PISA 2003
Data Analysis Manual
SAS® Users
The OECD’s Programme for International Student Assessment (PISA) surveys, which take place every three years, have been designed to collect information about 15-year-old students in participating countries. PISA examines how well students are prepared to meet the challenges of the future, rather than how well they master particular curricula. The data collected during each PISA cycle are an extremely valuable source of information for researchers, policy makers, educators, parents and students. It is now recognised that the future economic and social well-being of countries is closely linked to the knowledge and skills of their populations. The internationally comparable information provided by PISA allows countries to assess how well their 15-year-old students are prepared for life in a larger context and to compare their relative strengths and weaknesses.

The PISA 2003 database, on which this manual is focused, contains information on over a quarter of a million students from 41 countries. It includes not only information on their performance in the four main areas of assessment – reading, mathematics, science and problem solving – but also their responses to the Student Questionnaire that they complete as part of the assessment. Data from the school principals are also included.

The PISA 2003 Data Analysis Manual has evolved from the analytical workshops held in Sydney, Vienna, Paris and Bratislava, which exposed participants to the various techniques needed to correctly analyse the complex databases. It allows analysts to confidently replicate procedures used for the production of the PISA 2003 initial reports, Learning for Tomorrow’s World – First Results from PISA 2003 (OECD, 2004a) and Problem Solving for Tomorrow’s World – First Measures of Cross-Curricular Competencies from PISA 2003 (OECD, 2004b), and to accurately undertake new analyses in areas of special interest. In addition to the inclusion of the necessary techniques, the manual also includes a detailed account of the variables constructed from the student and school questionnaires. This information was previously published in the Manual for the PISA 2000 Database (OECD, 2002a).

The PISA 2003 Data Analysis Manual is in four parts – the first two sections give a detailed theoretical background and instructions for analysing the data; the third section lists the program codes (syntaxes and the macros), which are needed to carry out the analyses; and the fourth section contains a detailed description of the database.

PISA is a collaborative effort by the participating countries, and guided by their governments on the basis of shared policy-driven interests. Representatives of each country form the PISA Governing Board which decides on the assessment and reporting of results in PISA.

There are two versions of this manual – one for SPSS® users and one for SAS® users. The OECD recognises the creative work of Christian Monseur in preparing the text for both versions of the manual in collaboration with Sheila Krawchuk and Keith Rust, as well as his preparation of the program coding for the SAS® users’ manual. The coding for the SPSS® users’ manual was prepared by Wolfram Schulz and Eveline Gebhardt. The main editorial work was completed at the OECD Secretariat by Miyako Ikeda, Sophie Vayssettes, John Cresswell, Claire Shewbridge and Kate Lancaster. The PISA assessments and the data underlying the manuals were prepared by the PISA Consortium under the direction of Raymond Adams.
# Table of Contents

**Users’ Guide** .............................................................................................................................................. 9

**CHAPTER 1**

*The OECD’s Programme for International Student Assessment* ........................................................................... 11

- An overview of PISA ....................................................................................................................................... 12
- What makes PISA unique? ............................................................................................................................ 13
- How the assessment takes place .................................................................................................................. 15
- About this manual .......................................................................................................................................... 16

**CHAPTER 2**

*Sample Weights* ........................................................................................................................................... 19

- Introduction .................................................................................................................................................. 20
- Weights for simple random samples .......................................................................................................... 21
- Sampling designs for education surveys .................................................................................................... 22
- Why do the PISA weights vary? .................................................................................................................. 27
- Conclusions .................................................................................................................................................. 30

**CHAPTER 3**

*Replicate Weights* ....................................................................................................................................... 31

- Introduction .................................................................................................................................................. 32
- Sampling variance for simple random sampling ........................................................................................ 32
- Sampling variance for two-stage sampling ................................................................................................. 38
- Replication methods for simple random samples ..................................................................................... 44
- Resampling methods for two-stage samples .............................................................................................. 46
- The Jackknife for unstratified two-stage sample designs ........................................................................ 47
- The Jackknife for stratified two-stage sample designs ............................................................................ 48
- The Balanced Repeated Replication method ............................................................................................. 49
- Other procedures for accounting for clustered samples ........................................................................... 51
- Conclusions .................................................................................................................................................. 51

**CHAPTER 4**

*The Rasch Model* ......................................................................................................................................... 53

- Introduction .................................................................................................................................................. 54
- How can the information be summarised? .................................................................................................. 54
- The Rasch model for dichotomous items .................................................................................................... 56
- Other Item Response Theory models ......................................................................................................... 69
- Conclusions .................................................................................................................................................. 69
CHAPTER 5
Plausible Values.................................................................................................................71
Individual estimates versus population estimates.................................................................72
The meaning of plausible values..............................................................................................72
Comparison of the efficiency of Warm Likelihood Estimates, Expected A Posteriori
estimates and plausible values for the estimation of some population statistics..........76
How to perform analyses with plausible values.................................................................78
Conclusions.............................................................................................................................80

CHAPTER 6
Computation of Standard Errors............................................................................................81
Introduction .............................................................................................................................82
The standard error on univariate statistics for numerical variables......................................82
The SAS® macro for computing the standard error on a mean .............................................85
The standard error on percentages........................................................................................87
The standard error on regression coefficients.......................................................................90
The standard error on correlation coefficients.....................................................................92
Conclusions.............................................................................................................................93

CHAPTER 7
Analyses with Plausible Values...............................................................................................95
Introduction .............................................................................................................................96
Univariate statistics on plausible values................................................................................96
The standard error on percentages with plausible values......................................................101
The standard error on regression coefficients with plausible values..................................101
The standard error on correlation coefficients with plausible values..................................103
Correlation between two sets of plausible values.................................................................104
A fatal error shortcut ............................................................................................................109
An unbiased shortcut............................................................................................................110
Conclusions.............................................................................................................................111

CHAPTER 8
Use of Proficiency Levels.....................................................................................................113
Introduction .............................................................................................................................114
Generation of the proficiency levels......................................................................................114
Other analyses with proficiency levels................................................................................120
Conclusions.............................................................................................................................124

CHAPTER 9
Analyses with School Level Variables ..................................................................................125
Introduction .............................................................................................................................126
Limits of the PISA school samples.........................................................................................127
Merging the school and student data files ............................................................................128
Analyses of the school variables .........................................................................................128
Conclusions.............................................................................................................................130
CHAPTER 10

Standard Error on a Difference ................................................................. 131
  Introduction ......................................................................................... 132
  The standard error of a difference without plausible values .................. 134
  The standard error of a difference with plausible values ...................... 139
  Multiple comparisons ........................................................................ 141
  Conclusions ...................................................................................... 143

CHAPTER 11

OECD Average and OECD Total .............................................................. 145
  Introduction ........................................................................................ 146
  Recoding of the database for the estimation of the OECD total and OECD average ........................................................................... 146
  Duplication of the data for avoiding three runs of the procedure .......... 149
  Comparisons between OECD average or OECD total estimates and a country estimate ............................................................. 149
  Conclusions ........................................................................................ 152

CHAPTER 12

Trends ..................................................................................................... 153
  Introduction ........................................................................................ 154
  The computation of the standard error for trend indicators on variables other than performance ......................................................... 155
  The computation of the standard error for trend indicators on performance variables ................................................................. 158
  Conclusions ........................................................................................ 164

CHAPTER 13

Multilevel Analyses ................................................................................. 165
  Introduction ........................................................................................ 166
  Simple linear regression ...................................................................... 166
  Simple linear versus multilevel regression analyses ............................ 170
  Fixed effect versus random effect ........................................................ 173
  Some examples with SAS® ................................................................... 174
  Limitations of the multilevel model in the PISA context ...................... 190
  Conclusions ........................................................................................ 192

CHAPTER 14

Other Statistical Issues ............................................................................ 193
  Introduction ........................................................................................ 194
  Analyses by quarters ........................................................................... 194
  The concepts of relative risk and attributable risk .............................. 199
  Instability of the relative and attributable risks .................................... 201
  Computation of the relative risk and attributable risk ......................... 202
  Conclusions ........................................................................................ 202

CHAPTER 15

SAS® Macros .......................................................................................... 203
  Introduction ........................................................................................ 204
  Structure of the SAS® macros ............................................................. 204
# Table of Contents

- **Appendix 1**: PISA 2003 International Database ................................................................. 231
- **Appendix 2**: Student Questionnaire .................................................................................... 245
- **Appendix 3**: Educational Career Questionnaire ............................................................. 259
- **Appendix 4**: Information Communication Technology (ICT) Questionnaire .................. 261
- **Appendix 5**: School Questionnaire ..................................................................................... 265
- **Appendix 6**: Student Questionnaire Data File Codebook ............................................... 279
- **Appendix 7**: School Questionnaire Data File Codebook ................................................. 325
- **Appendix 8**: Student Cognitive Test Data File Codebook ............................................... 339
- **Appendix 9**: Student and School Questionnaire Indices .................................................. 369
- **Appendix 10**: Scores Allocated to the Items ..................................................................... 395

*References.................................................................................................................................................. 401*
**USERS’ GUIDE**

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

**SAS® users**
By running the SAS® control files, the PISA 2003 SAS® student data file and the PISA 2003 SAS® school data file are created. Please keep the both files in the same folder and run commands for assigning the folder as a SAS® library before starting analysis.

For example, if the student and school SAS® data files are saved in the folder of “c:\pisa2003\data\”, the following commands need to be run to create a SAS® library:

```sas
libname PISA2003 “c:\pisa2003\data\”; run;
```

The ten SAS® macros presented in Chapter 15 need to be saved under “c:\pisa2003\prg”.

**SAS® syntax and macros**
All syntaxes and macros used in this manual can be copied from the PISA Web site (www.pisa.oecd.org). 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**

<table>
<thead>
<tr>
<th>Country</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>AUT</td>
</tr>
<tr>
<td>Belgium</td>
<td>BEL</td>
</tr>
<tr>
<td>Brazil</td>
<td>BRA</td>
</tr>
<tr>
<td>Canada</td>
<td>CAN</td>
</tr>
<tr>
<td>Switzerland</td>
<td>CHE</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>CZE</td>
</tr>
<tr>
<td>Denmark</td>
<td>DNK</td>
</tr>
<tr>
<td>Spain</td>
<td>ESP</td>
</tr>
<tr>
<td>Finland</td>
<td>FIN</td>
</tr>
<tr>
<td>France</td>
<td>FRA</td>
</tr>
<tr>
<td>Greece</td>
<td>GRC</td>
</tr>
<tr>
<td>Hong Kong-China</td>
<td>HKG</td>
</tr>
<tr>
<td>Hungary</td>
<td>HUN</td>
</tr>
<tr>
<td>Indonesia</td>
<td>IDN</td>
</tr>
<tr>
<td>Ireland</td>
<td>IRL</td>
</tr>
<tr>
<td>Iceland</td>
<td>ISL</td>
</tr>
<tr>
<td>Italy</td>
<td>ITA</td>
</tr>
<tr>
<td>Japan</td>
<td>JPN</td>
</tr>
<tr>
<td>Korea</td>
<td>KOR</td>
</tr>
<tr>
<td>Liechtenstein</td>
<td>LIE</td>
</tr>
<tr>
<td>Latvia</td>
<td>LVA</td>
</tr>
<tr>
<td>Macao-China</td>
<td>MAC</td>
</tr>
<tr>
<td>Mexico</td>
<td>MEX</td>
</tr>
<tr>
<td>Netherlands</td>
<td>NLD</td>
</tr>
<tr>
<td>Norway</td>
<td>NOR</td>
</tr>
<tr>
<td>New Zealand</td>
<td>NZL</td>
</tr>
<tr>
<td>Poland</td>
<td>POL</td>
</tr>
<tr>
<td>Portugal</td>
<td>PRT</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>RUS</td>
</tr>
<tr>
<td>Slovakia</td>
<td>SVK</td>
</tr>
<tr>
<td>Sweden</td>
<td>SWE</td>
</tr>
<tr>
<td>Switzerland</td>
<td>SWE</td>
</tr>
<tr>
<td>Turkey</td>
<td>TUR</td>
</tr>
<tr>
<td>Tunisia</td>
<td>TUN</td>
</tr>
<tr>
<td>Turkey</td>
<td>TUR</td>
</tr>
<tr>
<td>Uruguay</td>
<td>URY</td>
</tr>
<tr>
<td>United States</td>
<td>USA</td>
</tr>
<tr>
<td>Serbia</td>
<td>YUG</td>
</tr>
</tbody>
</table>

**Socio-economic status**
The highest occupational status of parents (HISEI) is referred to as the socio-economic status of the students throughout this manual. It should be noted that occupational status is only one aspect of socio-economic status, which can also include education and wealth. The PISA 2003 database also includes a broader socio-economic measure called the index of Economic, Social and Cultural Status (ESCS), which is derived from the highest occupational status of parents, the highest educational level and an estimate related to household possessions.
Further documentation

For further information on the PISA 2003 results, see the PISA 2003 initial reports: *Learning for Tomorrow’s World – First Results from PISA 2003* (OECD, 2004a) and *Problem Solving for Tomorrow’s World – First Measures of Cross-Curricular Competencies from PISA 2003* (OECD, 2004b). For further information on the PISA assessment instruments and the method used in PISA, see the *PISA 2003 Technical Report* (OECD, forthcoming) and the PISA Web site (www.pisa.oecd.org).