POS 3713: Assignment 5
Assigned: Monday, 3/20/2000
Due date: In class, Monday, 3/27/2000
Tutorial Session: Thursday, 3/23/2000, 9am-10:45am; and Friday, 3/24/2000, 9am–10:45am
Professors Mitchell and Lubell

The purpose of this assignment is to introduce you to Analysis of Variance (ANOVA) and Chi-square hypothesis tests. We will be using the 1996 National Election Study for this exercise. To begin the assignment, open your saved NES file from your previous assignments. If you do not have a saved file, you will have to begin from scratch by downloading the original NES dataset from the Internet as detailed in the instructions to Assignment 1. Remember to fully answer all questions (typed) and include all relevant output in your final product.

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. SPSS generates a F-ratio for the above example equal to 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 (V960610) and party identification (V960420). The first step is to rename V960610 = *Educate* and V960420 = *PartyID2*. Also, recode the single minor party (value=7) respondent in *PartyID2* as system-missing. The main dependent variable for this analysis will be the *ClinTher* (V960272) from previous exercises (if you are using a raw dataset, rename the variable appropriately). Follow these steps to conduct the ANOVA test for *Educate* and *PartyID2*:

- Select "Analyze", "Compare Means", "One-way ANOVA" from the menu. This will open up the ANOVA interface.

- Click the options button, and then place a checkmark in the “Descriptive” box, and click “Continue”. This tells SPSS to produce a table of descriptive statistics by group, so you can better understand the substantive results.
- Move *ClinTher* into the dependent variable text box using the arrow button. Move *Educate* into the factor text box using the arrow button. You have now defined your dependent and independent variables. Click “OK” to execute the ANOVA.
- For *PartyID2*, repeat the same exact steps, but replace *Educate* with *PartyID* in the factor text box.

Exercise A: Use the information calculated above to answer the following questions:

1) Copy the descriptive statistics and ANOVA tables from each analysis into your Word file using CTRL-K (copy objects) and CTRL-V (paste).
2) State the null and alternative hypotheses for each case.
3) 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?
4) 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	30 (22)	10 (18)	40
worker	54.5% (40%)	22% (40%)	40%
Not working as a social	25 (33)	35 (27)	60
worker	45% (60%)	77% (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 NES data to see if party identification (V960417) and educational (V960610) level are statistically independent, or if there is some relationship between these two variables. You already renamed V960610= *Educate* in Part A of this assignment, and if you are using a prior dataset you will have already renamed V960417= *PartyID* (note the difference between *PartyID* and *PartyID2*!) and recoded the basic party identification variables to include only Republicans, Democrats, and Independents. Check the frequency distributions for *PartyID* to make sure you have the right values; if you don't code all values other than Republicans, Democrats, and Independents (4-9) as system-missing. Follow these steps to compute the chi-square test:

- Select "Analyze", "Descriptive Statistics", "Cross-tabs" from the menu. This will access the cross-tab interface.
- Click on the "Statistics" button, enter a check next to "Chi-square", and click "Continue." Click the "Cells" button, enter checks next to "Observed Frequencies", "Expected Frequencies", and "Column Percentages", and click "Continue".
- Enter *Educate* into the "Rows" text box and *PartyID* into the "Columns" text box using the arrow button. Click "OK" to run the analysis. Note that we are thinking of *Educate* as the independent variable and *PartyID* as the dependent variable. This violates the "convention" mentioned by Healey concerning the placement of independent variables in the columns and the dependent in the row, but makes for a neater table in this case.

<u>Exercise B</u>: Use the information calculated above to answer the following questions:

1) Copy the descriptives and Chi-square table into your Word file, or attach the SPSS output. The relevant row in the Chi-square table for this exercise is labeled "Pearson's Chi-square"; you can ignore the other rows.
2) 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? 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.