Example

8

Factor Analysis

Introduction

This example demonstrates confirmatory common factor analysis.

About the Data

Holzinger and Swineford (1939) administered 26 psychological tests to 301 seventh- and eighth-grade students in two Chicago schools. In the present example, we use scores obtained by the 73 girls from a single school (the Grant-White school). Here is a summary of the six tests used in this example:

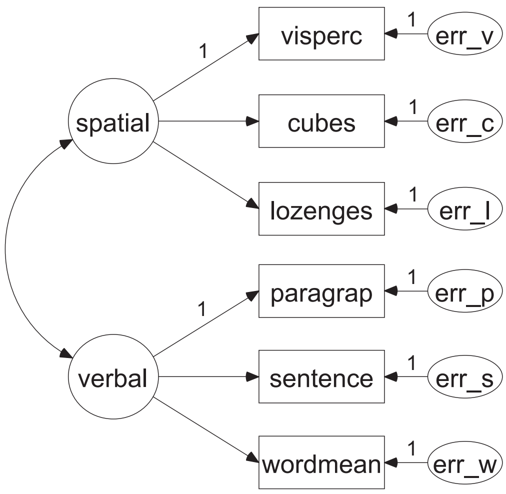
Test	Explanation
visperc	Visual perception scores
cubes	Test of spatial visualization
lozenges	Test of spatial orientation
paragraph	Paragraph comprehension score
sentence	Sentence completion score
wordmean	Word meaning test score

The file *Grnt_fem.sav* contains the test scores:

	visperc	cubes	lozenges	paragrap	sentence	wordmean
1	33.00	22.00	17.00	8.00	17.00	10.00
2	30.00	25.00	20.00	10.00	23.00	18.00
3	36.00	33.00	36.00	17.00	25.00	41.00
4	28.00	25.00	9.00	10.00	18.00	11.00
5	30.00	25.00	11.00	11.00	21.00	8.00
6	20.00	25.00	6.00	9.00	21.00	16.00
7	17.00	21.00	6.00	5.00	10.00	10.00
8	33.00	31.00	30.00	11.00	23.00	18.00

A Common Factor Model

Consider the following model for the six tests:



Example 8
Factor analysis: Girls' sample
Holzinger and Swineford (1939)
Model Specification

This model asserts that the first three tests depend on an unobserved variable called *spatial*. *Spatial* can be interpreted as an underlying ability (spatial ability) that is not directly observed. According to the model, performance on the first three tests depends on this ability. In addition, performance on each of these tests may depend on something other than spatial ability as well. In the case of *visperc*, for example, the unique variable *err_v* is also involved. *Err_v* represents any and all influences on *visperc* that are not shown elsewhere in the path diagram. *Err_v* represents error of measurement in *visperc*, certainly, but also socioeconomic status, age, physical stamina, vocabulary, and every other trait or ability that might affect scores on *visperc* but that does not appear elsewhere in the model.

The model presented here is a common factor analysis model. In the lingo of common factor analysis, the unobserved variable *spatial* is called a **common factor**, and the three unobserved variables, *err_v*, *err_c*, and *err_l*, are called **unique factors**. The path diagram shows another common factor, *verbal*, on which the last three tests depend. The path diagram also shows three more unique factors, *err_p*, *err_s*, and *err_w*. The two common factors, *spatial* and *verbal*, are allowed to be correlated. On the other hand, the unique factors are assumed to be uncorrelated with each other and with the common factors. The path coefficients leading from the common factors to the observed variables are sometimes called **factor loadings**.

Identification

This model is identified except that, as usual, the measurement scale of each unobserved variable is indeterminate. The measurement scale of each unobserved variable can be established arbitrarily by setting its regression weight to a constant, such as 1, in some regression equation. The preceding path diagram shows how to do this. In that path diagram, eight regression weights are fixed at 1, which is one fixed regression weight for each unobserved variable. These constraints are sufficient to make the model identified.

The proposed model is a particularly simple common factor analysis model, in that each observed variable depends on just one common factor. In other applications of common factor analysis, an observed variable can depend on any number of common factors at the same time. In the general case, it can be very difficult to decide whether a common factor analysis model is identified or not (Davis, 1993; Jöreskog, 1969, 1979). The discussion of identifiability given in this and earlier examples made the issue appear simpler than it actually is, giving the impression that the lack of a natural unit of measurement for unobserved variables is the sole cause of non-identification. It is true that the lack of a unit of measurement for unobserved variables is an ever-present cause of non-identification. Fortunately, it is one that is easy to cure, as we have done repeatedly.

But other kinds of under-identification can occur for which there is no simple remedy. Conditions for identifiability have to be established separately for individual models. Jöreskog and Sörbom (1984) show how to achieve identification of many models by imposing equality constraints on their parameters. In the case of the factor analysis model (and many others), figuring out what must be done to make the model identified requires a pretty deep understanding of the model. If you are unable to tell whether a model is identified, you can try fitting the model in order to see whether Amos reports that it is unidentified. In practice, this empirical approach works quite well, although there are objections to it in principle (McDonald and Krane, 1979), and it is no substitute for an *a priori* understanding of the identification status of a model. Bollen (1989) discusses causes and treatments of many types of non-identification in his excellent textbook.

Specifying the Model

Amos analyzes the model directly from the path diagram shown on p. 138. Notice that the model can conceptually be separated into *spatial* and *verbal* branches. You can use the structural similarity of the two branches to accelerate drawing the model.

Drawing the Model

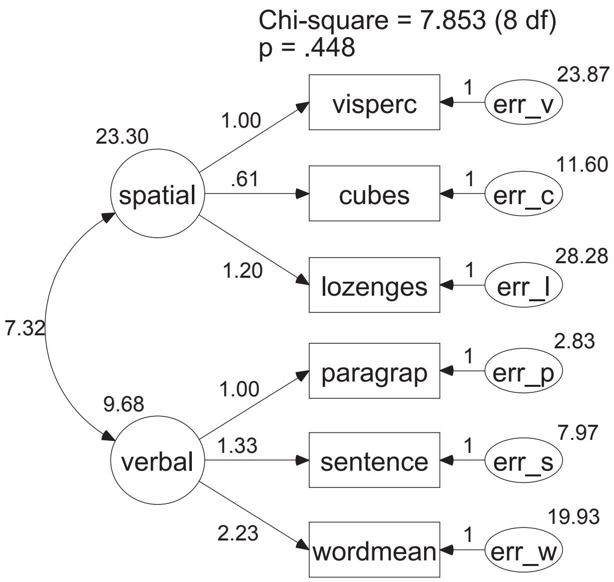
After you have drawn the first branch:

- From the menus, choose Edit → Select All to highlight the entire branch.
- To create a copy of the entire branch, from the menus, choose Edit → Duplicate and drag one of the objects in the branch to another location in the path diagram.

Be sure to draw a double-headed arrow connecting *spatial* and *verbal*. If you leave out the double-headed arrow, Amos will assume that the two common factors are uncorrelated. The input file for this example is *Ex08.amw*.

Results of the Analysis

Here are the unstandardized results of the analysis. As shown at the upper right corner of the figure, the model fits the data quite well.



Example 8
Factor analysis: Girls' sample
Holzinger and Swineford (1939)
Unstandardized estimates

As an exercise, you may wish to confirm the computation of degrees of freedom.

Computation of degrees of freedom: (Default model)	
Number of distinct sample moments:	21
Number of distinct parameters to be estimated:	13
Degrees of freedom (21 – 13):	8

The parameter estimates, both standardized and unstandardized, are shown next. As you would expect, the regression weights are positive, as is the correlation between spatial ability and verbal ability.

Regression Weights: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
visperc	<---	spatial	1.000				
cubes	<---	spatial	.610	.143	4.250	***	
lozenges	<---	spatial	1.198	.272	4.405	***	
paragrap	<---	verbal	1.000				
sentence	<---	verbal	1.334	.160	8.322	***	
wordmean	<---	verbal	2.234	.263	8.482	***	

Standardized Regression Weights: (Group number 1 - Default model)

			Estimate
visperc	<---	spatial	.703
cubes	<---	spatial	.654
lozenges	<---	spatial	.736
paragrap	<---	verbal	.880
sentence	<---	verbal	.827
wordmean	<---	verbal	.841

Covariances: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
spatial	<-->	verbal	7.315	2.571	2.846	.004	

Correlations: (Group number 1 - Default model)

			Estimate
spatial	<-->	verbal	.487

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
spatial	23.302	8.123	2.868	.004	
verbal	9.682	2.159	4.485	***	
err_v	23.873	5.986	3.988	***	
err_c	11.602	2.584	4.490	***	
err_l	28.275	7.892	3.583	***	
err_p	2.834	.868	3.263	.001	
err_s	7.967	1.869	4.263	***	
err_w	19.925	4.951	4.024	***	

Obtaining Standardized Estimates

To get the standardized estimates shown above, do the following before you perform the analysis:

- ▶ From the menus, choose View → Analysis Properties.
- ▶ In the Analysis Properties dialog box, click the Output tab.
- ▶ Select Standardized estimates (a check mark appears next to it).

▶ Also select Squared multiple correlations if you want squared multiple correlation for each endogenous variable, as shown in the next graphic.

▶ Close the dialog box.

Squared Multiple Correlations: (Group number 1 - Default model)

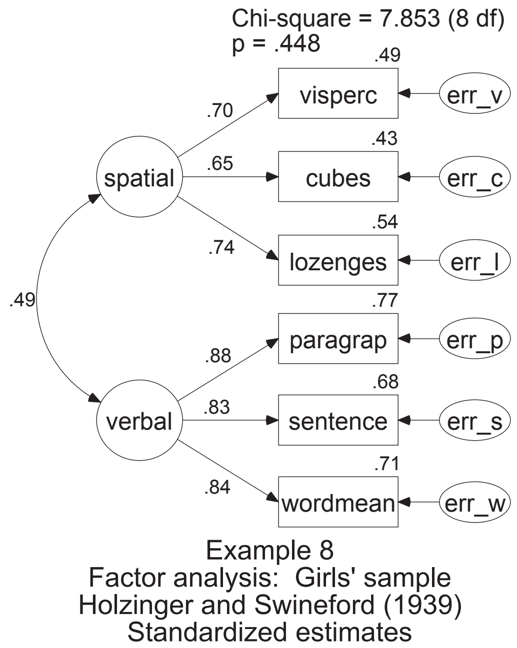
	Estimate
wordmean	.708
sentence	.684
paragrap	.774
lozenges	.542
cubes	.428
visperc	.494

Viewing Standardized Estimates

▶ In the Amos Graphics window, click the Show the output path diagram button.

▶ Select Standardized estimates in the Parameter Formats panel at the left of the path diagram.

Here is the path diagram with standardized estimates displayed:



The squared multiple correlations can be interpreted as follows: To take *wordmean* as an example, 71% of its variance is accounted for by verbal ability. The remaining 29% of its variance is accounted for by the unique factor *err_w*. If *err_w* represented measurement error only, we could say that the estimated reliability of *wordmean* is 0.71. As it is, 0.71 is an estimate of a lower-bound on the reliability of *wordmean*.

The Holzinger and Swineford data have been analyzed repeatedly in textbooks and in demonstrations of new factor analytic techniques. The six tests used in this example are taken from a larger subset of nine tests used in a similar example by Jöreskog and Sörbom (1984). The factor analysis model employed here is also adapted from theirs. In view of the long history of exploration of the Holzinger and Swineford data in the factor analysis literature, it is no accident that the present model fits very well. Even more than usual, the results presented here require confirmation on a fresh set of data.

Modeling in VB.NET

The following program specifies the factor model for Holzinger and Swineford's data. It is saved in the file *Ex08.vb*.

```
Sub Main()
    Dim Sem As New AmosEngine
    Try
        Sem.TextOutput()
        Sem.Standardized()
        Sem.Smc()

        Sem.BeginGroup(Sem.AmosDir & "Examples\Grnt_fem.sav")
        Sem.AStructure("visperc   = (1) spatial + (1) err_v")
        Sem.AStructure("cubes     =     spatial + (1) err_c")
        Sem.AStructure("lozenges  =     spatial + (1) err_l")

        Sem.AStructure("paragrap = (1) verbal  + (1) err_p")
        Sem.AStructure("sentence  =     verbal  + (1) err_s")
        Sem.AStructure("wordmean  =     verbal  + (1) err_w")
        Sem.FitModel()
    Finally
        Sem.Dispose()
    End Try
End Sub
```

You do not need to explicitly allow the factors (*spatial* and *verbal*) to be correlated. Nor is it necessary to specify that the unique factors be uncorrelated with each other and with the two factors. These are default assumptions in an Amos program (but not in Amos Graphics).