**CJFS 3710 Statistics Dataset iProject Part 1: Problem Selection**

(50 points possible, due by 1:30pm 5/24/2017) HO #5

Summer 2017

Student Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. State a specific problem (i.e. unanswered open-ended question) you’d like to pursue with respect to the assigned individual project in this class. Ultimately for part 2 due in November, your problem will include at least 5 concepts that relate to something involving crime within the http://[www.city-data.com](http://www.city-data.com), http://[www.areaconnect.com](http://www.areaconnect.com), http://[www.bestplaces.net](http://www.bestplaces.net), <http://nces.ed.gov/collegenavigator> , <http://www.clrsearch.com> , <http://bjs.ojp.usdoj.gov/index.cfm?ty=daa> or other websites with instructor approval. Criteria specified in HO#6 detail what you will need to do for the Dataset iProject part 2.

Note: Concepts are abstract variables, they define an issue/topic. A variable is something measurable that has more than one category. The UCR is a variable for a city. Each of the 8 crimes within the UCR are variables too, with each of their categories being either the number of crimes in a city or the number of crimes per 100,000 people in a city. Other variable examples include: percent of people unemployed, percent of people living in poverty, median household income, median age, average temperature in July, number of full-time law enforcement officers per 1,000 population….Examples of questions are: how do economics influence crime?, how does weather influence crime?, what factors matter when looking at motor vehicle theft? how do sunbelt cities compare to rustbelt cities in terms of crime? how do certain campus characteristics impact crime on campus?...

2) In one short, simple sentence, describe one hypothesis that relates to your problem in #1. Hypotheses are proposed answers to problems. In short, hypotheses are the bones of a theory concerning a problem. Take whatever question you pose in #1 and provide a short answer to it.

3) Write a 2-paragraph newspaper-editorial-style-article (like those in the Cleveland Plain Dealer or Youngstown Vindicator or Mansfield News Journal) that summarizes your views on your problem. In writing this editorial, keep in mind your audience and the need for you to both “sell” the newspaper and to do justice to your problem. You are shooting straight from the hip here, no filtering needed, let it fly!

4) Ask 3 people you know quite well to read this editorial and write down their feedback in 3 sentences total. Simply ask them to describe what they thought of your editorial. The reason for this part is for you to receive some immediate critical thinking on your problem. This rationale is related to the 3rd leg of science (application) and that whatever it is we learn through using theory and doing research should be shared with others in order to advance science. Do not specify any names of the 3 people you polled.

5) Use three sentences to describe why you picked this problem and its importance.

6) Reference three prior works on your selected problem area. Use <http://www.citationmachine.net> for help with citing any reference material in terms of content and format (choose American Psychological Association, APA, style or see <http://www.apastyle.org> ). These three “references” can be either virtual or paper resource. Under each citation, provide one sentence description of its content in your own words.

**CJFS 3710 Statistics Dataset Project Part 2 (200 points, due by 1:30pm 6/14/2017) HO#6**

Summer 2017

1. Describe 5 variables related to your problem, include their units, dates, and categories/codes

1 dichotomous

1 nominal with 3 or more categories

1 ordinal

2 interval

2. Enter 20 cases of info for these 5 variables from any of public-use/accessible website or from another source with professor permission only

 Specify the source:

3. Create and display frequency distributions for the dichotomous, nominal, and ordinal variables

4. Calculate the most appropriate central tendency and dispersion values for the ordinal & interval variables only

5. Interpret the central tendency and dispersion calculations in #4 by doing the following:

a. describe the two central tendency values and what they mean

b. describe the two dispersion values and what they mean as well

6. Create an appropriate graph for each of the 5 variables and adhere to the graphic advice

(meaningful/enticing title, number of cases used in the analysis specified, dates, units, categories, comparison groups all clearly marked, source provided, right graph used-pie for dichotomous, nominal, and ordinal—not for interval!...for interval variables, graph min, max, median, mean, and standard deviation)

 Dichotomous variable graph

 Nominal variable graph Ordinal variable graph

 Interval one variable graph Interval two variable graph

7. Compare two groups on any one interval variable and show the results, hint use “=ttest(range,range,2,2)” where “range” specifies where the data are located

8. Comparison Interpretation--describe the two groups being compared in #7, specify the means of the interval variable being used to compare the two groups, specify the p-value calculated, indicate the alpha and whether or not the p-value calculated was less than or greater than the alpha, and provide an overall decision on whether a difference exists between the two groups

9. Test to see if 2 interval variables relate to one another and show the Pearson’s r correlation coefficient, hint use “=correl(range,range)” where “range” tells where the data are located

10. Correlation Interpretation--describe the two interval level variables being tested for a connection or pattern in #9, specify what the Pearson's r correlation coefficient is between these two variables, indicate if a correlation exists, describe how large the correlation is, and describe the direction of the connection