Psychology 350
An introduction to R for psychological research

William Revelle
Swift 315
email: revelle@northwestern.edu
Office Hours T:2-4, F 2:3

Dana Chan
Swift 015 or Zoom (by appointment)
DanaChan2024@u.northwestern.edu
M 1:30-3, W: 11:00-12:30

May 24, 2023

0.1 News about syllabus updates

Today is May 24, 2023

March 27: First draft of syllabus is on the server and on Canvas

March 29: Minor change to March 28 slides.

March 29: improved the correlation slides.

April 2: Added some more material to the correlation slides. Also fixed the week 2 handout so that it actually works.

April 4; Added an Assignment (Homework 1-a) to the Assignments section of Canvas. This is basically to apply the examples in the week 2.c handout to a new data set. You may use your own data, or choose a data set from 350 data sets.

April 5: Added some more R references, April 5: Added further explanations and improvements of our boot function, as well as the differences between character and numeric data.

April 8: Fixed the bug in describeBy that we discovered last week. The new version of psych (same version number, new date) is on the pmc server. use `packageDate("psych")` to find out the date.

April 10: Updated the week 3 slides and added slides on Reliability

April 12: Updated the week 3 slides for factor analysis and principal components analysis.

April 12: Added a homework assignment (see the Week 3 detailed notes)

April 16: Improved the UseR slides and added links to the *psych* sourcecode.

April 19: Improved the factor analysis html and Rmd files. Added a homework assignment.

April 24: Improved the comparisons of t and F and the HTML and Rmd files.

April 26 Modified the The general linear model slides to include data from a recent paper comparing univariate and multivariate measures of effect sizes.

May 1: Added a modifying code.rmd modifying code HTML file on programming

May 3: Improved the dynamics slides and associated HTML/RMD files. Added a small homework assignment using what we have discussed today.

May 8: Added updates to *psychTools* and *psych* which are on the server. Added a discussion of how to manipulate data. Following a suggestion/request from Yaelle Pierre, I added the `recode` function.

May 10: Further small fixes to *psych* and *psychTools*. Same version numbers (2.3.5) but dates show when last updated (05/09). Added a RMD and HTML working through scale construction, validation, and cross validation.

May 15: Further small fixes to *psych*. Same version number (2.3.5) but dates show when last updated (05/14). Added a new reading about scale construction. Updates to test theory HTML and Rmd files. Updated the laavan slides.

Note: *psych* has been updated to version 2.3.5 on the pmc server This has the improvements discussed in class on Wednesday.

May 17: Added slides to discuss how to write and use some simple graphic functions html and Rmd.

May 22: Another update to *psych* (packageDate 05/22/2023). Updated the Advanced programming slides, Rmd and HTML.

May 24: Improved the final summary lecture notes;

1 Syllabus as a table

1.1 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.

1.2 The syllabus

Week	Topic/function	Statistical notes	R Notes/functions	Homework/examples
1	Computers and Psychology	R guide for psychology	A short history of computing R: overview and R: Intro part 2	Install R and Rstudio
		Introduction to R	R Reference Card	Problem Set 1 RMD html
1b	Data Entry	Packages and objects	The psych package	Problem set 2 RMD html
	Descriptive Statistics	Help menus		Importing from SPSS
		Correlation	Vignettes	Qualtrics, etc.
		A diversion	Descriptives stats html and Rmd	
	Final part of Introduction starting at slide 51		Using the objects from a function	Distributions (html) and 2a.Rmd
2	Correlation and graphics	Confidence Intervals vs. "magic astericks"	`error.dots`, `error.bars` Reading Code	Handout 2 html rmd *psych* source code zip or *psych* source code
		the bootstrap starting at page 28	`t2d`, `fisherz` `corr.test` and `corPlot`	Handout 2c (homework 1a)html Rmd
		More on sampling html	`corPlotUpperLowerCi` and `multi.hist`	Issues in data html
3	Scales and Reliability	Reliability	`by head tail headTail`	Handout 3 Rmd
		α to ω	`splitHalf alpha`	Handout 3a Rmd
		Reliability theory	`scoreItems scoreOverlap`	
		Why not use α	`omega reiiability`	How to use omega
		factor analysis	`tetrachoric` and `polychoric`	Handout 3b Rmd
		advanced notes on Factor Analysis		
	Item Response Theory		`irt.fa` and `scoreIrt`	
4	UseRs vs. ProgrammeRs	UseR vs. ProgrammeR	`testRetest splitHalf alpha`	Reliability (html) and Reliability (Rmd)
		reliability appendix	`scoreItems scoreOverlap`	
	Factor analysis	factor analysis How to do factor analysis	`fa` `fa.diagram` `fa.congurence` `iclust`	fa Rmd file
				fa html file
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 scatterHist`	the html file
			`corPlot corCi`	and Rmd file
				(Eagly and Revelle, 2022)
6 a	More on the linear model		%in% subset `outliers`	data manipulation (html) Rmd
				Detecting outliers Rmd
		Mediation/Moderation	mediation/moderation `mediate`	mediation (html) Rmd matReg setCor.diagram modifying code.rmd modifying code HTML
7	Writing functions	More on regression	`lm` and `setCor`	
				programming Rmd
	Multilevel modeling	modeling dynamics	`multilevel.reliability`	MLM Rmd
		3 levels of analysis	`lattice nlme`	homework answers
		Final project datasets		Final project 2.7.1Homework
8	Writing functions (2)	Manipulating data. Writing functions	`alpha scoreItems scoreFast`	html and Rmd file
		Debugging (an example)		debugging html Rmd
	data manipulation	Scoring scales	`table` `%in%` `subset` `merge`	data manipulation html Rmd
			`corPlot matSort`	scale construction,
	scale construction	Test Theory	`irt.fa scoreIrt scoreIrt.2pl`	Reliability Homework - answers
9	Item Response Theory (IRT)	Test Theory (continued)	`ICC cohen.kappa`	html and Rmd file
		More on Reliability		
	Confirmatory Factor Analysis (ÇFA)	Using *lavaan*	functions: `irt.fa` `scoreIrt`	html and Rmd file
	Graphics		packages: *ltm MIRT lavaan*	
		Function development	`diagram`	html Rmd
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 a	Review (continued)	Review of R		Sara Weston Tutorial

2 Detailed Notes

2.1 Week 1

The history and current use of statistical analyses and computer programming in psychology (Revelle et al., 2020)

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, 2023) 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.

2.2 Week 2

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

2.2.1 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.

2.2.2 Getting your data into R

The *psych* package (Revelle, 2023) 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.

2.2.3 Homework for week 2

As discussed in the Handout 2c, adapt that code to do the following:

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.

Turn this in on Canvas by Sunday night.

2.2.4 More comments on class notes

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

Issues in treating character versus numeric data. See the data html

Steps towards improving a function. The example of our boot function

2.2.5 Some interesting web resources

While browsing the web, I cam across several interesting links

1. Best coding practices for R
2. Big Book of R (a compilation of 300 links to various R related readings.)

2.3 Week 3

2.3.1 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 (Revelle and Condon, 2019)

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

2.3.2 Homework for Week 3 – Graded for 10 points

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* (see the data sets listed in factor analysis p 19 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. Conduct a factor analysis of your data. How many factors best represent the data?

6. Form the items into scales that best represent these factors. What are various estimates of reliability of your scales? (e.g. $\omega_h, \alpha, \omega_t$, split half estimates, etc.) . Why do these estimate differ?

Turn this in on Canvas by Sunday night.

2.4 Week 4b

Multivariate analysis includes `principal` components and *factor analysis*. See the "HowTo" use the *psych* package for factor analysis. Look at the extended Rmd and html files on factor analysis.

2.4.1 Homework for week 4: Graded for 5 points

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* (see the data sets listed in factor analysis p 19 if you want) These data can be the same you used last week..
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.
4. read the data into R (show your work)
5. Report basic descriptive statistics of the data set.
6. Conduct a factor analysis of your data. How many factors best represent the data?
7. Factor analysis is a process of model comparison. This means you should try different solutions to the same data set. Try one more and one fewer than the solutions you got before.
8. Try a `principal` components as well as an `iclust` solution.

   Compare these solutions using `fa.congruence`

Turn this in on Canvas by Sunday night.

2.5 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.

2.5.1 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.

2.6 Week 6

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

2.6.1 Homework for week 6

Using the complete Fisher data set discussed in class, try to form different composites than the ones done in class, and then do the multi-level graphics and analysis as discussed. Look at the discussion of scoring

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

2.7 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, 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).

2.7.1 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.

2.8 Week 8

More on reliability and daa manipulation. A homework assignment to compare various estimates of reliability and to create a short function to find coefficient alpha. Note that the answers are given in the assignment.

2.9 Week 9

Even more on test theory and reliability.

2.10 Week 10

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

3 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.

4 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.

Eagly, A. H. and Revelle, W. (2022). Understanding the Magnitude of Psychological Differences Between Women and Men Requires Seeing the Forest and the Trees. *Perspectives on Psychological Science*, 17(5):1339–1358.

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

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, https://CRAN.r-project.org/package=psych, 2.3.3 edition. R package version 2.3.3.

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 Condon, D. M. (2019). Reliability: from alpha to omega. *Psychological Assessment*, 31(12):1395–1411.

Revelle, W., Elleman, L. G., and Hall, A. (2020). Statistical analyses and computer programming in personality. In Corr, P. J., editor, *Cambridge University Press Handbook of Personality*, pages 495–534. Cambridge University Press.

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.

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