Simulate data for SEM class use

Brian P. Flaherty

October 10, 2018
Version: 1.1

1 Introduction

In our class, we will do a fair amount of analysis of our own data. As we discussed in class, you may want to tinker with models, play around, and learn on simulated data, such that you don't sully your science when you get to the point of working with your real data. This brief document describes how to do this.

At least initially, it will generate simple multivariate normal data. Perhaps sometime down the road, it can be expanded to include generating ordinal data.

2 R library `mvtnorm`

I'll use the command `rmvnorm()` from the R library `mvtnorm`. To check if it is already installed, simply type "`help(rmvnorm)`" or "`library(mvtnorm)`" at the R prompt. If you get something like this

```
> library(mvtnorm)
Error in library(mvtnorm) : there is no package called mvtnorm
```

it is likely not installed.

To install, simply run "`install.packages(‘‘mvtnorm’’)`". (`mvtnorm` is a character string and thus enclosed in double-quotes in the R session.)

Once installed, `library(mvtnorm)` should work.

3 What do you need?

In order to simulate data for our classwork, you need a vector of means and covariance matrix for all the items you plan to use. For example, if you have fives scales, each with ten items, you need to obtain a vector of 50 means and the $50x50$ sample covariance matrix.

Remember, if you have any missing data, for your covariance calculation, do not do pairwise deletion. Do listwise so you get a well-behaved covariance

matrix. (Of course, you are welcome to do something more sophisticated like and EM algorithm, but I'm not assuming that knowledge here.)

4 Generating the data

First, you need to get your means and covariance matrix into R. You can cut and paste or you could save each as text files that you will then read into R.

Let's say I have a file called `means.txt` that contains my item means. I can read those means into R with the following syntax,

```
data.means <- scan("means.txt")
```

If the file was in a different place than the R current working directory, you need to include that path information in the scan command,

```
data.means <- scan("/the/path/to/my/file/means.txt")
```

You can check the current working directory of your R session with the command `getwd()`.

Read the covariance matrix in the same way,

```
tmp.cov <- scan("cov.txt")
```

However, `tmp.cov` is a vector and you want a matrix.

```
data.cov <- matrix(tmp.cov, byrow=T, ncol=50)
```

(Remember, we are pretending to have five scales with 10 items each, thus the 50 for the number of columns in the `matrix` command above.)

I recommend you check that everything has worked and verify that your means and covariance match the numbers you had at the start.

Once we have all this information, generating the data is relatively simple. Here, I'm generating data for 300 people.

```
working.data <- rmvnorm(n=300, mean=data.means, sigma=data.cov)
```

The object working.data now contains raw simulated data from 300 people from a population with your set of means and your given covariance matrix. You can use that raw data matrix in further analyses, or you can create and save a covariance matrix.

```
working.covmat <- cov(working.data)
```

5 Multiple groups

Last, let's say I really want to work with theoretical models, but within three groups. I could repeat all the above, but use the within group means and covariance matrices to simulate data from the three groups separately. I can then do my analyses within group or as a multiple groups problem.