DIFFERENTIAL VARIABILITY OF TEST SCORES AMONG SCHOOLS: A MULTILEVEL ANALYSIS OF THE 5TH GRADE INVALSI TEST USING HETEROSCEDASTIC RANDOM EFFECTS

Leonardo Grilli                        Claudia Sani

Outline

- Objectives
- Description of sample
- Analysis
  - Preliminary analysis
  - Multilevel analysis
    - Results
    - Analysis of predicted random effects
- Conclusions and future developments
Objectives

Evaluation of the performance of the Italian school system through the learning levels of the pupils

Is the system fair?

Investigation of individual and contextual factors that affect the variability of the scores across schools

Methodological novelty: we use a multilevel model with heteroschedastic random effects

Dataset

We analyse data from INVALSI. The dataset is composed by:

- Results of the mathematics test administered to 5th grade pupils (about 11 years old) at the end of the 2008/2009 year (41 items) → summarized by the Rasch score
- Results of pupil’s questionnaire for measuring socio-economic factors
- Some demographic features of pupils (provided by school offices)

The sample includes about 1000 schools and 40000 pupils
Steps of the analysis

1. Preliminary analysis for describing pupil and school features, and how they are related to the Rasch score of the math test:
   - Box-plot
   - Anova
   - Scheffé test for multiple comparisons

2. Multilevel regression models

   Dependent variable: Rasch score of the math test

   Note: analysis carried out using the software Stata

Analysed variables

- **Demographic variables**
  - Gender
  - Foreigner vs Italian
  - Year of birth
- **Socio-cultural variables**
  - Availability of computer, encyclopaedia, internet
  - Number of books at home
  - Help with homework
  - Hours playing video games
  - Hours reading
- **Home environment**
  - With whom pupil lives
  - Language spoken at home
- **Wealth**
  - Presence of alarm at home
  - Number of bathrooms at home
  - Number of cars at home
- **School climate**
  - Unease score (scale 0 - 4)
- **Geographical area of schools**
### Distribution of Rasch math score

<table>
<thead>
<tr>
<th>Geographical Area</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Percentil 25%</th>
<th>Percentil 50%</th>
<th>Percentil 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NorthWest</td>
<td>-5.664</td>
<td>4.683</td>
<td>0.009</td>
<td>0.944</td>
<td>-0.626</td>
<td>-0.135</td>
<td>0.515</td>
</tr>
<tr>
<td>NorthEast</td>
<td>-2.322</td>
<td>1.322</td>
<td>0.93</td>
<td>0.94</td>
<td>-0.389</td>
<td>0.077</td>
<td>0.777</td>
</tr>
<tr>
<td>Center</td>
<td>-2.513</td>
<td>2.109</td>
<td>1.04</td>
<td>0.93</td>
<td>-0.626</td>
<td>0.026</td>
<td>0.626</td>
</tr>
<tr>
<td>South</td>
<td>-5.927</td>
<td>4.187</td>
<td>0.88</td>
<td>1.04</td>
<td>-0.688</td>
<td>-0.135</td>
<td>0.515</td>
</tr>
<tr>
<td>SouthIsles</td>
<td>-5.664</td>
<td>4.683</td>
<td>0.009</td>
<td>0.944</td>
<td>-0.626</td>
<td>-0.135</td>
<td>0.515</td>
</tr>
</tbody>
</table>

Note: Basilicata and Calabria belong to South-Isles area.

### Distribution of school mean Rasch math score

<table>
<thead>
<tr>
<th>Geographical Area</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NorthWest</td>
<td>0.30</td>
</tr>
<tr>
<td>NorthEast</td>
<td>0.34</td>
</tr>
<tr>
<td>Center</td>
<td>0.41</td>
</tr>
<tr>
<td>South</td>
<td>0.67</td>
</tr>
<tr>
<td>SouthIsles</td>
<td>0.54</td>
</tr>
</tbody>
</table>
Multilevel model

Two-level linear model with heteroschedastic random effects

\[ Y_{ij} = \gamma_{00} + \sum_{p=1}^{r} \beta_p X_{pij} + \sum_{l=1}^{s} \gamma_{0l} Z_{ij} + u_{0j} + e_{ij} \]

- level 1 variables
- level 2 variables
- level 2 errors
- level 1 errors

\[ e_{ij} \sim N(0; \sigma_{e}^2) \]
\[ u_{0j} \sim N(0; \sigma_{m}^2) \]

\[ i = 1, 2, \ldots, 38708 \rightarrow \text{pupils (level 1)} \]
\[ j = 1, 2, \ldots, 932 \rightarrow \text{schools (level 2)} \]
\[ m = 1, 2 \rightarrow \text{male, female} \]
\[ k = 1, 2, \ldots, 5 \rightarrow \text{North-West, North-East, Center, South, South-Isles} \]

Model selection

1. Empty models
   - with homoschedastic errors
   - with pupil-level errors depending on gender
   - with school-level errors depending on the geographical area
2. Models with pupil-level covariates (heteroschedastic errors)
3. Models with pupil-level and school-level covariates (heteroschedastic errors)
Significant pupil-level covariates

- **Demographic variables**
  - Gender
  - Foreigner vs Italian

- **Socio-cultural variables**
  - Availability of encyclopaedia
  - Hobby of reading
  - Help with homework
  - Number of books at home

- **Wealth**
  - Number of bathrooms at home

- **School climate**
  - Unease score (scale 0 - 4)

Note: statistical significance of coefficients has been assessed using the Wald test at 5% level

Significant school-level covariates

- **Geographical areas**
  - North-West
  - North-East
  - Center (reference cat.)
  - South
  - South-Isles

- **Contextual variables**
  - MS-Encyclopaedia
  - MS-Reading
  - MS-Bathrooms
  - MS-Unease

Obtained as school means of pupil variables

Note: statistical significance of coefficients has been assessed using the Wald test at 5% level
Coefficients of pupil-level covariates

Parameters estimated via Maximum Likelihood (command xtmixed)

Coefficients of school-level covariates

MS-Ency, MS-Read, MS-Bath: +10% in the school percentage
MS-Uneasiness: +0.1 in the school mean of the score
Standard deviations of model errors

<table>
<thead>
<tr>
<th>Std. Dev. pupil-level errors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0.886</td>
</tr>
<tr>
<td>Female</td>
<td>0.805</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Std. Dev. school-level errors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>North-West</td>
<td>0.197</td>
</tr>
<tr>
<td>North-East</td>
<td>0.237</td>
</tr>
<tr>
<td>Center</td>
<td>0.351</td>
</tr>
<tr>
<td>South</td>
<td>0.652</td>
</tr>
<tr>
<td>South-Isles</td>
<td>0.486</td>
</tr>
</tbody>
</table>

Parameters estimated via Maximum Likelihood (command xtmixed)

Intraclass Correlation Coefficient (ICC)

\[
\rho_{m,k} = \frac{\left(\sigma_{u_k}^{(k)}\right)^2}{\left(\sigma_{u_k}^{(k)}\right)^2 + \left(\sigma_{\epsilon}^{(m)}\right)^2} = \frac{\text{variance between schools}}{\text{total variance}}
\]

\[\rho_{m,k} \in [0,1]\]

<table>
<thead>
<tr>
<th>ICC</th>
<th>North-West</th>
<th>North-East</th>
<th>Center</th>
<th>South</th>
<th>South-Isles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>4.7%</td>
<td>6.7%</td>
<td>13.5%</td>
<td>35.1%</td>
<td>23.1%</td>
</tr>
<tr>
<td>Females</td>
<td>5.7%</td>
<td>8.0%</td>
<td>16.0%</td>
<td>39.6%</td>
<td>26.7%</td>
</tr>
</tbody>
</table>

In the South area about 40% of the residual variance depends on the school attended, compared to about 5% in the North area.
Predicted random effects

Predictions obtained with the Empirical Bayes method: the shrinkage property prevents schools with few pupils from having extreme values.

Comparison between models with and without outlying schools

<table>
<thead>
<tr>
<th></th>
<th>N schools</th>
<th>N positive outliers</th>
<th>N negative outliers</th>
<th>ICC Model with outliers</th>
<th>ICC Model without outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>North-West</td>
<td>216</td>
<td>0</td>
<td>0</td>
<td>4.7%</td>
<td>4.7%</td>
</tr>
<tr>
<td>North-East</td>
<td>179</td>
<td>5</td>
<td>2</td>
<td>6.7%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Center</td>
<td>186</td>
<td>4</td>
<td>0</td>
<td>13.5%</td>
<td>7.2%</td>
</tr>
<tr>
<td>South</td>
<td>175</td>
<td>13</td>
<td>1</td>
<td>35.1%</td>
<td>11.9%</td>
</tr>
<tr>
<td>South-Isles</td>
<td>176</td>
<td>8</td>
<td>0</td>
<td>23.1%</td>
<td>14.5%</td>
</tr>
</tbody>
</table>

Outlying schools in the South area explain a great part of residual variance.
Best and worst schools

- The ranking of the schools based on the mean math score is similar to the ranking based on predicted random effects (i.e. adjusted for the covariates).
- In particular, the positions of the top 20 schools are nearly unchanged.
- Both best and worst schools belong to South and South-Isles areas (a consequence of the high variability).

Differences in the expected score: underprivileged vs. privileged pupil; ineffective vs. effective school

<table>
<thead>
<tr>
<th>Mean Score</th>
<th>Observed variables</th>
<th>Not observed variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pupil underprivileged → privileged</td>
<td>School ineffective → effective</td>
</tr>
<tr>
<td>North-West</td>
<td>+0.707</td>
<td>+0.569</td>
</tr>
<tr>
<td>North-East</td>
<td>+0.707</td>
<td>+0.569</td>
</tr>
<tr>
<td>Center</td>
<td>+0.707</td>
<td>+0.569</td>
</tr>
<tr>
<td>South</td>
<td>+0.707</td>
<td>+0.569</td>
</tr>
<tr>
<td>South-Isles</td>
<td>+0.707</td>
<td>+0.569</td>
</tr>
</tbody>
</table>

Observed variables (model covariates)
- Underprivileged/ineffective → Negative pattern of covariates
- Privileged/effective → Positive pattern of covariates

Not observed variables (model errors)
- Underprivileged/ineffective → -2 Std. Dev.
- Privileged/effective → +2 Std. Dev.
Conclusions /1

- The math test score depends on several pupil-level factors (gender, to be foreigner, availability of encyclopaedias, hobby of reading, help with homework, number of books at home, number of bathrooms at home, uneasiness at school), as well as the geographical area of school and the contextual variables

- The South-Isles area has the lowest mean score (-0.195 versus 0.097 of the North-East area), whereas the South area has a mean score similar to Northern Italy but a huge between-school standard deviation (0.652 versus 0.197 of the North-West)

- In the South area about 40% of residual variance depends on the school, whereas in Northern Italy about 5% → The variance among schools increases when going from North to South → The goal of fairness is not attained in Southern Italy

Conclusions /2

- In the South area, a great part of variance depends on 14 outlying schools (8%): all of them – except one – have exceptionally positive results
  - the self-selection process seems to be asymmetric (‘positive’ selection into excellent schools more than ‘negative’ selection into worse schools)
  - the existence of positive outlying schools responsible for a substantial part of the variability makes the overall picture less problematic than what might appear at first sight
Future developments

- Analyse the Invalsi test score on Italian language and compare it with the math test (similar conclusions?)
- Analyse the test scores for several years (are the results stable over time?)
- Extend the model to more than two levels (for example, province as third level)
- Use models with random effects having an asymmetric distribution (to account for schools with exceptionally positive results)

References