HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

CASE: Multilevel Modelling in the PISA Survey

Risto Lehtonen
University of Helsinki

Lecture notes, 26 and 29 September 2011

Case: PISA 2000

- Lehtonen R. and Pahkinen E. Practical Methods for Design and Analysis of Complex Surveys. Second Edition. Chichester: John Wiley & Sons

  - **Section 9.4.**

  - **MULTILEVEL MODELLING IN EDUCATIONAL SURVEY**

PISA 2000

Programme for International Student Assessment

- Data collection in 2000
  - 32 countries
- Topics
  - **Reading literacy**
  - Mathematics
  - Science
- Countries selected
  - Brazil, Finland, Germany, Hungary, Republic of Korea, United Kingdom, and United States

- Hierarchical structure of data in each target country
  - Level 1: Student
  - Level 2: School
- Typical sampling design
  - Stratified two-stage cluster sampling
  - Clusters: Schools
  - School sampling
    - Systematic PPS (*Sampling with probabilities proportional to size*)

PISA – Analysis strategy

Model-based analysis

- Modelling the hierarchical structure of the data

- **Mixed models**
- **Multilevel models**

- **Complexities to be accounted for**
  - Weighting
  - Stratification
  - Clustering effect

- **Computation**
  - **SAS Procedures**
    - MIXED, GLIMMIX, NLMIXED
- **MLwiN** (Harvey Goldstein)

http://www.mlwin.com/

- **HML** Hierarchical Linear and Nonlinear Modeling

http://www.ssicentral.com/hlm/

- **Mplus** for multilevel models

http://www.statmodel.com/index.shtml

Why multilevel modelling?

- Hierarchical structure of the data
  - School level
  - Student level within schools
- Cluster sampling design
  - First stage:
    School sample
  - Second stage:
    Student sample out of the sampled schools
- Clustering by schools introduces intra-cluster correlation (ICC)
- Accounting for ICC by multilevel/mixed model

PISA 2000 – Weighting

- Element-level design weight
  - Inclusion probabilities
    - School level
    - Student level
  - Adjustment for nonresponse
  - Country-specific features
- Indexing
  - School *i*
  - Student *k*

- **Rescaling of design weights by country**
  - Analysis weights
  - Sum of weights = *n* (sample size by country)
  - Mean weight = 1
  - Details, see: OECD (2002b)

Weighting procedure (design weight)

Weight w_{ik} for student k in school i:

$$w_{ik} = w_{1i} \times w_{2ik} \times f_i, \quad i = 1,...,m \text{ and } k = 1,...,n_i,$$

where

$w_{1i} = 1/(\pi_i \hat{\theta}_i)$ is the reciprocal of the product of the inclusion probability π_i and the estimated participation probability $\hat{\theta}_i$ of school i;

$w_{2ik} = 1/(\pi_{k|i} \hat{\theta}_{k|i})$ is the reciprocal of the product of the conditional inclusion probability $\pi_{k|i}$ and estimated conditional response probability $\hat{\theta}_{k|i}$ of student k from within the selected school i;

f_i is an adjustment factor for school i to compensate any country-specific refinements in the survey design, and
m is the number of sample schools in a given country and n_i is the number of sample students in school i.

PISA 2000 – Study variable

- Study variable *y*
- *Student's combined reading literacy score*
- Combined variable using five variables measuring different aspects of reading skills
- Scaling
  - Mean over participating OECD countries = 500
  - S.D = 100
- Minimum = 402 (Brazil)
- Maximum = 550 (Finland)

PISA 2000 – Descriptive

Table 9.8 Descriptive statistics for combined reading literacy score in the PISA 2000 Survey by country (in alphabetical order).

Country	Combined reading literacy score: Mean	Standard error	Overall design effect	Design-effect accounting for stratification and clustering	Effective sample size of students	Number of observations in data set: Students	Schools
Brazil	402.9	3.82	8.33	5.17	476	3961	290
Finland	550.7	2.15	2.79	2.74	1600	4465	147
Germany	497.4	5.68	13.47	11.68	305	4108	183
Hungary	485.7	6.02	20.00	16.20	231	4613	184
Republic of Korea	526.6	3.66	12.99	11.67	351	4564	144
United Kingdom	531.4	4.08	14.08	7.16	564	7935	328
United States	517.0	5.16	6.93	5.46	354	2455	112

Data source: OECD PISA database, 2001.

PISA 2000 – Design effects

- *Overall design effect (1)*
  - Measures the effect of:
    - Stratification
    - Clustering
    - Weighting

    on variance estimate of the mean estimate
  - SRS variance estimate is for **unweighted** mean estimate

- *Deff accounting for stratification and clustering (2)*
  - Measures the effect of:
    - Stratification
    - Clustering

    on variance estimate of the mean estimate
  - SRS variance estimate is for **weighted** mean estimate

Design effect *Deff*

Design effect, deff (Kish 1965) measures the magnitude of the clustering effect to variance (standard error) estimate

Estimated overall deff (1):

$$deff(\bar{y}^*) = \frac{\hat{v}(\bar{y}^*)}{\hat{v}_{srs}(\bar{y})}$$

where

$\bar{y}^*$ is weighted mean estimate and $\bar{y}$ is the corresponding unweighted mean estimate

$\hat{v}(\bar{y}^*)$ is based on the actual sampling design

$\hat{v}_{srs}(\bar{y})$ is the SRS-based variance estimate

Deff (2):

$$deff(\bar{y}^*) = \frac{\hat{v}(\bar{y}^*)}{\hat{v}_{srs}(\bar{y}^*)}$$

PISA 2000 – Effective sample size

- *Effective sample size*
  The original student-level sample size divided by the design effect

- Effective sample size n_{eff} gives the SRS-based sample size that produces the same precision (measured by variance or standard error) as obtained for student-level sample size n under the actual cluster sampling design

- Example: Hungary

$$n_{eff} = \frac{n}{deff} = \frac{4613}{20.00} = 231$$

  - Strong intra-cluster correlation (large deff) introduces decreasing effective sample size!

PISA – Two-level hierarchical linear model

Fitting a Two-Level Hierarchical Linear Model

- *Study variable y*: Combined scaled reading literacy score
- Predictors:
  - School level
    - School size (SSIZE)
    - Teacher autonomy (AUTONOMY)

  - Standardization

  Mean (over countries) = 0

  Variance = 1

.

- Student level
  - FEMALE (1 is for females and 0 is for males)
  - Socio-economic background (SEB)
  - Engagement in reading (ENGAGEMENT)
  - Achievement press (ACHPRESS)

  - Standardization

  Mean (over countries) = 0

  Variance = 1

PISA – Linear two-level model

$$
\begin{aligned}
y_{ik} &= \text{INTERCEPT} + \gamma_1 \times \text{SSIZE}_i + \gamma_2 \times \text{AUTONOMY}_i \\
&\quad + \beta_1 \times \text{FEMALE}_{ik} + \beta_2 \times \text{SEB}_{ik} + \beta_3 \times \text{ENGAGEMENT}_{ik} \\
&\quad + \beta_4 \times \text{ACHPRESS}_{ik} + u_i + e_{ik}
\end{aligned}
$$

Index k : Level 1 elements (students)
Index i: Level 2 elements (schools)

Fixed effects γ and β:
Regression coefficients at school and student levels

Random effects:
u_i : School level *random intercept*
Normal distribution with mean 0 and variance σ_u^2

e_{ik} : Student-level random term (residual)
Normal distribution with mean 0 and variance σ_e^2

Random terms u_i and e_{ik} assumed independent
Student-level analysis weights w_{ik}

PISA – Intra-cluster correlation

Intra-cluster correlation

Skinner et al. (1989), Goldstein (2003), Snijders & Bosker (2002)

$$\hat{\rho}_{ICC} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}^2}$$

Overall variance estimate $\hat{\sigma}^2$ is decomposed into two parts:

Between-school variance $\hat{\sigma}_u^2$

Within-school variance $\hat{\sigma}_e^2$

ICC measures the pair-wise correlation of elements (students) belonging to the same cluster (school)

PISA – Linear two-level model

Baseline model (a) Table 9.9

$$y_{ik} = \text{INTERCEPT} + u_i + e_{ik}$$

Model including predictors (b) Table 10.10

$$\begin{aligned} y_{ik} = {} & \text{INTERCEPT} + \gamma_1 \times \text{SSIZE}_i + \gamma_2 \times \text{AUTONOMY}_i \\ & + \beta_1 \times \text{FEMALE}_{ik} + \beta_2 \times \text{SEB}_{ik} + \beta_3 \times \text{ENGAGEMENT}_{ik} \\ & + \beta_4 \times \text{ACHPRESS}_{ik} + u_i + e_{ik} \end{aligned}$$

Index k: Level 1 elements (students)
Index i: Level 2 elements (schools)

PISA – ICC for model (a)

Example for Hungary

(a) Baseline model (*multilevel model with only intercept and residuals at both levels*), estimated ICC
(Hungary in Table 9.9)

$$\hat{\rho}_{ICC} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}^2} = \frac{6093.7}{6093.7+3148.3} = 0.659$$

PISA – ICC for model (b)

(b) Model including predictors
Residual intra-school correlation coefficient
(Hungary in Table 9.10)

$$\hat{\rho}_{ICC} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}^2} = \frac{4744.2}{4744.2+2897.4} = 0.621$$

PISA – Null (baseline) model

- Model (a) including fixed intercept and random intercepts

- Table 9.9 Estimates of two-level variance component models (null models) for combined reading literacy score in the PISA 2000 Survey by country (ordered by the size of the estimated intra-school correlation coefficient).

Table 9.9 Estimates of two-level variance component models (null models) for combined reading literacy score in the PISA 2000 Survey by country (ordered by the size of the estimated intra-school correlation coefficient).

Country	Intra-school correlation coefficient	Variance components: School level	Variance components: Student level	Intercept	Standard error
Hungary	0.659	6093.7	3148.3	464.1	5.84
Germany	0.553	5572.2	4507.8	496.1	5.61
Brazil	0.428	3146.9	4201.4	387.9	3.61
Republic of Korea	0.375	1828.6	3043.0	520.9	3.74
United States	0.241	2318.2	7315.5	503.3	4.97
United Kingdom	0.212	1917.5	7126.5	529.0	2.88
Finland	0.063	470.7	6960.9	550.6	2.18

Data source: OECD PISA database, 2001.

PISA – More advanced model

- Model (b) including predictors
- Table 9.10 Estimates of two-level models for combined reading literacy score in the PISA 2000 Survey by country

	Hungary	Germany	Brazil	Republic of Korea	United States	United Kingdom	Finland
Random effects: Variance component							
School level	4744.2	3501.6	2730.5	1387.3	1770.6	999.6	394.8
Student level	2897.4	3981.9	3830.6	2809.6	6094.1	5779.0	4984.3
Residual intra-school correlation coefficient	0.621	0.468	0.416	0.331	0.225	0.147	0.073
Proportional reduction in variance components, compared to null model (%)							
School level	22.1	37.2	13.2	24.1	23.6	47.9	16.1
Student level	8.0	11.7	8.8	7.7	16.7	18.9	28.4
Total	17.3	25.8	10.7	13.8	18.4	25.0	27.6

PISA – Some conclusions

- Multilevel modelling offers a powerful tool for analysis when there is a hierarchical structure in the data set
- Multilevel modelling provides explicit information about group (cluster) differences
  - More information is obtained for the interpretation of the results
- By multilevel modelling it is possible to account for the complexities of the research design
  - Stratification, clustering, weighting
- Additional levels can be introduced
  - Time…

Literature

- Chambers R.L. and Skinner C.J. (Eds.) (2004). *Analysis of Survey Data*. Chichester: Wiley.
- Demidenko E. (2004). *Mixed Models. Theory and Applications*. New York: Wiley.
- Diggle, P. J., Liang, K.-Y. & Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford: Oxford University Press.
- Goldstein, H. (2003). *Multilevel Statistical Models. 3rd Edition*. London: Edward Arnold. http://www.cmm.bristol.ac.uk/MLwiN/index.shtml
- Lehtonen R. and Pahkinen E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*. Second Edition. Chichester: Wiley.
- OECD (2002a). PISA 2000 Technical Report. Paris: OECD. http://www.pisa.oecd.org/
- Snijders, T. and Bosker, R. (2002). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling.* London: Sage Publications.