Multilevel model: comparison with conditional regression

Contents

1. Introduction

In this example, we consider an example where the categorical variable has several categories with unequal numbers of observations. Verbal skill is often considered to be the core feature of intelligence in studies of cognitive ability. The differences in this skill over individuals are also strongly associated with individual performance in many mental tasks. However, the correlation between verbal skill and performance may also depend on other personal traits.

The file **Math on Reading by Career.lsf** contains three variables. Data and syntax files can be found in the **MVABOOK\Chapter4** folder.

The variable Career is coded as follows:

- 1 = trades
- 2 = police or security
- 3 = business management
- 4 = sales
- 5 = military service
- 6 = teacher training
- 7 = industrial operations
- 8 = undecided
- 9 = real estate management

Math on Reading by Career.lsf

	Career	Math	Reading
1	1.00	48.71	15.24
2	1.00	43.49	6.33
3	1.00	44.08	15.00
4	1.00	47.50	23.00
5	1.00	63.88	34.67
6	1.00	45.62	15.43
7	1.00	43.77	12.69
8	1.00	49.49	13.20
9	1.00	42.89	13.94
10	1.00	49.69	8.91
11	1.00	42.23	9.19
12	1.00	56.15	17.19
13	2.00	57.87	29.27
14	2.00	47.02	14.00
15	2.00	38.26	7.45
16	2.00	47.26	10.28
17	2.00	48.75	15.63
18	2.00	46.05	25.82
19	2.00	37.50	3.46
20	2.00	38.08	4.41

2. Conditional regression

In another example, we fitted a simple linear regression to these data, using the mathematics score as outcome and the reading score as single predictor. The model can be expressed as

$$Math_{ij} = \alpha_i + \gamma_i Reading_{ij} + z_{ij}, \qquad i = 1, 2, \ldots 9, \; j = 1, 2, \ldots, n_i,$$

where $Math_{ij}$ and $Reading_{ij}$ are the scores for student *j* in career group *j* and z_{ij} is the error in regression. Results for the 9 career groups are shown below.

```
Math = 38.033 + 0.655*Reading + Error, R² = 0.591
Math = 36.149 + 0.707*Reading + Error, R² = 0.864
Math = 40.512 + 0.383*Reading + Error, R² = 0.376
Math = 36.085 + 0.736*Reading + Error, R² = 0.855
Math = 39.527 + 0.630*Reading + Error, R² = 0.581
Math = 34.430 + 0.827*Reading + Error, R² = 0.821
Math = 37.439 + 0.653*Reading + Error, R² = 0.752
Math = 41.436 + 0.676*Reading + Error, R² = 0.755
Math = 35.292 + 0.647*Reading + Error, R² = 0.935
```

3. Multilevel model

In this example, we use the same data but opt to fit a model that accommodates all career groups simultaneously. We use the 9 career groups as level-2 units, each with students nested within it. The syntax for this model is shown below.

mathread1.prl

```
OPTIONS OUTPUT=ALL ;
TITLE=Math on Reading by Career;
SY='Math on Reading by Career.lsf';
ID2=Career;
RESPONSE=Math;
FIXED=intcept Reading;
RANDOM1=intcept;
RANDOM2=intcept Reading;
```

We allow for random variation in both the intercept and the reading score over the career groups and use Career as the level-2 ID. Note that by using the OUTPUT = ALL option on the OPTIONS command we will also obtain the residuals under this model.

The first part of the output file shows the number of students nested within each career group:

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

 NUMBER OF LEVEL 2 UNITS :       9
 NUMBER OF LEVEL 1 UNITS :     147

 N2  :       1       2       3       4       5       6       7       8
 N1  :      12      23      19      24      22       9      12      17

 N2  :       9
 N1  :       9
 ======================================================================
```

After convergence, the following estimates are obtained:

```
                       +-----------------------+
                       | FIXED PART OF MODEL   |
                       +-----------------------+
------------------------------------------------------------------------------
  COEFFICIENTS              BETA-HAT       STD.ERR.      Z-VALUE       PR > |Z|
------------------------------------------------------------------------------
  intcept                   37.68531       0.76864       49.02825      0.00000
  Reading                    0.65545       0.04066       16.12127      0.00000

                       +-----------------------+
                       | RANDOM PART OF MODEL  |
                       +-----------------------+

------------------------------------------------------------------------------
  LEVEL 2                         TAU-HAT      STD.ERR.     Z-VALUE    PR > |Z|
------------------------------------------------------------------------------
  intcept /intcept                0.91092      2.32327      0.39208    0.69500
  Reading /intcept                0.00319      0.10399      0.03068    0.97552
  Reading /Reading                0.00494      0.00669      0.73802    0.46050
```

```
--------------------------------------------------------------------
  LEVEL 1                       TAU-HAT     STD.ERR.    Z-VALUE   PR > |Z|
--------------------------------------------------------------------
  intcept /intcept             13.55506     1.67523     8.09146   0.00000
```

From the results for the random part of the model, we note that while there is significant variation in the level-1 intercept, the same is not true at career level. The variation in intercept and reading scores between career groups is not statistically significant. This leads to the conclusion that the 9 regression lines are the same.

4. Bayes residuals

The file **mathread1.ba2** contains the Bayes estimates.

mathread1.BA2 - Notepad

File Edit Format View Help

```
           1         1     0.10707         0.61606      intcept
           1         2     0.72738E-02     0.26697E-02  Reading
           2         1    -0.35773         0.56775      intcept
           2         2     0.63561E-02     0.13663E-02  Reading
           3         1    -0.32494         0.62351      intcept
           3         2    -0.11737         0.16012E-02  Reading
           4         1    -0.41321         0.49411      intcept
           4         2     0.18879E-01     0.18271E-02  Reading
           5         1     0.52911         0.57515      intcept
           5         2     0.30752E-01     0.20819E-02  Reading
           6         1    -0.17362         0.66952      intcept
           6         2     0.19155E-01     0.24943E-02  Reading
           7         1    -0.69131E-01     0.68260      intcept
           7         2    -0.68889E-02     0.18396E-02  Reading
           8         1      1.2686         0.60906      intcept
           8         2     0.92458E-01     0.17909E-02  Reading
           9         1    -0.56617         0.69489      intcept
           9         2    -0.50620E-01     0.19943E-02  Reading
```

Ln 1, Col 1 | 100% | Windows (CRLF) | UTF-8

The first line of each pair of lines represents the estimate of the error term associated with the intercept and the associated standard error, the second the estimate of the error term and standard error associated with the slope of the predictor Reading. Note that the standard errors are larger than the estimates in most cases, again confirming our previous conclusion that these rand terms are not statistically significant.

In order to obtain estimates of the fixed intercept and reading slope for each career group, we can use these results and calculate each career group's estimated intercept and slope by adding the Bayes estimates of the error terms to the estimates of the fixed effects for intercept and reading slope, namely 37.68531 and 0.65545 respectively.

$$Group = 1:$$

$$37.68531 + 0.10707 = 37.79238$$

$$0.65545 + 0.00727 = 0.66272$$

$$\ldots$$

$$Group = 9:$$

$$37.68531 - 0.56617 = 37.11914$$

$$0.65545 - 0.05062 = 0.60483$$

The conditional regression results for these two groups were:

```
Math = 38.033 + 0.655*Reading + Error, R² = 0.591
Math = 35.292 + 0.647*Reading + Error, R² = 0.935
```

One would have expected the results to be more similar, given that the multilevel model analysis showed no significant variation in either intercept or reading slopes over the career groups. However, it should be kept in mind that the underlying assumptions of the two models considered here are quite different. Whereas conditional regression estimates the intercept and slope as fixed effects for each career group, assuming a random sample for each level-2 unit, multilevel modeling assumes that the level-2 units come from a population of level-2 units which implies that intercepts and slopes are random variables.