Statistics 13V
Fall 2006
Final Exam KEY, Friday, Dec 15th, 2006

Name: ________________________________

Last 6 digits of Student ID#: ________________

Open book and notes. If you need more space use the back of the page or a separate sheet of paper. You should have 4 questions on 4 pages, make sure you have them all.

1. (30 pts total) *The following scenarios are modified from various quizzes throughout the quarter (but you don't need your quizzes to answer the questions here)*. For each scenario, do the following:
   - (3 pts each) Identify the parameter of interest (mean, proportion, etc.) by giving the *correct symbol*
   - (4 pts each) *Explain in words* what the symbol represents in context, including the population to which it applies
   - (3 pts each) *Identify* all of the following that would be appropriate:
     A confidence interval, a one-tailed hypothesis test, a two-tailed hypothesis test.

   a. (From Quiz 1) Researchers would like to compare meditation and exercise to see whether one is more effective for reducing blood pressure, and if so, how much more effective. One hundred people who suffer from high blood pressure volunteer to participate in a study for ten weeks. Participants either will be given a 10-week course in meditation or will participate in a 10-week exercise program. Change in blood pressure over the 10 weeks will be measured. Assume the volunteers are representative of all adults with high blood pressure.

   Symbol: ___ **$\mu_1 - \mu_2$** _____

   Explanation:

   μ_1 = mean change in blood pressure for the population of adults with high blood pressure, if they were to participate in a 10-week course in meditation.
   μ_2 = mean change in blood pressure for the population of adults with high blood pressure, if they were to participate in a 10-week exercise program.

   Relevant inference method(s) (circle all that apply): [**Correct answers in bold**]
   Confidence interval One-tailed hypothesis test **Two-tailed hypothesis test**

   b. (From Quiz 2) A researcher is studying romantic relationships among college students between the ages of 17 and 22. One question of interest is how long, on average, their relationships last. He gives a survey to a large class, which includes the question "What is the longest time you have been involved in a romantic relationship with someone? Answer in number of months."

   Symbol: ____ **μ** _______

   Explanation:

   μ = mean length of the longest romantic relationship for the population of college students between the ages of 17 and 22.

   Relevant inference method (circle all that apply): [**Correct answer in bold**]
   Confidence interval One-tailed hypothesis test Two-tailed hypothesis test

Question 1, continued…

c. (From Quiz 5) In a study reported in Chapter 3 of the textbook, researchers were able to get information about voting behavior of a sample of registered voters. They knew whether or not these people voted in the November 1986 election and wanted to know if those who actually voted would be more honest about their voting than those who did not, and if so, to what extent. Seven months after the election, they surveyed these people and asked them whether or not they had voted in that election. Of those who actually *had* voted, 96% told the truth and said that they did. Of those who had *not* voted, 60% told the truth and admitted that they did not. Suppose these people are representative of the populations of registered voters who do and do not vote.

Symbol: $p_1 - p_2$

Explanation:

p_1 *= the proportion that would tell the truth about voting in the population of all registered voters who actually vote.*
p_2 *= the proportion that would tell the truth about not voting in the population of all registered voters who do not actually vote.*

Relevant inference method (circle all that apply): [**Correct answers in bold**]
Confidence interval **One-tailed hypothesis test** Two-tailed hypothesis test

2. (25 points total) As indicated in one of the presentations shown in class, the M&M Mars Company claims on its website that 20% of its plain chocolate M&M candies are orange. The total number of M&M candies in the class sample was $n = 4893$.

a. (10 pts) Describe the sampling distribution of the sample proportion of orange M&Ms in a random sample of this size, if indeed 20% (.2) of all M&Ms are orange. Include the shape as well as numerical values for the mean and standard deviation, carried to 4 decimal places.

The sampling distribution is approximately normal, with mean = p = .2 and standard deviation = $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.2(.8)}{4893}} = .0057.$

b. (10 pts) In the class project, there were 1007 orange M&Ms, so the proportion was $\frac{1007}{4893} = .2058.$

Find the probability of observing a sample proportion this large or larger, if the sampling distribution you described in part (a) is correct.

$$P(\hat{p} \geq .2058) = P(z \geq \frac{.2058 - .2}{.0057}) = P(z \geq 1.02) = .1539$$

Question 2, continued....

c. (5 pts) Using the data from the class survey, a 95% confidence interval for the true proportion of orange M&Ms is .194 to .217. Based on this result, do you think the class results were consistent or were inconsistent with the information given on the company website for orange M&Ms, which claimed that 20% are orange? Explain.

   The class results were consistent with the 20% reported on the website. The confidence interval provides a range of values that is likely to cover the true proportion, and .20 is included.

3. (20 points total) Write a few sentences to answer each of the following questions.

   a. (5 pts) A researcher compares two different treatments for depression. The result is not statistically significant so the researcher says that the two treatments are equally effective. Do you agree or disagree with the researcher's comment? Explain.

      Disagree. Just because the null hypothesis that the treatments are equal cannot be rejected, it doesn't mean that it can be accepted. The correct interpretation would be to say something like "based on the statistical evidence, we cannot conclude that the treatments differ." There are many ways you could say this.

   b. (5 pts) An observational study found a statistically significant relationship between regular consumption of tomato products (yes, no) and development of prostate cancer (yes, no), with lower risk for those consuming tomato products. Based on this result, can we conclude that regular consumption of tomato products lowers the risk of developing prostate cancer? Explain.

      No. Because the results are based on an observational study, we can conclude that there is a relationship, but we cannot conclude that it is "cause and effect." There may be confounding variables that can explain the relationship.

   c. (5 pts) There are many words that have a different or more precise meaning in statistics than they do in common usage. One example is "significant." Give *two* other examples. No explanation is necessary, unless you think it is required to make sure you are understood.

      <u>Some possibilities</u>: *normal, correlation, independent, blind, sensitivity, residual, density, expected value, variance, block.*

   d. (5 pts) What is the relationship between a parameter and a statistic in statistical inference?

      A statistic is a value from the sample that is used to estimate a parameter, which is a value representing a feature of a population.

4. (25 points total) Example 11.7 and Exercise 13.38 in the book (not required to solve this problem) both present data on the amount of sleep students reported getting the night before, in surveys of Statistics 10 (n = 25) and Statistics 13 (n = 148) students. (Statistics 10 is a GE course called "Statistical Thinking.") Assume these students are representative of all past and future students in these courses. The question of interest is whether the population means differ. The display below contains data summaries and partial Minitab output (to save you from heavy computation) and you are welcome to use the results in the display if they are relevant. Although graphs are not shown, there were no extreme outliers or skewness in the data. Carry out the five steps of hypothesis testing in this situation, which are summarized briefly below. Use $\alpha = .05$.

```
Sample    N  Mean  StDev  SE Mean
1        25  7.66   1.34     0.27
2       148  6.81   1.73     0.14
Difference = mu (1) - mu (2)
Estimate for difference:  0.850000
95% CI for difference:  (0.235816, 1.464184)
T-Test of difference = 0 (vs not =): T-Value = 2.80 DF = 38
```

Step 1 (5 pts) *Specify the null and alternative hypotheses*:

H_0: $\mu_1 - \mu_2 = 0$
Ha: $\mu_1 - \mu_2 \neq 0$
where μ_1 = the mean hours of sleep for the population of Statistics 10 students (past, current and future) and μ_2 = the same thing for Statistics 13 students.

Step 2 (5 pts) *Verify necessary conditions and compute the test statistic*:

Conditions are that the sample sizes are large *or* the populations are bell-shaped; this latter condition is satisfied as long as there are no outliers or extreme skewness. In this case, we are told that's the case, so even though one sample size is only 25, the conditions to proceed are met.

Test statistic is given in the computer output as t = 2.80.

Step 3 (5 pts) *Find the p-value [or p-value range]*:

From the computer output, df = 38. The test is two-sided, so p-value = 2 × area above 2.80 for a t-distribution with df = 38. If you use your calculator, you can find this exactly as .008. Using Table A.3, we find that the area above 2.80 is between .003 and .007 (or .002 and .008), so the p-value is between .006 and .014.

Step 4 (5 pts) *Make a conclusion about statistical significance*:

Because the p-value is less than .05, reject the null hypothesis. The difference is statistically significant.

Step 5 (5 pts) *Report the conclusion in context*:

There is a statistically significant difference in mean hours of sleep a night for the populations of Statistics 10 and Statistics 13 students.

END OF EXAM. HAVE A GREAT BREAK!