

# Stat 505: Linear Models or Advanced Regression I

## Introduction

Jim Robison-Cox

Department of Mathematical Sciences  
Montana State University  
jimrc@math.montana.edu

August 29, 2011

## Double-edged Sword of Stat Computing

- Computers are frustrating. I have to remember lots of little tricks.

BUT

- They are fast and accurate. When computer responds to my commands and provides insight to a difficult project, it's fun. I get a sense of accomplishment.
- A computer does exactly what it's told to do, not what I meant for it to do.
- To achieve the sense of satisfaction and accomplishment you have to learn the rules of the computer. Conform in order to conquer.

Stat computing uses some different tools, and has different end results, but the basics are the same as those of computer science.

Jim Robison-Cox

Stat 505 Intro

## Computing Overview

Input (data and models)

Process (using R or SAS)

Output (analysis and plots)

## Common Issues

- Storage – where do things reside in the computer?  
Manage storage so things are easy to find
- Planning – top-down programing is much more efficient  
Clearly specify the inputs and desired results.
- Reproducibility – if you come back in a year, will your code make any sense?
- Interface – how do we direct the computer to do a task?
  - Graphical User Interface (GUI): we point, click, select options
  - Command Line Interface (CLI): we type in commandsR and SAS use CLI, though there are several GUI interfaces for R (Rcmdr, JGR, pmg, SciViews) and SAS (JMP, Assist)

Jim Robison-Cox

Stat 505 Intro

Jim Robison-Cox

Stat 505 Intro

- What folders do you have on your computer?
- Do you have an mp3 player with folders? A phone?

Your Turn: list folders you know and compare with a neighbor.

- My Documents
- My Pictures
- My Videos
- Hidden folders with game configurations?

Need to organize to be able to find things.

Repeated structures are easier to remember. For example:

- Classes
  - Stat 501
  - Stat 505
    - data
    - notes
    - plots
    - Rcode !! Save your code !! You will reuse it!!
    - SAScode !! ditto !!
- Consulting
  - WhirlingDisease
    - 7 rivers, 2-5 sites each, 2 species
  - Proteomics
  - RangeScience
  - Ecology
    - data
    - plots
    - R code
    - Reference articles

- We find the right icon and (double) click to start a program.
- Or – select a file, double-click and it opens up in the right application. How does the computer know?
- File extensions are usually hidden. So are some folders and files.
- Script files of commands (.r, .s, .R) or .sas
- Data files
  - .txt space or tab separated
  - .csv comma separated values

Spreadsheets are great for data input. Then 'Save As' csv or txt.

- R uses .RData to store the 'image' of your session for next session.

- Create a folder for each class and for each project.
  - In MSWin, you can copy the R icon on the Desktop and give it a different "Start In" directory (right-click and go to Properties). Make one for each class or project. Change its title to say where it will start.
- .RData will live in that folder and contain data specific to the project or class. Should you save the image when exiting?
  - Advantage: Importing and processing a big dataset takes some time, so with big datasets do save. When exiting R click Yes or type y to save the image.
  - Disadvantage: Old stuff hangs around, and might cause confusion (which x? which y?). Do use descriptive names for variables and datasets. To not save, type n or click "no".
  - To clean it all out use command `rm(list=ls())` to remove all objects in memory. (Or just remove stuff you won't need again.) Then `save.image()` will copy over the top of the old image.

The official homework is online. Find it on the course webpage.

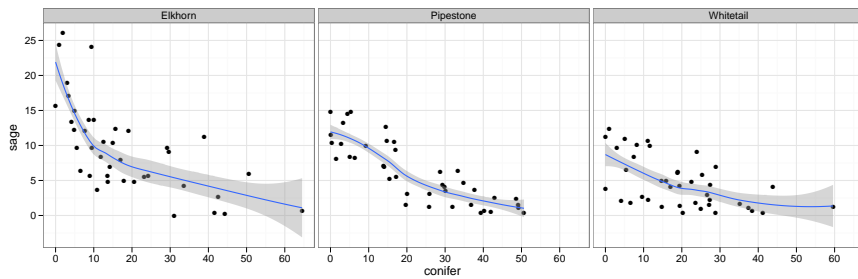
- Install (or update) R and either Rstudio or Emacs + ESS on a computer.  
Go to the home page for this course:  
<http://www.math.montana.edu/~jimrc/classes/stat505>  
Bookmark the page. Assignments will appear here. Click “Setting Up R” under Resources and follow the instructions.
- Set up a folder for Stat 505. If you use MSWin, copy the R icon setting “Start In” directory to Stat 505.
- Email me to answer these questions:
  - Will you do assignments on your computer, departmental computers, or in Reid/Roberts?
  - And a few other questions to see you got R installed. See HW 0.5 web page.

Use the email you check most often so I can contact you if needed.

- A programming environment
- A way to run stat analyses
- Built of functions and objects.
- Great at making complex plots (not necessarily easy)
- A project involving work from hundreds of people.
- Evolving and expanding. Over 3000 packages are available to extend R in all sorts of directions. +80/week

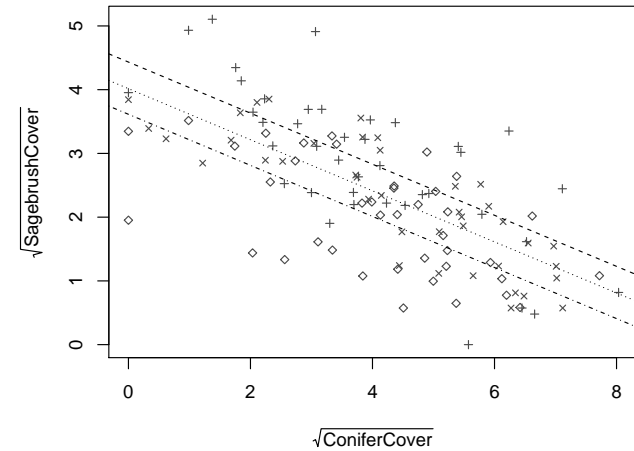
- A spreadsheet.
- A database.
- A place to enter data from the field directly.
- A point-and-click environment.
- A commercial product with professional support staff.

- For complex stat analyses.
  - Mixed Models
  - Spatial/Temporal correlation
  - Cutting edge stat techniques
- Complex graphics showing important facets of the data. Some commercial software does a good job, but R sets a high bar.
- Reproducibility. Save code to rerun an analysis with new data, or change things and rerun an analysis. Sweave and ODFweave package together the analysis, writeup, and output.



What is the relationship between conifer cover and sagebrush cover?

Is it the same for the three sites?



If we take square roots of both variables, we get a good fitting linear model. Slope is the same for each site.

## Organization

### Basic Principle

You must save your code. You will want to look at it months, perhaps years from now. It should be commented so it makes sense.

### Readability:

- Use comments to explain what you are doing.
- See guidelines for code readability from Google.
- For homework, include complete R code and output as an Appendix. In the body, write a report based on the stat package output showing that you know which parts to extract. (Output  $\neq$  Report)
- Use a fixed font like `courier` for computer code and output so that columns line up.