10

Analyses with School-Level Variables

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

INTRODUCTION

The target population in PISA is 15-year-old students. This population was chosen because, at this age in most OECD countries, students are approaching the end of their compulsory schooling. Thus, PISA should be able to indicate the cumulative effect of a student's education.

As described in detail in Chapter 3, a two-stage sampling procedure is used in PISA. After the population is defined, school samples are selected with a probability proportional to size. Subsequently, 35 students are randomly selected from each school. As the target population is based on age, it is possible that students will come from a variety of grades.

Table 10.1 presents the distribution of 15-year-olds in each country, by grade and by the International Standard Classification of Education (ISCED) level in PISA 2006.

In a few countries, most of the 15-year-old population tends to be in a modal grade, whereas in others, the 15-year-old population is spread across several grades.

Table 10.1

Percentage of students per grade and ISCED level, by country (PISA 2006)

	Grade								
	7	8	9	10	11	12	13	ISCED 2	ISCED 3
AUS		0.1	9.2	70.8	19.8	0.1		80.1	19.9
AUT	0.3	6.4	44.6	48.7	0			6.7	93.3
BEL	0.4	4.4	31.1	63.2	1			6.8	93.2
CAN	0	1.7	13.3	83.8	1.2	0		15	85
CHE	0.8	16.1	62.6	20.3	0.3	0		82.8	17.2
CZE	0.6	3.5	44.3	51.5				50.4	49.6
DEU	1.6	12.3	56.5	29.3	0.3			97.3	2.7
DNK	0.2	12	85.3	1.4	1.1			98.9	1.1
ESP	0.1	7	33	59.8	0			100	0
FIN	0.2	11.7	88.1	0				100	0
FRA	0	5.2	34.8	57.5	2.4	0		40	60
GBR			0	0.9	98.4	0.7		0.5	99.5
GRC	0.5	2.1	5.3	78.8	13.3			8	92
HUN	2.2	5.5	65.7	26.6	0			7.7	92.3
IRL	0	2.7	58.5	21.2	17.5			61.3	38.7
ISL			0.2	99.2	0.6			99.4	0.6
ITA	0.3	1.5	15	80.4	2.8			1.7	98.3
JPN				100					100
KOR			2	97.3	0.7			2	98
LUX	0.2	11.8	53.4	34.4	0.1			64.2	35.8
MEX	2.3	8.1	33.5	48.9	5.1	2		44.5	55.5
NLD	0.1	3.7	44.9	50.7	0.4	0		73.5	26.5
NOR			0.5	99	0.5			99.5	0.5
NZL			0	6.2	89.4	4.4	0	6.2	93.8
POL	0.6	3.8	95	0.6				99.4	0.6
PRT	6.6	13.1	29.5	50.7	0.2			50.2	49.8
SVK	0.7	2.2	38.5	58.7				38.6	61.4
SWE		1.9	95.9	2.2				97.8	2.2
TUR	0.8	4.5	38.4	53.7	2.6			5.3	94.7
USA	0.8	1	10.7	70.9	16.5	0.1		12.4	87.6

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

The PISA target population can be spread over several grades for different reasons:

- If the student does not pass a particular grade examination, he or she has to repeat the grade. For example in some countries there may be up to about 35% of students who have already repeated at least one grade.
- Even if grade retention is not used, the 15-year-old population might be separated at the time of testing into two grades. For logistical reasons, PISA testing takes place in a single calendar year. As the recommended testing window is around April (in the northern hemisphere), the PISA target population is defined as all students between 15 years and 3 months old and 16 years and 2 months old at the beginning of the testing period. If the entrance rules for compulsory education are defined in terms of complete calendar years, then the PISA target population will attend just one grade.

Further, in most countries, the PISA population is spread across the two levels of secondary education, *i.e.* lower secondary (ISCED 2) and upper secondary (ISCED 3).

As the 15-year-old population attends different grades in most OECD countries, the within-school samples can only consist of a random sample of students. Consequently, the PISA participating students are attending several grades and within a particular grade, are in several classes, depending on the school size. Largely because the PISA sample is not class-based, PISA collects data at the school level. This chapter describes how and why school-level data should be analysed.

Since the PISA target population attends several grades in most countries, it would be interesting to compute the average performance growth between two adjacent grades, so that performance differences between countries could be translated into school-year effect. However, this would certainly lead to an overestimation of the performance growth: 15-year-olds attending lower grades are either lower achievers or younger students, and 15-year-olds attending higher grades are either high achievers or older students. Therefore, caution is required in comparing subpopulations in different grades.

LIMITS OF THE PISA SCHOOL SAMPLES

As mentioned earlier, the following statement is valid for both PISA and other international studies:

> Although the student's samples were drawn from within a sample of schools, the school sample was designed to optimize the resulting sample of students, rather than to give an optimal sample of schools. For this reason, it is always preferable to analyse the school-level variables as attributes of students, rather than as elements in their own right. (Gonzalez and Kennedy, 2003)

This advice is particularly important in PISA as the target population is not defined as a grade, but as students of a particular age.

In some countries, lower secondary and upper secondary education are provided by the same school, whereas in others, this is not the case. In these countries, usually, the transition between lower and upper secondary education occurs around the age of 15, *i.e.* at the end of compulsory education (in most cases). As PISA focuses on the 15-year-old population, it means that one part of the target population is attending upper secondary education, while the other is attending lower secondary education (see Table 10.1 for the percentages of student by level of secondary education). Consequently, in several countries, 15-year-olds can be in different educational institutions.

As discussed in Chapter 3, schools are selected from the school sample frame by the probability proportional to size (PPS) sampling method, *i.e.* proportionally to the number of 15-year-olds attending the school.

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

This might mean, for example, that upper secondary schools only attended by students over the PISA age of 15 should not be included in the school sample frame. Similarly, lower secondary schools without any 15-year-olds should not be included in the school sample frame.

Thus, neither the lower secondary school population, nor the upper secondary school population represents the 15-year-old school population. In other words, the PISA school target population does not necessarily match the school population(s) within a particular country.

This lack of perfect match between the usual school population(s) and the PISA school population affects the way school data should be analysed. To avoid biases for the population estimates, school data have to be imported into the student data files and have to be analysed with the student final weight. This means, for example, that one will not estimate the percentage of public schools versus private schools, but will estimate the percentage of 15-year-olds attending private schools versus public schools. From a pedagogical and/or policy point of view, what is really important is not the percentage of schools that present such characteristics, but the percentage of students who are affected by these characteristics.

MERGING THE SCHOOL AND STUDENT DATA FILES

Box 10.1 provides the SAS® syntax for merging the student and school data files. Both files need first to be sorted by the identification variables, *i.e.* CNT, SCHOOLID and STIDSTD in the student data file and CNT and SCHOOLID in the school data file. Afterwards, the two sorted data files can be merged according to the common identification variables, *i.e.* CNT and SCHOOLID.

Box 10.1 **SAS® syntax for merging the student and school data files (*e.g.* PISA 2006)**

```
libname PISA2006 "c:\pisa\2006\data\";
options nofmterr notes;

data temp1;
        set pisa2006.stu;
run;
proc sort data=temp1;
        by cnt schoolid stidstd;
run;
data temp2;
        set pisa2006.sch;
run;
proc sort data = temp2;
        by cnt schoolid;
run;
data pisa2006.alldata;
        merge temp1 temp2;
        by cnt schoolid;
run;
```

ANALYSES OF THE SCHOOL VARIABLES

After merging the student and school data files, school data can be analysed as any student-level variables since the school-level variables are now considered as attributes of students. However, in this case, it is even more critical to use the replicate weights to compute sampling errors. Failure to do so would give a completely misleading inference.

The remainder of this chapter explains the methods for computing the percentage of students by school location, and their respective standard errors, as well as the student average performance in science, by school location.

Box 10.2 sets out the school-location question found in the school questionnaire.

Box 10.2 **Question on school location in PISA 2006**

Q7 WHICH OF THE FOLLOWING BEST DESCRIBES THE COMMUNITY IN WHICH YOUR SCHOOL IS LOCATED?

Tick only one box in each row.

A <village, hamlet or rural area> (fewer than 3 000 people)	☐ 1
A <small town> (3 000 to about 15 000 people)	☐ 2
A <town> (15 000 to about 100 000 people)	☐ 3
A <city> (100 000 to about 1 000 000 people)	☐ 4
A large <city> with over 1 000 000 people	☐ 5

Box 10.3 provides the SAS® syntax for computing the percentage of students and the average performance in science, by school location. As previously indicated, the SAS® macro might be CPU-consuming; thus it is advised to keep only the variables indispensable for the analyses.

Box 10.3 **SAS® syntax for computing the percentage of students and the average performance in science, by school location (*e.g.* PISA 2006)**

```
data temp3;
   set pisa2006.alldata;
   if (cnt="DEU");
   if (not missing (sc07q01));
   w_fstr0=w_fstuwt;
   scie1=pv1scie;
   scie2=pv2scie;
   scie3=pv3scie;
   scie4=pv4scie;
   scie5=pv5scie;
   keep   cnt schoolid stidstd
          w_fstr0-w_fstr80 scie1-scie5 sc07q01;
run;
%include "c:\pisa\macro\proc_freq_no_pv.sas";
%BRR_FREQ(   INFILE=temp3,
             REPLI_ROOT=w_fstr,
             BYVAR=cnt,
             VAR=sc07q01,
             LIMIT=no,
             LIMIT_CRITERIA=,
             ID_SCHOOL=,
             OUTFILE=exercise1);
run;
%include "c:\pisa\macro\proc_means_pv.sas";
%BRR_PROCMEAN_PV(INFILE=temp3,
                 REPLI_ROOT=w_fstr,
                 BYVAR=cnt sc07Q01,
                 PV_ROOT=scie,
                 STAT=mean,
                 LIMIT=no,
                 LIMIT_CRITERIA=,
                 ID_SCHOOL=schoolid,
                 OUTFILE=exercise2);
run;
```

Table 10.2 and Table 10.3 present the structure of the output data files exercise1 and exercise2.

Table 10.2
Output data file exercise1 from Box 10.3

CNT	SC07Q01	STAT	SESTAT
DEU	1	3.8	1.1
DEU	2	27.6	2.9
DEU	3	45.0	3.5
DEU	4	15.7	2.6
DEU	5	7.9	1.6

Table 10.3
Output data file exercise2 from Box 10.3

CNT	SC07Q01	STAT	SESTAT
DEU	1	453.1	14.9
DEU	2	515.9	7.7
DEU	3	526.4	7.7
DEU	4	521.1	13.6
DEU	5	487.5	17.8

As a reminder, the school data was analysed at the student level and weighted by the student final weight. Therefore, results should be interpreted as: 27.6% of the 15-year-olds are attending a school located in a village with less than 3 000 people. The students attending a school located in a small village on average perform at 515.9.

As the percentages for some categories might be small, the standard error will be large for the mean estimates.

All the SAS® macros described in the previous chapters can be used on the school variables once they have been imported in the student data file.

CONCLUSION

For statistical and pedagogical reasons, the data collected through the school questionnaire, as well as the variables derived from that instrument, have to be analysed at the student level.

All the SAS® macros developed can be used without any modifications. The interpretation of the results should clearly state the analysis level, *i.e.* for instance the percentage of students attending a school located in a small village and not the percentage of schools located in a small village.

References

Beaton, A.E. (1987), *The NAEP 1983-1984 Technical Report*, Educational Testing Service, Princeton.

Beaton, A.E., *et al.* (1996), *Mathematics Achievement in the Middle School Years, IEA's Third International Mathematics and Science Study*, Boston College, Chestnut Hill, MA.

Bloom, B.S. (1979), *Caractéristiques individuelles et apprentissage scolaire*, Éditions Labor, Brussels.

Bressoux, P. (2008), *Modélisation statistique appliquée aux sciences sociales*, De Boek, Brussels.

Bryk, A.S. and **S.W. Raudenbush** (1992), *Hierarchical Linear Models for Social and Behavioural Research: Applications and Data Analysis Methods*, Sage Publications, Newbury Park, CA.

Buchmann, C. (2000), *Family structure, parental perceptions and child labor in Kenya: What factors determine who is enrolled in school?* a*Soc. Forces*, No. 78, pp. 1349-79.

Cochran, W.G. (1977), *Sampling Techniques*, J. Wiley and Sons, Inc., New York.

Dunn, O.J. (1961), "Multilple Comparisons among Menas", *Journal of the American Statistical Association*, Vol. 56, American Statistical Association, Alexandria, pp. 52-64.

Kish, L. (1995), *Survey Sampling*, J. Wiley and Sons, Inc., New York.

Knighton, T. and **P. Bussière** (2006), "Educational Outcomes at Age 19 Associated with Reading Ability at Age 15", Statistics Canada, Ottawa.

Gonzalez, E. and **A. Kennedy** (2003), *PIRLS 2001 User Guide for the International Database*, Boston College, Chestnut Hill, MA.

Ganzeboom, H.B.G., P.M. De Graaf and **D.J. Treiman** (1992), "A Standard International Socio-economic Index of Occupation Status", *Social Science Research* 21(1), Elsevier Ltd, pp 1-56.

Goldstein, H. (1995), *Multilevel Statistical Models*, 2nd Edition, Edward Arnold, London.

Goldstein, H. (1997), "Methods in School Effectiveness Research", *School Effectiveness and School Improvement* 8, Swets and Zeitlinger, Lisse, Netherlands, pp. 369-395.

Hubin, J.P. (ed.) (2007), *Les indicateurs de l'enseignement*, 2nd Edition, Ministère de la Communauté française, Brussels.

Husen, T. (1967), *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, Almqvist and Wiksells, Uppsala.

International Labour Organisation (ILO) (1990), *International Standard Classification of Occupations: ISCO-88.* Geneva: International Labour Office.

Lafontaine, D. and **C. Monseur** (forthcoming), "Impact of Test Characteristics on Gender Equity Indicators in the Assessment of Reading Comprehension", *European Educational Research Journal*, Special Issue on PISA and Gender.

Lietz, P. (2006), "A Meta-Analysis of Gender Differences in Reading Achievement at the Secondary Level", *Studies in Educational Evaluation* 32, pp. 317-344.

Monseur, C. and **M. Crahay** (forthcoming), "Composition académique et sociale des établissements, efficacité et inégalités scolaires : une comparaison internationale – Analyse secondaire des données PISA 2006", *Revue française de pédagogie.*

OECD (1998), *Education at a Glance – OECD Indicators*, OECD, Paris.

OECD (1999a), *Measuring Student Knowledge and Skills – A New Framework for Assessment*, OECD, Paris.

OECD (1999b), *Classifying Educational Programmes – Manual for ISCED-97 Implementation in OECD Countries*, OECD, Paris.

OECD (2001), *Knowledge and Skills for Life – First Results from PISA 2000*, OECD, Paris.

OECD (2002a), *Programme for International Student Assessment – Manual for the PISA 2000 Database*, OECD, Paris.

OECD (2002b), *Sample Tasks from the PISA 2000 Assessment – Reading, Mathematical and Scientific Literacy,* OECD, Paris.

OECD (2002c), *Programme for International Student Assessment – PISA 2000 Technical Report,* OECD, Paris.

OECD (2002d), *Reading for Change: Performance and Engagement across Countries – Results from PISA 2000,* OECD, Paris.

OECD (2003a), *Literacy Skills for the World of Tomorrow – Further Results from PISA 2000,* OECD, Paris.

OECD (2003b), *The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills,* OECD, Paris.

OECD (2004a), *Learning for Tomorrow's World – First Results from PISA 2003,* OECD, Paris.

OECD (2004b), *Problem Solving for Tomorrow's World – First Measures of Cross-Curricular Competencies from PISA 2003,* OECD, Paris.

OECD (2005a), *PISA 2003 Technical Report,* OECD, Paris.

OECD (2005b), *PISA 2003 Data Analysis Manual,* OECD, Paris.

OECD (2006), *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006,* OECD, Paris.

OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World,* OECD, Paris.

OECD (2009), *PISA 2006 Technical Report,* OECD, Paris.

Peaker, G.F. (1975), *An Empirical Study of Education in Twenty-One Countries: A Technical report. International Studies in Evaluation VIII,* Wiley, New York and Almqvist and Wiksell, Stockholm.

Rust, K.F. and **J.N.K. Rao** (1996), "Variance Estimation for Complex Surveys Using Replication Techniques", *Statistical Methods in Medical Research,* Vol. 5, Hodder Arnold, London, pp. 283-310.

Rutter, M., *et al.* (2004), "Gender Differences in Reading Difficulties: Findings from Four Epidemiology Studies", *Journal of the American Medical Association* 291, pp. 2007-2012.

Schulz, W. (2006), *Measuring the socio-economic background of students and its effect on achievement in PISA 2000 and PISA 2003,* Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.

Wagemaker, H. (1996), *Are Girls Better Readers. Gender Differences in Reading Literacy in 32 Countries,* IEA, The Hague.

Warm, T.A. (1989), "Weighted Likelihood Estimation of Ability in Item Response Theory", *Psychometrika,* Vol. 54(3), Psychometric Society, Williamsburg, VA., pp. 427-450.

Wright, B.D. and **M.H. Stone** (1979), *Best Test Design: Rasch Measurement,* MESA Press, Chicago.

Table of contents

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

LIST OF BOXES

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

LIST OF FIGURES

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

LIST OF TABLES

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009

User's Guide

Preparation of data files

All data files (in text format) and the SAS® control files are available on the PISA website (*www.pisa.oecd.org*).

SAS® users

By running the SAS® control files, the PISA data files are created in the SAS® format. Before starting analysis, assigning the folder in which the data files are saved as a SAS® library.

For example, if the PISA 2000 data files are saved in the folder of "c:\pisa2000\data\", the PISA 2003 data files are in "c:\pisa2003\data\", and the PISA 2006 data files are in "c:\pisa2006\data\", the following commands need to be run to create SAS® libraries:

```
libname PISA2000 "c:\pisa2000\data\";
libname PISA2003 "c:\pisa2003\data\";
libname PISA2006 "c:\pisa2006\data\";
run;
```

SAS® syntax and macros

All syntaxes and macros in this manual can be copied from the PISA website (*www.pisa.oecd.org*). The 17 SAS® macros presented in Chapter 17 need to be saved under "c:\pisa\macro\", before staring analysis. Each chapter of the manual contains a complete set of syntaxes, which must be done sequentially, for all of them to run correctly, within the chapter.

Rounding of figures

In the tables and formulas, figures were rounded to a convenient number of decimal places, although calculations were always made with the full number of decimal places.

Country abbreviations used in this manual

AUS	Australia	FRA	France	MEX	Mexico
AUT	Austria	GBR	United Kingdom	NLD	Netherlands
BEL	Belgium	GRC	Greece	NOR	Norway
CAN	Canada	HUN	Hungary	NZL	New Zealand
CHE	Switzerland	IRL	Ireland	POL	Poland
CZE	Czech Republic	ISL	Iceland	PRT	Portugal
DEU	Germany	ITA	Italy	SVK	Slovak Republic
DNK	Denmark	JPN	Japan	SWE	Sweden
ESP	Spain	KOR	Korea	TUR	Turkey
FIN	Finland	LUX	Luxembourg	USA	United States

PISA DATA ANALYSIS MANUAL: SAS® SECOND EDITION – ISBN 978-92-64-05624-4 – © OECD 2009