Multivariate analysis

Contents

1. Introduction

Snijders and Bosker analyzed the data used in the example extensively. The data are on school students in the Netherlands, and information is available on various test scores. To illustrate the fitting of a three-level multivariate model, we use a subset of these data. The data are given **in nlschools.lsf**. Data and syntax files can be found in the **MVABOOK\Chapter4** folder.

nlschools.lsf

	School	Student	IQ_verb	IQ_perf	aritpret	aritpost	langpret	langpost
1	1.00	1.00	15.00	12.33	14.00	24.00	36.00	46.00
2	1.00	2.00	14.50	10.00	12.00	19.00	36.00	45.00
3	1.00	3.00	9.50	11.00	10.00	24.00	33.00	33.00
4	1.00	4.00	11.00	10.00	13.00	26.00	29.00	46.00
5	1.00	5.00	8.00	6.67	8.00	9.00	19.00	20.00
6	1.00	6.00	9.50	9.00	8.00	13.00	22.00	30.00
7	1.00	7.00	9.50	10.33	7.00	13.00	20.00	30.00
8	1.00	8.00	13.00	14.33	17.00	30.00	44.00	57.00
9	1.00	9.00	9.50	8.67	10.00	23.00	34.00	36.00
10	1.00	10.00	11.00	15.00	14.00	22.00	31.00	36.00
11	1.00	11.00	5.50	9.00	11.00	19.00	18.00	29.00
12	1.00	12.00	14.00	9.33	10.00	23.00	36.00	40.00
13	1.00	13.00	9.00	9.00	10.00	18.00	31.00	41.00
14	1.00	14.00	10.50	13.00	12.00	22.00	34.00	47.00
15	1.00	15.00	10.00	8.67	9.00	15.00	31.00	33.00
16	1.00	16.00	11.00	12.33	13.00	21.00	32.00	37.00
17	1.00	17.00	6.50	7.33	9.00	13.00	23.00	29.00
18	1.00	18.00	4.00	9.33	9.00	12.00	20.00	26.00
19	1.00	19.00	11.00	6.67	7.00	11.00	27.00	37.00
	1.00	20.00	11.00	11.00	16.00	23.00	31.00	40.00

The variables are:

- School: the school the student is from, used as the level-3 ID in this analysis.
- Student: The student ID, used as level-2 ID, with the multiple test scores nested within each student.
- IQ_verb: an IQ test score

- IQ_perf: an additional IQ test score
- aritpret: the first of two arithmetic test scores
- aritpost: the second arithmetic test score
- langpret: the first of two language test scores
- langpost: the second language test score

All students were in the eighth grade at the time of measurement. The variables in this data are a subset of that used by Snijders & Bosker.

2. Multivariate model

We are interested in exploring the between and within variation in scores. To fit the multivariate model, we consider the response variable to be

$$\mathbf{y}_{ij} = (IQ\_verb_{ij}, IQ\_perf_{ij}, aritpret_{ij}, aritpost_{ij}, langpret_{ij}, langpost_{ij})$$

with our model written as

$$\mathbf{y}_{ij} = \boldsymbol{\alpha} + \mathbf{u}_i + \mathbf{e}_{ij},$$

where *i* indicates a student and *j* a school. As the test scores are nested within students, the student ID is used as level-2 ID. Students are in turn nested within schools, and the school number serves as level-3 ID. The vector $\boldsymbol{\alpha}$ represents the grand mean vector, and $\mathbf{u}_i$ and $\mathbf{e}_{ij}$ the random effects associated with student *i* and school *j*. Syntax for this model is shown below.

nlschools2.PRL

```
OPTIONS SUMMARY=NONE COVBW=YES;
TITLE=Within and Between Covariance Matrix for 6 Test Scores;
SY=nlschools.lsf;
ID3=School;
ID2=Student;
RESPONSE=IQ_verb IQ_perf aritpret aritpost langpret langpost;
FIXED=intcept;
RANDOM1=intcept;
RANDOM2=intcept;
RANDOM3=intcept;
```

To obtain the estimates of the between and within group covariance matrices, the COVBW option is used on the Options command. Results are given below.

```
                    +-----------------------+
                    | FIXED PART OF MODEL   |
                    +-----------------------+

------------------------------------------------------------------------------
  COEFFICIENTS             BETA-HAT      STD.ERR.      Z-VALUE       PR > |Z|
------------------------------------------------------------------------------
  intcept1                11.75097       0.07663     153.33900       0.00000
  intcept2                10.98571       0.06392     171.85698       0.00000
  intcept3                11.73904       0.15190      77.28386       0.00000
  intcept4                18.95263       0.33508      56.56127       0.00000
  intcept5                33.88039       0.26608     127.33025       0.00000
  intcept6                40.35982       0.41861      96.41444       0.00000
```

The fixed part of the model represents the estimates of α, and we see that all effects are highly significant. This information is followed by the level-3 and level-2 covariance matrices, i.e., the between- and within-school matrices. We note that the variation over scores within schools (over students) is higher than between schools. When the *z*-values associated with the variances and covariances are considered (given in output, not shown here) we note that there is significant variation and covariation between all 6 scores.

```
LEVEL 3 COVARIANCE MATRIX

              intcept1    intcept2    intcept3    intcept4    intcept5

 intcept1      0.52351
 intcept2      0.26196     0.25167
 intcept3      0.87099     0.45958     2.37204
 intcept4      1.93179     1.24817     4.19631    12.53172
 intcept5      1.61381     0.94915     3.21832     6.93619     6.73197
 intcept6      2.50844     1.58392     4.93494    14.26373     9.03740


              intcept6

 intcept6     18.66449


       LEVEL 2 COVARIANCE MATRIX

              intcept1    intcept2    intcept3    intcept4    intcept5

 intcept1      3.83022
 intcept2      1.71733     4.62199
 intcept3      2.49665     3.26362     9.82812
 intcept4      5.59173     6.74055    10.94488    32.27615
 intcept5      7.36608     5.61583     9.49776    19.02252    38.96517
 intcept6      9.27566     6.88912    12.31897    28.66060    35.96826


              intcept6

 intcept6     64.69375
```

To use this information, we opted to use the COVBW option as noted previously. Through use of this option, the between- and within-schools covariance matrices are automatically saved to the text files **nlschools2_between.cov** and **nlschools2_within.cov**. These files can serve as input for, for example, multiple-group structural equation modeling. If this is the intended use, note that the sample sizes can be found in the first part of the output file.

```
                          +--------------+
                          | DATA SUMMARY |
                          +--------------+

 NUMBER OF LEVEL 3 UNITS :      131
 NUMBER OF LEVEL 2 UNITS :     2287
 NUMBER OF LEVEL 1 UNITS :    13722


     Adjusted between cluster sample size=     17
     Within cluster sample size=     2287
```