**Department of Sociology and Anthropology**

**Carleton College**

Explorations in Social Data Analysis Peter D. Brandon

Soan 239 Leighton Hall 229

Spring Term 2011 Phone: 222-7199

 Email: pbrandon@carleton.edu

Assignments are intended to help you gain firsthand experience using statistical methods that are applicable to understanding social data. Also, assignments give you practice using STATA. Note that one or more of an assignment’s questions will appear on the following week’s Friday examination.

**ASSIGNMENT 1:**

**A & F: Chapter 1**

1. **Complete exercises 1.2, 1.4, 1.6, 1.8, 1.10, 1.14, 1.16**
2. **Complete the following exercise:**

Open the “AidsinAustralia.dta” data and tell me:

* 1. How many variables are in the data set?
	2. How many observations are in the data set?
	3. How do you think this data was collected?
	4. Why would you collect such data? Why might it prove useful?

**ASSIGNMENT 2:**

**A & F: Chapter 2**

1. **Complete exercises 2.2, 2.4, 2.6, 2.8, 2.10, 2.14, 2.16**
2. **Complete the following exercise:**

Open the “Recid.dta” data and tell me:

* 1. Which variables are
		1. Quantitative
		2. Categorical
		3. Discrete
		4. Continuous
	2. Is this an observation study
	3. Describe the sample based on what your learned in Chapter 2.
	4. True or false: The nominal variables could be transformed into continuous variables.

**ASSIGNMENT 3:**

**A & F: Chapter 3**

1. **Complete exercises 3.1 – 3.9, 3.15 – 3.26**
2. **Complete the following exercise:**

Open the “Labor unions in America, CPS88.dta“ data, pick four variables, and complete the following:

* 1. Tell me their levels of measurement.
	2. Graph the four variables carefully.
	3. Give descriptive statistics for the 4 variables and frequency tabulations.
	4. Prove me with the mean, mode, and median for the 4 variables. Are they the same?

**ASSIGNMENT 4:**

**A & F: Chapter 4**

1. **Complete exercises 4.1 – 4.20, 4.47, 4.49 – 4.53**
2. **Complete the following exercise:**

Once again open the “Labor unions in America, CPS88.dta“ data, pick two continuous variables, and complete the following:

* 1. Graph the variables and describe the distribution of the variables.
	2. Using STATA, computer the standard error of the means of the two variables.
	3. In sequence, take 10, 20, and 40 percent samples from the data and repeat parts (a) and (b) for your two continuous variables. (HINT: There is a STATA command with documentation for taking random samples from data.)
	4. Present your results of (b) and (c) in a table.

**ASSIGNMENT 5**

**A & F: Chapter 5**

1. **Complete exercises 5.1 – 5.15, 5.25, 5.49 – 5.52**
2. **Complete the following exercise:**

Open the data on mothers’ allocations of time to household work, “Timeuse.dta“, pick five variables, and complete the following:

* 1. Graph the variables, describe the distribution of the variables, and tell me their levels of measurement.
	2. For the 5 variables you have chosen given me their point estimates and discuss what these estimates infer.
	3. Provide 90, 95, and 99 percent confidence intervals for the 5 variables. Interpret your results based upon what you have learned in Chapter 5 of Agresti.
	4. Tell me one way to decrease the outer bounds of the 99 percent confidence interval?
	5. Create one-sided confidence intervals for 2 of your 5 variables and justify the choice of side.

**ASSIGNMENT 6**

**A & F: Chapter 6**

1. **Complete exercises 6.1 – 6.22, 6.37, 6.41 – 6.49**
2. **Complete the following exercise:**

Open the data on child care, “Childcare.dta“, chose 3 variables, and complete the following:

* 1. Graph the variables, describe their distributions, and tell me their levels of measurement.
	2. For the 3 variables you have chosen given me the point estimates and discuss what these estimates infer.
	3. Provide 90, 95, and 99 percent confidence intervals for the variables. Interpret your results based upon what you have learned in Chapter 5 of Agresti.
	4. For the three variables, take the means, multiply each of the means by 0.33 to get three new values. Create hypothesis tests based on those new values by pretending that they are the mean values and determine if the new values are plausible means for the population under study based on the significance test techniques discussed in Chapter 6. Use STATA to test your hypotheses since it will provide *p*-values. All work can be done in STATA, except for your final judgment on the plausibility of the fictitious mean values.

**ASSIGNMENT 7**

**A & F: Chapter 7**

1. **Complete exercises 7.1 – 7.20, 7.32 – 7.38**
2. **Complete the following exercise:**

Open the data on women and work, “Womenandwork.dta“. Complete the following:

* 1. Create a table showing presence or absence of children and whether mothers’ worked.
	2. Create a table showing presence or absence of children and average number of maternal hours worked.
	3. Create box plots displaying presence or absence of children and number of maternal hours worked.
	4. Test whether working mothers have fewer children than non-working mothers.
	5. Create a table showing high school education or less and above a high school education and wives’ average hourly earnings (in 1975 dollars).
	6. Create box plots displaying high school education or less and above a high school education and wives’ average hourly earnings (in 1975 dollars).
	7. Test whether less-educated wives earn significantly less than more educated wives.

**ASSIGNMENT 8**

**A & F: Chapter 8**

1. **Complete exercises 8.1, 8.5, 8.7 – 8.20, 8.29 – 8.33**
2. **Complete the following exercise:**

Open the data on indigenous Australians, “aborginaldeathsincustody.dta“, and complete the following:

* 1. Graph the variables, describe the distribution of the variables, and tell me their levels of measurement.
	2. Compare the rates of deaths in custody between the general population and that of aboriginals.
	3. Construct a table based on (b) and reflecting what you have learned in A & F.
	4. Do aboriginal people in custody die at a higher rate than non-aboriginal people?
	5. What conclusions can you draw from your analyses?

**ASSIGNMENT 9**

**A & F: Chapter 12—Part I**

1. **Complete exercises 12.1 – 12.9**
2. **Complete the following exercise:**

Open the data on child care, “Childcare.dta”, and complete the following:

* 1. Is the average price of child care, (avccpce), equal across regions?
	2. Specify the null and alternative hypotheses for (a).
	3. Is the evidence strong or weak for your conclusion? Explain. What significance level is used and why?
	4. Discuss your test statistics and your relevant degrees of freedom (df).
	5. For questions (a) – (d) STATA provides a table of output. Please explain these aspects of the table, including showing how the numbers are derived: Number of observations, Partial Sum of Squares, Root MSE, Residual Sum of Squares, Model Sum of Squares, Total Sum of Squares, MS, and the *p*-value for the *F* statistic.
	6. Is family income, (faminc), equal across regions?
	7. Specify the null and alternative hypotheses for (f).
	8. Is the evidence strong or weak for your conclusion? Explain. What level of significance is use and why?
	9. Discuss your test statistics and your relevant degrees of freedom (df).
	10. For questions (f) – (i) STATA provides a table of output. Please explain these aspects of the table, including showing how the numbers are derived: Number of observations, Partial Sum of Squares, Root MSE, Residual Sum of Squares, Model Sum of Squares, Total Sum of Squares, MS, and the *p*-value for the *F* statistic.

**ASSIGNMENT 10**

**A & F: Chapter 12—Part II**

1. **Complete exercises 12.28, 12.30, 12.31, 12.35, 12.38, 12.43**
2. **Complete the following exercise:**

Open the data, “CPS78.dta” and complete the following:

* 1. Wages might depend on gender and marriage. Construct an ANOVA table and conduct F-tests for the presence of nonzero gender effects, marriage effects, and interaction effects. Repeat this for p-values level of .001, .05, and .10. Explain and interpret your results.
	2. Wages might depend on gender and marriage and union status. Construct an ANOVA table and conduct F-tests for the presence of nonzero gender effects, marriage effects, and union effects as well as for interaction effects between marriage and gender and gender and union status. Explain and interpret your results.
	3. Wages might NOT depend on marriage, but on gender, race, and union status. Construct an ANOVA table and conduct F-tests for the presence of nonzero gender effects, race effects, and union effects as well as for interaction effects between race and gender and race and union status. Explain and interpret your results.
	4. Based on your findings reported for questions (a), (b), and (c), what can you infer about the determinants of wages without accounting for education levels and other factors.