Spring 2-1-2018

BIOB 595.07: ST: Using R for Biostatistics

William Michael Janousek
University of Montana, Missoula

BIOB 595-07 (CRN 39312) Using R for Biostatistics

Mondays 3:00 – 4:50pm in NAC 014 computer lab.

Will Janousek PhD Candidate, Wildlife Biology Program, Forestry Building (FOR 303), 243-2035,

Office Hours: Tuesday / Friday 11am in FOR 303

Websites

R project - http://www.r-project.org/ [you can download R v 3.4.3 (Kite-Eating Tree…seriously) from this site]

CRAN [comprehensive R archive network] - http://cran.wustl.edu/

R examples - http://www.mayin.org/ajayshah/KB/R/index.html

Community Blogs - https://www.r-bloggers.com

Wiki with links to tutorials - http://openwetware.org/wiki/R_Statistics

Books

There are many now available. See the list via Amazon here. None required for this course, because there are so many good resources online.

Outlook

R is an increasingly popular free programming language for statistics. In this seminar, I will introduce the basics of data input and manipulation, show how to do common kinds of statistical analyses in biology (chi-sq, t tests, ANOVA, regression, PCA, and linear mixed-effects models), discuss how to fit and evaluate models, introduce R's graphical capabilities, and lay out some of R's more useful programming aspects (how to write scripts, loops, and functions). The course will consist of lectures and demonstrations coupled with lots of hands-on coding by you. Weekly assignments will be emailed out [problems to solve using R] whose answers you will email back to me. Hopefully we can go almost entirely paperless (except for this document!). Please also bring your own datasets, and I can tailor sections of the course to those—and derive homework problems based on them. In addition, to practice our visualization skills, each person will also do a piece of artwork based on a found data set.

Learning Outcomes

Upon completion of this class the student will have:

1. Gained a broad perspective of the versatility of program R as a statistical and graphing tool.
2. Gained experience in preforming basic statistical analyses including but not limited to t-tests, anova, and generalized linear models.
3. Gained experience in visualizing data using several R packages including but not limited to base plot, ggplot2, tmap, and ggmap.

Grading

Entirely credit/no credit. If you make a reasonable effort to do and turn in > 70% of the home works and do the final project, you will pass. No exams or extra credit given.

Schedule

Week 1 (1/22) First things first. Examples of what R can do, course outlook, installing R on your own computer. Basics like using the command prompt as a calculator and manipulating data. Objects → vectors, lists, and data.frames. How to get HELP! Saving R workspaces.

Week 2 (1/29) Functions. Learn how functions work and how to write them. Coding style for functions. Getting and using packages.

Week 3 (2/5) Getting your data in and out. Getting your raw data file(s) into shape, reading them into R, setting row & column names, accessing the data once it's in R. Also, how to assemble data frames from manipulations you do in R. How to write data back out to files. What to do with missing values. Dealing with file names and data structure.

Week 4 (2/12) Control-flow. How to control the flow of commands in a script. How to construct loops (for, while, loop variables) and set up conditional statements (if/else). Also how to do vectorized computations (lapply, tapply, sapply, apply), which speed up computations enormously. Setting up script files for more complicated programs, using functions to handle tedious repeated code.

(2/19) No class—President's Day

Week 5 (2/26) Accessing and manipulating your data. How to access, filter, and sort the contents of complex data types. Introduction to scales (nominal, ordinal, discrete, continuum), factors in R (factor, ordered.factor), and manipulating factors (changing labels, adding values).

Week 6 (3/5) Initial explorations: summarizing your data. How to find means, plot histograms and scatterplots, analyze outliers, and do basic linear regression. Some initial components of plotting.

Week 7 (3/12) Customizing your plots. More histograms, box plots, scatterplots, contour plots of density, and trellis plots. Creating multi-panel plots. Overlaying data on previously drawn plots using points. Putting in lines, adding fits, adding text, and annotating axes.

Week 8 (3/19) More graphics. Extracting data from complex datasets and plotting it at high density. Visualizing Big Data. How to plot raster data. Talk about art project!

(3/26) No class—Spring Break.

Week 9 (4/2) Linear and logistic regression, ANOVA. Linear models and how to deal with multiple variables. How to code formulas in models, specify interactions, and transform variables. How to read summaries and ANOVA tables. Dealing with factors & contrasts.

Week 10 (4/9) Model comparison & linear mixed-effects models. Comparison of models using AIC (tradeoff between additional variance explained & model complexity). Balanced designs and nesting. LMEs explained.

Week 11 (4/16) Spatial Data. Dealing with geographic data - *raster* package.

Week 12 (4/23) Bayesian Models. Introduction to Markov chain Monte Carlo (MCMC) sampling?

Week 13 (4/30) Presentation of Student Data. Each of you will do a 5 – 10 minute presentation on one of your own data sets. The presentation should cover what the question is, what you did and collected, and the structure of the data obtained. Then you should show how you analyzed it in R, along with some graphics to display it. Depending on total # of students, we may have to schedule some extra time for this. Creative projects like mapping spatial data are also an option.

**** Schedule Subject To Change ****

Policy for Accommodating Disabilities

The University of Montana assures equal access to instruction through collaboration between students with disabilities, instructors, and Disability Services for Students. If you have a disability that adversely affects your academic performance, and you have not already registered with Disability Services, please contact Disability Services in Lommasson Center 154 or 406.243.2243. I will work with you and Disability Services to provide an appropriate modification.