Sociology 592 - Research Statistics I
Exam 2 Answer Key
November 17, 1993

Where appropriate, show your work - partial credit may be given. (On the other hand, don't waste a lot of time on excess verbiage.) Do not spend too much time on any one problem. It is legitimate (and probably essential) to refer to results that have previously been proven in class or homework, without re-proving them - for example, you wouldn't need to prove that $P(-1.96 \leq Z \leq 1.96) = .95$, since we have already shown that in class. Likewise, you are free to refer to anything that was demonstrated in the homework or handouts.

1. (10 points each, 30 points total). You have been asked to serve as a statistical consultant for several proposed projects. For <u>each</u> 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).

(ii) the null and alternative hypotheses

(iii) whether a Z, T, or chi-square 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 have to tell what the acceptance region is.

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. Census figures show that, in 1990, 32 percent of the residents in Census tract #6 lived in poverty. The mayor is confident that, thanks to his strong leadership and neighborhood revitalization efforts, the poverty rate has fallen since then. To test his hypothesis, the poverty status of 250 randomly selected residents of the tract will be determined.

Solution.
(i) This falls under single sample tests, Case II, test of the binomial parameter p.
(ii) H_0: $p = .32$
H_A: $p < .32$
(iii) Z statistic is appropriate

b. The Dean contends that religion (Catholic or NonCatholic) has no effect on faculty salaries. Disgruntled NonCatholics counter-argue that they are being discriminated against. Hence, the religion and salary of 60 randomly selected faculty are determined.

Solution.
(i) Two sample tests, case II, variances unknown but assumed equal. Note that Anova is inappropriate because the alternative is one-tailed.
(ii) H_0: $\mu_1 = \mu_2$ (where group 1 is Catholics, Group 2 is nonCatholics)
H_A: $\mu_1 > \mu_2$

(iii) T test, with 58 d.f.

c. Undaunted by his devastating defeat in last Tuesday's NAFTA debate with Al Gore, Ross Perot maintains that opposition to NAFTA has actually increased this week. To prove his point, Perot will compare the results of independent surveys taken last Monday and again today. In each survey, 1,000 Americans were asked whether or not they supported NAFTA.

Solution.
(i) Two sample tests, Case V, test of $p_1 = p_2$
(ii) H_0: $p_1 = p_2$
H_A: $p_1 < p_2$
(iii) Z statistic is appropriate

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 political scientist maintains that labor union members are more likely to support health care reform than are others. She collects the following data:

	Supports reform	Opposes reform
Labor union member	337	663
Not in labor union	578	422

Based on these differences, her claim that the null hypothesis should be rejected in favor of her theory is strongly supported.

Solution. FALSE. You don't even need to do any calculations. Things are obviously just the opposite of what she claimed: labor union members are much LESS likely to support reform.

b. A researcher has 15 husbands and their wives in her sample. Her 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 2.09. If $\alpha = .05$, she should reject the null hypothesis.

Solution. FALSE. Note that this falls under two sample tests, case IV, matched pairs. The d.f. therefore equal 14 and the critical value is 2.145. Since 2.09 is less than 2.145, do not reject.

c. The null and alternative hypotheses are

H_0: $\mu = 10$
H_A: $\mu < 10$

The 95% confidence interval ranges from 7.1 to 10.1. Therefore, the researcher should not reject the null hypothesis.

Solution. FALSE. Since the alternative hypothesis is one tailed, the confidence interval should not be used for hypothesis testing. Indeed, given the c.i., it looks like 10 probably falls outside of the one-tailed acceptance region.

d. A researcher is interested in the relationship between gender (male, female), social class (Upper, Middle, Lower) and evaluation of the Bill Clinton (Favorable or Unfavorable). She believes that feelings towards Clinton are independent of gender and social class, even though gender and social class are not independent of each other; that is, she believes in the model of conditional independence. She gets a pearson chi-square of 11. If $\alpha = .05$, she should not reject the model of conditional independence.

Solution. TRUE. The d.f. = v = rcl - 1 - (rc - 1) - (l - 1) = 2*3*2 - 1 - (2*3 - 1) - (2 - 1) = 5. For a chi-square with 5 d.f., the critical value is 11.07. Since 11 is less than 11.07, she should not reject.

Answer two of the following three questions. You can get 10 points extra credit if you do all three correctly.

3. (25 points; or, 10 points extra credit) The critical NAFTA vote is just days away. Although he has made substantial progress, Clinton still lacks the votes to guarantee passage. In order to help him determine his strategy for the final days before Congress votes, Clinton wants to know how support for NAFTA differs across the country. He will then use this information when lobbying Congress. His pollsters collect the following data:

Region/Voting preference	Support NAFTA	Oppose NAFTA	Undecided	**TOTAL**
Western States	25	20	15	**60**
Eastern States	10	20	10	**40**
TOTAL	**35**	**40**	**25**	**100**

Using our 5-step hypothesis testing procedure, determine whether support for NAFTA significantly differs across the various regions of the country. Use $\alpha = .01$.

Solution.

Step 1.

H_0: Support for NAFTA is the same in all regions
H_A: Support for NAFTA differs by region

or, equivalently,

H_0: $P(A_i \cap B_j) = P(A_i)P(B_j)$ (Model of independence)
H_A: $P(A_i \cap B_j) <> P(A_i)P(B_j)$

Step 2. An appropriate test statistic is

$\chi^2_v = \Sigma\Sigma\ (O_{ij} - E_{ij})^2/E_{ij}$, $v = rc-1-(r-1)-(c-1) = (r-1)(c-1) = 2$

Step 3. For $\alpha = .01$ and $v = 2$, accept H_0 if

$\chi^2_v \leq 9.21$

Step 4. The computed value of the test statistic is

Region/Voting preference	Support NAFTA	Oppose NAFTA	Undecided	**TOTAL**
Western States	(25 - 21)²/21 = 16/21	(20 - 24)²/24 = 16/24	(15 - 15)²/15 = 0	**1.428**
Eastern States	(10 - 14)²/14 = 16/14	(20 - 16)²/16 = 1	(10 - 10)²/10 = 0	**2.143**
TOTAL	**1.905**	**1.667**	**0**	**3.57**

Step 5. Do not reject the null, 3.57 is less than 9.21.

4. (25 points; or, 10 points extra credit) A company wants to make its employees more sensitive to issues of sexual harassment. Fifty people are combined into 25 pairs of "near-twins". In each pair, one person sees a film on sexual harassment (group A) while the other does not (group B). Sensitivity to issues of sexual harassment is then measured on a scale that runs from 0 to 25. For each pair, the researcher computes $D_i = X_{Ai} - X_{Bi}$. The researcher finds that $\Sigma D_i = 100$ and $\Sigma D_i^2 = 1000$.

a. Compute the mean and standard deviation of D.

Solution. Note that this falls under 2 sample tests, case IV, Matched pairs.
a. $\overline{D} = \Sigma D_i/N = 100/25 = 4.0$,

$$s_D = \sqrt{\frac{1}{N-1}(\Sigma d_i^2 f_i - N\bar{d}^2)} = \sqrt{\frac{1}{24}(1000 - 25 * 4^2)} = \sqrt{25} = 5$$

b. Using our 5 step hypothesis testing procedure, determine whether seeing the film significantly increases sensitivity to issues of sexual harassment.

Solution.

Step 1: H_0: $\mu_D = 0$
H_A: $\mu_D > 0$

Step 2: The appropriate test statistic is

$$t_{24} = \frac{\bar{d} - \mu_{D_0}}{\frac{s_D}{\sqrt{N}}} = \frac{\bar{d}}{\frac{\sqrt{25}}{\sqrt{25}}} = \bar{d}$$

The d.f. = 24.

Step 3: Accept H_0 if $t \leq 1.711$

Step 4: The computed value of the test statistic is

$$t_{29} = \frac{\bar{d} - \mu_{D_0}}{\frac{s_D}{\sqrt{N}}} = \frac{\bar{d}}{\frac{\sqrt{25}}{\sqrt{25}}} = \bar{d} = 4$$

Step 5: Reject H_0.

5. (25 points; or, 10 points extra credit) University administrators believe that cultural sensitivity (measured on a scale that ranges from 0, not sensitive, to 50, very sensitive) varies by the geographic region a student is from (North, South, East, or West) and by the student's gender (male or female). For each combination of region and gender, 25 students are interviewed. The university discovers that cultural sensitivity has a mean of 30 and a standard deviation of 5. 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 (Region)	300			
B (Gender)				10.0
AB (or 2-way interaction)				
A + B + AB (or explained)				
Error (or residual)			20	
Total				

Solution.

Note that

J = 4, K = 2, N = 4 * 2 * 25 = 200,
MST = s^2 = 5^2 = 25, DFT = N - 1 = 199,
SST = MST * DFT = 199 * 25 = 4975
DFE = N - JK = 200 - 8 = 192, SSE = DFE * MSE = 192 * 20 = 3840
SS Explained = SST - SS Residual = 4975 - 3840 = 1135,
MS Columns = F Columns * MSE = 10 * 20 = 200
DF Columns = K - 1 = 1
SS Columns = MS Columns * DF Columns = 200 * 1 = 200
SS Main = SS Columns + SS Rows = 200 + 300 = 500
SS Interaction = SS Explained - SS Main = 1135 - 500 = 635

The rest falls easily into place:

Source	SS	D.F.	Mean Square	F
A + B (or Main Effects)	SS Main = 500	J + K - 2 = 4	SS Main / (J + K - 2) = 125	MS Main / MS Error = 6.25
A (or main effect of A)	SS Rows = **300**	J - 1 = 3	SS Rows / (J - 1) = 100	MS Rows / MS Error = 5
B (or main effect of B)	SS Columns = 200	K - 1 = 1	SS Columns / (K - 1) = 200	MS Cols / MS Error = **10**
AB (or 2-way interaction)	SS Intraction = 635	(J - 1) * (K - 1) = 3	SS Intr / (J-1)(K-1) = 211.67	MS Intrction / MS Error = 10.58
A + B + AB (or explained)	SS Cells = 1135	(J * K) - 1 = 7	SS Cells / (J * K) - 1 = 162.14	MS Cells / MS Error = 8.11
Error (or residual)	SS Error = 3840	N - (J * K) = 192	SS Error / (N - J * K) = **20**	
Total	SS Total = 4975	N - 1 = 199	SS Total / (N - 1) = 25	