Sociology 592 - Research Statistics I
Exam 2 Answer Key
November 3, 1995

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. [NOTE: You must do three of the following, and you'll receive 5 points extra credit for each additional problem you get right.]

a. A prosecutor believes that black women will be harder on alleged criminals in domestic violence cases than will white women. A sample of 30 black women and 30 white women will be drawn. For each respondent, attitudes toward domestic violence will be measured on a scale that ranges from 0 (cares little about domestic violence) to 100 (feels very strongly about domestic violence).

Two sample tests, case II, sigmas unknown but assumed equal. Counter to what some said, this is NOT matched pairs; these are two independent samples. It would be matched pairs if, say, each black woman were paired with a white woman of similar education and income.

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

T statistic with 58 d.f.

b. Colin Powell wants to know whether he will have better luck running as a Republican or as an independent for President. Thirty respondents are asked to rate, on a scale ranging from 0 to 100, how likely it is they would support Powell if he ran as an independent. Using that same scale, they are also asked how likely it is they would support Powell if he ran as a Republican.

Two sample tests, case 4, matched pairs. Note that support is measured on a continuous scale rather than as a dichotomy. In effect, each respondent receives two "treatments" -- Powell as Republican and Powell as Independent -- and the scores on each treatment are then compared.

H_0: $\mu_{Rep} = \mu_{Ind}$
H_A: $\mu_{Rep} < > \mu_{Ind}$

T test, 29 d.f.

c. The new Provost wants to find out about gender disparities in the rank and tenure of Notre Dame faculty. A random sample of 100 faculty will be drawn. For each faculty member, the gender (male or female) and rank (assistant professor, associate professor, and full professor) will be recorded.

Nonparametric tests, case II, tests of association. Both variables are categorical. Even if you wanted to really stretch things and say that rank was continuous, this would then be a 1-way ANOVA problem, not 2-way, as some said.

H_0: Occupational distribution is the same for men and women
H_A: Occupational distribution differs

Chi-square statistic with 2 d.f.

2. (5 points each, 20 points total). For <u>each</u> 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 collected data from 36 respondents on their religion (Catholic or nonCatholic), gender (male or female), and life satisfaction (measured on a scale ranging from 1 to 50). She computes

$$F_{(J-1)(K-1),N-JK} = \frac{SS\ Interaction/(J-1)(K-1)}{SS\ Error/(N-JK)} = 4$$

If she is using the .05 level of significance, she should conclude that neither religion nor gender has a significant effect on life satisfaction.

False. This says there are no interaction effects, but there could still be main effects of either religion or gender. Substitute SS Cells for SS Interaction in the above equation.

b. A researcher is interested in the relationship between race (3 categories), education (4 categories), and party affiliation (3 categories), where party affiliation is her dependent variable. She estimates the model of conditional independence. If $\alpha = .05$ and her Pearson chi-square statistic = 38, she should not reject the model of conditional independence.

False. d.f. for model of conditional independence = rcl - 1 - (rc - 1) - (l - 1) = 36 - 1 - 11 - 2 = 22. Critical value is 33.92, 38 is greater than this, so reject the null.

c. The null and alternative hypotheses are

H_0: $\mu_1 - \mu_2 = 10$
H_A: $\mu_1 - \mu_2 \neq 10$

The 95% confidence interval ranges from 3 to 11. Ergo, the null hypothesis should not be rejected.

True. 10 falls within the confidence interval.

d. A political scientist believes that black males who participated in the Million Man March (group 1) are more liberal than black males who did not (group 2). She constructs a liberalism scale which ranges from 1 (very conservative) to 100 (very liberal). She draws a sample of 100 black males who attended and 100 black males who did not. She finds that

$$\hat{\mu}_1 = 57,\ \hat{\mu}_2 = 62,\ s_1 = 10,\ s_2 = 9$$

She should not reject the null hypothesis.

True, because the marchers' mean of 57 is less than the non-marchers' mean of 62, which is the exact opposite of what the political scientist hypothesized.

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

3. (25 points or 10 points extra credit) Suppose that the body weight at birth of children within the United States is approximately normally distributed and has a mean of 7.2 pounds. A demographer believes that the birth weights of children born of mothers who are drug addicts may be lower on the average than for the population as a whole. In order to test this hypothesis, he secures records of the birth weights of a random sample of 25 children from mothers who are drug addicts. The mean of this sample is 7 pounds with s = 0.5. Evaluate the demographer's hunch, using our 5-step hypothesis testing procedure and $\alpha = .05$. Also compute the 95% confidence interval.

This falls under single sample tests, case III, sigma unknown.

Step 1: The null and alternative hypotheses are

H_0: $\mu = 7.2$
H_A: $\mu < 7.2$

Step 2: The appropriate test statistic is

$$T_{N-1} = \frac{\hat{\mu} - \mu_0}{s/\sqrt{N}} = \frac{\hat{\mu} - 7.2}{.5/5} = \frac{\hat{\mu} - 7.2}{.1}$$

d.f. = 24

Step 3. Acceptance region.

Accept if T > = -1.711, otherwise reject.

Step 4: The test statistic equals

$$T_{N-1} = \frac{\hat{\mu} - \mu_0}{s/\sqrt{N}} = \frac{\hat{\mu} - 7.2}{.5/5} = \frac{\hat{\mu} - 7.2}{.1} = \frac{7 - 7.2}{.1} = -2$$

Step 5: Reject the null.

For the 95% c.i., the critical value of T is 2.064. Ergo, the c.i. is

$7 - .1 * 2.064 \leq \mu \leq 7 + .1 * 2.064$, i.e.
$6.7936 \leq \mu \leq 7.2064$

4. (25 points or 10 points extra credit) The University is concerned that not all of its students are "computer literate." It believes that computer literacy differs by College (Arts and Letters, Business, Other) and race (white, black, other). For each combination of College and race, 20 students are interviewed. Computer literacy is measured on a scale that ranges from 0 (knows nothing about computers) to 100 (computing genius). The University discovers that computing literacy has a mean of 50 and a standard deviation of 10. Complete the following ANOVA table. You do NOT need to indicate whether or not the F values are statistically significant.

Source	SS	D.F.	M. S.	F
A + B (or Main Effects)				
A (College)			80	
B (Race)	144			
AB (or 2-way interaction)				
A + B + AB (or explained)				
Error (or residual)	1368			
Total				

Note that J = K = 3 (Arts and Letters is one college, not two!). N = 20 * 3 * 3 = 180. Also, MST = s^2 = 100 (although I meant to say that the variance, rather than the standard deviation, equals 10, the problem is still solvable -- but the results are a little weird!) If you fill in the degrees of freedom and first solve for the error and total rows, everything else follows pretty easily.

Source	SS	D.F.	Mean Square	F
A + B (or Main Effects)	SS Main = 304	J + K - 2 = 4	$\frac{\text{SS Main}}{(J + K - 2)} = 76$	$\frac{\text{MS Main}}{\text{MS Error}} = 9.5$
A (or main effect of A)	SS Rows = 160	J - 1 = 2	$\frac{\text{SS Rows}}{(J - 1)} = 80$	$\frac{\text{MS Rows}}{\text{MS Error}} = 10$
B (or main effect of B)	SS Columns = 144	K - 1 = 2	$\frac{\text{SS Columns}}{(K - 1)} = 72$	$\frac{\text{MS Cols}}{\text{MS Error}} =$ **9**
AB (or 2-way interaction)	SS Intraction = 16,228	(J - 1) * (K - 1) = 4	$\frac{\text{SS Intr}}{(J-1)(K-1)} = 4057$	$\frac{\text{MS Intrction}}{\text{MS Error}} = 507.125$
A + B + AB (or explained)	SS Cells = 16,532	(J * K) - 1 = 8	$\frac{\text{SS Cells}}{(J * K) - 1} = 2066.5$	$\frac{\text{MS Cells}}{\text{MS Error}} = 258.3125$
Error (or residual)	SS Error = 1368	N - (J * K) = 171	$\frac{\text{SS Error}}{(N - J * K)} =$ **8**	
Total	SS Total = 17,900	N - 1 = 179	$\frac{\text{SS Total}}{(N - 1)} = 100$	

5. (25 points or 10 points extra credit) It is Summer, 1996. Democratic Presidential nominee Bill Clinton is set to do battle with his Republican opponent, Colin Powell. Ross Perot is now debating whether to enter the race as a third party candidate. Perot believes that he would draw more voters from Clinton than he would Powell. To test his idea, 100 voters are asked who they currently support (Clinton or Powell) and whether they would switch to Perot if he entered the race. The results are as follows:

CURRENT Current choice by SWITCH Switch to Perot?

		SWITCH		
	Count	Yes	No	
		1.00	2.00	Row Total
CURRENT	1.00	20	30	50
Clinton				50.0
	2.00	11	39	50
Powell				50.0
	Column Total	31	69	100
		31.0	69.0	100.0

Using our 5-step hypothesis testing procedure and the .05 level of significance, determine whether the data support Perot's claim that he will steal more voters from Clinton than he will Powell.

The critical thing is to realize that this falls under 2 sample tests, case V, test of P1 = P2. There are two groups here, Clinton supporters and Powell supporters. Perot believes that he will be more successful (i.e. that P, the probability of switching, is greater) with Clinton supporters than with Powell supporters. Note that you can't treat this as a chi-square problem, because a chi-square test will only tell you that the probability of switching is different in the two groups, it won't tell you which group members are more likely to switch.

Step 1:

H_0: $p_1 = p_2$
H_A: $p_1 > p_2$

Step 2: A z-statistic is appropriate, specifically,

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{N_1 + N_2}{N_1 N_2}\right)\left(\frac{X_1 + X_2}{N_1 + N_2}\right)\left(1 - \frac{X_1 + X_2}{N_1 + N_2}\right)}} = \frac{\frac{X_1}{N_1} - \frac{X_2}{N_2}}{\sqrt{\left(\frac{N_1 + N_2}{N_1 N_2}\right)\left(\frac{X_1 + X_2}{N_1 + N_2}\right)\left(1 - \frac{X_1 + X_2}{N_1 + N_2}\right)}}$$

Step 3. Accept the null hypothesis if the test statistic is less than or equal to 1.65, otherwise reject.

Step 4. The computed value of the test statistic is

$$z = \frac{\frac{X_1}{N_1} - \frac{X_2}{N_2}}{\sqrt{\left(\frac{N_1 + N_2}{N_1 N_2}\right)\left(\frac{X_1 + X_2}{N_1 + N_2}\right)\left(1 - \frac{X_1 + X_2}{N_1 + N_2}\right)}} = \frac{\frac{20}{50} - \frac{11}{50}}{\sqrt{\left(\frac{50 + 50}{50 * 50}\right)\left(\frac{20 + 11}{50 + 50}\right)\left(1 - \frac{20 + 11}{50 + 50}\right)}} = 1.946$$

Step 5. Reject the null. It does appear that Perot will draw more supporters from Clinton than from Powell. Note that if the alternative was two-tailed, we would not reject.