



Regression models with latent variables: Head Start summer program

Sörbom (1981) reanalyzed some data from the Head Start summer program previously reanalyzed by Magidson (1977). Sörbom used data on 303 children consisting of a Head Start sample ($N = 148$) and a matched Control sample ($N = 155$). The correlations, standard deviations and means are given in the table below.

Table: Correlations, standard deviations and means for the Head Start data

Control Group

Variable	Correlations						Std. Dev.	Means
x_1	1.000						1.360	3.839
x_2	0.484	1.000					1.195	3.290
x_3	0.224	0.342	1.000				1.193	2.600
x_4	0.268	0.215	0.387	1.000			3.239	6.435
y_1	0.230	0.215	0.196	0.115	1.000		3.900	20.415
y_2	0.265	0.297	0.234	0.162	0.635	1.000	2.719	10.070

Head Start Group

Variable	Correlations						Std. Dev.	Means
x_1	1.000						1.332	3.520
x_2	0.441	1.000					1.281	3.081
x_3	0.220	0.203	1.000				1.075	2.088
x_4	0.304	0.182	0.377	1.000			2.648	5.358
y_1	0.274	0.265	0.208	0.084	1.000		3.764	19.672
y_2	0.270	0.122	0.251	0.198	0.664	1.000	2.677	9.562

The children were matched on gender and kindergarten attendance but no attempt had been made to match on social status variables. The variables used in Sörbom's reanalysis were:

x_1 = Mother's education

x_2 = Father's education

x_3 = Father's occupation

x_4 = Family income

y_1 = Score on the Metropolitan Readiness Test

y_2 = Score on the Illinois Test of Psycholinguistic Abilities

We want to do the following:

- Test whether x_1 , x_2 , x_3 , and x_4 can be regarded as indicators of a single construct **Ses** (socioeconomic status) for both groups. Is the measurement model the same for both groups? Is there a difference in the mean of **Ses** between groups?
- Assuming that y_1 and y_2 can be used as indicators of another construct **Ability** (cognitive ability), test whether the same measurement model applies to both groups. Test the hypothesis of no difference in the mean of **Ability** between groups.
- Estimate the structural equation:

$$\text{Ability} = \alpha + \gamma \text{Ses} + z$$

Is γ the same for the two groups? Test the hypothesis $\alpha = 0$. Interpret the results.

The file **EX16.DAT** is:

```

1
.484 1
.224 .342 1
.268 .215 .387 1
.230 .215 .196 .115 1
.265 .297 .234 .162 .635 1
1.360 1.195 1.193 3.239 3.900 2.719
3.839 3.290 2.600 6.435 20.415 10.070
1
.441 1
.220 .203 1
.304 .182 .377 1
.274 .265 .208 .084 1
.270 .122 .251 .198 .664 1
1.332 1.281 1.075 2.648 3.764 2.677
3.520 3.081 2.088 5.358 19.672 9.562

```

The model for Problem A is similar to that in the thirteenth example. We have four indicators of a single latent variable *Ses* and we want to estimate the mean difference in *Ses*. The input file **EX16A.SPL** (see the **SIMPLIS Examples** folder) is shown below:

```

Group = Control. Head Start Summer Program - Problem A
Observed Variables: MOTHEduc FATHEDUC FATHOCU FAMILINC MRT ITPA
Correlations from File EX16.DAT
Standard Deviations from File EX16.DAT
Means from File EX16.DAT
Sample Size: 149
Latent Variable: Ses
Relationships:
    MOTHEduc          = CONST + 1*Ses
    FATHEDUC - FAMILINC = CONST +   Ses

Group = Head Start. Head Start Summer Program - Problem A
Correlations from File EX16.DAT
Standard Deviations from File EX16.DAT
Means from File EX16.DAT
Sample Size: 156
Relationship:
    Ses = CONST
Set the Error Variances of MOTHEduc - FAMILINC free
Set the Variance of Ses free
Path Diagram
End of Problem

```

Note that six variables are read but only four variables are used in the model. The program automatically selects the variables for analysis on the basis of the variables included in the relationships.

For Problem A we obtain an overall Chi-square measure of goodness-of-fit of

Degrees of Freedom for (C1)-(C2)	10
Maximum Likelihood Ratio Chi-Square (C1)	35.914 (P = 0.0001)

Indicating that the fit of the model is not very good. An examination of the modification indices reveals that the errors of MOTHEduc and FATHEDuc may be correlated. This indicates that Parents' education levels correlate more than can be explained by social status. This, to add the covariance between the errors of MOTHEduc and FATHEDuc, we add the line

Let the Errors of MOTHEduc FATHEDuc correlate

in both groups (see **EX16B.SPL**). This model has an acceptable fit:

Degrees of Freedom for (C1)-(C2)	8
Maximum Likelihood Ratio Chi-Square (C1)	6.477 (P = 0.5939)

The difference in degrees of freedom from the previous model is two, as we have added two parameters. The output shows that the groups differ significantly in Ses:

Mean Vector of Independent Variables

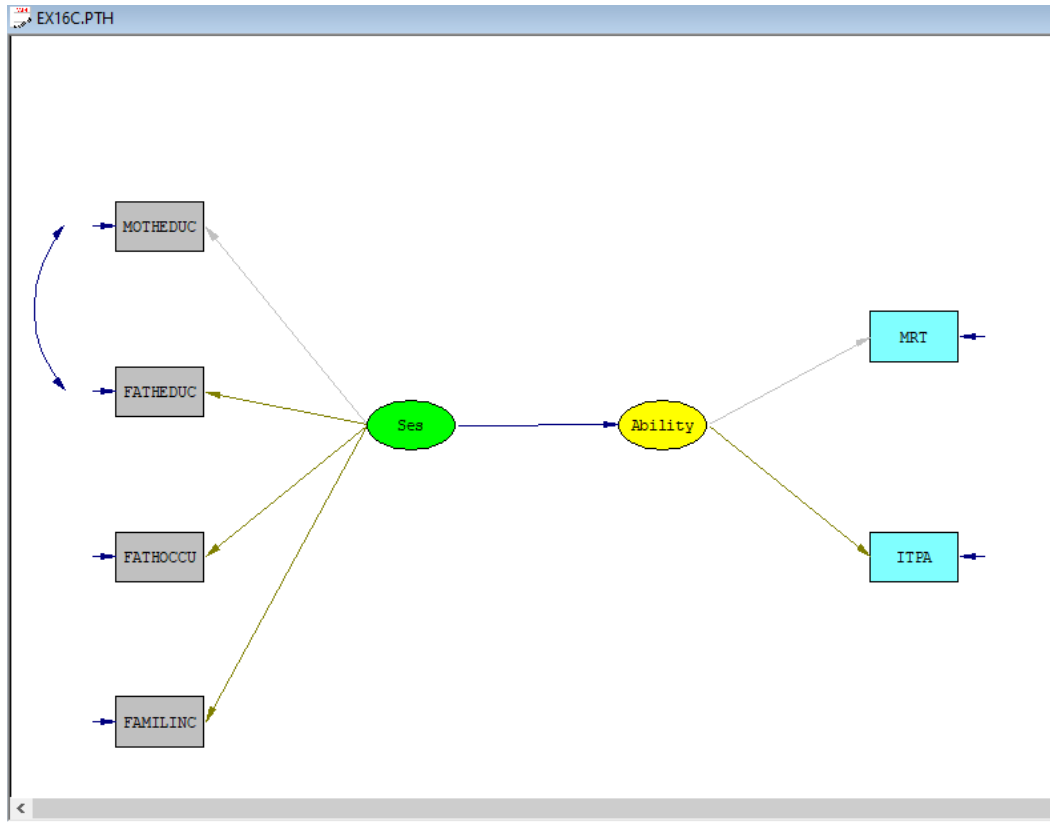
Ses

-0.342
(0.097)
-3.534

The children of the Head Start group have a lower Ses on average than the children of the Control group.

As criteria. Magidson (1977), used two cognitive ability tests – the Metropolitan Readiness Test (MRT) and the Illinois Test of Psycholinguistic Abilities (ITPA). Magidson made separate analyses for the two tests, but here, as suggested by Bentler & Woodward (1978), we will use the two tests to define the construct cognitive ability. This model is the same as in Problem A, except that there are only two observed variables. We leave it to the reader to set up the input file. The model has no degrees of freedom, so one could compute the estimates simply by equating the first and second order moments implied by the model to their observed counterparts. Also, in cognitive ability the Head Start group is below the Control group in the sense that the estimated mean difference is negative (- 0.74). However, the difference is not significant, the standard error of the estimate equals 0.44.

For Problem C we use the combined model as depicted in the figure below. The main focus is on the structural equation where $g = 1$ for the Control group and $g = 2$ for the Head Start group. We set $\alpha^{(1)} = 0$ and estimate $\alpha^{(2)}$, $\gamma^{(1)}$ and $\gamma^{(2)}$.



The input file is **EX16C.SPL**:

Group = Control. Head Start Summer Program - Problem C
 Observed Variables: MOTHEduc FATHEDUC FATHOCCU FAMILINC MRT ITEA
 Correlations from File EX16.DAT
 Standard Deviations from File EX16.DAT
 Means from File EX16.DAT
 Sample Size: 149
 Latent Variables: Ses Ability

Relationships:

MOTHEduc = CONST + 1*Ses
 FATHEDUC - FAMILINC = CONST + Ses
 MRT = CONST + 1*Ability
 ITEA = CONST + Ability
 Ability = Ses

Let the Errors of MOTHEduc and FATHEDUC correlate

Group = Head Start. Head Start Summer Program - Problem C
 Correlations from File EX16.DAT
 Standard Deviations from File EX16.DAT
 Means from File EX16.DAT
 Sample Size: 156

Relationship:

Ses = CONST
 Ability = CONST + Ses

Set the Error Variances of MOTHEduc - ITPA free
 Set the Variance of Ses free
 Set the Error Variance of Ability free
 Let the Errors of MOTHEduc and FATHEDuc correlate
 Path Diagram
 End of Problem

The chi-square for the combined model equals

Degrees of Freedom for (C1)-(C2)	22
Maximum Likelihood Ratio Chi-Square (C1)	27.965 (P = 0.1768)

So that the fit of the model is acceptable. Examination of the γ parameters in the two groups shows that

they are probably equal, since $\hat{\gamma}^{(1)} = 2.02$ and $\hat{\gamma}^{(2)} = 2.29$ with estimated standard errors equal to 0.63 and 0.73, respectively. Thus, the final model is a model with the γ 's constrained to be equal. To specify this, we replace the line (See **EX16D.SPL**)

$$\text{Ability} = \text{CONST} + \text{Ses}$$

with the line

$$\text{Ability} = \text{CONST}$$

The chi-square for this model is

Degrees of Freedom for (C1)-(C2)	23
Maximum Likelihood Ratio Chi-Square (C1)	28.078 (P = 0.2129)

Note that the *P*-value increased despite the reduction of parameters. The difference in chi-square for the last two models can be used as a test of the hypothesis that the γ 's are equal. Chi-square with one degree of freedom is 0.11. Thus we can treat the γ 's as equal. As the regression lines are parallel, it is meaningful to talk about $\alpha^{(2)}$ as a measure of the effect of Head Start.

There seems to be no significant effect of the Head Start program when controlling for social status, although the inclusion of social class has changed the negative effect to be positive. The estimate of $\alpha^{(2)}$ is 0.19 with a standard error equal to 0.38.

In the more general case, when there are more than two groups and/or more than one dependent variable, one can test the hypothesis of no effect by re-estimation of the model after adding the restriction $\alpha^{(g)} = 0$, and then compare the chi-squares. In the above case, we obtain a chi-square equal to 28.23 with 24 degrees of freedom. Thus, the test of no effect results in a chi-square with one degree of freedom equal to 0.24, which in this case is the same as the square of the *t*-value of α .