Psychology 314: Special Topics
An introduction to R for psychological research

William Revelle
Swift 315
email: revelle@northwestern.edu

November 26, 2018

1 Overview

If not already, R is well on its way to becoming the lingua franca of statistical analysis. It is open source, free, and extraordinarily powerful. Most importantly, more and more *packages* are being contributed to core R (R Core Team, 2018). As of today there are at least 13,127 packages that add to the functionality of R. More packages are added daily. R is a data analysis system that is both open source and is also extensible. Open source means that the actual computer code behind all operations is available to anyone to examine and to reuse, within the constraints of the GPL 2.0 (GNU General Public License, 1991). It is free software in the meaning of free speech in that everyone can use it, everyone can examine the code, everyone can distribute it, and everyone can add to it. Psychologists around the world are learning to take advantage of R for their research and this course will allow you to do so as well.

2 Objectives

1. To become familiar with the R language for modern statistical analysis is psychology. This will include using standard packages for data entry, data analysis, and scientific reports.
2. To learn how to write basic "R scripts" to combine several functions together and to document them in markup languages such as *RMarkdown*. This will help your research be reproducible, in that your analyses will be public.
3. To learn basic programming techniques in R. This will include learning how to read and adapt other peoples' functions as well as writing your own. This will lead to the development of new functions for psychological data analysis.

3 Text, readings, and requirements

3.1 Text

Much of this will be hands on work, and no published text will be used. However, readings from web based tutorials will be done extensively.

3.2 Readings

Multiple web based readings including, but not limited to the ones listed in the references. This list will be added to throughout the quarter.

Syllabus and handouts available at http://personality-project.org/courses/314/314.syllabus.pdf

3.3 Requirements

Motivation and interest. You do not need to have any experience in writing or using computer programs. If you can use an iPhone or Android, you can use R.

Because we will be doing modern and advanced statistics, Psychological statistics (Psych 201) (or its equivalent) is required and research methods (Psych 205) is highly recommended.

If you have a laptop computer, I suggest bringing it to class to work through the examples as they are discussed.

3.4 Accessibility

Any student requesting accommodations related to a disability or other condition is required to register with AccessibleNU (accessiblenu@northwestern.edu; 847-467-5530) and provide me with an accommodation notification from AccessibleNU, preferably within the first two weeks of class. All information will remain confidential.

3.5 Office Hours

Tuesdays (and most Thursdays) from 2-5. Swift Hall 315.

3.6 Evaluation

Homework assignments will be given weekly. These are for your benefit and will be graded on a completed, not completed basis. These are for your benefit. Doing the assignments will help you understand both the basic statistics involved, as well as how to do complex analyses.

Students will be expected to write a short paper demonstrating the use of R applied to their particular research interests. They will also be asked to present their use of R for data analysis or as a new function in short (5-10 minute) presentations in the last few weeks of the course.

This is a hands on course. You will be expected to try the various programs on simulated and real data sets.

This is a new course and we will all work together to make it a useful experience. Suggestions for improvements will be greatly appreciated.

4 Outline (to be added to frequently – keep checking)

We will be doing two things in parallel: learning modern statistical techniques and learning how to use, read and write R. Thus, each class will be about a certain statistical technique and how it is implemented in R, as well as developing expertise in useR, readR and writeR.

Week	Topic/function	Statistical notes	R Notes/functions	Homework
1	Introduction	R guide	A short course	Install R and Rstudio Problem set 1
2	Data Entry Descriptive Statistics	Packages and objects Help menus Correlation	The psych package Vignettes html and Rmd file	Problem set 2 Importing from SPSS Qualtrics, etc.

3	Correlation
	and graphics
		Confidence Intervals
		vs. "magic asteriks"
		the bootstrap
			Using the objects from a function
			Reading Code
			`corr.test` and `corPlot`
			`corPlotUpperLowerCi` and `multi.hist`
				Handout 3
				psych source code
				Handout 3a

4	Scales and Reliability
	
	Item Response Theory
		α to ω
		Reliability theory
		Why not use α
			`scoreItems`
			`alpha` and `omega`
			
			`tetrachoric` and `polychoric`
			`irt.fa` and `scoreIrt`
				Handout 4 Rmd
				factor analysis
				Handout 4b Rmd

5a
 	
5b
	More on factor analysis
	Central Limit Theorem
	
	ANOVA and the
	linear model
		fa and Iterative fit
		How to do factor analysis
		
		t and F tests
			`fa` `fa.diagram`
			
			`t.test` `anova` `lm`
				Handout 5a
				Rmd file
				
				Handout 5b
				The Rmd file
				5c html fileThe Rmd file

6	general linear model	Interactions as products
		of 0 centered scores
		Correlation and Regression
		mediation/moderation
			`lm` `setCor`
			`dummy.code`
			`corPlot` `corCi`
			`mediate`
				The Rmd file
				the html file
				outliers and Rmd file
				mediation and Rmd file

7	Writing functions
	
	Multilevel modeling
		More on regression
		
		modeling dynamics
			`lm` and `setCor`
			
			`multilevel.reliability`
			`lattice` `nlme`
				html and Rmd file
				programming html and Rmd file
				mlm html and Rmd file
				homework answers

8	Writing functions (part 2) Test Theory	Writing functions Test Theory	`alpha` `scoreItems` `ICC` `cohen.kappa`	html and Rmd file html and Rmd file Homework
9	data manipulation .	More on Reliability	`table` `%in%` `corPlot` `matSort`	data manipulation html Rmd .
10			Review of R	Sara Weston Tutorial

5 Detailed Notes

5.1 Week 1

Introduction to R. What is it, where did it come from, why use it. Why other statistical systems (e.g., SPSS, JMP, SAS) should be discouraged.

R (R Core Team, 2018) is an object oriented programming language. Just think of R like having a conversation with a specific person. They (R) have their own language, and you need to learn how to speak it. (adapted from Sara Weston – see A short course pages 36-64)

Downloading R, RStudio, and Rmarkdown

Objects and functions. Everything is an object.

5.2 Week 2

Functions are verbs, parameters are adverbs.

5.2.1 Packages What are they and why use them?

Installing the packages you need. Using `library` to make them active. Many packages have "vignettes" which describe what the package does and has some nice examples. The *psych* package has three vignettes. To find the vignettes for a particular package, e.g., the *psych* package you can just browse them.

R code

```
browseVignettes("psych")
```

On a Mac, if running R.app rather than RStudio, just go to the help menu and choose vignettes.

For a brief discussion of packages and functions. see Packages and objects.

5.2.2 Getting your data into R

The *psych* package (Revelle, 2018) is a basic toolkit for data analysis, with particular applications for psychology.

The `read.file` command will read from text, csv., or sav files. See the detailed discussion on data entry and the Problem set 2 demonstration of using RMarkdown.

`describe` to get basic descriptive statistics.

Using *Rmarkdown* and *Rstudio* to annotate your work.

5.3 Week 3

Using functions: Functions return objects which may be acted upon by other functions: Graphical displays of data and confidence intervals of the mean as well as the correlation. See the Handout for week 3

The "new statistics" Confidence intervals vs. "magic asteriks" (Cumming, 2013)

String functions together to do useful analyses.

What is packed in the object that a function returns? The `str` and `names` command.

Using the `by` and `apply` functions. Using `describeBy` and `statsBy` to get descriptive statistics by group. See the 2nd handout for week 3

5.4 Week 4

Scales are typically formed as composites of items. Methods for summing items or finding their means are straight forward applications (e.g., `scoreItems`). Alternative measures of internal consistency of these scales include $\alpha = \lambda_3$ (Cronbach, 1951; Guttman, 1945) and $\omega_h < \omega_t$ (Revelle and Zinbarg, 2009).

See the "How to" find ω

The discussion of reliability From alpha to omega is a fairly thorough treatment of reliability theory.

Debugging a function may be done using the `debug` or `browser` functions.

5.5 Week 5

Multivariate analysis includes `principal` components and *factor analysis*. See the "HowTo" use the *psych* package for factor analysis.

5.6 Week 6

Regression and the linear model using the `lm` function can also be done using the `setCor` function. A simple extension of `lm` is the application for doing mediation or moderation analysis. See the "How to " for mediation and moderation.

5.7 Week 7

Writing functions, using more functions for reliability and scale construction.

The study of test theory and the many kinds of reliabilities one can find.

A discussion of how to score single or multiple scales using `scoreItems` and other functions is found in the "How To" score scales.

Multilevel analysis considers data collected (e.g.) within subjects over time. We review these kind of data (Revelle and Wilt, 2017; Wilt and Revelle, 2017) and include a tutorial on multilevel modeling,

5.8 Week 8

More on reliability. A homework assignment to compare various estimates of reliability and to create a short function to find coefficient alpha

5.9 Week 9

Course review and further notes (taken from Sara Weston's introduction to R)

6 R advice

The R tutorial gives a short introduction to the use of R.

- (Macs and PCs) For this, or any other package to work, you must activate it by either using the Package Manager or the "library" command:
  - type library(psych)
  - If loading the psych package works, function such as `describe` and `pairs.panels` should work (or at least give an error message that is NOT "could not find function").
  - entering ?psych will give a list of the functions available in the psych package.

7 R guides and cheat sheets

See excellent tutorial by Sara Weston at the Open Science Framework https://osf.io/m5ja3/

The Rpad 6 page summary of most commands.

The Rstudio cheat sheets including Rmakrkdown cheat sheet.

Is R suitable for biostatisticians and clinical research?

Garrett Grolemund and Hadley Wickham have a very useful book describing R for Data Science which is available as a web book. It emphasizes a somewhat different philosophy from Core-R and introduces the concept of tidy R. This is set of packages that work well together but do not necessarily play well with others. It is worth exploring.

References

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16:297–334.

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

GNU General Public License (1991). Free software foundation. *License used by the Free Software Foundation for the GNU Project. See http://www. fsf. org/copyleft/gpl*.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4):255–282.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, https://cran.r-project.org/web/packages=psych. R package version 1.8.4.

Revelle, W. and Wilt, J. A. (2017). Analyzing dynamic data: a tutorial. *Personality and Individual Differences*.

Revelle, W. and Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: comments on Sijtsma. *Psychometrika*, 74(1):145–154.

Wilt, J. and Revelle, W. (2017). The big five, situational context, and affective experience. *Personality and Individual Differences*.

Visit the sapa-project.org to measure your personality.