



Programme for International Student Assessment

# PISA 2003 Data Analysis Manual

SPSS® Users

**OECD**

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT





# Foreword

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

|  |    |
|--|----|
| <b>Plausible Values</b> .....  | 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

|  |    |
|--|----|
| <b>Computation of Standard Errors</b> .....                              | 81 |
| Introduction .....   | 82 |
| The standard error on univariate statistics for numerical variables..... | 82 |
| The SPSS® 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

|   |     |
|---|-----|
| <b>Analyses with Plausible Values</b> .....                               | 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..... | 104 |
| Correlation between two sets of plausible values .....                    | 104 |
| A fatal error shortcut.....   | 108 |
| An unbiased shortcut .....  | 109 |
| Conclusions.....  | 110 |

## CHAPTER 8

|   |     |
|---|-----|
| <b>Use of Proficiency Levels</b> .....      | 111 |
| Introduction .....                          | 112 |
| Generation of the proficiency levels.....   | 112 |
| Other analyses with proficiency levels..... | 118 |
| Conclusions.....                            | 121 |

## CHAPTER 9

|   |     |
|---|-----|
| <b>Analyses with School Level Variables</b> ..... | 123 |
| Introduction .....                                | 124 |
| Limits of the PISA school samples .....           | 125 |
| Merging the school and student data files.....    | 126 |
| Analyses of the school variables .....            | 126 |
| Conclusions.....                                  | 128 |



|  |     |
|--|-----|
| CHAPTER 10   |     |
| Standard Error on a Difference .....   | 129 |
| Introduction .....   | 130 |
| The standard error of a difference without plausible values .....                                    | 132 |
| The standard error of a difference with plausible values .....                                       | 137 |
| Multiple comparisons .....   | 140 |
| Conclusions .....  | 141 |
| CHAPTER 11   |     |
| OECD Average and OECD Total .....  | 143 |
| Introduction .....   | 144 |
| Recoding of the database for the estimation of the OECD total and OECD average .....                 | 144 |
| Duplication of the data for avoiding three runs of the procedure .....                               | 147 |
| Comparisons between OECD average or OECD total estimates and a country estimate .....                | 147 |
| Conclusions .....  | 150 |
| 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 SPSS® .....   | 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 .....  | 198 |
| Instability of the relative and attributable risks .....   | 200 |
| Computation of the relative risk and attributable risk .....   | 201 |
| Conclusions .....  | 201 |
| CHAPTER 15   |     |
| SPSS® Macros .....   | 203 |
| Introduction .....   | 204 |
| Structure of the SPSS® macros .....  | 204 |



|   |     |
|---|-----|
| Appendix 1: PISA 2003 International Database.....                         | 237 |
| Appendix 2: Student Questionnaire .....                                   | 251 |
| Appendix 3: Educational Career Questionnaire .....                        | 265 |
| Appendix 4: Information Communication Technology (ICT) Questionnaire..... | 267 |
| Appendix 5: School Questionnaire .....                                    | 271 |
| Appendix 6: Student Questionnaire Data File Codebook .....                | 285 |
| Appendix 7: School Questionnaire Data File Codebook.....                  | 331 |
| Appendix 8: Student Cognitive Test Data File Codebook.....                | 345 |
| Appendix 9: Student and School Questionnaire Indices .....                | 375 |
| Appendix 10: Scores Allocated to the Items .....                          | 401 |
| <i>References</i> .....   | 407 |





## USERS' GUIDE

### Preparation of data files

All data files (in text format) and the SPSS® control files are available on the PISA Web site ([www.pisa.oecd.org](http://www.pisa.oecd.org)).

### SPSS® users

The PISA 2003 student data file and the PISA 2003 school data file need to be saved under “c:\PISA\Data2003” before running the syntax in the following chapters.

Student data file: c:\PISA\Data2003\INT\_stui\_2003.sav

School data file: c:\PISA\Data2003\INT\_schi\_2003.sav

The SPSS® macros presented in Chapter 15 need to be saved under “c:\PISA\macros”.

### SPSS® syntax and macros

All syntaxes and macros used in this manual can be copied from the PISA Web site ([www.pisa.oecd.org](http://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

|     |                |     |                 |     |               |     |                    |
|-----|----------------|-----|-----------------|-----|---------------|-----|--------------------|
| AUS | Australia      | FRA | France          | KOR | Korea         | PRT | Portugal           |
| AUT | Austria        | GBR | United Kingdom  | LIE | Liechtenstein | RUS | Russian Federation |
| BEL | Belgium        | GRC | Greece          | LUX | Luxembourg    | SVK | Slovakia           |
| BRA | Brazil         | HKG | Hong Kong-China | LVA | Latvia        | SWE | Sweden             |
| CAN | Canada         | HUN | Hungary         | MAC | Macao-China   | THA | Thailand           |
| CHE | Switzerland    | IDN | Indonesia       | MEX | Mexico        | TUN | Tunisia            |
| CZE | Czech Republic | IRL | Ireland         | NLD | Netherlands   | TUR | Turkey             |
| DEU | Germany        | ISL | Iceland         | NOR | Norway        | URY | Uruguay            |
| DNK | Denmark        | ITA | Italy           | NZL | New Zealand   | USA | United States      |
| ESP | Spain          | JPN | Japan           | POL | Poland        | YUG | Serbia             |
| FIN | Finland        |     |                 |     |               |     |                    |

### 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](http://www.pisa.oecd.org)).