Chapter title: Analysing PIAAC data with IDB Analyzer (SPSS and SAS)

Author's information:
Andres Sandoval-Hernandez, University of Bath, A.Sandoval@bath.ac.uk
Diego Carrasco, Centro de Medición MIDE UC, Pontificia Universidad Católica de Chile, dacarras@uc.cl

Abstract:
This chapter provides a step-by-step guide for readers to perform both simple and complex analyses with PIAAC data using the IEA International Database (IDB) Analyzer. The IDB Analyzer is a Windows-based tool that generates SPSS and SAS syntax. Using this syntax, corresponding analyzes can be conducted in SPSS and SAS. The chapter will present the data-merging module as well as the analysis module. Potential analyses with the IDB Analyzer will be demonstrated, e.g., the calculation of percentages, averages, proficiency levels, linear regression, correlations, and percentiles.

Table of Contents

6. Analysing PIAAC data with IDB Analyzer (SPSS and SAS) 3

6.1. The IDB Analyzer 3

6.2. Merging files with the IDB analyzer 5

6.3. Examples Analyses with the IDB Analyzer 9

6.3.1. Means with plausible values 9

6.3.2. Means with other variables 14

6.3.3. Percentiles 17

6.3.4. Percentages 20

6.3.5. Linear Regression 25

6.3.6. Correlations 29

6.3.7. Proficiency Levels 32

6.4. Concluding Remarks 36

List of Figures

Figure 6.1 IDB Analyzer main window 4

Figure 6.2 IDB Analyzer Merge Module: Select Data Files and Participants 6

Figure 6.3 IDB Analyzer Merge Module: Selecting all countries 7

Figure 6.4 IDB Analyzer Merge Module: Select Data files and Participants 8

Figure 6.5 IDB Analyzer Merge Module: Selecting all variables 8

Figure 6.6 Analysis of means by group setup 11

Figure 6.7 Analysis of mean by group output 11

Figure 6.8 Analysis of means setup 15

Figure 6.9 Analysis of means output 15

Figure 6.10 Percentile setup 18

Figure 6.11 Percentile output 19

Figure 6.12 Selecting a newly generated variable 22

Figure 6.13 Percentage setup 22

Figure 6.10 Percentage output 23

Figure 6.15 Regression setup 26

Figure 6.16 Regression output 27
Figure 6.17 Correlation setup 30
Figure 6.18 Correlation output 31
Figure 6.19 Benchmark setup 34
Figure 6.20 Benchmark output 35

List of Tables

Table 6-1 Generated files by an analysis of means 13
Table 6-2 Means of literacy scores for Female and Males in each country 13
Table 6-3 Generated files by an analysis of means 16
Table 6-4 Means of lifetime years in schooling 17
Table 6-5 Generated files by an analysis of percentiles 19
Table 6-6 Percentiles (25th, 50th and 75th) for total years of schooling by country 20
Table 6-3 Generated files by percentage analysis 23
Table 6-4 Proportion of participants without upper secondary education 24
Table 6-9 Generated files by a regression analysis 27
Table 6-7 Standardized Regression coefficients and model fit 28
Table 6-11 Generated files by a correlation analysis 31
Table 6-12 Correlations between literacy and numeracy scale scores 32
Table 6-13 Levels of proficiency 33
Table 6-14 Generated files by a correlation analysis 35
Table 6-15 Percentage of participants below proficiency level 1 35

6. Analysing PIAAC data with IDB Analyzer (SPSS and SAS)

This chapter describes the general use of IEA's International Database Analyzer (IDB Analyzer) for analyzing PIAAC data (IEA, 2019). The IDB Analyzer provides a user-friendly interface to easily merge the data files of the different countries participating in PIAAC. Furthermore, it seamlessly takes into account the sampling information and the multiple imputed achievement scores to produce accurate statistical results (see chapter 2 in this volume for details about PIAAC's complex sample and assessment design).

This chapter is subdivided into three main sections. In the first section, we will provide a brief overview of the software[1]. Sections two and three will be dedicated to the Merge and Analysis modules of the IDB Analyzer, respectively. For each of these two sections, we will provide a description of the functionalities of the respective modules and examples to illustrate some of the capabilities of the IDB Analyzer (version 4.0) to merge files and to compute a variety of statistics, including the calculation of percentages, averages, benchmarks (proficiency levels), linear regression, logistic regression, correlations, and percentiles.

6.1. The IDB Analyzer

Developed by the IEA Hamburg, the IDB Analyzer is an interface that creates syntax for SPSS (IBM, 2013) and SAS (SAS, 2012). The IDB Analyzer was originally designed to allow users to combine and analyze data from IEA's large-scale assessments, but it has been adapted to work with data from most major large-scale assessment surveys, including those conducted by the Organisation for Economic Co-operation and Development (OECD), such as PIAAC, PISA and TALIS.

The IDB Analyzer generates SPSS or SAS syntax files that take into account information from the complex sampling design of the study, to produce population estimates. In addition, the generated syntax makes appropriate use of plausible values for calculating estimates of achievement scores, combining both sampling variance and imputation variance. Using either SPSS or SAS to analyze PIAAC data considering its complex sample and complex assessment design without the IDB Analyzer, would require the user to have programing knowledge in order to create their own macros. The IDB analyser automatically generates these macros (syntax files) in a user-friendly environment that allows their customization according to the purposes of the intended analysis.

The IDB Analyzer consists of two modules, the merge module and the analysis module. These two modules are integrated and executed in one common application. When working

[1] Most of the information for this section is adapted from the last version of the Help Manual for the IDB analyzer (IEA, 2019)

with PIAAC data, the merge module is used to create analysis datasets by combining data files from different countries and selecting subsets of variables for analysis. The analysis module provides procedures for computing various statistics and their standard errors.
Once the IDB Analyzer application is launched[2], the main window will appear, as shown in Figure 6.1. Users have then the option of choosing either SPSS or SAS as their statistical software of choice. For the examples in this chapter, we will use the SPSS software. The main window also has options to select the "Merge Module", the "Analysis Module", the "Help Manual"; or to Exit the application.

Figure 6.1 IDB Analyzer main window

There are at least two ways to access guidance on how to use IDB Analyzer: video tutorials made by the IEA and the main Help guide that accompanies this software installation. An easy way to get you started with IDB Analyzer is to watch IEA video tutorials. These were made available at the following link: https://www.iea.nl/training#IDB_Analyzer_Video_Tutorials
These videos were shared via youtube and cover step by step examples of how to estimate correlations, percentiles, percentages and means, logistic regression, linear regression and benchmarks.
A second way to get help and guidance is to consult the "Help" manual via de main Menu in IDB Analyzer. This official manual can be accessed by clicking on the third button present in the main menu. Figure 6.1 shows how this main menu looks like.

[2] The latest version of the IDB Analyzer (version 4.0) and instruction to install it, are available from the IEA website https://www.iea.nl/index.php/data-tools/tools

The IDB Analyzer will work on most IBM-compatible computers using current Microsoft Windows[3] operating system. The IDB Analyzer is licensed free of cost and is for use only in accordance with the terms of the licensing agreement. While the IDB Analyzer is free, the user must own a valid license of at least one of the software packages used as statistical engine (i.e. SPSS Version 18 or later or SAS Version 9 or later). Additionally, the user should have a valid license of Microsoft Excel 2003 or later version (since outputs are also produced in this format). The IDB Analyzer's license expires at the end of each calendar year. So, every year, users have to download and reinstall the most current version of the software and agree to the new license.

6.2. Merging files with the IDB analyzer

PIAAC Public Use Files containing both responses to the background questionnaire and the cognitive assessment are available for downloading for each of the participant countries separately. The Merge Module of the IDB Analyzer allows users to combine datasets from more than one country into a single data file for cross-country analyses. For the purposes of this chapter, we will assume all data files have been copied within a folder named "C:\Data\PIAAC\." PIAAC data files are available in both SPSS and SAS from the PIAAC website[4]. Users should download the data files in the format of their preference.

The Merge Module recognizes the data files for PIAAC by reading the file names in the selected directory and matching them to the file naming convention pre-specified in the IDB Analyzer configuration files. For this reason, in order to ensure that the IDB Analyzer will correctly identify the different files contained in the PIAAC data sets, as well as the user-generated files:

- Users should not change the name of the files once downloaded from the PIAAC website.
- Users should not save the merged file in the same directory where the source files are located.
- Users should keep files from different studies and years in separate directories.

The following steps will create an SPSS or SAS data file with data from multiple countries and/or multiple file types:

1. Open IDB Analyzer.

[3] Currently there is no standalone Mac version of the IDB Analyzer. However, the software can be used on Mac through a virtual machine and Windows installed on it. The current version was tested using Windows installed on Parallels Desktop for Mac (http://www.parallels.com/products/desktop/).

[4] http://www.oecd.org/skills/piaac/

2. Select the Statistical Software you want to work with (Choose between SAS or SPSS).
3. Select the Merge Module of the IDB Analyzer.
4. Click the Merge Module button. The Merge Module interface is divided into two different tabs. In the first one, you can select the countries, and edit country labels. In the second tab, you can select the variables you want to include in your analysis and specify the name of the merged file.
5. Under the "Select Data Files and Participants" tab and in the "Select Directory" field, browse to the folder where all data files are located. For example, in Figure 6.2, all SPSS data files are located in the folder "C:\Data\PIAAC\." The program will automatically recognize and complete the "Select Study" and "Select Cycle" fields and list all countries available in this folder as possible candidates for merging.

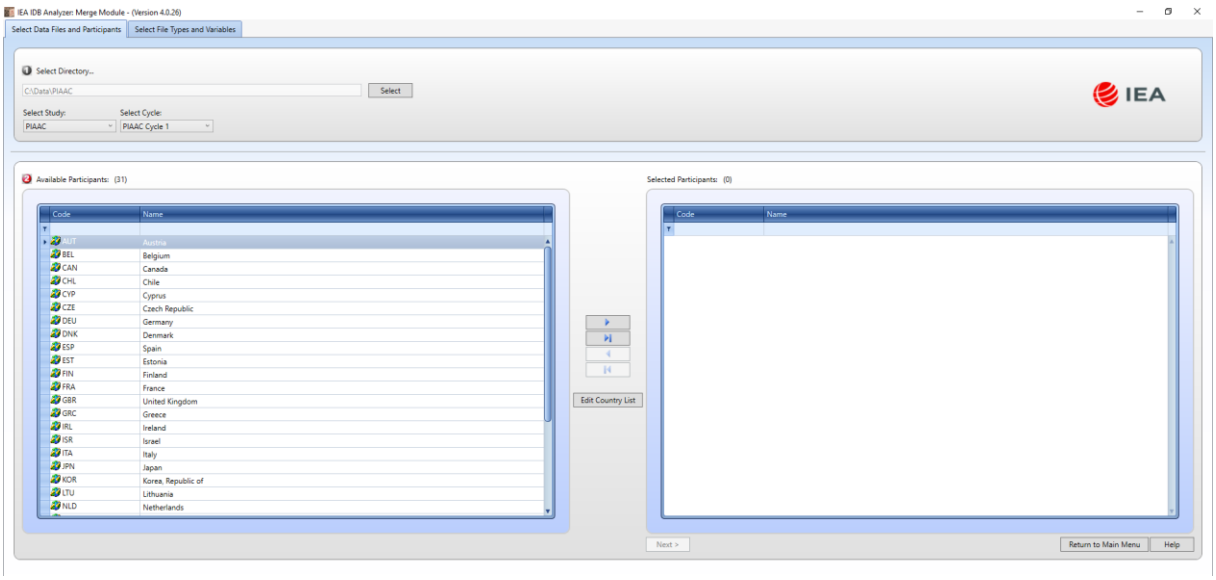


Figure 6.2 IDB Analyzer Merge Module: Select Data Files and Participants

6. Click the countries of interest from the "Available Participants" list and click the right arrow button (▻) to move them to the "Selected Participants panel" on the right. Individual countries can also be moved directly to the "Selected Participants" panel by double-clicking on them. To select multiple countries, hold the CTRL-key of the keyboard when clicking on countries. Click the tab-right arrow button (⧐) to move all countries to the Selected Participants panel. For this example, we selected all the countries available.

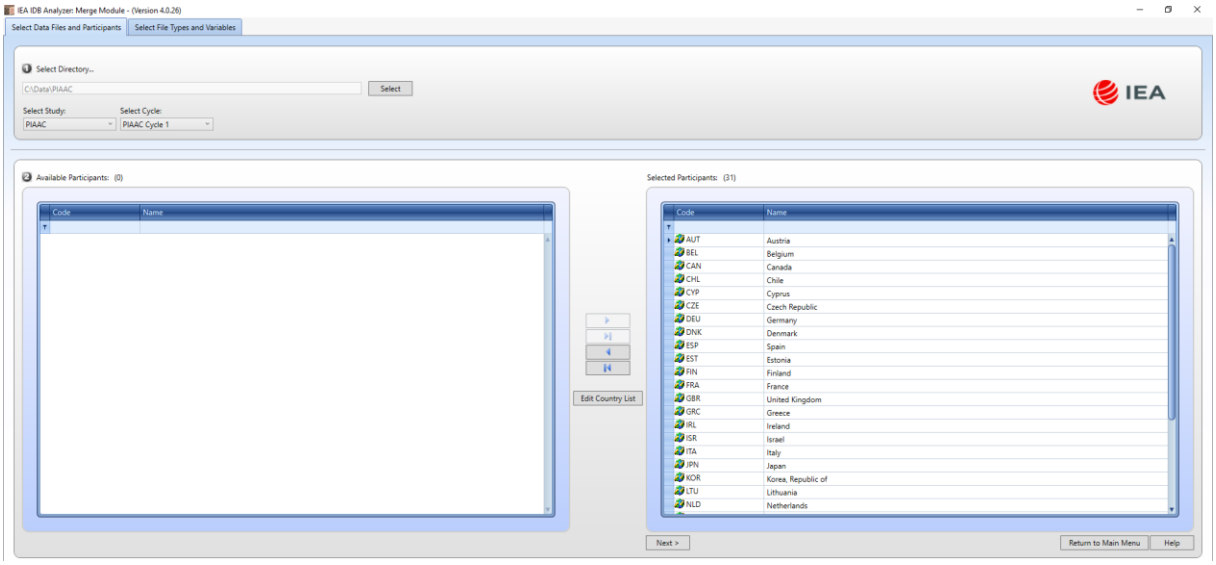


Figure 6.3 IDB Analyzer Merge Module: Selecting all countries

7. Click the "Next >" button to proceed to the next step. The software will open the "Select File Types and Variables" tab of the merge module (see **Error! Reference source not found.**), to select the file types and the variables to be included in the merged data file.
8. Select the files for merging by checking the appropriate boxes to the left of the window. For example, in **Error! Reference source not found.**, the "General Response File" has been selected[5]. Checking this box will automatically populate the "Selected Variables" panel with the three scores available in PIAAC (i.e. Literacy Scale Score, Numeracy Scale Score and Problem-Solving Scale Score), as well as with all the ID (e.g. Country ID) and sampling variables (e.g. sampling and replicate weights) needed for the corresponding analyses.

[5] With other studies such as PISA and TALIS there are more options. In the case of PIAAC, there is only one option.

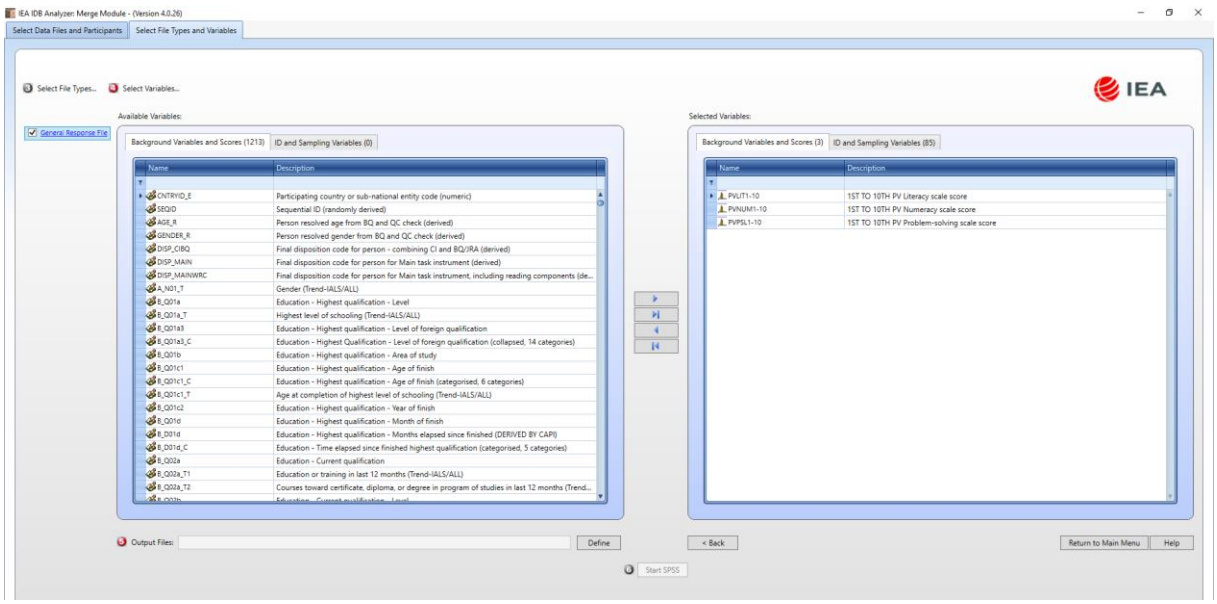


Figure 6.4 IDB Analyzer Merge Module: Select Data files and Participants

9. Select the variables of interest from the "Available Variables" list in the left panel. In SPSS, you can *right-click* on the variable names to open a menu with details about each of the available variables (i.e. variable name, label, measurement level and value labels). Variables are selected by clicking on them and then clicking the right arrow (▻) button. Clicking the tab-right arrow (⭲) button selects all variables.

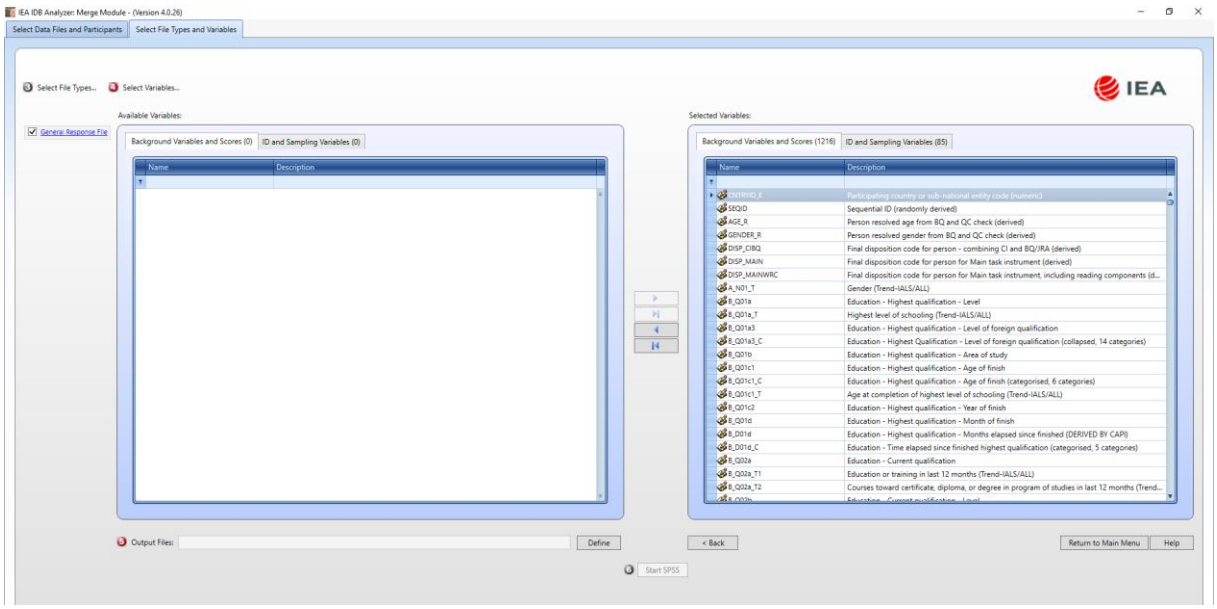


Figure 6.5 IDB Analyzer Merge Module: Selecting all variables

10. When selecting the variables, you can search variables by variable name, or by variable label using the filter boxes (blue space between column header and list of variables) in the "Available Variables" list and "Selected Variables" list.

11. Note that the IDB Analyzer assumes that files have the same structure and the variables have the same properties (e.g. variables, formats, labels) in each of these files. Any deviation from this can cause unexpected results. Should you want to modify the contents of a file for a country, or set of them, it is recommended to do this on the resulting merged file, after the merge is completed.
12. In the "Output Files" field, click on the "Define" button to specify the name for the merged data file and the folder where it will be saved. The IDB Analyzer also will create an SPSS syntax file (*SPS) (or a SAS syntax file, *.SAS, if you are using this software) of the same name and in the same folder with the code necessary to perform the merge. In the example shown in **Error! Reference source not found.**, the merged data file "merge_piaac.sav" and the syntax file "merge_piaac.sps" both will be created and stored in the folder titled "C:\Data\". The merged data file will contain all the variables listed in the "Selected Variables" panel, and if all available variables were selected the resulting merge file should be about 622 megabytes of size.
13. Click the "Start SPSS" button to create the SPSS syntax file. An SPSS Syntax Editor window with the created syntax code will be automatically opened. The syntax file can be executed by opening the "Run" menu of SPSS and selecting the "All" menu option. Alternatively, you can also submit the code for processing with the keystrokes Ctrl+A (to select all), followed by Ctrl+R (to run the selection). In SAS, the syntax file can be executed selecting the "Submit" option from the "Run" menu.

Once SPSS or SAS has completed its execution, it is important to check the SPSS output window or SAS log for possible warnings. If warnings appear, they should be examined carefully because they might indicate that the merge process was not performed properly and that the resulting merged data file might not include all the relevant variables or countries.

6.3. Examples Analyses with the IDB Analyzer

In the following section, we will describe step-by-step instructions to produce means, percentiles, percentages, linear regressions, correlations and benchmarks, using the latest PIAAC public-use data files. In each subsection, a sequence of steps was included as a numbered list. These steps are reiterated for each analysis routine. In this way, each subsection is self-contained, and the reader does not need to consult any other part of the chapter to complete the steps she or he needs to follow to produce means, percentiles, percentages, linear regressions, correlations or benchmarks.

6.3.1. Means with plausible values

In this section, we illustrate how to estimate the means of literacy scores by country. The first example contains a variable with plausible values. In PIAAC there are three variables with plausible values: the literacy scale scores, the numeracy scale score, and the problem-solving scale score. Each of these variables consists of ten different columns of values within the PIAAC data set. For each test, plausible values are generated as random draws of the posterior distribution of the participant's proficiency (Wu, 2005). To produce population estimates with these scores, IDB Analyzer computes the results for each plausible and combine these estimates using Rubin-Shaffer rules (Rutkowski, Gonzalez, Joncas, & von Davier, 2010). The following steps produce mean estimates of literacy proficiency by country, for females and males.

1. Open IDB Analyzer.
2. Select the Statistical Software you want to work with (Choose between SAS or SPSS).
3. Open the Analysis Module of the IDB Analyzer
4. For this example, specify the data file “merge_piaac.sav” as the Analysis File (see section 6.2 in this chapter for the details on how this file was created).
5. Select “PIAAC (using final full sample weight)” as the Analysis Type.
6. Select “Percentages and Means” as the Statistic Type.
7. Under the “Plausible Values Options”, select “Use PVs”.
8. Click on the “Separate Tables by” section at the right-hand side of the software window. This section will become active and highlighted in light yellow.
9. Go to the "Select variables" section and click on the “GENDER_R” variable in the fourth row of the name list.
10. Drag the “GENDER_R” variable to the “Separate Tables by” section.
11. Click on the “Plausible Values” section at the right-hand side of the software window. This section will become active and highlighted in light yellow.
12. Go to the "Select variables" section and click on the “PVLIT1-10” variable in the first row of the name list.
13. Drag the “PVLIT1-10” variable to the “Plausible Values“ section.
14. The Weight Variable is automatically selected by the software. SPFTWT0 is selected by default, this variable contains the final sampling weight.
15. Specify the name and the folder of the output files in the “Output Files” field by clicking the Define/Modify button. For this example, we use the term “mean_with_pv”.

After all these steps, the reached setup should look similar to Figure 6.6:

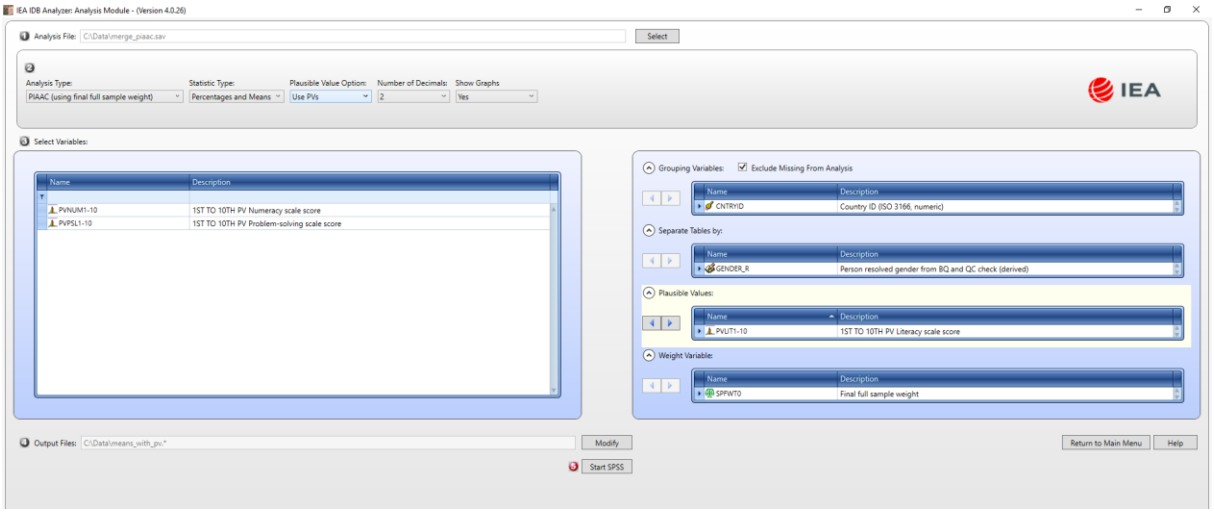
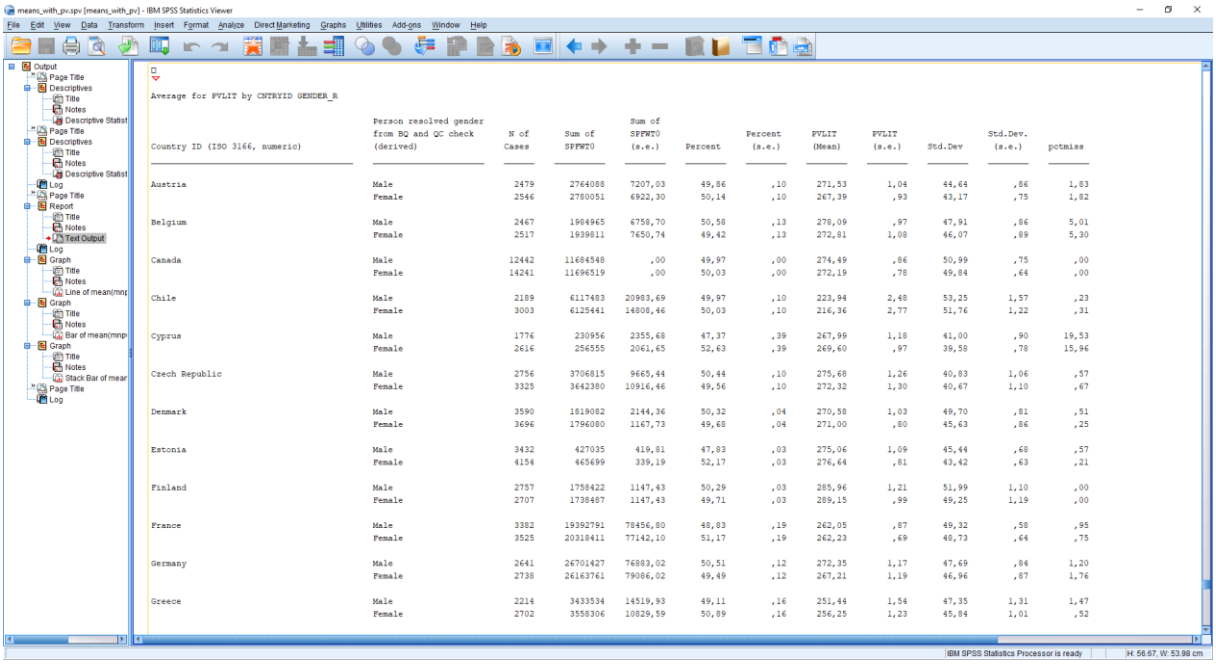


Figure 6.6 Analysis of means by group setup

16. Then, click the Start SPSS button. This will create an SPSS syntax file and open it in an SPSS editor window.
17. To start the computations, one needs to press the following keys combinations. CTRL+A first, to select the entire generated code present in the syntax window, and then CTRL+R to run these commands. The output of these analyses is depicted in Figure 6.7.



Average for PVLIT by CNTRYID GENDER_R

Country ID (ISO 3166, numeric)	Person resolved gender from BQ and QC check (derived)	N of Cases	Sum of SPFWT0	Sum of SPFWT0 (s.e.)	Percent	Percent (s.e.)	PVLIT (Mean)	PVLIT (s.e.)	Std.Dev	Std.Dev. (s.e.)	pctmiss
Austria	Male	2479	2764088	7207,03	49,86	,10	271,53	1,04	44,64	,86	1,83
	Female	2546	2780051	6922,30	50,14	,10	267,39	,93	43,17	,75	1,82
Belgium	Male	2467	1984965	6758,70	50,58	,13	278,09	,97	47,91	,86	5,01
	Female	2517	1939811	7650,74	49,42	,13	272,81	1,08	46,07	,89	5,30
Canada	Male	12442	11684548	,00	49,97	,00	274,49	,86	50,99	,75	,00
	Female	14241	11696519	,00	50,03	,00	272,19	,78	49,84	,64	,00
Chile	Male	2189	6117483	20983,69	49,97	,10	223,94	2,48	53,25	1,57	,23
	Female	3003	6125441	14808,46	50,03	,10	216,36	2,77	51,76	1,22	,31
Cyprus	Male	1776	230956	2355,68	47,37	,39	267,99	1,18	41,00	,90	19,53
	Female	2616	256555	2061,65	52,63	,39	269,60	,97	39,58	,78	15,96
Czech Republic	Male	2756	3706815	9665,44	50,44	,10	275,68	1,26	40,83	1,06	,57
	Female	3325	3642380	10916,46	49,56	,10	272,32	1,30	40,67	1,10	,67
Denmark	Male	3590	1819082	2144,36	50,32	,04	270,58	1,03	49,70	,81	,51
	Female	3696	1796080	1167,73	49,68	,04	271,00	,80	45,63	,86	,25
Estonia	Male	3432	427035	419,81	47,83	,03	275,06	1,09	45,44	,68	,57
	Female	4154	465699	339,19	52,17	,03	276,64	,81	43,42	,63	,21
Finland	Male	2757	1758422	1147,43	50,29	,03	285,96	1,21	51,99	1,10	,00
	Female	2707	1738487	1147,43	49,71	,03	289,15	,99	49,25	1,19	,00
France	Male	3382	19392791	78456,80	48,83	,19	262,05	,87	49,32	,58	,95
	Female	3525	20318411	77142,10	51,17	,19	262,23	,69	48,73	,64	,75
Germany	Male	2641	26701427	76883,02	50,51	,12	272,35	1,17	47,69	,84	1,20
	Female	2738	26163761	79086,02	49,49	,12	267,21	1,19	46,96	,87	1,76
Greece	Male	2214	3433534	14519,93	49,11	,16	251,44	1,54	47,35	1,31	1,47
	Female	2702	3558306	10829,59	50,89	,16	256,25	1,23	45,84	1,01	,52

Figure 6.7 Analysis of mean by group output

In the generated output, the first column contains the list of countries. The second column presents the categorical values of the "GENDER_R" variable: "Male" and "Female". In the

third column, the nominal sample size is presented for each group, within each country. In the fourth column, the sum of survey weights is included. These later numbers represent the survey population to which the estimates are projected to (Heeringa, West, & Berglund, 2009). Additionally, IDB Analyzer generates standard errors for the survey population size (sixth column). In the "Percent" column, the estimate of the proportion of each group in the population is presented. These point estimates are accompanied by its standard errors in the "Percent (s.e.)" column. In the column "PVLIT (Mean)" we find the point estimates of the literacy scores. Each country has two values, one for males, and one for females. These point estimates present uncertainty, due to measurement error and due to sampling error. This uncertainty is summarized in the "PVLIT (s.e.)" column. Standard Deviations of these means are included in the "Std.Dev" column. Similarly to previous estimates, at its right, standard errors of the standard deviations are provided, in the column "Std.Dev. (s.e.)". Finally, the last column, "pctmiss", contains the percentage of missing cases in the variables involved in the analysis ("PVLIT1-10" and "GENDER_R").

IDB Analyzer creates six files after an analysis of means with plausible values is complete. Table 6-3 details these files and their content.

Table 6-1 Generated files by an analysis of means

Generated files	File type	Content
means_with_pv.sps	SPSS	Syntax to run the means computations.
means_with_pv.spv	SPSS	Output of the means computations.
means_with_pvGENDER_R.sav means_with_pvGENDER_R.xlsx	SPSS Excel	Contains the means estimates and their standard errors.
means_with_pv_PVLIT_by_GENDER_R_Sig.sav means_with_pv_PVLIT_by_GENDER_R_Sig.xlsx	SPSS Excel	Contains a group within-country comparison for the estimated means and percentages, providing t statistics for these comparisons.

Using the results provided in the file “means_with_pvGENDER_R.xlsx”, we created

Table 6-4 to present the computed results. Means are presented and its standard errors are included in parenthesis.

Table 6-2 Means of literacy scores for Female and Males in each country

Country	Female	Male	Country	Female	Male
Austria	267.39	271.53	Korea, Republic of	269.43	275.72
	(0.93)	(1.04)		(0.87)	(0.75)
Belgium	272.81	278.09	Lithuania	268.47	264.97
	(1.08)	(0.97)		(1.20)	(1.32)
Canada	272.19	274.49	Netherlands	280.92	287.06
	(0.78)	(0.86)		(0.94)	(1.08)
Chile	216.36	223.94	New Zealand	280.69	280.66
	(2.77)	(2.48)		(1.06)	(1.20)
Cyprus	269.60	267.99	Norway	276.43	280.34
	(0.97)	(1.18)		(0.91)	(0.97)
Czech Republic	272.32	275.68	Poland	270.08	263.66
	(1.30)	(1.26)		(0.86)	(0.97)
Denmark	271.00	270.58	Russian Federation	277.37	272.90
	(0.80)	(1.03)		(2.88)	(2.98)
Estonia	276.64	275.06	Singapore	253.89	261.42
	(0.81)	(1.09)		(1.01)	(0.98)
Finland	289.15	285.96	Slovak Republic	274.22	273.47
	(0.99)	(1.21)		(0.82)	(0.86)
France	262.23	262.05	Slovenia	257.67	255.17
	(0.69)	(0.87)		(0.99)	(1.08)
Germany	267.21	272.35	Spain	249.45	254.11
	(1.19)	(1.17)		(1.04)	(1.00)
Greece	256.25	251.44	Sweden	277.54	280.88
	(1.23)	(1.54)		(1.10)	(1.08)
Ireland	265.43	267.71	Turkey	220.89	231.98
	(1.10)	(1.17)		(1.35)	(1.56)
Israel	255.04	255.45	United Kingdom	271.03	273.90
	(0.96)	(1.14)		(1.29)	(1.37)
Italy	250.61	250.36	United States	269.47	270.16
	(1.32)	(1.50)		(1.33)	(1.21)
Japan	294.69	297.78	Table Average	266.34	267.96
	(1.01)	(0.88)		(0.22)	(0.23)

IDB Analyzer produces a “Table Average”, which contains an overall mean between all countries, with its standard error. These estimates are presented in Table 6.2 in the last row, in the second column. The illustrated routine can be replicated with the Numeracy scale scores and with the problem-solving scores present in PIAAC study.

6.3.2. Means with other variables

The following example is simpler than its previous counterpart. In the next example, we compute the mean of total years of schooling in each country. In the PIAAC study, a total of years in schooling was derived using different responses of participants regarding their educational participation during their lifetime. These values can be found in the `YRSQUAL_T` variable. Using the IDB Analyzer we need to follow the next steps:

1. Open IDB Analyzer.
2. Select the Statistical Software you want to work with (Choose between SAS or SPSS).
3. Open the Analysis Module of the IDB Analyzer
4. For this example, specify the data file “merge_piaac.sav” as the Analysis File (see section 6.2 in this chapter for the details on how this file was created).
5. Select “PIAAC (using final full sample weight)” as the Analysis Type.
6. Select “Percentages and Means” as the Statistic Type.
7. Under the “Plausible Values Options”, select “None Used”.
8. Click on the “Analysis Variables” section at the right-hand side of the software window. This section will become active and highlighted in light yellow.
9. Go to the "Select variables" section, and under the “Description” heading click on it, and type in “total years”. This action would look for all the variables containing “total” and “year” in their description field.
10. Specify the variable YRSQUAL_T as the analysis variable by clicking the “Analysis Variables” field to activate it. Select YRSQUAL_T from the list of available variables present in the “Select Variables” section and move it to the “Analysis variables” by clicking the right arrow button in this section.
11. The Weight Variable is automatically selected by the software. SPFTWT0 is selected by default, this variable contains the final sampling weight.
12. Specify the name and the folder of the output files in the “Output Files” field by clicking the Define/Modify button. For this example, we use the term “mean”.
    After all these steps, the reached setup should look similar to Figure 6.8:

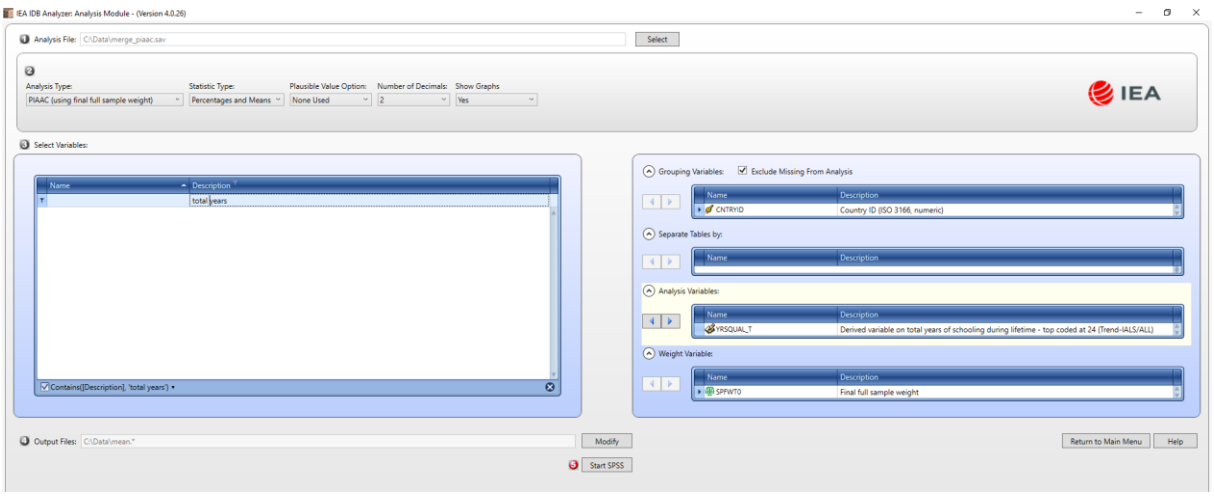
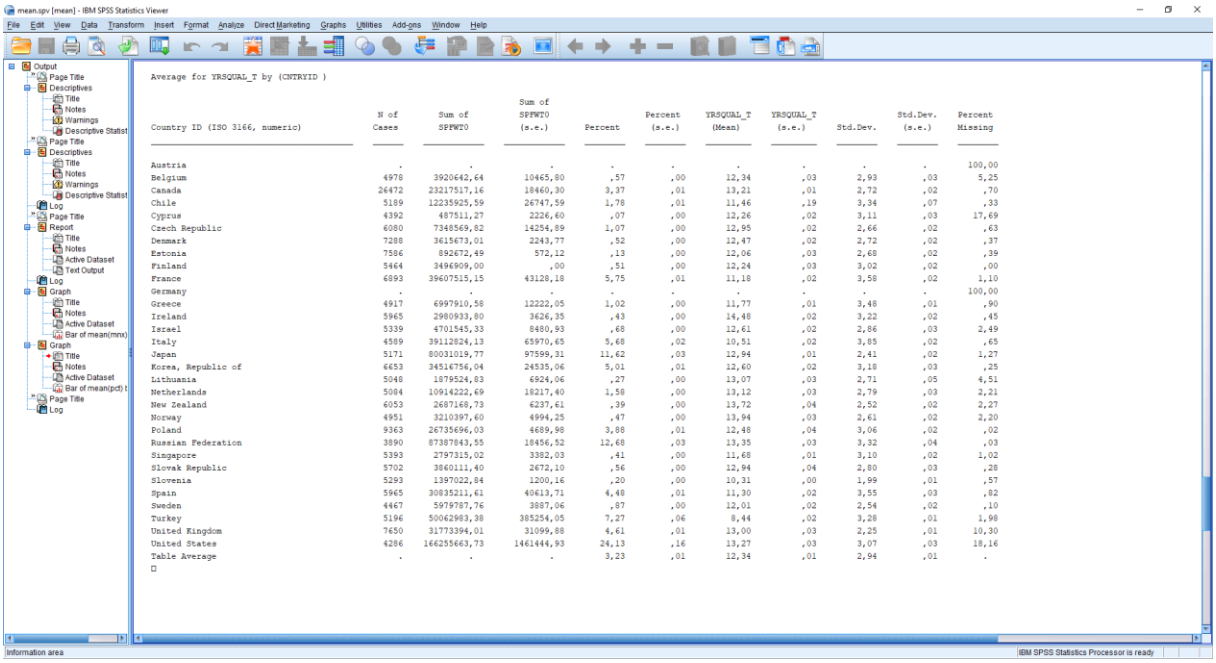


Figure 6.8 Analysis of means setup

13. Then, click the Start SPSS button. This will create an SPSS syntax file and open it in an SPSS editor window.
14. To start the computations, one needs to press the following keys combinations. CTRL+A first, to select the entire generated code present in the syntax window, and then CTRL+R to run these commands. The output of these analyses is depicted in Figure 6.9.



Average for YRSQUAL_T by (CNTRYID)

Country ID (ISO 3166, numeric)	N of Cases	Sum of SPFWT0	Sum of SPFWT0 (s.e.)	Percent	Percent (s.e.)	YRSQUAL_T (Mean)	YRSQUAL_T (s.e.)	Std.Dev.	Std.Dev. (s.e.)	Percent Missing
Austria	.	.	.	.	.	.	.	.	.	100,00
Belgium	4978	3920642,64	10465,80	,57	,00	12,34	,03	2,93	,03	5,25
Canada	26472	23217517,16	18460,30	3,37	,01	13,21	,01	2,72	,02	,70
Chile	5189	12235925,59	26747,59	1,78	,01	11,46	,19	3,34	,07	,33
Cyprus	4392	487511,27	2226,60	,07	,00	12,26	,02	3,11	,03	17,69
Czech Republic	6080	7348569,82	14254,89	1,07	,00	12,95	,02	2,66	,02	,63
Denmark	7288	3615673,01	2243,77	,52	,00	12,47	,02	2,72	,02	,37
Estonia	7586	892672,49	572,12	,13	,00	12,06	,03	2,68	,02	,39
Finland	5464	3496909,00	,00	,51	,00	12,24	,03	3,02	,02	,00
France	6893	39607515,15	43128,18	5,75	,01	11,18	,02	3,58	,02	1,10
Germany	.	.	.	.	.	.	.	.	.	100,00
Greece	4917	6997910,58	12222,05	1,02	,00	11,77	,01	3,48	,01	,90
Ireland	5965	2980933,80	3626,35	,43	,00	14,48	,02	3,22	,02	,45
Israel	5339	4701545,33	8480,93	,68	,00	12,61	,02	2,86	,03	2,49
Italy	4589	39112824,13	65970,65	5,68	,02	10,51	,02	3,85	,02	,65
Japan	5171	80031019,77	97599,31	11,62	,03	12,94	,01	2,41	,02	1,27
Korea, Republic of	6653	34516756,04	24535,06	5,01	,01	12,60	,02	3,18	,03	,25
Lithuania	5048	1879524,83	6924,06	,27	,00	13,07	,03	2,71	,05	4,51
Netherlands	5084	10914222,69	18217,40	1,58	,00	13,12	,03	2,79	,03	2,21
New Zealand	6053	2687168,73	6237,61	,39	,00	13,72	,04	2,52	,02	2,27
Norway	4951	3210397,60	4994,25	,47	,00	13,94	,03	2,61	,02	2,20
Poland	9363	26735696,03	4689,98	3,88	,01	12,48	,04	3,06	,02	,02
Russian Federation	3890	87387843,55	18456,52	12,68	,03	13,35	,03	3,32	,04	,03
Singapore	5393	2797315,02	3382,03	,41	,00	11,68	,01	3,10	,02	1,02
Slovak Republic	5702	3860111,40	2672,10	,56	,00	12,94	,04	2,80	,03	,28
Slovenia	5293	1397022,84	1200,16	,20	,00	10,31	,00	1,99	,01	,57
Spain	5965	30835211,61	40613,71	4,48	,01	11,30	,02	3,55	,03	,82
Sweden	4467	5979787,76	3887,06	,87	,00	12,01	,02	2,54	,02	,10
Turkey	5196	50062983,38	385254,05	7,27	,06	8,44	,02	3,28	,01	1,98
United Kingdom	7650	31773394,01	31099,88	4,61	,01	13,00	,03	2,25	,01	10,30
United States	4286	166255663,73	1461444,93	24,13	,16	13,27	,03	3,07	,03	18,16
Table Average	.	.	.	3,23	,01	12,34	,01	2,94	,01	.

Figure 6.9 Analysis of means output

Similar to the previous example, the generated output presents several columns. The first column is the list of countries. In the second column is the nominal sample size of each country. Notice that Austria and Germany do not have observations for this variable and present 100 per cent of missing. The third column contains the sum of survey weights, which represent the survey population size (Heeringa et al., 2009), and in the fourth column, IDB Analyzer includes the standards errors of the survey population size. In the "Percent" column, the proportion of the survey population size is depicted. For example, the United States projects its number of cases (4286) to a survey population of more than 166 million people, and its resulting proportion in the table is of "24,13"; whereas Canada has a larger nominal sample of 26472 cases, yet projected to a survey population of more than 23 million people, and hence its proportion in the table is of "3,37". These percentages are accompanied by its standard errors included in the 6 column. In the seventh column, the estimates of interest are included: the mean of total years of schooling per country, under heading "YRSQUAL_T (Mean)". Next to it, in the eight column, we can find the standard error of this estimates,

below the heading “YRSQUAL_T (s.e.)”. The “Std.Dev” column contains the standard deviations of the analysis variable and the “Std.Dev (s.e.)” contains the standard deviations standard errors. The last column of the table presents the percentage of missing of the analysed variable.

When the analysis of means is complete IDB Analyzer generates six files. Table 6-3 details these files and their content.

Table 6-3 Generated files by an analysis of means

Generated files	File type	Content
mean.sps	SPSS	Syntax to run the means computations.
mean.spv	SPSS	Output of the means computations.
mean.sav mean.xlsx	SPSS Excel	Contains the means estimates and their standard errors.
mean_YRSQUAL_T_by_CNTRYID_Sig.sav mean_YRSQUAL_T_by_CNTRYID_Sig.xlsx	SPSS Excel	Contains a country by country comparison for the estimated means, providing t statistics.

Using the results provided in the file “mean.xlsx”, we created

Table 6-4 to present the computed results.

Table 6-4 Means of lifetime years in schooling

Country	Mean	Standard Error	Country	Mean	Standard Error
Belgium	12.34	0.03	Lithuania	13.07	0.03
Canada	13.21	0.01	Netherlands	13.12	0.03
Chile	11.46	0.19	New Zealand	13.72	0.04
Cyprus	12.26	0.02	Norway	13.94	0.03
Czech Republic	12.95	0.02	Poland	12.48	0.04
Denmark	12.47	0.02	Russian Federation	13.35	0.03
Estonia	12.06	0.03	Singapore	11.68	0.01
Finland	12.24	0.03	Slovak Republic	12.94	0.04
France	11.18	0.02	Slovenia	10.31	0.00
Greece	11.77	0.01	Spain	11.30	0.02
Ireland	14.48	0.02	Sweden	12.01	0.02
Israel	12.61	0.02	Turkey	8.44	0.02
Italy	10.51	0.02	United Kingdom	13.00	0.03
Japan	12.94	0.01	United States	13.27	0.03
Korea, Republic of	12.60	0.02	Table Average	12.34	0.01

Considering that the population average might not be the most informative location parameter to describe the variable's distribution, in the next section we describe how to obtain percentiles of a continuous variable.

6.3.3. Percentiles

Means and percentiles are different location parameters in a distribution (Wilcox, 2017). The arithmetic mean is the expected location of the value with the least difference to the rest of the values within a distribution. In contrast, percentiles are any location under which there is a certain proportion of cases. Mean are informative for symmetric distributions, such as the normal distribution. However, when distributions depart from normality, medians (percentile 50th) or other location parameters could be of interest. For the following example, we choose percentile 25th, 50th and 75th, for the same variable. We will repeat the steps 1-3 from the previous routine, but we will change the Statistic type.

1. Open the Analysis Module of the IDB Analyzer.
2. For this example, specify the data file "merge_piaac.sav" as the Analysis File (see section 6.2 in this chapter for the details on how this file was created).
3. Select "PIAAC (using final full sample weight)" as the Analysis Type.

4. Select “Percentiles” as the Statistic Type.
5. Under the “Plausible Values Options”, select “None Used”.
6. Click on the “Analysis Variables” section at the right-hand side of the software window. This section will become active and highlighted.
7. Go to the "Select variables" section, and under the “Description” heading click on it, and type in “years”. This action would look for all the variables containing “years” in their description field.
8. Specify the variable YRSQUAL_T as the analysis variable by clicking the “Analysis Variables” field to activate it. Select YRSQUAL_T from the list of available variables present in the “Select Variables” section and move it to the “Analysis variables” by clicking the right arrow button in this section. In this step, is also possible to select more than one variable in this routine. However, for the sake of simplicity, in this example, we are including only one variable.
9. In the “Percentiles” section type in “25 50 75”, all separated by a space.
10. Specify the name and the folder of the output files in the Output Files field by clicking the Define/Modify button. For this example, we use the term “percentile”.

The generated setup should be similar to the screenshot presented in Figure 6.10.

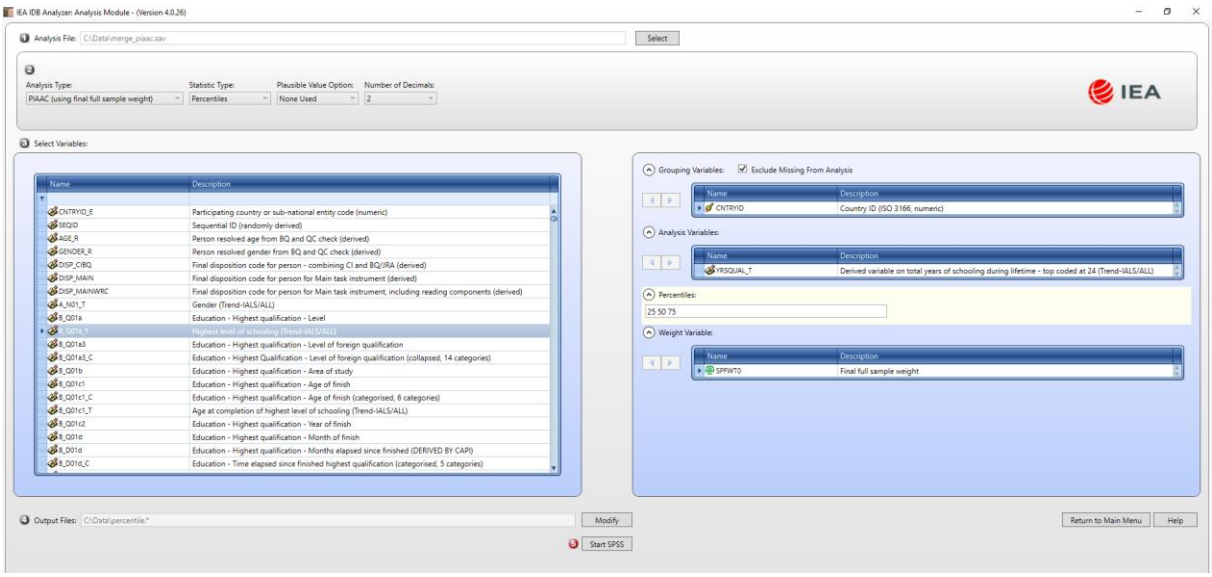
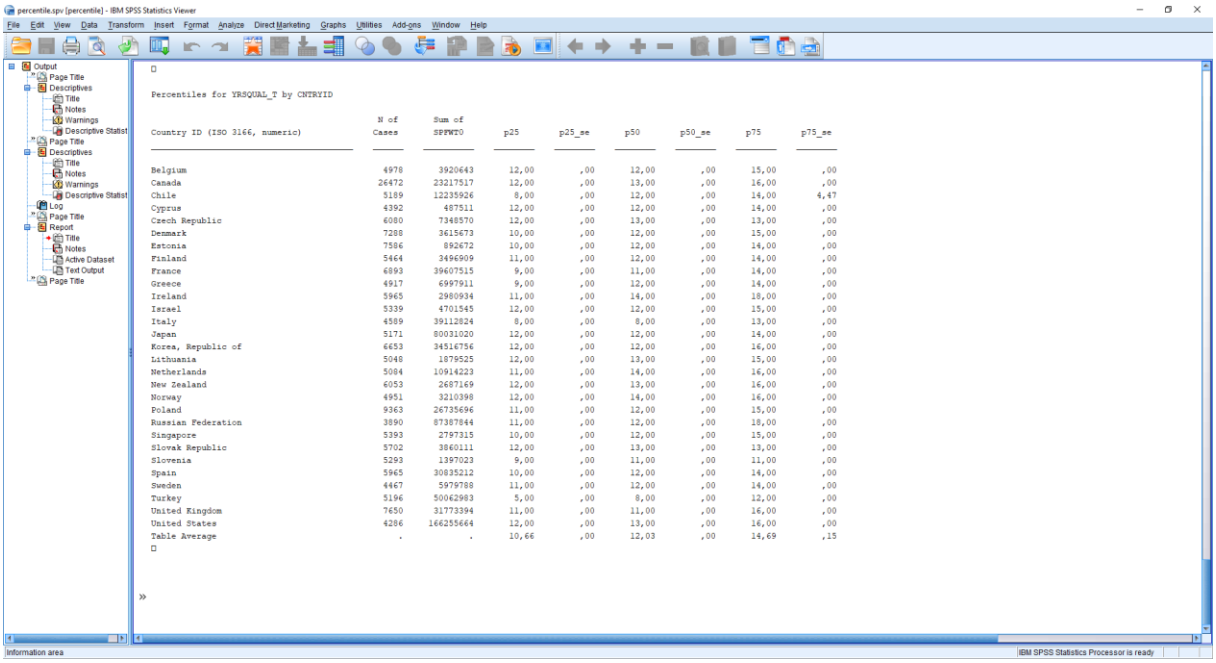


Figure 6.10 Percentile setup

11. Afterwards, click the “Start SPSS” button, run the syntax and wait for the results to appear in the output window. The output from this routine is presented in Figure 6.11.



Percentiles for YRSQUAL_T by CNTRYID

Country ID (ISO 3166, numeric)	N of Cases	Sum of SPFWT0	p25	p25_se	p50	p50_se	p75	p75_se
Belgium	4978	3920643	12,00	,00	12,00	,00	15,00	,00
Canada	26472	23217517	12,00	,00	13,00	,00	16,00	,00
Chile	5189	12235926	8,00	,00	12,00	,00	14,00	4,47
Cyprus	4392	487511	12,00	,00	12,00	,00	14,00	,00
Czech Republic	6080	7348570	12,00	,00	13,00	,00	13,00	,00
Denmark	7288	3615673	10,00	,00	12,00	,00	15,00	,00
Estonia	7586	892672	10,00	,00	12,00	,00	14,00	,00
Finland	5464	3496909	11,00	,00	12,00	,00	14,00	,00
France	6893	39607515	9,00	,00	11,00	,00	14,00	,00
Greece	4917	6997911	9,00	,00	12,00	,00	14,00	,00
Ireland	5965	2980934	11,00	,00	14,00	,00	18,00	,00
Israel	5339	4701545	12,00	,00	12,00	,00	15,00	,00
Italy	4589	39112824	8,00	,00	8,00	,00	13,00	,00
Japan	5171	80031020	12,00	,00	12,00	,00	14,00	,00
Korea, Republic of	6653	34516756	12,00	,00	12,00	,00	16,00	,00
Lithuania	5048	1879525	12,00	,00	13,00	,00	15,00	,00
Netherlands	5084	10914223	11,00	,00	14,00	,00	16,00	,00
New Zealand	6053	2687169	12,00	,00	13,00	,00	16,00	,00
Norway	4951	3210398	12,00	,00	14,00	,00	16,00	,00
Poland	9363	26735696	11,00	,00	12,00	,00	15,00	,00
Russian Federation	3890	87387844	11,00	,00	12,00	,00	18,00	,00
Singapore	5393	2797315	10,00	,00	12,00	,00	15,00	,00
Slovak Republic	5702	3860111	12,00	,00	13,00	,00	13,00	,00
Slovenia	5293	1397023	9,00	,00	11,00	,00	11,00	,00
Spain	5965	30835212	10,00	,00	12,00	,00	14,00	,00
Sweden	4467	5979788	11,00	,00	12,00	,00	14,00	,00
Turkey	5196	50062983	5,00	,00	8,00	,00	12,00	,00
United Kingdom	7650	31773394	11,00	,00	11,00	,00	16,00	,00
United States	4286	166255664	12,00	,00	13,00	,00	16,00	,00
Table Average	.	.	10,66	,00	12,03	,00	14,69	,15

Figure 6.11 Percentile output

The generated output presents nine columns. The first is the list of countries, the second is the nominal sample size for each country, and in the third column, we can find the sum of survey weights, which represent survey population size (Heeringa et al., 2009). In the "Percent" column, IDB Analyzer includes the proportion that the survey population size represents within the output table. Then, for each requested percentile (p25, p50, p75) we can find the point estimates and its standard error at its right (p25_se, p50_se, p75_se).

For the computation of percentiles, IDB Analyzer generates four files. Table 6-5 details these files and their content.

Table 6-5 Generated files by an analysis of percentiles

Generated files	File type	Content
percentile.sps	SPSS	Syntax to run the means computations.
percentile.spv	SPSS	Output of the means computations.
percentile.sav percentile.xlsx	SPSS Excel	Contains the means estimates and their standard errors.

Using the results provided in the file “percentile.xlsx”, we created Table 6-6 to present the computed results. The estimated percentiles are included for each country, alongside their standard errors in parenthesis.

Table 6-6 Percentiles (25th, 50th and 75th) for total years of schooling by country

Country	P25	P50	P75
Belgium	12.00 (0.00)	12.00 (0.00)	15.00 (0.00)
Canada	12.00 (0.00)	13.00 (0.00)	16.00 (0.00)
Chile	8.00 (0.00)	12.00 (0.00)	14.00 (4.47)
Cyprus	12.00 (0.00)	12.00 (0.00)	14.00 (0.00)
Czech Republic	12.00 (0.00)	13.00 (0.00)	13.00 (0.00)
Denmark	10.00 (0.00)	12.00 (0.00)	15.00 (0.00)
Estonia	10.00 (0.00)	12.00 (0.00)	14.00 (0.00)
Finland	11.00 (0.00)	12.00 (0.00)	14.00 (0.00)
France	9.00 (0.00)	11.00 (0.00)	14.00 (0.00)
Greece	9.00 (0.00)	12.00 (0.00)	14.00 (0.00)
Ireland	11.00 (0.00)	14.00 (0.00)	18.00 (0.00)
Israel	12.00 (0.00)	12.00 (0.00)	15.00 (0.00)
Italy	8.00 (0.00)	8.00 (0.00)	13.00 (0.00)
Japan	12.00 (0.00)	12.00 (0.00)	14.00 (0.00)
Korea, Republic of	12.00 (0.00)	12.00 (0.00)	16.00 (0.00)
Lithuania	12.00 (0.00)	13.00 (0.00)	15.00 (0.00)
Netherlands	11.00 (0.00)	14.00 (0.00)	16.00 (0.00)
New Zealand	12.00 (0.00)	13.00 (0.00)	16.00 (0.00)
Norway	12.00 (0.00)	14.00 (0.00)	16.00 (0.00)
Poland	11.00 (0.00)	12.00 (0.00)	15.00 (0.00)
Russian Federation	11.00 (0.00)	12.00 (0.00)	18.00 (0.00)
Singapore	10.00 (0.00)	12.00 (0.00)	15.00 (0.00)
Slovak Republic	12.00 (0.00)	13.00 (0.00)	13.00 (0.00)
Slovenia	9.00 (0.00)	11.00 (0.00)	11.00 (0.00)
Spain	10.00 (0.00)	12.00 (0.00)	14.00 (0.00)
Sweden	11.00 (0.00)	12.00 (0.00)	14.00 (0.00)
Turkey	5.00 (0.00)	8.00 (0.00)	12.00 (0.00)
United Kingdom	11.00 (0.00)	11.00 (0.00)	16.00 (0.00)
United States	12.00 (0.00)	13.00 (0.00)	16.00 (0.00)
Table Average	10.66 (0.00)	12.03 (0.00)	14.69 (0.15)

From the generated results we can notice most of the participating countries have a median lifetime of schooling of 12 years. Ireland, The Netherlands and Norway reach at least 14 years of schooling for half of their population of participants. In the lower end, Italy and Turkey presented a median schooling lifetime of 8 years.

6.3.4. Percentages

In the next example, we will create a new variable, not present in the merged files, to then retrieve percentage estimates at the population level for each country. We will use PIAAC data

to estimate the proportion of the population in each participating country that has reached at least upper-secondary education. To do this, we first need to recode a derived variable present in the public use file of the study. We will recode variable EDCAT8 into a dummy variable. EDCAT8 contains codes from the International Standard Classification of Education (ISCED) to express the highest level of formal education of the participants (OECD, 2015).

Using the following syntax code (see

Table 6-4), we can create a dummy variable, which differentiates between the participants who hold upper secondary qualification (coded as one), and the participants who present lower educational qualifications, such as primary degrees or incomplete secondary degrees (coded as zero).

To include this new variable in the generated merged, the user needs to open the merged file in SPSS. Then, open a new syntax window, type in the syntax code included in Code 6-1; press CTRL+A and CTRL+R to create this variable. Click on the window with the merged data, and press CTRL+S to save this variable in the merged file.

Code 6-1 Recoding highest educational level to a dummy variable

```
if (EDCAT8 <= 2) edu_usl = 0 .
if (EDCAT8 >= 3) edu_usl = 1 .
execute .

VARIABLE LABELS   edu_usl   'Population with upper secondary education
(1=yes, 0=no)'.

VALUE LABELS edu_usl
0 'No'
1 'Yes'.
```

With the merge file closed, one can open IDB Analyzer, and used this new variable for further analysis. In the next example, we will estimate what proportion of the population of the participant countries has at least upper secondary educational qualifications. Similarly, to previous examples, we start by opening IDB Analyzer.

1. Open the Analysis Module of the IDB Analyzer.
2. For this example, specify the data file “merge_piaac.sav” as the Analysis File (see section 6.2 in this chapter for the details on how this file was created).
3. Select “PIAAC (using final full sample weight)” as the Analysis Type.
4. Select “Percentages only” as the Statistic Type.
5. Click on the “Grouping Variables” section.
6. Go to the "Select variables" section, and under the “Description” heading click on it, and type in “upper”. This action would look for all the variables containing “upper” in their description field. This is presented in Figure 6.12

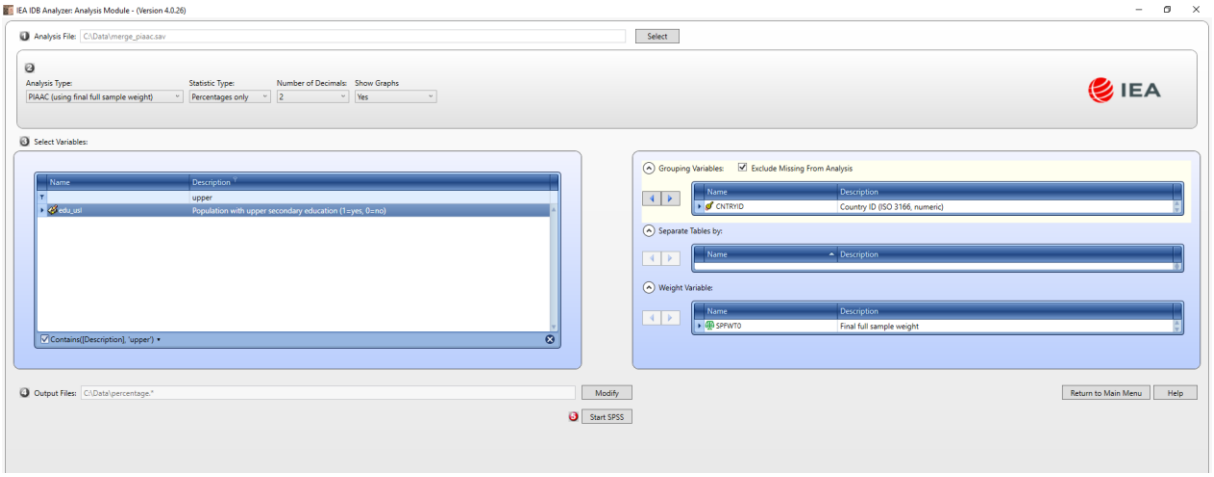


Figure 6.12 Selecting a newly generated variable

7. Drag the variable “edu_usl” to the “Grouping variable” section. By clicking the “Analysis Variables” field to activate it. Select “edu_usl” from the list of available variables present in the “Select Variables” section and move it to the “Grouping Variables” field by clicking the right arrow button in this section.
8. Specify the name and the folder of the output files in the Output Files field by clicking the Define/Modify button. In this example, we will use the term “percentage”. This setup is presented in Figure 6.13.

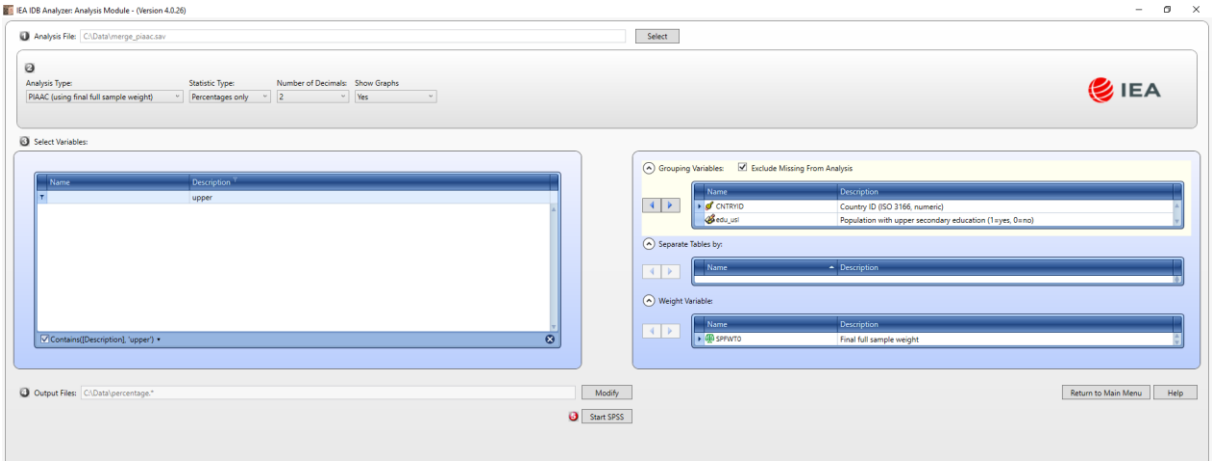
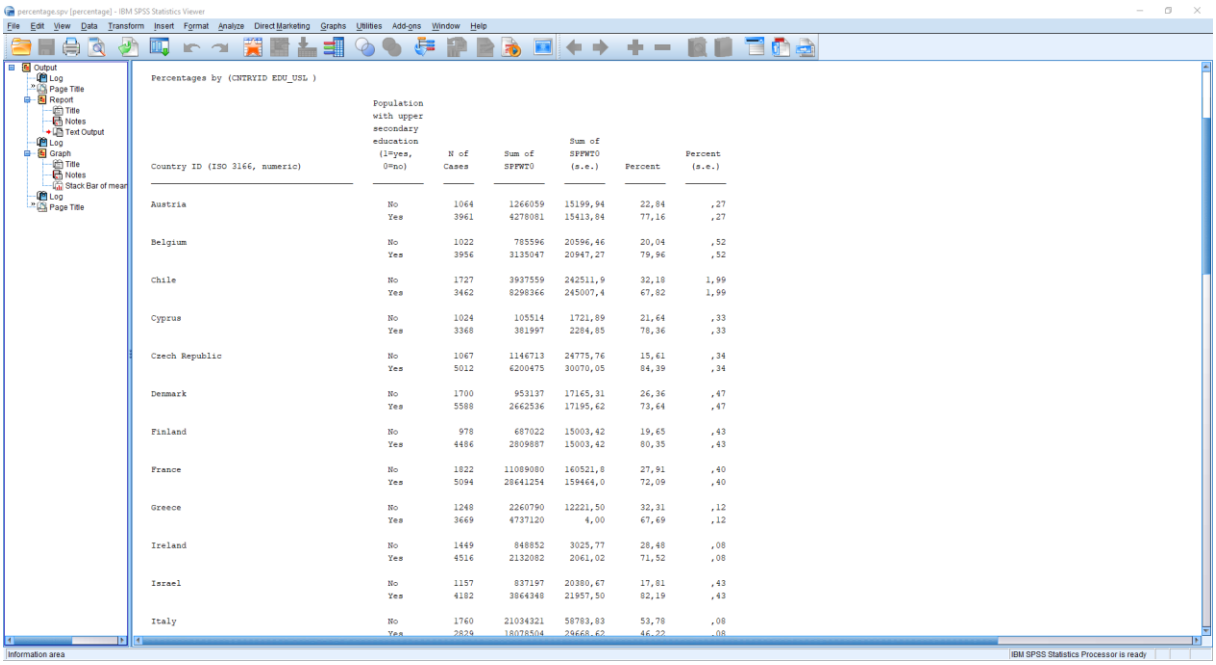


Figure 6.13 Percentage setup

9. Click the Start SPSS button to create the SPSS syntax file and open it in an SPSS editor window.
10. After the user has executed the generated syntax, by pressing the sequence of keys CTRL+A and CTRL+R, IDB Analyzer will start to run their macros to compute the requested percentages.

Once the calculations are finished, the SPSS output window would present the following results (see Figure 6.14).



Percentages by (CNTRYID EDU_USL)

Country ID (ISO 3166, numeric)	Population with upper secondary education (1=yes, 0=no)	N of Cases	Sum of SPFWT0	Sum of SPFWT0 (s.e.)	Percent	Percent (s.e.)
Austria	No	1064	1266059	15199,94	22,84	,27
	Yes	3961	4278081	15413,84	77,16	,27
Belgium	No	1022	785596	20596,46	20,04	,52
	Yes	3956	3135047	20947,27	79,96	,52
Chile	No	1727	3937559	242511,9	32,18	1,99
	Yes	3462	8298366	245007,4	67,82	1,99
Cyprus	No	1024	105514	1721,89	21,64	,33
	Yes	3368	381997	2284,85	78,36	,33
Czech Republic	No	1067	1146713	24775,76	15,61	,34
	Yes	5012	6200475	30070,05	84,39	,34
Denmark	No	1700	953137	17165,31	26,36	,47
	Yes	5588	2662536	17195,62	73,64	,47
Finland	No	978	687022	15003,42	19,65	,43
	Yes	4486	2809887	15003,42	80,35	,43
France	No	1822	11089080	160521,8	27,91	,40
	Yes	5094	28641254	159464,0	72,09	,40
Greece	No	1248	2260790	12221,50	32,31	,12
	Yes	3669	4737120	4,00	67,69	,12
Ireland	No	1449	848852	3025,77	28,48	,08
	Yes	4516	2132082	2061,02	71,52	,08
Israel	No	1157	837197	20380,67	17,81	,43
	Yes	4182	3864348	21957,50	82,19	,43
Italy	No	1760	21034321	58783,83	53,78	,08
	Yes	2829	18078504	29668,62	46,22	,08

Figure 6.14 Percentage output

Similarly, to the procedure of means estimation, the procedure to estimate percentages produces six files as outputs. These files and their contents are described in Table 6-7.

Table 6-7 Generated files by percentage analysis

Generated files	File type	Content
percentage.sps	SPSS	Syntax to run the percentage computations.
percentage.spv	SPSS	Output of the percentage computations.
percentage.sav percentage.xlsx	SPSS Excel	Contains the percentage estimates and their standard errors.
percentage__by_EDU_USL_Sig.sav percentage__by_EDU_USL_Sig.xlsx	SPSS Excel	Contains a country by country comparison for the estimated percentages.

Inspecting the generated output file in excel format, “percentage.xlsx”, we can filter and order the results to produce Table 6-8 and display the proportions of participants without upper secondary education for each participating country in PIAAC.

Table 6-8 Proportion of participants without upper secondary education

Country	Estimate	Standard Error
Turkey	64.51	0.17
Italy	53.78	0.08
Spain	47.46	0.05
Greece	32.31	0.12
Chile	32.18	1.99
Netherlands	31.02	0.63
Ireland	28.48	0.08
France	27.91	0.40
Norway	27.43	0.53
Denmark	26.36	0.47
United Kingdom	24.09	0.59
Sweden	23.74	0.39
Slovenia	23.63	0.20
Austria	22.84	0.27
New Zealand	22.63	0.72
Korea, Republic of	21.66	0.49
Cyprus	21.64	0.33
Slovak Republic	20.61	0.62
Belgium	20.04	0.52
Finland	19.65	0.43
Singapore	18.92	0.17
Israel	17.81	0.43
Czech Republic	15.61	0.34
Poland	15.34	0.42
Japan	14.78	0.40
United States	14.74	0.28
Lithuania	11.93	0.47
Russian Federation	7.03	0.78
Table Average	25.29	0.11

In the following section, we will use the dummy variable we have created “edu_usl” and estimate its relation to literacy scores in the population of each country.

6.3.5. Linear Regression

Apart from descriptive estimates such as means, percentiles and percentages, IDB Analyzer can also estimate regression models and logistic regression models (IEA, 2019). In the following

example, we will estimate the relationships between educational qualifications and literacy in each country. Specifically, we will estimate the gap in literacy scores between those who hold at least upper secondary education, and the rest of the population. Although this gap can be obtained with a mean comparison, we want to retrieve more estimates than the mean differences between the two groups. We will use the linear regression routine for these purposes, and get this difference as a standardized effect, while also retrieving a measure of explained variance. These results can answer "how much difference in literacy skills there is between those with and without upper secondary education?". To estimate a regression analysis, we need to follow the next steps in IDB Analyzer:

1. Open the Analysis Module of the IDB Analyzer.
2. For this example, specify the data file "merge_piaac.sav" as the Analysis File (see section 6.2 in this chapter for the details on how this file was created).
3. Select "PIAAC (using final full sample weight)" as the Analysis Type.
4. Select "Regression" as the Statistic Type.
5. Under the "Plausible Values Options", select "Use PVs".
6. On the right-hand side of the window, click on the area of "Dependent variables". This will become highlighted once is clicked.
7. Then, select "Plausible Values" in the righthand side window.
8. Move the cursor to the left-hand side of the window and click on the "PVLIT1-10" variable to select the literacy scores.
9. Go back to the right-hand side and click on the right arrow to move the "PVLIT1-10" variables, to the "Dependent variables" section.
10. Move the cursor to the "Independent Variables" section, and click on the "Categorical Variables" to active this section.
11. Move the cursor to "Select Variables" section on the left. Just right before the variable list, in the first row under the description section, type in: "upper". This will filter all present variables from the merge file.
12. Select the variable "edu_usl", and move it to the right-hand side, by clicking in the right arrow, under "Independent Variables", specifically using the right arrow from the "Categorical Variables" subsection.
13. Specify the name and the folder of the output files in the Output Files field by clicking the Define/Modify button. In this example, we will name the syntax file as "regression".

Once all previous steps are complete, the regression setup should look like Figure 6.15.

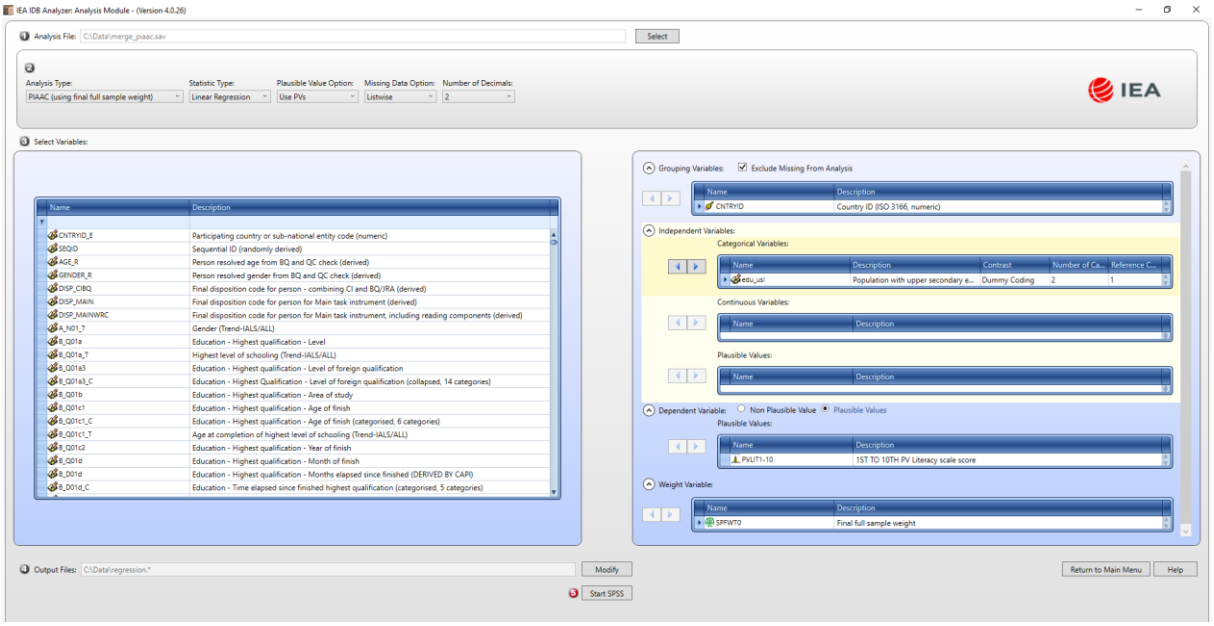


Figure 6.15 Regression setup

14. Click the Start SPSS button. This action will open SPSS and create the syntax to run the regression model.
15. To execute the generated syntax, select all the written commands in the syntax editor, and run these commands using the “Run Selection” button. Alternatively, press CTRL+A, to select all the commands, and then press CTRL+R to execute the syntax. This action would make SPSS run the regression analysis.

Because this analysis involves plausible values, it may take considerably longer in comparison to examples without plausible values in their calculations. This is because the regression analysis needs to be computed for each plausible value once, and then these results are synthetically presented using Rubin-Shaffer rules (Rutkowski et al., 2010). As such, this routine takes may take ten times more than a regression analysis without the use of plausible values.
Once the regression analysis is done, SPSS will present the results in its output window. Figure 6.16 depicts how these results are displayed.

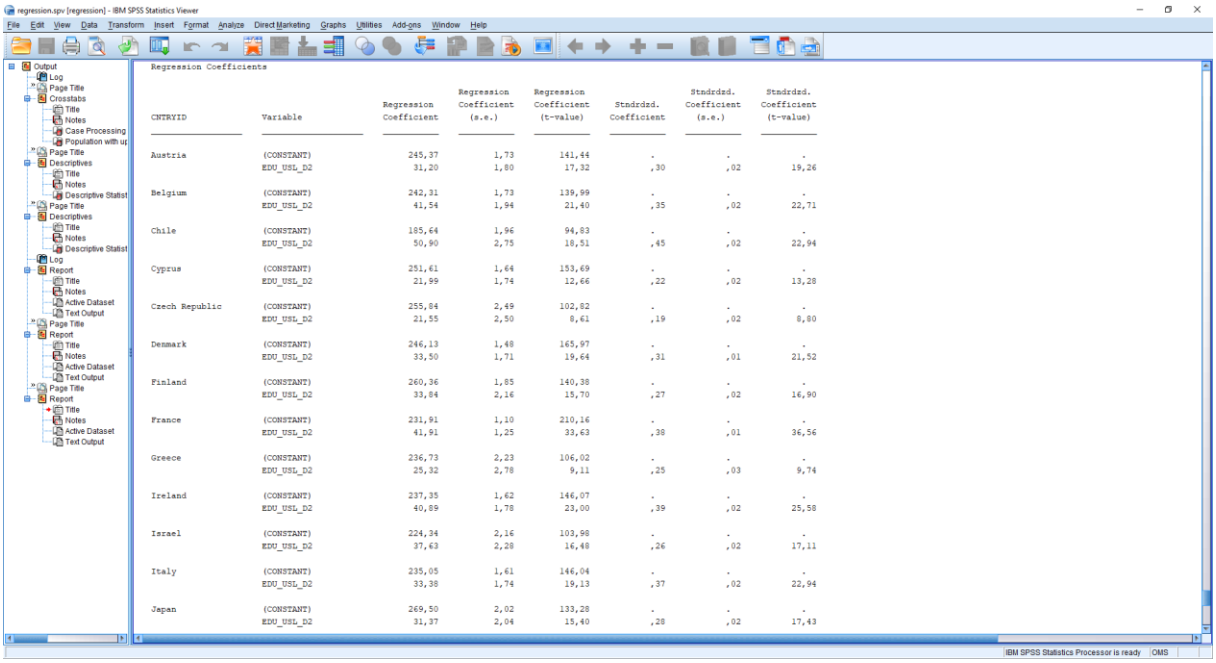


Figure 6.16 Regression output

Once the analysis is concluded, IDB Analyzer will generate 8 files. These files include the syntax, the output, the model fit, the coefficients of the regression and the descriptives of the included variables in the model. In Table 6-9 is the list of the 8 generated files and a description of their contents.

Table 6-9 Generated files by a regression analysis

Generated files	File type	Content
regression.sps	SPSS	Syntax to run the regression analysis
regression.spv	SPSS	Output of the regression analysis
regression_Model.sav regression_Model.xlsx	SPSS Excel	Contains r-square of the regression model, and its adjusted r-squared.
regression_Coef.sav regression_Coef.xlsx	SPSS Excel	Contains unstandardized and standardized regression coefficients, their standard errors.
regression_Desc.sav regression_Desc.xlsx	SPSS Excel	Contains descriptive statistics for all the variables included in the regression model. These include means, standard deviations and variances.

Using the estimates present in "regression_Coef.xlsx" and in "regression_Model.xlsx", we created Table 6-10, to glance the general results of the fitted model. These results are ranked in descending order using the R^2, a measure of explained variance (see, for example, Field, A., 2013 for more information about regression analysis).

Table 6-10 Standardized Regression coefficients and model fit

Countries	Standardized Estimate	Standard Error	R^2
Singapore	0.55	0.01	0.30
Spain	0.46	0.01	0.21
Chile	0.45	0.02	0.20
Netherlands	0.42	0.02	0.18
Ireland	0.39	0.02	0.15
United Kingdom	0.39	0.01	0.15
Turkey	0.39	0.02	0.15
France	0.38	0.01	0.15
Italy	0.37	0.02	0.14
New Zealand	0.37	0.02	0.13
Korea, Republic of	0.36	0.02	0.13
Belgium	0.35	0.02	0.12
Sweden	0.35	0.01	0.12
United States	0.33	0.01	0.11
Slovak Republic	0.33	0.02	0.11
Denmark	0.31	0.01	0.10
Austria	0.30	0.02	0.09
Norway	0.30	0.01	0.09
Slovenia	0.29	0.02	0.09
Japan	0.28	0.02	0.08
Finland	0.27	0.02	0.07
Israel	0.26	0.02	0.07
Greece	0.25	0.03	0.06
Cyprus	0.22	0.02	0.05
Czech Republic	0.19	0.02	0.04
Poland	0.16	0.02	0.03
Russian Federation	0.15	0.03	0.02
Lithuania	0.06	0.02	0.00
Table Average	0.32	0.00	0.11

Inspecting the regression coefficients present in “regression_Coef.xlsx” and their t values, we can conclude all estimated differences are above the sampling error, all beta.t are larger than two. Thus, in all countries, those who hold at least upper secondary education obtain higher literacy scores in the PIAAC test. The average difference of all participating countries is of .32

standard deviations of literacy scores. The estimated gap varies between countries. For example, in Singapore, Spain and Chile is larger than .45 standard deviations. In contrast, in Lithuania, the Russian Federation and Poland this difference is less than or equal to .16 standard deviations of literacy scores.

6.3.6. Correlations

In the PIAAC study, Literacy, Numeracy and Problem-solving in technology-rich environments were measured. How these different skills are related to each other? That is, to what extent these two variables fluctuate together? In the OECD (2016a) report, "Skills Matter", these were reported as highly and positively correlated, with correlations of .86 for Literacy and Numeracy for the OECD partners (see, for example, Field, 2013 for more information about correlation analysis). In the following example, we will estimate the correlation between proficiency in literacy, numeracy and problem-solving in rich environments. To compute these correlations, we need to follow the next steps:

1. Open the Analysis Module of the IDB Analyzer.
2. For this example, specify the data file "merge_piaac.sav" as the Analysis File (see section 6.2 in this chapter for the details on how this file was created).
3. Select "PIAAC (using final full sample weight)" as the Analysis Type.
4. Select "Correlations" as the Statistic Type.
5. Under the "Plausible Values Options" select "Use PVs".
6. Under the "Missing Data option" select "Pairwise".
7. Click on the "Plausible Values" section at the right-hand side of the software window.
8. Go to the "Select variables" section and select the three plausible values variables.
9. Move all the selected variables, by clicking the right arrow in the righthand side window under the "Plausible Values" subsection.
10. Specify the name and the folder of the output files in the Output Files field by clicking the Define/Modify button. In this example, we define the syntax as "correlation".

The final setup should resemble the presented setup in Figure 6.17.

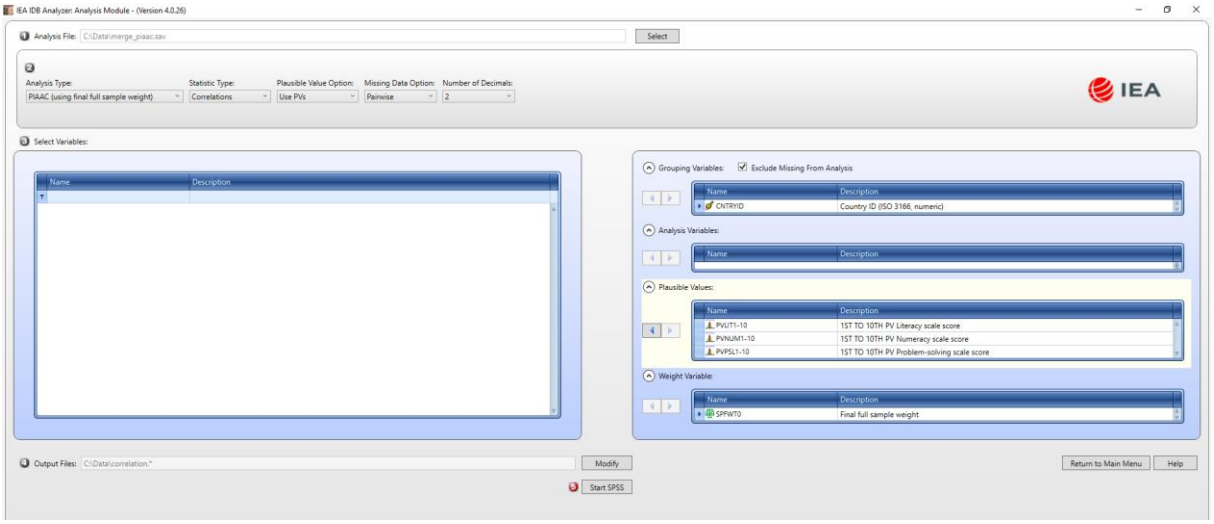


Figure 6.17 Correlation setup

11. Then, click the Start SPSS button. This will create an SPSS syntax file and open it in an SPSS editor window.
12. To start the computations, one needs to press the following keys combinations. CTRL+A first, to select the entire generated code present in the syntax window, and then CTRL+R to run these commands. The output of these analyses is depicted in Figure 6.18.

Because these computations involve the plausible values of the three proficiency scores, its estimation will take longer in comparison to correlations between variables with no plausible values. When the computations are done, six files are generated. These files are described in

Table 6-11.

Table 6-11 Generated files by a correlation analysis

Generated files	File type	Content
correlation.spv	SPSS	Syntax to run the correlation analysis
correlation.sps	SPSS	Output of the correlation analysis
correlation_Corr.sav correlation_Corr.xlsx	SPSS Excel	Contains the correlation estimates and their standard errors.
correlation_Desc.sav correlation_Desc.xlsx	SPSS Excel	Contains descriptive for all the variables included in the correlation analysis. These include means, standard deviations and variances.

The output of these computations is displayed in Figure 6.18.

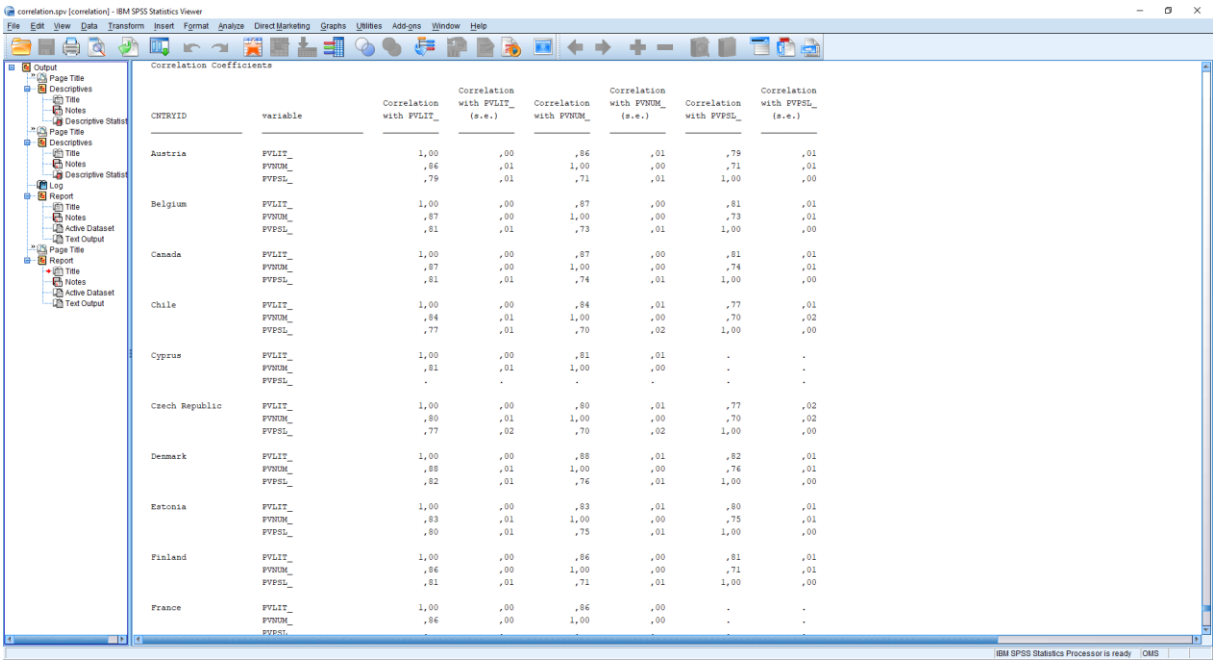


Figure 6.18 Correlation output

These results match those shown in Table A2.7 of the report Skills Matter: Further Results from the Survey of Adult Skills (OECD, 2016a). In Table 6-12, we include only the matching countries from the OECD report and the countries present in the current merge file. Thus, the correlations from Australia, Northern Ireland and Jakarta (Indonesia) are excluded in the present table.

Table 6-12 Correlations between literacy and numeracy scale scores

	OECD estimates	IDB Estimates
Austria	0.86	0.86
Belgium	0.87	0.87
Canada	0.87	0.87
Chile	0.84	0.84
Cyprus	0.81	0.81
Czech Republic	0.80	0.80
Denmark	0.88	0.88
Estonia	0.83	0.83
Finland	0.86	0.86
France	0.86	0.86
Germany	0.87	0.87
Greece	0.81	0.81
Ireland	0.87	0.87
Israel	0.86	0.86
Italy	0.83	0.83
Japan	0.86	0.86
Korea, Republic of	0.88	0.88
Lithuania	0.84	0.84
Netherlands	0.89	0.89
New Zealand	0.87	0.87
Norway	0.89	0.89
Poland	0.85	0.85
Russian Federation	0.79	0.79
Singapore	0.93	0.93
Slovak Republic	0.85	0.85
Slovenia	0.88	0.88
Spain	0.89	0.89
Sweden	0.89	0.89
Turkey	0.85	0.85
United Kingdom	0.87	0.87
United States	0.89	0.89

6.3.7. Proficiency Levels

PIAAC study present proficiency levels, these are segments of scores used to describe the skills of literacy, numeracy and problem-solving in technology-rich environments, at different levels of ability. These are ranges of scores to describe in qualitative terms what participants can do at different levels of proficiency. In general terms, those participants with higher scores in each domain, are more likely to resolve more difficult task, than their counterparts with lower scores (OECD, 2016a).

Literacy scale scores have six proficiency levels. These proficiency levels are briefly described in Table 6-13, and more details can be found in "The Survey of Adults Skills. Reader's Companion" (OECD, 2016b).

Table 6-13 Levels of proficiency

Level	Cut scores	Brief descriptions of more likely resolved tasks
Below level 1	Below 176	Basic reading comprehension with a basic vocabulary.
Level 1	From 176 to Below 226	Reading tasks resolve at this level includes the integration of information, using identical or synonymous terms.
Level 2	From 226 to Below 276	Reading task from this level requires the integration of information of similar meaning, via low inference or paraphrase, and discerning competing information.
Level 3	From 276 to Below 326	Readings tasks required the participant to read through larger pieces of text and construct meaning across paragraphs.
Level 4	From 326 to Below 376	Reading task from this level, require complex inferences and application of background knowledge. The participants need to evaluate subtle evidence-claims or persuasive discourse relationships.
Level 5	At or Above 376	Tasks may require the respondent to search for information and integrate information of similar and contrasting ideas, points of view or evaluate evidence-based arguments. Evaluating the reliability of evidentiary sources and selecting key information is frequently a requirement.

1. Open the Analysis Module of the IDB Analyzer.
2. For this example, specify the data file "merge_piaac.sav" as the Analysis File (see section 6.2 in this chapter for the details on how this file was created).
3. Select "PIAAC (using final full sample weight)" as the Analysis Type.
4. Select "Benchmarks" as the Statistic Type.
5. Under the "Benchmarks Options" select "Discrete". This option will retrieve what proportion of the population falls within each proficiency level. Other options include:

"Cumulative", which compute the proportion of people at or above the cut score; "Discrete with analysis variables" option permits the user to calculates the mean of an analysis variable for those within each proficiency level. For this example, we will use the "Discrete" option.

6. Click on the "Plausible Values" section at the right-hand side of the software window.
7. Move the cursor to the left-hand side of the window and click on the "PVLIT1-10" variable to select the literacy scores.
8. Move the selected variable, by clicking in the right arrow in the right-hand side window, under the "Plausible Values" subsection.
9. Under the "Achievement Benchmarks" section, select the corresponding scores for the Literacy scores, these are "176 226 276 326 376".
10. Specify the name and the folder of the output files in the Output Files field by clicking the Define/Modify button. Here we define the syntax as "benchmark".

The setup of this analysis is depicted in Figure 6.19.

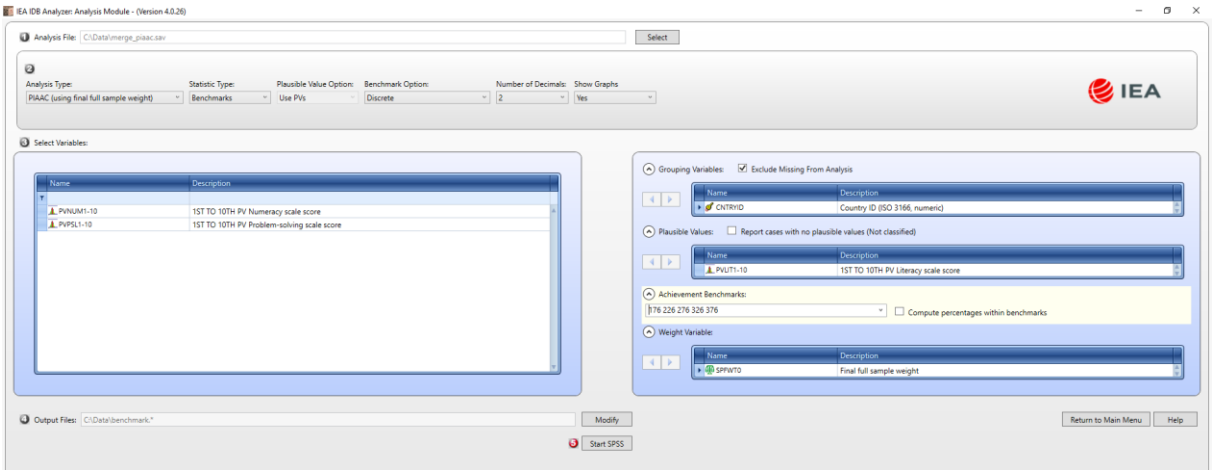
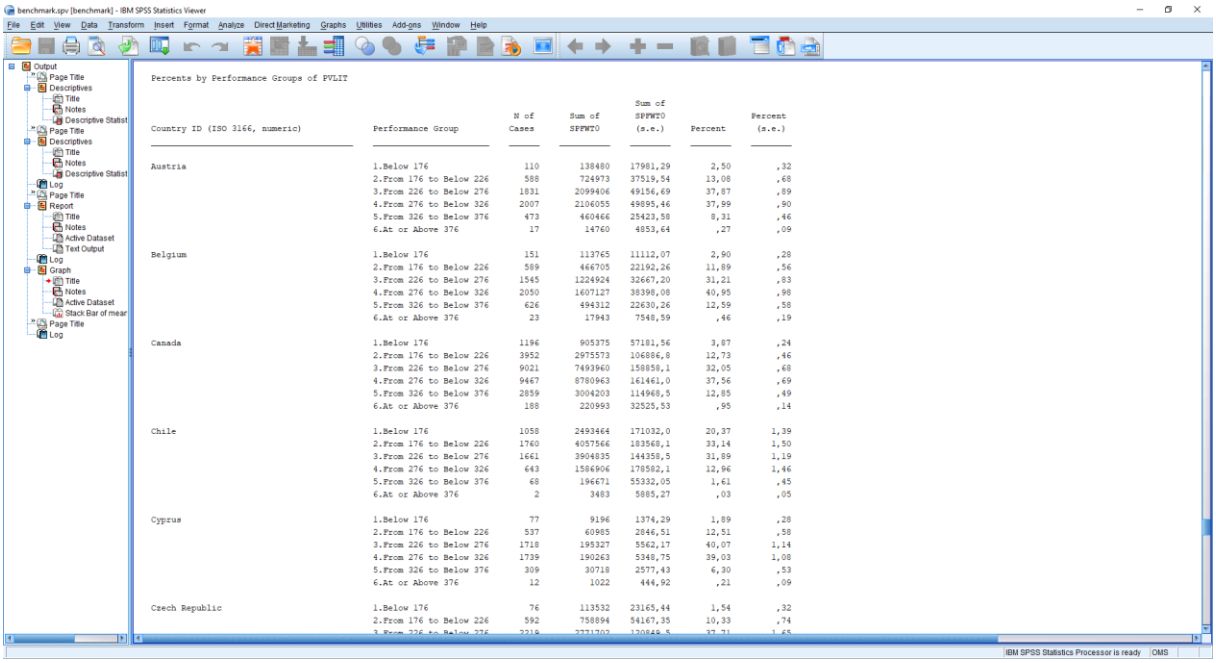


Figure 6.19 Benchmark setup

11. Click the Start SPSS button. This action will open SPSS and create the syntax to run the regression model.
12. To execute the generated syntax press CTRL+A, to select all the commands, and then press CTRL+R to execute the syntax. Now, SPSS will compute the proportion of case at each benchmark.

Results are displayed in Figure 6.20, as they appear in SPSS.



Percents by Performance Groups of PVLIT

Country ID (ISO 3166, numeric)	Performance Group	N of Cases	Sum of SPFWT0	Sum of SPFWT0 (s.e.)	Percent	Percent (s.e.)
Austria	1.Below 176	110	138480	17981,29	2,50	,32
	2.From 176 to Below 226	588	724973	37519,54	13,08	,68
	3.From 226 to Below 276	1831	2099406	49156,69	37,87	,89
	4.From 276 to Below 326	2007	2106055	49895,46	37,99	,90
	5.From 326 to Below 376	473	460466	25423,58	8,31	,46
	6.At or Above 376	17	14760	4853,64	,27	,09
Belgium	1.Below 176	151	113765	11112,07	2,90	,28
	2.From 176 to Below 226	589	466705	22192,26	11,89	,56
	3.From 226 to Below 276	1545	1224924	32667,20	31,21	,83
	4.From 276 to Below 326	2050	1607127	38398,08	40,95	,98
	5.From 326 to Below 376	626	494312	22630,26	12,59	,58
	6.At or Above 376	23	17943	7548,59	,46	,19
Canada	1.Below 176	1196	905375	57101,56	3,87	,24
	2.From 176 to Below 226	3952	2975573	106886,8	12,73	,46
	3.From 226 to Below 276	9021	7493960	158858,1	32,05	,68
	4.From 276 to Below 326	9467	8780963	161461,0	37,56	,69
	5.From 326 to Below 376	2859	3004203	114968,5	12,85	,49
	6.At or Above 376	188	220993	32525,53	,95	,14
Chile	1.Below 176	1058	2493464	171032,0	20,37	1,39
	2.From 176 to Below 226	1760	4057566	183568,1	33,14	1,50
	3.From 226 to Below 276	1661	3904835	144358,5	31,89	1,19
	4.From 276 to Below 326	643	1586906	178582,1	12,96	1,46
	5.From 326 to Below 376	68	196671	55332,05	1,61	,45
	6.At or Above 376	2	3483	5885,27	,03	,05
Cyprus	1.Below 176	77	9196	1374,29	1,89	,28
	2.From 176 to Below 226	537	60985	2846,51	12,51	,58
	3.From 226 to Below 276	1718	195327	5562,17	40,07	1,14
	4.From 276 to Below 326	1739	190263	5348,75	39,03	1,08
	5.From 326 to Below 376	309	30718	2577,43	6,30	,53
	6.At or Above 376	12	1022	444,92	,21	,09
Czech Republic	1.Below 176	76	113532	23165,44	1,54	,32
	2.From 176 to Below 226	592	758894	54167,35	10,33	,74

Figure 6.20 Benchmark output

What do these results mean? We need to consider the procedure the benchmark routine is doing to explain this output. Each cut score is the lower bound value for each defined range (IEA, 2019). We used the following cut scores "176 226 276 326 376". Thus, it computes all the cases below 176 points, all the cases between 176-226, between 226-276, between 276-326, between 326-376, and finally all the cases above 376. In the last two columns of the output, the estimates are the percentage of cases and their standard errors that fall into the specified ranges. This procedure generates the following files, in the specified location (see Table 6-14).

Table 6-14 Generated files by a correlation analysis

Generated files	File type	Content
benchmark.spv	SPSS	Syntax to run the benchmark computations
benchmark.sps	SPSS	Output of the benchmark computations
benchmark.sav	SPSS	Contains the benchmark estimates and their standard errors.
benchmark.xlsx	Excel	
benchmark_by_CNTRYID_Sig.sav	SPSS	Contains a country by country comparison for the percentages at each benchmark.
benchmark_by_CNTRYID_Sig.xlsx	Excel	

Using the information contained in "benchmark.xlsx", we created **Error! Reference source not found.**. This table displays the proportions of participants who performed below level 1 in the literacy proficiency scale.

Table 6-15 Percentage of participants below proficiency level 1

Country	Estimate	Standard Error	Country	Estimate	Standard Error
Chile	20.37	1.39	United Kingdom	3.32	0.38
Turkey	12.92	0.85	Norway	3.05	0.3
Singapore	10.21	0.39	Belgium	2.9	0.28
Israel	8.25	0.41	Finland	2.66	0.23
Spain	7.27	0.47	Netherlands	2.62	0.27
Slovenia	6.01	0.41	New Zealand	2.57	0.28
Italy	5.56	0.57	Austria	2.5	0.32
France	5.37	0.31	Lithuania	2.26	0.38
Greece	4.95	0.53	Korea, Republic of	2.23	0.2
Ireland	4.3	0.43	Estonia	2.02	0.19
United States	4.09	0.47	Slovak Republic	1.89	0.24
Poland	3.94	0.32	Cyprus	1.89	0.28
Canada	3.87	0.24	Russian Federation	1.56	0.54
Denmark	3.83	0.29	Czech Republic	1.54	0.32
Sweden	3.68	0.33	Japan	0.57	0.15
Germany	3.33	0.39	Table Average	4.57	0.08

In Chile, 20% are below level 1, whereas in Japan, Czech Republic, Russian Federation, Cyprus, and the Slovak Republic there are fewer than 2% below level 1.

6.4. Concluding Remarks

In this chapter, we demonstrated how to perform both simple and complex analysis with PIAAC data using the IEA International Database (IDB) Analyzer. We showed examples of how to combine datasets from more than one country into a single data file for cross-country analyses. We also described and illustrated in a step-by-step fashion how to run descriptive statistical analyses including means, percentiles and percentages; as well as inferential analyses such as correlations and regressions.

All the examples included in this chapter used data from OECD's PIAAC, but it is important to mention that the IDB Analyzer can be used to analyse not only PIAAC data but many other international large-scale assessments, such as the OECD's Programme for International Student Assessment (PISA) and Teaching and Learning International Survey (TALIS), as well as the IEA's Trends in International Mathematics and Science Study (TIMSS), Progress

in International Reading Literacy Study (PIRLS), International Civic and Citizenship Study (ICCS), among several others.

The IDB Analyser is certainly not the only tool available to obtain correct estimates when analyzing PIAAC data, but it is probably the most user-friendly one. As mentioned before, the IDB Analyzer is a windows-based tool that creates SAS code or SPSS syntax to perform analysis with PIAAC data. The code or syntax generated by the IDB Analyzer automatically takes into account the complex sample (e.g. sampling weights, replicate weights) and complex assessment design (e.g.plausible values) of PIAAC to compute analyses with the correct standard errors. It enables researchers to test statistical hypotheses in the population without having to write any programming code.

References

Field, A. (2013). *Discovering Statistics Using IBM SPSS.* London: Sage Publications.

Heeringa, S. G., West, B., & Berglund, P. A. (2009). *Applied Survey Data Analysis*. Boca Raton, London, New York: Taylor & Francis Group.

IBM. (2013). *IBM SPSS Statistics (Version 22.0)*. Somers, NY: IBM corporation.

IEA. (2019). *Help Manual for the IEA IDB Analyzer (Version 4.0)*. Hamburg, Germany. Retrieved from www.iea.nl/data.htm

OECD. (2015). Codebook for derived variables for PIAAC public database (with SAS code). Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/skills/piaac/codebook for DVs 3_16 March 2015.docx

OECD. (2016a). *Skills Matter: Further Results from the Survey of Adult Skills*. Paris, France: OECD Publishing. https://doi.org/10.1787/9789264258051-en

OECD. (2016b). *The survey of adult skills: Reader's companion*. Paris: OECD Publishing. https://doi.org/10.1787/9789264204256-en

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher*, *39*(2), 142–151. https://doi.org/10.3102/0013189X10363170

SAS. (2012). *SAS System for Windows (Version 9.4)*. Cary, NC: SAS Institute.

Wilcox, R. R. (2017). *Understanding and applying basic statistical methods using R*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, *31*(2–3), 114–128. https://doi.org/10.1016/j.stueduc.2005.05.005