Name Key
(Please ALSO write your name on the back of the last page.)

Ag 400 - Quiz 3
Makeup - Fall 2008

A. Suppose that you have surveyed a random sample of 160 Mudville residents concerning their views on a number of political issues. To ascertain whether your sample differs significantly from all Mudville voters in regard to political affiliation, you collect information from voter registrations and find that 20% of all registered voters in Mudville are Republicans, 70% are Democrats, and 10% are Independents. The distribution of political affiliation in your sample is as follows:

	Pop %	Number of Cases	Sample %
Republicans	20%	35	21.9
Democrats	70%	122	76.2
Independents	10%	3	1.9
Total		160	100.0

(22) 1. Test the statistical significance of the difference between your sample and the population of Mudville registered voters in regard to political affiliation. Use the .05 level to determine statistical significance. Specify the following:

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

H_0: the distrib of polit. affil. in the Mudville population is the same as the distrib of polit affil. in the pop. we sampled.

H_A: these two distribs differ

b. Show the necessary calculations.

	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
Rep	35	32	3	9	.281
Dem	122	112	10	100	.893
Indep	3	16	-13	169	10.562
	160		0		11.736 = χ^2

c. Specify degrees of freedom.

df = 2

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

5.99

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?

$\beta = 0$

h. Interpret the meaning of your findings precisely in terms of the problem.

The samples did not come like a simple random sample from the given population. Independents were underrepresented in the sample.

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

F a. The above test uses a 2x3 contingency chi-square to test the significance of the relationship between the sample and the given population in regard to political affiliation.

F b The Null Hypothesis for the above test could be stated as follows:

H_o: In the Mudville population, the numbers of Republicans, Democrats, and Independents in the population are equal.

F c. Chi-square can be calculated here using the percentages of Republicans, Democrats, and Independents in the sample as the "observed frequencies" and the percentages of Republicans, Democrats, and Independents in the voter registration records as the "expected frequencies."

F d. Cramer's V can correctly be used here to measure the strength of the relationship between the distribution of political affiliation in the sample and the distribution of political affiliation in the population.

F e. If all of the original data were available, including the political affiliation of each of the individuals in the sample, the difference between the sample and the total population in regard to political affiliation could have been tested for significance using a one sample t-test.

F f. The statistical unit in the above analysis is a registered voter in Mudville. The independent variable (political affiliation) is measured by a nominal scale. The dependent variable (number of registered voters) is measured by an interval scale.

B. Using data from the above sample, you compile the following table showing the relationship between political affiliation and how a person reported voting on Issue X.

Political Affiliation	Vote on Issue X					
	Yes		No		Abstain	
	----------number of voters----------					
Republicans	17	12	13	18	30	5
Democrats	43	48	77	72	120	0
~~Independents~~	5		0		0	
Total	~~65~~ 60		90		150	5

(14) 1. Test the statistical significance of the relationship between Political Affiliation and Vote on Issue X. Use the .05 level to determine statistical significance.

a. State the Null Hypothesis and the Alternative Hypothesis

H_0: there is no difference between Reps + Dems in how they voted on Issue X

H_A: there is a difference

b. Show your work.

f_o	f_e	$f_o - f_e$	$\lvert f_o - f_e \rvert - .5$	$(\lvert f_o - f_e \rvert - .5)^2 / f_e$
17	12	5	4.5	1.6875
43	48	-5	4.5	.4570
13	18	-5	4.5	1.1250
77	72	+5	4.5	.2812
				3.55

$df = 1$

b. Reject or Do Not Reject the Null Hypothesis.

Do Not Reject H_0

c. What can you conclude about the relationship between Political Affiliation and Vote on Issue X.?

Based on these data, we cannot conc. that Reps + Dems differ in how they voted on Issue X

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

T a. The statistical unit is a Mudville registered voter.

T b. The independent variable, political affiliation, is measured by a nominal scale. The dependent variable, vote on Issue X is measured by a nominal scale.

F c. The Null Hypothesis for the test shown in B1 can be stated as follows:

H_o: In the population, there is no difference between political affiliation and vote on Issue X.

F d. In the sample, Republicans were less likely than Democrats to vote "yes" on Issue X.

F e. The above analysis is a one sample chi square in which a single sample provides data for testing the relationship between political affiliation and vote on Issue X.

F f. If you reject the Null Hypothesis in B1 above, you would be wrong and you would make a Type I error.

F g. The Null Hypothesis being tested above can be stated as follows:

H_o: there is no relationship between Political Affiliation and Vote on Issue X in this sample.

T h. In a chi-square analysis:

$\Sigma fo = \Sigma fe$

F i. If all of the original data on which the above table was compiled were available, including the political affiliation and vote of each sample member, the significance of the relationship between Political Affiliation and Vote on Issue X could be tested using a t-test for the difference between two independent sample means where:

$\overline{X}_1$ = mean political affiliation

$\overline{X}_2$ = mean vote on Issue X

T j. Both Cramer's V and the Contingency Coefficient (C) calculated from the above data would be greater than 0.00 and less than 1.00.

T k. Cramer's V and the Contingency Coefficient are descriptive statistics that measure the strength of the relationship between two variables in a sample. Chi-Square is an inferential statistic.

(30) C. The attached output was obtained using SPSS and data compiled from a random sample of persons living in Littletown.

The following variables were used:

SEX
1 Male
2 Female

AGE
1 <40 years
2 40 - 64 years
3 65 and older

ANSWER to Question A
1 Agree
2 Undecided
3 Disagree

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

T 1. There are 120 Littletown residents in this sample and the data file has 120 rows.

F 2. The Null Hypothesis for testing the relationship on Page 1 of the output could be stated as follows:

H_o: the relationship between AGE and ANSWER is not statistically significant.

T 3. Based on this output, you would reject the following Null Hypothesis at the .05 level:

H_o: In the population there is no relationship between AGE and ANSWER to Question A.

T 4. If you conclude that AGE is related to ANSWER to Question A based on this analysis, $\alpha = .034$; $\beta = 0.00$

F 5. Persons who were in the age category "40-64 years" were the most likely to "Disagree" with Question A; those in the "<40 years" category were the least likely to "Disagree."

T 6. The relationship between AGE and ANSWER in this sample can be described as follows:

> The likelihood of persons "agreeing" with Question A was greatest (47.2%) for those < 40 years of age and least (12.5%) for those 65 years and older.

F 7. The relationship between AGE and ANSWER in this sample can be described as follows:

> There is no relationship between age and ANSWER for persons in the "40-64 yr." age category. For those who are less than 40 years of age, as age increases, the percentage answering "agree" declines. For those "65 years and older," as age increases, the percentage of "disagree" answers decreases.

T 8. One third of all persons in the sample in the "40-64 yr." age category answered "Agree"; one third answered "Undecided," and one third answered "Disagree" to Question A.

T 9. On PAGE1 of the output, Phi and Cramer's V have different values. You should use Cramer's V here to measure the strength of this relationship on a scale from 0 (meaning no relationship) to 1.00 (meaning a perfect relationship).

F 10. On Page 2, the relationship between SEX and ANSWER in the sample can be described as follows:

> While 65% of the males "Agree" with Question A, only 35% of the females "Agree." Also, 40% of the males "Disagree" and 60% of females "Disagreed."

F 11. There is no relationship between ANSWER and SEX in the sample because there are equal numbers of males and females and respondents were equally likely to answer "Agree," "Undecided" or "Disagree."

T 12. Chi square testing the statistical significance of the relationship between SEX and ANSWER is not statistically significant at the .05 level.

F 13. If you do not reject the following Null Hypothesis, $\alpha = 0$; $\beta = .061$.

> H_o: there is no relationship between SEX and ANSWER to Question A.

F 14. Based on the analysis on Page 2 of the output, you should conclude that, for males and females, the distributions of ANSWER are the same.

F 15. The relationship of AGE to ANSWER in this sample is stronger than the relationship of SEX to ANSWER.

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
AGE * ANSWER	120	100.0%	0	.0%	120	100.0%

AGE * ANSWER Crosstabulation

			ANSWER			Total
			1 Agree	2 Undecided	3 Disagree	
AGE	1 <40 years	Count	17	12	7	36
		Expected Count	12.0	12.0	12.0	36.0
		% within AGE	47.2%	33.3%	19.4%	100.0%
		% within ANSWER	42.5%	30.0%	17.5%	30.0%
	2 40-64 years	Count	20	20	20	60
		Expected Count	20.0	20.0	20.0	60.0
		% within AGE	33.3%	33.3%	33.3%	100.0%
		% within ANSWER	50.0%	50.0%	50.0%	50.0%
	3 65 yrs & over	Count	3	8	13	24
		Expected Count	8.0	8.0	8.0	24.0
		% within AGE	12.5%	33.3%	54.2%	100.0%
		% within ANSWER	7.5%	20.0%	32.5%	20.0%
Total		Count	40	40	40	120
		Expected Count	40.0	40.0	40.0	120.0
		% within AGE	33.3%	33.3%	33.3%	100.0%
		% within ANSWER	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.417[a]	4	.034
Likelihood Ratio	11.035	4	.026
Linear-by-Linear Association	10.119	1	.001
N of Valid Cases	120		

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

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.295	.034
	Cramer's V	.208	.034
N of Valid Cases		120	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ANSWER * SEX	120	100.0%	0	.0%	120	100.0%

ANSWER * SEX Crosstabulation

			SEX		
			1 Males	2 Females	Total
ANSWER	1 Agree	Count	26	14	40
		Expected Count	20.0	20.0	40.0
		% within ANSWER	65.0%	35.0%	100.0%
		% within SEX	43.3%	23.3%	33.3%
	2 Undecided	Count	18	22	40
		Expected Count	20.0	20.0	40.0
		% within ANSWER	45.0%	55.0%	100.0%
		% within SEX	30.0%	36.7%	33.3%
	3 Disagree	Count	16	24	40
		Expected Count	20.0	20.0	40.0
		% within ANSWER	40.0%	60.0%	100.0%
		% within SEX	26.7%	40.0%	33.3%
Total		Count	60	60	120
		Expected Count	60.0	60.0	120.0
		% within ANSWER	50.0%	50.0%	100.0%
		% within SEX	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.600[a]	2	.061
Likelihood Ratio	5.668	2	.059
Linear-by-Linear Association	4.958	1	.026
N of Valid Cases	120		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 20.00.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.216	.061
	Cramer's V	.216	.061
N of Valid Cases		120	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.