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

Kayla Garner
Cresap 109 or Zoom (by appointment)
kalyagarner@u.northwestern.edu
W from 3-3:30pm & 5:30-6pm
Th from 2:30-3:30pm

March 26, 2024

1 Overview

If not already, the R language is well on its way to becoming the lingua franca of statistical analysis. The language is an open source, free, and extraordinarily powerful computer programming environment. R is a data analysis system that is both open source and is also extensible. Most importantly, more and more *packages* are being contributed to core R (R Core Team, 2023). As of today there are at least 21,365 packages that add to the functionality of R. More packages are added daily. 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. It is extensible means that users can add new features to it and then share their code. 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.

R may be downloaded for free from the Comprehensive R Archive Network (CRAN which may be found at https://cran.r-project.org) and is available for PCs, MacOS, and Linux/Unix operating systems. For purposes of speed, much of core-R is written in Fortran or C++, but most of the packages for R are written in R itself. For R is more than a statistical system, it is a programming language. This means that R is extensible in that anyone can add *packages* to the CRAN as well as other repositories such as GitHub or BioConductor (https://bioconductor.org). Perhaps more importantly, one can write short *scripts* for personal used. CRAN has certain quality assurance tests that guarantee the contributed programs have consistent documentation, including examples, and will not fail while running these examples. CRAN *does not check* the validity or utility of submitted packages, that is up to the contributor as well as the users of the packages. As of this writing, several thousand contributors have added more than 21,365 *packages* to core-R and this number increases daily.

2 Course Objectives

1. To become familiar with the R language for modern statistical analysis in psychology. This will include using standard packages for data entry, data analysis, and scientific reports. A particularly useful Integrative Development Environment (IDE) is available from R Studio (now called posit) which has a free desktop version to download.

2. To learn how to write basic "R scripts" to combine several functions together and to document them in markup languages such as R Markdown or Sweave. This will help your research be reproducible, in that your analyses will be public and well documented.

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 are available at https://personality-project.org/courses/350.syllabus.pdf and will be added to as the course progresses. When you find a particularly helpful one, let me know and we will add it to this syllabus.

Although a copy of the syllabus will be available on the NU CANVAS website, the material will also appear on the https://personality-project.org/courses/350/ page. This follows the basic principle of open science that materials should be shared.

An Introduction to R is a valuable resource. It may be purchased as a 100 page book, or just downloaded for free.

Some of the readings are "vignettes". These are helpful tutorials about how to use particular packages.

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.

The assumption for this course is that you have access to a computer. I suggest using it in class to work through the examples as they are discussed. (Obviously, although you can attend the lectures with just a smart phone or tablet, the actual assignments require the use of a computer).

3.4 Accessibility

Northwestern University is committed to providing the most accessible learning environment as possible for students with disabilities. Should you anticipate or experience disability-related barriers in the academic setting, please contact AccessibleNU to move forward with the university's established accommodation process (e: accessiblenu@northwestern.edu; p: 847-467-5530). If you already have established accommodations with AccessibleNU, please let me know as soon as possible, preferably within the first two weeks of the term, so we can work together to implement your disability accommodations. Disability information, including academic accommodations, is confidential under the Family Educational Rights and Privacy Act.

3.5 Office Hours

Tuesdays (and most Thursdays) from 2-4. To meet physically, my office is Swift Hall 315. Otherwise I am available (by appointment) for "office hours" by Zoom. I strongly encourage you to take advantage of this time to ask additional questions. Asking questions by email at any time or CampusWire is fine .

Past experience suggests that coming to office hours, either with me or Kayla, will help you understand the material.

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.

In addition, we will have two to three graded assignments (two short, one a bit longer) showing your ability to analyze data using R.

Students will be expected to write a short paper demonstrating the use of R applied to their particular research interests. They might 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.

Although homework will be graded individually, working together on assignments is encouraged. Realizing that you are not alone in your confusion is probably the most important part of study groups. We will also take advantage of CampusWire, an interactive help system to contact me and our Teaching Assistant, Kayla Garner.

4 Thoughts about learning programming and statistics

Few people find programming immediately obvious. Fewer still find statistics beautiful. Think of this course as learning two new languages. Like any new language, this will take time. It will take practice.

Programming in particular is a matter of trying. Your first and second, and even 10th attempt will likely not work. Don't worry, it is not you, it is the fact that computers are particularly picky about what they want. View the class as a computer game where you are trying to get to the next level. If you can master any computer game, you can learn R. Modern statistics is meant to be interactive: you are trying to make sense of the world by focusing on the signal and ignoring the noise. R will let you do that.

One word of warning: R can be addictive. Like any slot machine or psychologically well designed system, it is randomly reinforcing. After many attempts, you are suddenly rewarded. That release of Dopamine can overcome a great deal of the frustration of getting there and makes you want to keep going at the expense of not doing other, more important things. Don't forget to eat, sleep, or talk to your friends. Although I will deny it, your other courses and your "real life" are important as well.

5 Provost statements

The provost has asked that all instructors include the following in all syllabi:

5.1 Academic Integrity

Students in this course are required to comply with the policies found in the booklet, "Academic Integrity at Northwestern University: A Basic Guide". All papers submitted for credit in this course must be submitted electronically unless otherwise instructed by the professor. Your written work may be tested for plagiarized content. For details regarding academic integrity at Northwestern or to download the guide, visit: https://www.northwestern.edu/provost/policies-procedures/academic-integrity/index.html

Any form of cheating, including improper use of content generated by artificial intelligence, constitutes a violation of Northwestern's academic integrity policy.

5.2 Religious Observance Statement

Northwestern is committed to fostering an academic community respectful and welcoming of persons from all backgrounds. To that end, the policy on academic accommodations for religious holidays stipulates that students will not be penalized for class absences to observe religious holidays. If you will observe a religious holiday during a class meeting, scheduled exam, or assignment deadline, please let me know as soon as possible, preferably within the first two week of class. If exams or assignment deadlines on the syllabus fall on religious holidays you observe, please reach out so that we can discuss that coursework.

5.3 Course Details subject to change

Please note that the specifics of this course syllabus are subject to change in the case of unforeseen circumstances. Instructors will notify students of any changes as soon as possible. Students will be responsible for abiding by the changes.

5.4 Exceptions to class modality

Class sessions for this course will occur in person. Individual students will not be granted permission to attend remotely except as the result of an Americans with Disabilities Act (ADA) accommodation as determined by AccessibleNU.

Community health remains our priority. If you are experiencing symptoms of COVID-19, do not attend class and follow the steps outlined by the CDC for testing and isolation. Contact meas soon as possible to make plans to complete your coursework. Students who experience other personal emergencies should contact tme as soon as possible to arrange to complete coursework.

Should public health recommendations prevent in-person class from being held on a given day, I or the university will notify students.

5.5 Prohibition of recording of class sessions by students

Unauthorized student recording of classroom or other academic activities (including advising sessions or office hours) is prohibited. Unauthorized recording is unethical and may also be a violation of University policy and state law. Students requesting the use of assistive technology as an accommodation should contact AccessibleNU. Unauthorized use of classroom recordings – including distributing or posting them – is also prohibited. Under the University's Copyright Policy, faculty own the copyright to instructional materials – including those resources created specifically for the purposes of instruction, such as syllabi, lectures and lecture notes, and presentations. Students cannot copy, reproduce, display, or distribute these materials. Students who engage in unauthorized recording, unauthorized use of a recording, or unauthorized distribution of instructional materials will be referred to the appropriate University office for follow-up.

5.6   Support for Wellness and Mental Health

Northwestern University is committed to supporting the wellness of our students. Student Affairs has multiple resources to support student wellness and mental health. If you are feeling distressed or overwhelmed, please reach out for help. Students can access confidential resources through the Counseling and Psychological Services (CAPS), Religious and Spiritual Life (RSL) and the Center for Awareness, Response and Education (CARE). All Northwestern students are also eligible to access support at no cost though TimelyCare, a virtual mental health platform that provides counseling, health coaching and 24/7 on-demand services.

Additional information on the resources mentioned above can be found here:

https://www.northwestern.edu/counseling/

https://www.northwestern.edu/religious-life/

https://www.northwestern.edu/care/

https://www.northwestern.edu/studentaffairs/timelycare.html

5.7   The Writing Place

When working on writing assignments for this class, I encourage you to visit the Writing Place, Northwestern's peer writing center. You will work with juniors and seniors who have been trained to provide you feedback and assistance on any type of writing at any stage in the writing process. They will not edit your work. Rather, they will work with you to brainstorm ideas, organize or outline an essay, clarify your argument, document your sources correctly, or refine grammar and style.

To book an appointment, register for an account at https://northwestern.mywconline.com/.

6 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.

6.1 News about syllabus updates

Today is March 26, 2024

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

7 Syllabus as a table

7.1 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.

7.2 The syllabus

Week	Topic/function	Statistical notes	R Notes/functions	Homework/examples
1 1b	Computers and Psychology Data Entry Descriptive Statistics	R guide for psychology Introduction to R Packages and objects Help menus Correlation A diversion	A short history of computing R: overview and R: Intro part 2 R Reference Card The psych package as a Swiss Army Knife Vignettes Descriptives stats html and Rmd	Install R and Rstudio Problem Set 1 RMD html Problem set 2 RMD html Importing from SPSS Qualtrics, etc.
2	Final part of Introduction starting at slide 51 Correlation and graphics	Confidence Intervals vs. "magic astericks" the bootstrap starting at page 28 More on sampling html	Using the objects from a function error.dots, error.bars Reading Code t2d, fisherz corr.test and corPlot corPlotUpperLowerCi and multi.hist	Distributions (html) and 2a.Rmd Handout 2 html rmd psych source code zip or psych source code Handout 2c (homework 1a)html Rmd Issues in data html
3	Scales and Reliability Item Response Theory	Reliability α to ω Reliability theory Why not use α factor analysis advanced notes on Factor Analysis	by head tail headTail splitHalf alpha scoreItems scoreOverlap omega reiiability tetrachoric and polychoric irt.fa and scoreIrt	Handout 3 Rmd Handout 3a Rmd How to use omega Handout 3b Rmd
4	UseRs vs. Programmers Factor analysis	UseR vs. ProgrammeR reliability appendix factor analysis How to do factor analysis	testRetest splitHalf alpha scoreItems scoreOverlap fa fa.diagram	Reliability (html) and Reliability (Rmd) fa Rmd file fa html file
5a 5b	ANOVA and the linear model general linear model	t and F tests The general linear model of 0 centered scores	t.test anova lm lm setCor dummy.code corPlot corCi	Handout 5 The Rmd file 5b html fileThe Rmd file The Rmd file the html file and Rmd file
6 a	More on the linear model	Mediation/Moderation	%in% subset outliers mediation/moderation mediate	data manipulation (html) Rmd Detecting outliers Rmd mediation (html) Rmd matReg setCor.diagram
7	Writing functions Multilevel modeling	More on regression modeling dynamics 3 levels of analysis Final project datasets	lm and setCor multilevel.reliability lattice nlme	programming html and Rmd file mlm html and Rmd file homework answers Final project 8.6.1Homework
8	Writing functions (2) data manipulation	Writing functions Debugging (an example) Scoring scales Test Theory	alpha scoreItems scoreFast table %in% subset merge corPlot matSort irt.fa scoreIrt scoreIrt.2pl	html and Rmd file debugging html Rmd data manipulation html Rmd Reliability Homework - answers
9	Item Response Theory (IRT) Confirmatory Factor Analysis (ÇFA)	Test Theory (continued) More on Reliability Using lavaan	ICC cohen.kappa functions: irt.fa scoreIrt packages: ltm MIRT lavaan	html and Rmd file html and Rmd file
10	data manipulation Review	Advanced programming Review of R	table %in% grep sub order match corPlot matSort dfOrder Sara Weston Tutorial	Advanced programming html Rmd
11 a	Review (continued)	Review of R		Sara Weston Tutorial

8 Detailed Notes

8.1 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 conver-

sation 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.

8.2 Week 2

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

8.2.1 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.

8.2.2 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.

8.2.3 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.

8.2.4 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

8.2.5   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.)

8.3   Week 3

8.3.1   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.

8.3.2   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.

8.4 Week 4b

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

8.5 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.

8.5.1 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.

8.5.2 Week 6

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

8.6 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,

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

8.6.1 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.

8.7 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.

8.8 Week 9

Even more on test theory and reliability.

8.9 Week 10

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

9   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.

10   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 (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.