Name Key
Please ALSO put your name on the back of the last page.

Ag 400 - Quiz 3
Fall 2008

A. Let us suppose that you have collected data from a random sample of 100 Mudville residents concerning whether they are satisfied or dissatisfied with living in their local community. These data are cross-tabulated with income level of the subjects to obtain the following table.

Income	Satisfied with Local Community			Dissatisfied with Local Community		
		----------number of persons----------				
			fe		fe	
Less than $25,000	30%	6	12	14	8	20
$25,000 - $49,999	60%	36	36	24	24	60
$50,000 or more	90%	18	12	2	8	20
		60		40		100

(20) 1. Test the statistical significance of the relationship between income and whether people are satisfied with their local community. Use the .05 level to determine statistical significance. Specify the following:

a. State the Null Hypothesis and the Alternative Hypothesis in words, not statistical symbols.

H_0: there is no relationship between income and satisfaction with the local community

H_A: there is a rel. . . .

b. Show the necessary calculations.

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
6	12	-6	36	3.0
36	36	0	0	0
18	12	+6	36	3.0
14	8	+6	36	4.50
24	24	0	0	0
2	8	-6	36	4.50
				15.00 $= \chi^2$

c. Specify degrees of freedom.

2

d. Indicate the critical value of the test statistic at the .05 level.

e. Reject or do not reject the Null Hypothesis.

Reject H_0

f. What is the probability that you have made a Type I error in (e) above?

g. What is the probability that you have made a Type II error in (e) above?

h. Make your conclusion precisely in terms of the problem.

There is a relationship between income and satisfaction with the local community

Of those earning < $20,000, only 30% were satisfied. However, as income increased, the % reporting they were satisfied increased with 60% of those earning $25,000 - 49,999 being satisfied & 90% of those earning $50,000+ reporting they were satisfied

(4) 2. Calculate an appropriate measure of the strength of the relationship between income and whether or not a person is satisfied with the local community. Show your work.

$$\phi = V = \sqrt{\frac{15}{100}} = \sqrt{.15} = .387$$

(20) 3. Indicate whether each of the following statements is true (T) or false (F) in terms of these data. If any part of a statement is untrue, it should be marked false (F). Add comments to clarify your responses if you wish.

F a. The statistical unit in this problem is a dollar of income.

F b. The dependent variable, number of persons, is measured by an interval scale.

T c. The figures calculated in (A1) above are statistics because they are calculated from sample data.

F d. The above test of significance is called a one sample chi square because it utilizes data from only a single sample.

T e. If the number of cases in each of the six cells in the above table were doubled, the calculated chi square would also be doubled, but Cramer's V would remain the same.

F f. In the above sample, those with incomes of between $25,000 and $49,999 are the most likely of the three income groups to report that they are "satisfied" with the local community.

T g. For the above table, the Phi Coefficient and Cramer's V are equal.

F h. The "expected frequencies" calculated for the chi square analysis here are those frequencies we would expect in the cells if the sample was perfectly representative of the Mudville population.

F i. The Null Hypothesis for the test in (A1) above could be stated as follows:

H_o: In all three income categories, people are equally as likely to be satisfied as to be dissatisfied.

F j. Chi-square is never negative when correctly calculated, and is never greater than 1.0.

B. Suppose that Dr. Ima Knowitall asserts that three fourths of all people in Mudville who have income between $25,000 and $49,999 are "satisfied" with their local community.

(10) 1. Based on the data given in Section A above, should you doubt her claim? Use the .05 level to determine statistical significance.

a. Show your work.

	f_o	f_e	$(f_o - f_e)$	Yates $-.5$	
Satis	36	45	9	8.5	1.606
Dissatis	24	15	9	8.5	4.817
	60	60			$6.423 = \chi^2$

$df = 1$ Sig

b. Make your conclusion precisely in terms of the problem

Dr. Ima Knowitall is wrong. Fewer than 75% of those who have incomes between $25,000 + 49,999 are satisfied with their local community

(10) 2. Indicate whether each of the following statements is true (T) or false (F) in terms of these data. If any part of a statement is untrue, it should be marked false (F). Add comments, if you wish, to clarify your answers.

T a. The Null Hypothesis for the test in (1) above can be stated as follows:

H_o: Dr. Knowitall is correct.

T b. The Alternative Hypothesis for the above test could be stated as follows:

H_A: the percentage of people with incomes between $25,000 and $49,999 in Mudville who are "satisfied" with their local community is <u>not</u> 75%.

T c. It would not be appropriate to calculate a Phi Coefficient here to measure the strength of the relationship tested in (B1) above.

T d. If you reject the Null Hypothesis at the .05 level in (B1) above, you cannot make a Type II error, and the probability of making a Type I error is less than .05.

F e. Based on the above analysis you should conclude (at the .05 level of significance) that there is a relationship between the sampled population and Dr. Knowitall's assertion.

(36) C. The attached output was obtained using **graduate students only** from the STATDATA.sav file. AGE and PREVMATH were recoded to obtain the following variables:

AGE3
1 Less than 25 years
2 25 - 29 years
3 30 years and older

MATH (number of previous math courses)
0 None
1 One
2 Two or more

Answer the following questions by indicating whether the following statements are true (T) or false (F) in terms of this analysis. If any part of a statement is untrue, it should be marked false (F).

T 1. The analysis presented here is a 3x3 Contingency analysis with Chi Square.

T 2. The total number of graduate students in the data file is 417, but 4 of these are omitted from the contingency analysis because of missing data on one or both of the variables of AGE3 and MATH.

F 3. Those graduate students with zero codes for math should be deleted from this analysis of the relationship between age and number of math courses because these students had no math courses.

F 4. To obtain this output, you would click on

Analyze → Nonparametric Tests → Chi Square

and make the appropriate variable, statistics, and cell selections. Click Ok.

F 5. The Null Hypothesis for the test shown on the attached output can be stated as follows:

H_o: In the population of graduate students, there is no difference between age (as measured by these three categories) and number of previous math courses (none, one, two or more).

F 6. In the sample, 42.7% of those students less than 25 years of age had taken two or more previous math courses; 27.4% of those 30 years of age and older had done so.

F 7. A Yates Correction for Continuity should be applied in calculating Chi-Square in this analysis. The calculated Chi Square, corrected for continuity, would be 15.243 here.

T 8. The expected frequencies for each of the nine cells in this analysis are calculated by the formula:

$$expected\ frequency = \frac{(row\ total)\ (column\ total)}{Grand\ total}$$

F 9. Percentages are used to describe the relationship between AGE3 and MATH in this analysis. Chi Square can be calculated using these percentages in place of the observed frequencies.

T 10. Degrees of freedom for analysis of a 3x3 contingency table are always 4.

T 11. If you rejected the following Null Hypothesis in favor of the following Alternative Hypothesis here, $\alpha = .004$; $\beta = 0$.

H_o: In the population, there is no relationship between age and number of previous math courses.

H_A: In the population there is a relationship between age and number of previous math courses.

T 12. If you did not reject the Null Hypothesis based on this analysis, you could not make a Type I error ($\alpha = 0.00$).

T 13. Nearly half of the students 30 years of age or more had no previous math courses, while about 25% had one previous math course and 25% had two or more previous math courses.

F 14. The relationship between age and number of previous math courses as shown in this output can be described as follows:

For all three age categories, students were more likely to have no previous math courses than they were to have 2 or more previous math courses.

T 15. In this analysis, chi square is an inferential statistic;

T 16. An appropriate measure of the strength of the relationship between AGE3 and MATH here is Cramer's V.

T 17. If both Phi and Cramer's V are calculated for the same contingency table, they are *always* either equal or Cramer's V is less than Phi.

T 18. The Contingency Coefficient calculated for this analysis would equal .189.

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
MATH * AGE3	417	99.0%	4	1.0%	421	100.0%

MATH * AGE3 Crosstabulation

			AGE3			Total
			1	2	3	
MATH	0	Count	52	39	63	154
		Expected Count	58.7	48.4	46.9	154.0
		% within MATH	33.8%	25.3%	40.9%	100.0%
		% within AGE3	32.7%	29.8%	49.6%	36.9%
	1	Count	57	57	32	146
		Expected Count	55.7	45.9	44.5	146.0
		% within MATH	39.0%	39.0%	21.9%	100.0%
		% within AGE3	35.8%	43.5%	25.2%	35.0%
	2	Count	50	35	32	117
		Expected Count	44.6	36.8	35.6	117.0
		% within MATH	42.7%	29.9%	27.4%	100.0%
		% within AGE3	31.4%	26.7%	25.2%	28.1%
Total		Count	159	131	127	417
		Expected Count	159.0	131.0	127.0	417.0
		% within MATH	38.1%	31.4%	30.5%	100.0%
		% within AGE3	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15.447[a]	4	.004
Likelihood Ratio	15.243	4	.004
Linear-by-Linear Association	4.038	1	.044
N of Valid Cases	417		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 35.63.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} =$$

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.192	.004
	Cramer's V	.136	.004
N of Valid Cases		417	