Statistics 13V   Name:_____
Fall 2002
Final Exam   Last six digits of Student ID#:_____

Open book and notes. If you need more space for any question you may use the back of the page.
You should have 6 questions on 4 pages. Please check to make sure you have all pages.

1. (5 pts each, 20 total). Write a few sentences to answer each of the following questions.
    a. Under what circumstances can the results of a study be extended to a population *and* a cause-and-effect conclusion be made?

    b. Explain why is it important to consider the size of the sample when interpreting the results of a significance test

    c. Explain the difference between sampling error and nonsampling error in a survey.

    d. Suppose you wanted to know the mean weight of all players on a rival school's football team. If you knew the individual weights of all of the players, would you need to construct a confidence interval to estimate the mean weight? Explain.

2. (10 pts total) Based on data from nine randomly selected counties near nuclear plants, a regression equation was determined for predicting y = cancer mortality rates (out of 100,000 people) from x = "exposure index." The exposure index was based on factors such as location of the plant and proximity to potentially contaminated water. The results were:
$$\hat{y} = 115 + 9.2x \quad \text{and} \quad r^2 = 85.8\%$$
   a. (3 pts) What is the correlation between exposure index and cancer mortality rate?

   b. (7pts) What is the predicted cancer mortality rate for a county with an exposure index of 3.0? Show your work, then write your answer in a sentence describing how many cancer deaths would be predicted per 100,000 people.

3. (15 pts total) A waste removal company in one community charges customers $1 extra a week for curbside recycling pickup. They know that 40% (.40) of their customers buy this service. A nearby community includes recycling pickup in the basic cost of waste removal, so there is no way to know how many households use the service. The city manager wants to know if the proportion of households who do so is similar to the .40 found in the community that charges customers. He asks a random sample of 900 households, and 45% say they use the service.
   a. (7 pts) Give the shape, mean and standard deviation of the sampling distribution of the sample proportion for a survey of 900 households if in fact 40% of the customers use this service. Show your work.

   b. (5 pts) Based on the sampling distribution you found in part (a), assuming the true proportion is .4, how unlikely would a sample proportion of .45 or higher be in this survey of 900 households? Show your work.

   c. (3 pts) Based on your answer in part (b), do you think the proportion of customers who use the service is higher in the community where it is included in the basic cost? Explain.

4. (25 pts total) As part of the 2001 Youth Risk Behavior Surveillance System done biannually by the U.S. Government, a random sample of 12th graders was asked how often they wear a seat belt while driving. Of 1302 females, 964 said "most times or always." Of 1443 males, 924 said "most times or always." Test whether there is a difference in the proportions of 12th grade males and females in the population who would answer that they wear seat belts most times or always. Go through the 5 steps of hypothesis testing.
(*Data source:* http://www.cdc.gov/nccdphp/dash/yrbs/)

Step 1 (5 pts): *Specify the hypotheses.*

Step 2 (10 pts): *Verify necessary conditions and compute the test statistic.*

Step 3 (5 pts): *Find the p-value.*

Step 4 (2 pts): *Make a conclusion about statistical significance.*

Step 5 (3 pts): *Make a conclusion in context.*

5. (15 pts total) Exercise 2.17 in the textbook gives the ages of the CEOs of the 60 top-ranked small businesses in the US in 1993 (by *Forbes* magazine). Here is a stem-and-leaf plot of the ages, with the number of leaves in each stem counted for you already:

```
Stem Leaves              Number of leaves        Five-number summary
 |3| 23                         2
 |3| 678                        3
 |4| 013344                     6
 |4| 55556677788889            14
 |5| 000000112333             12
 |5| 555666677889             12
 |6| 0111223                    7
 |6| 99                         2
 |7| 04                         2
                               60 Total
```

a. (10 pts) In the space to the right of the stem-and-leaf plot and counts, provide a five-number summary of the ages.

b. (5 pts) Based on the stem-and-leaf plot and the five-number summary, write a few sentences describing the ages, as if you were writing a short news story about the ages of these CEOs. Someone with no training in statistics should be able to understand it.

6. (15 pts total) In a survey of 75 Penn State students one question asked was "How many minutes do you spend talking on the phone in a typical week?" The mean of the responses was 132.6 minutes and the standard deviation was 140.5 minutes.

a. (5 pts) Explain how you can tell that the data are not bell-shaped.

b. (10 pts) Assuming these students are equivalent to a random sample of all college students, find a 95% confidence interval for the mean number of minutes college students spend talking on the phone in a typical week.