

# Psychology 350

## An introduction to R for psychological research

William Revelle

Swift 315

email: [revelle@northwestern.edu](mailto:revelle@northwestern.edu)

Office Hours T:2-4, F 2:3

Kayla Garner

Cresap 109 or Zoom (by appointment)

[kalyagarner@u.northwestern.edu](mailto:kalyagarner@u.northwestern.edu)

W from 3-3:30pm

Th from 2:30-3:30pm

Friday 11:30-12:30 (zoom)

May 7, 2024

## 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 7, 2024

March 24: First draft of syllabus is on the server and on Canvas

March 28: Minor improvements to [Data Entry](#) slides

March 31: Minor improvements in [Correlation](#), [week 2 Rmd](#) and [html](#) files.

April 7th: Added a discussion of [conflicts between packages](#), particularly with Tibbles from TidyVerse.

April 10th: Modified the examples in [Handout 3 Rmd](#)

April 15th: Elaborated the discussion of [UseRs vs. ProgrammeRs](#). Updated the [Reliability \(html\)](#) and [Reliability](#) (Rmd) handouts.

April 16th: Serious additions to the [fa html file](#) factor analysis slides. I work through a few simple R functions to actually do a factor analysis from a correlation matrix. Also show examples of “hand” rotation of factors.

April 21st: Psych has been updated. Please reinstall from the [personality-project.org](https://personality-project.org) repository and check the date;

R code

```
install.packages("psych", repos="https://personality-project.org/r")
quit()
packageDate(pkg="psych")
```

Updated the slides for week 5.

April 23rd. Another update to psych. Please reinstall (again) and check the date. It should be April 23.

Updated the week 5 slides. [5b html file](#)The Rmd file on the linear model.

April 28th: Yet another minor update to *psychTools*. Current date is 04-28-24. Updated the Mediation/moderation notes.

April 29th Updated the example slides for week 6

May 1: Added a discussion of `merge` and `maps` with an up to date demonstration of using maps to show epidemiological trends. You will need to install the package *maps*.

R has been updated to version 4.4.0 This requires reinstalling R and the relevant packages.

```
install.packages("maps")
```

R code

May 5: Another small update to *psych* (still version 2.4.4 but the date is

```
packageDate(pkg="psych")  
[1] "2024-05-03"
```

R code

The change was to allow `corPlot` to display “magic astericks” . Also had a change in `cor2` and `fa.sort`.

May 8th Updated the [data manipulation](#) slides.

## 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/examples
1	Computers and Psychology	R guide for psychology	A short history of computing R: overview and R: Intro part 2 R Reference Card	Install R and Rstudio
1b	Data Entry Descriptive Statistics	Introduction to R Packages and objects Help menus Correlation A diversion	The psych package as a Swiss Army Knife Vignettes Descriptives stats <a href="#">html</a> and <a href="#">Rmd</a>	Problem Set 1 <a href="#">RMD html</a> Problem set 2 <a href="#">RMD html</a> Importing from SPSS Qualtrics, etc.
2	Final part of <a href="#">Introduction</a> starting at slide 51 <a href="#">Correlation</a> and graphics	Confidence Intervals vs. “magic astericks”  <a href="#">the bootstrap</a> starting at page 28 More on <a href="#">sampling html</a>	Using the objects from a function  <code>error.dots</code> , <code>error.bars</code> Reading Code  <code>t2d</code> , <code>fisherz</code> <code>corr.test</code> and <code>corPlot</code> <code>corPlotUpperLowerCi</code> and <code>multi.hist</code>	<a href="#">Distributions</a> (html) and <a href="#">2a.Rmd</a>  Handout 2 <a href="#">html rmd</a> <a href="#">psych</a> <a href="#">source code</a> <a href="#">zip</a> or <a href="#">psych</a> <a href="#">source code</a> Handout 2c (homework 1a) <a href="#">html</a> <a href="#">Rmd</a> Issues in data <a href="#">html</a>
3	Scales and Reliability      Item Response Theory	Reliability $\alpha$ to $\omega$ Reliability theory Why not use $\alpha$ <a href="#">factor analysis</a> advanced notes on <a href="#">Factor Analysis</a>	by head tail <code>headTail</code> <code>splitHalf</code> <code>alpha</code> <code>scoreItems</code> <code>scoreOverlap</code> <code>omega</code> <code>reliability</code> <code>tetrachoric</code> and <code>polychoric</code>      <code>irt.fa</code> and <code>scoreIrt</code>	<a href="#">conflicts</a> <a href="#">Handout 3 Rmd</a> <a href="#">Handout 3a Rmd</a>     <a href="#">How to use omega</a> <a href="#">Handout 3b Rmd</a>
4	UseRs vs. Program- meRs  Factor analysis	UseR vs. ProgrammeR <a href="#">reliability appendix</a>  <a href="#">factor analysis</a> <a href="#">How</a> <a href="#">to do factor analysis</a>	<code>testRetest</code> <code>splitHalf</code> <code>alpha</code>   <code>scoreItems</code> <code>scoreOverlap</code> <code>fa</code> <code>fa.diagram</code>	<a href="#">Reliability</a> (html) and <a href="#">Reliability</a> (Rmd)  <a href="#">fa</a> <a href="#">Rmd</a> file  <a href="#">fa</a> <a href="#">html</a> file
5a	ANOVA and the linear model	<a href="#">t and F tests</a>	<code>t.test</code> <code>anova</code> <code>lm</code>	<a href="#">Handout 5</a> <a href="#">The Rmd file</a> <a href="#">5b html file</a> <a href="#">The Rmd file</a> <a href="#">The Rmd file</a> <a href="#">the html file</a> and <a href="#">Rmd</a> file
5b	general linear model	<a href="#">The general linear model</a> of 0 centered scores	<code>lm</code> <code>lmCor</code> <code>dummy.code</code> <code>corPlot</code> <code>corCi</code>	
6 a	More on the linear model		<code>%in%</code> <code>subset</code> <code>outliers</code>	<a href="#">data manipulation</a> (html) <a href="#">Rmd</a> <a href="#">Detecting outliers</a> <a href="#">Rmd</a> <a href="#">mediation</a> (html) <a href="#">Rmd</a> <a href="#">matReg</a> <a href="#">lmCor.diagram</a> <a href="#">programming</a> <a href="#">html</a> and <a href="#">Rmd</a> file
6b		<a href="#">Mediation/Moderation</a>  writing functions	<a href="#">mediation/moderation</a> <code>mediate</code>  <code>merge</code> and <code>map</code>	
7	Writing functions  Multilevel modeling	More on regression  <a href="#">modeling dynamics</a> <a href="#">3 levels of analysis</a> <a href="#">CTA.RST</a> <a href="#">data manipulation</a>	<code>lm</code> and <code>setCor</code>   <code>multilevel.reliability</code> <code>lattice</code> <code>nlme</code> <code>rbind</code> <code>cbind</code> <code>merge</code> <code>recocde</code> <code>vJoin</code> <code>scrub</code>	<a href="#">programming</a> <a href="#">html</a> and <a href="#">Rmd</a> file <a href="#">mlm</a> <a href="#">html</a> and <a href="#">Rmd</a> file <a href="#">homework answers</a> <a href="#">Final project 3.6.1</a> <a href="#">Homework</a> <a href="#">data manipulation</a> <a href="#">html</a> <a href="#">Rmd</a>
8	Writing functions (2)  data manipulation	Writing <a href="#">functions</a> Debugging ( <a href="#">an example</a> ) <a href="#">Scoring scales</a>  <a href="#">Test Theory</a>	<code>alpha</code> <code>scoreItems</code> <code>scoreFast</code>   <code>table</code> <code>%in%</code> <code>subset</code> <code>merge</code> <code>corPlot</code> <code>matSort</code> <code>irt.fa</code> <code>scoreIrt</code> <code>scoreIrt.2pl</code>	<a href="#">html</a> and <a href="#">Rmd</a> file <a href="#">debugging</a> <a href="#">html</a> <a href="#">Rmd</a> <a href="#">data manipulation</a> <a href="#">html</a> <a href="#">Rmd</a>
9	Item Response Theory (IRT) Confirmatory Factor Analysis (CFA)	<a href="#">Test Theory</a> (continued) More on Reliability <a href="#">Using lavaan</a>	ICC <code>cohen.kappa</code>   functions: <code>irt.fa</code> <code>scoreIrt</code> packages: <code>ltm</code> <code>MIRT</code> <code>lavaan</code>	<a href="#">html</a> and <a href="#">Rmd</a> file   <a href="#">html</a> and <a href="#">Rmd</a> file
10	data manipulation	<a href="#">Advanced programming</a>	<code>table</code> <code>%in%</code> <code>grep</code> <code>sub</code> <code>order</code> <code>match</code>   <code>corPlot</code> <code>matSort</code> <code>dfOrder</code> <a href="#">Sara Weston Tutorial</a>	<a href="#">Advanced programming</a> <a href="#">html</a> <a href="#">Rmd</a>
11 a	Review Review (continued)	<a href="#">Review of R</a> <a href="#">Review of R</a>		<a href="#">Sara Weston Tutorial</a>

## 3 Detailed Notes

### 3.1 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.

## 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.

```
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, 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.

### 3.2.3 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 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.

Turn this in on Canvas by Sunday night.

### 3.2.4 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](#)

### 3.2.5 Some interesting web resources

While browsing the web, I came 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.)

## 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  $\omega$

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.

### 3.3.2 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.

### 3.4 Week 4b

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

### 3.5 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](#).

#### 3.5.1 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.

#### 3.5.2 Week 6

More on mediation, moderation, and how to detect outliers. A more extensive discussion of the linear model. Added a discussion of the use of `merge` and the `maps` package.

### 3.6 Week 7

Changes in *psych*2.4.4 to allow ‘better’ graphics output.

R code

```
cors <- cor2(bfi[1:25],bfi[26:28], pval=TRUE, show=FALSE)
cors <- fa.sort(cors) #sort them by size of correlation
corPlot(cors, cex=1,stars=TRUE_
```

[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](#)).

#### 3.6.1 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.

### 3.7 Week 8

More on 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 the answers are given in the assignment.

### 3.8 Week 9

Even more on test theory and reliability.

### 3.9 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 Rmarkdown 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.