#################################################################

################ Computer Practical 1 ###########################

###### Structural Equation Modeling and Causal Inference ########

########################################### 06Nov2023, Ozan Aksoy

Q1: Recursive path model

# Status attainment (“Blau and Duncan model”)

In this question, you will fit a variant of the path model discussed during the lecture, focusing on the interpretation of direct and indirect effects. We will use data from the famous Blau & Duncan (Blau & Duncan (1967) *The American Occupational Structure*. New York: The Free Press) on the determinants of occupational status. The variables used in the B&D analysis are:

FaEd Father’s Education;

FaOcc Father's Occupational status;

Educ Respondent’s Education;

Occ1 Respondent's first-job Occupational status;

Occ2 Respondent's 1962 Occupational status.

The correlation matrix reported in the book is:

1.0000

0.5160 1.0000

0.4530 0.4380 1.0000

0.3320 0.4170 0.5380 1.0000

0.3220 0.4050 0.5960 0.5410 1.0000

Number OBS = 20700

Names = faed faocc educ occ1 occ2

In the path analysis, Educ, Occ1, and Occ2 will be used as the dependent (endogenous) variables; FaEd, and FaOcc act as predictor (exogenous) variables. The equations of the model are:

Educ = FaEd + FaOcc (+ e\_Educ)

Occ1 = FaOcc + Educ (+ e\_Occ1)

Occ2 = FaOcc + Educ + Occ1 (+ e\_Occ2)

1. Sketch the path diagram.
2. Get the data in R and fit the Blau & Duncan path model.
3. Make sure that the direct and indirect effects of FaOcc on Occ2 are estimated. What is the direct effect of FaOcc on Occ2? And the indirect effect? What is the total effect? Interpret the effect of FaOcc on Occ2 in terms of direct and indirect effects.
4. We are now looking into **inference** on the indirect and total effects, e.g., test that the indirect effect is non-zero. To do so, we need to estimate standard errors of indirect and total effects (for the direct effects, standard errors are already reported, right?). By default, R uses the asymptotic delta method. Some scholars cast doubt on the accuracy of the (fast) delta method -- the argument is that the indirect effect is the (sum of) product of coefficients, if the estimates of the coefficients are approximately normal distributed, then the product of two (dependent) normal variates may be quite non-normal distributed, and hence a standard deviation and a central confidence interval may give misleading impressions of uncertainty in the indirect effect. As a solution they propose to use the (computer intensive) bootstrap method. In lavaan, you may request bootstrapping by adding the statement **se = "bootstrap"** in the fitting function. However, we’ll need raw data for bootstrapping. Here, we only have correlation matrix. So normally we are not able to implement bootstrapping. For the sake of the exercise, we will simulate raw data using the correlation matrix we have, assuming multivariate normality. Simulate 20700 observations for the five variables in the model using the correlation matrix above. When simulating the data, set the means of the variables to zero. You can use various functions in R ( e.g. **mvrnorm()** function of the MASS package or probably better, the **kd()** function of semTools). Fit the same model you fitted above, but now use the simulated raw data and calculate standard errors using the bootstrap method. Beware: bootstrapping is a computationally intensive process, hence it will take a while before R produces the estimates. What is your conclusion about the indirect effect?
5. Give an interpretation of the **model fit statistics** (carefully select the fit indices to be reported!)
6. Note that in analyzing the Blau & Duncan Path Model, the number of observations was 20,700. This will affect some of the fit indices. If we divide the number of observations by 10, some of the fit indices will change. Reduce the number of observations to 2,070 in the model you fitted using the correlation matrix. What do you conclude?
7. Scholars have long interpreted indirect effects as causal mediation, that is, the causal effect of an exogenous variable is mediated through the intervening variable. However, such an interpretation relies on very strong assumptions. What assumptions are needed to be met to be able to interpret the above path diagram as a causal mediation model?

# Q2: Non-recursive Path Model

# Peer influence

Duncan, Haller & Portes (1968), *Peer Influences on Aspirations: A Reinterpretation*, American Journal of Sociology, 74:119-137, present an interesting application of non-recursive path models. The article can be found online through UCL library.

a) Carefully inspect the models I and II (slightly modified Model I) on pages 122 and 126. For your convenience, I also paste below the path diagrams for models I and II. Get the data in R using the correlation matrix below and fit model I and II using R-lavaan. Argue why the models are identified. Model IV involves latent variables. If you want, you can also fit model III. Note that DHP fit the model using two-stage least squares, while R uses maximum likelihood, thus your parameter estimates will slightly differ from the results reported by DHP.

The correlation matrix is pasted below for N=329 ego-alter dyads. The variables are:

occasp occupational aspiration of ego

edasp educational aspiration of ego

fedasp educational aspiration of friend/alter

foccasp occupational aspiration of friend/alter

parasp ego’s parental aspirations

iq intelligence quotient of ego

ses SES of ego

fses SES of friend/alter

fiq intelligence quotient of friend/alter

fparasp. Friend/alter’s parental aspirations

Correlation matrix is:

1.0000

.6247 1.0000

.3269 .3669 1.0000

.4216 .3275 .6404 1.0000

.2137 .2742 .1124 .0839 1.0000

.4105 .4043 .2903 .2598 .1839 1.0000

.3240 .4047 .3054 .2786 .0489 .2220 1.0000

.2930 .2407 .4105 .3607 .0186 .1861 .2707 1.0000

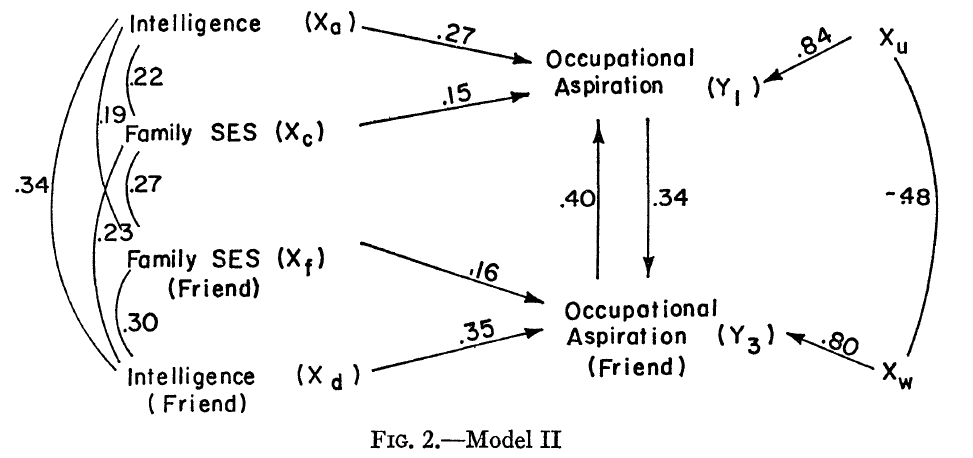
.2995 .2863 .5191 .5007 .0782 .3355 .2302 .2950 1.0000

.0760 .0702 .2784 .1988 .1147 .1021 .0931 -.0438 .2087 1.0000

Number obs = 329

Var names = occasp edasp fedasp foccasp parasp iq ses fses fiq fparasp

****



**b)** Theoretically the model for ego and friend should be “symmetric” from the perspectives of ego and his/her friend. For instance, the (unstandardized) effect of ego’s ambition on his friend’s ambition should be the same as the (unstandardized) effect of friend’s ambition on ego’s ambition. Consider carefully which parameters of the model are affected by this symmetry argument. Test whether these equality constraints are met. (Actually, this refined analysis requires individual data or the covariance matrix, while we only have the correlation matrix. In an analysis of the correlation matrix, we are actually constraining the standardized coefficients rather than the unstandardized coefficients. I ask you to do anyways…)