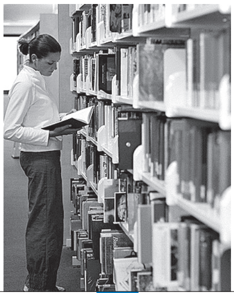
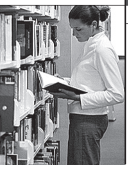


8

Analyses with Plausible Values

PISA DATA ANALYSIS MANUAL: SPSS® SECOND EDITION – ISBN 978-92-64-05626-8 –



INTRODUCTION

As described in Chapters 5 and 6, the cognitive data in PISA are scaled with the Rasch Model and the performance of students is denoted with plausible values (PVs). For minor domains, only one scale is included in the international databases. For major domains,[1] a combined scale and several subscales are provided. For each scale and subscale, five plausible values per student are included in the international databases. This chapter describes how to perform analyses with plausible values.

Since plausible values are mainly used for reporting student performance on cognitive tests, this chapter is mainly useful when conducting analyses on achievement data and their relationships with student or school characteristics.

UNIVARIATE STATISTICS ON PLAUSIBLE VALUES

The computation of a statistic with plausible values always consists of six steps, regardless of the required statistic.

1. The required statistic and its respective standard error have to be computed for each plausible value. In Chapter 7, it was mentioned that 81 estimates were necessary to get the final estimate and its standard error. Therefore, any analysis that involves five plausible values will require 405 estimates. If a mean needs to be estimated, then 405 means will be calculated. The means estimated with the final weights are denoted $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_3$, $\hat{\mu}_4$ and $\hat{\mu}_5$. From the 80 replicates applied on each of the five plausible values, five sampling variances are estimated, denoted respectively $\sigma^2_{(\hat{\mu}_1)}$, $\sigma^2_{(\hat{\mu}_2)}$, $\sigma^2_{(\hat{\mu}_3)}$, $\sigma^2_{(\hat{\mu}_4)}$ and $\sigma^2_{(\hat{\mu}_5)}$. These five mean estimates and their respective sampling variances are provided in Table 8.1.

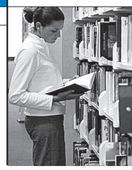
Table 8.1
The 405 mean estimates

Weight	PV1	PV2	PV3	PV4	PV5
Final	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\mu}_5$
Replicate 1	$\hat{\mu}_{1\_1}$	$\hat{\mu}_{2\_1}$	$\hat{\mu}_{3\_1}$	$\hat{\mu}_{4\_1}$	$\hat{\mu}_{5\_1}$
Replicate 2	$\hat{\mu}_{1\_2}$	$\hat{\mu}_{2\_2}$	$\hat{\mu}_{3\_2}$	$\hat{\mu}_{4\_2}$	$\hat{\mu}_{5\_2}$
Replicate 3	$\hat{\mu}_{1\_3}$	$\hat{\mu}_{2\_3}$	$\hat{\mu}_{3\_3}$	$\hat{\mu}_{4\_3}$	$\hat{\mu}_{5\_3}$
....	....	....	....	....	....
....	....	....	....	....	....
Replicate 80	$\hat{\mu}_{1\_80}$	$\hat{\mu}_{2\_80}$	$\hat{\mu}_{3\_80}$	$\hat{\mu}_{4\_80}$	$\hat{\mu}_{5\_80}$
Sampling variance	$\sigma^2_{(\hat{\mu}_1)}$	$\sigma^2_{(\hat{\mu}_2)}$	$\sigma^2_{(\hat{\mu}_3)}$	$\sigma^2_{(\hat{\mu}_4)}$	$\sigma^2_{(\hat{\mu}_5)}$

2. The final mean estimate is equal to the average of the five mean estimates, *i.e.* $\hat{\mu} = \frac{1}{5}(\hat{\mu}_1 + \hat{\mu}_2 + \hat{\mu}_3 + \hat{\mu}_4 + \hat{\mu}_5)$.

3. The final sampling variance is equal to the average of the five sampling variances,
   i.e. $\sigma^2_{(\hat{\mu})} = \frac{1}{5}(\sigma^2_{(\hat{\mu}_1)} + \sigma^2_{(\hat{\mu}_2)} + \sigma^2_{(\hat{\mu}_3)} + \sigma^2_{(\hat{\mu}_4)} + \sigma^2_{(\hat{\mu}_5)})$

4. The imputation variance, also denoted measurement error variance, is computed as $\sigma^2_{(test)} = \frac{1}{4}\sum_{i=1}^{5}(\hat{\mu}_i - \hat{\mu})^2$. Indeed, as PISA returns five plausible values per scale, then $\sigma^2_{(test)} = \frac{1}{M-1}\sum_{i=1}^{M}(\hat{\mu}_i - \hat{\mu})^2 = \frac{1}{4}\sum_{i=1}^{5}(\hat{\mu}_i - \hat{\mu})^2$ with M being the number of plausible values. This formula is similar to the one used for the estimation of a population variance, except that in this particular case, observations are not compared with the population mean, but each PV mean is compared with the final mean estimate.



5. The sampling variance and the imputation variance are combined to obtain the final error variance as

$$\sigma^2_{(error)} = \sigma^2_{(\hat{\mu})} + \left(1.2\sigma^2_{(test)}\right)$$

Indeed, $\sigma^2_{(error)} = \sigma^2_{(\hat{\mu})} + \left(\left(1+\frac{1}{M}\right)\sigma^2_{(test)}\right) = \sigma^2_{(\hat{\mu})} + \left(\left(1+\frac{1}{5}\right)\sigma^2_{(test)}\right) = \sigma^2_{(\hat{\mu})} + \left((1.2)\sigma^2_{(test)}\right)$

6. The standard error is equal to the square root of the error variance.

The mean estimate on the science scale and its respective standard error for the PISA 2006 Belgian data can be computed. The macro described in Chapter 7 and labelled MCR_SE_UNIV can be sequentially used five times, and the results can be combined in an Excel® spreadsheet. Table 8.2 presents the different PV means, their respective sampling variances, as well as the mean estimates on the first and last replicates.

Table 8.2

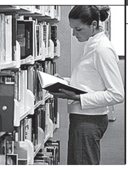
Mean estimates and their respective sampling variances on the science scale for Belgium (PISA 2006)

Weight	PV1	PV2	PV3	PV4	PV5
Final	510.18	510.58	510.36	510.62	510.09
Replicate 1	509.06	509.63	509.64	509.72	509.03
....	....	....	....	....	....
Replicate 80	511.38	512.03	511.63	512	511.28
Sampling variance	$(2.47)^2$	$(2.42)^2$	$(2.50)^2$	$(2.45)^2$	$(2.52)^2$

Box 8.1 presents the SPSS® syntax for running sequentially the MCR_SE_UNIV macro described in Chapter 7.

Box 8.1 **SPSS® syntax for computing the mean on the science scale by using the MCR_SE_UNIV macro (*e.g.* PISA 2006)**

```
GET FILE=“C:\PISA\2006\DATA\INT_STU06_DEC07.SAV”.

SELECT IF CNT=“BEL”.
SAVE OUTFILE=“C:\TEMP\BEL.SAV”.

INSERT FILE=“C:\PISA\MACRO\MCR_SE_UNIV.SPS”.

SET MPRINT=YES.

UNIVAR  STAT=MEAN/ DEP=PV1SCIE/ PSU=SCHOOLID/ LIMIT_CRITERIA=50 5 3 1/
INFILE=“C:\TEMP\BEL.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS1.SAV”.
UNIVAR  STAT=MEAN/ DEP=PV2SCIE/ PSU=SCHOOLID/ LIMIT_CRITERIA=50 5 3 1/
INFILE=“C:\TEMP\BEL.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS2.SAV”.
UNIVAR  STAT=MEAN/ DEP=PV3SCIE/ PSU=SCHOOLID/ LIMIT_CRITERIA=50 5 3 1/
INFILE=“C:\TEMP\BEL.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS3.SAV”.
UNIVAR  STAT=MEAN/ DEP=PV4SCIE/ PSU=SCHOOLID/ LIMIT_CRITERIA=50 5 3 1/
INFILE=“C:\TEMP\BEL.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS4.SAV”.
UNIVAR  STAT=MEAN/ DEP=PV5SCIE/ PSU=SCHOOLID/ LIMIT_CRITERIA=50 5 3 1/
INFILE=“C:\TEMP\BEL.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS5.SAV”.
```



The final mean estimate for Belgium on the combined science scale is computed as:

$$\hat{\mu} = \frac{1}{5}(\hat{\mu}_1 + \hat{\mu}_2 + \hat{\mu}_3 + \hat{\mu}_4 + \hat{\mu}_5),\ i.e.\ \hat{\mu} = \frac{(510.18 + 510.58 + 510.36 + 510.62 + 510.09)}{5} = 510.4$$

The final sampling variance on the mean estimate for the combined science literacy scale is equal to:

$$\sigma^2_{(\hat{\mu})} = \frac{1}{5}(\sigma^2_{(\hat{\mu}_1)} + \sigma^2_{(\hat{\mu}_2)} + \sigma^2_{(\hat{\mu}_3)} + \sigma^2_{(\hat{\mu}_4)} + \sigma^2_{(\hat{\mu}_5)}),\ i.e.\ \sigma^2_{(\hat{\mu})} = \frac{(2.47)^2 + (2.42)^2 + (2.50)^2 + (2.45)^2 + (2.52)^2}{5} = 6.11$$

The imputation variance is equal to:

$$\sigma^2_{(test)} = \frac{1}{4}\sum_{i=1}^{5}(\hat{\mu}_i - \hat{\mu})^2,\ i.e.\ \sigma^2_{(test)} = \frac{\left[(510.18 - 510.4)^2 + (510.58 - 510.4)^2 + \ldots + (510.09 - 510.4)^2\right]}{4} = 0.055$$

The final error variance is equal to:

$$\sigma^2_{(error)} = \sigma^2_{(\hat{\mu})} + \left(1.2\sigma^2_{(test)}\right),\ i.e.\ \sigma^2_{(error)} = 6.11 + (1.2 * 0.055) = 6.17$$

The final standard error is therefore equal to:

$$SE = \sqrt{\sigma^2_{(error)}} = \sqrt{6.17} = 2.48$$

Sequentially running the MCR_SE_UNIV macro five times and combining the results can be avoided: a SPSS® macro has been developed for dealing with plausible values (see Box 8.2). This macro also computes the:

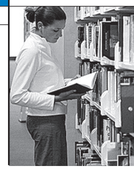
- five mean estimates,
- five sampling variances,
- imputation variance,
- final standard error by combining the final sampling variance and the imputation variance.

Box 8.2 **SPSS® syntax for computing the mean and its standard error on PVs (*e.g.* PISA 2006)**

```
INSERT FILE="C:\PISA\MACRO\MCR_SE_PV.SPS".

SET MPRINT=YES.

PV    STAT=MEAN/
      DEP=PV1SCIE PV2SCIE PV3SCIE PV4SCIE PV5SCIE/
      INFILE="C:\TEMP\BEL.SAV"/
      GRP=CNT/                                     /* default=NOGRP */
      PSU = SCHOOLID/                              /* is default */
      LIMIT_CRITERIA=100 10 5 1/                   /* default=0 */
      WGT=W_FSTUWT/                                /* is default */
      RWGT=W_FSTR/                                 /* is default */
      CONS=0.05/.                                  /* is default */
```

Besides the eight arguments common to most SPSS® macros previously described in this manual, *i.e.* *(i)* `NREP=`; *(ii)* `GRP=`; *(iii)* `WGT=`; *(iv)* `RWGT=`; *(v)* `CONS=`; *(vi)* `LIMIT_CRITERIA=`; *(vii)* `PSU=`, and *(viii)* `INFILE=`, two arguments need to be specified: the dependent variables (`DEP=`) and the requested statistic (`STAT=`) :

- The name of the 5 variables with the plausible values needs to be listed after the DEP argument.
- The STAT argument to specify the requested statistics. Available statistics have been described in Chapter 7.



The structure of the output data file is presented in Table 8.3.

Table 8.3
Output data file from Box 8.2

CNT	Statistic	STAT	SE	FLAG_STD	FLAG_SCH	FLAG_PCT
BEL	MEAN	510.4	2.48	0	0	0

Similar to the SPSS® macros described in the previous chapter, more than one breakdown variable can be used. For instance, if one wants to determine whether the dispersion of the science performance is larger for females than for males, the macro MCR_SE_PV can be used as shown in Box 8.3.

Box 8.3 **SPSS® syntax for computing the standard deviation and its standard error on PVs by gender (*e.g.* PISA 2006)**

```
PV   STAT=STDDEV/
     DEP=PV1SCIE PV2SCIE PV3SCIE PV4SCIE PV5SCIE/
     INFILE="C:\TEMP\BEL.SAV"/
     GRP=CNT ST04Q01/
     PSU = SCHOOLID/
     LIMIT_CRITERIA=100 10 5 1/.
```

The structure of the output data file is presented in Table 8.4.

Table 8.4
Output data file from Box 8.3

CNT	ST04Q01	Statistic	STAT	SE	FLAG_STD	FLAG_SCH	FLAG_PCT
BEL	1	STDDEV	96.1	2.23	0	0	0
BEL	2	STDDEV	102.9	2.50	0	0	0

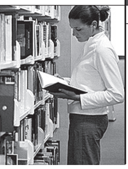
According to Table 8.4, the standard deviation (STAT) is larger for males than for females. Unfortunately, the information of the two standard errors in SE is not enough to conduct a test of the equality for these two standard deviation coefficients since the standard deviation estimates for males and females may be correlated. The detail of the standard error on difference will be presented in Chapter 11.

THE STANDARD ERROR ON PERCENTAGES WITH PVS

The MCR_SE_GRPPCT first presented in Chapter 7 was developed for the computation of percentages and their respective standard errors. Chapter 9 will deal with the application of this macro to plausible values: an entire chapter needs to be devoted to this type of analysis because of the issues involved.

THE STANDARD ERROR ON REGRESSION COEFFICIENTS WITH PVS

Suppose that the statistical effect of gender and student socio-economic background on the performance in science needs to be estimated. Just like estimating a mean, this question can be solved by sequentially applying five times the MCR_SE_REG macro described in Chapter 7. Box 8.4 presents the SPSS® syntax for such an approach.



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)**

```
GET FILE=“C:\PISA\2006\DATA\INT_STU06_DEC07.SAV”.
SELECT IF CNT=“BEL”.
RECODE ST04Q01 (1=1)(2=0) INT GENDER.
SAVE OUTFILE=“C:\TEMP\BEL2.SAV”.

INSERT FILE=“C:\PISA\MACRO\MCR_SE_REG.SPS”.

REGnoPV IND=GENDER HISEI/ DEP=PV1SCIE/ PSU=SCHOOLID/
LIMIT_CRITERIA=100 10 5 1/ INFILE=“C:\TEMP\BEL2.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS6.SAV”.
REGnoPV IND=GENDER HISEI/ DEP=PV2SCIE/ PSU=SCHOOLID/
LIMIT_CRITERIA=100 10 5 1/ INFILE=“C:\TEMP\BEL2.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS7.SAV”.
REGnoPV IND=GENDER HISEI/ DEP=PV3SCIE/ PSU=SCHOOLID/
LIMIT_CRITERIA=100 10 5 1/ INFILE=“C:\TEMP\BEL2.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS8.SAV”.
REGnoPV IND=GENDER HISEI/ DEP=PV4SCIE/ PSU=SCHOOLID/
LIMIT_CRITERIA=100 10 5 1/ INFILE=“C:\TEMP\BEL2.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS9.SAV”.
REGnoPV IND=GENDER HISEI/ DEP=PV5SCIE/ PSU=SCHOOLID/
LIMIT_CRITERIA=100 10 5 1/ INFILE=“C:\TEMP\BEL2.SAV”/.
SAVE OUTFILE=“C:\TEMP\RESULTS10.SAV”.
```

Just like the computation of a mean and its standard error, the computation of regression coefficients and their respective standard errors consists of six steps:

1. For each plausible value and for each explanatory variable, regression coefficients are computed with the final and the 80 replicate weights. Thus, 405 regression coefficients per explanatory variable will be computed. The MCR_SE_REG macro applied sequentially five times will return, per explanatory variable, five estimates, denoted, $\hat{\beta}_1, \ldots, \hat{\beta}_5$ and five standard errors, denoted $\sigma_{(\hat{\beta}_1)}, \ldots, \sigma_{(\hat{\beta}_5)}$. Table 8.5 presents the mathematical expression for these 405 estimates and Table 8.6 presents some of the values for the 405 regression coefficients obtained on the Belgian data for the HISEI (international socio-economic index of occupational status) variable.

2. The final regression coefficient estimate is equal to $\hat{\beta} = \frac{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_5}{5}$

   i.e. for HISEI $\hat{\beta} = \frac{2.22+2.24+2.25+2.24+2.26}{5} = 2.24$

3. The final sampling variance estimate is equal to $\sigma^2_{(\hat{\beta})} = \frac{1}{5}\left(\sigma^2_{(\hat{\beta}_1)} + \sigma^2_{(\hat{\beta}_2)} + \sigma^2_{(\hat{\beta}_3)} + \sigma^2_{(\hat{\beta}_4)} + \sigma^2_{(\hat{\beta}_5)}\right)$

   i.e. for HISEI $\sigma^2_{(\hat{\beta})} = \frac{(0.09)^2 + (0.09)^2 + (0.09)^2 + (0.09)^2 + (0.09)^2}{5} = 0.0081$

4. The imputation variance is equal to $\sigma^2_{(test)} = \frac{1}{4}\sum_{i=1}^{5}(\hat{\beta}_i - \hat{\beta})^2$

   i.e. for HISEI $\sigma^2_{(test)} = \frac{(2.22-2.24)^2 + (2.24-2.24)^2 + \ldots + (2.26-2.24)^2}{4} = 0.0002$

5. The final error variance is equal to $\sigma^2_{(error)} = \sigma^2_{(\hat{\beta})} + (1.2\sigma^2_{(test)})$

   i.e. for HISEI $\sigma^2_{(error)} = 0.0081 + (1.2*0.0002) = 0.0084$

6. The final standard error is equal to $SE = \sqrt{\sigma^2_{(error)}} = \sqrt{0.0084} = 0.10$

As 2.24 divided by 0.10 is about 22.4, the regression coefficient for HISEI is significantly different from 0.

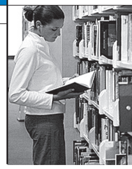


Table 8.5
The 450 regression coefficient estimates

Weight	PV1	PV2	PV3	PV4	PV5
Final	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Replicate 1	$\hat{\beta}_{1\_1}$	$\hat{\beta}_{2\_1}$	$\hat{\beta}_{3\_1}$	$\hat{\beta}_{4\_1}$	$\hat{\beta}_{5\_1}$
Replicate 2	$\hat{\beta}_{1\_2}$	$\hat{\beta}_{2\_2}$	$\hat{\beta}_{3\_2}$	$\hat{\beta}_{4\_2}$	$\hat{\beta}_{5\_2}$
Replicate 3	$\hat{\beta}_{1\_3}$	$\hat{\beta}_{2\_3}$	$\hat{\beta}_{3\_3}$	$\hat{\beta}_{4\_3}$	$\hat{\beta}_{5\_3}$
....	....	....	....	....	....
....	....	....	....	....	....
Replicate 80	$\hat{\beta}_{1\_80}$	$\hat{\beta}_{2\_80}$	$\hat{\beta}_{3\_80}$	$\hat{\beta}_{4\_80}$	$\hat{\beta}_{5\_80}$
Sampling variance	$\sigma^2_{(\hat{\beta}_1)}$	$\sigma^2_{(\hat{\beta}_2)}$	$\sigma^2_{(\hat{\beta}_3)}$	$\sigma^2_{(\hat{\beta}_4)}$	$\sigma^2_{(\hat{\beta}_5)}$

Table 8.6
HISEI regression coefficient estimates and their respective sampling variance on the science scale in Belgium after accounting for gender (PISA 2006)

Weight	PV1	PV2	PV3	PV4	PV5
Final	2.22	2.24	2.25	2.24	2.26
Replicate 1	2.22	2.22	2.24	2.23	2.27
....	....	....	....	....	....
Replicate 80	2.14	2.16	2.18	2.18	2.18
Sampling variance	$(0.09)^2$	$(0.09)^2$	$(0.09)^2$	$(0.09)^2$	$(0.09)^2$

A SPSS® macro MCR_SE_REG_PV has also been developed for regression analyses with plausible values as dependent variables. The SPSS® syntax is presented in Box 8.5.

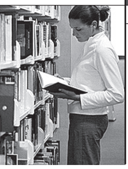
Box 8.5 **SPSS® syntax for running the simple linear regression macro with PVs (*e.g.* PISA 2006)**

```
INSERT FILE="C:\PISA\MACRO\MCR_SE_REG_PV.SPS".

SET MPRINT=YES.

REG_PV INFILE="C:\TEMP\BEL2.SAV"/
       IND=HISEI GENDER/
       DEP=PV1SCIE PV2SCIE PV3SCIE PV4SCIE PV5SCIE/
       GRP=CNT/
       NREP=80/
       WGT=W_FSTUWT/
       RWGT=W_FSTR/
       CONS=0.05 /
       PSU=SCHOOLID/
       LIMIT_CRITERIA=100 10 5 1/.
```

Besides the arguments common to most macros, the name of the 5 plausible value variables has to be specified, in addition to the list of independent variables. The structure of the output data file is presented in Table 8.7.

Table 8.7
Output data file from Box 8.5

CNT	Ind	STAT	SE	FLAG_STD	FLAG_SCH	FLAG_PCT
BEL	INTERCEPT	403.39	5.74	0	0	0
BEL	HISEI	2.24	0.10	0	0	0
BEL	GENDER	0.10	3.24	0	0	0
BEL	R_SQUARE	0.15	0.01	0	0	0



A quick overview of these results shows that the intercept and the regression coefficient for HISEI are significantly different from 0. However, the regression coefficient for GENDER does not statistically differ from 0.

Similar to the MCR_SE_REG macro, the MCR_SE_REG_PV macro also returns the outcomes of the sampling size requirement analysis in a separate file. The file name consists of: *(i)* the name of the output files that contains the regression parameters; and *(ii)* the name of the output file followed with _CRITERIA.

THE STANDARD ERROR ON CORRELATION COEFFICIENTS WITH PVS

A SPSS® macro has also been developed to compute the correlation between a set of plausible values and another variable. The SPSS syntax for running this macro is presented in Box 8.6 and the structure of the output data file is presented in Table 8.8.

Box 8.6 **SPSS® syntax for running the correlation macro with PVs (*e.g.* PISA 2006)**

```
INSERT FILE="C:\PISA\MACRO\MCR_SE_COR_1PV.SPS".

COR_1PV NOPV=HISEI/
        PV=PV1SCIE PV2SCIE PV3SCIE PV4SCIE PV5SCIE /
        INFILE="C:\TEMP\BEL2.sav"/
        GRP=CNT/
        LIMIT_CRITERIA=100 10 5 1/
        PSU = SCHOOLID /
        WGT=W_FSTUWT/
        RWGT=W_FSTR/
        NREP=80/
        CONS=0.05/.
```

Table 8.8
Output data file from Box 8.6

CNT	STAT	SE	FLAG_STD	FLAG_SCH	FLAG_PCT
BEL	0.38	0.02	0	0	0

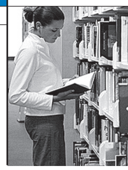
CORRELATION BETWEEN TWO SETS OF PLAUSIBLE VALUES

Some researchers may be interested in the correlation between the different PISA domains and subdomains. For instance, some might want to compute the correlation between the reading subdomains or between the mathematics subdomains, or between reading and mathematics using the PISA 2000, PISA 2003 and PISA 2006 databases.

As described in Chapter 5, the PISA assessment used incomplete assessment designs, *i.e.* students are required to answer a subset of the item battery. Further, while all students were assessed in the major domain, only a subset of students was assessed in minor domains.

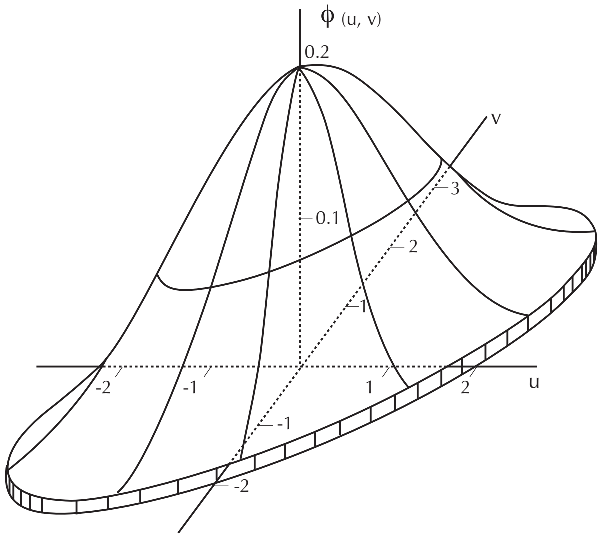
PISA 2000 included plausible values of a minor domain only for students who answered questions for that minor domain. Therefore, using the PISA 2000 database to compute the correlation between reading and mathematics, for example, would require working on a subset of students.[2]

To facilitate secondary analyses, PISA 2003 returned plausible values for all domains and for all students, regardless of whether they were actually assessed or not. Ignoring the assessment status is possible, because the cognitive data in PISA are scaled according to multi-dimensional models.



Since this is easier to illustrate graphically, let's suppose that only two domains were assessed, *i.e.* mathematics/quantity and mathematics/space and shape. If the mathematics/quantity and mathematics/space and shape materials were scaled independently, the correlation between the two subdomains would be largely underestimated. In order to avoid this problem, both materials are scaled together. The model will build a two-dimensional posterior distribution, instead of two one-dimensional posterior distributions as described in Chapter 6. Figure 8.1 graphically presents a two-dimensional normal distribution.

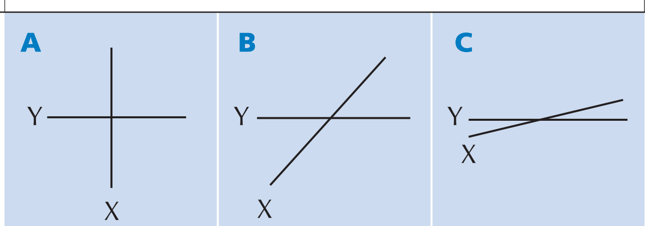
Figure 8.1
A two-dimensional distribution

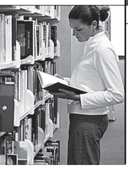


To correctly describe such distributions, two means, two variances, and one correlation are needed. If the correlation is equal to 0, then the two axes will be orthogonal. As the absolute value of the correlation starts to increase, the angle formed by the two axes becomes less than 90 degrees (see Figure 8.2, part A).[3] Two axes perfectly overlapping would represent a correlation of 1.0 (or –1.0). These different cases are illustrated in Figure 8.2.

With a two-dimensional model, the first plausible value for mathematics/quantity will be drawn at the same time as the first plausible value for mathematics/space and shape. Per student, this will consist of randomly drawing one dot in the scatter plot. The values of the two plausible values will be the coordinates of the dot on the two axes. The same procedure is applied for the second, third, fourth, and fifth plausible values.

Figure 8.2
Axes for two-dimensional normal distributions





As the PISA domains and subdomains highly correlate, as shown by the graph on the far right in Figure 8.2, it is very unlikely for a student to get a high score for the first plausible value in mathematics/quantity (PV1MATH4) and a low score for the first plausible value in mathematics/space and shape (PV1MATH1). If plausible values were drawn independently for these two mathematics subdomains, such a case would be possible and therefore the correlation would be underestimated.

Since each draw is independent, to calculate the correlation between the two domains, the correlation between each set of plausible value, *i.e.* correlation between two plausible values with the same number, needs to be computed. In the PISA 2003 example of correlation between PV1MATH1 to PV5MATH1 and PV1MATH4 to PV5MATH4, the following values need to be computed:

- PV1MATH1 and PV1MATH4,
- PV2MATH1 and PV2MATH4,
- PV3MATH1 and PV3MATH4,
- PV4MATH1 and PV4MATH4,
- PV5MATH1 and PV5 MATH4.

Table 8.9 presents all possible 25 correlation coefficients between the 5 plausible values in mathematics/quantity (MATH4) and mathematics/space and shape (MATH1), respectively, for Germany in PISA 2003. As expected, the correlation coefficients on the diagonal of the square matrix are substantially higher than the other correlation coefficients. The final correlation estimate between these two mathematics subdomains will be the average of the five correlation coefficients on the diagonal.

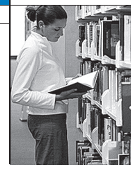
Table 8.9
Correlation between the five plausible values for each domain, mathematics/quantity and mathematics/space and shape

	PV1MATH1	PV2MATH1	PV3MATH1	PV4MATH1	PV5MATH1
PV1MATH4	0.90	0.83	0.84	0.84	0.83
PV2MATH4	0.83	0.90	0.84	0.84	0.83
PV3MATH4	0.84	0.83	0.90	0.84	0.83
PV4MATH4	0.83	0.83	0.84	0.90	0.83
PV5MATH4	0.83	0.83	0.84	0.84	0.90

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)**

```
GET FILE="C:\PISA\2003\DATA\INT_STUI_2003.SAV".
SELECT IF CNT="DEU".
SAVE OUTFILE="C:\TEMP\DEU.SAV".

INSERT FILE="C:\PISA\MACRO\MCR_SE_COR_2PV.SPS".

COR_2PV PV1=PV1MATH1 PV2MATH1 PV3MATH1 PV4MATH1 PV5MATH1/
        PV2=PV1MATH4 PV2MATH4 PV3MATH4 PV4MATH4 PV5MATH4/
        INFILE="C:\TEMP\DEU.SAV"/
        GRP=CNT/
        LIMIT_CRITERIA=100 10 5 1/
        PSU=SCHOOLID/
        WGT=W_FSTUWT/
        RWGT=W_FSTR/
        NREP=80/
        CONS=0.05/.
```


PISA DATA ANALYSIS MANUAL: SPSS® SECOND EDITION – ISBN 978-92-64-05626-8 – © OECD 2009



The standard error on this correlation estimate can be easily obtained by using the SPSS® macro denoted MCR_SE_COR_2PV especially designed for the computation of the correlation between two sets of plausible values. The SPSS® syntax is given in Box 8.7 and the output data file is presented in Table 8.10.

Table 8.10
The correlation estimate between mathematics/quantity and mathematics/space and shape and their respective sampling variance

CNT	STAT	SE	FLAG_STD	FLAG_SCH	FLAG_PCT
DEU	0.897	0.004132	0	0	0

The computation of the correlation between two domains or between a subdomain and a domain might be problematic in some cases in the PISA databases. PISA 2000 used two scaling models:

- a three-dimensional model with mathematics, reading and science;
- a five-dimensional model with mathematics, reading/retrieving information, reading/interpreting texts, reading/reflection and evaluation, and science.

PISA 2003 also used two scaling models:

- a four-dimensional model with mathematics, problem solving, reading and science;
- a seven-dimensional model with mathematics/space and shape, mathematics/change and relationships, mathematics/uncertainty, mathematics/quantity, problem solving, reading and science.

PISA 2006 used two scaling models as well:

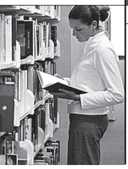
- a five-dimensional model with mathematics, reading, science and the two attitudinal scales;
- a five-dimensional model with mathematics, reading, and the three science scales (the identifying scientific issues scale, the explaining phenomena scientifically scale, and the using scientific evidence scale).

The PISA databases should contain two sets of plausible values for each of the minor domains. As this would be too confusing, only one set was provided. Therefore, the correlation coefficients are underestimated.

This can be confirmed by examining the data. In the case of a minor domain and a subscale of the major domain, the correlation coefficients on the diagonal do not differ from the other correlations, since these two sets of plausible values were generated by two different models.

In PISA 2006, as well as in PISA 2000 and in PISA 2003, the plausible values for the minor domains included in the databases were generated with the major domain as a combined scale. This means that:

- The correlation between a minor domain and the combined scale of the major domain can be computed.
- The correlation between two minor domains can be computed.
- The correlation between the subdomains can be computed.
- It is not possible to compute the correlation between minor domains and one of the subscales of the major domain.



A FATAL ERROR SHORTCUT

A common fatal error when analysing with plausible values involves computing the mean of the five plausible values, before further analysis.

In Chapter 6, the expected *a posteriori* (EAP) student performance estimator was described. As a reminder, the EAP estimator is equal to the mean of the posterior distribution. Therefore, computing the mean of the five plausible values at the student level is more or less equal to the EAP estimate.

In Chapter 6, the efficiency of the EAP estimator was also compared with the weighted likelihood estimate (WLE) and the plausible values for some statistics estimations. It was indicated that the EAP estimator:

- underestimates the standard deviation,
- overestimates the correlation between the student performance and some background variables,
- underestimates the within-school variance.

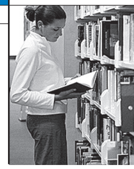
Therefore, computing the mean of the five plausible values and then computing statistics on this new score would bias the results just as the EAP does. Table 8.11 provides, per country, the standard deviation of the combined mathematics scale in PISA 2003 using the correct method, as described in this chapter, and the incorrect method of averaging the five plausible values at the student level and then computing the standard deviation on this new score. The result of the latter is denoted as pseudo-EAP.

As shown by Table 8.11, the pseudo-EAP underestimates the standard deviation.

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)

	Plausible values	Pseudo-EAP
AUS	95.42	91.90
AUT	93.09	89.91
BEL	109.88	106.65
CAN	87.11	83.37
CHE	98.38	94.97
CZE	95.94	92.50
DEU	102.59	99.54
DNK	91.32	87.52
ESP	88.47	84.52
FIN	83.68	79.77
FRA	91.70	88.07
GBR	92.26	89.18
GRC	93.83	89.49
HUN	93.51	89.71
IRL	85.26	82.03
ISL	90.36	86.55
ITA	95.69	92.00
JPN	100.54	96.96
KOR	92.38	89.07
LUX	91.86	88.28
MEX	85.44	80.52
NLD	92.52	89.89
NOR	92.04	88.31
NZL	98.29	95.07
POL	90.24	86.49
PRT	87.63	83.91
SVK	93.31	89.86
SWE	94.75	91.07
TUR	104.74	100.79
USA	95.25	92.12



The analysis process should always aggregate the results of the five plausible values at the latest stage, *i.e.* the statistic that has to be reported is computed five times, then these five statistics are combined.

AN UNBIASED SHORTCUT

Table 8.1 and Table 8.5 respectively give the 405 mean and regression coefficient estimates needed for computing a mean or regression coefficient final estimate and the respective standard errors.

On average, analysing one plausible value instead of five plausible values provides unbiased population estimates as well as unbiased sampling variances on these estimates. It will not be possible to estimate the imputation variance using this method, however.

Therefore, an unbiased shortcut consists of:

- computing, using one of the five plausible values, the statistical estimate and its sampling variance by using the final student weight as well as the 80 replicate weights;
- computing the statistical estimate by using the final student weight on the four other plausible values;
- computing the final statistical estimate by averaging the plausible value statistical estimates;
- computing the imputation variance, as previously described;
- combining the imputation variance and the sampling variance, as previously described.

This unbiased shortcut is presented in Table 8.12 for the estimation of a mean and its standard error. This shortcut only requires the computation of 85 estimates instead of 405. The final estimate of this shortcut will be equal to the one obtained with the long procedure, but the standard error might differ slightly.

Table 8.12
Unbiased shortcut for a population estimate and its standard error

Weight	PV1	PV2	PV3	PV4	PV5
Final	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\mu}_5$
Replicate 1	μ_{1_1}				
Replicate 2	μ_{1_2}				
Replicate 3	μ_{1_3}				
....	....				
....	....				
Replicate 80	μ_{1_80}				
Sampling variance	$\sigma^2_{(\hat{\mu}_1)}$				

Table 8.13 presents the standard errors for the country mean estimate in student performance on the science scale (PISA 2006) as well as the immigrant subpopulations. Two standard errors are provided per estimate: the standard error that results from the strict application of the recommended procedure; and the standard errors using the described shortcut. In all or nearly all cases, the difference between the two standard error estimates is negligible, even based on less than 50 students.

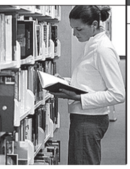


Table 8.13
Standard errors from the full and shortcut computation (PISA 2006)

	Mean estimate in science		Immigrant students mean estimate in science		
	Full computation	Shortcut computation	Full computation	Shortcut computation	
	S.E.	S.E.	S.E.	S.E.	Student flag (50)
AUS	2.26	2.21	5.75	5.77	0
AUT	3.92	3.98	10.92	11.00	0
BEL	2.48	2.49	8.34	8.65	0
CAN	2.03	2.00	5.24	5.25	0
CHE	3.16	3.25	6.94	7.13	0
CZE	3.48	3.40	15.89	15.48	0
DEU	3.80	3.80	8.80	8.62	0
DNK	3.11	3.06	8.05	8.06	0
ESP	2.57	2.60	7.23	7.43	0
FIN	2.02	2.05	16.28	15.76	0
FRA	3.36	3.34	10.05	9.94	0
GBR	2.29	2.23	14.74	15.07	0
GRC	3.23	3.29	10.33	10.73	0
HUN	2.68	2.63	14.57	13.15	0
IRL	3.19	3.18	14.56	13.95	0
ISL	1.64	1.59	13.94	13.73	0
ITA	2.02	2.02	8.18	8.30	0
JPN	3.37	3.45	36.57	37.54	1
KOR	3.36	3.41			
LUX	1.05	1.14	3.73	3.82	0
MEX	2.71	2.64	7.61	6.78	0
NLD	2.74	2.77	10.16	10.18	0
NOR	3.11	3.07	11.21	11.34	0
NZL	2.69	2.67	6.58	6.79	0
POL	2.34	2.37	53.97	49.11	1
PRT	3.02	3.02	11.09	10.64	0
SVK	2.59	2.57	47.09	49.10	1
SWE	2.37	2.28	8.09	8.29	0
TUR	3.84	3.82	18.20	18.18	1
USA	4.22	4.20	7.89	8.15	0

CONCLUSION

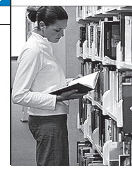
This chapter described the different steps for analysing data with plausible values. It also provided some SPSS® macros to facilitate the computations.

Attention was also drawn to a common error that consists of computing the average of the plausible values at the student level and adding this value to the database to be used as the student score in analyses. The correct method involves the averaging process which should always occur at the latest stage, that is on the statistic that will be reported.

The particular issue of analysing two sets of plausible values was also presented in the case of a correlation. The procedure that was applied can also be extended to other linear or non-linear modelling, such as a linear regression analysis.

Finally, an unbiased shortcut was described, one that is useful for time-consuming procedures, *e.g.* multilevel procedures.

PISA DATA ANALYSIS MANUAL: SPSS® SECOND EDITION – ISBN 978-92-64-05626-8 – © OECD 2009



Notes

1. Reading was the major domain in PISA 2000, mathematics in PISA 2003 and science in PISA 2006.

2. For more information, see the *Manual for the PISA 2000 Database* (OECD, 2002b).

3. A correlation coefficient can be expressed by the cosines of the angle formed by the two variables.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table of contents

PISA DATA ANALYSIS MANUAL: SPSS® SECOND EDITION – ISBN 978-92-64-05626-8 – © OECD 2009

LIST OF BOXES

LIST OF FIGURES

LIST OF TABLES

PISA DATA ANALYSIS MANUAL: SPSS® SECOND EDITION – ISBN 978-92-64-05626-8 – © OECD 2009

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

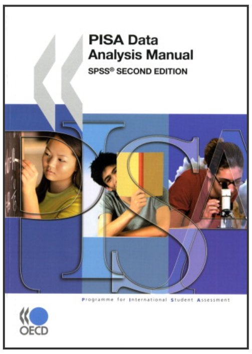
All syntaxes and macros in this manual can be copied from the PISA website (*www.pisa.oecd.org*). These macros were developed for SPSS 17.0. The 19 SPSS® macros presented in Chapter 17 need to be saved under "c:\pisa\macro\", before staring analysis. Each chapter of the manual contains a complete set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

Rounding of figures

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

Country abbreviations used in this manual

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



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), “Analyses with Plausible Values”, in *PISA Data Analysis Manual: SPSS, Second Edition*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264056275-9-en