Sociology 592 - Research Statistics I
Exam 2 Answer key
November 9, 2001

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. Notre Dame football coach Bob Davie claims that student support for him is as strong now as it was at the beginning of the season. Highly critical sports columnists who are calling for Davie to be fired do not believe him. When the season began, 25 randomly selected students were asked whether they approved or disapproved of Davie's performance as coach. This weekend, another random sample of 25 students will be asked the same question.

2 sample tests, case V, test of p1 = p2. We want to test whether the probability of supporting Davie has declined across time. The null and alternative hypotheses are

H_0: $P_1 = P_2$
H_A: $P_1 > P_2$

The alternative is one-tailed because sportswriters claim that Davie had more support at the beginning of the season than he does now.

A Z statistic is appropriate.

b. Recent national statistics show that 43% of all first marriages break up within 15 years. A researcher wants to know whether cohabitation (i.e. living together before getting married) affects the stability of a marriage. She has identified a random sample of 28 couples married in October who cohabited beforehand. In 15 years, she will determine how many of those couples have gotten divorced.

Single sample tests, Case II, binomial parameter p. The researcher wants to see whether the probability of divorce is different for cohabiters than for the population as a whole. The null and alternative are

H_0: $P_1 = .43$
H_A: $P_1 \neq .43$

Alternative is 2-tailed because the researcher does not state whether cohabiters will have more stable or less stable marriages. A Z statistic is appropriate.

This also falls under Nonparametric Tests, Case I, sample distribution compared to population distribution. Null and alternative hypotheses are

H_0: Distribution of divorce and non-divorce is the same for cohabitors as it is for general population
H_A: Distribution of divorce and non-divorce is not the same for cohabitors as it is for general population

A chi-square statistic with one d.f. is appropriate.

c. A researcher is interested in how Earnings (High, Medium, Low) and Sex (Male, Female) are related to Attitudes Toward Abortion (Pro-life, Pro-choice, Undecided). She realizes that earnings and sex are almost certainly related to each other (i.e. men make more than women). Hence, she wants to test whether attitudes toward abortion are independent of sex and earnings, without making the dubious assumption that sex and earnings are independent of each other. She will gather data from 36 randomly selected adults in South Bend.

This falls under nonparametric tests, Case II, Tests of Association. Specifically, the researcher wants to test the model of Conditional Independence. The null and alternative hypotheses are

H_0: $P(A_i \cap B_j \cap C_k) = P(A_i \cap B_j) * P(C_k)$ for all i, j, k
H_A: $P(A_i \cap B_j \cap C_k) <> P(A_i \cap B_j) * P(C_k)$ for some i, j, k

A chi-square statistic is appropriate.
DF = rcl - 1 - (rc - 1) - (l - 1) = 3*2*3 – 1 –(3*2 – 1) – (3-1) = 18 – 1 – 5 –2 =10.

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 believes that a particular problem falls under 2 sample tests, case II, $\sigma_1 = \sigma_2 = \sigma$, σ unknown. He collects 2 samples, each of size 30, and finds that his computed T statistic equals 2.93. Another researcher argues that the problem really falls under 2 sample tests, case III, σ_1 and σ_2 unknown and not assumed equal. Using the same data as the first researcher, the T statistic the second researcher will report is 2.47.

False. Because the two samples are the same size, the value of the T statistic will be the same either way, so the second researcher will also get t = 2.93. However, the degrees of freedom and hence the acceptance regions may differ.

b. A geneticist wants to see whether a new drug has any effect on the likelihood that laboratory mice will produce male rather than female offspring. When she tests the equi-probability model on a sample of 80 births, she gets a χ^2 statistic of 5. This means that there must have been 50 male births and 30 female births.

False. You would indeed get a χ^2 of 5 if there were 50 male births and 30 female births:

Sex	**O**	**E**	**$(O – E)^2/E$**
Male	50	40	2.5
Female	30	40	2.5
χ^2			5

BUT, you would also get a χ^2 of 5 if the opposite were true and there were 30 male births and 50 female births. The chi-square test by itself does not tell you the direction of differences.

c. A researcher has collected data from 28 respondents on their education (college graduate or not college graduate), race (white or nonwhite), and feelings of personal efficacy (measured on a scale that ranges from a low of 1 to a high of 63). She computes

$$F_{(J-1)(K-1),N-JK} = \frac{SS\ Interaction/(J-1)(K-1)}{SS\ Error/(N-JK)} = 3$$

If she is using the .01 level of significance, she should conclude that neither education nor race has a significant effect on feelings of self-efficacy.

False. The above shows that there are no significant interaction effects between education and race. There could still be significant main effects. The researcher ought to be using MS Cells rather than MS Interaction in the numerator.

d. Advisors to President Bush are wondering how different segments of the country would react if the military attacks against Afghanistan were expanded to include Iraq. A Pentagon analyst believes that military veterans (group 1) are more likely to support military action against Iraq than are non-veterans (group 2). She draws a sample of 50 veterans and 200 non-veterans. She finds that

$$\hat{p}_1 = .12,\ \ \hat{p}_2 = .32$$

where $\hat{p}_i$ = the proportion of each group supporting military action against Iraq. If she is using the .01 level of significance, she should reject the null hypothesis.

False. Things have come out exactly the opposite of what the analyst predicted; veterans are less likely to support military action. The null and alternative are

H_0: P1 = P2
H_A: P1 > P2

Since P1 is less than P2 in the sample, the alternative hypothesis is clearly not supported.

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

3. ISTEP is a standardized test given to Indiana grade school students. Last year, the mean score was 33.5 with σ = 10. This year, the test was given during the week of September 11. Educators fear that the terrorist attacks adversely affected student performance, although they do not expect σ to have changed. Complete results will not be available for several weeks yet, so a random sample of 25 students has been drawn. Results show that students in the sample have an average score of 30 on the ISTEP.

a. Construct the 95% confidence interval

b. Using our 5-step hypothesis testing procedure and the .05 level of significance, determine whether the educators' concerns appear to be warranted or not.

This falls under single sample tests, case I, σ known. The critical value for Z is 1.96. The sample mean is 30, N = 25, σ = 10.

a. The 95% c.i. is

$$\bar{x} \pm (z_{\alpha/2} * \sigma/\sqrt{N}), i.e.,$$
$$30 - (1.96 * 10/5) \le \mu \le 30 + (1.96 * 10/5), i.e.,$$
$$26.08 \le \mu \le 33.92$$

b. Hypothesis testing procedure.

Step 1:

H_0: μ = 33.5
H_A: μ < 33.5

Alternative is one-tailed because Educators fear that scores will be lower because of the terrorist attacks. Hence, note that you should NOT use the confidence interval for hypothesis testing.

Step 2: The appropriate test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{N}} = \frac{\bar{x} - 33.5}{10/\sqrt{25}} = \frac{\bar{x} - 33.5}{2}$$

Step 3: Accept H_0 if Z >= -1.65

Step 4: The computed value of the test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{N}} = \frac{\bar{x} - 33.5}{10/\sqrt{25}} = \frac{\bar{x} - 33.5}{2} = \frac{30 - 33.5}{2} = -1.75$$

Step 5: Reject the Null. The Educators' concerns appear to be warranted.

4. A criminologist asks 20 blacks, 160 whites, and 20 members of other races whether they support or oppose the death penalty. She finds that 5 blacks, 70 whites, and 5 members of other races support the death penalty while the rest oppose it. Using our 5-step hypothesis testing procedure and the .05 level of significance, determine whether support for the death penalty differs by race.

This falls under Nonparametric Tests, Case II, Tests of Association. The researcher wants to test whether the model of independence holds. Note that, from the information given, we can easily construct the following table:

Race/Death Penalty	Supports Death Penalty	Opposes Death Penalty	Σ
Black	5	15	20
White	70	90	160
Other	5	15	20
Σ	80	120	200

From the above, we see that
P(Black) = .10, P(White) = .80, P(Other) = .10,
P(Support) = .40, P(Oppose) = .60

Step 1: The null and alternative hypotheses are

H_0: Blacks, whites and others do not differ in their support for the death penalty
H_A: Blacks, whites and others do differ in their support for the death penalty

or, equivalently,

H_0: $P(A_i \cap B_j) = P(A_i)P(B_j)$ for all i, j (Model of independence)
H_A: $P(A_i \cap B_j) <> P(A_i)P(B_j)$ for some i, j

Step 2: An appropriate test statistic is

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

where $E_{iJ} = P(A_i) * P(B_J) * N$

Step 3: Accept H_0 if $\boldsymbol{\chi}^2_2 \leq 5.9915$

Step 4: Test Stat. We compute

Race/Death Penalty	Observed	Expected	$(O_{ij} - E_{ij})^2/E_{ij}$
Black Supporter	5	8 = P(Black)*P(Support)*N = .10*.40*200	1.125
Black Opponent	15	12 = P(Black)*P(Oppose)*N = .10*.6*200	.75
White Supporter	70	64 = P(White)*P(Support)*N = .80*.40*200	.5625
White Opponent	90	96 = P(White)*P(Oppose)*N = .80*.6*200	.375
Other Supporter	5	8 = P(Other)*P(Support)*N = .10*.40*200	1.125
Other Opponent	15	12 = P(Other)*P(Oppose)*N = .10*.6*200	.75
χ^2			4.6875

Step 5: Do not reject the null.

Here is an SPSS solution to the above.

```
DATA LIST FREE / Race Death Freq.
BEGIN DATA.
1.00   1.00   5.00
1.00   2.00   15.00
2.00   1.00   70.00
2.00   2.00   90.00
3.00   1.00   5.00
3.00   2.00   15.00
End Data.
Weight by Freq.
Var Labels Race "Race of Respondent"/ Death "Views on Death Penalty".
Value Labels Race 1 "Black" 2 "White" 3 "Other"/
         Death 1 "Supports Death Penalty" 2 "Opposes Death Penalty".

CROSSTABS
  /TABLES=race  BY death
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ
  /CELLS= COUNT EXPECTED .
```

Crosstabs

RACE Race of Respondent * DEATH Views on Death Penalty Crosstabulation

			DEATH Views on Death Penalty		
			1.00 Supports Death Penalty	2.00 Opposes Death Penalty	Total
RACE Race of Respondent	1.00 Black	Count	5	15	20
		Expected Count	8.0	12.0	20.0
	2.00 White	Count	70	90	160
		Expected Count	64.0	96.0	160.0
	3.00 Other	Count	5	15	20
		Expected Count	8.0	12.0	20.0
Total		Count	80	120	200
		Expected Count	80.0	120.0	200.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.688[a]	2	.096
Likelihood Ratio	4.917	2	.086
Linear-by-Linear Association	.000	1	1.000
N of Valid Cases	200		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.00.

5. (25 points) At recent Congressional hearings, the American Red Cross was heavily criticized for its handling of donations after the Sept. 11 terrorist attacks. It is afraid that future contributions will suffer unless it can quickly come up with successful new fundraising techniques. In order to determine what types of strategies will be most effective, and with who, it has developed three different types of fundraising literature. This literature has been sent both to previous donors and to people who have not been donors in the past. The design is balanced and the total number of subjects in the study is 108. The amount (if any) donated by each recipient has been recorded. Complete the following F table. You do NOT need to indicate whether the F values are significant or not.

Source	SS	D.F.	M. S.	F
A + B (or Main Effects)	642			
A (Type of fundraising literature)				25
B (Past donor or not)				
AB (or 2-way interaction)				
A + B + AB (or explained)				
Error (or residual)			6	
Total	1605			

Solution. Note that N = 108, J = 3, K = 2. Once you fill in the known information and the degrees of freedom, the numbers are obtained fairly easily.

Source	SS	D.F.	Mean Square	F
A + B (or Main Effects)	SS Main = **642**	J + K - 2 = **3**	$\frac{\text{SS Main}}{(J + K - 2)}$ = **214**	$\frac{\text{MS Main}}{\text{MS Error}}$ = **35.67**
A (Type of fundraising literature)	SS Rows = **300**	J - 1 = **2**	$\frac{\text{SS Rows}}{(J - 1)}$ = **150**	$\frac{\text{MS Rows}}{\text{MS Error}}$ = **25**
B (Past donor or not)	SS Columns = **342**	K - 1 = **1**	$\frac{\text{SS Columns}}{(K - 1)}$ = **342**	$\frac{\text{MS Columns}}{\text{MS Error}}$ = **57**
AB (or 2-way interaction)	SS Intraction = **351**	(J - 1) * (K - 1) = **2**	$\frac{\text{SS Intrction}}{(J - 1)(K - 1)}$ = **175.5**	$\frac{\text{MS Intrction}}{\text{MS Error}}$ = **29.25**
A + B + AB (or explained)	SS Cells = **993**	(J * K) - 1 = **5**	$\frac{\text{SS Cells}}{(J * K) - 1}$ = **198.6**	$\frac{\text{MS Cells}}{\text{MS Error}}$ = **33.1**
Error (or residual)	SS Error = **612**	N - (J * K) = **102**	$\frac{\text{SS Error}}{(N - J * K)}$ = **6**	
Total	SS Total = **1605**	N - 1 = **107**	$\frac{\text{SS Total}}{(N - 1)}$ = **15**	