

Using R to analyze a simple data set

Katharine Funkhouser
Psychology Research Methods: Fall, 2007

Abstract

Using R to analyze data from a psychology study such as the 205 project 2 is simpler than it seems. This “how to” guide addresses the use of R as seen by an undergraduate just learning R

Introduction

R may seem intimidating at first because of its lack of a GUI interface (a GUI interface is the box that pops up in Excel or SPSS and asks what data is the independent variable, dependent variable, what kind of test will be run, etc). These steps must be typed directly into R. This might seem new or foreign, but it is really no different from typing numbers directly into a TI graphing calculator. It might be helpful to think of R as a (free!) graphing calculator on your computer that is especially useful for statistics.

Getting R

1. Download and install R. You can download R from <http://cran.r-project.org>. Install and open R.
2. Install the **psych** package.
 1. On a PC, look click on the packages option on the toolbar and select install packages: Select a mirror (one in the US: e.g. <http://www.stathy.com/cran>) and double click on it. Next, a tall, skinny menu will pop up with a list of packages. Select **psych** and install it.
 2. On a Mac, go to the “package installer” in the “Packages & Data” menu and double click. A list of potential “mirrors” will appear. Choose one nearby (e.g., e.g. <http://www.stathy.com/cran>) and then “Get list” and search for “psych”. (Using the “Q” search in the R Package Installer makes this easier, just type in psych in the “Q” bar and you will be shown the most recent version of **psych**. Click on it.

Data analysis with R

In the following list, R commands are preceded by a “>”. Do not enter this, just the command. In addition, the lines are commented by adding a “#” sign.

1. Make the **psych** package active:

```
> library{psych} #this loads the psych package
```

2. Load the data into R by copying your data into the clipboard and then reading the clipboard into R.

```
> sim.data <- read.clipboard() #if you do not have data lables, then you should say  
> sim.data <- read.clipboard(header=FALSE)
```

3. Find basic descriptive statistics for the data using the **describe** function.

```
>describe(sim.data)
```

4. For a basic graphic view of your data, use the **pairs.panels** function.

```
> pairs.panels(sim.data) #show a SPLOM of all the data. or  
> pairs.panels(sim.data[2:9]) #show a splom of just the 2nd through 9th variable
```

5. To analyze your data using the linear model, you first need to center the data around 0. The **scale** function will do this, but it returns a matrix. This needs to be converted back to a data frame.

```
> cen.data <- data.frame(scale(sim.data,scale=FALSE)) #center the data around the mean  
> describe(cen.data) # notice how the means have been changed to zero
```

6. Test your hypothesis using the linear model function **lm**

```
> model1 <- lm(arousal ~ drug * time, data = cen.data)  
> summary(model1)
```

7. Graph the results

```
> attach(sim.data)  
> plot(time,arousal,pch=20+drug)  
> by(sim.data,drug,function(x) abline(lm(arousal~time,data=x)))
```

For more detail

For more detail with sample output, consult the longer handout at <http://personality-project.org/r/r.short.pdf>