Confirmatory factor analysis of reader reliability

Contents

1. Introduction

Confirmatory factor analysis, in contrast to exploratory factor analysis where both the number of factors and the meaning of the factors are initially unknown, begins by defining the latent variables of interest. The observable variables are then constructed to measure these latent variables. In essence, in a confirmatory factor analysis the number of factors is known and equal to the number of latent variables.

In this example, we consider the special case where there is a single latent variable. The latent variable of interest here is reader reliability in essay scoring in the case of 126 respondents who were given a three-part English Composition examination. Scores are available on four aspects:

- The original part 1 essay (ORIGPRT1)
- a handwritten copy of the original part 1 essay (WRITCOPY)
- a carbon copy of the handwritten copy, and (CARBCOPY)
- the original part 2 essay (ORIGPRT2).

These scores were assigned by a group of readers. The question of interest here is whether the four scores can be used interchangeably or whether scores on the copies (the second and third scores) were less reliable than for the originals (scores one and four). The covariance matrix of the four scores is used as input and is as given in the table below. These data are from Votaw (1948).

	x_1	x_2	x_3	x_4
ORIGPRT1	25.0704			
WRITCOPY	12.4363	28.2021		
CARBCOPY	11.7257	9.2281	22.7390	
ORIGPRT2	20.7510	11.9732	12.0692	21.8707

2. Congeneric model

We wish to test the hypotheses that these measurements are parallel, tau-equivalent and congeneric, respectively. We will use maximum likelihood in all three cases.

To estimate the congeneric model, we use the following syntax (**votaw1a.spl**)

```
votaw1a.spl
Analysis of Reader Reliability in Essay Scoring; Votaw's Data
Congeneric model estimated by ML
Observed Variables: ORIGPRT1  WRITCOPY  CARBCOPY  ORIGPRT2
Covariance Matrix
25.0704
12.4363     28.2021
11.7257      9.2281     22.7390
20.7510     11.9732     12.0692     21.8707
Sample Size: 126
Latent Variable: Ability
Relationship
ORIGPRT1 - ORIGPRT2 = Ability
Path Diagram
End of Problem
```

For this model, the following goodness-of-fit measures are obtained:

```
                              Log-likelihood Values

                        Estimated Model          Saturated Model
                        ---------------          ---------------
Number of free parameters(t)          8                       10
-2ln(L)                        1840.667                 1838.387
AIC (Akaike, 1974)*            1856.667                 1858.387
BIC (Schwarz, 1978)*           1879.294                 1886.670

*LISREL uses AIC= 2t - 2ln(L) and BIC = tln(N)- 2ln(L)


                           Goodness-of-Fit Statistics

Degrees of Freedom for (C1)-(C2)                      2
Maximum Likelihood Ratio Chi-Square (C1)              2.280 (P = 0.3198)
Browne's (1984) ADF Chi-Square (C2_NT)                2.218 (P = 0.3299)
```

Note that this model can also be fitted using LISREL syntax: see **votaw1b.lis**.

3. Tau-equivalent model

Turning to the hypothesis of tau-equivalence, we use slightly amended syntax (see **votaw2b.lis**). The key difference between these two LISREL syntax files is the addition of the EQ command in the syntax for the tau-equivalence model.

votaw2b.lis

```
Analysis of Reader Reliability in Essay Scoring; Votaw's Data
Tau-equivalent model estimated by ML
DA NI=4 NO=126
LA
ORIGPRT1  WRITCOPY  CARBCOPY  ORIGPRT2
CM
25.0704
12.4363     28.2021
11.7257      9.2281     22.7390
20.7510     11.9732     12.0692     21.8707
MO NX=4 NK=1 LX=FR
LK
Ability
EQ LX(1) - LX(4)
PD
OU
```

For this model, the following goodness-of-fit measures are obtained:

```
                                  Log-likelihood Values

                         Estimated Model          Saturated Model
                         ---------------          ---------------
 Number of free parameters(t)          5                       10
 -2ln(L)                        1878.811                 1838.387
 AIC (Akaike, 1974)*            1888.811                 1858.387
 BIC (Schwarz, 1978)*           1902.952                 1886.670

*LISREL uses AIC= 2t - 2ln(L) and BIC = tln(N)- 2ln(L)


                            Goodness-of-Fit Statistics

 Degrees of Freedom for (C1)-(C2)                       5
 Maximum Likelihood Ratio Chi-Square (C1)               40.423 (P = 0.0000)
 Browne's (1984) ADF Chi-Square (C2_NT)                 46.375 (P = 0.0000)
```

4. Parallel measures model

In the case of the third hypothesis, we consider the hypothesis that measurements are parallel. The syntax for this is as follows. Note the changes to the EQ statement(s).

```
votaw3b.lis
Analysis of Reader Reliability in Essay Scoring; Votaw's Data
Parallel model estimated by ML
DA NI=4 NO=126
LA
ORIGPRT1  WRITCOPY  CARBCOPY  ORIGPRT2
CM
25.0704
12.4363     28.2021
11.7257      9.2281     22.7390
20.7510     11.9732     12.0692     21.8707
MO NX=4 NK=1 LX=FR
LK
Ability
EQ LX(1) - LX(4)
EQ TD(1) - TD(4)
PD
OU
```

For this model, the following goodness-of-fit measures are obtained:

```
                                Log-likelihood Values

                        Estimated Model          Saturated Model
                        ---------------          ---------------
 Number of free parameters(t)        2                       10
 -2ln(L)                      1947.510                 1838.387
 AIC (Akaike, 1974)*          1951.510                 1858.387
 BIC (Schwarz, 1978)*         1957.167                 1886.670

*LISREL uses AIC= 2t - 2ln(L) and BIC = tln(N)- 2ln(L)


                           Goodness-of-Fit Statistics

 Degrees of Freedom for (C1)-(C2)                        8
 Maximum Likelihood Ratio Chi-Square (C1)                109.123 (P = 0.0000)
 Browne's (1984) ADF Chi-Square (C2_NT)                  63.684 (P = 0.0000)
```

When the results for these three models are compared, we note that the hypothesis of a congeneric model is the most acceptable with a chi-square of 2.298 and 2 degrees of freedom. Both the hypotheses of parallel measurements and a tau-equivalent model are rejected.