

6

Plausible Values

Individual estimates versus population estimates.....	94
The meaning of plausible values (PVs).....	94
Comparison of the efficiency of WLEs, EAP estimates and PVs for the estimation of some population statistics.....	97
How to perform analyses with plausible values.....	100
Conclusion.....	101



INDIVIDUAL ESTIMATES VERSUS POPULATION ESTIMATES

Education assessments can have two major purposes:

1. To measure the knowledge and skills of particular students. The performance of each student usually will have an impact on his or her future (school career, admission to post-secondary education, and so on). It is therefore particularly important to minimise the measurement error associated with each individual's estimate.
2. To assess the knowledge or skills of a population. The performance of individuals will have no impact on their school career or professional life. In such a case, the goal of reducing error in making inferences about the target population is more important than the goal of reducing errors at the individual level.

National and international education surveys belong to the second category.

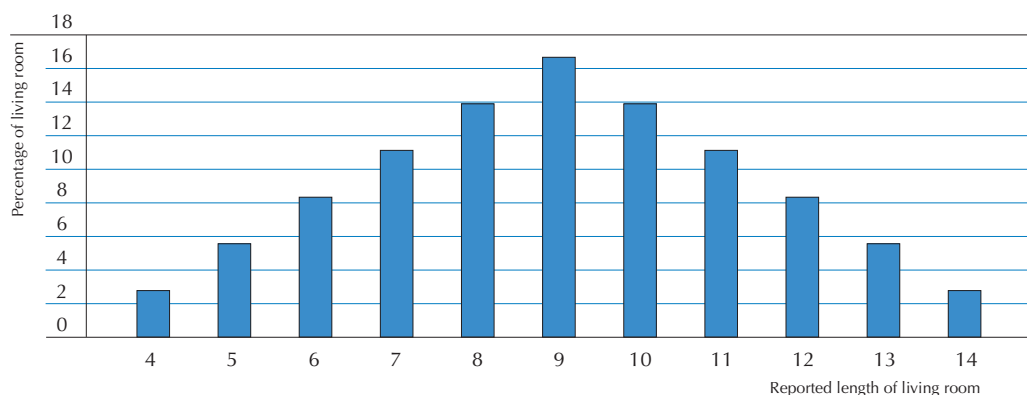
International surveys such as PISA report student performance through plausible values.¹ This chapter will explain the conceptual meaning of plausible values and the advantage of reporting with them. Individual estimators (such as the *weighted likelihood estimate* [WLE] defined in Chapter 5) will be compared with plausible values for the purposes of estimating a range of population statistics.

THE MEANING OF PLAUSIBLE VALUES (PVs)

An example taken from the physical sciences measurement area can help illustrate the complex concept of plausible values. Suppose that a city board decides to levy a new building tax to increase the city's revenue. This new tax will be proportional to the length of the family house living room. Inspectors visit all city houses to measure the length of the living rooms. They are given a measuring tape and are instructed to record the length in term of integers only, *i.e.* 1 metre, 2 metres, 3 metres, 4 metres and so on.

The results of this measure are shown in Figure 6.1. About 2% of the living rooms have a reported length of 4 metres; slightly over 16% of the living rooms have a reported length of 9 metres and so on.

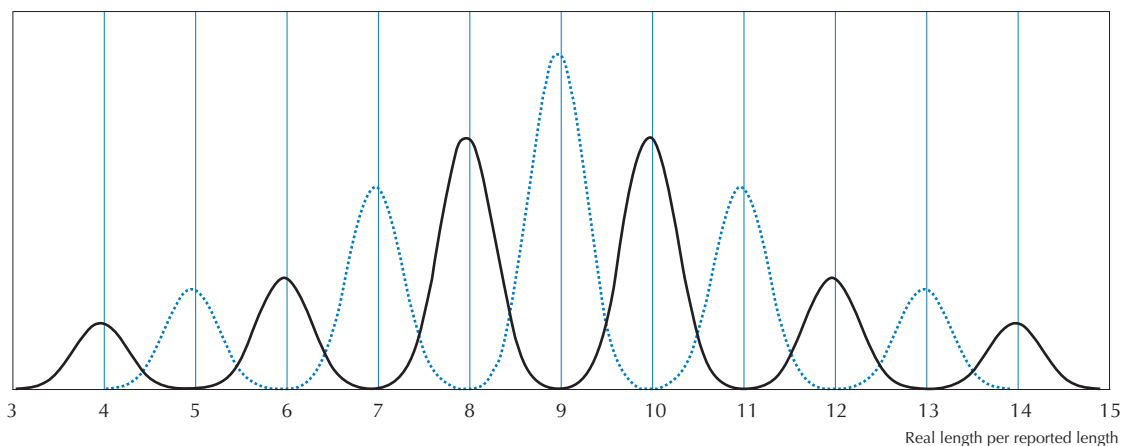
Figure 6.1
Living room length expressed in integers



Of course, the reality is quite different, as length is a continuous variable. With a continuous variable, observations can take any value between the minimum and the maximum. On the other hand, with a discontinuous variable, observations can only take a predefined number of values. Figure 6.2 gives the length distribution of the living rooms per reported length.



Figure 6.2
Real length per reported length



All living rooms with a reported length of 5 metres are not exactly 5 metres long. On average, they are 5 metres long, but their length varies around the mean. The difference between reported length and real length is due to the rounding process and measurement error. An inspector might incorrectly report 5 metres for a particular living room, when it really measures 4.15 metres. If the rounding process were the only source of error, then the reported length should be 4 metres. The second source of error, the error in measuring, explains the overlap in the distribution.

In this particular example, the lengths of the living rooms are normally distributed around the mean, which is also the reported length. If the difference between the length and the closest integer is small, then the probability of not reporting this length with the closest integer is very small. For instance, it is unlikely that a length of 4.15 be reported as 5 metres or 3 metres. However, as the distance between the real length and the closest integer increases, the probability of not reporting this length with the closest integer will also increase. For instance, it is likely that a length of 4.95 will be reported as 5 metres, whereas a length of 4.50 will be reported equally as many times as 4 metres as it is 5 metres.

The methodology of plausible values consists of:

- mathematically computing distributions (denoted as posterior distributions) around the reported values and the reported length in the example; and
- assigning to each observation a set of random values drawn from the posterior distributions.

Plausible values can therefore be defined as random values from the posterior distributions. In the example, a living room of 7.154 metres that was reported as 7 metres might be assigned any value from the normal distribution around the reported length of 7. It might be 7.45 as well as 6.55 or 6.95. Therefore, plausible values should not be used for individual estimation.

This fictitious example from the physical sciences can be translated successfully to the social sciences. For example, with a test of six dichotomous items, a continuous variable (*i.e.* mental ability) can be transformed into a discontinuous variable. The discontinuous variable will be the student raw score or the number of correct answers. The only possible scores are: 0, 1, 2, 3, 4, 5 and 6.



Contrary to most measures in the physical sciences, psychological or education measures encompass substantial measurement errors because:

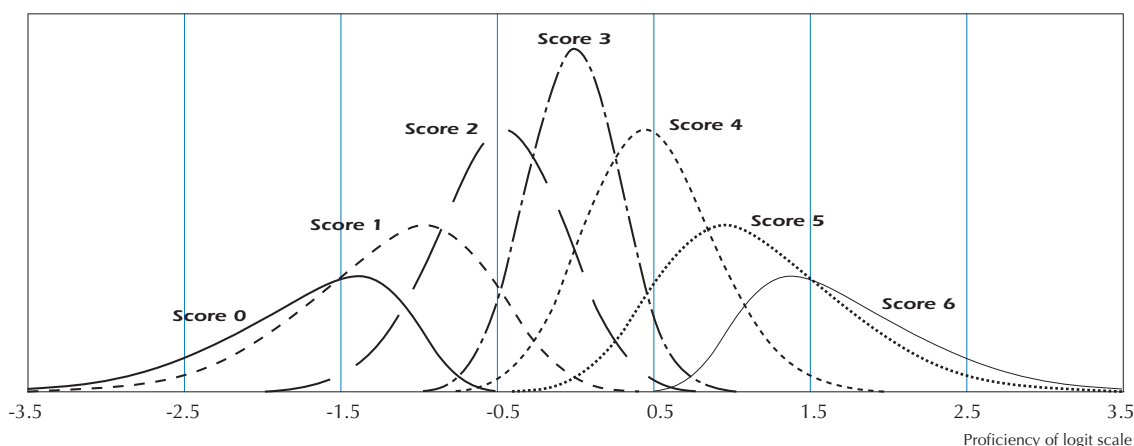
- The concept to be measured is broader.
- They might be affected by the mental and physical dispositions of the students on the day of the assessment.
- The conditions in which students are tested might also affect the results.

This means that there are large overlaps in the posterior distributions, as shown in Figure 6.3.

Further, with the example of the living room, the measurement error of the posterior distributions can be considered as independent of the living room.² In education, the measurement error is not always independent of the proficiency level of the students. It may be smaller for average students, and larger for low and high achievers, depending on the test average difficulty.

Further, in this particular example, the posterior distributions for score 0 and score 6 are substantially skewed, as the posterior distributions of the living rooms with a reported length of 4 and 14 metres would be, if all living rooms smaller than 4 metres were reported as 4, and if all living rooms longer than 14 metres were reported as 14. This means that the posterior distributions are not normally distributed, as shown in Figure 6.3.

Figure 6.3
A posterior distribution on a test of six items



Generating plausible values on an education test consists of drawing random numbers from the posterior distributions. This example clearly shows that plausible values should not be used for individual performance. Indeed, a student who scores 0 might get -3, but also -1. A student who scores 6 might get 3, but also 1.

It has been noted that:

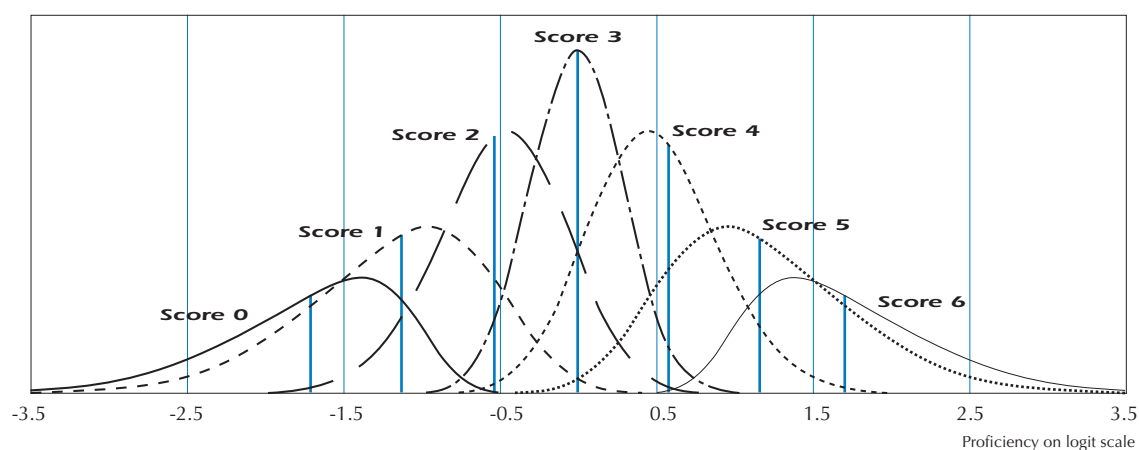
“The simplest way to describe plausible values is to say that plausible values are a representation of the range of abilities that a student might reasonably have. [...] Instead of directly estimating a student’s ability θ , a probability distribution for a student’s θ is estimated. That is, instead of obtaining a point estimate for θ , like a WLE, a range of possible values for a student’s θ , with an associated probability for each of these values is estimated. Plausible values are random draws from this (estimated) distribution for a student’s θ ” (Wu and Adams, 2002).³



All this methodology aims at building a continuum from a collection of discontinuous variables (*i.e.* the test score). It is meant to prevent biased inferences occurring as a result of measuring an unobservable underlying ability through a test using a relatively small number of items.

Finally, an individual estimate of student ability can also be derived from the posterior distributions. This derived individual estimate is called the *expected a posteriori* estimator (EAP). Instead of assigning a set of random values from the posterior distributions, the mean of the posterior distributions is assigned. Therefore, the EAP can be considered as the mean of an infinite set of plausible values for a particular student. See Figure 6.4.

Figure 6.4
EAP estimators



As only one value is assigned per posterior distribution, the EAP estimator is also a discontinuous variable.⁴ However, EAP estimates and WLEs differ as the former requires a population distribution assumption, which is not the case for the latter. Further, while any raw score for a particular test will always be associated with one and only one WLE, different EAP values can be associated with a particular raw score, depending on the regressors used as conditioning variables.

Researchers not used to working with plausible values might consider this apparent randomisation as a source of imprecision. The comparison of the different types of Rasch ability estimators (WLE, EAP and PV) through the estimation of population statistics will overcome this perception. The PISA database only includes PVs⁵ for student performance but uses WLE for contextual indices derived from the student or school questionnaire. Although PISA does not include any EAP in its databases, the comparison will incorporate EAP estimates to show biases that occur when data analysts average the plausible values at the student levels to obtain one score value per student.

COMPARISON OF THE EFFICIENCY OF WLEs, EAP ESTIMATES AND PVs FOR THE ESTIMATION OF SOME POPULATION STATISTICS⁶

A comparison between different student ability estimators could be performed on real data. Such a comparison will outline differences, but it will not identify the best estimators for a particular population statistic. Therefore, a simulation will be used to illustrate and evaluate the differences in various estimators.



The simulation consists of three major steps:

- The generation of a dataset including a continuous variable that represents the student abilities (*i.e.* denoted as the latent variable), some background variables including the gender and an index of social background, denoted HISEI, and a pattern of item responses coded 0 for an incorrect answer and 1 for a correct answer. The results presented hereafter are based on a fictitious test of 15 items.⁷
- The computation of the student ability estimator, in particular the WLEs, EAP estimates and PVs.⁸
- The estimation of some population parameters using the student ability (*i.e.* latent variable) and the different student ability estimators. A comparison will be made for:
 - mean, variance and percentiles,
 - correlation,
 - between- and within-school variance.

The dataset of this simulation contains 5 250 students distributed in 150 schools with 35 students per school. Table 6.1 presents the structure of the simulated dataset before the importation of the Rasch student ability estimators.

Table 6.1
Structure of the simulated data

School ID	Student ID	Sex	HISEI	Item 1	Item 2	...	Item 14	Item 15
001	01	1	32	1	1		0	0
001	02	0	45	1	0		1	0
...	...							
150	5 249	0	62	0	0		1	1
150	5 250	1	50	0	1		1	1

Table 6.2 presents the mean and the variance of the latent variable, the WLEs, the EAP estimates and the five plausible values. The average of the 5 PV mean is also included.

Table 6.2 shows that a good estimate of the population's mean (*i.e.* the latent variable estimate) is obtained regardless of the type of latent variable used (WLEs, EAP estimates or PVs). It can be empirically demonstrated that none of the estimates significantly differs from the expected mean, *i.e.* 0.00 in this particular case (Wu and Adams, 2002). Additionally, it can also be shown that the mean of the WLEs will not be biased if the test is well targeted, *i.e.* if the average of the item difficulties is around 0 on the Rasch scale (Wu and Adams, 2002). That is, on a well-targeted test, students will obtain a raw score of about 50% correct answers. If the test is too easy then the mean of the WLEs will be underestimated (this is called the ceiling effect), while if it is too difficult then the mean of the WLEs will be overestimated (this is called the floor effect).

Table 6.2
Means and variances for the latent variables and the different student ability estimators

	Mean	Variance
Latent variable	0.00	1.00
WLE	0.00	1.40
EAP	0.00	0.75
PV1	0.01	0.99
PV2	0.00	0.99
PV3	0.00	1.01
PV4	0.00	1.01
PV5	-0.01	0.00
Average of the 5 PV statistics	0.00	1.00



These last results explain why the mean of the WLEs provided in the PISA 2000 database differs from the mean of the plausible values, especially for the partner countries. For the reading/reflection and evaluation scale, the means obtained for Canada using WLEs and PVs are 538.4 and 542.5, respectively, which are very close. In contrast, the means obtained for Peru, using WLEs and PVs are 352.2 and 322.7, respectively, a difference of about 0.3 standard deviations. This shows that there is bias when WLEs are used to estimate the mean if the test is not well targeted.

For the population variance, Table 6.2 shows that PVs give estimates closest to the expected value, while WLEs overestimate it and the EAP underestimates it. These results are consistent with other simulation studies.

Table 6.3 presents some percentiles computed on the different ability estimators. For example, because the variance computed using plausible values is not biased, the percentiles based on PVs are also unbiased. However, because the EAP estimates and WLEs variances are biased, the percentiles, and in particular, extreme percentiles will also be biased. These results are consistent with other simulation studies previously cited.

Table 6.4 presents the correlation between the social background index (HISEI), gender and the latent variables and the different estimators of student abilities. The correlation coefficients with the WLEs are both underestimated, while the correlation coefficients with the EAP estimates are overestimated. Only the correlation coefficients with the plausible values are unbiased.⁹

It should be noted that the regression coefficients are all unbiased for the different types of estimators. Nevertheless, as variances are biased for some estimators, residual variances will also be biased. Therefore, the standard error on the regression coefficients will be biased in the case of the WLEs and the EAP estimates.

Table 6.3
Percentiles for the latent variables and the different student ability estimators

	P5	P10	P25	P50	P75	P90	P95
Latent variable	-1.61	-1.26	-0.66	0.01	0.65	1.26	1.59
WLE	-2.15	-1.65	-0.82	-0.1	0.61	1.38	1.81
EAP	-1.48	-1.14	-0.62	-0.02	0.55	1.08	1.37
PV1	-1.68	-1.29	-0.71	-0.03	0.64	1.22	1.59
PV2	-1.67	-1.31	-0.69	-0.03	0.62	1.22	1.58
PV3	-1.67	-1.32	-0.70	-0.02	0.64	1.21	1.56
PV4	-1.69	-1.32	-0.69	-0.03	0.63	1.23	1.55
PV5	-1.65	-1.3	-0.71	-0.02	0.62	1.2	1.55
Average of the 5 PV statistics	-1.67	-1.31	-0.70	-0.03	0.63	1.22	1.57

Table 6.4
Correlation between HISEI, gender and the latent variable, the different student ability estimators

	HISEI	GENDER
Latent variable	0.40	0.16
WLE	0.33	0.13
EAP	0.46	0.17
PV1	0.41	0.15
PV2	0.42	0.15
PV3	0.42	0.13
PV4	0.40	0.15
PV5	0.40	0.14
Average of the 5 PV statistics	0.41	0.14



Table 6.5
Between- and within-school variances

	Between-school variance	Within-school variance
Latent variable	0.33	0.62
WLE	0.34	1.02
EAP	0.35	0.38
PV1	0.35	0.61
PV2	0.36	0.60
PV3	0.36	0.61
PV4	0.35	0.61
PV5	0.35	0.61
Average of the 5 PV statistics	0.35	0.61

Finally, Table 6.5 presents the between- and within-school variances. Between-school variances for the different estimators do not differ from the expected value of 0.33. However, WLEs overestimate the within-school variance, while the EAP estimates underestimate it. These results are consistent with other simulation studies (Monseur and Adams, 2002).

As this example shows, plausible values provide unbiased estimates.

HOW TO PERFORM ANALYSES WITH PLAUSIBLE VALUES

As stated in the previous section, a set of plausible values, usually five, are drawn for each student for each scale or subscale. Population statistics should be estimated using each plausible value separately. The reported population statistic is then the average of each plausible value statistic. For instance, if one is interested in the correlation coefficient between the social index and the reading performance in PISA, then five correlation coefficients should be computed and then averaged.

Plausible values should never be averaged at the student level, *i.e.* by computing in the dataset the mean of the five plausible values at the student level and then computing the statistic of interest once using that average PV value. Doing so would be equivalent to an EAP estimate, with a bias as described in the previous section.

Mathematically, secondary analyses with plausible values can be described as follows. If θ is the population statistic and θ_i is the statistic of interest computed on one plausible value, then:

$$\theta = \frac{1}{M} \sum_{i=1}^M \theta_i$$

where M is the number of plausible values.

Plausible values also allow computing the uncertainty in the estimate of θ due to the lack of precision in the measurement test. If a perfect test could be developed, then the measurement error would be equal to zero and the five statistics from the plausible values would be exactly identical. Unfortunately, perfect tests do not exist and never will. This measurement variance, usually denoted imputation variance, is equal to:

$$B_M = \frac{1}{M-1} \sum_{i=1}^M (\theta_i - \theta)^2$$

It corresponds to the variance of the five plausible value statistics of interest. The final stage is to combine the sampling variance and the imputation variance as follows:

$$V = U + \left(1 + \frac{1}{M}\right) B_m$$

where U is the sampling variance.



Chapter 7 will demonstrate how to compute sampling variances and imputation variances and how to combine them, using the PISA databases.

CONCLUSION

This chapter was devoted to the meaning of plausible values and the steps required to analyse data with PVs. A comparison between PVs and alternate individual ability estimates was presented to demonstrate the superiority of this methodology for reporting population estimates.

Notes

1. The methodology of plausible values was first implemented in the National Assessment of Educational Progress (NAEP) studies. For more information, see Beaton (1987).
2. The measurement error will be independent of the length of the living rooms if the inspectors are using a measuring instrument that is at least 15 metres long (such as a measuring tape). If they are using a standard metre, then the overall measurement error will be proportional to the length of the living room.
3. The probability distribution for a student's θ can be based on the cognitive data only, *i.e.* the item response pattern, but can also include additional information, such as student gender, social background, and so on. The probability distribution becomes therefore conditioned by this additional information. A mathematical explanation of the model used for the scaling of the PISA 2000 scaling can be found in the *PISA 2000 Technical Report* (OECD, 2002c).
4. If several regressors are used as conditioning variables, then the EAP estimator tends to be a continuous variable.
5. PISA 2000 data files include both WLEs and PVs.
6. PV and EAP estimators can be computed with or without regressors. As the PVs in PISA were generated based on all variables collected through the student questionnaires, this comparison will only include PVs and EAP estimators with the use of regressors.
7. The data generation starts with a factorial analysis on a 3 by 3 squared correlation matrix. The correlation between the latent variable and gender was set at 0.20, the correlation between the latent variable and the social background indicator was set at 0.40 and the correlation between gender and the social background indicator was set at 0.00. Three random variables are drawn from normal distributions and combined according to the factorial regression coefficients to create the three variables of interest, *i.e.* reading, gender and social background. Based on the student score on the latent variable and a predefined set of 20 item difficulties; probabilities of success are computed according to the Rasch Model. These probabilities are then compared to uniform distribution and recoded into 0 and 1. Finally, gender is recoded into a dichotomous variable.
8. The estimators were computed with the Conquest Software® developed by M.L. Wu, R.J. Adams and M.R. Wilson.
9. The results on the EAP and PV correlation coefficients are observed when the probability distributions are generated with conditioning variables. Without the conditioning, the correlation with the plausible values would be underestimated.



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 Boeck, 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?* *aSoc. Forces*, No. 78, pp. 1349-79.
- Cochran, W.G.** (1977), *Sampling Techniques*, J. Wiley and Sons, Inc., New York.
- Dunn, O.J.** (1961), "Multiple 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

FOREWORD	3
USER'S GUIDE	17
CHAPTER 1 THE USEFULNESS OF PISA DATA FOR POLICY MAKERS, RESEARCHERS AND EXPERTS ON METHODOLOGY	19
PISA – an overview	20
▪ The PISA surveys.....	20
How can PISA contribute to educational policy, practice and research?	22
▪ Key results from PISA 2000, PISA 2003 and PISA 2006.....	23
Further analyses of PISA datasets	25
▪ Contextual framework of PISA 2006.....	28
▪ Influence of the methodology on outcomes.....	31
CHAPTER 2 EXPLORATORY ANALYSIS PROCEDURES	35
Introduction	36
Weights	36
Replicates for computing the standard error	39
Plausible values	43
Conclusion	45
CHAPTER 3 SAMPLE WEIGHTS	47
Introduction	48
Weights for simple random samples	49
Sampling designs for education surveys	51
Why do the PISA weights vary?	55
Conclusion	56
CHAPTER 4 REPLICATE WEIGHTS	57
Introduction	58
Sampling variance for simple random sampling	58
Sampling variance for two-stage sampling	63
Replication methods for simple random samples	68
Replication methods for two-stage samples	70
▪ The Jackknife for unstratified two-stage sample designs.....	70
▪ The Jackknife for stratified two-stage sample designs.....	71
▪ The Balanced Repeated Replication method.....	72
Other procedures for accounting for clustered samples	74
Conclusion	74

CHAPTER 5 THE RASCH MODEL	77
Introduction	78
How can the information be summarised?	78
The Rasch Model for dichotomous items	79
▪ Introduction to the Rasch Model.....	79
▪ Item calibration.....	83
▪ Computation of a student's score.....	85
▪ Computation of a student's score for incomplete designs.....	89
▪ Optimal conditions for linking items.....	90
▪ Extension of the Rasch Model.....	91
Other item response theory models	92
Conclusion	92
 CHAPTER 6 PLAUSIBLE VALUES	 93
Individual estimates versus population estimates	94
The meaning of plausible values (PVs)	94
Comparison of the efficiency of WLEs, EAP estimates and PVs for the estimation of some population statistics	97
How to perform analyses with plausible values	100
Conclusion	101
 CHAPTER 7 COMPUTATION OF STANDARD ERRORS	 103
Introduction	104
The standard error on univariate statistics for numerical variables	104
The SPSS® macro for computing the standard error on a mean	107
The standard error on percentages	110
The standard error on regression coefficients	112
The standard error on correlation coefficients	114
Conclusion	115
 CHAPTER 8 ANALYSES WITH PLAUSIBLE VALUES	 117
Introduction	118
Univariate statistics on plausible values	118
The standard error on percentages with PVs	121
The standard error on regression coefficients with PVs	121
The standard error on correlation coefficients with PVs	124
Correlation between two sets of plausible values	124
A fatal error shortcut	128
An unbiased shortcut	129
Conclusion	130
 CHAPTER 9 USE OF PROFICIENCY LEVELS	 133
Introduction	134
Generation of the proficiency levels	134
Other analyses with proficiency levels	139
Conclusion	141



CHAPTER 10 ANALYSES WITH SCHOOL-LEVEL VARIABLES	143
Introduction	144
Limits of the PISA school samples	145
Merging the school and student data files	146
Analyses of the school variables	146
Conclusion	148
CHAPTER 11 STANDARD ERROR ON A DIFFERENCE	149
Introduction	150
Statistical issues and computing standard errors on differences	150
The standard error on a difference without plausible values	152
The standard error on a difference with plausible values	157
Multiple comparisons	161
Conclusion	162
CHAPTER 12 OECD TOTAL AND OECD AVERAGE	163
Introduction	164
Recoding of the database to estimate the pooled OECD total and the pooled OECD average	166
Duplication of the data to avoid running the procedure three times	168
Comparisons between the pooled OECD total or pooled OECD average estimates and a country estimate	169
Comparisons between the arithmetic OECD total or arithmetic OECD average estimates and a country estimate	171
Conclusion	171
CHAPTER 13 TRENDS	173
Introduction	174
The computation of the standard error for trend indicators on variables other than performance	175
The computation of the standard error for trend indicators on performance variables	177
Conclusion	181
CHAPTER 14 STUDYING THE RELATIONSHIP BETWEEN STUDENT PERFORMANCE AND INDICES DERIVED FROM CONTEXTUAL QUESTIONNAIRES	183
Introduction	184
Analyses by quarters	184
The concept of relative risk	186
▪ Instability of the relative risk	187
▪ Computation of the relative risk	188
Effect size	191
Linear regression and residual analysis	193
▪ Independence of errors	193
Statistical procedure	196
Conclusion	197



CHAPTER 15 MULTILEVEL ANALYSES	199
Introduction	200
Two-level modelling with SPSS®	202
▪ Decomposition of the variance in the empty model.....	202
▪ Models with only random intercepts.....	205
▪ Shrinkage factor.....	207
▪ Models with random intercepts and fixed slopes.....	207
▪ Models with random intercepts and random slopes.....	209
▪ Models with Level 2 independent variables.....	214
▪ Computation of final estimates and their respective standard errors.....	217
Three-level modelling	219
Limitations of the multilevel model in the PISA context	221
Conclusion	222
 CHAPTER 16 PISA AND POLICY RELEVANCE – THREE EXAMPLES OF ANALYSES	 223
Introduction	224
Example 1: Gender differences in performance	224
Example 2: Promoting socio-economic diversity within school?	228
Example 3: The influence of an educational system on the expected occupational status of students at age 30	234
Conclusion	237
 CHAPTER 17 SPSS® MACRO	 239
Introduction	240
Structure of the SPSS® Macro	240
 REFERENCES	 321
 APPENDICES	 323
Appendix 1 Three-level regression analysis.....	324
Appendix 2 PISA 2006 International database.....	332
Appendix 3 PISA 2006 Student questionnaire.....	341
Appendix 4 PISA 2006 Information communication technology (ICT) Questionnaire.....	350
Appendix 5 PISA 2006 School questionnaire.....	352
Appendix 6 PISA 2006 Parent questionnaire.....	359
Appendix 7 Codebook for PISA 2006 student questionnaire data file.....	363
Appendix 8 Codebook for PISA 2006 non-scored cognitive and embedded attitude items.....	407
Appendix 9 Codebook for PISA 2006 scored cognitive and embedded attitude items.....	427
Appendix 10 Codebook for PISA 2006 school questionnaire data file.....	439
Appendix 11 Codebook for PISA 2006 parents questionnaire data file.....	450
Appendix 12 PISA 2006 questionnaire indices.....	456



LIST OF BOXES

Box 2.1	WEIGHT statement in SPSS®.....	37
<hr/>		
Box 7.1	SPSS® syntax for computing 81 means (e.g. PISA 2003).....	104
Box 7.2	SPSS® syntax for computing the mean of HISEI and its standard error (e.g. PISA 2003).....	107
Box 7.3	SPSS® syntax for computing the standard deviation of HISEI and its standard error by gender (e.g. PISA 2003).....	109
Box 7.4	SPSS® syntax for computing the percentages and their standard errors for gender (e.g. PISA 2003).....	110
Box 7.5	SPSS® syntax for computing the percentages and its standard errors for grades by gender (e.g. PISA 2003).....	112
Box 7.6	SPSS® syntax for computing regression coefficients, R^2 and its respective standard errors: Model 1 (e.g. PISA 2003).....	113
Box 7.7	SPSS® syntax for computing regression coefficients, R^2 and its respective standard errors: Model 2 (e.g. PISA 2003).....	114
Box 7.8	SPSS® syntax for computing correlation coefficients and its standard errors (e.g. PISA 2003).....	114
<hr/>		
Box 8.1	SPSS® syntax for computing the mean on the science scale by using the MCR_SE_UNIV macro (e.g. PISA 2006).....	119
Box 8.2	SPSS® syntax for computing the mean and its standard error on PVs (e.g. PISA 2006).....	120
Box 8.3	SPSS® syntax for computing the standard deviation and its standard error on PVs by gender (e.g. PISA 2006).....	131
Box 8.4	SPSS® syntax for computing regression coefficients and their standard errors on PVs by using the MCR_SE_REG macro (e.g. PISA 2006).....	122
Box 8.5	SPSS® syntax for running the simple linear regression macro with PVs (e.g. PISA 2006).....	123
Box 8.6	SPSS® syntax for running the correlation macro with PVs (e.g. PISA 2006).....	124
Box 8.7	SPSS® syntax for the computation of the correlation between mathematics/quantity and mathematics/space and shape by using the MCR_SE_COR_2PV macro (e.g. PISA 2003).....	126
<hr/>		
Box 9.1	SPSS® syntax for generating the proficiency levels in science (e.g. PISA 2006).....	135
Box 9.2	SPSS® syntax for computing the percentages of students by proficiency level in science and its standard errors (e.g. PISA 2006).....	136
Box 9.3	SPSS® syntax for computing the percentage of students by proficiency level in science and its standard errors (e.g. PISA 2006).....	138
Box 9.4	SPSS® syntax for computing the percentage of students by proficiency level and its standard errors by gender (e.g. PISA 2006).....	138
Box 9.5	SPSS® syntax for generating the proficiency levels in mathematics (e.g. PISA 2003).....	139
Box 9.6	SPSS® syntax for computing the mean of self-efficacy in mathematics and its standard errors by proficiency level (e.g. PISA 2003).....	140
<hr/>		
Box 10.1	SPSS® syntax for merging the student and school data files (e.g. PISA 2006).....	146
Box 10.2	Question on school location in PISA 2006.....	147
Box 10.3	SPSS® syntax for computing the percentage of students and the average performance in science, by school location (e.g. PISA 2006).....	147
<hr/>		
Box 11.1	SPSS® syntax for computing the mean of job expectations by gender (e.g. PISA 2003).....	152
Box 11.2	SPSS® macro for computing standard errors on differences (e.g. PISA 2003).....	155



Box 11.3	Alternative SPSS® macro for computing the standard error on a difference for a dichotomous variable (e.g. PISA 2003).....	156
Box 11.4	SPSS® syntax for computing standard errors on differences which involve PVs (e.g. PISA 2003).....	158
Box 11.5	SPSS® syntax for computing standard errors on differences that involve PVs (e.g. PISA 2006).....	160
<hr/>		
Box 12.1	SPSS® syntax for computing the pooled OECD total for the mathematics performance by gender (e.g. PISA 2003).....	166
Box 12.2	SPSS® syntax for the pooled OECD average for the mathematics performance by gender (e.g. PISA 2003).....	167
Box 12.3	SPSS® syntax for the creation of a larger dataset that will allow the computation of the pooled OECD total and the pooled OECD average in one run (e.g. PISA 2003).....	168
<hr/>		
Box 14.1	SPSS® syntax for the quarter analysis (e.g. PISA 2006).....	185
Box 14.2	SPSS® syntax for computing the relative risk with five antecedent variables and five outcome variables (e.g. PISA 2006).....	189
Box 14.3	SPSS® syntax for computing the relative risk with one antecedent variable and one outcome variable (e.g. PISA 2006).....	190
Box 14.4	SPSS® syntax for computing the relative risk with one antecedent variable and five outcome variables (e.g. PISA 2006).....	190
Box 14.5	SPSS® syntax for computing effect size (e.g. PISA 2006).....	192
Box 14.6	SPSS® syntax for residual analyses (e.g. PISA 2003).....	196
<hr/>		
Box 15.1	Normalisation of the final student weights (e.g. PISA 2006).....	203
Box 15.2	SPSS® syntax for the decomposition of the variance in student performance in science (e.g. PISA 2006).....	203
Box 15.3	SPSS® syntax for normalising PISA 2006 final student weights with deletion of cases with missing values and syntax for variance decomposition (e.g. PISA 2006).....	206
Box 15.4	SPSS® syntax for a multilevel regression model with random intercepts and fixed slopes (e.g. PISA 2006).....	208
Box 15.5	Results for the multilevel model in Box 15.4.....	208
Box 15.6	SPSS® syntax for a multilevel regression model (e.g. PISA 2006).....	210
Box 15.7	Results for the multilevel model in Box 15.6.....	211
Box 15.8	Results for the multilevel model with covariance between random parameters.....	212
Box 15.9	Interpretation of the within-school regression coefficient.....	214
Box 15.10	SPSS® syntax for a multilevel regression model with a school-level variable (e.g. PISA 2006).....	214
Box 15.11	SPSS® syntax for a multilevel regression model with interaction (e.g. PISA 2006).....	215
Box 15.12	Results for the multilevel model in Box 15.11.....	216
Box 15.13	SPSS® syntax for using the multilevel regression macro (e.g. PISA 2006).....	217
Box 15.14	SPSS® syntax for normalising the weights for a three-level model (e.g. PISA 2006).....	219
<hr/>		
Box 16.1	SPSS® syntax for testing the gender difference in standard deviations of reading performance (e.g. PISA 2000).....	225
Box 16.2	SPSS® syntax for computing the 5th percentile of the reading performance by gender (e.g. PISA 2000).....	227
Box 16.3	SPSS® syntax for preparing a data file for the multilevel analysis.....	230



Box 16.4	SPSS® syntax for running a preliminary multilevel analysis with one PV	231
Box 16.5	Estimates of fixed parameters in the multilevel model.....	231
Box 16.6	SPSS® syntax for running preliminary analysis with the MCR_ML_PV macro.....	233
Box 17.1	SPSS® macro of MCR_SE_UNI.sps.....	243
Box 17.2	SPSS® macro of MCR_SE_PV.sps.....	247
Box 17.3	SPSS® macro of MCR_SE_PERCENTILES_PV.sps	251
Box 17.4	SPSS® macro of MCR_SE_GrpPct.sps.....	254
Box 17.5	SPSS® macro of MCR_SE_PctLev.sps.....	257
Box 17.6	SPSS® macro of MCR_SE_REG.sps	261
Box 17.7	SPSS® macro of MCR_SE_REG_PV.sps.....	265
Box 17.8	SPSS® macro of MCR_SE_COR.sps.....	270
Box 17.9	SPSS® macro of MCR_SE_COR_1PV.sps.....	273
Box 17.10	SPSS® macro of MCR_SE_COR_2PV.sps.....	277
Box 17.11	SPSS® macro of MCR_SE_DIFF.sps.....	281
Box 17.12	SPSS® macro of MCR_SE_DIFF_PV.sps.....	285
Box 17.13	SPSS® macro of MCR_SE_PV_WLEQRT.sps.....	290
Box 17.14	SPSS® macro of MCR_SE_RR.sps.....	295
Box 17.15	SPSS® macro of MCR_SE_RR_PV.sps.....	298
Box 17.16	SPSS® macro of MCR_SE_EFFECT.sps.....	302
Box 17.17	SPSS® macro of MCR_SE_EFFECT_PV.sps	306
Box 17.18	SPSS® macro of MCR_ML.sps.....	311
Box 17.19	SPSS® macro of MCR_ML_PV.sps	315
Box A1.1	Descriptive statistics of background and explanatory variables.....	326
Box A1.2	Background model for student performance.....	327
Box A1.3	Final net combined model for student performance.....	328
Box A1.4	Background model for the impact of socio-economic background.....	329
Box A1.5	Model of the impact of socio-economic background: “school resources” module.....	330
Box A1.6	Model of the impact of socio-economic background: “accountability practices” module	331
Box A1.7	Final combined model for the impact of socio-economic background.....	331

LIST OF FIGURES

Figure 1.1	Relationship between social and academic segregations.....	27
Figure 1.2	Relationship between social segregation and the correlation between science performance and student HISEI	27
Figure 1.3	Conceptual grid of variable types.....	29
Figure 1.4	Two-dimensional matrix with examples of variables collected or available from other sources	30
Figure 2.1	Science mean performance in OECD countries (PISA 2006).....	37
Figure 2.2	Gender differences in reading in OECD countries (PISA 2000).....	38
Figure 2.3	Regression coefficient of ESCS on mathematic performance in OECD countries (PISA 2003).....	38
Figure 2.4	Design effect on the country mean estimates for science performance and for ESCS in OECD countries (PISA 2006)	41
Figure 2.5	Simple random sample and unbiased standard errors of ESCS on science performance in OECD countries (PISA 2006)	42



Figure 4.1	Distribution of the results of 36 students.....	58
Figure 4.2	Sampling variance distribution of the mean.....	60
Figure 5.1	Probability of success for two high jumpers by height (dichotomous).....	80
Figure 5.2	Probability of success for two high jumpers by height (continuous).....	81
Figure 5.3	Probability of success to an item of difficulty zero as a function of student ability.....	81
Figure 5.4	Student score and item difficulty distributions on a Rasch continuum.....	84
Figure 5.5	Response pattern probabilities for the response pattern (1, 1, 0, 0).....	86
Figure 5.6	Response pattern probabilities for a raw score of 1.....	87
Figure 5.7	Response pattern probabilities for a raw score of 2.....	88
Figure 5.8	Response pattern probabilities for a raw score of 3.....	88
Figure 5.9	Response pattern likelihood for an easy test and a difficult test.....	89
Figure 5.10	Rasch item anchoring.....	90
Figure 6.1	Living room length expressed in integers.....	94
Figure 6.2	Real length per reported length.....	95
Figure 6.3	A posterior distribution on a test of six items.....	96
Figure 6.4	EAP estimators.....	97
Figure 8.1	A two-dimensional distribution.....	125
Figure 8.2	Axes for two-dimensional normal distributions.....	125
Figure 13.1	Trend indicators in PISA 2000, PISA 2003 and PISA 2006.....	175
Figure 14.1	Percentage of schools by three school groups (PISA 2003).....	194
Figure 15.1	Simple linear regression analysis versus multilevel regression analysis.....	201
Figure 15.2	Graphical representation of the between-school variance reduction.....	209
Figure 15.3	A random multilevel model.....	210
Figure 15.4	Change in the between-school residual variance for a fixed and a random model.....	212
Figure 16.1	Relationship between the segregation index of students' expected occupational status and the segregation index of student performance in reading (PISA 2000).....	236
Figure 16.2	Relationship between the segregation index of students' expected occupational status and the correlation between HISEI and students' expected occupational status.....	236

LIST OF TABLES

Table 1.1	Participating countries/economies in PISA 2000, PISA 2003, PISA 2006 and PISA 2009.....	21
Table 1.2	Assessment domains covered by PISA 2000, PISA 2003 and PISA 2006.....	22
Table 1.3	Correlation between social inequities and segregations at schools for OECD countries.....	28
Table 1.4	Distribution of students per grade and per ISCED level in OECD countries (PISA 2006).....	31
Table 2.1	Design effect and type I errors.....	40
Table 2.2	Mean estimates and standard errors.....	44



Table 2.3	Standard deviation estimates and standard errors.....	44
Table 2.4	Correlation estimates and standard errors.....	45
Table 2.5	ESCS regression coefficient estimates and standard errors.....	45
<hr/>		
Table 3.1	Height and weight of ten persons	50
Table 3.2	Weighted and unweighted standard deviation estimate	50
Table 3.3	School, within-school, and final probability of selection and corresponding weights for a two-stage, simple random sample with the first-stage units being schools of equal size.....	52
Table 3.4	School, within-school, and final probability of selection and corresponding weights for a two-stage, simple random sample with the first-stage units being schools of unequal size	52
Table 3.5	School, within-school, and final probability of selection and corresponding weights for a simple and random sample of schools of unequal size (smaller schools)	53
Table 3.6	School, within-school, and final probability of selection and corresponding weights for a simple and random sample of schools of unequal size (larger schools)	53
Table 3.7	School, within-school, and final probability of selection and corresponding weights for PPS sample of schools of unequal size	54
Table 3.8	Selection of schools according to a PPS and systematic procedure.....	55
<hr/>		
Table 4.1	Description of the 630 possible samples of 2 students selected from 36 students, according to their mean.....	59
Table 4.2	Distribution of all possible samples with a mean between 8.32 and 11.68.....	61
Table 4.3	Distribution of the mean of all possible samples of 4 students out of a population of 36 students.....	62
Table 4.4	Between-school and within-school variances on the mathematics scale in PISA 2003.....	65
Table 4.5	Current status of sampling errors.....	65
Table 4.6	Between-school and within-school variances, number of participating schools and students in Denmark and Germany in PISA 2003	66
Table 4.7	The Jackknifes replicates and sample means.....	68
Table 4.8	Values on variables X and Y for a sample of ten students.....	69
Table 4.9	Regression coefficients for each replicate sample.....	69
Table 4.10	The Jackknife replicates for unstratified two-stage sample designs.....	70
Table 4.11	The Jackknife replicates for stratified two-stage sample designs.....	71
Table 4.12	Replicates with the Balanced Repeated Replication method.....	72
Table 4.13	The Fay replicates	73
<hr/>		
Table 5.1	Probability of success when student ability equals item difficulty.....	82
Table 5.2	Probability of success when student ability is less than the item difficulty by 1 unit.....	82
Table 5.3	Probability of success when student ability is greater than the item difficulty by 1 unit	82
Table 5.4	Probability of success when student ability is less than the item difficulty by 2 units	83
Table 5.5	Probability of success when student ability is greater than the item difficulty by 2 units.....	83
Table 5.6	Possible response pattern for a test of four items.....	85
Table 5.7	Probability for the response pattern (1, 1, 0, 0) for three student abilities.....	85
Table 5.8	Probability for the response pattern (1, 0) for two students of different ability in an incomplete test design.....	89
Table 5.9	PISA 2003 test design	91



Table 6.1	Structure of the simulated data.....	98
Table 6.2	Means and variances for the latent variables and the different student ability estimators.....	98
Table 6.3	Percentiles for the latent variables and the different student ability estimators.....	99
Table 6.4	Correlation between HISEI, gender and the latent variable, the different student ability estimators.....	99
Table 6.5	Between- and within-school variances.....	100
<hr/>		
Table 7.1	HISEI mean estimates	105
Table 7.2	Squared differences between replicate estimates and the final estimate.....	106
Table 7.3	Output data file from Box 7.2.....	108
Table 7.4	Available statistics with the UNIVAR macro	109
Table 7.5	Output data file from Box 7.3.....	109
Table 7.6	Output data file from Box 7.4.....	110
Table 7.7	Percentage of girls for the final and replicate weights and squared differences.....	111
Table 7.8	Output data file from Box 7.5.....	112
Table 7.9	Output data file from Box 7.6.....	113
Table 7.10	Output data file from Box 7.7.....	114
Table 7.11	Output data file from Box 7.8.....	114
<hr/>		
Table 8.1	The 405 mean estimates.....	118
Table 8.2	Mean estimates and their respective sampling variances on the science scale for Belgium (PISA 2006).....	119
Table 8.3	Output data file from Box 8.2.....	121
Table 8.4	Output data file from Box 8.3.....	121
Table 8.5	The 450 regression coefficient estimates.....	123
Table 8.6	HISEI regression coefficient estimates and their respective sampling variance on the science scale in Belgium after accounting for gender (PISA 2006).....	123
Table 8.7	Output data file from Box 8.5.....	123
Table 8.8	Output data file from Box 8.6.....	124
Table 8.9	Correlation between the five plausible values for each domain, mathematics/quantity and mathematics/space and shape.....	126
Table 8.10	The five correlation estimates between mathematics/quantity and mathematics/space and shape and their respective sampling variance.....	127
Table 8.11	Standard deviations for mathematics scale using the correct method (plausible values) and by averaging the plausible values at the student level (pseudo-EAP) (PISA 2003).....	128
Table 8.12	Unbiased shortcut for a population estimate and its standard error	129
Table 8.13	Standard errors from the full and shortcut computation (PISA 2006).....	130
<hr/>		
Table 9.1	The 405 percentage estimates for a particular proficiency level	136
Table 9.2	Estimates and sampling variances per proficiency level in science for Germany (PISA 2006)	137
Table 9.3	Final estimates of the percentage of students, per proficiency level, in science and its standard errors for Germany (PISA 2006).....	137
Table 9.4	Output data file from Box 9.3.....	138
Table 9.5	Output data file from Box 9.4.....	138
Table 9.6	Mean estimates and standard errors for self-efficacy in mathematics per proficiency level (PISA 2003).....	141
Table 9.7	Output data file from Box 9.6.....	141



Table 10.1	Percentage of students per grade and ISCED level, by country (PISA 2006)	144
Table 10.2	Output data file from the first model in Box 10.3	148
Table 10.3	Output data file from the second model in Box 10.3	148
<hr/>		
Table 11.1	Output data file from Box 11.1	153
Table 11.2	Mean estimates for the final and 80 replicate weights by gender (PISA 2003)	153
Table 11.3	Difference in estimates for the final weight and 80 replicate weights between females and males (PISA 2003)	155
Table 11.4	Output data file from Box 11.2	156
Table 11.5	Output data file from Box 11.3	157
Table 11.6	Gender difference estimates and their respective sampling variances on the mathematics scale (PISA 2003)	157
Table 11.7	Output data file from Box 11.4	158
Table 11.8	Gender differences on the mathematics scale, unbiased standard errors and biased standard errors (PISA 2003)	159
Table 11.9	Gender differences in mean science performance and in standard deviation for science performance (PISA 2006)	159
Table 11.10	Regression coefficient of HISEI on the science performance for different models (PISA 2006)	160
Table 11.11	Cross tabulation of the different probabilities	161
<hr/>		
Table 12.1	Regression coefficients of the index of instrumental motivation in mathematics on mathematic performance in OECD countries (PISA 2003)	165
Table 12.2	Output data file from Box 12.1	166
Table 12.3	Output data file from Box 12.2	167
Table 12.4	Difference between the country mean scores in mathematics and the OECD total and average (PISA 2003)	170
<hr/>		
Table 13.1	Trend indicators between PISA 2000 and PISA 2003 for HISEI, by country	176
Table 13.2	Linking error estimates	178
Table 13.3	Mean performance in reading by gender in Germany	180
<hr/>		
Table 14.1	Distribution of the questionnaire index of cultural possession at home in Luxembourg (PISA 2006)	184
Table 14.2	Output data file from Box 14.1	186
Table 14.3	Labels used in a two-way table	186
Table 14.4	Distribution of 100 students by parents' marital status and grade repetition	187
Table 14.5	Probabilities by parents' marital status and grade repetition	187
Table 14.6	Relative risk for different cutpoints	187
Table 14.7	Output data file from Box 14.2	189
Table 14.8	Mean and standard deviation for the student performance in reading by gender, gender difference and effect size (PISA 2006)	191
Table 14.9	Output data file from the first model in Box 14.5	197
Table 14.10	Output data file from the second model in Box 14.5	197
Table 14.11	Mean of the residuals in mathematics performance for the bottom and top quarters of the PISA index of economic, social and cultural status, by school group (PISA 2003)	195

Table 15.1	Between- and within-school variance estimates and intraclass correlation (PISA 2006).....	204
Table 15.2	Fixed parameter estimates	211
Table 15.3	Variance/covariance estimates before and after centering.....	213
Table 15.4	Output data file of the fixed parameters file.....	215
Table 15.5	Average performance and percentage of students by student immigrant status and by type of school.....	216
Table 15.6	Variables for the four groups of students	216
Table 15.7	Comparison of the regression coefficient estimates and their standard errors in Belgium (PISA 2006).....	218
Table 15.8	Comparison of the variance estimates and their respective standard errors in Belgium (PISA 2006)	218
Table 15.9	Three-level regression analyses.....	220
<hr/>		
Table 16.1	Differences between males and females in the standard deviation of student performance (PISA 2000).....	226
Table 16.2	Distribution of the gender differences (males – females) in the standard deviation of the student performance	226
Table 16.3	Gender difference on the PISA combined reading scale for the 5 th , 10 th , 90 th and 95 th percentiles (PISA 2000)	227
Table 16.4	Gender difference in the standard deviation for the two different item format scales in reading (PISA 2000)	228
Table 16.5	Random and fixed parameters in the multilevel model with student and school socio-economic background.....	229
Table 16.6	Random and fixed parameters in the multilevel model with socio-economic background and grade retention at the student and school levels	233
Table 16.7	Segregation indices and correlation coefficients by country (PISA 2000).....	234
Table 16.8	Segregation indices and correlation coefficients by country (PISA 2006).....	235
Table 16.9	Country correlations (PISA 2000).....	237
Table 16.10	Country correlations (PISA 2006).....	237
<hr/>		
Table 17.1	Synthesis of the 19 SPSS® macros.....	241
<hr/>		
Table A2.1	Cluster rotation design used to form test booklets for PISA 2006	332
<hr/>		
Table A12.1	Mapping of ISCED to accumulated years of education	457
Table A12.2	ISCO major group white-collar/blue-collar classification	459
Table A12.3	ISCO occupation categories classified as science-related occupations	459
Table A12.4	Household possessions and home background indices.....	463
Table A12.5	Factor loadings and internal consistency of ESCS 2006 in OECD countries.....	473
Table A12.6	Factor loadings and internal consistency of ESCS 2006 in partner countries/economies.....	474



User's Guide

Preparation of data files

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

SPSS® users

By running the SPSS® control files, the PISA data files are created in the SPSS® format. Before starting analysis in the following chapters, save the PISA 2000 data files in the folder of "c:\pisa2000\data\", the PISA 2003 data files in "c:\pisa2003\data\", and the PISA 2006 data files in "c:\pisa2006\data\".

SPSS® syntax and macros

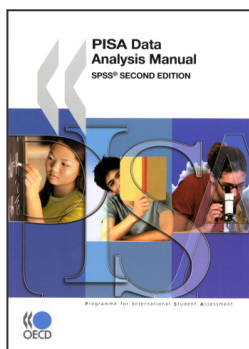
All syntaxes and macros in this manual can be copied from the PISA website (www.pisa.oecd.org). These macros were developed for SPSS 17.0. The 19 SPSS® macros presented in Chapter 17 need to be saved under "c:\pisa\macro\", before starting analysis. Each chapter of the manual contains a complete set 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



From:
PISA Data Analysis Manual: SPSS, Second Edition

Access the complete publication at:
<https://doi.org/10.1787/9789264056275-en>

Please cite this chapter as:

OECD (2009), "Plausible Values", in *PISA Data Analysis Manual: SPSS, Second Edition*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264056275-7-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.