

Psychology 350: Special Topics

An introduction to R for psychological research

William Revelle

Swift 315 (as if that were useful)

email: revelle@northwestern.edu

April 27, 2020

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

To make it easier, I have made a hyper link directly [to this section](#)

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.

1.1 News about syllabus updates

Today is April 27, 2020

April 8: Minor tweaks to the website to make the links work.

April 13: Modified Week 2 a little to include the last part of Week 1 intro (slides 51-80)

April 15: Added another handout for week 2. Also added a brief homework assignment in detailed notes section.

April 20: Updates and revisions

April 22: Revisions to the Handout 3a Rmd and html files, addition of omega tutorial

April 27th: Reorganized week 4 to slow us down a bit, trying to work through more userR applications rather than programmeR tricks.

Syllabus as a table

Week	Topic/function	Statistical notes	R Notes/functions	Homework
1		R guide Computers and Psychology	A short course	Install R and Rstudio Problem set 1
1b	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.
2a	Final part of Introduction starting at slide 51		Using the objects from a function	Distributions (html) and 2a.Rmd
2b	Correlation and graphics	Confidence Intervals vs. “magic asteriks” the bootstrap	error.dots , error.bars Reading Code t2d , fisherz corr.test and corPlot corPlotUpperLowerCi and multi.hist	Handout 2 psych source code Handout 2c
3	Scales and Reliability Item Response Theory	α to ω Reliability theory Why not use α factor analysis advanced notes on Factor Analysis	by head tail headTail scoreItems alpha omega tetrachoric and polychoric irt.fa and scoreIrt	Handout 3 Rmd Handout 3a Rmd How to use omega Handout 3b Rmd
4a	UseRs vs. ProgrammeRs Central Limit Theorem	UseR vs. ProgrammeR reliability appendix How to do factor analysis	testRetest splitHalf alpha scoreItems scoreOverlap fa fa.diagram	Reliability (html) and Reliability (Rmd) Rmd file
4b	ANOVA and the linear model	t and F tests	t.test anova lm	Handout 5b The Rmd file 5c html file The Rmd file
5	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 Homework with answers
9	data manipulation	More on Reliability	table %in% corPlot matSort	data manipulation html Rmd
10			Review of R	Sara Weston Tutorial

Detailed Notes

Week 1

The [history and current use](#) of statistical analyses and computer programming in psychology.

[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, 2019](#)) 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.

Week 2

Functions are verbs, parameters are adverbs. ([Introduction](#) slides 51-80)

3.2.1 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.

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

R code

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](#).

3.2.2 Getting your data into R

The *psych* package (Revelle, 2020) is a basic toolkit (a Swiss Army Knife) for data analysis, with particular applications for psychology. Some of these functions have been moved to the *psychTools* package which can be downloaded from CRAN or from the local repository.

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.

3.2.3 Homework for week 2

In a short R Markdown document:

1. Choose a data set (ideally one of yours, but you can use one of the ones in *psych* if you want).
2. In a paragraph, describe the data set the way you would in a paper. Who are the subjects, what are the variables of interest.
3. read the data into R (show your work)
4. Report basic descriptive statistics of the data set.
5. Graphically display the correlations of no more than 8 of your variables.
6. Find the “significance” of your correlations.

Send this me at revelle@northwestern.edu by the end of this weekend.

3.2.4 Week 3 a

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](#)

3.3 Week 3

3.3.1 Week 3 a

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](#) 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.

3.4 Week 5

Multivariate analysis includes `principal` components and *factor analysis*. See the “HowTo” use the *psych* package for [factor analysis](#).

3.5 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](#).

3.6 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, 2019](#); [Wilt and Revelle, 2019](#)) and include a tutorial on multilevel modeling,

3.7 Week 8

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

3.8 Week 9

Course [review](#) and further notes (taken from Sara Weston’s [introduction to R](#))

4 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.

5 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 Rmakrkdwn cheat sheet.

Is [R suitable for biostatisticians](#) and clinical research?

Garrett Golemund 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 (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revelle, W. (2020). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, <https://CRAN.r-project.org/package=psych>. R package version 2.0.1.
- Revelle, W. and Wilt, J. A. (2019). Analyzing dynamic data: a tutorial. *Personality and Individual Differences*, 136(1):38–51.
- 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. (2019). The big five, everyday contexts and activities, and affective experience. *Personality and Individual Differences*, 136(1):140–147.
- Visit the sapa-project.org to measure your personality.