



## Multivariate regression

### Contents

1. Introduction .....	1
2. Testing hypotheses using LISREL syntax .....	2
3. Testing that error terms are uncorrelated .....	3

### 1. Introduction

To illustrate hypothesis testing in the case of multivariate regression, we use data collected by Dr. William Rohwer (see Timm, 1975). Data were collected on children in kindergarten and the question of interest was to predict test scores from paired-associated (PA) learning proficiency developed by Dr. Rohwer.

The data file **rohwer.isf** shown below contains information on 37 children with low socio-economic status and 32 children with high socio-economic status. The data and syntax files can be found in the MVABOOK examples\Chapter2 folder.

The variables are:

- SES: indicator of socio-economic status. Code 0 for children with low SES, 1 for children with high SES.
- SAT: score on a Student Achievement Test
- PPVT: score on the Peabody Picture Vocabulary Test
- Raven: score on the Raven Progressive Matrices Test
- named: performance on a 'named' PA test
- Still: performance on a 'Still' PA task
- NamedSti: performance on a 'named still' PA task
- NamedAct: performance on a 'named action' PA task
- SentStil: performance on a 'sentence still' PA task.

	SES	SAT	PPVT	Raven	named	Still	NamedSti	Namedact	SentStil
1	0.00	49.00	48.00	8.00	1.00	2.00	6.00	12.00	16.00
2	0.00	47.00	76.00	13.00	5.00	14.00	14.00	30.00	27.00
3	0.00	11.00	40.00	13.00	0.00	10.00	21.00	16.00	16.00
4	0.00	9.00	52.00	9.00	0.00	2.00	5.00	17.00	8.00
5	0.00	69.00	63.00	15.00	2.00	7.00	11.00	26.00	17.00
6	0.00	35.00	82.00	14.00	2.00	15.00	21.00	34.00	25.00
7	0.00	6.00	71.00	21.00	0.00	1.00	20.00	23.00	18.00
8	0.00	8.00	68.00	8.00	0.00	0.00	10.00	19.00	14.00
9	0.00	49.00	74.00	11.00	0.00	0.00	7.00	16.00	13.00
10	0.00	8.00	70.00	15.00	3.00	2.00	21.00	26.00	25.00
11	0.00	47.00	70.00	15.00	8.00	16.00	15.00	35.00	24.00
12	0.00	6.00	61.00	11.00	5.00	4.00	7.00	15.00	14.00
13	0.00	14.00	54.00	12.00	1.00	12.00	13.00	27.00	21.00
14	0.00	30.00	55.00	13.00	2.00	1.00	12.00	20.00	17.00
15	0.00	4.00	54.00	10.00	3.00	12.00	20.00	26.00	22.00
16	0.00	24.00	40.00	14.00	0.00	2.00	5.00	14.00	8.00
17	0.00	19.00	66.00	13.00	7.00	12.00	21.00	35.00	27.00
18	0.00	45.00	54.00	10.00	0.00	6.00	6.00	14.00	16.00
19	0.00	22.00	64.00	14.00	12.00	8.00	19.00	27.00	26.00
20	0.00	16.00	47.00	16.00	3.00	9.00	15.00	18.00	10.00

We are interested in using the five performance assessments to predict the scores on the SAT, PPVT and Raven scores. Another question to answer is whether the effective predictors in such an analysis are the same for the two socio-economic groups.

## 2. Testing hypotheses using LISREL syntax

The syntax file **rohwer1b1.lis** illustrates the regression of SAT, PPVT and Raven on named, Still, NameSti, Namedact and SentStil. Under the assumption that the three error terms are correlated, we first test the hypothesis that

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 & x & x & 0 \\ x & 0 & 0 & 0 & x & 0 \\ x & 0 & x & 0 & 0 & 0 \end{bmatrix}$$

where 0 is a fixed zero and  $x$  is a free parameter.

```
rohwer1b1.lis
!Testing hypothesis on GAMMA
DA NI=9
RA=rohwer.lsf
SE
SAT PPVT Raven SES named Still NamedSti Namedact SentStil|
MO NY=3 NX=6 AL=FR KA=FR PS=SY,FR
PA GA
0 0 0 1 1 0
1 0 0 0 1 0
1 0 1 0 0 0
OU
```

To specify the hypothesis on  $\Gamma$  one can use a pattern matrix where 0 represents a fixed parameter and 1 represents a free parameter. All fixed parameters are 0 by default.

The following results were obtained. Judging by the chi-square values and associated  $p$ -values the hypothesis cannot be rejected. This suggest that many of the predictor variables do not have any significant effect on SAT, PPVT and Raven scores.

Log-likelihood Values

	Estimated Model	Saturated Model
Number of free parameters (t)	42	54
-2ln(L)	2388.458	2373.507
AIC (Akaike, 1974)*	2472.458	2481.507
BIC (Schwarz, 1978)*	2566.290	2602.149

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

Goodness-of-Fit Statistics

Degrees of Freedom for (C1)-(C2)	12
Maximum Likelihood Ratio Chi-Square (C1)	14.950 (P = 0.2442)
Due to Covariance Structure	14.950
Due to Mean Structure	0.00
Browne's (1984) ADF Chi-Square (C2_NT)	14.269 (P = 0.2838)

GAMMA

	SES	named	Still	NamedSti	Namedact	SentStil
SAT	--	--	--	-2.239 (0.779) -2.875	2.980 (0.699) 4.266	--
PPVT	16.751 (2.559) 6.545	--	--	--	1.268 (0.207) 6.141	--
Raven	1.216 (0.666) 1.826	--	0.231 (0.076) 3.030	--	--	--

Note that named and SentStil have no effect on the three outcome variables. Still only has an effect on the Raven score, and NamedSti on the SAT score.

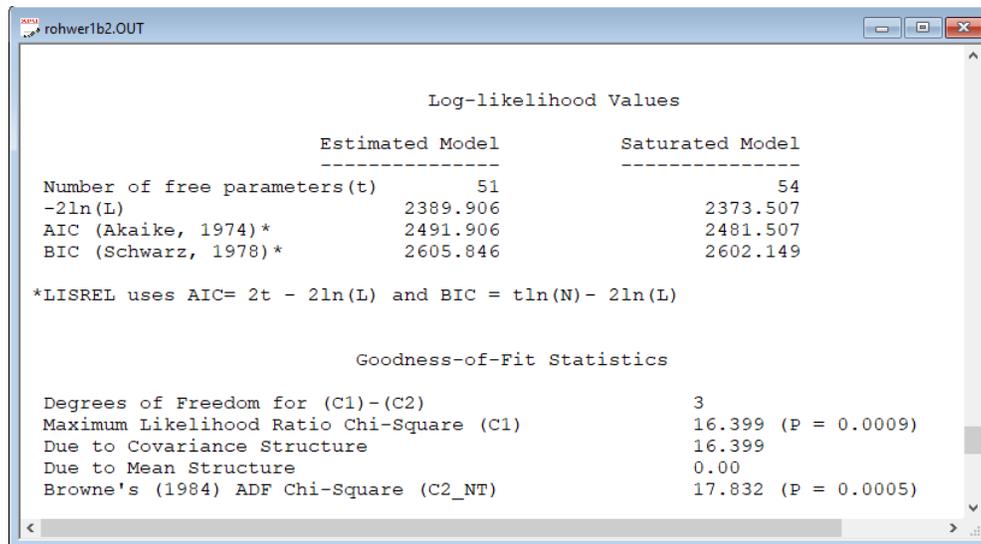
### 3. Testing that error terms are uncorrelated

We now assume that  $\Gamma$  is free and focus on testing the hypothesis that  $\Psi$  is diagonal, which implies that the error terms are uncorrelated. LISREL assumes PSI to be diagonal by default, so the model can be estimated using the simpler syntax shown below.

```

rohwer1b2.lis
!Rohwer Data: Regression on Pooled Data
!Testing hypothesis on PSI
DA NI=9
RA=rohwer.lsf
SE
SAT PPVT Raven SES named Still NamedSti Namedact SentStil
MO NY=3 NX=6 AL=FR KA=FR
OU
  
```

The output obtained is as follows:



rohwer1b2.OUT

Log-likelihood Values

	Estimated Model	Saturated Model
Number of free parameters(t)	51	54
-2ln(L)	2389.906	2373.507
AIC (Akaike, 1974)*	2491.906	2481.507
BIC (Schwarz, 1978)*	2605.846	2602.149

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

Goodness-of-Fit Statistics

Degrees of Freedom for (C1)-(C2)	3
Maximum Likelihood Ratio Chi-Square (C1)	16.399 (P = 0.0009)
Due to Covariance Structure	16.399
Due to Mean Structure	0.00
Browne's (1984) ADF Chi-Square (C2_NT)	17.832 (P = 0.0005)

Both chi-squares are highly significant, indicating that the hypothesis should be rejected.