

**CMPSC 390
Data Analytics
Fall 2017**

**Lab 7: Statistical Analysis of Psychological Data
Using R and the R *Psych* package
Save this lab assignment to: labs/lab7**

Objectives

To explore statistical tools which are relevant for the evaluation of psychological data. In particular, to be able to research how to use new R-statistics software packages and apply them to particular contexts for which they were designed. To extract knowledge from the produced visualizations and extracted interpretation of results.

Reading Assignment

Please review the technical manual for the *Psych* package for use with R. It can be found at the following link: <http://personality-project.org/r/overview.pdf>. This document will be used to guide your coding for testing and visualization.

Groupwork

You are to work in a group of not more than three people for this lab. Be sure to discuss each of the tasks and proceed after the group has come to a complete agreement. **Each person is to turn in his or her own report and code, however all lab partners should be listed in the submission.**

Psychology Data Analysis

In this lab you will utilize the *Psych* package to analyze data from the psychology discipline. In particular your deliverables will be composed of the following items.

1. **Obtain your data:** You may obtain your data from any online source as long as it is a credible source and that the data stems from the psychology discipline. For an idea, you could select the data available by the Cornell University website: <https://www.ciser.cornell.edu/ASPs/datasource.asp>.
2. **Description of Data:** You are to write a short report to describe your data. Discuss what the data contains and its purpose (i.e., why was this data collected, for what purpose?). You may need to look this information up online from sources other than the one where you found it.
3. **Tools and Software:** Using your credible data, you are to use the *psych* package to perform a correlation analysis over variables that you will study in connection to your research questions.

4. **Five Answerable Questions:** Determine some four or five basic questions which you are able to answer from your data. You will use programming from the *psych* package to respond to these questions.
5. **Analytics:** To respond to your questions, after you have run a correlation analysis, you are to implement at least two t.tests and two linear regression analyses tests. Please provide visualizations to help describe and explain the results of these tests. For example, plot the points of your linear regression model, along with the mean and regression lines to explain your model's performance and result. In addition, you are to provide a brief summary, included in your report, of your interpretation of the correlation analysis and tests.
6. **Summary:** Your summary of each test is to include discussion of its *p*-value. Be sure to explain how this value is useful for understanding the test's result.

Important Details

Lab directory structure: Make sure you have placed your submission materials for this lab into labs/lab7 directory in your Bitbucket repository (cs390f2017-billb).

Note: Please remember to include your name on everything you submit for the class. Although the instructor collects your work from Bitbucket, each work must be graded outside of the Bitbucket directory and so without adding your name, the instructor will be unable to award you credit for your work.

Required Deliverables

By the 10th November at lab time (2:30pm), submit electronic versions of the following deliverables through your Bitbucket repository (cs390f2017-billb) by correctly using using appropriate Git commands, such as `git add -A`, `git commit -m "your message"` and `git push`. When you have finished, please ensure that the Bitbucket Web site has your pushed work. Please contact the instructor if you have any questions about assignment submission.

1. The R source code that you used to answer your questions.
2. Report that includes the questions, the data descriptions, exploratory questions, correlation analysis, at least two t.tests and at least two linear regression tests. Write your report so that it is clear which parts of your code and discussion addresses what particular question. In addition, this report is to contain the visualizations and your brief written evaluation after each to describe the knowledge that you have gained from each visualizations and /or test, in light of the *p*-values.

Your report document should be in an open office format and your visualizations should be included directly into your report document.

3. You are also to include a reflection portion to your report document where you describe how data analysis research is different (likely different) between psychology and political science. For instance, you can describe how the software tools and packages may differ between both disciplines. Follow your notes from the talk given by Dr. Lydia Jackson and include insights from her talk in your reflection document.