Index of useful S043/Stat151 resources

This document attempts to organize the many resources of the course to help you find what you need. This only lists handouts and coding resources. It does not list lecture readings or lecture code.

How to Find it: In general this is how I would find help for a specific thing:

1) Scan the list below to see if there is a handout on that specific thing. That may solve your problem.
2) If not, find a lecture where that specific thing happens. Then look at the R code for that lecture and/or the reading for that lecture.
3) You can also check section for that week---there is often useful code and tips in the section handouts.
4) Email a TF or instructor to get a pointer as to where to look.

Also, I strongly recommend downloading the library reserves scans onto your own computer. Then the names of those files will make finding things easier. Also grab the section handouts.

Generally speaking, I think the case study documents are possibly the most useful handouts.

Using the lmer() command

Handout: extracting information from lmer models – This is a useful reference guide for coding. It gives all sorts of ways to get information out of a fit model (this also includes linear regression models, actually).

Faraway, Chapter 8 (pp153-183) – This chapter is basically an R cookbook for using lmer. It has a few interesting case studies and a few tidbits of coding that may be useful. Worth a skim. It also broadens the concept of what one might do with random effects (e.g., analysis of blocks in randomized experiments). It also has a bit more technical detail of these methods.

Handout: Optimizing your optimization in R – It can be a horrible moment when you get "failure to optimize" type messages when you fit complex models with lots of moving parts. These models have less easy to identify "best fit" parameter values. This handout discusses this issue and some things you might do to avoid it.

Handout: Complex error structure (reproducing R&B chapter 6) – This handout gives all the R code to replicate the various models in the case study from Chapter 6. It is a good reference for how to specify these error structures in R.

Making nifty plots and tables

Handout: Intro to ggplot – Quick overview of how to plot data using ggplot.

Making regression tables (regression_table_demo.pdf) – Walks through making regression tables using screenreg, texreg, and stargazer.

Making plots with expand.grid (plotting_demo.pdf) – Walks through how to make trajectory plots and how to use predict() to make plots.

Core R, non-clustered regression, and debugging code

Handout: Linear regression in R – This tells you how to fit simple regression models in R. It shows the formula notation of such models, and gives useful snippits of code.

R and Stata Translation Guide – This has a series of side-by-side charts showing R code and Stata code that do the same thing. Useful for Stata users.

Spector, Chapter 6 (on lists and vectors) – This comes from a very nice, simple book called "Data Manipulation in R." It focuses on simple ways of manipulating data using the core R functionality. This chapter talks about subscripting, which is basically extracting chunks out of lists of values. Reading this will give you more a sense of what R is doing when you see "a[b]" or "a[[b]]" in code.

Matloff, Chapter 13 (on debugging) – This is a chapter from "The Art of R Programming," and it discusses more fancy ways of debugging code.

Handout: Introduction to glms in R – The code you need to fit a simple (non-hierarchical) glm in R.

Using R Markdown

Intro to R Markdown (getting-started-with-markdown.pdf) – Intro to how to use Markdown with some initial thoughts and how-tos. (In RMarkdown Help folder)

R Markdown Cheat sheets and reference guides – These tell you the markdown notation to get various effects. This includes customizing your r code chunks.

Case studies that illustrate specific models

Handout: lec 2.4 model_cheat_sheet.pdf – Connects output, math, and lmer command for the simple random slope model.

Handout: The HSB Example – This handout, which is actually a very commented and detailed R script, replicates all the code in Chapter 4 of R&B. Continue to work through this handout as you read more and more of Chapter 4 to get a sense of how to connect the ideas in Chapter 4 to R itself. ***This is a key reference document.***

Handout: Kenya case study – Another case study, this one illustrating three level models with a longitudinal component. More demonstration of useful R code and interpretation of that code. Very useful reference for getting code for final projects.

Handout: pefr case study – This extra case study illustrates fitting and interpreting a three level clustered model. It targets the R code needed to fit the model, and connecting parameters in the mathematical representation to the code output.

Writing your own math

math reference handout – This handout is in the Files/handouts folder, and has the Latex you need to write equations easily for your reports. Check it out to help with PS 1!

Notation

Handout: mathematical representations of multilevel models (model_representations.pdf) – Useful guide for how to look at notation some more.

Handout: defining pooling – Quick compare and contrast on language of "pooling."

Handout: A reference for interpreting GLMs – This gives a quick overview on interpreting the coefficients of a Poison model and logistic regression model.

Math Derivations

Handout: Residual covariance derivation – This gives the mathematical derivation of the covariance terms for a specific multilevel model to illustrate how marginal models are regression models with complex error structure.

Missing Data

Handout on some code for missing data (MissingData.pdf) – This handout has some useful code for dropping cases with data, doing single mean imputation, and for exploring missing data.

Unwin, Chapter 9 (on missing data and outlier detection) – As you start to explore your final project data, this quick guide for missing data and outlier detection has a few useful ideas and coding tricks. This is listed here in the syllabus as an approximate early-side date; keep this in mind when diving into real, more messy data