

Sociology 592 - Research Statistics I
Exam 2
November 5, 2004

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, σ 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. σ) assume they are unknown; and if two or more unknown σ 's are involved, assume they are equal.

a. The Democratic Party believes that President Bush's support will decline over the next year. Three hundred voters will be interviewed and asked their level of support for the President, on a scale that ranges from a low of 0 to a high of 70. One year from now, those same 300 voters will be asked to again rate the President.

b. A nationwide study of state-funded Universities has found that 25% of the students at those schools favor the right of gays to get married. Notre Dame wants to see how its students compare. Therefore, 60 Notre Dame students will be asked whether they approve or disapprove of gay marriage.

c. Doctors are worried about the nationwide shortage of flu vaccine. They have developed a new vaccine that can be quickly manufactured in mass quantities. But, first they must see whether it is more effective, equally effective, or less effective than the old vaccine in preventing the flu. Five hundred randomly selected subjects will receive the new vaccine while another 500 will receive the old. After a month, the doctors will measure how many members of each group have gotten the flu.

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. When there are many cells in a table and the expected cell frequencies are very large, e.g. several hundred in each cell, Fisher's Exact Test should be used instead of a chi-square test.

b. A One-way ANOVA with 8 groups produces a highly significant F value. The researcher wants to determine which groups differ from each other. An LSD (Least Significant Difference) test is desirable because such a test controls for the overall probability of rejecting the hypotheses that some pairs of means are different, when in fact they are equal.

c. Using a balanced design, a researcher has collected data from 36 respondents on their race (white, black, other), religion (Catholic, NonCatholic) and political conservatism (measured on a continuous scale ranging from 0 to 30). She finds that

$$F_{JK-1, N-JK} = \frac{SS\ Cells / (JK - 1)}{SS\ Error / (N - JK)} = \frac{MS\ Cells}{MS\ Error} = 58$$

She should therefore conclude that the political conservatism of individuals is affected by both their race and their religion.

d. A researcher is interested in the relationship between gender and attitudes toward abortion. Specifically, she believes that men are more likely to be pro-life than women are. She interviews 50 men and 50 women, and asks them whether they are pro-life or pro-choice. She obtains the following:

```
. tab gender abortion [fw=freq], chi2
```

gender	Attitudes toward Abortion		Total
	pro-life	pro-choic	
Male	20	30	50
Female	35	15	50
Total	55	45	100

```
Pearson chi2(1) = 9.0909 Pr = 0.003
```

If she is using the .01 level of significance, she should conclude that the data support her position.

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. Once again, the Republican Presidential candidate has carried the state of Indiana. Democrats are trying to find out whether support for President Bush was equally strong in all parts of the state or whether his support varied by geographic region. A survey of 100 Indiana voters produces the following results.

```
. tab candidate region [fw=freq]
```

Who did you vote for?	Region of the State			Total
	North	Central	South	
Bush	15	20	25	60
Kerry	20	15	5	40
Total	35	35	30	100

Using our five-step hypothesis testing procedure and the .05 level of significance, determine whether or not support for Bush varied by region. 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 did support for Bush differ across the state?

4. A major university wants to strengthen its undergraduate statistics course. In the past, this class has been taught using a lecture format with little if any student discussion. Under this format, students have gotten a mean score of 70 on a standardized exam with a population standard deviation of $\sigma = 14$. Sixteen randomly selected students are now in a new statistics course that emphasizes class participation. Instructors believe that this format will produce higher test scores, although they are confident that σ will continue to equal 14. Results show that students in the new course receive an average score of 76 on the exam.

a. Construct the 95% confidence interval. Based on the confidence interval, should you reject the null?

b. Using our 5-step hypothesis testing procedure and the .05 level of significance, test whether the instructors' belief is supported by the data.

5. Most couples want two children. For those who have two children, a demographer believes that their desire for a third (measured on an interval-level scale) may be related to their social status (high, medium, or low) and their religion (Catholic, Protestant, Jewish, or Other). To test her ideas, she collects random samples of 10 individuals (all of whom currently have two children) for each possible combination of social status and religion. She finds that SS Rows= 120, SS Columns = 144, and SS Interaction = 240. When she tests the hypothesis that there are no differences in desires across social status (i.e. when she tests the hypothesis the row effects are all 0) she gets an F of 5.0.

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 (Social Status)				
B (Religion)				
AB (or 2-way interaction)				
A + B + AB (or explained)				
Error (or residual)				
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.