

**Sociology 592 - Research Statistics I**  
**Exam 2**  
**November 7, 2003**

**1.** (10 points each, 30 points total.) You have been asked to serve as a statistical consultant for several proposed projects. For each of the following, your employers want you to tell them:

- (i) Which of the cases we have studied their problem falls under (e.g. one sample tests, case I,  $\sigma$  known; nonparametric tests, case II, tests of association). Briefly explain why.
- (ii) the null and alternative hypotheses
- (iii) whether a Z, T, chi-square, or F test is appropriate; where applicable, also tell what the degrees of freedom for the test are. You DO NOT have to give the formula for the test statistic, nor do you need to specify the acceptance region.

If values for population parameters are not specified (e.g.  $\sigma$ ) assume they are unknown; and if two or more unknown  $\sigma$ 's are involved, assume they are equal.

a. Notre Dame wonders what causes some graduate students to take longer to finish than others do. It suspects that field of study may have something to do with it. It will therefore draw random samples of 30 recent Ph.Ds each from the humanities, social sciences, and science and engineering. For each Ph.D., the number of years spent in graduate school will be determined.

b. A professor being considered for tenure needs proof that students are learning something from him. On the first day of his Introductory Sociology class, his 100 students will be given an exam that measures general knowledge of Sociology (like most exams, the possible scores range from 0 to 100). Students will not be told how they scored or what the correct answers were. On the last day of class, the students will once again take this same exam.

c. A prosecutor has several domestic violence cases coming up. To aid her with her jury selection, she wants to know if she is correct in believing that women care more about domestic violence than men do. A sample of 60 men and 60 women will be drawn. Each person will be asked whether or not they feel domestic violence is an important problem.

**2.** (5 points each, 20 points total). For each of the following, indicate whether the statement is true or false. If you think the statement is false, indicate how the statement could be corrected.

NOTE: These are all pretty easy, but you could waste a great deal of time on some of them or make stupid mistakes if you don't happen to see what the easiest way to approach each problem is.

a. A researcher has a sample consisting of 522 subjects. For each subject, she has collected information on 3 variables, each of which has 3 categories. When she tests the model of conditional independence for these three variables, she gets a chi-square value of 27. If she is using the .05 level of significance, she should reject the model of conditional independence.

b. The null and alternative hypothesis are

$$H_0: p = .4$$

$$H_A: p < .4$$

Data are collected from 115 cases. The researcher computes a chi-square test statistic value of -4. Using the .05 level of significance, she should reject the null.

c. A researcher has collected data from 36 respondents on their religion (Catholic or non-Catholic), gender (male or female), and support for the Democratic Party (measured on a scale that ranges from -80 to +80). She computes

$$F_{J-1, N-JK} = \frac{SS\ Rows/(J-1)}{SS\ Error/(N-JK)} = \frac{MS\ Rows}{MS\ Error} = 347$$

If religion is her row variable, she should conclude that religion affects support for the Democratic Party but gender does not.

d. A researcher believes that a problem falls under 2 sample tests, case II,  $\sigma_1$  and  $\sigma_2$  are unknown and assumed equal. In reality, the problem falls under 2 sample tests, case IV, matched pairs. Fortunately, so long as  $N_1 = N_2$ , both approaches will yield the same computed value for the test statistic although the degrees of freedom for the two tests may differ.

**Answer two of the following three questions. (25 points each; you will get up to 10 points extra credit if you answer all three correctly.)**

**3.** In recent years, more and more Americans have turned to non-traditional, alternative forms of health care, such as chiropractic care and acupuncture. Medical researchers are interested in determining whether the use of such alternative methods is related to education. One hundred randomly selected individuals are asked whether or not they use alternative forms of healthcare. Their survey yields the following information:

**ALTRNATV Does respondent use alternative health care? \* EDUC Level of Education  
Crosstabulation**

| Count   |          | EDUC Level of Education |             |          | Total |
|---|----------|-------------------------|-------------|----------|-------|
|   |          | 1.00 High               | 2.00 Medium | 3.00 Low |       |
| ALTRNATV Does respondent use alternative health care? | 1.00 No  | 10                      | 25          | 30       | 65    |
|   | 2.00 Yes | 15                      | 15          | 5        | 35    |
| Total   |          | 25                      | 40          | 35       | 100   |

Using our five-step hypothesis testing procedure and the .05 level of significance, determine whether or not education is related to the use of alternative medicine. If there is a significant relationship, use the information you have been given to explain exactly what you think that relationship is, i.e. how does the use of alternative health care appear to differ by levels of education?

**4.** A school district is very concerned by its low mean score of 60 on an achievement test of its 8<sup>th</sup> graders. An outside firm contends that it can do a better job with the district's students than the current schools are. The district and the firm agree to conduct a 1 year pilot study to test the firm's claim. Thirty randomly selected eighth graders are enrolled in the firm's program. At the end of the year the firm's students get an average score of 67 on the achievement test with a sample standard deviation of 20.

a. Using our 5-step hypothesis testing procedure and the .05 level of significance, determine whether the firm's claim that it does a better job is supported.

b. The teachers' union opposes hiring the outside firm. To support its argument, it notes that the district's score of 60 falls within the 95% confidence interval. Is the union correct in saying that the confidence interval includes 60? If so, is its argument valid? Why or why not?

**5.** With his fundraising continuing to go well and support from labor unions starting to grow, Howard Dean is more confident than ever that he will be the Democratic nominee for President. Nonetheless, he wonders whether his support differs across key demographic groups. Specifically, he wants to know whether his support differs by race (coded as white, black, other) and age (coded as young, middle-aged, and elderly). For each combination of race and age, 20 voters are interviewed. Support for Dean is measured on a scale that ranges from a low of -30 to a high of 30. The survey shows that support for Dean has a mean of 5 and a variance of 125.

a) Complete the following Anova table. You do NOT need to indicate whether the F values are statistically significant or not.

| Source                    | SS   | D.F. | M. S. | F |
|---------------------------|------|------|-------|---|
| A + B (or Main Effects)   |      |      |       |   |
| A (Race of voter)         |      |      | 100   |   |
| B (Age)                   | 180  |      |       |   |
| AB (or 2-way interaction) |      |      |       |   |
| A + B + AB (or explained) |      |      |       |   |
| Error (or residual)       | 1710 |      |       |   |
| Total                     |      |      |       |   |

b) Explain what significant interaction terms might mean. Be specific; don't just talk about interaction terms in general, rather, talk about what interactions involving the variables in this analysis might be due to.