Reflection on Y750 Final Project

Why a simulation study?

For the final project, I decided to do a simulation study for the very first time. I had never done a simulation study and saw this as an opportunity to learn *how* to conduct a simulation by working on a very small project. The topic for this study marries an area that I'm interested in exploring further for my dissertation--multilevel models, with a topic that I've had some exposure to, latent growth curve models.

Figuring out exactly what is entailed in a simulation study

The first, and most obvious source of difficulty is figuring out exactly how one would conduct a simulation. I used readings from Bandalos & Leite (2013) and Bandalos & Gagne (2012) which provided guidelines on conducting simulations in structural equation modeling. Because I was new to MLGCMs, I needed to understand the model and learn how to parameterize the model in Mplus. I found a dissertation (Jiang, 2014) that investigated the effects of different treatments of missing data for multilevel latent growth curve models. This dissertation provided me with detailed background information in how to parameterize the MLGCM. In addition, Jiang (2014) used the ECLSK:2011 for the population model with Math scores as the outcome. Because of my familiarity with the ECLSK:2011, I decided to use it to generate population parameters for my study, using reading scores as my outcome of interest. In addition, because it was my time using the MLGCM and I relied heavily on the Mplus User's Guide (Chapter 9) as a reference. I adapted the syntax in chapter 9.12

What is the "best" way to do this?

I quickly learned that there are many ways in which researchers have conducted simulation studies and there isn't a *correct* way of conducting a simulation. For example one could choose from different software to generate data, such as SAS, R, Mplus, and then decide on a software to use to analyze the simulated datasets. I spent approximately four days looking at over and trying out code from SAS, R, and Mplus before deciding to use Mplus to generate and analyze my data. This seemed to be the most straightforward option given that I had approximately three weeks to complete the study and write up the results. I would eventually like to learn to use R to conduct a simulation as it offers more flexibility.

Learning the nuts and bolts of a simulation study

The resource I used the most when trying to figure out how to do the simulation in Mplus was the Mplus User's Guide (Chapter 12). Reading the examples in Chapter 12 helped me understand the different types of simulations one could run in Mplus. When I started to write my Mplus code, I learned that needed to create separate syntax files in Mplus for each treatment condition. To do this in an efficient manner, and to minimize human error, I decided to learn how to use Mplus Automation (MA), an R package, for the simulation. Although I used MA this past summer for my work with Latent Class Analysis, using it for simulations was new to me.

I used a paper by Hallquist & Wiley (2018) which helped me with the syntax for creating input files in R. It seemed that the syntax for MA was straightforward but after encountering an error

message, I learned was that I had to use a text file and not an R script file to use as an initial file to create Mplus Input. This is something that I learned after searching the MA Google Group site where Joshua Wiley (2013) replied a user who had the same error message as me:

> *Did you put your code in a *separate* template file? None of it will work entered straight into R. It has to be in a file that you process with the createModels() function.*

Another issue I found was that my output files did not give me CFI or TLI results. I emailed Mplus support and learned that CFI/TLI statistics are not provided for internal Monte Carlo studies (where data is generated and interpreted in one go using Mplus) and are only provided for external Monte Carlo studies such as the one found in example 12.6 of the Mplus User's Guide. Because I wanted to compare CFI results as one of the outcomes in my study, I decided to adapt my Mplus code so that the simulation study I conduct would be an external one. To do so, I modified the syntax used to generate population parameters so that the population parameters would be saved as an external data file. I added an extra line at the end of the syntax:

```
SAVEDATA: Estimates=/Users/melissa/Documents/Box Sync/-Active/Classes Fa18/ Y750
Adv SEM/ResearchProject/mplus_data_gen/templateExample/750estimates.dat;
```

Then I modified example 12.7 from the Mplus User's Guide to tell Mplus to read-in the data file containing population parameters.

```
Population = y750estimates.dat;
```

I started by simulating and analyzing 10 datasets. I found that m population parameters were not correctly specified in output and I could not get them to show up correctly. After a lot of trial and error, I figured out that I had to specify the same data file for my population and coverage values.

```
Population = y750estimates.dat;
Coverage = y750estimates.dat;
```

Issues with collating the results

After running the simulations, I had to figure out how to compile the results. To do so, I used both syntax specific to MA and R syntax. After running the `readModels` command for MA, I had difficulty putting the results in a table format. I found answers in the MA Google Group and ended up modifying the following syntax provided by a user named Nicholas Bishop:

```
mySummaries <-readModels("C:/aim 1", what="summaries", recursive=TRUE)

combined_sum <- do.call(dplyr::bind_rows, lapply(1:length(mySummaries), function(m) {
      df <- mySummaries[[m]]$summaries
      return(df)
    }))
write.table(combined_sum,paste('model fit summaries','.csv',sep=""),row.names=FALSE,
col.names=TRUE, quote=F, sep=',')
```

Another issue I had was compiling results that MA did not have functions for. For example, I could not figure out how to extract Tech9 results. Because Tech9 results are in text format, I could not get code using `dplyr` to work (shown above). I asked Dr. Rutkowski for help and Dr. Rutkowski shared R code for her project with Dr. Svetina where parameters were extracted by scanning Mplus output files. I also asked a friend proficient in R for help. My friend helped me with writing for-loops in R to extract Tech9 as well as as ICC results.

For loop written to extract ICCs:

```
filenames <- (list.files("/Volumes/STUFF/Y750
Project/mplus_data_gen/mplus_data_gen/y750project_1107", pattern="*.out",
full.names=TRUE))

     # Extract ICCs
     row_icc = array(1:7)
     row_tech9 = array(1:6)
     row_savedata = array(1:6)
     info_icc = array(1:7)
     info_tech9 = array(1:6)

     num_files = length(filenames)
     # Create ICC file
     for (i in 1:num_files){
       #print(i)
       trigger = 0
       row = 0
       temp1<-readLines(filenames[i])
       while(trigger == 0){
         row <- row+1
         if(temp1[row]="     Variable  Correlation   Variable  Correlation   Variable
     Correlation")
         {row_icc[i] = row
         print(temp1[(row_icc[i]+2):(row_icc[i]+3)])
         trigger=1
         }
       }
     }
```

Concluding thoughts

I cannot be more glad that I decided to do a simulation for my final project. The process was frustrating at times and there were moments where I was not sure if I would even get to the results (it was a very low point where my code in Mplus was just not working). Two things kept me going. Firstly, Dr. Rutkowski's support throughout this process. I was never made to feel like I was a bad coder or that my simple simulation was too simple. She listened to my many questions and shared her suggestions. This was enough for me to keep trying (and trying). Second, what helped me see the light at the end of the tunnel was my friend who helped me with for loops in R. If not for his help, I would likely still be copying and pasting from my 6000 output files.

I've learned that in terms of simulations, I am currently limited by my lack of R programming skills. I hope to spend some time during the winter break working this.

References

Bandalos, D. L. & Leite, W. L. (2013). Use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course (2nd Ed.)*. (pp.564-666) Greenwich, CT: Information Age Publishing.

Bandalos, D.L. & Gagne, P. (2012). Simulation methods in structural equation modeling. In R.H. Hoyle, (Ed.), *Handbook of Structural Equation Modeling.* New York: Guilford Publications.

Hallquist, M. N., & Wiley, J. F. (2018). Mplus Automation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling*, 25(4), 621-638. http://dx.doi.org/10.1080/10705511.2017.1402334

Jiang, H. (2014). *Missing data treatments in multilevel latent growth model : A Monte Carlo simulation study* (Order No. 3672176). Available from Dissertations & Theses @ CIC Institutions; ProQuest Dissertations & Theses Global. (1647164295). Retrieved from http://proxyiub.uits.iu.edu/login?url=https://search.proquest.com/docview/1647164295?accountid=11620

Joshua Wiley (2013, February 2). Getting error in init section: Error: unexpected '[[' in "[[" [Msg 2]. Message posted to https://groups.google.com/forum/#!msg/mplusautomation/Hl6SUhBlPTI/jgVHhSWqKLsJ;context-place=forum/mplusautomation

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén

Nicholas Bishop (2018, July 19). Extracting model summaries using runModels [Msg 6]. Message posted to https://groups.google.com/forum/#!msg/mplusautomation/R6sic5atpXM/HB86uY6XBwAJ;context-place=msg/mplusautomation/8ooMoB0QzmE/OIitEAEGAQAJ