

# BIOS 600 · Computer Skills and Data Analysis

## Preliminary Info

**Project Details:** On November 8th I will give you data and instructions for a simple data analysis. The data analysis will require several computer skills from the list below. As we cover these topics in class, you should teach yourself how to perform these tasks in an approved computer software. This is an independent study project, but you may work together to learn a computer software until November 8th. Once the instructions and dataset for the data analysis are posted, you should not (a) discuss the project or (b) share code with your classmates.

The data analysis will be a take-home project. It will be open book, open notes, open internet (with limits), open instructor. You do not need to memorize computer syntax. In fact, I encourage you to NOT memorize but to compile examples of code for each the of computer skills.

**Approved Software:** An important concept in scientific research is reproducibility. Good research practice requires documenting every step of your analysis so that the results can be reproduced by another researcher. Un-scripted software (like Excel) does not provide a step-by-step log of your analysis. Because of this, analyses in Excel fail to meet the standard of reproducibility.

The following software meet the standard of reproducibility:

1. R
2. SAS
3. Matlab
4. SPSS
5. minitab
6. Stata
7. Shazam

**Help:** The TAs and I can provide help with SAS and R. You are free to choose another statistical software, but please do not expect the TAs or myself to offer you programming help.

**Computer Skills**

You should know how to perform the following tasks

1. Load data from a .txt or .csv file
2. Summarize the variables in a dataset
3. Generate appropriate plots
  - scatter plot
  - box plot
  - bar/line plot
  - histogram
4. Use probability distribution functions
5. Generate random numbers
6. Perform standard data analyses
  - t-test
  - two sample t-test
  - ANOVA
  - linear regression
  - $\chi^2$  tests for contingency tables
  - Fisher's exact test
  - Cochran-Mantel-Haenszel test
7. Perform nonparametric data analyses
  - sign test
  - Wilcoxon signed rank test
  - Wilcoxon rank sum test
  - Kruskal-Wallis test