14

Studying the Relationship between Student Performance and Indices Derived from Contextual Questionnaires

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

INTRODUCTION

The PISA initial reports have used the following tools to describe the relationship between student performance and questionnaire indices: *(i)* dividing the questionnaire indices into quarters and then reporting the mean performance by quarter; *(ii)* the relative risk; *(iii)* the effect size; and *(iv)* the linear regression. This chapter discusses technical issues related to these four tools and presents some SAS® macros that facilitate their computation.

ANALYSES BY QUARTERS

As described in Chapter 5, the indices derived from questionnaire data were generated with the Rasch Model, and students' estimates were reported with the *weighted likelihood estimates* (WLEs). As previously mentioned, WLE individual estimates constitute a discontinuous variable. Indeed, with the Rasch Model, the raw score is a sufficient statistic.

Table 14.1 presents the distribution of the questionnaire index of **cultural possessions at home** from the PISA 2006 data in Luxembourg. This table shows the discontinuous character of the variable.

Table 14.1
Distribution of the questionnaire index of cultural possession at home in Luxembourg (PISA 2006)

WLE	%
-1.60	17.52
-1.57	0.13
-1.49	0.10
-0.93	0.13
-0.84	0.02
-0.58	24.53
-0.50	0.27
-0.23	0.02
-0.20	0.15
-0.16	0.38
0.13	0.14
0.21	23.74
0.50	0.07
0.58	0.82
1.11	0.09
1.19	0.05
1.23	31.84

The cultural-possession-at-home scale consists of three dichotomous items. There are therefore four possible scores, ranging from 0 to 3. Thus, 98% of the students (17.52 + 24.53 + 23.74 + 31.84) are distributed among four WLEs. All the other WLEs, with negligible percentages, represent response patterns with at least one missing value.

How can this distribution be divided into four quarters, especially given that approximately 32% of the students have the highest WLE?

There are two ways of recoding the data:
1. smaller versus equal and greater;
2. smaller and equal versus greater.

Depending on the procedure adopted, the percentages of students in the bottom quarter, second quarter, third quarter, and top quarter will vary. Further, neither of these two procedures generate four equal quarters of 25% students each. Since the percentages of students in each quarter can vary among countries, no international comparisons can be made.

It is therefore necessary to distribute the students with a WLE equal to one of the three cutpoint percentiles into the two respective adjacent quarters. For instance, if 31.84% of the students get a score equal to percentile 75, it is then necessary to sample 6.84% of these students who will be allocated to the third quarter. The remaining 25% will be allocated to the fourth quarter.

This random subsampling process is implemented by adding a small random variable to the questionnaire index. That random noise will in a sense transform this discontinuous variable into a pseudo-continuous variable. The three new percentiles on this pseudo-continuous variable will divide the index variable into quarters, each including exactly 25% of the students.

This random allocation of some parts of the population to one of the four quarters adds an error component to the standard error. Indeed, in the example, the composition of the 6.84% of the students allocated to the third quarter is likely to differ between two runs of the procedure.

To account for this new error component, the statistical approach adopted for the analyses of plausible values can be implemented. It will therefore consist of:

- computing a set of five plausible quarters for each student;
- computing the required statistic and its respective sampling variance by using the final and 80 replicate weights, per plausible quarter;
- averaging the five estimates and their respective sampling variances;
- computing the imputation variance;
- combining the sampling variance and the imputation variance to obtain the final error variance.

As the dependent variable is exclusively, or nearly exclusively, a set of plausible values, the procedure described in Chapter 8 will be used, except that each plausible value will be analysed with a different plausible quarter.

Box 14.1 **SAS® syntax for the quarter analysis (*e.g.* PISA 2006)**

```
libname PISA2006  “c:\pisa\2006\data\”;
options nofmterr notes;
run;
data temp1;
   set pisa2006.stu;
   if (cnt=“LUX”);
   scie1=pv1scie;
   scie2=pv2scie;
   scie3=pv3scie;
   scie4=pv4scie;
   scie5=pv5scie;
   w_fstr0=w_fstuwt;
   keep cnt schoolid stidstd scie1-scie5 w_fstr0-w_fstr80 cultposs;
run;
%include “c:\pisa\macro\quartile_pv.sas”;
%QUARTILE_PV(INFILE=temp1,
             REPLI_ROOT=w_fstr,
             BYVAR =cnt ,
             PV_ROOT =scie,
             INDEX =cultposs,
             LIMIT=yes,
             LIMIT_CRITERIA=100 20 20 1,
             ID_SCHOOL=schoolid,
             OUTFILE =exercise1);
```

A SAS® macro was developed to facilitate these computations, as largely used in the three PISA initial reports. Box 14.1 presents the SAS® syntax and Table 14.2 the output data file.

The nine arguments of the quarter SAS® macro have been extensively described in other chapters. The output data file contains: *(i)* the mean of the contextual index (the independent variable) per quartile and its corresponding standard error; and *(ii)* the average performance in science (the dependent variable) and its standard errors.

Table 14.2
Output data file exercise1 from Box 14.1

CNT	CAT	INDEX_STAT	INDEX_SESTAT	PV_STAT	PV_SESTAT	FLAG_STUD	FLAG_SCH	FLAG_PCT
LUX	1	-1.31	0.01	451.40	3.09	0	0	0
LUX	2	-0.35	0.01	471.64	3.37	0	0	0
LUX	3	0.51	0.01	494.24	3.00	0	0	0
LUX	4	1.23	0.00	531.71	3.52	0	0	0

THE CONCEPT OF RELATIVE RISK

The notion of relative risk is a measure of association between an antecedent factor and an outcome factor (Cornfield, 1951). The relative risk is simply the ratio of two risks, *i.e.* the risk of observing the outcome when the antecedent is present, and the risk of observing the outcome when the antecedent is not present.

Table 14.3 presents the notation that will be used.

$p_{..}$ is equal to $\frac{n_{..}}{n_{..}}$, with $n_{..}$ the total number of students and $p_{..}$ is therefore equal to 1; $p_{i.}$, $p_{.j}$ respectively represent the marginal probabilities for each row and for each column. The marginal probabilities are equal to the marginal frequencies divided by the total number of students. Finally, the p_{ij} values represent the probabilities for each cell and are equal to the number of observations in a particular cell divided by the total number of observations.

Table 14.3
Labels used in a two-way table

		Outcome measure		
		Yes	No	Total
Antecedent measure	Yes	p_{11}	p_{12}	$p_{1.}$
	No	p_{21}	p_{22}	$p_{2.}$
	Total	$p_{.1}$	$p_{.2}$	$p_{..}$

In this chapter, the conventions for the two-way table are:

- The rows represent the antecedent factor with:
  – the first row having the antecedent; and
  – the second row not having the antecedent.
- The columns represent the outcome with:
  – the first column having the outcome; and
  – the second column not having the outcome.

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

With these conditions, the relative risk is equal to:

$$RR = \frac{(p_{11} / p_{1.})}{(p_{21} / p_{2.})}$$

Let's suppose that a psychologist wants to analyse the risk of a student repeating a grade if the parents recently divorced. The psychologist draws a simple random sample of students in grade 10. In this example, the outcome variable is present if the child is repeating grade 10 and the antecedent factor is considered present if the student's parents divorced in the past two years. The results are found in Table 14.4 and Table 14.5.

Table 14.4
Distribution of 100 students by parents' marital status and grade repetition

	Repeat the grade	Does not repeat the grade	Total
Parents divorced	10	10	20
Parents not divorced	5	75	80
Total	15	85	100

Table 14.5
Probabilities by parents' marital status and grade repetition

	Repeat the grade	Does not repeat the grade	Total
Parents divorced	0.10	0.10	0.20
Parents not divorced	0.05	0.75	0.80
Total	0.15	0.85	1.00

The relative risk is therefore equal to:

$$RR = \frac{(p_{11} / p_{1.})}{(p_{21} / p_{2.})} = \frac{(0.10/0.20)}{(0.05/0.80)} = \frac{0.5}{0.0625} = 8$$

This means that the probability of repeating grade 10 is eight times greater if the parents recently divorced than if they had stayed together.

Instability of the relative risk

The relative risk was developed for dichotomous variables. More and more often, this coefficient is extended and is used with continuous variables. However, to apply the coefficient to continuous variables, a cutpoint needs to be set for each variable and the continuous variables need to be dichotomised.

It is important to recognise that when applied to dichotomised variables, the computed values of the relative risk will depend on the value of the chosen cutpoint.

Table 14.6
Relative risk for different cutpoints

Percentile	Relative risk
10	2.64
15	2.32
20	1.90
25	1.73
30	1.63

To demonstrate the influence of the cutpoint on the relative risk, two random variables were generated with a correlation of 0.30. These two variables were then transformed into dichotomous variables by using the 10^{th}, 15^{th}, 20^{th}, 25^{th} and 30^{th} percentiles respectively as cutpoints. Table 14.6 presents the relative risk for a range of choices for the cutpoints.

Table 14.6 shows that the relative risk coefficient is dependent on the setting of the cutpoints; thus, the values should be interpreted in light of this.

Such a comparison of the relative risks was computed for the PISA 2000 data to identify the changes depending on the cutpoint location. The antecedent factor was the mother's educational level and the outcome variable was student performance in reading. Low reading performance was successively defined within countries as being below the 10th, 15th, 20th, 25th, 30th and 35th percentiles.

In PISA 2000, the relative risks for these different cutpoints are on average (across OECD countries) equal to 2.20, 1.92, 1.75, 1.62, 1.53, and 1.46, respectively. In PISA, it was decided to use the 25th percentile as the cutpoint for continuous variables when calculating the relative risk.

Computation of the relative risk

Depending on the variables involved in the computation of the relative risk, the procedure might differ. Indeed, the relative risk concept requires as input two dichotomous variables, such as gender.

However, most of the variables in the PISA databases are not dichotomous; they are categorical or continuous variables.

The recoding of a categorical into a dichotomous variable does not raise any specific issues. From a theoretical point of view, the purpose of the comparison needs to be decided; the recoding will follow. For instance, in PISA 2003, the education levels of the parents are reported by using the ISCED classification (OECD, 1999b). If the comparison is based on the distinction between tertiary versus non-tertiary education, then the categorical variable can be recoded into a dichotomous variable.

Numerical variables also have to be recoded into dichotomous variables. As stated earlier, the OECD has decided to divide numerical variables based on the 25th percentile.

If plausible values are involved as outcome measures, after the recoding of the five estimates of the student performance into dichotomous variables, five relative risks will be computed and then combined.

In the PISA databases however, most numerical variables are discontinuous variables. To ensure that the 25th percentile will divide the variables into two categories that will include 25% and 75% respectively, a random component has to be added to the initial variable, as described in the previous section on analyses per quarter. Five relative risk estimates are computed and then combined.

Box 14.2 presents the SAS® syntax for computing the increased likelihood of the students in the bottom quarter of HISEI (international socio-economic index of occupational status) scoring the bottom quarter of the science performance distribution, with the PISA 2006 data in France. As HISEI is a discontinuous variable with a limited number of values, it is necessary to add a random component. This example therefore has five antecedent variables and five outcome variables.

The first macro devoted to the computation of the relative risk requires five dummy variables as antecedents and five dummy variables as outcomes. Value 1 will be assigned if the risk is present; otherwise the assigned value will be 0. Value 1 will also be assigned if the outcome is present; otherwise it will be 0. It is of prime importance to respect these application conditions. Inverting the values will not stop the macro running, but it will change the meaning of the results.

Table 14.7 shows that a student in the bottom quarter of the international socio-economic index of occupational status has 2.38 times more chance of appearing in the bottom quarter of the science performance distribution.

Box 14.2 **SAS® syntax for computing the relative risk with five antecedent variables and five outcome variables (*e.g.* PISA 2006)**

```
data temp2;
    set pisa2006.stu;
    w_fstr0=w_fstuwt;
    if (cnt in (“FRA”));
    array a1 (5) scie1-scie5;
    array a2 (5) level1-level5;

    if (st04q01=1) then gender=0;
    if (st04q01=2) then gender=1;

    if (intscie <= 0) then int=1;
    if (intscie > 0 ) then int=0;
    if (intscie in (.,.I,.M,.N)) then int=.;

    ses1=hisei+(0.01*normal(-12));
    ses2=hisei+(0.01*normal(-23));
    ses3=hisei+(0.01*normal(-34));
    ses4=hisei+(0.01*normal(-45));
    ses5=hisei+(0.01*normal(-56));

    if (escs in (.,.I,.M,.N)) then delete;
run;
proc means data=temp2 vardef=wgt noprint;
    var ses1-ses5 pv1scie pv2scie pv3scie pv4scie pv5scie;
    by cnt;
    weight w_fstr0;
    output out=temp3 p25=ses25_1-ses25_5 pv25_1-pv25_5;
run;
data temp4;
    merge temp2 temp3;
    by cnt;
    array a1 (5) ses1-ses5;
    array a2 (5) ses25_1-ses25_5 ;
    array a3 (5) ses_risk1-ses_risk5;
    array a4 (5)  pv1scie pv2scie pv3scie pv4scie pv5scie;
    array a5 (5) pv25_1-pv25_5;
    array a6 (5) pv_out1-pv_out5;
    do i=1 to 5;
       if (a1(i) <= a2(i)) then a3(i)=1;
       if (a1(i) > a2(i)) then a3(i)=0;
       if (a1(i) in (.,.M,.I,.N)) then a3(i)=.;
    end;
    do i=1 to 5;
       if (a4(i) <= a5(i)) then a6(i)=1;
       if (a4(i) > a5(i)) then a6(i)=0;
       if (a4(i) in (.,.M,.I,.N)) then a6(i)=.;
    end;
run;
%include “c:\pisa\macro\relative_risk_no_pv.sas”;
%include “c:\pisa\macro\relative_risk_pv.sas”;
%BRR_RR_PV(INFILE=temp4,
           REPLI_ROOT=w_fstr,
           BYVAR=cnt,
           ANTECEDENT_ROOT=ses_risk,
           OUTCOME_ROOT=pv_out,
           LIMIT=no,
           LIMIT_CRITERIA= ,
           ID_SCHOOL=schoolid,
           OUTFILE=exercise2);
run;
```

Table 14.7
Output data file exercise2 from Box 14.2

CNT	STAT	SESTAT
FRA	2.38	0.19

A second macro presented in Box 14.3 has been developed for analyses that involve only one antecedent variable and one outcome variable.

Box 14.3 **SAS® syntax for computing the relative risk with one antecedent variable and one outcome variable (*e.g.* PISA 2006)**

```
%BRR_RR(INFILE=temp4,
        REPLI_ROOT=w_fstr,
        BYVAR=cnt,
        ANTECEDENT=gender,
        OUTCOME=int,
        LIMIT=no,
        LIMIT_CRITERIA=,
        ID_SCHOOL=schoolid,
        OUTFILE=exercise3);
run;
```

No macro has been developed for analyses that involve one antecedent variable and five outcome variables. However, Box 14.4 presents the SAS® syntax for computing the relative risk in that case. It consists of running the macro for relative risk without plausible values five times and then, as usual, combining the results.

Box 14.4 **SAS® syntax for computing the relative risk with one antecedent variable and five outcome variables (*e.g.* PISA 2006)**

```
data temp5;
        set temp4;
        array a1 (5) pv1scie pv2scie pv3scie pv4scie pv5scie;
        array a2 (5) low1-low5;
        do i=1 to 5;
           if (a1(i) > 409.5) then a2(i)=0;
           if (a1(i) < 409.5) then a2(i)=1;
        end;
run;
%macro rr;
%do rr=1 %to 5;
%BRR_RR(INFILE=temp5,
           REPLI_ROOT=w_fstr,
           BYVAR=cnt,
           ANTECEDENT=gender,
           OUTCOME=low&rr,
           LIMIT=no,
           LIMIT_CRITERIA=,
           ID_SCHOOL=schoolid,
           OUTFILE=out&rr);
run;
data out&rr;
        set out&rr;
        stat&rr=stat;
        se&rr=sestat;
        keep cnt stat&rr se&rr;
run;
%end;
data out;
        merge out1 out2 out3 out4 out5;
        stat=(stat1+stat2+stat3+stat4+stat5)/5;
        mesvar= (((stat1-stat)**2)+((stat2-stat)**2)+((stat3-stat)**2)+
                ((stat4-stat)**2)+((stat5-stat)**2))/4;
        sampvar=((se1**2)+(se1**2)+(se1**2)+(se1**2)+(se1**2))/5;
        var=sampvar+(1.2*mesvar);
        se=var**0.5;
        keep cnt stat se;
run;
%mend;
%rr;
run;
```

EFFECT SIZE

An effect size is a measure of the strength of the relationship between two variables. PISA requires sampling a substantial number of students from each participating country. As standard errors are inversely proportional to the number of observations, small differences will be statistically different from 0. It is therefore recommended to analyse the strength of a relationship that is statistically significant. In other words, the effect size helps researchers decide whether a statistically significant difference is of practical concern.

The term effect size is commonly used to refer to standardised differences. Standardising a difference is useful when the metric has no intrinsic meaning. With variables that have an intrinsic meaning, it is preferable to use non-standardised differences. For instance, the differences between male and female averages of height or weight are more meaningful if they are expressed in metres or kilos than if they are expressed in standardised differences.

Mathematically, the effect size is equal to:

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

$\hat{\mu}_1$ and $\hat{\mu}_2$ respectively represent the mean estimates for groups 1 and 2 and σ_1^2, σ_2^2 their variance.

Effect sizes are particularly interesting in PISA. Firstly, as differences are also compared across countries, using the difference or using the effect size might change the interpretation of the results. Differences will mainly affect countries with large or small standard deviations. For instance, the mean and the standard deviation of the student performance in reading for males and for females are presented in Table 14.8.

Table 14.8

Mean and standard deviation for the student performance in reading by gender, gender difference and effect size (PISA 2006)

	Performance in reading							
	Mean		Standard deviation		Difference in means between females and males			
	Females	Males	Females	Males	Dif.	Effect size	Rank of the difference	Rank of the effect size
AUS	531.8	494.9	86.9	96.5	36.9	0.40	14	16
AUT	512.9	468.3	102.8	108.8	44.6	0.42	24	22
BEL	521.7	482.0	102.1	113.5	39.7	0.37	17	12
CAN	543.0	511.1	90.7	98.9	31.9	0.34	7	6
CHE	515.2	484.4	91.2	94.3	30.8	0.33	5	5
CZE	508.6	462.8	107.5	110.0	45.8	0.42	25	21
DEU	516.6	474.6	105.6	113.9	42.0	0.38	22	14
DNK	509.3	479.5	86.0	90.0	29.8	0.34	3	8
ESP	478.7	443.3	82.4	91.4	35.4	0.41	13	19
FIN	572.0	521.4	73.4	80.9	50.6	0.66	28	29
FRA	504.6	469.8	97.1	107.9	34.9	0.34	11	9
GBR	509.5	480.4	95.3	106.3	29.2	0.29	2	2
GRC	488.1	431.6	87.9	108.3	56.6	0.57	29	28
HUN	503.0	463.4	87.1	96.8	39.6	0.43	16	24
IRL	534.0	500.2	87.0	94.7	33.8	0.37	10	13
ISL	508.9	460.4	87.6	99.8	48.5	0.52	27	27
ITA	489.0	447.7	100.6	112.8	41.3	0.39	20	15
JPN	513.3	482.7	95.0	107.1	30.6	0.30	4	3
KOR	573.8	538.8	82.4	90.4	35.0	0.41	12	18
LUX	495.4	463.7	95.2	102.5	31.7	0.32	6	4
MEX	426.7	393.1	91.7	96.8	33.6	0.36	9	10
NLD	519.0	494.9	93.1	98.4	24.2	0.25	1	1
NOR	508.0	462.1	95.5	108.8	45.9	0.45	26	25
NZL	539.1	501.7	99.2	107.9	37.4	0.36	15	11
POL	527.6	487.4	92.6	103.5	40.1	0.41	18	20
PRT	488.2	455.3	94.1	100.8	32.8	0.34	8	7
SVK	487.8	446.1	98.5	107.0	41.7	0.40	21	17
SWE	528.1	487.6	92.6	99.3	40.5	0.42	19	23
TUR	471.0	427.3	84.6	94.8	43.7	0.49	23	26

The differences and the effect sizes have been sorted and ranked. In most cases, the two rankings presented in the two last columns of Table 14.8 are equal or close. However, for some countries, they differ substantially. For instance, Germany is ranked 22 in the difference, and 14 in the effect size; Hungary ranks 16 and 24, respectively. The German and Hungarian researchers and policy makers would certainly interpret the results differently if their analyses were based on the absolute difference or on the effect size.

An effect size also allows a comparison of differences across measures that differ in their metric. For example, it is possible to compare effect sizes between the PISA indices and the PISA test scores, as for example, gender differences in performance in science compared to gender differences in several indices.

Two SAS® macros have been developed for the computation of the effect size, depending on whether the independent variables consist of plausible values or not. Box 14.5 presents the SAS® syntax for running effect size analyses.

Box 14.5 **SAS® syntax for computing effect size (*e.g.* PISA 2006)**

```
data temp6;
        set pisa2006.stu;
        w_fstr0=w_fstuwt;
        scie1=pv1scie;
        scie2=pv2scie;
        scie3=pv3scie;
        scie4=pv4scie;
        scie5=pv5scie;
        if(cnt in ("AUT"));
run;
%include "c:\pisa\macro\effect_size_no_pv.sas";
%include "c:\pisa\macro\effect_size_pv.sas";
%BRR_EFFECT_PV(INFILE=temp6,
                REPLI_ROOT=w_fstr,
                BYVAR=cnt,
                PV_ROOT=scie,
                EFFECT=st04Q01 1 2,
                OUTFILE=exercise4);
run;
%BRR_EFFECT(INFILE=temp6,
             REPLI_ROOT=w_fstr,
             BYVAR=cnt,
             VAR=instscie,
             EFFECT=st04q01 1 2,
             OUTFILE=exercise5);
run;
```

The dependent variable is listed in the PV_ROOT or VAR argument and the independent variable is listed in the EFFECT argument. The two values that define the two subgroups follow the name of the independent variable in the EFFECT argument. The order of two values is important: with `st04q01 1 2`, the effect size will be equal to:

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\dfrac{\sigma_1^2 + \sigma_2^2}{2}}}$$

With `st04q01 2 1`, the effect size will be equal to:

$$\frac{\hat{\mu}_2 - \hat{\mu}_1}{\sqrt{\dfrac{\sigma_1^2 + \sigma_2^2}{2}}}$$

Table 14.9 and Table 14.10 present the structure of the exercise4 and exercise5 output data files.

Table 14.9
Output data file exercise4 from Box 14.5

CNT	STAT	SESTAT
AUT	-0.08	0.05

Table 14.10
Output data file exercise5 from Box 14.5

CNT	STAT	SESTAT
AUT	-0.21	0.04

LINEAR REGRESSION AND RESIDUAL ANALYSIS

This section is devoted to linear regression analyses. As it will be demonstrated, such models can be used, even if there is some dependency between the errors within schools. Further, analyses of regression residuals might also provide researchers an opportunity to investigate composition effects outside complex multilevel modelling, for instance.

Independence of errors

As expressed by P. Bressoux (2008), "*qui se ressemble s'assemble mais aussi qui s'assemble tend à se ressembler*" (all that appear similar, group together; but also, all that group together tend to appear similar).

Selected students that attend the same school cannot be considered as independent observations as they are usually more similar to each other than students that attend different schools. This assumption of independence of errors can also be translated as a requirement of the absence of intraclass correlation. The PISA initial and thematic reports have reported the school variances and the intraclass correlation coefficients. In some countries such as Austria, Belgium or Germany, more than 50% of the variance is accounted for by the schools. The assumption of independence between errors cannot be maintained.

The violation of this assumption does not bias the regression coefficient estimates, but underestimates its standard error, which leads to an increase of type I error (Bressoux, 2008, p. 108). Fortunately, standard errors in PISA are estimated by replication techniques and are therefore unbiased.

The PISA initial and thematic reports have extensively used the linear regression, mainly for reporting the change in the student performance score per unit of indices derived from contextual questionnaires. The detailed description of conducting linear regression analysis with using replicates and plausible values is presented in Chapters 7 and 8.

This section will describe an alternative use of the linear regression. It consists of analysing the residual according to a particular criterion. An illustration from Monseur and Crahay (forthcoming) is also provided to show the potential of such analyses.

The underlying hypothesis of the illustration relates to the impact of social segregation in schools on student performance. Typically, this hypothesis should be tested with a multilevel regression analysis. However, such models are quite complex and their results are not always easy to explain.

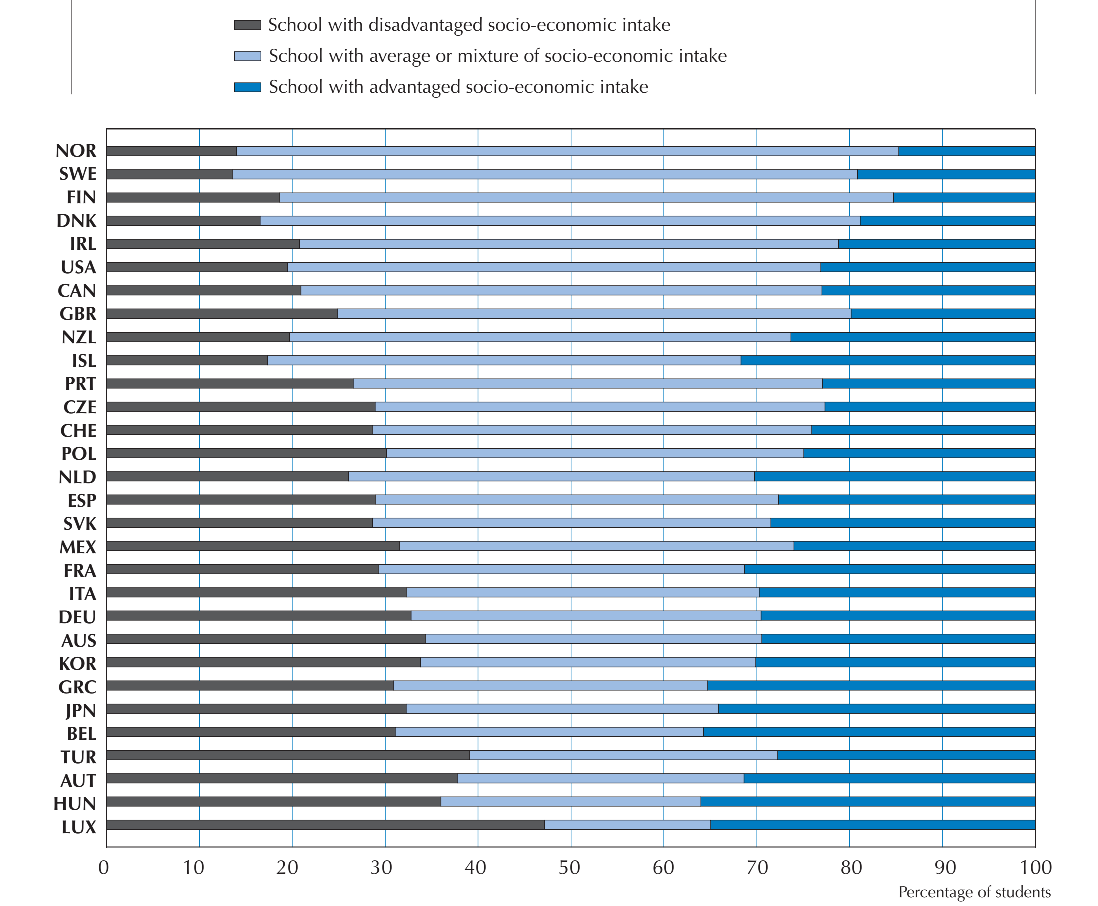
The first step implemented by Monseur and Crahay (forthcoming) was to allocate schools to one of the three following groups: *(i)* schools mainly attended by students with disadvantaged socio-economic backgrounds; *(ii)* schools mainly attended by students with average socio-economic backgrounds, or students with a mixture of disadvantaged and advantaged socio-economic backgrounds; and *(iii)* schools mainly attended by students with advantaged socio-economic backgrounds.

As the within-school samples are simple random samples, the standard error of the school average of the student socio-economic background can be estimated only with the student final weights without replicates. Then, the difference of a school average and the country average is statistically tested. If the null hypothesis is accepted, the school is allocated to the second group, *i.e.* the average or mixed schools; otherwise, the school is allocated to the first or to the third group, depending on whether the difference is negative or positive.

Figure 14.1 presents the percentage of students by the three school groups in PISA 2003.

Northern European countries, and to a lesser extent English-speaking countries, tend to have a large percentage of mixed schools. Highly tracked educational systems, such as Belgium or Germany, present small percentages of the average or mixed schools.

Figure 14.1
Percentage of schools by three school groups (PISA 2003)



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

The mathematics performance of the students was predicted by the PISA index of economic, social and cultural status (ESCS) of the student. The residuals were then saved and their means were computed by school categories and by the national quartiles of ESCS. Table 14.11 presents the average residual for the students in the bottom quarter of ESCS and for the students in the top quarter of ESCS, by the three school groups.

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)

	(i) School with disadvantaged socio-economic intake				*(ii)* School with average or mixture of socio-economic intake				*(iii)* School with advantaged socio-economic intake			
	Students in the bottom quarter of ESCS		Students in the top quarter of ESCS		Students in the bottom quarter of ESCS		Students in the top quarter of ESCS		Students in the bottom quarter of ESCS		Students in the top quarter of ESCS	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
AUS	-12.7	(5.4)	-32.5	(7.9)	10.3	(6.4)	-13.3	(4.4)	47.7	(11.6)	18.6	(4.1)
AUT	-20.1	(5.3)	-79.3	(8.6)	25.5	(7.7)	-25.5	(8.0)	87.3	(11.9)	20.4	(4.9)
BEL	-30.3	(4.4)	-71.9	(8.6)	22.2	(4.7)	-18.9	(5.0)	97.3	(5.7)	15.3	(3.5)
CAN	-8.5	(3.6)	-19.2	(5.9)	4.0	(2.7)	-9.5	(3.0)	23.9	(7.7)	17.1	(4.2)
CHE	-13.9	(6.3)	-43.3	(11.6)	-0.3	(4.7)	-24.6	(5.1)	72.1	(17.1)	21.9	(5.2)
CZE	-23.2	(5.5)	-71.8	(11.8)	6.6	(4.0)	-32.0	(4.4)	78.6	(15.1)	29.4	(5.4)
DEU	-30.2	(5.2)	-85.1	(9.2)	25.8	(5.8)	-32.0	(7.3)	109.3	(9.8)	24.9	(3.4)
DNK	-15.6	(7.6)	-18.5	(14.7)	2.5	(4.0)	-1.6	(3.8)	19.8	(13.7)	9.2	(5.7)
ESP	-7.2	(5.4)	-31.9	(9.3)	9.8	(4.5)	-9.8	(4.2)	43.0	(8.9)	13.8	(3.7)
FIN	1.6	(4.2)	-7.5	(7.5)	-0.3	(3.2)	-0.1	(3.8)	-15.8	(12.7)	1.6	(5.9)
FRA	-29.7	(5.2)	-70.8	(15.9)	27.4	(6.6)	-8.5	(7.0)	69.6	(7.7)	9.7	(5.2)
GBR	-12.4	(4.0)	-30.0	(8.0)	11.0	(4.3)	-10.9	(3.8)	47.5	(11.3)	25.9	(5.3)
GRC	-16.3	(5.5)	-63.8	(11.4)	21.3	(6.9)	-10.4	(5.6)	57.6	(9.4)	14.0	(4.8)
HUN	-21.2	(4.8)	-84.6	(10.2)	36.7	(5.8)	-37.0	(8.2)	76.8	(11.8)	14.4	(4.3)
IRL	-22.3	(5.7)	-38.8	(12.1)	10.0	(3.8)	-9.2	(4.0)	25.9	(11.3)	9.2	(5.7)
ISL	2.5	(4.9)	-0.2	(15.9)	-0.4	(3.8)	4.9	(4.5)	-1.1	(9.6)	0.9	(3.7)
ITA	-27.5	(5.6)	-71.9	(7.4)	24.2	(7.3)	-22.7	(6.9)	71.1	(12.4)	12.5	(5.9)
JPN	-27.7	(7.6)	-75.4	(11.0)	19.1	(7.9)	-41.1	(6.8)	78.3	(11.0)	26.6	(8.7)
KOR	-25.0	(5.5)	-79.8	(10.5)	41.8	(6.2)	-9.6	(5.1)	59.3	(9.5)	20.1	(8.4)
LUX	-9.4	(2.7)	-51.2	(6.3)	14.7	(6.6)	0.0	(8.2)	73.9	(8.1)	25.3	(3.0)
MEX	-13.9	(4.7)	-56.8	(9.3)	20.8	(5.2)	-30.5	(5.2)	71.7	(7.1)	22.4	(4.5)
NLD	-30.3	(6.1)	-87.7	(8.4)	15.1	(7.8)	-28.3	(6.2)	105.3	(7.5)	34.0	(5.0)
NOR	-0.5	(7.2)	-14.2	(15.3)	-1.4	(3.3)	-2.1	(4.3)	-11.4	(15.8)	5.6	(5.1)
NZL	-19.1	(6.3)	-45.0	(12.8)	7.7	(4.0)	-2.3	(4.1)	30.9	(10.6)	19.2	(3.9)
POL	-8.3	(5.0)	-19.5	(7.9)	0.7	(4.8)	-4.2	(4.7)	23.4	(14.2)	4.8	(3.8)
PRT	-22.3	(5.5)	-20.1	(11.3)	20.0	(6.4)	-5.4	(5.8)	58.6	(12.8)	15.1	(3.8)
SVK	-28.6	(4.9)	-75.2	(12.2)	10.3	(5.3)	-27.8	(6.3)	59.0	(9.1)	18.2	(5.2)
SWE	-12.4	(8.6)	-24.4	(12.5)	6.4	(3.5)	-2.4	(4.0)	15.7	(10.3)	9.5	(7.0)
TUR	0.6	(5.0)	-79.3	(10.2)	30.0	(7.0)	-38.7	(5.4)	94.4	(13.6)	38.1	(13.2)
USA	-22.3	(6.0)	-53.5	(12.5)	13.2	(3.6)	-1.6	(3.6)	27.3	(10.5)	15.3	(4.2)

In all countries with significant differences:

- Students attending schools with disadvantaged socio-economic intake perform on average lower than what would be predicted based on their socio-economic background.
- Students attending schools with advantaged socio-economic intake perform on average higher than what would be predicted based on their socio-economic background.
- Students attending schools with average or a mixture of socio-economic intake and from the bottom quarter of ESCS on average perform higher than what would be predicted based on their socio-economic background; the reverse is observed for students from the top quarter.

These results confirm a school socio-economic composition effect on student performance in mathematics. These composition effects appear small in northern European countries but large in countries such as Belgium, Czech Republic and Germany and the Netherlands.

As stated previously, composition effect can be estimated by multilevel regression modelling. Such modelling does not require allocating schools to some groups. On the other hand, residual analyses can easily be explained to policy makers.

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

STATISTICAL PROCEDURE

Box 14.6 presents the SAS® syntax for running a residual analysis that involves plausible values. First, five regression models need to be computed, each of them on one of the five plausible values. These five regressions can be run via a short SAS® macro. The regression residuals are saved and then combined with the original database. These five regression residuals are then analysed according to a similar procedure as any plausible values.

Box 14.6 [1/2] **SAS® syntax for residual analyses (*e.g.* PISA 2003)**

```
data temp7;
       set pisa2003.stud;
       if (cnt in
       ('AUS','AUT','BEL','CAN','CZE','DNK','FIN','FRA','DEU',
       'GRC','HUN','ISL','IRL','ITA','JPN','KOR','LUX','MEX',
       'NLD','NZL','NOR','POL','PRT','SVK','ESP','SWE','CHE',
       'TUR','GBR','USA'));
       w_fstr0=w_fstuwt;
       math1=pv1math;
       math2=pv2math;
       math3=pv3math;
       math4=pv4math;
       math5=pv5math;
       keep cnt schoolid stidstd w_fstr0-w_fstr80 math1-math5 escs;
run;
proc sort data=temp7;
       by cnt schoolid;
run;
proc means data=temp7 noprint vardef=wgt;
       var escs;
       by cnt schoolid;
       weight w_fstr0;
       output out=school1 mean=mu_escs;
run;
proc means data=temp7 noprint ;
       var escs;
       by cnt schoolid;
       output out=school2 stderr=err_escs;
run;
data school3;
       merge school1 school2;
       by cnt schoolid;
       drop _type_ _freq_;
run;
proc means data=temp7 noprint vardef=wgt;
       var escs;
       by cnt ;
       weight w_fstr0;
       output out=cnt mean=cnt_escs p25=per25 p75=per75 ;
run;
data school4;
       merge school3 cnt;
       by cnt;
       if ( err_escs > 0) then do;
       t=(mu_escs-cnt_escs)/err_escs;
       end;
       if (t <= -1.96) then schl_type=1;
       if (t > -1.96 and t < 1.96) then schl_type=3;
       if (t >= 1.96) then schl_type=3;
       if (t=. ) then schl_type=.;
       keep cnt schoolid schl_type per25 per75;
run;
data temp8;
       merge temp7 school4;
       by cnt schoolid;
       if (escs <= per25) then statut=1;
       if (escs > per25 and escs <= per75) then statut=2;
       if (escs > per75) then statut=3;
       if (escs in (.,.I,.M,.N)) then statut=.;
run;
```


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

Box 14.6 [2/2] **SAS® syntax for residual analyses (*e.g.* PISA 2003)**

```
%include "c:\pisa\macro\proc_freq_no_pv.sas";
%BRR_FREQ(INFILE=temp8,
          REPLI_ROOT=w_fstr,
          BYVAR=cnt,
          VAR=schl_type,
          LIMIT=no,
          LIMIT_CRITERIA=,
          ID_SCHOOL=,
          OUTFILE=exercise6);
run;
%macro residuals;
%do i=1 %to 5;
        proc reg data=temp8 noprint;
           model math&i=escs;
           by cnt;
           weight w_fstr0;
           output out=out&i r=res&i;
        run;

        data out&i;
           set out&i;
           keep cnt schoolid stidstd res&i;
        run;
%end;
data temp9;
        merge temp8 out1 out2 out3 out4 out5;
        by cnt schoolid stidstd;
run;
%mend;
%residuals;
run;
%include "c:\pisa\macro\proc_means_pv.sas";
%BRR_PROCMEAN_PV(INFILE=temp9,
                 REPLI_ROOT=w_fstr,
                 BYVAR=cnt schl_type statut,
                 PV_ROOT=res,
                 STAT=mean,
                 LIMIT=yes,
                 LIMIT_CRITERIA=100 10 5 1,
                 ID_SCHOOL=schoolid,
                 OUTFILE=exercise7);
run;
```

CONCLUSION

This chapter was devoted to some statistical issues related to the way the OECD reported the relationship between questionnaire indices and student performance in the initial reports.

The PISA initial and thematic reports extensively use the linear regression, in particular for estimating the impact of contextual indices on student performance. An alternative use of the linear regression that consists of analysing the residuals was also presented.

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

LIST OF BOXES

LIST OF FIGURES

LIST OF TABLES

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

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

User's Guide

Preparation of data files

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

SAS® users

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

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

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

SAS® syntax and macros

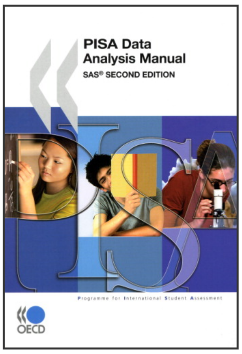
All syntaxes and macros in this manual can be copied from the PISA website (*www.pisa.oecd.org*). The 17 SAS® macros presented in Chapter 17 need to be saved under "c:\pisa\macro\", before staring analysis. Each chapter of the manual contains a complete set of syntaxes, which must be done sequentially,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

Rounding of figures

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

Country abbreviations used in this manual

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


PISA Data
Analysis Manual
SAS® SECOND EDITION
OECD