Structural Equation Modeling

P.08 - MIMIC Models and Instrumental Variables

November 15, 2022 (13:37:52)

Lab Description

For this practical you will need the following packages: `lavaan` and `semPlot`. You can install and load these packages using the following code:

```
# Install packages.
install.packages(c("lavaan", "semPlot", "mvtnorm", "GGally"))

# Load the packages.
library(lavaan)
library(semPlot)
```

Exercise 1

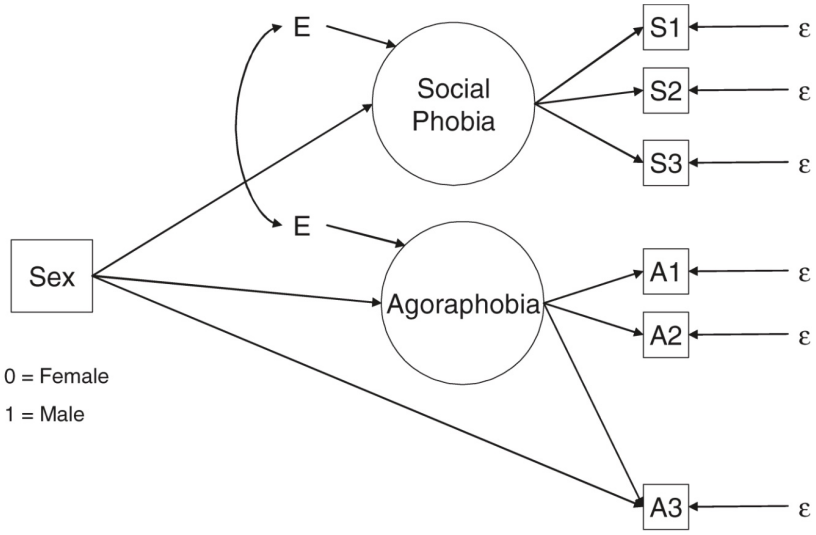
Estimate the model in *Figure 1* in `lavaan` and examine if there is evidence of Differential Item Functioning (DIF) in the measurement instruments. To help you get started, you are provided with the code that contains the correlations and standard deviations corresponding to the model depicted in *Figure 1*.

Standard deviations and correlations.

```
# Standard deviations.
sd <- "2.26 2.73 2.11 2.32 2.61 2.44 0.50"

# Correlations.
cor <- "
    1.000
    0.705   1.000
    0.724   0.646   1.000
    0.213   0.195   0.190   1.000
    0.149   0.142   0.128   0.521   1.000
    0.155   0.162   0.135   0.557   0.479   1.000
    -0.019  -0.024  -0.029  -0.110  -0.074  -0.291   1.000
"

# Get covariances.
cov <- getCov(cor, sds = sd, names = c("S1", "S2", "S3", "A1", "A2", "A3", "sex"))
```



Sample Correlations and Standard Deviations (*SD*s); *N* = 730 (365 males, 365 females)

	S1	S2	S3	A1	A2	A3	Sex
S1	1.000						
S2	0.705	1.000					
S3	0.724	0.646	1.000				
A1	0.213	0.195	0.190	1.000			
A2	0.149	0.142	0.128	0.521	1.000		
A3	0.155	0.162	0.135	0.557	0.479	1.000	
Sex	-0.019	-0.024	-0.029	-0.110	-0.074	-0.291	1.000
SD:	2.260	2.730	2.110	2.320	2.610	2.440	0.500

FIGURE 7.5. MIMIC model of Social Phobia and Agoraphobia. S1, giving a speech; S2, meeting strangers; S3, talking to people; A1, going long distances from home; A2, entering a crowded mall; A3, walking alone in isolated areas. (All questionnaire items rated on 0–8 scales, where 0 = no fear and 8 = extreme fear.)

Figure 1: Reproduction of Figure 7.5 from Brown (2014, p. 275)

Exercise 2

Open the dataset `card.csv` available on Canvas in the folder corresponding to the current practical. This dataset contains several variables used by David Card (1995) to estimate the causal effect of education on wages using proximity to college as an instrumental variable. You can find more information about this dataset at this link.

a. Estimate a model in which you only regress `lwage` on `educ` (i.e., without including an instrumental variable). What do you conclude from this regression?

b. Re-estimate the model at point *(a)*, but this time with the following control variables added: `exper`, `expersq`, `black`, `south`, `fatheduc`, and `motheduc`.

The problem with treating the direct association between `educ` and `lwage` as a causal effect is that there are likely many omitted variables that affect both education and wages. We could control for those variables by measuring them and including them in the model (i.e., as we did at point *b*). But there is no way we can control for all possible confounding variables, especially because some variables are difficult to measure (e.g., ability). It is therefore likely that education is correlated with the error term in the regression (i.e., a form of endogeneity), and that our regression coefficient is, in turn, biased to an unknown degree. David Card proposed to solve this problem by introducing proximity to college as an instrumental variable. Specifically, `nearc4` was a dummy indicator variable for whether or not the person was raised in a local labor market that included a four-year college.

c. Re-estimate the model at point *(b)* with the following additions:
   - add `nearc4` as an instrumental variable for for `educ`, while controlling for `fatheduc` and `motheduc`
   - add a covariance between the error terms of `educ` and `lwage`

Does this model provide evidence of endogeneity of educ? Why (not)?

d. Evaluate whether `nearc4` is a weak or strong instrument for dealing with the endogeneity of the variable `educ`. Specifically, consider the criteria that a strong instrument must meet in order to adequately correct for endogeneity.

References

Brown, T. A. (2014). *Confirmatory factor analysis for applied research.* Guilford Publications.