Psychology 371: Personality Research
Using R in Personality Research

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

June, 2014
Outline

1. What is R and how to get it
   - Installing R on your computer and adding packages
   - Basic R capabilities: Calculation, Statistical tables, Graphics

2. Basic data analysis
   - Reading the data

3. Scale construction and Reliability
   - Classical Test measures of reliability

4. Regression
   - t-test, ANOVA, $\chi^2$
   - Linear Regression

5. More help
   - R: a short guide
   - Scale construction, factor analysis guides
**R: What is it?**

1. **R: An international collaboration**
2. **R: The open source - public domain version of S+**
3. **R: Written by statistician (and all of us) for statisticians (and the rest of us)**
4. **R: Not just a statistics system, also an extensible language.**
   - This means that as new statistics are developed they tend to appear in R far sooner than elsewhere.
   - R facilitates asking questions that have not already been asked.
Statistical Programs for Psychologists

- General purpose programs
  - R
  - S+
  - SAS
  - SPSS
  - STATA
  - Systat

- Specialized programs
  - Mx
  - EQS
  - AMOS
  - LISREL
  - MPlus
  - Your favorite program
Statistical Programs for Psychologists

- General purpose programs
  - R
  - $+$
  - $A$
  - $P$
  - $TATA$
  - $ytat$

- Specialized programs
  - Mx (OpenMx is part of R)
  - EQ$
  - AMO$
  - LI$REL$
  - MP$u$

Your favorite program
“R is the lingua franca of statistical research. Work in all other languages should be discouraged.”

“This is R. There is no if. Only how.”

“Overall, SAS is about 11 years behind R and S-Plus in statistical capabilities (last year it was about 10 years behind) in my estimation.”

Q: My institute has been heavily dependent on SAS for the past while, and SAS is starting to charge us a very deep amount for license renewal.... The team is [considering] switching to R, ... I am talking about the entire institute with considerable number of analysts using SAS their entire career. ... What kind of problems and challenges have you faced?

A: One of your challenges will be that with the increased productivity of the team you will have time for more intellectually challenging problems. That frustrates some people.
R is open source, how can you trust it?

- Q: “When you use it [R], since it is written by so many authors, how do you know that the results are trustable?”
- A: “The R engine [...] is pretty well uniformly excellent code but you have to take my word for that. Actually, you don’t. The whole engine is open source so, if you wish, you can check every line of it. If people were out to push dodgy software, this is not the way they’d go about it.”
- Q: Are R packages bug free?
- A: No. But bugs are fixed rapidly when identified.
- Q: How does function x work? May I adapt it for my functions.
What is R?: Technically

- R is an open source implementation of S (The statistical language developed at Bell Labs). (S-Plus is a commercial implementation)
- R is a language and environment for statistical computing and graphics. R is available under GNU Copy-left
- R is a group project run by a core group of developers (with new releases semiannually). The current version of R is 3.1.0
- R is an integrated suite of software facilities for data manipulation, calculation and graphical display.

(Adapted from Robert Gentleman and the r-project.org web page)
R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It is:

1. an effective data handling and storage facility,
2. a suite of operators for calculations on arrays, in particular matrices,
3. a large, coherent, integrated collection of intermediate tools for data analysis,
4. graphical facilities for data analysis and display either on-screen or on hardcopy, and
5. a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

Many users think of R as a statistics system. We prefer to think of it of an environment within which statistical techniques are implemented. R can be extended (easily) via packages ... available through the CRAN family of Internet sites covering a very wide range of modern statistics.  (Adapted from r-project.org web page)
R: A brief history

- 1991-93: Ross Dhaka and Robert Gentleman begin work on R project for Macs at U. Auckland
- 1995: R available by ftp under the GPL
- 96-97: mailing list and R core group is formed
- 2000: John Chambers, designer of S joins the Rcore (wins a prize for best software from ACM for S)
- 2001-2014: Core team continues to improve base package with a new release every 6 months.
- Many others contribute “packages” to supplement the functionality for particular problems
  - 2003-04-01: 250 packages
  - 2004-10-01: 500 packages
  - 2007-04-12: 1,000 packages
  - 2009-10-04: 2,000 packages
  - 2011-05-12: 3,000 packages
  - 2012-08-27: 4,000 packages
  - 2014-05-16: 5,547 packages (on CRAN) + 824 bioinformatic packages on BioConductor
Rapid and consistent growth in packages contributed to R
http://r4stats.com/articles/popularity/ considers various measures of popularity

1. discussion groups
2. blogs
3. Google Scholar citations (> 14,000 citations, ≈ 1,800/year)
4. Google Page rank
R as a way of facilitating replicable science

1. R is not just for statisticians, it is for all research oriented psychologists.

2. R scripts are published in psychology journals to show new methods:
   - *Psychological Methods*
   - *Psychological Science*
   - *Journal of Research in Personality*

3. R based data sets are now accompanying journal articles:
   - The *Journal of Research in Personality* now accepts R code and data sets.
   - JRP special issue in R is coming this fall.

4. By sharing our code and data the field can increase the possibility of doing replicable science.
Reproducible Research: Sweave and KnitR

Sweave is a tool that allows to embed the R code for complete data analyses in LaTeX documents. The purpose is to create dynamic reports, which can be updated automatically if data or analysis change. Instead of inserting a prefabricated graph or table into the report, the master document contains the R code necessary to obtain it. When run through R, all data analysis output (tables, graphs, etc.) is created on the fly and inserted into a final LaTeX document. The report can be automatically updated if data or analysis change, which allows for truly reproducible research.


Supplementary material for journals can be written in Sweave/KnitR.
Misconception: R is hard to use

1. R doesn’t have a GUI (Graphical User Interface)
   - Partly true, many use syntax.
   - Partly not true, GUIs exist (e.g., R Commander, R-Studio).
   - Quasi GUIs for Mac and PCs make syntax writing easier.

2. R syntax is hard to use
   - Not really, unless you think an iPhone is hard to use.
   - Easier to give instructions of 1-4 lines of syntax rather than pictures of menu after menu to pull down.
   - Keep a copy of your syntax, modify it for the next analysis.

3. R is not user friendly: A personological description of R
   - R is introverted: it will tell you what you want to know if you ask, but not if you don’t ask.
   - R is conscientious: it wants commands to be correct.
   - R is not agreeable: its error messages are at best cryptic.
   - R is stable: it does not break down under stress.
   - R is open: new ideas about statistics are easily developed.
Misconceptions: R is hard to learn – some interesting facts

1. With a brief web based tutorial [http://personality-project.org/r](http://personality-project.org/r), 2nd and 3rd year undergraduates in psychological methods and personality research courses are using R for descriptive and inferential statistics and producing publication quality graphics.

2. More and more psychology departments are using it for graduate and undergraduate instruction.

3. R is easy to learn, hard to master
   - R-help newsgroup is very supportive
   - Multiple web based and pdf tutorials see (e.g., [http://www.r-project.org/](http://www.r-project.org/))
   - Short courses using R for many applications. (Look at APS program).

4. Books and websites for SPSS and SAS users trying to learn R (e.g., [http://r4stats.com/](http://r4stats.com/)) by Bob Muenchen (look for link to free version).
Ok, how do I get it: Getting started with R

- Download from R Cran (http://cran.r-project.org/)
  - Choose appropriate operating system and download compiled R
- Install R (current version is 3.1.0) (See a tutorial on how to install R and various packages at http://personality-project.org/r/psych)
- Start R
- Add useful packages (just need to do this once)
  - install.packages("ctv") #this downloads the task view package
  - library(ctv) #this activates the ctv package
  - install.views("Psychometrics") #among others
  - Take a 5 minute break
- Activate the package(s) you want to use today (e.g., psych)
  - library(psych) #necessary for most of today’s examples
- Use R
Installing R on your computer and adding packages

> install.packages("ctv")

> library(ctv)

> install.views("Psychometrics")

#or just install a few packages
> install.packages("psych")

> install.packages("GPArotation")
> install.packages("MASS")
> install.packages("mvtnorm")

- Install the task view installer package. You might have to choose a “mirror” site.
- Make it active
- Install all the packages in the “Psychometrics” task view. This will take a few minutes.
- Or, just install one package (e.g., psych) as well as a few suggested packages that add functionality for factor rotation, multivariate normal distributions, etc.
The R Project for Statistical Computing

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R, please choose your preferred CRAN mirror.
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News:

- R version 3.1.0 (Spring Dance) has been released on 2014-04-10.
- R version 3.0.3 (Warm Puppy) has been released on 2014-03-06.
- The R Journal Vol.5/2 is available.
- useR! 2013, took place at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013.
- R version 2.15.3 (Security Blanket) has been released on 2013-03-01.
Go to the Comprehensive R Archive Network (CRAN)

Download and Install R

Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!


- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).

- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.

- Source code of older versions of R is available here.

- Contributed extension packages

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the R project homepage for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN mirror nearest to you to minimize network load.
Download and install the appropriate version – PC

R for Windows

Subdirectories:

base

Binaries for base distribution (managed by Duncan Murdoch). This is what you want to install R for the first time.

contrib

Binaries of contributed packages (managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.

Rtools

Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.
Download and install the appropriate version – PC

R-3.1.0 for Windows (32/64 bit)

Download R 3.1.0 for Windows (54 megabytes, 32/64 bit)

Installation and other instructions
New features in this version

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the md5sum of the .exe to the true fingerprint. You will need a version of md5sum for windows: both graphical and command line versions are available.

Frequently asked questions

- How do I install R when using Windows Vista?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.

Other builds

- Patches to this release are incorporated in the r-patched snapshot build.
- A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build.
- Previous releases

Note to webmasters: A stable link which will redirect to the current Windows binary release is <CRAN MIRROR>/bin/windows/base/release.htm.

Last change: 2014-04-11, by Duncan Murdoch
Download and install the appropriate version – Mac

R for Mac OS X

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above), Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) here. Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the old directory.

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

R 3.1.0 "Spring Dance" released on 2014/04/10

This binary distribution of R and the GUI supports 64-bit Intel based Macs on Mac OS X 10.6 (Snow Leopard) or higher.

Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type

md5 R-3.1.0-snowleopard.pkg

in the Terminal application to print the MD5 checksum for the R-3.1.0-snowleopard.pkg image. On Mac OS X 10.7 and later you can also validate the signature using pkgutil --check-signature R-3.1.0-snowleopard.pkg

Files:

R-3.1.0-snowleopard.pkg

R 3.1.0 binary for Mac OS X 10.6 (Snow Leopard) and higher, signed package. Contains R 3.1.0 framework, R.app GUI 1.64 in 64-bit for Intel Macs. The above file is an Installer package which can be installed by double-clicking. Depending on your browser, you may need to press the control key and click on this link to download the file.

This package contains the R framework, 64-bit GUI (R.app) and Tcl/Tk 8.6.0 X11 libraries. The latter component is optional and can be omitted when choosing "custom install", it is only needed if you want to use the talk R package. GNU Fortran is NOT included (needed if you want to compile packages from sources that contain FORTRAN code) please see the tools directory.

R-3.1.0-mavericks.pkg

R 3.1.0 binary for Mac OS X 10.9 (Mavericks) and higher, signed package. It contains the same software versions as above, but this R build has been built with Xcode 5 to leverage new compilers and functionalities in Mavericks not available in earlier OS X versions.

Mac-GUI-1.64.tar.gz

Sources for the Rapp GUI 1.64 for Mac OS X. This file is only needed if you want to join the development of the GUI, it is not intended for regular users. Read the INSTALL file for further instructions.

NEWS (for Mac GUI)

The new Rapp Cocoa GUI has been written by Simon Urbanek and Stefano Iacus with contributions from many developers and translators world-wide, see "About R" in the GUI.

Subdirectories:

tools

Additional tools necessary for building R for Mac OS X:
Universal GNU Fortran compiler for Mac OS X (see R for Mac tools page for details).
Binaries of package builds for Mac OS X 10.6 or higher (Snow Leopard build)
Binaries of package builds for Mac OS X 10.9 or higher (Mavericks build)
Legacy binaries of universal (32-bit) package builds for Mac OS X 10.5 or higher
(Lion build)
Legacy binaries of universal (32-bit) package builds for Mac OS X 10.4 (Tiger build)

contrib

macosx

leopard

universal
Starting R on a PC

R version 3.1.0 (2014-04-10) -- "Spring Dance"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> sessionInfo()
R version 3.1.0 (2014-04-10)
Platform: i386-w64-mingw32/i386 (32-bit)
locale:
[1] LC_COLLATE=English_United_States.1252
[2] LC_CTYPE=English_United_States.1252
Installing a package (psych) on a PC by hand – note error

R version 3.1.0 (2014-04-10) -- "Spring Dance"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> install.packages(psych)
Error in install.packages(psych) : object 'psych' not found
> install.packages("psych")
Installing package into ‘C:/users/revelle/My Documents/R/win-library/3.1’
(as ‘lib’ is unspecified)
--- Please select a CRAN mirror for use in this session ---
trying URL 'http://cran.stat.ucla.edu/bin/windows/contrib/3.1/psych_1.4.5.zip'
Installing packages using the menu

R version 3.1.0 (2014-04-10) -- "Spring Dance"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> install.packages(psy7c)
Error in install.packages(psy7c) : object 'psy7c' not found
> install.packages("psy7c")
Installing package into 'C:/users/revelle/My Documents/R/win-library/3.1'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
trying URL 'http://cran.stat.ucla.edu/bin/windows/contrib/3.1/psy7c_1.4.5.zip'
Content type 'application/zip' length 2928284 bytes (2.8 Mb)
opened URL
downloaded 2.8 Mb
Start up R and get ready to play (Mac Development version)

R Under development (unstable) (2014-04-17 r65403) -- "Unsuffered Consequences"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.1.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.65 (6738) x86_64-apple-darwin13.1.0]

[Workspace restored from /Users/revelle/.RData]
[History restored from /Users/revelle/.Rapp.history]
Check the version number for R (should be $\geq 3.1.0$) and for psych ($\geq 1.4.5$)

```r
> library(psych)  #make the psych package active
> sessionInfo()  #what packages are active

R Under development (unstable) (2014-04-17 r65403)
Platform: x86_64-apple-darwin13.1.0 (64-bit)

locale:

attached base packages:
[1] stats graphics grDevices utils datasets methods base

other attached packages:
[1] psych_1.4.5
>
```
Various ways to run R

1. UNIX (and *NIX like) environments
   - Non interactive
   - Particularly fast if on remote processors
   - RStudio Server as “Integrated Development Environment” (IDE)
   - RStudio can be run remotely with a browser (e.g., even from an IPad)

2. PC
   - quasi GUI + text editor of choice
   - RStudio as “Integrated Development Environment” (IDE)

3. Mac
   - R.app + text editor of choice
   - RStudio as “Integrated Development Environment” (IDE)
R Studio is a useful “Integrated Development Environment” (IDE)
Basic R commands – remember don’t enter the >

R is just a fancy calculator. Add, subtract, sum, products, group

> 2 + 2

[1] 4

> 3^4

[1] 81

> sum(1:10)

[1] 55

> prod(c(1, 2, 3, 5, 7))

[1] 210

It is also a statistics table (the normal distribution, the t distribution)

> pnorm(q = 1)

[1] 0.8413447

> pt(q = 2, df = 20)

[1] 0.9703672
Table: To obtain the density, prefix with $d$, probability with $p$, quantiles with $q$ and to generate random values with $r$. (e.g., the normal distribution may be chosen by using dnorm, pnorm, qnorm, or rnorm.)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>base name</th>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>example application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>norm</td>
<td>mean</td>
<td>sigma</td>
<td></td>
<td>Most data</td>
</tr>
<tr>
<td>Multivariate normal</td>
<td>mvnorm</td>
<td>mean</td>
<td>$r$</td>
<td>sigma</td>
<td>Most data</td>
</tr>
<tr>
<td>Log Normal</td>
<td>lnorm</td>
<td>log mean</td>
<td>log sigma</td>
<td></td>
<td>income or reaction time</td>
</tr>
<tr>
<td>Uniform</td>
<td>unif</td>
<td>min</td>
<td>max</td>
<td></td>
<td>rectangular distributions</td>
</tr>
<tr>
<td>Binomial</td>
<td>binom</td>
<td>size</td>
<td>prob</td>
<td></td>
<td>Bernoulli trials (e.g., coin flips)</td>
</tr>
<tr>
<td>Student’s $t$</td>
<td>t</td>
<td>df</td>
<td></td>
<td>nc</td>
<td>Finding significance of a $t$-test</td>
</tr>
<tr>
<td>Multivariate $t$</td>
<td>mvt</td>
<td>df</td>
<td>corr</td>
<td>nc</td>
<td>Multivariate applications</td>
</tr>
<tr>
<td>Fisher’s $F$</td>
<td>f</td>
<td>df1</td>
<td>df2</td>
<td>nc</td>
<td>Testing for significance of $F$ test</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>chisq</td>
<td>df</td>
<td></td>
<td>nc</td>
<td>Testing for significance of $\chi^2$</td>
</tr>
<tr>
<td>Exponential</td>
<td>exp</td>
<td>rate</td>
<td></td>
<td></td>
<td>Exponential decay</td>
</tr>
<tr>
<td>Gamma</td>
<td>gamma</td>
<td>shape</td>
<td>rate</td>
<td>scale</td>
<td>distribution theory</td>
</tr>
<tr>
<td>Hypergeometric</td>
<td>hyper</td>
<td>m</td>
<td>n</td>
<td>k</td>
<td>Item Response Theory</td>
</tr>
<tr>
<td>Logistic</td>
<td>logis</td>
<td>location</td>
<td>scale</td>
<td></td>
<td>Count data</td>
</tr>
<tr>
<td>Poisson</td>
<td>pois</td>
<td>lambda</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>weibull</td>
<td>shape</td>
<td>scale</td>
<td></td>
<td>Reaction time distributions</td>
</tr>
</tbody>
</table>
Reading data from another program – using the clipboard

1. Read the data in your favorite spreadsheet or text editor
2. Copy to the clipboard
3. Execute the appropriate `read.clipboard` function with or without various options specified

```
my.data <- read.clipboard()  # assumes headers and tab or space delimited
my.data <- read.clipboard.csv()  # assumes headers and comma delimited
my.data <- read.clipboard.tab()  # assumes headers and tab delimited
  (e.g., from Excel)
my.data <- read.clipboard.lower()  # read in a matrix given the lower
my.data <- read.clipboard.upper()  # or upper off diagonal
my.data <- read.clipboard.fwf()  # read in data using a fixed format width
  (see `read.fwf` for instructions)
```

4. `read.clipboard()` has default values for the most common cases and these do not need to be specified. Consult `?read.clipboard` for details.
Perhaps the standard way of reading in data is using the `read` command.

- First must specify the location of the file
- Can either type this in directly or use the `file.choose` function. This goes to your normal system file handler.
- The file name/location can be a remote URL. (Note that `read.file` will not work on https files.)

Two examples of reading data

```r
file.name <- file.choose()  # this opens a window to allow you find the file
# or
data.filename="http://personality-project.org/r/datasets/R.appendix1.data"
my.data <- read.table(data.filename, header=TRUE)  # unless it is https (see above)
# or
data.ex1=read.https(data.filename, header=TRUE)  # read an https file

> dim(data.ex1)  # what are the dimensions of what we read?
[1] 18 2
> describe(data.ex1)  # do the data look right?

var  n  mean   sd  median trimmed  mad  min  max  range   skew  kurtosis   se
Dosage* 1 18 1.89 0.76   2  1.88 1.48   1  3   2  0.16  -1.12  0.18
Alertness 2 18 27.67 6.82  27 27.50 8.15  17 41  24 0.25  -0.68  1.61
```
Put it all together: read, show, describe

datafilename="http://personality-project.org/r/datasets/R.appendix1.data"
data.ex1<- read.table(datafilename,header=TRUE)  #unless it is https (see above)
dim(data.ex1)  #what are the dimensions of what we read?
data.ex1  #show the data
headTail(data.ex1)  #just the top and bottom lines
describe(data.ex1)  #descriptive stats

1. Read the data from a remote file
2. Show all the cases (problematic if there are are 100s – 1000s)
3. Just show the first and last (4) lines
4. Find descriptive statistics

---

Dosage  Alertness
1    a    30
2    a    38
...  (rows deleted by hand)
17   c    20
18   c    19

> headTail(data.ex1)  #just the top and bottom lines

<table>
<thead>
<tr>
<th></th>
<th>Dosage</th>
<th>Alertness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>a</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>a</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>a</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>a</td>
<td>41</td>
</tr>
<tr>
<td>15</td>
<td>c</td>
<td>17</td>
</tr>
<tr>
<td>16</td>
<td>c</td>
<td>21</td>
</tr>
<tr>
<td>17</td>
<td>c</td>
<td>20</td>
</tr>
<tr>
<td>18</td>
<td>c</td>
<td>19</td>
</tr>
</tbody>
</table>

> describe(data.ex1)  #descriptive stats

<table>
<thead>
<tr>
<th>vars</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dosage*</td>
<td>1</td>
<td>18</td>
<td>1.89</td>
<td>0.76</td>
<td>1.88</td>
<td>1.48</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0.16</td>
<td>-1.35</td>
<td>0.18</td>
</tr>
<tr>
<td>Alertness</td>
<td>2</td>
<td>18</td>
<td>27.67</td>
<td>6.82</td>
<td>27</td>
<td>8.15</td>
<td>17</td>
<td>41</td>
<td>24</td>
<td>0.25</td>
<td>-1.06</td>
<td>1.61</td>
</tr>
</tbody>
</table>
Get the data and look at it

Read in some data, look at the first and last few cases (using `headTail`), and then get basic descriptive statistics. For this example, we will use a built in data set.

```r
> headTail(epi.bfi)
```

```
epiE epiS epiImp epilie epiNeur bflagree bfcon bfext bfneur b fopen bdi traitanx stateanx
1   18   10    7   3   9    138   96  141    51   138   1    24   22
2   16    8    5   1  12    101   99  107   116   132    7    41   40
3    6    1    3   2   5    143  118  38    68    90    4    37   44
4   12    6    4   3  15    104  106  64   114   101    8    54   40
... ...     ... ... ... ...     ... ... ... ...     ... ...     ... ... ... ...
228  12    7    4   3  15    155  129  127    88   110    9    35  34
229  19  10    7   2  11    162  152  163   104   164    1    29   47
230    4    1    1   2  10    95  111  75   123   138    5    39  58
231    8    6    3   2  15    85  62  90   131    96   24    58  58
```

epi.bfi has 231 cases from two personality measures.
Now find the descriptive statistics for this data set

```r
> describe(epi.bfi)
```

<table>
<thead>
<tr>
<th>var</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>epiE</td>
<td>1</td>
<td>231</td>
<td>13.33</td>
<td>4.14</td>
<td>14</td>
<td>13.49</td>
<td>4.45</td>
<td>1</td>
<td>22</td>
<td>-0.33</td>
<td>-0.01</td>
<td>0.27</td>
</tr>
<tr>
<td>epiS</td>
<td>2</td>
<td>231</td>
<td>7.58</td>
<td>2.69</td>
<td>8</td>
<td>7.77</td>
<td>2.97</td>
<td>0</td>
<td>13</td>
<td>-0.57</td>
<td>0.04</td>
<td>0.18</td>
</tr>
<tr>
<td>epiImp</td>
<td>3</td>
<td>231</td>
<td>4.37</td>
<td>1.88</td>
<td>4</td>
<td>4.36</td>
<td>1.48</td>
<td>0</td>
<td>9</td>
<td>0.06</td>
<td>-0.59</td>
<td>0.12</td>
</tr>
<tr>
<td>epilie</td>
<td>4</td>
<td>231</td>
<td>2.38</td>
<td>1.50</td>
<td>2</td>
<td>2.27</td>
<td>1.48</td>
<td>0</td>
<td>7</td>
<td>0.66</td>
<td>0.30</td>
<td>0.10</td>
</tr>
<tr>
<td>epiNeur</td>
<td>5</td>
<td>231</td>
<td>10.41</td>
<td>4.90</td>
<td>10</td>
<td>10.39</td>
<td>4.45</td>
<td>0</td>
<td>23</td>
<td>0.06</td>
<td>-0.46</td>
<td>0.32</td>
</tr>
<tr>
<td>bfaagree</td>
<td>6</td>
<td>231</td>
<td>125.00</td>
<td>18.14</td>
<td>126</td>
<td>125.26</td>
<td>17.79</td>
<td>74</td>
<td>167</td>
<td>93</td>
<td>-0.21</td>
<td>-0.22</td>
</tr>
<tr>
<td>bfcon</td>
<td>7</td>
<td>231</td>
<td>113.25</td>
<td>21.88</td>
<td>114</td>
<td>113.42</td>
<td>22.24</td>
<td>53</td>
<td>178</td>
<td>125</td>
<td>-0.02</td>
<td>0.29</td>
</tr>
<tr>
<td>bfext</td>
<td>8</td>
<td>231</td>
<td>102.18</td>
<td>26.46</td>
<td>104</td>
<td>102.99</td>
<td>22.24</td>
<td>8</td>
<td>168</td>
<td>160</td>
<td>-0.41</td>
<td>0.58</td>
</tr>
<tr>
<td>bfneur</td>
<td>9</td>
<td>231</td>
<td>87.97</td>
<td>23.34</td>
<td>90</td>
<td>87.70</td>
<td>23.72</td>
<td>34</td>
<td>152</td>
<td>118</td>
<td>0.07</td>
<td>-0.51</td>
</tr>
<tr>
<td>bfopen</td>
<td>10</td>
<td>231</td>
<td>123.43</td>
<td>20.51</td>
<td>125</td>
<td>123.78</td>
<td>20.76</td>
<td>73</td>
<td>173</td>
<td>100</td>
<td>-0.16</td>
<td>-0.11</td>
</tr>
<tr>
<td>bdi</td>
<td>11</td>
<td>231</td>
<td>6.78</td>
<td>5.78</td>
<td>6</td>
<td>5.97</td>
<td>4.45</td>
<td>0</td>
<td>27</td>
<td>1.29</td>
<td>1.60</td>
<td>0.38</td>
</tr>
<tr>
<td>traitanx</td>
<td>12</td>
<td>231</td>
<td>39.01</td>
<td>9.52</td>
<td>38</td>
<td>38.36</td>
<td>8.90</td>
<td>22</td>
<td>71</td>
<td>49</td>
<td>0.67</td>
<td>0.54</td>
</tr>
<tr>
<td>stateanx</td>
<td>13</td>
<td>231</td>
<td>39.85</td>
<td>11.48</td>
<td>38</td>
<td>38.92</td>
<td>10.38</td>
<td>21</td>
<td>79</td>
<td>58</td>
<td>0.72</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Boxplots are a convenient descriptive device

Show the Tukey “boxplot” for the Eysenck Personality Inventory

Use the box plot function

```r
my.data <- epi.bfi
boxplot(my.data[1:5])
```
An alternative display is a 'violin' plot (available as `violinBy`)

```r
violinBy(my.data[1:5])
```
Plot the scatter plot matrix (SPLOM) of the first 5 variables using the `pairs.panels` function

```
Use the `pairs.panels` function from `psych`

```

```
pairs.panels(my.data[1:5])
```
Plot the scatter plot matrix (SPLOM) of the first 5 variables using the `pairs.panels` function but with smaller `pch` and jittering the points.

Use the `pairs.panels` function from `psych`

```r
pairs.panels(my.data[1:5], pch='.', jiggle=TRUE)
```
> round(cor(my.data, use = "pairwise"), 2)

<table>
<thead>
<tr>
<th></th>
<th>epiE</th>
<th>epiS</th>
<th>epiImp</th>
<th>epilie</th>
<th>epiNeur</th>
<th>bfagree</th>
<th>bfcon</th>
<th>bfext</th>
<th>bfneur</th>
<th>bfopen</th>
<th>bdi</th>
<th>traitanx</th>
<th>stateanx</th>
</tr>
</thead>
<tbody>
<tr>
<td>epiE</td>
<td>1.00</td>
<td>0.85</td>
<td>0.80</td>
<td>-0.22</td>
<td>-0.18</td>
<td>0.18</td>
<td>-0.11</td>
<td>0.54</td>
<td>-0.09</td>
<td>0.14</td>
<td>-0.16</td>
<td>-0.23</td>
<td>-0.13</td>
</tr>
<tr>
<td>epiS</td>
<td>0.85</td>
<td>1.00</td>
<td>0.43</td>
<td>-0.05</td>
<td>-0.22</td>
<td>0.20</td>
<td>0.05</td>
<td>0.58</td>
<td>-0.07</td>
<td>0.15</td>
<td>-0.13</td>
<td>-0.26</td>
<td>-0.12</td>
</tr>
<tr>
<td>epiImp</td>
<td>0.80</td>
<td>0.43</td>
<td>1.00</td>
<td>-0.24</td>
<td>-0.07</td>
<td>0.08</td>
<td>-0.24</td>
<td>0.35</td>
<td>-0.09</td>
<td>0.07</td>
<td>-0.11</td>
<td>-0.12</td>
<td>-0.09</td>
</tr>
<tr>
<td>epilie</td>
<td>-0.22</td>
<td>-0.05</td>
<td>-0.24</td>
<td>1.00</td>
<td>-0.25</td>
<td>0.17</td>
<td>0.23</td>
<td>-0.04</td>
<td>-0.22</td>
<td>-0.03</td>
<td>-0.20</td>
<td>-0.23</td>
<td>-0.15</td>
</tr>
<tr>
<td>epiNeur</td>
<td>-0.18</td>
<td>-0.22</td>
<td>-0.07</td>
<td>-0.25</td>
<td>1.00</td>
<td>-0.08</td>
<td>-0.13</td>
<td>-0.17</td>
<td>0.63</td>
<td>0.09</td>
<td>0.58</td>
<td>0.73</td>
<td>0.49</td>
</tr>
<tr>
<td>bfagree</td>
<td>0.18</td>
<td>0.20</td>
<td>0.08</td>
<td>0.17</td>
<td>-0.08</td>
<td>1.00</td>
<td>0.45</td>
<td>0.48</td>
<td>-0.04</td>
<td>0.39</td>
<td>-0.14</td>
<td>-0.31</td>
<td>-0.19</td>
</tr>
<tr>
<td>bfcon</td>
<td>-0.11</td>
<td>0.05</td>
<td>-0.24</td>
<td>0.23</td>
<td>-0.13</td>
<td>0.45</td>
<td>1.00</td>
<td>0.27</td>
<td>0.04</td>
<td>0.31</td>
<td>-0.18</td>
<td>-0.29</td>
<td>-0.14</td>
</tr>
<tr>
<td>bfext</td>
<td>0.54</td>
<td>0.58</td>
<td>0.35</td>
<td>-0.04</td>
<td>-0.17</td>
<td>0.48</td>
<td>0.27</td>
<td>1.00</td>
<td>0.04</td>
<td>0.46</td>
<td>-0.14</td>
<td>-0.39</td>
<td>-0.15</td>
</tr>
<tr>
<td>bfneur</td>
<td>-0.09</td>
<td>-0.07</td>
<td>-0.09</td>
<td>-0.22</td>
<td>0.63</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>1.00</td>
<td>0.29</td>
<td>0.47</td>
<td>0.59</td>
<td>0.49</td>
</tr>
<tr>
<td>bfopen</td>
<td>0.14</td>
<td>0.15</td>
<td>0.07</td>
<td>-0.03</td>
<td>0.09</td>
<td>0.39</td>
<td>0.31</td>
<td>0.46</td>
<td>0.29</td>
<td>1.00</td>
<td>-0.08</td>
<td>-0.11</td>
<td>-0.04</td>
</tr>
<tr>
<td>bdi</td>
<td>-0.16</td>
<td>-0.13</td>
<td>-0.11</td>
<td>-0.20</td>
<td>0.58</td>
<td>-0.14</td>
<td>-0.18</td>
<td>-0.14</td>
<td>0.47</td>
<td>-0.08</td>
<td>1.00</td>
<td>0.65</td>
<td>0.61</td>
</tr>
<tr>
<td>traitanx</td>
<td>-0.23</td>
<td>-0.26</td>
<td>-0.12</td>
<td>-0.23</td>
<td>0.73</td>
<td>-0.31</td>
<td>-0.29</td>
<td>-0.39</td>
<td>0.59</td>
<td>-0.11</td>
<td>0.65</td>
<td>1.00</td>
<td>0.57</td>
</tr>
<tr>
<td>stateanx</td>
<td>-0.13</td>
<td>-0.12</td>
<td>-0.09</td>
<td>-0.15</td>
<td>0.49</td>
<td>-0.19</td>
<td>-0.14</td>
<td>-0.15</td>
<td>0.49</td>
<td>-0.04</td>
<td>0.61</td>
<td>0.57</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Find the correlations for this data set, round off to 2 decimal places using `lowerCor`

```r
> lowerCor(my.data)

          epiE  epiS  epImp epili epiNr bfagr bfcon bfext bfner bfopn bdi trtnx sttnx
epiE  1.00  
epiS  0.85  1.00  
epImp 0.80  0.43  1.00  
epilie-0.22 -0.05 -0.24  1.00  
epiNeur-0.18 -0.22 -0.07 -0.25  1.00  
bfagree  0.18  0.20  0.08  0.17 -0.08  1.00  
bfcon 0.11  0.05 -0.24  0.23 -0.13  0.45  1.00  
bfext  0.54  0.58  0.35 -0.04 -0.17  0.48  0.27  1.00  
bfneur -0.09 -0.07 -0.09 -0.22  0.63 -0.04  0.04  0.04  1.00  
bfopen  0.14  0.15  0.07 -0.03  0.09  0.39  0.31  0.46  0.29  1.00  
bdi  -0.16 -0.13 -0.11 -0.20  0.58 -0.14 -0.18 -0.14  0.47 -0.08  1.00  
traitanx-0.23 -0.26 -0.12 -0.23  0.73 -0.31 -0.29 -0.39  0.59 -0.11  0.65  1.00  
stateanx-0.13 -0.12 -0.09 -0.15  0.49 -0.19 -0.14 -0.15  0.49 -0.04  0.61  0.57  1.00  
```
Test the significance and use Holm correction for multiple tests

```r
> corr.test(my.data)
Call: corr.test(x = my.data)
Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>epiE</th>
<th>epiS</th>
<th>epiImp</th>
<th>epilie</th>
<th>epiNeur</th>
<th>bfagree</th>
<th>bfcon</th>
<th>bfext</th>
<th>bfneur</th>
<th>bfopen</th>
<th>bdi</th>
<th>traitanx</th>
<th>stateanx</th>
</tr>
</thead>
<tbody>
<tr>
<td>epiE</td>
<td>1.00</td>
<td>0.85</td>
<td>0.80</td>
<td>-0.22</td>
<td>-0.18</td>
<td>0.18</td>
<td>-0.11</td>
<td>0.54</td>
<td>-0.09</td>
<td>0.14</td>
<td>-0.16</td>
<td>-0.23</td>
<td>-0.13</td>
</tr>
<tr>
<td>epiS</td>
<td>0.85</td>
<td>1.00</td>
<td>0.43</td>
<td>-0.05</td>
<td>-0.22</td>
<td>0.20</td>
<td>0.05</td>
<td>0.58</td>
<td>-0.07</td>
<td>0.15</td>
<td>-0.13</td>
<td>-0.26</td>
<td>-0.12</td>
</tr>
<tr>
<td>epiImp</td>
<td>0.80</td>
<td>0.43</td>
<td>1.00</td>
<td>-0.24</td>
<td>-0.07</td>
<td>0.08</td>
<td>-0.24</td>
<td>0.35</td>
<td>-0.09</td>
<td>0.07</td>
<td>-0.11</td>
<td>-0.12</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stateanx</td>
<td>-0.13</td>
<td>-0.12</td>
<td>-0.09</td>
<td>-0.15</td>
<td>0.49</td>
<td>-0.19</td>
<td>-0.14</td>
<td>-0.15</td>
<td>0.49</td>
<td>-0.04</td>
<td>0.61</td>
<td>0.57</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Sample Size

<table>
<thead>
<tr>
<th></th>
<th>epiE</th>
<th>epiS</th>
<th>epiImp</th>
<th>epilie</th>
<th>epiNeur</th>
<th>bfagree</th>
<th>bfcon</th>
<th>bfext</th>
<th>bfneur</th>
<th>bfopen</th>
<th>bdi</th>
<th>traitanx</th>
<th>stateanx</th>
</tr>
</thead>
<tbody>
<tr>
<td>epiE</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stateanx</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
<td>231</td>
</tr>
</tbody>
</table>

Probability values (Entries above the diagonal are adjusted for multiple tests.)

<table>
<thead>
<tr>
<th></th>
<th>epiE</th>
<th>epiS</th>
<th>epiImp</th>
<th>epilie</th>
<th>epiNeur</th>
<th>bfagree</th>
<th>bfcon</th>
<th>bfext</th>
<th>bfneur</th>
<th>bfopen</th>
<th>bdi</th>
<th>traitanx</th>
<th>stateanx</th>
</tr>
</thead>
<tbody>
<tr>
<td>epiE</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.27</td>
<td>0.27</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.59</td>
<td>0.02</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>epiS</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.04</td>
<td>0.08</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.62</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>epiImp</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>0.01</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>epilie</td>
<td>0.00</td>
<td>0.43</td>
<td>0.00</td>
<td>0.01</td>
<td>0.32</td>
<td>0.03</td>
<td>1.00</td>
<td>0.03</td>
<td>1.00</td>
<td>0.08</td>
<td>0.02</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>epiNeur</td>
<td>0.01</td>
<td>0.00</td>
<td>0.26</td>
<td>0.00</td>
<td>1.00</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>bfagree</td>
<td>0.01</td>
<td>0.23</td>
<td>0.01</td>
<td>0.21</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.95</td>
<td>0.00</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>bfcon</td>
<td>0.08</td>
<td>0.48</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>bfext</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.99</td>
<td>0.00</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>bfneur</td>
<td>0.15</td>
<td>0.30</td>
<td>0.18</td>
<td>0.00</td>
<td>0.50</td>
<td>0.50</td>
<td>0.57</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>bfopen</td>
<td>0.04</td>
<td>0.02</td>
<td>0.30</td>
<td>0.70</td>
<td>0.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>bdi</td>
<td>0.02</td>
<td>0.04</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>traitanx</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>stateanx</td>
<td>0.05</td>
<td>0.07</td>
<td>0.18</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.04</td>
<td>0.02</td>
<td>0.00</td>
<td>0.52</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
>
Psychometrics

1. Classical test theory measures of reliability
   - Scoring tests
   - Reliability (alpha, beta, omega)

2. Multivariate Analysis
   - Factor Analysis
   - Components analysis
   - Multidimensional scaling
   - Structural Equation Modeling

3. Item Response Theory
   - One parameter (Rasch) models
   - 2PL and 2PN models
1 Scoring tests

scoreItems Score 1 ... n scales using a set of keys and finding the simple sum or average of items. Reversed items are indicated by -1

score.multiple.choice Score multiple choice items by first converting to 0 or 1 and then proceeding to score the items.

2 Alternative estimates of reliability

alpha $\alpha$ reliability of a single scale finds the average split half reliability. (some items may be reversed keyed).

omega $\omega_h$ reliability of a single scale estimates the general factor saturation of the test.

guttman Find the 6 Guttman reliability estimates

splitHalf Find the range of split half reliabilities
# Finding coefficient $\alpha$ for a scale (see Revelle and Zinbarg, 2009, however, for why you should not)

Reliability analysis
Call: alpha(x = ability)

<table>
<thead>
<tr>
<th>raw_alpha</th>
<th>std.alpha</th>
<th>G6(smc)</th>
<th>average_r</th>
<th>S/N</th>
<th>ase</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.83</td>
<td>0.83</td>
<td>0.84</td>
<td>0.23</td>
<td>4.9</td>
<td>0.0086</td>
<td>0.51</td>
<td>0.25</td>
</tr>
</tbody>
</table>

lower alpha upper 95% confidence boundaries
0.81 0.83 0.85

Reliability if an item is dropped:

<table>
<thead>
<tr>
<th>raw_alpha</th>
<th>std.alpha</th>
<th>G6(smc)</th>
<th>average_r</th>
<th>S/N</th>
<th>alpha</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>reason.4</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.23</td>
<td>4.5</td>
<td>0.0093</td>
</tr>
<tr>
<td>reason.16</td>
<td>0.82</td>
<td>0.82</td>
<td>0.83</td>
<td>0.24</td>
<td>4.7</td>
<td>0.0091</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rotate.6</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.23</td>
<td>4.5</td>
<td>0.0092</td>
</tr>
<tr>
<td>rotate.8</td>
<td>0.82</td>
<td>0.82</td>
<td>0.83</td>
<td>0.24</td>
<td>4.6</td>
<td>0.0091</td>
</tr>
</tbody>
</table>

Item statistics

<table>
<thead>
<tr>
<th>n</th>
<th>r</th>
<th>r.cor</th>
<th>r.drop</th>
<th>mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>reason.4</td>
<td>1442</td>
<td>0.58</td>
<td>0.54</td>
<td>0.50</td>
<td>0.68</td>
</tr>
<tr>
<td>reason.16</td>
<td>1463</td>
<td>0.50</td>
<td>0.44</td>
<td>0.41</td>
<td>0.73</td>
</tr>
<tr>
<td>r...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Using `scoreItems` to score 25 Big 5 items (taken from the bfi example)

```r
> keys.list <- list(Agree=c(-1,2:5),Conscientious=c(6:8,-9,-10),Extraversion=c(-11,-12,13:15), Neuroticism=c(16:20),Openness = c(21,-22,23,24,-25))
> keys <- make.keys(bfi,keys.list)
> scores <- scoreItems(keys,bfi)

Call: score.items(keys = keys, items = bfi)

(Unstandardized) Alpha:

Agree Conscientious Extraversion Neuroticism Openness
alpha 0.7 0.72 0.76 0.81 0.6

Average item correlation:

Agree Conscientious Extraversion Neuroticism Openness
average.r 0.32 0.34 0.39 0.46 0.23

Guttman 6* reliability:

Agree Conscientious Extraversion Neuroticism Openness
Lambda.6 0.7 0.72 0.76 0.81 0.6

Scale intercorrelations corrected for attenuation
raw correlations below the diagonal, alpha on the diagonal
corrected correlations above the diagonal:

Agree Conscientious Extraversion Neuroticism Openness
Agree 0.70 0.36 0.63 -0.245 0.23
Conscientious 0.26 0.72 0.35 -0.305 0.30
Extraversion 0.46 0.26 0.76 -0.284 0.32
Neuroticism -0.18 -0.23 -0.22 0.812 -0.12
Openness 0.15 0.19 0.22 -0.086 0.60
...
### Item by scale correlations:

**corrected for item overlap and scale reliability**

<table>
<thead>
<tr>
<th></th>
<th>Agree</th>
<th>Conscientious</th>
<th>Extraversion</th>
<th>Neuroticism</th>
<th>Openness</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>-0.40</td>
<td>-0.06</td>
<td>-0.11</td>
<td>0.14</td>
<td>-0.14</td>
</tr>
<tr>
<td>A2</td>
<td>0.67</td>
<td>0.23</td>
<td>0.40</td>
<td>-0.07</td>
<td>0.17</td>
</tr>
<tr>
<td>A3</td>
<td>0.70</td>
<td>0.22</td>
<td>0.48</td>
<td>-0.11</td>
<td>0.17</td>
</tr>
<tr>
<td>A4</td>
<td>0.49</td>
<td>0.29</td>
<td>0.30</td>
<td>-0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>A5</td>
<td>0.62</td>
<td>0.23</td>
<td>0.55</td>
<td>-0.23</td>
<td>0.18</td>
</tr>
<tr>
<td>C1</td>
<td>0.13</td>
<td>0.53</td>
<td>0.19</td>
<td>-0.08</td>
<td>0.28</td>
</tr>
<tr>
<td>C2</td>
<td>0.21</td>
<td>0.61</td>
<td>0.17</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td>C3</td>
<td>0.21</td>
<td>0.54</td>
<td>0.14</td>
<td>-0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>C4</td>
<td>-0.24</td>
<td>-0.66</td>
<td>-0.23</td>
<td>0.31</td>
<td>-0.23</td>
</tr>
<tr>
<td>C5</td>
<td>-0.26</td>
<td>-0.59</td>
<td>-0.29</td>
<td>0.36</td>
<td>-0.10</td>
</tr>
<tr>
<td>E1</td>
<td>-0.30</td>
<td>-0.06</td>
<td>-0.59</td>
<td>0.11</td>
<td>-0.16</td>
</tr>
<tr>
<td>E2</td>
<td>-0.39</td>
<td>-0.25</td>
<td>-0.70</td>
<td>0.34</td>
<td>-0.15</td>
</tr>
<tr>
<td>E3</td>
<td>0.44</td>
<td>0.20</td>
<td>0.60</td>
<td>-0.10</td>
<td>0.37</td>
</tr>
<tr>
<td>E4</td>
<td>0.51</td>
<td>0.23</td>
<td>0.68</td>
<td>-0.22</td>
<td>0.04</td>
</tr>
<tr>
<td>E5</td>
<td>0.34</td>
<td>0.40</td>
<td>0.55</td>
<td>-0.10</td>
<td>0.31</td>
</tr>
<tr>
<td>N1</td>
<td>-0.22</td>
<td>-0.21</td>
<td>-0.11</td>
<td>0.76</td>
<td>-0.12</td>
</tr>
<tr>
<td>N2</td>
<td>-0.22</td>
<td>-0.19</td>
<td>-0.12</td>
<td>0.74</td>
<td>-0.06</td>
</tr>
<tr>
<td>N3</td>
<td>-0.14</td>
<td>-0.20</td>
<td>-0.14</td>
<td>0.74</td>
<td>-0.03</td>
</tr>
<tr>
<td>N4</td>
<td>-0.22</td>
<td>-0.30</td>
<td>-0.39</td>
<td>0.62</td>
<td>-0.02</td>
</tr>
<tr>
<td>N5</td>
<td>-0.04</td>
<td>-0.14</td>
<td>-0.19</td>
<td>0.55</td>
<td>-0.18</td>
</tr>
<tr>
<td>O1</td>
<td>0.16</td>
<td>0.20</td>
<td>0.31</td>
<td>-0.09</td>
<td>0.52</td>
</tr>
<tr>
<td>O2</td>
<td>-0.01</td>
<td>-0.18</td>
<td>-0.07</td>
<td>0.19</td>
<td>-0.45</td>
</tr>
<tr>
<td>O3</td>
<td>0.26</td>
<td>0.20</td>
<td>0.42</td>
<td>-0.07</td>
<td>0.61</td>
</tr>
<tr>
<td>O4</td>
<td>0.06</td>
<td>-0.02</td>
<td>-0.10</td>
<td>0.21</td>
<td>0.32</td>
</tr>
<tr>
<td>O5</td>
<td>-0.09</td>
<td>-0.14</td>
<td>-0.11</td>
<td>0.11</td>
<td>-0.53</td>
</tr>
<tr>
<td>gender</td>
<td>0.25</td>
<td>0.11</td>
<td>0.12</td>
<td>0.14</td>
<td>-0.07</td>
</tr>
<tr>
<td>education</td>
<td>0.06</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>age</td>
<td>0.22</td>
<td>0.14</td>
<td>0.07</td>
<td>-0.13</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Analysis of Variance

1. aov is designed for balanced designs, and the results can be hard to interpret without balance: beware that missing values in the response(s) will likely lose the balance.

2. If there are two or more error strata, the methods used are statistically inefficient without balance, and it may be better to use lme in package nlme.

```r
datafilename = "https://personality-project.org/r/datasets/R.appendix2.data"
data.ex2 = read.csv(datafilename, header = T)  # read the data into a table

# show the data
```

<table>
<thead>
<tr>
<th>Observation</th>
<th>Gender</th>
<th>Dosage</th>
<th>Alertness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>m</td>
<td>a</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>m</td>
<td>a</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>m</td>
<td>a</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>m</td>
<td>a</td>
<td>12</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>f</td>
<td>b</td>
<td>12</td>
</tr>
<tr>
<td>15</td>
<td>f</td>
<td>b</td>
<td>18</td>
</tr>
<tr>
<td>16</td>
<td>f</td>
<td>b</td>
<td>22</td>
</tr>
</tbody>
</table>
Do the analysis of variances and the show the table of results.

```r
aov.ex2 = aov(Alertness~Gender*Dosage,data=data.ex2) #do the analysis of variance
summary(aov.ex2) #show the summary table

> aov.ex2 = aov(Alertness~Gender*Dosage,data=data.ex2) #do the analysis of variance
> summary(aov.ex2) #show the summary table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1</td>
<td>76.562</td>
<td>76.562</td>
<td>2.9518</td>
<td>0.1115</td>
</tr>
<tr>
<td>Dosage</td>
<td>1</td>
<td>5.062</td>
<td>5.062</td>
<td>0.1952</td>
<td>0.6665</td>
</tr>
<tr>
<td>Gender:Dosage</td>
<td>1</td>
<td>0.063</td>
<td>0.063</td>
<td>0.0024</td>
<td>0.9617</td>
</tr>
</tbody>
</table>
```
```r
> print(model.tables(aov.ex2,"means"),digits=3)

Residuals    12 311.250  25.938

Tables of means
Grand mean

14.0625

  Gender
Gender
  f  m
16.25 11.88

  Dosage
Dosage
  a  b
13.50 14.62

  Gender:Dosage
    Dosage
Gender a  b
  f 15.75 16.75
  m 11.25 12.50
```
Somewhat more complicated because we need to convert “wide” data.frames to “long” or “narrow” data.frame.

This can be done by using the stack function. Some data sets are already in the long format.

A detailed discussion of how to work with repeated measures designs is at http://personality-project.org/r/r.anova.html and at http://personality-project.org/r

See also the tutorial by Jason French at http://jason-french.com/tutorials/repeatedmeasures.html
Analysis of variance within subjects

```r
> datafilename="http://personality-project.org/r/datasets/R.appendix5.data"
> data.ex5=read.table(datafilename,header=T) #read the data into a table
> #data.ex5 #show the data
> aov.ex5 =
+ aov(Recall~(Task*Valence*Gender*Dosage)+Error(Subject/(Task*Valence)))+
+ (Gender*Dosage),data.ex5)
> summary(aov.ex5)

Error: Subject

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1</td>
<td>542.26</td>
<td>542.26</td>
<td>5.6853</td>
</tr>
<tr>
<td>Dosage</td>
<td>2</td>
<td>694.91</td>
<td>347.45</td>
<td>3.6429</td>
</tr>
<tr>
<td>Gender:Dosage</td>
<td>2</td>
<td>70.80</td>
<td>35.40</td>
<td>0.3711</td>
</tr>
<tr>
<td>Residuals</td>
<td>12</td>
<td>1144.56</td>
<td>95.38</td>
<td></td>
</tr>
</tbody>
</table>

Signif. codes: 0 ^O***~O 0.001 ^O**~O 0.01 ^O*~O 0.05 ^O.~O 0.1 ^O ~O 1

Error: Subject:Task

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>1</td>
<td>96.333</td>
<td>96.333</td>
<td>39.8621</td>
</tr>
<tr>
<td>Task:Gender</td>
<td>1</td>
<td>1.333</td>
<td>1.333</td>
<td>0.5517</td>
</tr>
<tr>
<td>Task:Dosage</td>
<td>2</td>
<td>8.167</td>
<td>4.083</td>
<td>1.6897</td>
</tr>
<tr>
<td>Task:Gender:Dosage</td>
<td>2</td>
<td>3.167</td>
<td>1.583</td>
<td>0.6552</td>
</tr>
<tr>
<td>Residuals</td>
<td>12</td>
<td>29.000</td>
<td>2.417</td>
<td></td>
</tr>
</tbody>
</table>

... (lots more)
Multiple regression

1. Use the `sat.act` data set from `psych`
2. Do the linear model
3. Summarize the results

```r
mod1 <- lm(SATV ~ education + gender + SATQ, data=sat.act)
> summary(mod1, digits=2)
```

**Call:**
```r
lm(formula = SATV ~ education + gender + SATQ, data = sat.act)
```

**Residuals:**
```
Min 1Q Median 3Q Max
-372.91 -49.08 2.30 53.68 251.93
```

**Coefficients:**
```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 180.87348  23.41019   7.726  3.96e-14 ***
education    1.24043   2.32361   0.534   0.59363 
gender      20.69271   6.99651   2.958   0.00321 **
SATQ         0.64489   0.02891  22.309  < 2e-16 ***
```

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 86.24 on 683 degrees of freedom
(13 observations deleted due to missingness)

Multiple R-squared:  0.4231, Adjusted R-squared:  0.4205
F-statistic:  167 on 3 and 683 DF,  p-value: < 2.2e-16
Zero center the data before examining interactions

```r
> zsat <- data.frame(scale(sat.act,scale=FALSE))
> mod2 <- lm(SATV ~ education * gender * SATQ,data=zsat)
> summary(mod2)

Call:
lm(formula = SATV ~ education * gender * SATQ, data = zsat)

Residuals:
   Min     1Q Median     3Q    Max
-372.53  -48.76    3.33    51.24   238.50

Coefficients:
                         Estimate Std. Error t value  Pr(>|t|)  
(Intercept)               0.773576   3.304938   0.234   0.81500  
education                 2.517314   2.337889   1.077   0.28198  
gender                    18.485906   6.964694   2.654   0.00814 **  
SATQ                      0.620527   0.028925  21.453  < 2e-16 ***  
education:gender          1.249926   4.759374   0.263   0.79292  
education:SATQ            -0.101444   0.020100  -5.047  5.77e-07 ***  
gender:SATQ               0.007339   0.060850   0.121   0.90404  
education:gender:SATQ     0.035822   0.041192   0.870   0.38481  

---
Signif. codes:  < 0.001 ***  0.001 **  0.01 *  0.05 .  0.1  
```
Compare model 1 and model 2

Test the difference between the two linear models

> anova(mod1, mod2)

Analysis of Variance Table

Model 1: SATV ~ education + gender + SATQ
Model 2: SATV ~ education * gender * SATQ

<table>
<thead>
<tr>
<th>Res.Df</th>
<th>RSS</th>
<th>Df</th>
<th>Sum of Sq</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>683</td>
<td>5079984</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>679</td>
<td>4870243</td>
<td>4</td>
<td>209742</td>
<td>7.3104</td>
</tr>
</tbody>
</table>

Signif. codes:  0 ^O***®  0.001 ^O**®  0.01 ^O*®  0.05 ^O.®  0.1 ^O
Show the regression lines by gender

Verbal varies by Quant and gender

```r
> with(sat.act, plot(SATV ~ SATQ, col=c("blue", "red") [gender]))
> by(sat.act, sat.act$gender, function(x) abline
   (lm(SATV ~ SATQ, data=x),
    lty=c("solid", "dashed"))
> title("Verbal varies by Quant and gender")
```
http://personality-project.org/r/ A short guide to R

- Examples that take you through much of what you need

2 The `psych` package

- Most of the stats that you will need
- Remember to make psych active by `library(psych)` at the start of each session.
“How to”

1. Installing R and the *psych* package
   - [http://personality-project.org/r/psych/HowTo/getting_started.pdf](http://personality-project.org/r/psych/HowTo/getting_started.pdf)

2. Using the *psych* package to score personality scales
   - [http://personality-project.org/r/psych/HowTo/scoring.pdf](http://personality-project.org/r/psych/HowTo/scoring.pdf)

3. Using the *psych* package for factor analysis
   - [http://personality-project.org/r/psych/HowTo/factor.pdf](http://personality-project.org/r/psych/HowTo/factor.pdf)