Scientific Software International

Multilevel model: math achievement and SES

Contents

1. Introduction

In this example, we analyze data from Bryk & Raudenbush (1992). The High School and Beyond Study contain data on the student scores along with demographic information. In addition, information on school characteristics is also available.

The file **Mathach.lsf** contains six variables. Data and syntax files can be found in the **MVABOOK\Chapter4** folder.

These are:

- ses: a measure of the student's socio-economic status
- mathach: a standardized measure of the student's mathematics achievement
- school: the school identifier, to be used as level-2 ID in our analysis
- sector: a school characteristic that indicates whether the school is a public school (sector = 0) or a catholic school (sector = 1).
- meanses: a school characteristic representing the average of the student ses values for all students within that school
- cses: the difference between the individual student's ses and the meanses value of the school the student attends. In effect, the group-mean centered value of ses for the student.

mathach.lsf

	ses	mathach	school	sector	meanses	cses
1	-1.53	5.88	1224.00	0.00	-0.43	-1.10
2	-0.59	19.71	1224.00	0.00	-0.43	-0.16
3	-0.53	20.35	1224.00	0.00	-0.43	-0.10
4	-0.67	8.78	1224.00	0.00	-0.43	-0.24
5	-0.16	17.90	1224.00	0.00	-0.43	0.27
6	0.02	4.58	1224.00	0.00	-0.43	0.45
7	-0.62	-2.83	1224.00	0.00	-0.43	-0.19
8	-1.00	0.52	1224.00	0.00	-0.43	-0.57
9	-0.89	1.53	1224.00	0.00	-0.43	-0.46
10	-0.46	21.52	1224.00	0.00	-0.43	-0.03
11	-1.45	9.47	1224.00	0.00	-0.43	-1.02
12	-0.66	16.06	1224.00	0.00	-0.43	-0.23
13	-0.47	21.18	1224.00	0.00	-0.43	-0.04
14	-0.99	20.18	1224.00	0.00	-0.43	-0.56
15	0.33	20.35	1224.00	0.00	-0.43	0.76
16	-0.68	20.51	1224.00	0.00	-0.43	-0.25
17	-0.30	19.34	1224.00	0.00	-0.43	0.13
18	-1.53	4.14	1224.00	0.00	-0.43	-1.10
19	0.04	2.93	1224.00	0.00	-0.43	0.47
20	-0.08	16.41	1224.00	0.00	-0.43	0.35

2. Unconditional model

In order to explore differences in the mathematic scores in high schools, we start by fitting a simple model, also known as a fully unconditional model.

$$Mathach_{ij} = \alpha_i + u_i + e_{ij},$$

where $Mathach_{ij}$ is modelled as having a fixed effect α_i within each school *i* and with random variation both within and between schools (e_{ij} and u_i) respectively. The syntax file below specifies this model.

mathach1.PRL

```
OPTIONS;
TITLE=Mathach Model 1;
SY=mathach.lsf;
ID2=school;
RESPONSE=mathach;
FIXED=intcept;
RANDOM1=intcept;
RANDOM2=intcept;
```

For this model, the following results were obtained:

```
                 +-----------------------+
                 | FIXED PART OF MODEL   |
                 +-----------------------+
------------------------------------------------------------------------------
 COEFFICIENTS            BETA-HAT      STD.ERR.      Z-VALUE       PR > |Z|
------------------------------------------------------------------------------
 intcept                 12.63707      0.24362       51.87260      0.00000
```

```
                       +-----------------------+
                       |   -2 LOG-LIKELIHOOD   |
                       +-----------------------+

  DEVIANCE= -2*LOG(LIKELIHOOD) =    47115.8102215771
  NUMBER OF FREE PARAMETERS =                  3

                       +-----------------------+
                       | RANDOM PART OF MODEL  |
                       +-----------------------+
------------------------------------------------------------------------------
  LEVEL 2                        TAU-HAT      STD.ERR.     Z-VALUE    PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept               8.55346      1.06121      8.06010    0.00000

------------------------------------------------------------------------------
  LEVEL 1                        TAU-HAT      STD.ERR.     Z-VALUE    PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept              39.14840      0.66054     59.26709    0.00000
```

The mean mathematics score over the schools is estimated at 12.637 and is statistically significantly. We also note significant variation in scores over students (within schools, in other words) and between schools. Most of the variation in scores seem to reside at the student level.

3. Taking school characteristics into account

To see whether the socio-economic scores of the schools explain some of the variation in the math scores, we now fit a conditional model using meanses as a predictor:

mathach2.PRL

```
OPTIONS SUMMARY=NONE;
TITLE=Mathach Model 2;
SY=mathach.lsf;
ID2=school;
RESPONSE=mathach;
FIXED=intcept meanses;
RANDOM1=intcept;
RANDOM2=intcept;
```

After convergence, the following estimates are obtained:

```
                       +-----------------------+
                       | FIXED PART OF MODEL   |
                       +-----------------------+
------------------------------------------------------------------------------
  COEFFICIENTS                BETA-HAT       STD.ERR.      Z-VALUE       PR > |Z|
------------------------------------------------------------------------------
  intcept                    12.64974       0.14832       85.28728       0.00000
  meanses                     5.86292       0.35913       16.32541       0.00000
```

```
                      +-----------------------+
                      |   -2 LOG-LIKELIHOOD   |
                      +-----------------------+

  DEVIANCE= -2*LOG(LIKELIHOOD) =    46959.1085280252
  NUMBER OF FREE PARAMETERS =                4

                      +-----------------------+
                      | RANDOM PART OF MODEL  |
                      +-----------------------+
------------------------------------------------------------------------------
  LEVEL 2                       TAU-HAT      STD.ERR.     Z-VALUE    PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept              2.59319      0.39249      6.60710    0.00000

------------------------------------------------------------------------------
  LEVEL 1                       TAU-HAT      STD.ERR.     Z-VALUE    PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept             39.15732      0.66065     59.27091    0.00000
```

The estimate for meanses is highly significant. There is also still significant random variation both within and between schools. The deviance statistic for this model is 46959. For the previous, unconditional model, it was 47115. The differences between these can be used as a chi-square with 1 degree of freedom (the second model had one more parameter than the first) to gauge whether the addition of the predictor meanses contributed to the overall explanation of variation in outcome. We conclude from this model that the mean socio-economic status of a school plays a significant role explaining differences in math scores: schools with higher mean socio-economic status will also have higher expected mean mathematics scores as indicated by the positive estimated coefficient of 5.86292. When we take into account that the mean scores range from -1.188 to 0.831 over the schools in the sample, that would imply a suggested increase of approximately 2 (5.86292) in expected maths scores between schools with minimum SES and those with maximum SES.

4. Taking student characteristics into account

Given the relationship between the mean expected maths score for schools and their socio-economic status, it seems logical to also take a closer look at how the students' individual SES may impact their expected scores. We extend the model to include the individual measure of SES as well.

mathach3.PRL

```
OPTIONS SUMMARY=NONE;
TITLE=Mathach Model 3;
SY=mathach.lsf;
ID2=school;
RESPONSE=mathach;
FIXED=intcept cses;
RANDOM1=intcept;
RANDOM2=intcept cses;
```

```
                      +-----------------------+
                      | FIXED PART OF MODEL   |
                      +-----------------------+
------------------------------------------------------------------------------
  COEFFICIENTS             BETA-HAT      STD.ERR.      Z-VALUE       PR > |Z|
------------------------------------------------------------------------------
  intcept                 12.64943      0.24374      51.89774       0.00000
```

```
  cses                         2.19315       0.12784      17.15575       0.00000

                       +-----------------------+
                       |   -2 LOG-LIKELIHOOD    |
                       +-----------------------+

  DEVIANCE= -2*LOG(LIKELIHOOD) =     46710.9801965303
  NUMBER OF FREE PARAMETERS =                  6

                       +-----------------------+
                       | RANDOM PART OF MODEL   |
                       +-----------------------+
------------------------------------------------------------------------------
  LEVEL 2                       TAU-HAT      STD.ERR.     Z-VALUE    PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept              8.62107      1.06231      8.11537    0.00000
  cses    /intcept              0.05040      0.39315      0.12821    0.89798
  cses    /cses                 0.67827      0.28418      2.38677    0.01700

------------------------------------------------------------------------------
  LEVEL 1                       TAU-HAT      STD.ERR.     Z-VALUE    PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept             36.70004      0.62593     58.63283    0.00000
```

The estimate for cses is highly significant. The positive estimated effect indicates an expected increase in math score with higher individual SES. Note that, due to the centering performed on this variable, the estimated intercept in this model represents the expected mean math score for students with SES equal to the mean SES for their school. There is also still significant random variation both within and between schools, though no evidence of significant covariation between cses and the intercept term. It would probably be adequate to describe the level-2 variance/covariance with a diagonal matrix instead of a full matrix. We again also see a sizable decrease in the deviance statistic when compared to that for the previous model. Note that this model had 2 more parameters than the previous model.

5. Combined model

The question remains whether, after taking both individual and school level SES into account, the type of school a student attends plays a role in the modeling of the variation in the maths scores. To investigate this, we fit a final model in which we take all the previous predictors into account along with interactions between them.

The syntax for this model is given in **mathach4.prl**:

mathach4.PRL

```
OPTIONS SUMMARY=NONE;
TITLE=Mathach Model 4;
SY=mathach.lsf;
ID2=school;
RESPONSE=mathach;
FIXED=intcept cses meanses sector cses*meanses cses*sector;
RANDOM1=intcept;
RANDOM2=intcept cses;
```

For this model we obtain the following results:

```
                       +-----------------------+
                       | FIXED PART OF MODEL   |
                       +-----------------------+
------------------------------------------------------------------------------
  COEFFICIENTS              BETA-HAT      STD.ERR.      Z-VALUE       PR > |Z|
------------------------------------------------------------------------------
  intcept                  12.11360       0.19692      61.51554       0.00000
  cses                      2.93936       0.15350      19.14949       0.00000
  meanses                   5.33795       0.36569      14.59696       0.00000
  sector                    1.21694       0.30338       4.01132       0.00006
  cses    *meanses          1.04237       0.29604       3.52108       0.00043
  cses    *sector          -1.64387       0.23736      -6.92564       0.00000

                       +-----------------------+
                       |   -2 LOG-LIKELIHOOD   |
                       +-----------------------+

  DEVIANCE= -2*LOG(LIKELIHOOD) =    46496.4338826074
  NUMBER OF FREE PARAMETERS =              10

                       +-----------------------+
                       | RANDOM PART OF MODEL  |
                       +-----------------------+
------------------------------------------------------------------------------
  LEVEL 2                         TAU-HAT      STD.ERR.     Z-VALUE   PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept                2.31890      0.35535      6.52574   0.00000
  cses    /intcept                0.18810      0.19584      0.96048   0.33681
  cses    /cses                   0.06524      0.20764      0.31418   0.75338

------------------------------------------------------------------------------
  LEVEL 1                         TAU-HAT      STD.ERR.     Z-VALUE   PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept               36.72115      0.62593     58.66614   0.00000
```

All fixed effects are statistically significant. An increase in either cses or meanses is associated with a higher expected mean math score (holding all other variables constant). Catholic schools are expected to have a higher mean score too. There is also evidence of statistically significant interaction between the two SES measures and between individual SES and the type of school a student attends.

The four models fitted in this example are not strictly nested models. Here we opted to focus on specific questions. An alternative approach would have been to start with the fully unconditional model described in section 2 and then adding the SES indicators, SECTOR indicator and interaction terms of interest sequentially, using the deviance statistics to compare any pair of successive models for evidence of significant improvement in overall fit.