the mall
   b. The scatterplot below shows x = price of a widget and y = number of widgets sold. Circle the equation that best fits the data in the scatterplot. $y = -2.1x + 27.4$, $y = 2.1x+27.4$,   or $y = x$
   c. Use the equation you chose in (b) to predict the number of widgets sold when the price is $5.

   d. Which is the correlation coefficient: -1.00, -0.92, 0, 0.92 or 1.00?
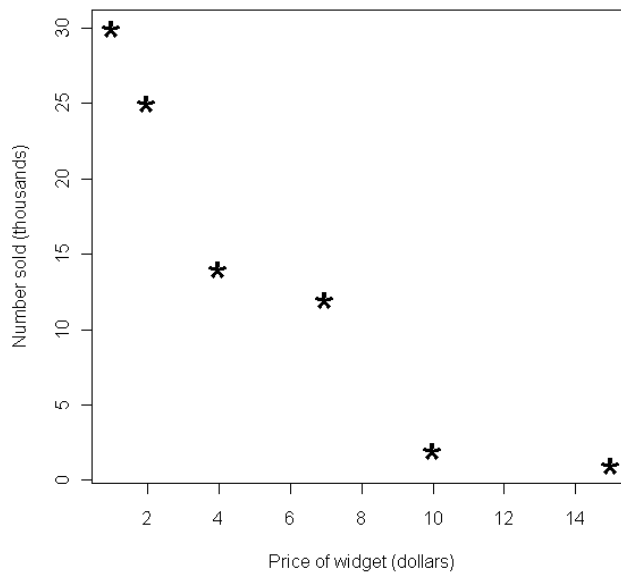
4. The weights of newborn babies is normally distributed with a mean of 7.2 pounds and standard deviation 1.1 pounds.

    a. What percent of newborn will weigh over 9 pounds?

    b. Suppose a newborn is considered "at risk" if his/her weight is among the lowest 1% of all newborn weights. Find the maximum weight of an "at risk" newborn.

    c. What is the probability a random sample of 100 newborns has mean weight over 7.5 pounds?

d. Indicate which graph (solid, dashed, dotted) matches the given distribution:
   1. Distribution of weights of newborns_____
   2. Distribution of sample means for samples of size 64 _____
   3. Distribution of sample means for samples of size 10_____

Newborn weight
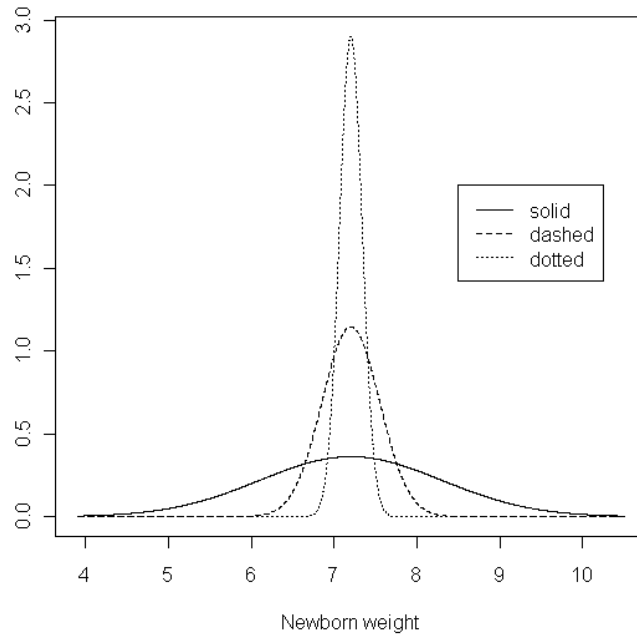
5. A life insurance policy costs $200. If the insured dies within the next year, the policy will pay $150,000 to the beneficiary. Suppose the probability the insured dies is 0.0001. Let the random variable X = the net profit to the insurance company.

   a. What are the possible values for X?

   b. Give the probability distribution of X.

   c. Find the expected value and standard deviation of X. (If you cannot solve part (b), you may make up a probability distribution and use it for this part.)

   d. If the company sells 100,000 of these insurance policies, what would you expect its net gain be?

6. Suppose that 10% of computers are infected with the "Disk Destruction" virus. If an office has 20 computers and we can assume they are infected with the virus independently, what is the probability that

    a. Two or less are infected with the virus?

    b. More than 2 are infected?

    c. If a large company has 1000 computers. On average, how many would you expect to be infected with the virus? What is the standard deviation of the number infected out of 1000?

7. A poll is conducted to determine the proportion of Americans who take vitamins. In a random sample of 200 Americans, 81 said they take vitamins.

    a. Construct a 90% confidence interval estimate of the true proportion who take vitamins.

    b. Can we conclude from part (a) that a majority of ALL Americans take vitamins? Why or why not?

    c. What effect does increasing the confidence level have on the width of the confidence interval? Circle one: Increases width or decreases width

    d. Suppose that of the 200 Americans randomly sampled, 3 said they own a motorcycle. Can we use the CI formula from part (a) to construct a CI (do not actually calculate the interval)? Why or why not?

    e. If the sample size is doubled and everything else remains the same (including $\hat{p}$), how does the width of the CI change? Give the precise percent by which it changes.

    f. What is $z_{\alpha/2}$ if $\alpha=0.14$?

8. Suppose that 15% of all Americans are left-handed. Suppose we randomly sample 3 Americans and observe whether they are left- or right-handed.
   a. List the sample space, S, for this experiment. (Hint: you might let L = the person in left-handed and R = the person is right-handed).

   b. Let X=number of left-handed people among the 3 selected. List the outcomes in the sample space from (a) corresponding to X=1.

   c. Are all outcomes in this sample space equally likely? i.e. is LLL just as likely s RRR?
   d. Calculate P(X=1) (be careful, consider your answer to part (c))

9. A drug manufacturer produces bottles of aspirin tablets. Each tablet is suppose to contain 200 mg of aspirin. The FDA wants to determine if the true mean amount of aspirin in a tablet differs from 200 mg. A random sample of 36 tables has mean aspirin content 198 mg with standard deviation 12 mg. Do the data provide evidence at the 0.05 level of significance that the true mean aspirin content differs from 200 mg?

10. A study was conducted to determine if the drug Prempro, widely used in hormone replacement therapy, increases the risk of dementia. Of 2260 women on Prempro, there were 40 cases of dementia. Of 2270 women on a placebo, there were 21 cases of dementia.

    a. Give a point estimate for the difference in the proportion of dementia in the Prempro and placebo groups

    b. Obtain a 95% confidence interval estimate for the difference in the proportions.

    c. Based on your answer to part (b), does there appear to be a difference in the proportion of dementia in the two groups? Do you think Prempro increases the chance of dementia? Justify your answer.

11. ABC Beverage Company produces 12 oz. cans of root beer. Consumers have been claiming that ABC Beverages is underfilling its cans of root beer. A sample of 36 cans yields a mean weight of 11.4 oz with a standard deviation of 0.68 oz. At the 0.05 level of significance test the claim that ABC is underfilling its cans of root beer. Note that the alternative hypothesis will represent the statement that ABC is underfilling its root beer.

12. Type I and type II error A paramedic arrives at the scene of a car crash and finds the injured and unconscious driver lying on the side of the road. She conducts the following hypothesis test:

~~Type I and type II error A paramedic arrives at the scene of a car crash and finds the injured and unconscious driver lying on the side of the road. She conducts the following hypothesis test:~~

~~H₀: The driver is still alive~~
~~Hₐ: The driver is dead~~

~~a.   Give the Type I and Type II errors in the context of this problem. State a possible negative consequence of each.~~

~~b.   Which would be preferable alpha = 0.02 and beta = 0.10 or alpha = 0.10 and beta = 0.02.~~

13. A bowl contains 20 balls. 5 are white and 15 are black. If three balls are drawn from the urn, what is the probability they are all black if they are drawn
    a. With replacement

    b. Without replacement

    c. Extra credit: what is the probability of 2 white and 3 black if five balls are drawn without replacement

14. (Extra Credit) A company bids on two contracts, Contract 1 and Contract 2. The company will earn $30,000 if it gets contract 1 and $50,000 if it gets contract 2.
    a. List the sample space S for this experiment
    b. Suppose the contracts are awarded independently. The probability the company gets contract 1 is 30% and contract 2 is 60%.   Let X = the profit from the contracts. Find the expected profit for the company.