Psychology 350
An introduction to R for psychological research

William Revelle
Swift 315
email: revelle@northwestern.edu

Elizabeth Dworak
Cresap 109 or Zoom (by appointment)
elizabeth.knowlton@northwestern.edu

May 25, 2022

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 May 25, 2022

March 22: First draft of syllabus is on sever and on Canvas

April; 4: Some revisions to the week 2 slides on correlation and the associated R studio slides

April 6: Added a discussion of correlational distributions and confidence intervals to the lecture notes (correlation) slides. Add some of this to the Rmd files as well.

I have modified *psych* to version 2.2.4 to include a fix to violin plotting of replicates from `corr.test`. However, it does not quite work yet.

April 11: Added lecture notes on the t-test and effect sizes. Updated how to get psych

April13; Added slides for [reliability] and updated the RMD file.

April 16: Updated the How to score scales and How to use omega tutorials.

April 17: Updated the factor analysis slides and Rmd file

Updated the *psych* and *psychTools* packages on the pmc server. Please install the latest versions (2.2.4 for both of them):

```R
install.packages(c("psych","psychTools"), repos="https://personality-project.org/r",type="source")
```

You must restart R and Rstudio to make the changes take effect!

April 20: Updated the UseR slides as well as the reliability slides. Added some information about Item Response Theory

April 25: Updated the t and F and the linear model slides as well as Week 5: the linear model

April 27: Note that version 4.2.0 of R has been released. You should update your version to this release. If you are using RStudio you will have to quit RStudio after you finish the update. The version of *psych* and

psychTools on the CRAN server are not up to date. To get versions 2.2.4 for both of these follow the update instructions for April 17th.

April 27th Added a discussion of process control loops and control

May 2: Updated the psych and psychTools packages (again) to version 2.2.5 for both. You should update your versions.

```
R code
install.packages(c("psych","psychTools"), repos="https://personality-project.org/r",type="source")
```

Added a new file for week 6 to show an actual application of data analysis for an in press paper .

May 4: Added a discussion of the Final project , updated the links to a recent paper (Widaman and Revelle, 2022) showing the power of simulation.

May 9: Updated the dynamics handout and the dynamics RMD/HTML files

A new release of psych is about to be sent to CRAN. The current release on the server is in the final stage of testing.

May 16: Version 2.2.5 is on CRAN. Today we show some minor functions as examples of using R to solve real problems.

May 18: Elaborated the homework assignment for this week.

May 25: Updated the advanced programing and review slides.

Final projects are due June 9th. I am happy to help you if want to assistance. Please schedule times.

2 Syllabus as a table

2.1 Using the Rmd files in the homework

The Homework is shown as both an html file (the markdown output) as well as an .Rmd file. In some browsers, if you click on the .Rmd file, it opens as a text file. This then needs to be saved on your computer using the .Rmd suffix. Then go to Rstudio and open the file using the RStudio File menu, open file option.

2.2 The syllabus

Week	Topic/function	Statistical notes	R Notes/functions	Homework
1	Computers and Psychology	R guide for psychology	R: Intro and R: Intro part 2	Install R and Rstudio
		Introduction to R	R Reference Card	Problem set 1
1b	Data Entry	Packages and objects	The psych package vignette	Problem set 2)
	Descriptive Statistics	Help menus	Vignettes	Importing from SPSS Qualtrics, etc.
2a		Correlation	html and Rmd file	
	Final part of Introduction starting at slide 51		Using the objects from a function	Distributions (html) and 2a.Rmd
2b	Correlation	Confidence Intervals	error.dots, error.bars	Handout 2
	and graphics	vs. "magic asteriks"	Reading Code	psych source code zip or psych source code
		the bootstrap	t2d, fisherz corr.test and corPlot	Handout 2c
			corPlotUpperLowerCi and multi.hist	
3	Descriptive vs. inferential statistics	The t tests and effect sizes	by head tail headTail	Handout 3 Rmd by and apply
				Handout 3 html
3b	Scales and Reliability	α to ω	scoreItems	Handout 3a Rmd
		Reliability theory	alpha How to score scales	
		Why not use α	omega How to use omega	
4		factor analysis	tetrachoric and polychoric	Handout 4a Rmd and html
		advanced notes on Factor Analysis	How to do a factor analysis	
	Item Response Theory	Reliability (again)	irt.fa and scoreIrt	
4b	UseRs vs. Programmers	UseR vs. ProgrammeR reliability appendix	testRetest splitHalf alpha	Reliability (html) and Reliability (Rmd)
			scoreItems scoreOverlap	
	Factor analysis	factor analysis How to do factor analysis	fa fa.diagram	fa Rmd file and fa html file
				Homework (see below)
5a	ANOVA and the linear model	t and F tests	t.test anova lm	Handout 5
				The Rmd file
				5b html fileThe Rmd file
5b	general linear model	The general linear model of 0 centered scores	lm setCor	The Rmd file
			dummy.code	the html file
		loops and control	corPlot corCi	and Rmd file
6 a	An example of analysis	Real data	scrub cohen.d reliability	the Rmd file
	More on the linear model		%in% subset outliers	data manipulation (html) Rmd
				Detecting outliers Rmd
		Mediation/Moderation	mediation/moderation mediate	mediation (html) Rmd the code matReg setCor.diagram
7	Writing functions	More on regression	lm and setCor	
				programming html and Rmd file
	Multilevel modeling	modeling dynamics See dynamics tutorial	multilevel.reliability	mlm html and Rmd file
		3 levels of analysis	lattice nlme	homework answers
		Final project datasets		Final project 3.7.1Homework
8	Writing functions (2)	Writing functions	alpha scoreItems scoreFast	html and Rmd file
		Two small functions		html and Rmd file
		Debugging (an example)		debugging html Rmd
	data manipulation	Scoring scales	table %in% subset merge	data manipulation html Rmd
			corPlot matSort	
		Test Theory part 1	irt.fa scoreIrt scoreIrt.2pl	Reliability Homework - answers
9	Item Response Theory (IRT)	Test Theory (part 2)	ICC cohen.kappa	html and Rmd file
		More on Reliability		
	Confirmatory Factor Analysis (CFA)	Using lavaan	functions: irt.fa scoreIrt packages: ltm MIRT lavaan	html and Rmd file
10	data manipulation	Advanced programming	table %in% grep sub order match	Advanced programming html Rmd
			corPlot matSort dfOrder	
	Review	Review of R	Sara Weston Tutorial	
11	Final papers due		Schedule Appointments for help	project due June 9th

3 Detailed Notes

3.1 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, 2022) 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.

3.2 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.

```
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.

3.2.2 Getting your data into R

The *psych* package (Revelle, 2022) 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 interst.

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)

Consider (Cohen, 1988, 1994) and the use of effect size estimates.

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 4b

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

3.5 Homework for week 4 (due on Sunday night)

Choose a data set that you find interesting (ideally, your own). Do the following analyses. Submit your report as a Rmd and PDF file.

1. Find the descriptive statistics

2. Find the structure of the items you are using. Is it one factor or more?

3. Form at least one composite of items to form a scale (or two or more)

4. What is their internal consistency?

5. Is the scale a good scale?

3.6 Week 5

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.

3.6.1 Homework for week 5

If you have any experimental or observational data, briefly describe it (in English), explain what the IVs and DVs are, and then compare an ANOVA approach to an linear model approach to your data. If you do not have any data, use the `Garcia` data set to test the effect of the IVs on the DVs. This should be done as a quasi paper: Introduction, Method, Results, Discussion, although these sections can be abbreviated to one sentence or so each.

3.6.2 Week 6

More on mediation, moderation, and how to detect outliers. A more extensive discussion of the linear model.

3.7 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.

Reliability is increased by aggregation, but the effects on validity are more complicated (Eagly and Revelle, 2022) 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,

An article (Revelle and Condon, 2015) describing why we use multiple levels to study the dynamics of personality (Revelle and Wilt, 2021).

3.7.1 Homework for week 7

In one paragraph, briefly outline your final project. This should include what data you will be examining, what kind of analyses you will be doing, and any hypotheses that you have.

3.8 Week 8

More on Test Theory, reliability, and data manipulation. A homework assignment to compare various estimates of reliability and to create a short function to find coefficient alpha. Note that some of the answers are given in the assignment, but that you are also asked to do these on a different set of data as well.

3.9 Week 9

Even more on test theory and reliability.

3.10 Week 10

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 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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates, Hillsdale, N.J., 2nd ed edition.

Cohen, J. (1994). The earth is round ($p < .05$). *American psychologist*, 49(12):997–1003.

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.

Eagly, A. H. and Revelle, W. (2022). Understanding the Magnitude of Psychological Differences between women and men requires seeing the forest and the trees (in press). *Perspectives in Psychological Science in press*.

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

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, psych, 2.2.3 edition. R package version 2.2.3.

Revelle, W. and Condon, D. M. (2015). A model for personality at three levels. *Journal of Research in Personality*, 56:70–81.

Revelle, W. and Wilt, J. A. (2019). Analyzing dynamic data: a tutorial. *Personality and Individual Differences*, 136(1):38–51.

Revelle, W. and Wilt, J. A. (2021). The history of dynamic approaches to personality. In Rauthman, J., Funder, D., and Sherman, R. A., editors, *The Handbook of Personality Dynamics and Processes*, chapter 1, pages 3–31. Elsevier.

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

Widaman, K. F. and Revelle, W. (2022). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*.

Wilt, J. and Revelle, W. (2019). The Big Five, Everyday Contexts and Activities, and Affective Experience. *Personality and Individual Differences*, 136(1):140–147.