

Citation for published version: Sandoval-Hernández, A & Carrasco, D 2019, Analysing PIAAC data with IDB Analyzer (SPSS). in DB Maehler & B Rammstedt (eds), Large-Scale Cognitive Assessment: Analysing PIAAC Data. Methodology of Educational Measurement and Assessment, Springer, Cham.

Publication date: 2019

Document Version Peer reviewed version

Link to publication

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

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

Author's information:

Andres Sandoval-Hernandez, University of Bath, <u>A.Sandoval@bath.ac.uk</u> Diego Carrasco, Centro de Medición MIDE UC, Pontificia Universidad Católica de Chile, <u>dacarras@uc.cl</u>

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	. Ana	lysing 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	
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 software¹. 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 launched², 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: <u>https://www.iea.nl/training#IDB_Analyzer_Video_Tutorials</u> 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 Windows³ 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 website⁴. 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 (<u>http://www.parallels.com/products/desktop/</u>).

⁴ <u>http://www.oecd.org/skills/piaac/</u>

- 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.

ata Files and Participan	nts Select File Types and Variables		
ielect Directory			
ata\PIAAC	Select		
t Study:	Select Cycle:		
×C ×	PIAAC Cycle 1		
	an.	Extended Destruction (D)	
valiable Participants:	(21)	selected hartscipants: (0)	
Code	Name	Code Name	
т		T	
N 🖉 AUT			
BEL	Beloium		
22 CAN	Canada		
СНІ	Chile		
20 CVP	Current		
2 CZE	Creck Beoublic		
20 DEU	Garmany		
20 DNK	Danmark		
20 ECD	Sasia		
20 EST	Setonia Estonia		
20 FIN	Sichard		
20 FRA	Frindrig		
2 CBR	France United Kineders	Edit Country List	
a car	Create	Lan County Con	
20 ID1	Greece		
20 ICP	ireiand .		
	isidei		
AT ION	itary lane		
1 KOR	Japan Kono Bouchto ef		
2 ITU	Norea, Republic or		
	Unhuania		
	1947579871087523		

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.

t Data Files and Participants Select File Types and Variables				
Select Directory CountPAAC Lect Study: Select Cycle AAC * PAAC Cycle 1 *	Select			<mark>ể</mark> IEA
Code Name		Sector Faces	Nume Austria Austria Beigum Canada Chila Canada Chila Cipan Canada Cipan Canada Cipan Cipa	
L		, And	Lithuania Netherlands	Return to Main Menu

Figure 6.3 IDB Analyzer Merge Module: Selecting all countries

- 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 selected⁵. 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.

ectrine types 😈 34	elect variables					
Ava	ailable Variables:		Se	lected Variables:		
ral Response File	Background Variables and Scores (1213)	ID and Sampling Variables (0)	ſ	Background Variables and Score	rs (3) ID and Sampling Variables (85)	
	Name	Description		Name	Description	
	*			1		
	CNTRYID_E	Participating country or sub-national entity code (numeric)		 A PVL/T1-10 	1ST TO 10TH PV Literacy scale score	
	SEQID	Sequential ID (randomly derived)		PVNUM1-10	1ST TO 10TH PV Numeracy scale score	
	AGE_R	Person resolved age from BQ and QC check (derived)		PVPSL1-10	1ST TO 10TH PV Problem-solving scale score	
	GENDER_R	Person resolved gender from BQ and QC check (derived)				
	BISP_CIBQ	Final disposition code for person - combining CI and BQ/JRA (derived)				
	BISP_MAIN	Final disposition code for person for Main task instrument (derived)				
	BISP_MAINWRC	Final disposition code for person for Main task instrument, including reading components (de				
	4 A_N01_T	Gender (Trend-IALS/ALL)				
	4 8_Q01a	Education - Highest qualification - Level	1 × 1			
		Highest level of schooling (Trend-IALS/ALL)	M			
	4 8_Q01a3	Education - Highest qualification - Level of foreign qualification				
	4 8_Q01a3_C	Education - Highest Qualification - Level of foreign qualification (collapsed, 14 categories)	14			
	4 8_Q01b	Education - Highest qualification - Area of study				
	4 B_Q01c1	Education - Highest qualification - Age of finish				
	@ 0_Q01c1_C	Education - Highest qualification - Age of finish (categorised, 6 categories)				
	38_001c1_T	Age at completion of highest level of schooling (Trend-IALS/ALL)				
	4 8_Q01c2	Education - Highest qualification - Year of finish				
	4 8_Q01d	Education - Highest qualification - Month of finish				
	4 8_D01d	Education - Highest qualification - Months elapsed since finished (DERIVED BY CAPI)				
	6_D01d_C	Education - Time elapsed since finished highest qualification (categorised, 5 categories)				
	∠ 8_Q02a	Education - Current qualification				
	45002a T1	Education or training in last 12 months (Trend-IALS/ALL)				
	2 8_Q02a_T2	Courses toward certificate, diploma, or degree in program of studies in last 12 months (Trend				
	1 AR 1 0000	Education Connet mulfication Land		L		

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.

lect File Types 🧕 Select Variables		🤘 IEA
Available Variables:	Selected Variables:	
eral Response File Background Variables and Scores (0) ID and Sampling Variables (0)	Background Variables and Scores (121	0 ID and Sampling Variables (85)
Name Description	Name	Description
×		
	28 SOD	Sequential ID (randomly derived)
	all act a	Berron resolved and from BO and OC shark (derived)
	GENDER R	Person resolved gender from 80 and 00 check (derived)
	AS DISP CIRO	Final disposition code for person - combining CI and RO/IRA (derived)
	DISP MAIN	Final disposition code for person for Main task instrument (derived)
	B DISP_MAINWRC	Final disposition code for person for Main task instrument including reading components (d.
	A NOT T	Gender (Trend-IALS/ALL)
	26 a. gota	Education - Highest gualification - Level
	H SciQota,T	Highest level of schooling (Trend-IALS/ALL)
	🗸 🥔 8_Q01a3	Education - Highest gualification - Level of foreign gualification
	4 B_Q01a3_C	Education - Highest Qualification - Level of foreign qualification (collapsed, 14 categories)
	3 8 Q01b	Education - Highest qualification - Area of study
	🚜 B_QQ1c1	Education - Highest qualification - Age of finish
	A gotel_c	Education - Highest qualification - Age of finish (categorised, 6 categories)
	₿ _Q01c1_T	Age at completion of highest level of schooling (Trend-IALS/ALL)
	🦓 8_Q01c2	Education - Highest qualification - Year of finish
	a s cond	Education - Highest qualification - Month of finish
	🤣 B_D01d	Education - Highest qualification - Months elapsed since finished (DERIVED BY CAPI)
	38_D014_C	Education - Time elapsed since finished highest qualification (categorised, 5 categories)
	🥵 8_Q02a	Education - Current qualification
		Education or training in last 12 months (Trend-IALS/ALL)
	6_Q02a_T2	Courses toward certificate, diploma, or degree in program of studies in last 12 months (Trend.
		Education Current nuclification I and
Output Files:	Define < Back	Return to Main Menu

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.
- 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".

nalysis File: C\Data\merge_piaac.sav			Select	
lysis Type: AC (using final full sample weight)	Statistic Type Plausble Value Option Number of Decimals: Show Graph • Precentages and Means: Use PN • 2 • Net	v		😂 IEA
flect Variables:	Description		Grouping Variables: Exclude Missing From Analysis Indexness	
T PVNUM1-10	1ST TO 10TH PV Numeracy scale score 1ST TO 10TH PV Problem-solving scale score		Cutterio Country ID (ISO 3166, numeric) Separate Tables by:	
			Name Description Geneticate Geneticate Person resolved gender from BQ and QC check Plaucible Values:	(derived)
			Posciption Posciption Posciption STID 10TH PV Literacy scale score	
		Y	Weight Variable Name Description Name Description Find full sample weight	
utput Files: C\Data\means_with_pv.		Modify		Return to Main Menu H

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.

a means_with_pv.spv [means_with_pv]	- IBM SPSS Statistics Viewer												- 0
File Edit View Data Transfor	m Insert Format Analyze Direct Marketing Graphs	Utilities Add-ons Window Help											
🧁 🔳 🖨 🙇 🤌) 🛄 🖛 🛥 🧱 🖿 📥 🗐 (🦦 🌭 🦆 📳 🗎) 💫 📧	• •	+ -		1	2					
Output Page Title	u												
Descriptives Descriptives	Average for PVLIT by CNTRYID GENDER R												
Notes		Received and second											
* A Page Title		from BO and OC check	N of	Sum of	sum or		Dercent	0.077.7.72	DOT TO		Std Dev		
Descriptives	Country ID (ISO 3166, numeric)	(derived)	Cases	SPFWT0	(8.0.)	Percent	(s.e.)	(Mean)	(8.8.)	Std.Dev	(s.e.)	potmiss	
- Im Title - B Notes													
Descriptive Statist													
Log	Austria	Male	2479	2764088	7207,03	49,86	,10	271,53	1,04	44,64	, 86	1,83	
- S Report		remaie	2046	2760051	6922,30	50,14	,10	267,39	,93	43,17	,75	1,02	
- Ittle - Notes	Belgium	Male	2467	1984965	6758,70	50,58	,13	278,09	,97	47,91	,86	5,01	
+ [] Text Output		Female	2517	1939811	7650,74	49,42	,13	272,81	1,08	46,07	,89	5,30	
- I Log	Canada	Male	12442	11684548	00	49 97	00	274 49	86	50.99	75	00	
- @ Title	Called a	Female	14241	11696519	,00	50,03	,00	272,19	,78	49,84	, 64	,00	
Notes													
Graph	Chile	Male	2189	6117483	20983,69	49,97	,10	223,94	2,48	53,25	1,57	,23	
Title		Female	3003	6125441	14808,46	50,03	,10	216,36	2,77	51,76	1,22	, 31	
Bar of mean(mnp)	Cyprus	Male	1776	230956	2355,68	47,37	, 39	267,99	1,18	41,00	, 90	19,53	
🖶 🛃 Graph		Female	2616	256555	2061,65	52,63	, 39	269,60	, 97	39,58	,78	15,96	
- In Notes													
Stack Bar of mean	Czech Republic	Male Remain	2756	3706815	9665,44	50,44	,10	275,68	1,26	40,83	1,06	, 57	
Page Title		remare	3323	3642300	10910,40	49,00	,10	212,32	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	
	Patonia	Mala	2422	427025	410 01	47.02	02	275.06	1.09	45 44	60	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	
		LENGTE	a/30	20103/01	19006,02	49,49	,14	601,61	1,19	40,90	,07	1,70	
	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.

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	SPSS	Contains the means estimates and their
means_with_pvGENDER_R.xlsx	Excel	standard errors.
means_with_pv_PVLIT_by_GENDER_R_Sig.	SPSS	Contains a group within-country
sav		comparison for the estimated means and
means_with_pv_PVLIT_by_GENDER_R_Sig.	Excel	percentages, providing t statistics for
xlsx		these comparisons.

Table 6-1 Generated files by an analysis of means

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.

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)

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

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:

Analysis File: C\Data\merge_piaac.sav			Select
Analysis Type: PIAAC (using final full sample weight) v	Statistic Type: Plausible Value Option: Number of Decimals: Show Graphs Percentages and Means ** [None Used **] [2 **] [165 **		送 IEA
Select Variables:			
Nane •	Perception *		Orouging Variables Disclude Missing From Analysis Image: Chromo Country U (50:0) 1560 1566 numeric) O Separate Tables by: Image: Chromo Country U (50:0) 1560 1566 numeric) O Analysis Variables Image: Chromo Country U (50:0) 1560 1566 numeric)
Contains[[Description], total years] •		Modify	Weight Variable Weight Variable Weight Variable Weight Variable Prent full sample weight Return to Main Menu Return to Main Menu Prent Variable Prent Variable Prent Variable Return to Main Menu Prent Variable Prent Variable

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.

at View Data Transfo	vm Insert Format Analyze Direct Marketing Graphs	Utilities Add-or	is <u>Window H</u> elp	🔈 🔳 🗸	⊨ → -	+ - 1	R 10 T	1 🕐 👌				
utput Page Title Descriptives	Average for YRSQUAL_T by (CNTRYID)											
C Title Notes Warnings Descriptive Statist	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	
Descriptives												
Title R Notes	Austria										100,00	
- KB Warnings	Belgium	4978	3920642,64	10465,80	, 57	,00	12,34	,03	2,93	,03	5,25	
- Descriptive Statist	Canada	26472	23217517,16	18460,30	3,37	,01	13,21	,01	2,72	,02	,70	
🖺 Log	Chile	5189	12235925,59	26747,59	1,78	,01	11,46	,19	3,34	,07	, 33	
Page Title	Cyprus	4392	487511,27	2226,60	,07	,00	12,26	,02	3,11	,03	17,69	
Report	Czech Republic	6080	7348569,82	14254,89	1,07	,00	12,95	,02	2,66	,02	,63	
- R Notes	Denmark	7288	3615673,01	2243,77	, 52	,00	12,47	,02	2,72	,02	,37	
- Active Dataset	Estonia	7386	892672,49	572,12	,13	,00	12,06	,03	2,68	,02	, 39	
Text Output	Finland	5164	20007515 15	42120.10	, 31	,00	12,29	,03	3,02	,02	,00	
Log	France	6073	39607313,13	43120,10	5,75	,01	11,10	,02	3,56	,02	100,00	
- P Title	Greece	4017	6007010 50	12222.05	1.02		11 22		2 40		100,00	
- R Notes	Traland	5945	2080023 80	2626 25	43	,00	14 49	02	3,10	02	45	
🕼 Active Dataset	Terael	5330	4701545 33	8480.93	,40	,00	12 61	02	2.86	,02	2 49	
Bar of mean(mrx)	Thely	4500	20112024 12	65970 65	5,60	02	10 51	02	2,00	02	65	
Graph	Japan	5171	80031019.77	97599.31	11.62	.03	12,94	.01	2,41	.02	1.27	
- Ch Notes	Korea, Republic of	6653	34516756.04	24535.06	5.01	.01	12,60	. 02	3,18	.03	. 25	
Active Dataset	Lithuania	5048	1879524.83	6924.06	.27	.00	13.07	.03	2.71	.05	4,51	
- 🛱 Bar of mean(pct) t	Netherlands	5084	10914222,69	18217,40	1,58	.00	13,12	.03	2,79	.03	2,21	
Page Title	New Zealand	6053	2687168.73	6237,61	. 39	.00	13.72	.04	2,52	.02	2,27	
🖆 Log	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.

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

Table 6-3 Generated files by an analysis of means

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

Table 6-4 to present the computed results.

		Standard			Standard
Country	Mean	Error	Country	Mean	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

Table 6-4 Means of lifetime years in schooling

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.
- 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.

x Number of Decinals	Grouping Variables: VI. Exclude VI. Variable Variable Variable Variable Variable Variable Variable	Kising From Analysis
e (rumeric) à	Grouping Variables Grouping Variables	Mising Fran Andysis Decorption
e (numeric)	Grouping Variables: Sclude I	Missing From Analysis Description
e (numeric) 🔶		Description
e (numeric)		Description
e (numeric)	CNTRHD	
rived)	there are a second s	Country ID (ISO 3166, numeric)
rived)	O A LINE I	
	Analysis Variables:	
(derived)	Name	Description
I and BQ/JRA (derived)	A P	Desired variable are total years of extension advice lifetime , too under at 24 (Teach-MI \$/A11)
instrument (derived)	CONTRACTION OF THE STOCK	Derived variable on total years or schooling during litesime - top coded at 24 (ireno-ulls/ALL)
instrument, including reading components (derived)	Percentiles	
	24 60 77	
	23 30 73	
	Weight Variable:	
ign qualification	(0)	
ign qualification (collapsed, 14 categories)	Name	Description
/	> @ SPFWTO	Final full sample weight
(categorised, 6 categories)		
(Trend-IALS/ALL)		
6		
ish		
sed since finished (DERIVED BY CAPI)		
qualification (categorised, 5 categories)		
((15 54 6	categorised, 6 categories) leno-GuSSALL) h di once finished (DERIVED BY CAPI) qualification (categorised, 5 categories) y	categorised 6 categories) Intere 44LS/ALU h h dires Finished (DERVED BY CAP) publication (categorised, 5 categories)

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.

eff up (a) (a) <t< th=""><th>percentile.spv [percentile] - IBM SP</th><th>PSS Statistics Viewer</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th>- 0</th></t<>	percentile.spv [percentile] - IBM SP	PSS Statistics Viewer										- 0
Organization Organization Sector Sector Sector Provide Sector Provide Sector Sector Provide Sector	File Edit View Data Transfo	orm Insert Format Analyze Direct Marketing Graphs	Utilities Add-or	is <u>Window H</u> elp				-	_			
Open Open Provide Formation Provide Series Provide Series <t< th=""><th>2 🗎 🖨 🔌 🚽</th><th>) 📖 🖛 🛥 🧮 🔚 🏭</th><th></th><th>(† 12 B</th><th>h 🔊 🔳</th><th>■ ◆ →</th><th>- + -</th><th></th><th></th><th></th><th></th><th></th></t<>	2 🗎 🖨 🔌 🚽) 📖 🖛 🛥 🧮 🔚 🏭		(† 12 B	h 🔊 🔳	■ ◆ →	- + -					
Image: Second	Output	D										
Bit of the second sec	Descriptives Descriptives Descriptives Descriptives	Percentiles for YRSQUAL_T by CNTRYID										
Bigum Consideration Description Description Description Bigum Casada 24472 2121737 12.00 .00 15.00 .00 Torse 4352 497731 12.00 .00 15.00 .00 Transe 4444 9467731 5.00 .00 11.00 .00 14.00 .00 Transe 4444 9467731 5.00 .00 15.00 .00 15.00 .00 Transe 11201 0.00 11.00 .00 15.00 .00	Warnings Bescriptive Statist Page Title	Country ID (ISO 3166, numeric)	N of Cases	Sum of SPFWT0	p25	p25_se	p50	p50_se	p75	p75_se		
Constance Constance <thconce< th=""> Constance <thconc< th=""><th>Descriptives Title</th><th>Belgium</th><th>4978</th><th>3920643</th><th>12 00</th><th>00</th><th>12 00</th><th>00</th><th>15.00</th><th>00</th><th></th><th></th></thconc<></thconce<>	Descriptives Title	Belgium	4978	3920643	12 00	00	12 00	00	15.00	00		
Los Los <thlos< th=""> <thlos< th=""> <thlo< th=""></thlo<></thlos<></thlos<>	Notes Warpings	Canada	26472	23217517	12,00	.00	13,00	.00	16,00	.00		
Cypes Cypes <th< th=""><th>Descriptive Statist</th><th>Chile</th><th>5189</th><th>12235926</th><th>8,00</th><th>,00</th><th>12,00</th><th>.00</th><th>14,00</th><th>4,47</th><th></th><th></th></th<>	Descriptive Statist	Chile	5189	12235926	8,00	,00	12,00	.00	14,00	4,47		
Provide Term Cach Republic Cach Republic Cach Republic Cach Republic Presents Trans 786 925721 1,00 .00 12,00 .00 14,00 .00 Presents Trans 786 926721 .00 .00 12,00 .00 14,00 .00 Trans 786 926721 .00 .00 12,00 .00 14,00 .00 Trans 786 926721 .00 .00 12,00 .00 14,00 .00 Trans 1203 .00 .00 12,00 .00 12,00 .00 14,00 .00 Trans 1203 .00 .00 12,00 .00 12,00 .00 14,00 .00 Trans 1203 .00 .00 12,00 .00 12,00 .00 12,00 .00 12,00 .00 12,00 .00 12,00 .00 12,00 .00 12,00 .00 12,00	Log	Cyprus	4392	487511	12,00	,00	12,00	,00	14,00	,00		
Press Press Total Total <th< th=""><th>- Page Title</th><th>Czech Republic</th><th>6080</th><th>7348570</th><th>12,00</th><th>,00</th><th>13,00</th><th>,00</th><th>13,00</th><th>,00</th><th></th><th></th></th<>	- Page Title	Czech Republic	6080	7348570	12,00	,00	13,00	,00	13,00	,00		
Face 756 99272 10,00 00 12,00 00 14,00 00 Proce 660 36609 10,00 00 12,00 00 14,00 00 Proce 660 36009 10,00 00 12,00 00 14,00 00 Trace 660 36009 10,00 00 12,00 00 14,00 00 Trace 660 36009 10,00 00 14,00 00 14,00 00 Testand 364 300120 11,00 00 11,00 00 15,00 00 Taty 459 301232 12,00 00 15,00 00 15,00 00 Taty 459 301232 12,00 00 15,00 00 15,00 00 Taty 459 301232 12,00 00 15,00 00 15,00 00 Taty 450 201320 10,00	Report	Denmark	7288	3615673	10,00	,00	12,00	,00	15,00	,00		
Print model Stef4 346600 11,00 .00 12,00 .00 14,00 .00 Prese Tree Greece 412 697315 5,00 .00 11,00 .00 14,00 .00 Treat 533 4701555 12,00 .00 12,00 .00 14,00 .00 Jace 533 4701555 12,00 .00 12,00 .00 13,00 .00 Jace 533 4701555 12,00 .00 12,00 .00 14,00 .00 Jace 533 4701555 12,00 .00 12,00 .00 14,00 .00 Jace 533 4701552 12,00 .00 12,00 .00 14,00 .00 Jace 504 107523 12,00 .00 12,00 .00 15,00 .00 BetherLand 603 247140 .00 12,00 .00 15,00 .00 BetherLand 603 247140 .00 12,00 .00 15,00 .00 Bet	+ Intle	Estonia	7586	892672	10,00	,00	12,00	,00	14,00	,00		
Image for closed France 660 3 900713 5 9,00 11,00 .00 14,00 .00 Irisand 964 507911 5,00 .00 11,00 .00 14,00 .00 Irisand 964 507911 5,00 .00 11,00 .00 15,00 .00 Irisand 964 507911 5,00 .00 11,00 .00 15,00 .00 Irisy 499 9 911343 1,00 .00 10,00 .00 15,00 .00 Irisy 499 9 911343 1,00 .00 10,00 .00 15,00 .00 Irisy 499 9 911342 1,00 .00 11,00 .00 15,00 .00 Pace Tail 601 9 101202 1,00 .00 15,00 .00 15,00 .00 Pace Tail 601 9 101202 1,00 .00 15,00 .00 15,00 .00 Pace Tail 503 207154 1,00 .00 15,00 .00 15,00 .00 Poland 963 2073546 11,00 .00 12,00 .00 15,00	- In Active Dataset	Finland	5464	3496909	11,00	,00	12,00	,00	14,00	,00		
Dresser Openation Openation <tho< th=""><th>Text Output</th><th>France</th><th>6893</th><th>39607515</th><th>9,00</th><th>,00</th><th>11,00</th><th>,00</th><th>14,00</th><th>,00</th><th></th><th></th></tho<>	Text Output	France	6893	39607515	9,00	,00	11,00	,00	14,00	,00		
Ireland 969 200934 11,00 .00 14,00 .00 Ireland 339 4701455 12,00 .00 13,00 .00 Ialy 489 3911224 1,00 .00 12,00 .00 13,00 .00 Kores, Repuise of 653 34911755 12,00 .00 12,00 .00 15,00 .00 Methorizania 5044 137555 12,00 .00 13,00 .00 15,00 .00 New Zeiland 5044 1091222 1,00 .00 14,00 .00 15,00 .00 New Zeiland 5043 320199 12,00 .00 13,00 .00 15,00 .00 New Zeiland 5043 320199 12,00 .00 13,00 .00 13,00 .00 Singeree 5093 279715 10,00 .00 12,00 .00 13,00 .00 Singeree 5093 319702 5,00 .00 13,00 .00 12,00 .00 12,00 .00	🐣 🕰 Page Title	Greece	4917	6997911	9,00	,00	12,00	,00	14,00	,00		
Iracel 333 4701454 12,00 ,00 13,00 ,00 Japas 5171 0701200 12,00 ,00 13,00 ,00 Japas 5171 0701200 12,00 ,00 14,00 ,00 Lithuania 564 117953 12,00 ,00 13,00 ,00 Meterianda 653 247146 12,00 ,00 14,00 ,00 Meterianda 953 279735 10,00 ,00 11,00 ,00 Signation 563 247746 12,00 ,00 11,00 ,00 Signation 563 247748 10,00 ,00 11,00 ,00 Signation 563		Ireland	5965	2980934	11,00	,00	14,00	,00	18,00	,00		
taly 489 9811214 0,00 0,00 10,00 00 dagan 5171 901200 12,00 0.00 12,00 0.00 Kores, Repuise of 653 34911754 12,00 0.00 12,00 0.00 Betherlands 5044 1075255 12,00 0.00 14,00 0.00 New Zeilad 5044 109120221 11,00 0.00 14,00 0.00 New Zeilad 5044 10912223 11,00 0.00 14,00 0.00 New Zeilad 5043 320199 12,00 0.00 15,00 0.00 New Zeilad 5039 2701744 11,00 0.00 12,00 0.00 15,00 0.00 Sinspece 5039 2707144 11,00 0.00 12,00 0.00 11,00 0.00 Siowsk Repulic 503 3600111 12,00 0.00 11,00 0.00 Siowsk Repulic 503 36012311 10,00 0.00 12,00 0.00 Bain 5645 30135212		Israel	5339	4701545	12,00	,00	12,00	,00	15,00	,00		
Japan 51.71 00011200 12,00 0.0 12,00 0.0 14,00 00 Lithazaia 564 137953 12,00 0.0 12,00 0.0 14,00 0.0 Net Kaladia 564 137953 12,00 0.0 13,00 0.0 14,00 0.0 Net Kaladia 653 247164 12,00 0.0 14,00 0.0 14,00 0.0 New Kaladia 653 247164 12,00 0.0 14,00 0.0 14,00 0.0 New Kaladia 953 247164 12,00 0.0 14,00 0.0 14,00 0.0 New Kaladia 953 2471364 11,00 0.0 11,00 0.0 14,00 0.0 Basess Protection 393 397144 11,00 0.0 11,00 0.0 11,00 0.0 11,00 0.0 11,00 0.0 11,00 0.0 11,00 0.0 11,00 0.0 11,00 <		Italy	4589	39112824	8,00	,00	8,00	,00	13,00	,00		
Kores, Republic of Lithosania Kores, Sepublic of Side H416755 12,00 .00 12,00 .00 12,00 .00 Hithoriania 5044 1075255 12,00 .00 13,00 .00 15,00 .00 New Zeiland 5044 10014223 11,00 .00 14,00 .00 15,00 .00 New Zeiland 5053 207149 12,00 .00 14,00 .00 15,00 .00 Russian Federation 3809 7710744 11,00 .00 12,00 .00 13,00 .00 Sinsprore 5939 279713 10,00 .00 13,00 .00 13,00 .00 Sinsprore 5939 219721 50,00 .00 11,00 .00 13,00 .00 Sinsprore 5939 219721 50,00 .00 11,00 .00 11,00 .00 Sinsprore 5934 50,00 .00 12,00 .00 12,00 .00		Japan	5171	80031020	12,00	,00	12,00	,00	14,00	,00		
I tablesting 5040 1079535 12,00 .00 13,00 .00 Methorizanda 5044 1079525 12,00 .00 14,00 .00 New Tealand 653 2671649 12,00 .00 14,00 .00 New Tealand 653 2671649 12,00 .00 14,00 .00 New Tealand 653 2671649 12,00 .00 14,00 .00 New Tealand 963 24713644 .00 .00 12,00 .00 Poland 963 24713644 .00 .00 12,00 .00 Simpport 350 2791744 .00 .00 12,00 .00 Simpport 350 2791744 .00 .00 12,00 .00 Simpport 350 2791744 .00 .00 12,00 .00 Simpport 110 .00 12,00 .00 14,00 .00 <th></th> <th>Korea, Republic of</th> <th>6653</th> <th>34516756</th> <th>12,00</th> <th>,00</th> <th>12,00</th> <th>,00</th> <th>16,00</th> <th>,00</th> <th></th> <th></th>		Korea, Republic of	6653	34516756	12,00	,00	12,00	,00	16,00	,00		
Methorizanda 504 10014223 11,00 .00 14,00 .00 Bwer Zeiland 603 207149 12,00 .00 14,00 .00 Bwer Zeiland 963 207149 12,00 .00 14,00 .00 Buer Zeiland 963 207159 11,00 .00 12,00 .00 15,00 .00 Buestan Federation 380 9710744 11,00 .00 12,00 .00 15,00 .00 Sinspace 593 279713 10,00 .00 12,00 .00 13,00 .00 Sinspace 593 319702 %,00 .00 11,00 .00 11,00 .00 Sinspace 593 319702 %,00 .00 11,00 .00 11,00 .00 Sinspace 594 3033212 10,00 .00 12,00 .00 12,00 .00 Tarkey 514 50023943 5,00 .00 14,00	1	Lithuania	5048	1879525	12,00	,00	13,00	,00	15,00	,00		
Bree Zealand 653 2671569 12,00 .00 13,00 .00 Norway 4561 3210398 12,00 .00 14,00 .00 Poland 943 26715644 11,00 .00 12,00 .00 Bassan Pederation 933 2797313 10,00 .00 12,00 .00 Sinsport 353 2797313 10,00 .00 12,00 .00 Sinsport 353 2797313 10,00 .00 12,00 .00 Sinsport 353 2797313 10,00 .00 12,00 .00 Sinsport 553 2797384 11,00 .00 12,00 .00 Sinsport 553 2797384 10,00 .00 12,00 .00 Sinsport 559789 10,00 .00 12,00 .00 Turkey 5164 5002393 5,00 .00 12,00 .00 United States 428 1662564 12,00 .00 14,00 .00 United States 428 16625644 12,00 .00 14,00 .00 United States 428 16625644 12,00 .00 14,60		Netherlands	5084	10914223	11,00	,00	14,00	,00	16,00	,00		
Norway 4951 3210399 12,00 .00 14,00 .00 Poland 963 2253566 11,00 .00 12,00 .00 15,00 .00 Bussan Pederation 3890 9710744 11,00 .00 12,00 .00 15,00 .00 Singepore 5993 279713 10,00 .00 12,00 .00 13,00 .00 Singepore 5993 319702 %,00 .00 11,00 .00 11,00 .00 Singepore 593 319702 %,00 .00 11,00 .00 11,00 .00 Singepore 593 319702 %,00 .00 12,00 .00 11,00 .00 Singepore 593 319702 %,00 .00 12,00 .00 11,00 .00 Singepore 5194 5190 .00 12,00 .00 12,00 .00 Dated K47 57173 1,00 .00 12,00 .00 12,00 .00 Dated K47 57173 1,00 .00 12,00 .00 14,00 .00 United States 4286 1642564 12,00		New Zealand	6053	2687169	12,00	,00	13,00	,00	16,00	,00		
Poland 963 26735664 11,00 .00 12,00 .00 15,00 .00 Reseatan Poteration 369 2777134 11,00 .00 12,00 .00 13,00 .00 Singapore 3593 2777135 10,00 .00 12,00 .00 13,00 .00 Singapore 3593 2777135 10,00 .00 11,00 .00 11,00 .00 Singapore 3593 2777135 10,00 .00 11,00 .00 11,00 .00 Singapore 3593 2777135 10,00 .00 11,00 .00 11,00 .00 Singapore 3593 2777313 10,00 .00 11,00 .00 11,00 .00 Singapore 1000 .00 11,00 .00 11,00 .00 12,00 .00 Singapore 5164 50063493 5,00 .00 13,00 .00 12,00 .00 United States 4286 16625644 12,00 .00 12,00 .00 14,00 .00 United States 4286 16625644 12,00 .00 12,00 .00 14,69 .15 <th></th> <th>Norway</th> <th>4951</th> <th>3210398</th> <th>12,00</th> <th>,00</th> <th>14,00</th> <th>,00</th> <th>16,00</th> <th>,00</th> <th></th> <th></th>		Norway	4951	3210398	12,00	,00	14,00	,00	16,00	,00		
Russian Pederation 3890 0730744 11,00 .00 12,00 .00 13,00 .00 Singegore 5393 237313 10,00 .00 13,00 .00 Singegore 5393 237313 10,00 .00 13,00 .00 Singerore 5393 237913 51,00 .00 13,00 .00 Singerore 5393 23901 10,00 .00 11,00 .00 Singerore 5393 23901 10,00 .00 11,00 .00 Singerore 5393 5401 10,00 .00 11,00 .00 Singerore 5393 5,00 .00 12,00 .00 14,00 .00 Tarkey 516 50023943 5,00 .00 15,00 .00 .00 United States 4286 16423544 12,00 .00 15,00 .00 United States 4286 16423544 12,00 .00 16,00 </th <th></th> <th>Poland</th> <th>9363</th> <th>26735696</th> <th>11,00</th> <th>,00</th> <th>12,00</th> <th>,00</th> <th>15,00</th> <th>,00</th> <th></th> <th></th>		Poland	9363	26735696	11,00	,00	12,00	,00	15,00	,00		
singspore 599 297313 10,00 .00 12,00 .00 13,00 .00 singspore 500 840111 12,00 .00 11,00 .00 11,00 .00 singspore 5293 1397021 5,00 .00 11,00 .00 11,00 .00 singspore 5293 1397021 5,00 .00 11,00 .00 11,00 .00 singspore 5461 577183 11,00 .00 11,00 .00 11,00 .00 Takis 5461 50624393 5,00 .00 10,00 .00 12,00 .00 United States 4286 16625564 12,00 .00 13,00 .00 16,00 .00 D . . 10,66 .00 12,00 .00 16,00 .00 D . . .066 .00 12,00 .00 15,00 .00 D . . .00 13,00 .00 16,00 .00 D . . .066 .00 12,00 .00 14,69 .15		Russian Federation	3890	87387844	11,00	,00	12,00	,00	18,00	,00		
slowsk Begulaice 502 3460111 12,00 ,00 13,00 ,00 slowsk Begulaice 5292 139702 5,00 ,00 11,00 ,00 gain 5454 319702 10,00 ,00 12,00 ,00 10,00 ,00 Beden 447 5797189 11,00 ,00 12,00 ,00 12,00 ,00 Tarkey 5164 5042943 5,00 ,00 10,00 ,00 12,00 ,00 12,00 ,00 United States 4286 164235644 12,00 ,00 16,00 ,00 Daited States 4286 164235644 12,00 ,00 16,00 ,00 D . . 15,66 ,00 12,03 ,00 14,69 ,15 D Daited States 		Singapore	5393	2797315	10,00	,00	12,00	,00	15,00	,00		
Slovenia 5593 1397021 5,00 .00 11,00 .00 Spice 545 305121 11,00 .00 11,00 .00 Spice 446 305121 11,00 .00 11,00 .00 Spice 446 305121 11,00 .00 11,00 .00 United States 4266 12,00 .00 12,00 .00 16,00 .00 United States 4286 1625564 12,00 .00 16,00 .00 D . .00,66 .00 12,00 .00 16,00 .00 D . .00,66 .00 12,00 .00 16,00 .00 D . .00,66 .00 12,03 .00 16,00 .00 D . .00,66 .00 12,03 .00 16,00 .00		Slovak Republic	5702	3860111	12,00	,00	13,00	,00	13,00	,00		
Spain 5943 30035212 10,00 .00 14,00 .00 Sweden 447 577783 11,00 .00 14,00 .00 Tarksy 5146 50042943 5,00 .00 6,00 .00 United Kingdom 7550 3173544 11,00 .00 11,00 .00 United States 4286 166235644 12,00 .00 16,00 .00 D - .01,66 .00 12,03 .00 16,00 .00 D - .01,66 .00 12,03 .00 14,69 .15		Slovenia	5293	1397023	9,00	,00	11,00	,00	11,00	,00		
Sweden ***/ 5#/7*8* 11,00 .00 12,00 .00 Turkey 5164 50/2493 5,00 .00 .00 12,00 .00 United Kingdom 7550 3177394 11,00 .00 11,00 .00 16,00 .00 United States 4286 16623564 .00 12,03 .00 14,69 .15 D . . 10,66 .00 12,03 .00 14,69 .15 D . . . 10,66 .00 12,03 .00 14,69 .15		spain	5965	30835212	10,00	,00	12,00	,00	14,00	,00		
Tatkey 5,44 30,42,493 5,00 4,00 1,00 ,00 Dailed Kingdom 7550 3173544 11,00 ,00 15,00 ,00 United Kingdom 7450 3173544 11,00 ,00 15,00 ,00 United Kingdom 7450 3173544 11,00 ,00 15,00 ,00 Dailed States 4286 166235644 12,00 ,00 16,69 ,15 D . .00,66 ,00 12,03 ,00 16,69 ,15		Sweden	4467	5979788	11,00	,00	12,00	,00	14,00	,00		
United Kingom /#50 31/3394 11,00 .00 14,00 .00 United States 4284 16623564 .00 12,03 .00 146,69 .15 D . . 10,66 .00 12,03 .00 146,69 .15		Turkey	2196	50062983	5,00	,00	8,00	,00	12,00	,00		
		United Kingdom	/650	31773394	11,00	,00	11,00	,00	16,00	,00		
180.6 average . . 10,66 .00 14,67		United States	1200	100200004	12,00	,00	13,00	,00	16,00	,00		
		Table Average			10,66	,00	12,03	,00	14,09	,15		
D C C C C C C C C C C C C C C C C C C C												
> Image: State Control of S												
> Image: Statute December 10 and												
		>>										
The second secon												
andia Ma												
metring and the second se	18	4										
	formation area										IDM ODOC Claticitics Drop	resear is ready

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	SPSS	Contains the means estimates and their
percentile.xlsx	Excel	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.

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)

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

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

Analysis File: C/\Data\umerge_piaac.sav	Select	
alysis Type: Statistic Type: Number of Decimals: Show Graphs AAC (using final full sample weight) ** (Percentages only **) 2 ** (Yes **)		🌔 IEA
Select Variables:		
Name Decomption 1 upper upper @datual Republicion with upper secondary education (Turyet, Ganc)	Concept Generation and a second former an	K rumeri) eght
Dutput Files: C1.Data/percentage."	Modify	Return to Main Menu
	3 Start SPSS	

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.

	2001	
Type: Statistic Type: Number of Decimals: Show Graphs aring final full sample weight() = Percentages only = 2 = 1968 = =		🌔 IEA
Variables:		
ame Description [†]	Grouping Variables: Reclude Missing From Analysis	
upper	Name Description	
	Country ID (ISO 3166, numeric)	ication (1=yes, 0=no)
	Separate Tables by:	
	Aame Description	
	Weight Variable:	
	Name Description	
ontains([Description], 'upper') •	Final full sample weight	
t Files: C\Data\percentage.*	Modify	Return to Main Menu H
	Character	

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).

Percentage.spv [percentage] - IBM	1 SPSS Statistics Viewer							- 0 ×
<u>File Edit View Data Transf</u>	orm Insert Format Analyze Direct Marketing Graphs	Utilities Add-ons y	lindow <u>H</u> elp					
😑 🔳 🖨 🗟 🥑	👌 📖 🖛 🛥 🧱 📰 📥 🗐 .	💊 🌒 🧦		🔳 💰 ا	+ +	+-		
Output Output Output Output Cog Page Title Report Other Notes	Percentages by (CNTRYID EDU_USL)	Population with upper						
Text Output Graph Graph Text Title	Country ID (ISO 3166, numeric)	<pre>secondary education (1=yes, 0=no)</pre>	N of Cases	Sum of SPFWT0	Sum of SPFWT0 (s.e.)	Percent	Percent (s.e.)	
Cog Cog Cog Cog Cog Cog Cog Cog Cog Cog	Austria	No Yes	1064 3961	1266059 4278081	15199,94	22,84 77,16	,27	
	Belgium	No Yes	1022 3956	785596 3135047	20596,46 20947,27	20,04 79,96	,52 ,52	
	Chile	No Yes	1727 3462	3937559 8298366	242511,9 245007,4	32,18 67,82	1,99 1,99	
	Cyprus	No Yes	1024 3368	105514 381997	1721,89 2284,85	21,64 78,36	,33 ,33	
*	Czech Republic	No Yes	1067 5012	1146713 6200475	24775,76 30070,05	15,61 84,39	,34 ,34	
	Denmark	No Yes	1700 5588	953137 2662536	17165,31 17195,62	26,36 73,64	,47 ,47	
	Finland	No Yes	978 4486	687022 2809887	15003,42 15003,42	19,65 80,35	,43 ,43	
	France	No Yes	1822 5094	11089080 28641254	160521,8 159464,0	27,91 72,09	,40 ,40	
	Greece	No Yes	1248 3669	2260790 4737120	12221,50 4,00	32,31 67,69	,12 ,12	
	Ireland	No Yes	1449 4516	848852 2132082	3025,77 2061,02	28,48 71,52	,08 ,08	
	Israel	No Yes	1157 4182	837197 3864348	20380,67 21957,50	17,81 82,19	,43 ,43	
4	Italy	No Yes	1760 2829	21034321 18078504	58783,83 29668,62	53,78 46.22	,08 .08	
Information area								IBM SPSS Statistics Processor is ready

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.

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

		Standard
Country	Estimate	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.
- 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.

gression.spv [regression] - IBM SF	PSS Statistics Viewer								_
e Edit View Data Transform Insert Format Analice Direct Marketing Graphs Utilities Add-ons Window Help									
🛚 🗏 🖨 🙇 🍠) 🛄 🖛 🛥	🧝 🖿 📥 🗐	Va 🌒 🦊	P 🖻 🗟	•	+ + -		👕 💼 🚘	
Output	Regression Coeffi	cients							
203 Page Title							an - 1 - 1 - 1	man da da d	
- S Crosstabs			Regression	Coefficient	Coefficient	Stadzdzd.	Coefficient	Coefficient	
- E Title	CNTRYID	Variable	Coefficient	(s.e.)	(t-value)	Coefficient	(5.0.)	(t-value)	
- Case Processing									
Population with up									
Page Title	Austria	(CONSTANT)	245,37	1,73	141,44				
- En Title		EDU_USL_D2	31,20	1,80	17,32	, 30	,02	19,26	
- 🔂 Notes									
Descriptive Statist	Belgium	(CONSTANT)	242,31	1,73	139,99			22.71	
Page file Descriptives		200_030_02	41,54	1,55	21,40	, 35	,02	22,11	
- @ Title	Chile	(CONSTANT)	185.64	1,96	94.83				
- C Notes		EDU USL D2	50,90	2,75	18,51	,45	,02	22,94	
- Ing Descriptive Statist									
- S Report	Cyprus	(CONSTANT)	251,61	1,64	153,69				
@ Title		EDU_USL_D2	21,99	1,74	12,66	, 22	,02	13,28	
Notes									
2h Text Output	Czech Republic	(CONSTANT)	255,84	2,49	102,82	· · · ·	•		
* 🕒 Page Title		EDU_USL_D2	21,55	2,50	8,61	,19	,02	8,80	
Report	Desmark	(COMPTANT)	246 12	1 40	165 67				
- Notes	Dennark	FDU USL D2	33,50	1,71	19.64	. 31	.01	21.52	
- Active Dataset		000_000_00	00,00	27.72	20/01	101	/**	62700	
Text Output	Finland	(CONSTANT)	260,36	1,85	140,38				
- B Report		EDU_USL_D2	33,84	2,16	15,70	, 27	,02	16,90	
+@ Title									
Notes	France	(CONSTANT)	231,91	1,10	210,16				
- La Active Dataset		EDU_USL_D2	41,91	1,25	33,63	, 38	,01	36,56	
		(0000000000)	226.22	2.22	106.02				
	Greece	(CONSTANT)	236,73	2,23	106,02		- 02	0.74	
		ED0_035_02	20,02	2,70	9,11	,20	,03	2, 14	
	Ireland	(CONSTANT)	237,35	1,62	146.07				
		EDU USL D2	40,89	1,78	23,00	, 39	,02	25,58	
	Israel	(CONSTANT)	224,34	2,16	103,98				
		EDU_USL_D2	37,63	2,28	16,48	,26	,02	17,11	
	Italy	(CONSTANT)	235,05	1,61	146,04	1	· · · ·		
		EDU_USL_D2	33,38	1,74	19,13	, 37	,02	22,94	
		(00000000000)	260 50	2.02	122 20				
	oapan	EDII USL D2	269,50	2,02	15,40	. 28	.02	17.43	
		200_000_02	31,37	2,04	10,40	,20	,02	17,43	

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	hy a	regression	analysis
1 <i>ubie</i> 0-9	Generalea	jues	<i>by u</i>	regression	unuiysis

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	SPSS	Contains r-square of the regression model, and its adjusted
regression_Model.xlsx	Excel	r-squared.
regression_Coef.sav	SPSS	Contains unstandardized and standardized regression
regression_Coef.xlsx	Excel	coefficients, their standard errors.
regression_Desc.sav	SPSS	Contains descriptive statistics for all the variables included
regression_Desc.xlsx	Excel	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 R², a measure of explained variance (see, for example, Field, A., 2013 for more information about regression analysis).

	Standardized	Standard	
Countries	Estimate	Error	\mathbb{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

Table 6-10 Standardized Regression coefficients and model fit

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.

Analysis File; CSDataImerge_plaac.sav	Select	
Onlysis Type: Statistic Type: Plausible Value Option Missing Data Option Number of Decimals: PRAdC (using final full sample weight) * (Overlations *) Use PVs *) Plansies *) 2 *)		送 IEA
Select Variables:		
	Grouping Variables: 🗹 Exclude Missing From Analysis	
Name Description	Name Description Source Country ID (ISO 3166, num	eńc) :
	Analysis Variables:	
	Name Description	
	Plausible Values:	
	Name Description L PUTI-10 1ST TO 10TH PV Literacy is L PVSL1-10 1ST TO 10TH PV Integration L PVSL1-10 1ST TO 10TH PV Problem-1	ale score scale score olving scale score
	Weight Variable:	
	Name Description Part SPRVT0 Final full sample weight	
Cooperment Coopercementor	Wickity	Neturn to Main Wend

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	SPSS	Contains the correlation estimates and their standard errors.
correlation_Corr.xlsx	Excel	
correlation_Desc.sav	SPSS	Contains descriptive for all the variables included in the
correlation_Desc.xlsx	Excel	correlation analysis. These include means, standard
		deviations and variances.

The output of these computations is displayed in Figure 6.18.

Correlation.spv [correlation] - IBM	SPSS Statistics Viewer		-							σ×
Ene Edit View Data Transf	form insert Format Analyz	is Direct Marketing Graphs	Daintes Ado-ons <u>Wi</u> n	dow Help		+ + -	10.10	7 8 8		
Output	Correlation Coeff:	Lcients				/ =				
Page Title Descriptives Title Notes Descriptive Statist Page Title	CNTRYID	variable	Correlation with PVLIT_	Correlation with PVLIT_ (s.e.)	Correlation with PVNUM_	Correlation with PVNUM_ (s.e.)	Correlation with PVPSL_	Correlation with PVPSL_ (s.e.)		
B Descriptives B Title B Notes Descriptive Statist	Austria	PVLIT_ PVNUM_ PVPSL_	1,00 ,86 ,79	,00 ,01 ,01	,86 1,00 ,71	,01 ,00 ,01	,79 ,71 1,00	,01 ,01 ,00		
Cog C	Belgium	PVNUM_ PVPSL_	1,00 ,87 ,81	,00 ,00 ,01	,87 1,00 ,73	,00 ,00 ,01	,81 ,73 1,00	,01 ,01 ,00		
Page Title	Canada	PVLIT_ PVNUM_ PVPSL_	1,00 ,87 ,81	,00 ,00 ,01	,87 1,00 ,74	,00 ,00 ,01	,81 ,74 1,00	,01 ,01 ,00		
- Carl Text Output	Chile	PVLIT_ PVNUM_ PVPSL_	1,00 ,84 ,77	,00 ,01 ,01	,84 1,00 ,70	,01 ,00 ,02	,77 ,70 1,00	,01 ,02 ,00		
	Cyprus	PVLIT_ PVNUM_ PVPSL_	1,00 ,81	,00 ,01	,81 1,00	,01 ,00		:		
	Czech Republic	PVLIT_ PVNUM_ PVPSL_	1,00 ,80 ,77	,00 ,01 ,02	,80 1,00 ,70	,01 ,00 ,02	,77 ,70 1,00	,02 ,02 ,00		
	Denmark	PVLIT_ PVNUM_ PVPSL_	1,00 ,88 ,82	,00 ,01 ,01	,88 1,00 ,76	,01 ,00 ,01	,82 ,76 1,00	,01 ,01 ,00		
	Estonia	PVLIT_ PVNUM_ PVPSL_	1,00 ,83 ,80	,00 ,01 ,01	,83 1,00 ,75	,01 ,00 ,01	,80 ,75 1,00	,01 ,01 ,00		
	Finland	PVLIT_ PVNUM_ PVPSL_	1,00 ,86 ,81	,00 ,00 ,01	,86 1,00 ,71	,00 ,00 ,01	,81 ,71 1,00	,01 ,01 ,00		
	France	PVLIT_ PVNUM_ PVDSI.	1,00,86	,00	,86 1,00	,00		:		
4	I	e (Pau						·····	IBM SPSS Statistics Processor is ready OI	JS JS

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.

	OECD	IDB
	estimates	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

Table 6-12 Correlations between literacy and numeracy scale scores

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.

alysis Type: St AAC (using final full sample weight) *							Select		
alysis Type: St AAC (using final full sample weight) *									
AAC (using final full sample weight) "	tatistic lype:	Plausible Value Option:	Benchmark Option:	Number of Decimals:	Show Graphs				
	Benchmarks ~	Use PVs ~	Discrete	2 ~	Yes	v			
jelect Variables:					_				
Name	Dercription					Grouping Variables: 🗹 Exclude Missing	g From Analysis		
T	ocampuon					Name	Description	(
A PVNUM1-10 1	1ST TO 10TH PV Numer	ocy scale score				CNTRYID	Country ID (ISO 3166, numeric)		
VVB0110 IST TO 10TH PV Palletsky side soce PVPSL1-10 IST TO 10TH PV Problem-solving scale score				Plausible Values: Report cases with	no plausible values (Not classified)				
						Name	Description		
						PVUT1-10	1ST TO 10TH PV Literacy scale score		
						 Achievement Benchmarks: 			
						h76 226 276 326 376	 Compute percentages within benchmarks 		
						Weight Variable:			
						Name	Description		
						· · · · · · · · · · · · · · · · · · ·	Final full sample weight		

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.

📻 benchmark.spv [benchmark] - IBM	SPSS Statistics Viewer									- ø ×
Ele Edit View Dala Transform Insert Format Analice Direct.Warkeling Graphs Window Help										
😑 🔳 🖨 🙇 🥑) 📖 🗠 🛥 🧮 🔚 🗄 🗐 '	🍛 🌭 🚑 🎥 🖻	ی 🔁	+ +	+ -		T 💼 📥			
Output Page Title	Percents by Performance Groups of PVLIT									
Descriptives										
- R Notes					Sum of					
Descriptive Statist	Country TD (TEO 2166 - putrola)	Desferrer Course	N of	Sum of	SPPWID	Description	Fercent			
Page Title	councry ib (150 5100, numeric)	Performance Group	Cases	SPEWIO	(5.6.)	reicenc	(3.0.)			
-@ Title										
Notes	Austria	1.Below 176	110	138480	17981,29	2,50	, 32			
-La Descriptive Statist		2.From 176 to Below 226	588	724973	37519,54	13,08	,68			
* 🕰 Page Title		3.From 226 to Below 276	1831	2099406	49156,69	37,87	,89			
🖶 🖲 Report		4.From 276 to Below 326	2007	2106055	49895,46	37,99	,90			
@ Title		5.From 326 to Below 376	473	460466	25423,58	8,31	,46			
- Ch Active Dataset		6.At or Above 376	17	14760	4853,64	, 27	,09			
-Un Text Output		1 - 1 104	151	1100.00						
- 🕅 Log	Beigium	1.Below 1/6	151	113765	11112,07	2,90	,28			
🖻 📲 Graph		2.From 176 to Below 226	589	1004004	22192,26	11,89	, 36			
+ En Title		A From 276 to Below 276	2050	1607127	32007,20	40.05	,03			
Active Dataset		5 From 326 to Below 326	676	494317	22630 26	12 59	58			
Ca Stack Bar of mean		6.ht or hhove 376	23	17943	7548.59	.46	. 19			
Page Title						,	,			
Log	Canada	1.Below 176	1196	905375	57181,56	3,87	,24			
1		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			
4	4	3 Prom 226 to Below 276	2210	2771702	120849 5	37 71	1.65			18

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
benchmark.xlsx	Excel	standard errors.
benchmark_by_CNTRYID_Sig.sav	SPSS	Contains a country by country comparison for
benchmark_by_CNTRYID_Sig.xlsx	Excel	the percentages at each benchmark.

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.

		Standard				
Country	Estimate	Error	Country	Estimate	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