Spring 2019

STAT512: Experimental Design and Data Analysis for Researchers II: R Software
Section 002
T/Th 8:00 – 9:50 Anatomy/Zoology W 205

Instructor:
Dr. Ben Sharp
Email: bensharp@colostate.edu
Office Hours: T/Th 9:50 – 10:30ish in same classroom
Office: Statistics Building 224

TAs/Graders:
Jialuo Liu jialuo.liu@colostate.edu

Objective:
This course builds on STAT 511 with further introduction to statistical thinking and statistical methods for graduate researchers. Emphasis continues to be on principles of design and inference, rather than the mechanics of computation. The course will continue to use R as the required statistical software package.

Prerequisite: STAT 511 or consent from instructor.

Canvas: You will need your eid and password to log onto Canvas. All course materials are available from Canvas. This will include lecture notes, examples, assignments and other material. You should strongly consider previewing and printing a copy of the notes to bring to class. A binder of physical material not only helps to follow along in class, but it also provides a reference for assignments and studying for exams.

Required Text: Ott and Longnecker, An Introduction to Statistical Methods and Data Analysis, Seventh Edition, Cengage Learning, 2016, ISBN 978-1-305-26947-7

Computing: R is a free software environment for statistical computing and graphics. To install R, go to www.r-project.org, and find the link to download R. Choose a CRAN mirror and select the appropriate platform you will be using.
RStudio is an integrated development environment for R that we will be using and is convenient for when coding R and developing graphics. It can be downloaded from www.rstudio.com.

Data Sets: From the Ott & Longnecker companion site (see Canvas), download the datasets in CSV format. Download the zip file for "ASCII-comma" and then unzip the file. Note: The file extension is .TXT even though the files are actually CSV! The files can be imported into R using the read.csv() function:

```
InData <- read.csv(file.choose(), quote = "'")
```

The file.choose() function allows files to be chosen interactively (without typing out the file path location). The quote option is used because the column names in the original data are (single) quoted.

Grading:

	Grade Percentage	Tentative Dates
Homework	30%	Fridays by 4 pm
Midterm 1	20%	In-Class, Thur 3/7
Midterm 2	20%	In-Class, Thur 4/25
Final Project	30%	Take-Home, Finals week

Academic Integrity: This course will adhere to the Academic Integrity Policy of the CSU General Catalog and the Student Conduct Code. On exams, students will sign a statement of the honor pledge "I have not given, received or used any unauthorized assistance."

Exams: Exams are open book, open notes. No computers or phones allowed. Calculator recommended. For any exam conflicts, please email the instructor at least one week prior to the scheduled exam date. Also see the instructor for exam grading questions.

Homework: Homework will be assigned weekly and typically due on Friday by 4:00 pm. You are encouraged to work together on homework, but the work turned in should be your own. All the computer output should be generated by you and the answers should be written by you. No late homework will be accepted (without prior approval). Homework is submitted in pdf format via Canvas, please submit all material in one file. Homework should be organized so that the grader can find your answers without searching through pages of computer output. Answer questions concisely. Write your own sentences to answer the question instead of just copying and pasting output. Only present the output that is related to the question. If the question does not require output, then do not include it. And, R code is not necessary unless specifically requested.

Final Project:
Students need to provide data for their project. Ideally these data would be part of your own research (or at least related). The analysis for this project needs to focus on a topic from the course: multiple regression, factorial ANOVA or mixed models. Notes that all of these topics include at least two predictor variables. Very small ($n<10$) or very large ($n>1000$) datasets will likely not be a good fit to use for analysis for the project.
A proposal will be due shortly after spring break. There will be at least one day of class in which we review and discuss your project. A concise write up (no more than a few pages) will include an introduction/background, discussion of variables/exploratory data analysis, design, and discussion of results. Figures and an appendix of code will also be part of the project write-up.

Other Notes about grades:
1. Withdrawal deadline is Monday 3/25 for a "W".
2. A weighted average of 80% or better will guarantee at least a B; 90% will guarantee an A.

3. There is no extra credit for the course.

Basic Course Outline:

1. Multiple Regression
    a. Basic model, assumptions, hypothesis testing, confidence/prediction intervals
    b. Polynomial terms, interactions, ANCOVA
    c. Model selection: stepwise selection and AICc/BIC
    d. Influence and collinearity diagnostics
    e. Generalized linear models: logistic and poison regression
2. Fixed Effects Models
    a. Blocked and latin square designs
    b. Factorial designs and pairwise comparisons
    c. Unbalanced and unreplicated designs.
3. Mixed Effect Models
    a. Random effects models
    b. Mixed models: crossed and nested factors, randomized complete block design, split-plot design, strip-plot design
    c. Repeated measures (one/two factors)