**INTRODUCTION TO STATISTICAL REASONING – Final Exam**

You must show your work and/or explain your steps (except on multiple choice questions) in order to get full credit. Carry calculations to completion. (You may leave fractions only if they are fully reduced.) You should also comment on the numerical results whenever it seems appropriate. Be sure to label all graphs.

You may not share a calculator, pencil, paper or anything else during the exam.

**Your Student ID:** | | | | | | | | | |

**Your First Name:** _____ **Your Last Name:** _____

**Your Section Name:** _____

**Your Signature:** _____

Grader use:

| page | points | earned | page | points | earned |
|---|---|---|---|---|---|
| 2 | 6 | | 10 | 8 | |
| 3 | 15 | | 11 | 15 | |
| 4 | 13 | | 12 | 22 | |
| 5 | 6 | | 13 | 15 | |
| 6 | 15 | | 14 | 15 | |
| 7 | 19 | | 15 | 15 | |
| 8 | 10 | | 16 | 15 | |
| 9 | 18 | | 17 | 25 | |
| subtotal | 102 | | Total | 232 | |

***** Do not turn this page until instructed to do so.*****

**Your Student ID:** | | | | | | | | | |

0. Please check that you have pages 2 through 17 plus a z-table.

1. There is a new drug that reduces the body's production of new blood vessels, thus starving the patients' tumors of blood.

   (a) (6 points) How many patients would we need to study in order to estimate the proportion of successful treatments (tumors disappearing) with a margin of error of $0.03$?

**Your Student ID:** | | | | | | | | | |

(b) (10 points) A study with 23 patients is performed, and the cancer disappears in 17 of these patients. Find a 95% confidence interval for the proportion of patients whose tumors disappear.

2. (5 points) The consumer price index for apparel (Oct 1998) is 135.6, with a base year of 1982-84. How much has the cost of apparel increased since 1982-84?

**Your Student ID:** | | | | | | | | | |

3. The nitrogen content of the leaves of a population of strawberry plants at a particular stage of growth has mean 2.43 parts per million (ppm) and standard deviation of 0.025 ppm. Assume the nitrogen content varies normally.

   (a) (5 points) Fill in the blanks. In a large sample of leaves, about $99.7\%$ of the leaves will have nitrogen content between

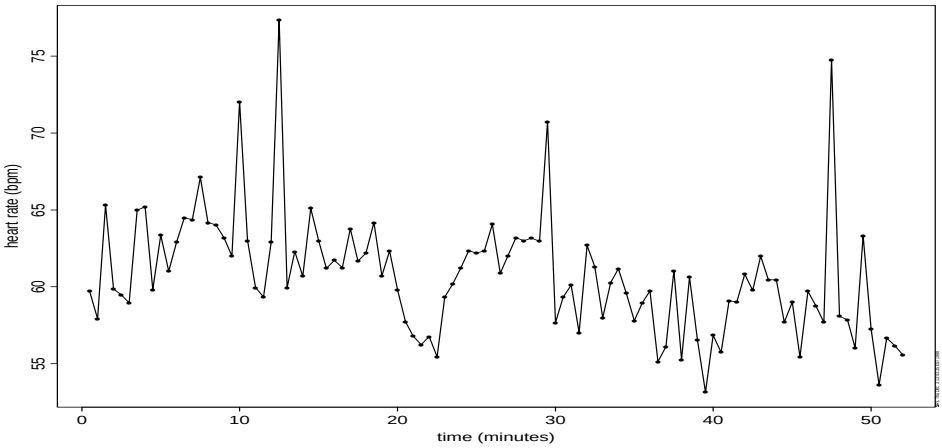
   _____ and _____.

   (Show your work.)

   (b) (8 points) What proportion (in a large sample) would have nitrogen content between 2.425 and 2.456 ppm?

**Your Student ID:** | | | | | | | | | |

4. (6 points) The plot below is of the heart rate of a healthy middle aged man while walking on a treadmill. Describe the data.



**Your Student ID:** | | | | | | | | | |

5. (15 points) My friend's family makes bets with each other about which team will win in each NFL football game during the fall season. During a particular week, my friend's dad was correct for 14 out of the 15 games of the week. Use a hypothesis test to decide whether dad actually has skill or could just have been guessing.

**Your Student ID:** | | | | | | | | | |

6. The Roper Organization (1992) conducted a study as part of a larger survey to ascertain the number of American adults who had seen a ghost. The results are shown in the figure below.

| | seen a ghost | not seen a ghost | total |
|---|---|---|---|
| aged 18 to 29 | 212 | 1,313 | 1,525 |
| aged 30 and over | 465 | 3,912 | 4,377 |
| Total | 677 | 5,225 | 5,902 |

(a) (3 points) What proportion of the older group reported seeing a ghost?

(b) (3 points) What are the odds of seeing a ghost for someone in the younger group?

(c) (8 points) What is the odds ratio of seeing a ghost in the younger group versus the older group?

(d) (5 points) The Minitab output below shows a hypothesis test. State the null and alternative hypotheses for this test (in English or in math notation).

**Your Student ID:** | | | | | | | | | |

```
MTB > chisq c1 c2


             C1       C2    Total
     1      212     1313     1525

     2      465     3912     4377

 Total      677     5225     5902

 ChiSq =  7.857 +  1.018 +
          2.737 +  0.355 = 11.967
 df = 1, p = 0.001
```

(e) (5 points) Use the above hypothesis test to draw conclusions.

(f) (5 points) The expected values were not shown in the above Minitab output. Assuming there is no association between age and having seen a ghost, what is the expected count of those under age 30 who have seen a ghost?

**Your Student ID:** | | | | | | | | | |

(g) (5 points) If the same study had been performed with only 590 people, and the proportions in each category had been about the same, what conclusions would you have drawn from the study?

7. Consider the following study designed to determine which is the best among candidate advertisements for an upcoming movie. Suppose there are two candidate advertisements. Randomly select 10 city blocks from a large city and interview each adult who lives on each of these 10 city blocks. Ask each respondent to say which advertisement he or she prefers.

(a) (5 points) What type of sampling is being used in this study? (Explain.)

(b) (3 points) Which of the following terms best describes this study?

___ a sample survey.

___ an observational study.

___ an experiment.

___ a meta-analysis.

(c) (5 points) What is the sampling frame in this study?

**Your Student ID:** | | | | | | | | | |

8. A poll on a web site asks which movie is your all-time favorite. Suppose that there were only two movies listed, Titanic and Star Wars. Suppose the poll also asked for the age of the respondent, and that the results came out as follows.

| | Titanic | Star Wars | Total |
|---|---|---|---|
| age $< 25$ | 505 | 585 | 1090 |
| age $\geq 25$ | 154 | 717 | 871 |

(a) (5 points) A 95% confidence interval for the odds ratio of preferring Titanic for young versus old respondents is $3.67$ to $4.40$. Interpret this confidence interval and draw conclusions about the differences in age groups.

(b) (3 points) This survey sample is an example of

___ a convenience sample

___ a simple random sample

___ a stratified sample

___ a volunteer sample.

**Your Student ID:** | | | | | | | | | |

9. Consider a marketing study to evaluate a new type of juice drink. The drink is a mixture of seltzer water and juice, but the researchers wish to determine how much juice to use. They recruit a large number of volunteers, each of whom will try just one drink sample and rate it on a scale of 1 to 100. For each volunteer, the proportion of juice in the drink is randomly selected to be either 5% or 10%. When the proportion of juice was 5%, the researchers added a bright food coloring. The age of the volunteer is recorded, along with the proportion of juice and the subject's rating of the drink. The results of the study are that the drink with 10% juice got the highest rating, and the preference was even stronger for younger subjects.

   (a) (3 points each) Which variable is...

      i. the explanatory variable? _____

      ii. the response? _____

      iii. a confounder? _____

      iv. an interacting variable? _____

   (b) (3 points) If the volunteer did not know what proportion of juice he or she was being given, but the experimenter did know, we can say that the study is

   _____.

**Your Student ID:** | | | | | | | | | |

10. The following is a random sample of 20 homework scores for homework 9 in this course: 43 70 97 88 90 93 102 99 93 92 86 86 88 58 101 0 0 0 46 85.

    (a) (10 points) Create a stem and leaf plot.

    (b) (12 points) Describe the distribution of scores.

**Your Student ID:** | | | | | | | | | |

(c) (15 points) The mean of these scores is 70.85 and the standard deviation is 34.81. Find a 95% confidence interval for the true population mean. (Assume the scores are normally distributed.)

**Your Student ID:** | | | | | | | | | |

(d) (15 points) In this course, 175 students attended both of labs 8 and 9, while the remaining 50 students missed at least one of labs 8 and 9. The mean score on homework 9 for all students who attended both labs 8 and 9 is 79.1 with standard deviation 26.5. The mean score on homework 9 for all students who missed at least one of labs 8 and 9 is 64.9 with standard deviation 37.96. Perform a hypothesis test to determine whether the two groups have different mean scores. (Consider the actual observed scores to be a random sample from the possible scores that could have occurred.)

**Your Student ID:** | | | | | | | | | |

(e) (10 points) Find a 99% confidence interval for the mean difference in scores in part (d) (and state what it means).

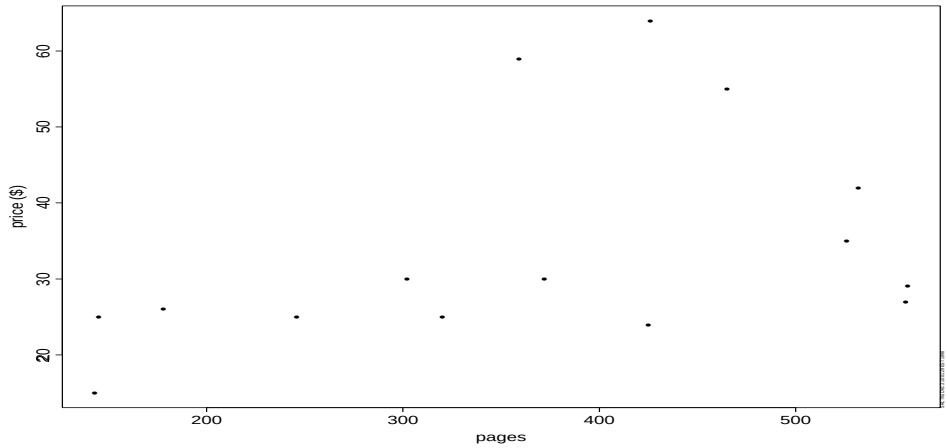
(f) (5 points) Draw conclusions based on your confidence interval.

**Your Student ID:** | | | | | | | | | |

11. The journal *Technometrics* (Feb. 1982 issue) reviewed 15 books. The relationship between the prices of these books and the number of pages in these books is summarized by the regression equation

$$\text{price } (\$) = 19.40 + .0396 \times \#\text{ pages}$$

and the data are shown in the figure below.




(a) (5 points) Use the equation to predict the price of a book that has 200 pages.

(b) (5 points) On the figure, sketch the regression relationship.

(c) (5 points) Interpret the slope.

**Your Student ID:** | | | | | | | | | |

12. (5 points) In the cyberspace study described in class, subjects were given computers, then their internet usage was monitored for two years.
    This is a ___retrospective ___case-control ___controlled ___prospective study. (Check one.)

13. (5 points) The length of time a subject is logged in to the internet is an example of a ___discrete ___continuous ___categorical variable. (Check one.)

14. (5 points) In a study about the effects of child abuse on crime, researchers recruited people convicted of violent crimes and interviewed them to determine the levels of abuse they were subjected to as children. The researchers recruited a comparable group of people who lived in the same neighborhoods the convicts had last lived in, and who were approximately the same ages as the convicts. The same interviews were performed with this second group, in order to see whether there would be less child abuse reported among non-convicts.
    This is ___a case study. ___a case-control study. ___an experiment. ___a prospective study. (Check one.)

15. (5 points) Suppose the 95% confidence interval for the mean height of male 201 students is reported as the interval from 68.7 inches and 69.3 inches.

    True or false: A randomly selected 201 male has a 95% chance of being between 68.7 and 69.3 inches tall.

16. (5 points) Which of the following could reasonably be a 95% confidence interval for the mean weight of CMU freshmen males, assuming the interval is based on a sample of size 100, and the sample mean is 165 lb.s?

    ___ 162 pounds to 168 pounds

    ___ 140 pounds to 190 pounds

Insert z-table here!