Sociology 592 - Research Statistics I
Exam 2 Answer Key [ROUGH DRAFT]
November 8, 2002

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 Federal Reserve Board is uncertain what effect, if any, the recent elections have had on the public's confidence in the economy. A week before the election, 80 Americans were asked to rate their confidence in the economy, on a scale ranging from a low of 0 to a high of 30. A week after the election, another random sample of 80 Americans will be asked to do the same thing.

Two Sample Tests, Case II, σ_1 and σ_2 are unknown but assumed to be equal (or else One Way Anova will work.)

H_0: $\mu_1 = \mu_2$
H_A: $\mu_1 \neq \mu_2$

Use either T test with df = 158 or F test with DF = 1, 158

b. President Bush feels he has a mandate for his tax policies. His advisors, however, are cautioning him that support for his policies is not universal. In particular, they warn that middle class voters are much less favorable to his tax policies than are upper class voters. One hundred middle class voters and 100 upper class voters will be asked whether they approve or disapprove of Bush's tax policies.

Two Sample Tests, Cave V, Difference between two proportions.

H_0: $p_{Middle} = p_{Upper}$
H_A: $p_{Middle} < p_{Upper}$

Z statistic.

c. Obesity is a growing health concern. A medical researcher knows that race and social class are related to each other, i.e. whites have higher incomes than do nonwhites. She also believes that obesity may be related to race and/or social class. Sixty individuals will be selected at random. For each individual, the researcher will measure (a) their weight – coded as obese or not obese (b) their race – white or nonwhite, and (c) their social class – coded as lower, middle, and upper.

Nonparametric Tests, Case 2, Tests of Association. Specifically, the model of Conditional Independence should be tested. Chi-Square statistic. v = rcl - 1 - (rc - 1) - (l - 1) = 2 * 3 * 2 – 1 – (2 * 3 - 1) – (2 -1) = 5.

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 12 boys and 12 girls in her sample. The null and alternative hypotheses are

H_0: $\mu_1 = \mu_2$
H_A: $\mu_1 > \mu_2$

The computed value of her test statistic is -2.517. If she is using the .01 level of significance, she should reject the null hypothesis.

False. The test statistic needs to be positive in order to reject the null. The negative value means that the sample mean for group 1 is less than the sample mean for group 2, which is the exact opposite of what the alternative hypothesis states.

b. The null and alternative hypothesis are

H_0: $p = .27$
H_A: $p > .27$

Data are collected from 61 cases. The chi-square test statistic equals 3.9. Using the .05 level of significance, she should reject the null.

False. While the chi-square value is large, it does not tell you whether p is significantly greater than .27 or significantly less than .27. She should use a Z statistic instead so she can discern the direction of the difference.

c. A researcher believes that five different TV shows are equally popular with viewers. She should estimate the model of independence to test her hypothesis.

False. The equi-probability model should be used.

d. Critics contend that standardized test scores have declined at Midwestern University. Twenty years ago, the average test score was 1218. A sample of 900 current students yields a mean of 1200 with a standard deviation of 300. Using the .05 level of significance, the null hypothesis should be rejected.

True. Note that the alternative hypothesis is one-tailed, so we will reject if the test statistic is <= -1.64 (not -1.96). The test stat = -1.8, so reject the null.

Answer two of the following three questions. You will get up to 10 points extra credit if you answer all three correctly.

3. A college professor is concerned that there may be a "gender gap" in understanding of statistics. That is, men and women may not understand statistics equally well. A random sample of 100 students is given a test to measure their understanding of statistics, yielding the following results:

GENDER Gender * STATS Understanding of Statistics Crosstabulation

Count

		STATS Understanding of Statistics			
		1.00 Low	2.00 Medium	3.00 High	Total
GENDER Gender	1.00 Male	8	14	28	50
	2.00 Female	12	20	18	50
Total		20	34	46	100

Using our five step hypothesis testing procedure, determine whether understanding of statistics significantly varies by gender. Use $\alpha = .05$. Do the results support the idea that there is a "gender gap" in the understanding of statistics?

Here is a more complete SPSS Analysis:

Gender * Understanding of Statistics Crosstabulation

			Understanding of Statistics			
			Low	Medium	High	Total
Gender	Male	Count	8	14	28	50
		Expected Count	10.0	17.0	23.0	50.0
	Female	Count	12	20	18	50
		Expected Count	10.0	17.0	23.0	50.0
Total		Count	20	34	46	100
		Expected Count	20.0	34.0	46.0	100.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.033[a]	2	.133
Likelihood Ratio	4.061	2	.131
Linear-by-Linear Association	3.275	1	.070
N of Valid Cases	100		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.00.

The chi-square statistic is not significant, so do not reject the null. There is insufficient evidence to conclude there is a gender gap in understanding of statistics.

4. The Democratic Party is very concerned by voter apathy and poor turnout in the last election. It feels that it must do more to get its likely supporters interested in politics. Twenty-five (25) non-voters from areas that went heavily Democratic are selected at random. Their interest in politics is measured on a scale that ranges from a low of 0 to a high of 100. Over the next 6 months, each voter receives a series of mailings that the party believes will increase their interest in politics. At the end of six months, their interest in politics is again measured. For each respondent, the Party computes D = Score at time 1 – Score at time 2. It finds that the mean of D is -7.7 and the standard deviation of D is 15.

a. Compute the 95% confidence interval for D.

$$s_{\bar{D}} = \frac{15}{\sqrt{25}} = 3$$

Critical value for t with 24 d.f. is 2.064, so c.i ranges from -13.892 to -1.508

b. Using our 5 step hypothesis testing procedure, determine whether the mailings appear to be having the desired effect. Use $\alpha = .01$.

Reject if t is less than -2.492. Test stat = -7.7/3 = -2.57, so reject the null. Mailings seem to be having desired effect.

5. A lender is concerned by its failure to make more home mortgage loans to minorities and to minority neighborhoods. It believes that part of the problem is that minority applicants and applicants from minority neighborhoods tend to have lower credit scores than do others. Using a balanced design, it draws a random sample of 1800 loans. Credit score is measured on a continuous scale that ranges from -9 to 9 with a sample mean of 0 and a sample standard deviation of 2. The race of each applicant is coded white, black, or other. The racial composition of the neighborhood the applicant wants to buy a house in is coded Large minority population, Moderate minority population, or Small minority population.

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 borrower)	2500			
B (Racial composition of neighborhood)				
AB (or 2-way interaction)				25
A + B + AB (or explained)	3614			
Error (or residual)				
Total				

Source	SS	D.F.	M. S.	F
A + B (or Main Effects)	3414	4	853.5	426.75
A (Race of borrower)	2500	2	1250	625
B (Racial composition of neighborhood)	914	2	457	228.5
AB (or 2-way interaction)	200	4	50	25
A + B + AB (or explained)	3614	8	451.75	225.875
Error (or residual)	3582	1791	2	
Total	7196	1799	4	

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

One possibility is that blacks in black neighborhoods are especially like to have poor credit scores. For example, being black might reduce your credit score by 10 points. Living in a minority neighborhood might reduce your credit score by 7 points. Being both black and living in a minority neighborhood might reduce your credit score by an additional 5 points, i.e. blacks in minority neighborhoods could have credit scores that are 22 points below average. Those blacks who live in minority neighborhoods might have exceptionally large problems with their credit; perhaps lenders are more likely to report such individuals to credit bureaus when they have problems.