

**CMPSC 301
Data Analytics
Fall 2020**

**Lab 5: Statistical Analysis of Psychological Data
Using R and the R *Psych* package**

Objectives

To determine a how to use correlations, *t*-tests and linear models during research. To gain experience using the computational methods from the R *Psych* library package. 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 pairs – not more than two people for this lab. Be sure to discuss each of the tasks and proceed after you and your partner 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.**

GitHub Starter Link for Groups

STOP! STOP!

Not everyone will be clicking this link at this time!

Only the team leader will be clicking the link to create the repository!!

<https://classroom.github.com/g/tXHTeu2S>

Creating Your Repository

We will use a group assignment functionality of GitHub Classroom for this assignment. For group assignments **only one person will be creating the team while the other team member will join that team.** Please form a team of **no more than two people** and select one person to create the repository.

The selected person of the team should go into the link to the lab in the assignment sheet. Copy this link and paste it into your web browser. Now, you should accept the laboratory assignment and create a new team with a unique and descriptive team name (under “Or Create a new team”).

Now the other members of the team can click on the assignment link and select their team from the list under “Join an Existing Team”. When other team members join their group in GitHub Classroom, a team is created in our GitHub organization. Every team member will be able to push and pull to their team’s repository.

To use this link, please follow the steps below.

- Click on the link and accept the assignment
- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this lab,
- Clone this repository (bearing your name) and work locally
- As you are working on your lab, you are to commit and push regularly. The commands are the following.

```
– git add -A
– git commit -m ‘‘Your notes about commit here’’
– git push
```

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.

Obtaining Data

- **Obtain Your Data Set:** 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 one of the data sets available on the *Openpsychometrics* website: https://openpsychometrics.org/_rawdata/ concerning personality types (Big Five Personality Test), Anxiety (Taylor Manifest Anxiety Scale), as well as other types of data stemming from psychological inquiry.
- **Describe Your Data:** You are to write a short report to describe your data. In your own words, discuss what the data is about: what it contains (what it measures) and why it was collected. Be sure to locate the data’s *code-book* to help you learn what the columns of data contain and to provide you with extra data about the set in general such as how the data was collected, from whom, and similar issues. In psychology, data sets are likely to have an associated article in which the data was featured. Look for articles for this extra information and use them if they exist.

Method Research

Often in data analytics work, you will have to read library documentation to determine how to use particular methods for your work. In this section, you will spend some time studying the *Psych* library documentation (link: <http://personality-project.org/r/overview.pdf>) to learn how to run types of tests.

1. **Choose Six Columns From Your Data Set:** Perform correlation analysis over six columns of data using plots and textual outputs. Explain how to understand these results. Can you pick other columns which have stronger correlations to answer this particular problem?

Correlations: According to the documentation, what are correlations between two columns of data (of your chosen set) and how should one interpret the output score? What are *high correlations* and what are *low-correlations*? You could also investigate using `pairs.panels()` to run some of your tests and you are to how to interpret these plots, if you employ them.

Justify Your Correlation Results: Once you have found correlations in your data set, try to justify why the correlation exists. Remember that the data will not answer the *why* part of your work and so you may need to investigate academic articles to provide grounded theories to explain your results. You and your group are to use logical reasoning to give plausible reasoning to provide explanations about these correlation(s). Note: If no correlations were found in your data, then you could go back to pick new columns to run the test again or, you could explain why you think that correlations are not present.

2. Statistical tests: You are to use the `t.test` or `lm` (linear model regression) to provide insight into your analysis of correlations from above.

t-Tests: For correlating variables, write `t-test` code in R to determine results of this test.

Linear Models: Now, for the variables that have correlations, try creating linear models containing the variables to provide insight into your analysis. What do you see? You are invited to use your slides, online resources or RStudio's online help to get you started (i.e., type `?lm` and `?t.test` in R for help).

3. **p-Values:** Your summary of each test from above will return a *p*-value. Explain how to understand this value from the test and interpret its results (i.e., using the *t*-test and linear model) using this *p*-value. Write your conclusions for each of the above tests.
4. The *Psych* package: Choose an two interesting statistical tests or plotting techniques and apply them to your data set. Here, you are gaining experience learning a method on your own and writing code to make the method work with your data.

Important Details

Note: Please remember to include your name, as well as the names of your group members, on everything you submit for the class.

Required Deliverables

1. File `src/analysis.r`: The R source code that you used to answer your questions.
2. File `writing/report.md`: A Markdown report document listing all the outcomes of the above tasks. Namely, your report is to include the following.
3. File(s) `data/*`: Leave any data files that you used in this work.
 - The source of data: please include a link to the online source,

- The descriptions of the variables in your data set,
- The correlation analysis: state which variables were in what test,
- Make an attempt to explain the reasoning behind any correlation in the variables that you find. Be creative; the actual reason may not be known at this time but still consider an explanation.
- The statistical tests, their interpretations and conclusions
- What tools, plots or method did you choose from the *psych* package? What did you learn about your data from this new insight?

Write your report so that it is clear which parts of your code and discussion addresses what particular question. Also, in your code, offer comments to state which code is for what exploratory problem.