Paper Assignment Statistics in Health Care Research Fall 2020 Dr. David Phillippi

**Paper due:** 12/7/2020 (Midnight)

Requirements:

**Research Question:** You need to identify a question on a topic of interest that you will seek to answer. **Although this is very much backwards from how it should be done**…I strongly suggest that you choose your question based on data that you know is freely available on the internet (in other words…find the data first and then phrase your question).

It should be possible (if the data meet the assumptions) to answer your research question using one of the following tests: Independent Samples t-Test, One-Way ANOVA, Chi-Square, Paired t-Test, or Repeated Measures ANOVA (see me for possible exceptions involving correlation).

**Data Requirements:** You will collect (and turn in) a data set in SPSS. The data should consist of (at least) 3 columns of information:

1. A variable with a list of the sample elements. Your sample could consist of U.S. states, years, teams or even individual players in a sport.
2. Two variables to compare. Each variable will need to consist of measurements or properties of the items in your sample. Your independent variable will need to be a categorical (grouping) variable in order to run any of the tests mentioned. Your dependent variable could be either categorical (so using a Chi-Square) or scale (using t-Tests or ANOVAs).

For example, my sample might be U.S. states (so that my first column would be a list of states. My independent variable could be the region of the country that state is in. My dependent variable might be the percent of the population below the poverty line (scale level). My research question would then be: Is there an association between the state-level poverty rate and the regions of the country that state is in? This question can be assessed by an ANOVA.

As another example, suppose my sample consists of pitchers in major league baseball (and my first column would contain the names of each pitcher. My independent variable could be whether they are left-handed or right-handed pitchers and the dependent variable could be whether they were a starting pitcher or a relief pitcher. My research question would then be: Are left-handed pitchers more or less likely to be starting pitchers? This question can be assessed using a Chi-Square test.

Here are a few links to sources of data, but feel free to google search your topic of interest.

Demographic data: [Census Data](https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t#none)

Sports Data: [Women's Basketball](http://www.espn.com/womens-college-basketball/statistics), [Cricket](https://www.bbc.com/sport/cricket/world-cup/averages), [Boston Marathon](http://registration.baa.org/cfm_Archive/iframe_ArchiveSearch.cfm?mode=results&RequestTimeout=600&snap=96255445&)

Financial Data: [S and P 500 Stock Index](https://finance.yahoo.com/quote/%5EGSPC/history/), [Gold Price](https://www.macrotrends.net/1333/historical-gold-prices-100-year-chart), [Inflation and Consumer Price Index](https://www.bls.gov/data/)

**Analysis Requirements:** You will need a copy of your data that SPSS can use. However, it may be convenient to create a copy of the data and “clean” it in Excel and copy this to SPSS once it is ready to be analyzed. Check the websites to see if it is possible to download your data (note that a comma delimited or CVC file can be understood by Excel). If not, it may be possible to copy and paste tables into Excel. You may need to then delete some rows and columns to get the sample and variables you need. Your final data should contain at least 3 columns. One column will list the items in the sample, while the other two will contain the values for the variable.

Once your data is ready you will need to use SPSS to run the statistical test. One of our “activity” hours will be used for this purpose. ***The data file and output file from your analysis will be submitted as part of the project.***

**Paper requirements:** Your paper should be a two to three page analysis of your results. I want you to follow the ten step study plan (summarized in the box on page 8 of the text and described in detail in Chapter 1) but you do not need to address steps 2 or 10. For the other eight steps, you should address these using appropriate statistics terminology and conventions (don’t forget the golden takeaways). Pay particular attention to:

***#7 Description of the sample and how it was obtained.*** This section may need to be a bit longer as you will need to describe the steps you took to acquire and adjust your data including (in broad terms) any deletions made to the data. This section should also describe your sources of data. I would appreciate the web address to be included, but I’m not requiring a formal “References” section.

You will be graded on the accuracy of your process and calculations as well as the descriptions of your results.