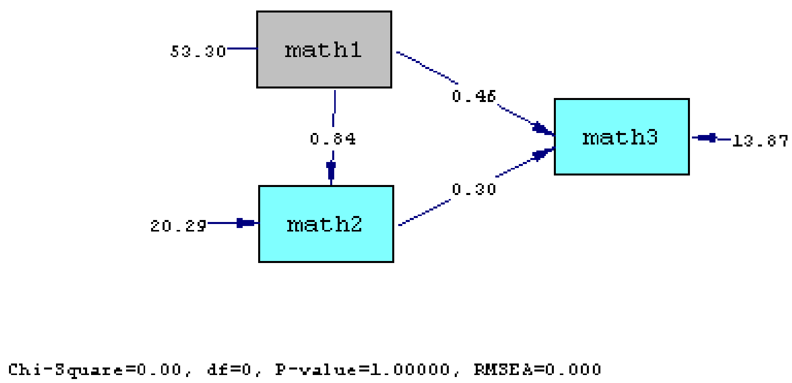
Structural Equation Modeling Lab 4

In Class Path Analysis Example

1. Path diagram and syntax




```
TI Path Analysis Example
 Math3 on Math1 and Math2 with path from Math1 to Math2
 DA NI=10 NO=0 MA=CM
 RA FI='C:\YOUR PATH HERE\jsp162_lab1_2-11-09.psf'
 SE
 6 7 5 /
 MO NX=1 NY=2 BE=FU GA=FI PS=SY
 FR BE(2,1) GA(1,1) GA(2,1)
 PD
 OU ND=4 EF SS
```

Note. You could reverse the order of selected variables (SE 7 6 5 for example), this would produce the same estimates but different parameter specifications. So in the case of variable seven being the first endogenous variable in order to get the same model you would need to specify BE(1,2) instead.

2. *Calculating degrees of freedom*

Degrees of freedom are determined by subtracting the number of free parameters we are estimating from the number of observed terms we are inputting.

For example, in this model the *df*=0. This is because we are inputting 6 unique terms with our covariance matrix

df = (variances + covariances) – (parameters being estimated)

$$df = \left(3\,\text{var} + \frac{3\times(3-1)}{2}\text{cov}\right) - \left(\beta_{21},\ \gamma_{21},\ \gamma_{11},\ \text{var}(x_1),\ \text{var}(\zeta_1),\ \text{var}(\zeta_2)\right)$$

$$df = (3+3) - (3+2+1) = 6 - 6 = 0$$

Zero degrees of freedom will result in a perfect model fit because we have exactly as many things to estimate as we are observing. This can be likened to solving the following equations:

$$2x + y = 10$$
$$x + 4y = -9$$

With two unknowns (*x* and *y*), and two equations, we can find one unique solution (*x*=7, *y*=-4) that reproduce the values exactly/perfectly. However, if we reduce the number of equations to one:

$$2x + y = 10$$

Now we have two unknowns with only one equation. This would result in negative degrees of freedom (when we have more unknown terms in our model that observed terms). In this case, there would be an infinite number of solutions, and no way to determine which was better than any other. If we ran a model with negative *df*, we would receive a “model fails to converge” error message. If we increase the number of equations (assuming no equations are linearly dependent)

$$2x + y = 10$$
$$x + 4y = -9$$
$$3x + 7y = 21$$

we have positive *df*. This will be what you typically deal with in SEM. The goal would be to find values for *x* and *y* that come as close to reproducing the equations as best as possible. There are different ways to evaluate how good the solution for the unknowns is, and this is where goodness-of-fit statistics come into play.

3. Fitted regression equations

For x_1=Math1, y_1=Math2 and y_3=Math3:

$$\hat{y}_1 = 4.7819 + 0.8414x_1$$
$$\hat{y}_2 = 10.6072 + 0.4586x_1 + 0.3046y_1$$

4. Path diagram equations

$$y_1 = \alpha_1 + \gamma_{11}x_1 + \zeta_1$$
$$y_2 = \alpha_2 + \gamma_{21}x_1 + \beta_{21}y_1 + \zeta_2$$
$$x_1 = x_1$$

These are the equations which are used to reproduce the originally observed variance covariance matrix.

	math3	math2	math1
	--------	--------	--------
math3	42.9972		
math2	38.2426	58.0238	
math1	38.1051	44.8473	53.2988

5. Equations used to reproduce the covariance matrix

$$\mathrm{var}(x_1) = \mathrm{var}(x_1)$$
$$\mathrm{cov}(x_1, y_1) = \gamma_{11}\,\mathrm{var}(x_1)$$
$$\mathrm{cov}(x_1, y_2) = \gamma_{21}\,\mathrm{var}(x_1) + \beta_{21}\,\mathrm{cov}(x_1, y_1)$$
$$\mathrm{var}(y_1) = \gamma_{11}^2\,\mathrm{var}(x_1) + \mathrm{var}(\zeta_1)$$
$$\mathrm{cov}(y_1, y_2) = \gamma_{11}\gamma_{21}\,\mathrm{var}(x_1) + \beta_{21}\gamma_{11}\,\mathrm{cov}(x_1, y_1) + \beta_{21}\,\mathrm{var}(\zeta_1)$$
$$\mathrm{var}(y_2) = \gamma_{21}^2\,\mathrm{var}(x_1) + 2(\beta_{21}\gamma_{21}\,\mathrm{cov}(x_1, y_1)) + \beta_{21}^2\,\mathrm{var}(y_1) + \mathrm{var}(\zeta_2)$$

Note. All terms in the above equations should be taken from the LISREL estimates (found after the iteration statement, and not the initial solutions).

6. Reduced form

Consider Math3 in our model. Math2 has a direct effect on Math3(β_{21}). Math1 also has a direct effect on Math3(γ_{21}). However, Math1 also has an indirect effect of Math3 that is mediated by its relationship through Math2. Since our model is recursive, we can express the entire relationship of math3 by substituting the equation for math2 within it. Take our two equations that express our model.

$$y_1 = \alpha_1 + \gamma_{11}x_1 + \zeta_1$$
$$y_2 = \alpha_2 + \gamma_{21}x_1 + \beta_{21}y_1 + \zeta_2$$

To create our reduced form, substitute the y_1 equation for the y_1 term in the y_2 equation:

$$y_2 = \alpha_2 + \gamma_{21}x_1 + \beta_{21}(\alpha_1 + \gamma_{11}x_1 + \zeta_1) + \zeta_2$$

Now expand the β_{21} parentheses:

$$y_2 = \alpha_2 + \gamma_{21}x_1 + \beta_{21}\alpha_1 + \beta_{21}\gamma_{11}x_1 + \beta_{21}\zeta_1 + \zeta_2$$

Reorganizing the intercept regression (path) coefficient for x_1 and error terms together, let's look explicitly at the regression of y_2 on x_1.

$$y_2 = (\alpha_2 + \beta_{21}\alpha_1) + (\gamma_{21}x_1 + \beta_{21}\gamma_{11}x_1) + (\beta_{21}\zeta_1 + \zeta_2)$$

$$y_2 = (\alpha_2 + \beta_{21}\alpha_1) + x_1(\gamma_{21} + \beta_{21}\gamma_{11}) + (\beta_{21}\zeta_1 + \zeta_2)\,.$$

This is the model for the reduced form described above.

Note. that there is no reduced form for y_1 because y_1 is only expressed in terms of the exogenous variable and has no indirect effects. It is within this reduced form for y_2 and more specifically in the combination of coefficients specific to x_1 that lies the key to delineating the direct, indirect and total effects of x_1 on y_2.

7. Standardized coefficients

LISREL will provide standardized solutions for all of the estimated parameters. It will do so by computing the product between the unstandardized estimate and the ratio between the corresponding standard deviations. Let us consider an example. LISREL reported the path coefficient between the two endogenous variables as follows:

```
BETA
            math3        math2
           --------     --------
math3         - -          - -
math2       0.3046         - -
           (0.0654)
            4.6595
```

From this it follows that the standardized form of this estimate can be computed as follows:

$$\beta_{21}{}^{*} = \beta_{21} \times \frac{\sqrt{\mathrm{var}(Math2)}}{\sqrt{\mathrm{var}(Math3)}} = \beta_{21} \times \frac{\sigma_{Math2}}{\sigma_{Math3}}$$

$$\beta_{21}{}^{*} = .3046 \times \frac{\sqrt{58.0238}}{\sqrt{42.9972}} = .3046 \times \frac{7.6173}{6.5572} = .3046 \times 1.1617 = .3538$$

This standardization can be verified in at least three ways: a) changing the path diagram model display to standardized estimates which will display the standardized solutions on the path diagram, b) double-checking the LISREL output which (due to our request in the output tag OU SS) will have produced the standardized solutions following the model fit, or c) resubmitting the model syntax with the substitution of KM for CM in the DA command MA=CM tag, which effectively analyzes a correlation matrix as opposed to the covariance matrix and thus produces already standardized solutions. For our example let us consider the output provided:

```
Standardized Solution
BETA
              math3      math2
            --------   --------
    math3      - -        - -
    math2    0.3539       - -
```

We can see that our computation is accurate to the 3rd decimal place. This slight divergence is due to rounding error since we have used rounded (to the fourth decimal place) estimates.

8. Calculating squared multiple correlation (*R*²)

Unlike in multiple regression where we compute a single squared multiple correlation coefficient (coefficient of determination) in SEM. Here in our path analysis example, we have one for each regressive equation. Note that in the case of y_1 the squared multiple correlation will be reduced to the square of the bivariate correlation between y_1 and x_1. The squared multiple correlation for the first equation can be generally expressed in the form:

$$R^2_{y1.x1} = 1 - \frac{\mathrm{var}(\zeta_1)}{\mathrm{var}(y_1)}$$

Again, note that the R^2 will be different for y_2 in the reduced form as opposed to the y_2 in the original equations? Look at our reduced form y_2 equation. You can see by changing the terms explicitly to express the x_1 relationship, reproducing the $\mathrm{var}(y_2)$ will be using different terms and will thus give you a different numerical result for the error variance, $\mathrm{var}(\beta_{21}\zeta_1 + \zeta_2) = \beta_{21}^2 \times \mathrm{var}(\zeta_1) + \mathrm{var}(\zeta_2) = \beta_{21}^2 \times \psi_{11} + \psi_{22}$.

9. Direct effects

The direct effects in the model are easily identified for those are the estimated path coefficients we see in the path diagram. Therefore, by examining the diagram we note that there are three direct effects, two of the exogenous variable on the two endogenous variables and one of the first endogenous variable on the second one.

10. Indirect and total effects

Indirect effects are not explicitly stated on the path diagram since those are effects mediated by other variables. In our example there is only one indirect effect, that of the exogenous variable on the second endogenous variable through the first endogenous variable. Meaning, there is an indirect effect of x_1 on y_2 though y_1. Let's look at the term associated with x_1 from the reduced form:

$$x_1(\gamma_{21} + \beta_{21}\gamma_{11})$$

In this term γ_{21} represents the direct effect of math1 (x_1) on math3 (y_2). This can be verified by checking the path diagram and sure enough, γ_{21} is the path coefficient for that direct relationship. Conversely, $\beta_{21}\gamma_{11}$ represents the indirect effect of math1 (x_1) on math3 (y_2) mediated by math2 (y_1). Subsequently, the total effect of math1 on math3 will be the sum of the direct and indirect effects $(\gamma_{21} + \beta_{21}\gamma_{11})$, as denoted in the reduced form for math3 (y_2). Direct, indirect and total effects can be separately analyzed in LISREL by requesting the portioned effects to be printed in the output using the EF option (OU EF). These estimate matrixes allow us to make inferences regarding the mediation of variables through other variables. In our example let us examine our one indirect effect of math1 (x_1) on math3 (y_2).

```
Total and Indirect Effects
Total Effects of X on Y

            math1
          --------
math2       0.84
          (0.05)
           17.25
math3       0.71
          (0.04)
           16.63
```

This is the total effect of x_1 on y_1: $\gamma_{11} = 0.8414$. Note that x_1 only has this direct relationship with y_1, so the direct effect is the same as the total effect.

This is the total effect of x_1 on y_2 : $\gamma_{21} + \beta_{21}\gamma_{11}$, the sum of the direct and indirect effects (see the second equation in number 3) $0.4586 + 0.2562 = 0.7148$.

```
Indirect Effects of X on Y

            math1
          --------
math2       - -
math3       0.26
          (0.06)
            4.50
```

As noted above, there is no indirect effect of x_1 on y_1.

However, there is an indirect effect of x_1 on y_2. Based on our reduced form, this should be $\beta_{21}\gamma_{11} = 0.3046 * 0.8414 = 0.2562$.

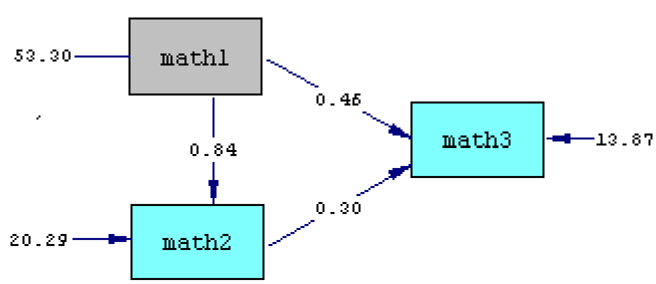
```
Total Effects of Y on Y

         math2     math3
        ------    ------
math2    - -       - -
math3    0.30      - --
        (0.07)
         4.66
```

Here is the total direct effect of y_1 on y_2 :$\beta_{21} = 0.3046$.

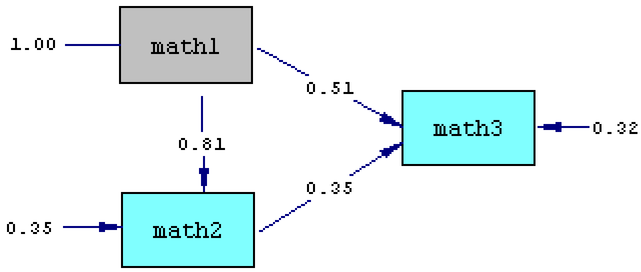
11. Equivalent models & temporal relationships

As you may have already discovered from the distributed readings and the course lecture SEM models are generally hypothesis driven. It is therefore imperative to have a clear and concise idea in mind when building these relational models. Temporal aspects often provide a significant logical indicator on how certain relationship ought to be modeled. In the discussed example we have three math achievement scores. These scores were collected from a sample on three separate occasions; math1 was collected when the sample was in 3rd grade, math2 in 5th grade and math3 in 8th grade. It is illogical in such a situation to model these variables retrospectively, meaning that math3 should predict math1 when math3 was temporally collected after math1. This particular situation also provides suggestions regarding mediation. It is plausible to conceive that math3 scores will be predicted by math1 and math2 scores, but also that math1 will have a separate impact on math2 and subsequently mediated indirect effect on math3.

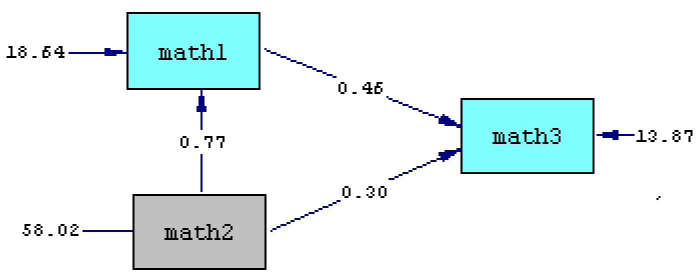
The issue of equivalent models is usually discussed when regarding the dilemma of directionalities of modeled relationships. In our example math1 predicts math2. Regard the original path diagram:



and its standardized solutions.



From regression analysis it follows that when one variable is regressed on another single predictor the resulting relationship accounts for the same amount of variability as when the relationship is reversed (i.e. regressing the predictor on the earlier response variable). This is due to the property that the coefficient of determination is reduced to the square of the correlation between the two, which does not change even if the relationship is reversed. Consequently, the unstandardized regression coefficient for the simple linear regression is appropriate for the scale of the predictor and will therefore be different if the relationship is reversed. However, since these relationships are equivalent one can verify it by standardizing the regression coefficient to see that it is the same in both cases. Consider now the earlier example. We could reverse the relationship between math2 and math1, making math2 the exogenous variable predicting math1 and math3. The unstandardized coefficient is going to be appropriate for the scale of math2.



However, the standardized coefficient, as well as all others, is the same. This is because the only altered relationship is that between math2 and math1 which in essence is a simple linear regression and therefore equivalent no matter which direction it is.