POS 3713: Homework Assignment #4
Spring 2001
Due on Friday, March 30th

Instructions: **Type** your answers to the following questions. You are permitted to do any calculations by hand on paper and attach the work to your typed responses. You should, however, report the final results of your calculations in the typed portion of your assignment.

The purpose of this assignment is to introduce you to Analysis of Variance (ANOVA) and Chi-square hypothesis tests. We will be using data from the 1996 National Election Study for this assignment.

Part A: One-Way ANOVA

One-way ANOVA allows you to test the differences between the means of ordinal or ratio level dependent variables for multiple groups. One-way ANOVA receives its name because the groups that you compare differ on a single independent variable of theoretical interest, such as religion, race, educational level, etc. For example, the table below presents the mean number of days reading the newspaper by educational level among 1711 respondents to the NES survey:

Mean Number of Days Reading News Paper by Education

< 8th Grade	9-11 Grades	H.S. Grad.	Some College	Junior degree	BA/BS Degree	Advanced Degree
2.56	3.10	3.28	3.23	3.17	3.86	4.01

ANOVA tests the null hypothesis that population means for each educational level are equal against the research hypothesis that at least one of the means is different:

H_o: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7$
H_1: The mean of at least one of the groups is different from the others, or
$\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6 \neq \mu_7$.

To determine if there is a significant difference, ANOVA compares the variance within groups to the variance between groups and computes an F-ratio test statistic based on the ratio of between group variance relative to within group variance. If the between group variance is large enough relative to the within group variance, then the F-ratio will exceed the critical level for $\alpha = .05$ (95% confidence), and you can reject the null hypothesis of equal means. The F-ratio for the above example equals 4.424, which exceeds the critical F value of approximately 2.09 for d.f. (within) = 1710, and d.f. (between) = 6. The degrees of freedom (within) is calculated as N - k, while the degrees of freedom (between) is calculated as k - 1, where k is the number of groups you are comparing. Thus, you can reject the null hypothesis and conclude that at least one of the educational levels has a mean significantly different from the others.

The goal of this exercise is to determine if there is a significant difference in presidential approval by educational level and party identification. We will be measuring presidential approval by using the Clinton feeling thermometer (described in assignment #2). Education level is measured on an ordinal scale that takes on values from 1 (8th grade or less education) to 7 (advanced degree). Party identification is measured on an ordinal scale as well, ranging from 0 (strong democrat) to 6 (strong republican). The descriptive statistics and ANOVA output are presented below; the dependent variable in both cases is the Clinton feeling thermometer.

Descriptives and ANOVA Output for the Clinton feeling thermometer based on Education Levels

Descriptives

Clinton thermometer

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1. 8 grades or less and no diploma or eq	71	69.44	27.59	3.27	62.91	75.97	0	100
2. 9-11 grades, no further schooling (in	154	68.08	28.65	2.31	63.52	72.65	0	100
3. High school diploma or equivalency te	545	61.61	29.94	1.28	59.09	64.13	0	100
4. More than 12 years of schooling, no h	307	56.01	28.59	1.63	52.80	59.22	0	100
5. Junior or community college level deg	155	59.48	29.37	2.36	54.82	64.14	0	100
6. BA level degrees; 17+ years, no postg	306	53.48	29.37	1.68	50.18	56.79	0	100
7. Advanced degree, including LLB [13-1	164	56.54	28.95	2.26	52.07	61.00	0	100
Total	1702	59.37	29.56	.72	57.96	60.77	0	100

ANOVA

Clinton thermometer

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	36994.35	6	6165.725	7.211	.000
Within Groups	1449399	1695	855.103		
Total	1486394	1701			

Descriptives and ANOVA Output for the Clinton feeling thermometer based on Party Identification

Descriptives

Clinton thermometer

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
0. Strong Democrat (1,1,0 in K1, K1a/b,	326	85.56	14.35	.79	84.00	87.12	0	100
1. Weak Democrat (1,5/8/9,0 in K1, K1a/b	331	73.20	17.80	.98	71.27	75.12	0	100
2. Independent-Democrat (3/4/5,0,5 in K1	233	69.17	19.46	1.27	66.66	71.68	0	100
3. Independent-Independent (3,0,3/8/9 in	144	56.79	24.01	2.00	52.84	60.75	0	100
4. Independent-Republican (3/4/5,0,1 in	183	41.78	28.23	2.09	37.66	45.89	0	100
5. Weak Republican (2,5/8/9,0 in K1, K1a	257	43.67	26.24	1.64	40.44	46.89	0	100
6. Strong Republican (2,1,0 in K1, K1a/b	213	23.20	22.75	1.56	20.12	26.27	0	100
Total	1687	59.41	29.64	.72	57.99	60.83	0	100

ANOVA

Clinton thermometer

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	708993.9	6	118165.646	256.918	.000
Within Groups	772692.1	1680	459.936		
Total	1481686	1686			

Question A1: State the null and alternative hypotheses for each case (for the analysis of approval by education levels and approval by party identification).
Question A2: What are the mean square between and the mean square within for each analysis? How are these numbers used to construct the F-ratio test? Based on these numbers, which analysis do you predict will produce a statistically significant F-ratio test?
Question A3: Interpret the output for each analysis. Is there a significant difference in support for Clinton according to educational level, or according to party identification? What are the theoretical explanations for the differences or lack of differences? Which independent variable has a more powerful influence on Clinton support? Be sure to discuss the specific numbers in the tables as evidence for your answers to these questions.

Part B: Chi-Square Analysis

Chi-square tests are used to test the null hypothesis that two variables are statistically independent. Two variables are independent if the classification of a case into a particular category of one variable has no effect on the probability that a case will fall into any particular category of another variable. Chi-square tests are computed by analyzing a bivariate table (usually called a “contingency table” or a “cross-tabulation”), and comparing the observed frequencies of cases in each cell to the expected frequencies under the assumption that the variables are independent. For example, the table below shows the observed and expected frequencies (expected values in parentheses) and column percentages for a cross-tabulation of employment status and college accreditation status for a sample of 100 social work majors:

Cross-tabulation of Employment and Accreditation Status

	Accreditation Status		
Employment Status	Accredited	Not Accredited	Totals
Working as a social worker	30 (22) 54.5% (40%)	10 (18) 22% (40%)	40 40%
Not working as a social worker	25 (33) 45% (60%)	35 (27) 77% (60%)	60 60%
Totals	55	45	100

As can be clearly seen, the observed percentage of social working majors actually working as a social worker is much lower than would be expected in non-accredited colleges and much higher than would be expected in accredited colleges, under the assumption of independence. Based on these calculations, you obtain a χ^2 (obtained)= 10.79, which exceeds the critical value of 3.84 with 1 d.f. and a 95% confidence level. Hence, you can reject the null hypothesis of independence and conclude there is some causal relationship between the two variables. In this case, you might conclude that accredited colleges are doing a much better job of producing people with marketable social work skills.

For this exercise, you will use the 1996 NES data to see if party identification and educational level are statistically independent, or if there is some relationship between these two variables. You should think about party identification as the dependent variable and education as the independent variable in this problem.

The first table below is the crosstabulation of education by party identification. In this problem, we are using the nominal version of party identification (rather than the ordinal measure used above), which simply classifies respondents as Democrats, Republicans, or Independents. The second table contains the calculated “Pearson’s Chi-square”; you can ignore the other information in the table.

EDUCATION * PARTY ID Crosstabulation

			PARTY ID			
			1. Democrat	2. Republican	3. Independent	Total
EDUCATION	1. 8 grades or less and no diploma or eq	Count	39	7	17	63
		Expected Count	26.4	18.7	17.8	63.0
		% within PARTY ID	5.9%	1.5%	3.8%	4.0%
	2. 9-11 grades, no further schooling (in	Count	71	25	49	145
		Expected Count	60.9	43.1	41.0	145.0
		% within PARTY ID	10.7%	5.3%	11.0%	9.2%
	3. High school diploma or equivalency te	Count	232	128	131	491
		Expected Count	206.1	146.0	138.9	491.0
		% within PARTY ID	35.0%	27.3%	29.4%	31.1%
	4. More than 12 years of schooling, no h	Count	106	99	77	282
		Expected Count	118.4	83.9	79.8	282.0
		% within PARTY ID	16.0%	21.1%	17.3%	17.9%
	5. Junior or community college level deg	Count	63	40	48	151
		Expected Count	63.4	44.9	42.7	151.0
		% within PARTY ID	9.5%	8.5%	10.8%	9.6%
	6. BA level degrees; 17+ years, no postg	Count	95	118	75	288
		Expected Count	120.9	85.7	81.5	288.0
		% within PARTY ID	14.4%	25.2%	16.8%	18.3%
	7. Advanced degree, including LLB [13-1	Count	56	52	49	157
		Expected Count	65.9	46.7	44.4	157.0
		% within PARTY ID	8.5%	11.1%	11.0%	10.0%
Total		Count	662	469	446	1577
		Expected Count	662.0	469.0	446.0	1577.0
		% within PARTY ID	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	56.289[a]	12	.000
Likelihood Ratio	58.124	12	.000
Linear-by-Linear Association	9.345	1	.002
N of Valid Cases	1577		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 17.82.

<u>Question B1:</u> Examine the observed and expected frequencies, the row percentages, and the Chi-square test. Based on these numbers, do you think education and party identification are independent? In other words, does there seem to be a large difference between the actual count and expected counts in the first table?

<u>Question B2:</u> Does the Chi-square test allow you to reject the null hypothesis of independence? If you do reject the null, what appears to be the relationship between education and party identification, and what might theoretically account for that relationship? Make direct references to the numbers in the table when discussing your interpretation.