

## Working with the Data The IEA IDB Analyzer

---

PIAAC International Database Training  
Prague, Czech Republic  
May 13-15, 2014



## About the IEA's IDB Analyzer

---

- Standalone application that creates SPSS syntax to 1) merge and 2) conduct analysis with large-scale assessment data
- Works with TIMSS, PIRLS, ICCS, SITES, TALIS, PISA ... and PIAAC!
- All calculations are performed with SPSS
- Creates output in Excel, SPSS data file, and SPSS pivot tables
- Knows PIAAC sample design and implements corresponding and correct variance estimation



## Installing the IDB Analyzer

---

- Run the IDB Analyzer installation
  - ✓ Current version is 3.1.8
- Accept the licensing agreement!
- Choose the default directory for the installation
  - ✓ C:\Users\
- Once the installation process is completed, you are ready to use the IDB Analyzer
- If you do not have the Microsoft .NET Framework 4 on your computer, things will take “a bit” longer



(c) IEA-ETS Research Institute ([www.IERInstitute.org](http://www.IERInstitute.org))

2

## Install Demonstration

---



## Using the Merge Module

---

- Follow the numbers...
  - ✓ 1) Select the directory where the data are stored
    - All data must be in the same directory
    - The IDB Analyzer will recognize the study (and year)
  - ✓ 2) Select the countries of interest and move them over to the panel on the right
    - You can edit the country list (translate labels, add labels, etc.)
    - Edits are saved for use next time
  - ✓ Go to the next panel by clicking on "NEXT" or on the "Select File Types and Variables" tab at the top of your screen



## Using the Merge Module

---

- Keep following the numbers...
  - ✓ 3) Select the "File Types" of interest by clicking on the box
    - Just one file type in PIAAC (easy!)
  - ✓ 4) Move the variables of interest to the panel on the right
  - ✓ 5) Provide the name of the output files
    - This name is used for ALL output files
  - ✓ 6) "Start SPSS"



## Using the Merge Module

---

- After SPSS starts, you will see a syntax file that, when processed, generates an SPSS file with the name and location you specified in step 5
- HELP is always available by pressing the F1 key, or clicking on the HELP button



(c) IEA-ETS Research Institute ([www.IERInstitute.org](http://www.IERInstitute.org))

6

## Merge Module Demonstration

---

Two countries – Austria and United States



## Between 'Merge' and 'Analysis'

---

- After merging the files, you can do the following
  - ✓ Create indices
  - ✓ Adjust country IDs as necessary to match OECD reporting etc.
  - ✓ Create selection or subgroup variables
  - ✓ Create variables for contrast coding
- There is no "naming convention" for new variables, but it is always recommended to keep them simple
- Data files should not be sliced/filtered



## Using the Analysis Module

---

- Available analysis types
  - ✓ Percentages
  - ✓ Percentages, means and standard deviations
  - ✓ Regression
  - ✓ Correlation
  - ✓ Benchmarks
  - ✓ Percentiles



## Analysis Types – Percentages (only)

---

- Computes percentages by user-selected subgroups
- Sample questions
  - ✓ What are the percentages of males and females in the population?
  - ✓ What is the percent of people within the population with different levels of education



## Analysis Types – Percentages and Means

---

- Computes percentages, means and standard deviations for selected variables by user-selected subgroups
- Option with or without plausible values
- Sample questions
  - ✓ What is the percent of males and females in the population and their average numeracy score?
  - ✓ What is the percent of people within the population with different levels of education and their average income?



## Analysis Types - Regression

---

- Computes regression equation and model statistics
  - ✓ Listwise or pairwise deletion
- Capability of including plausible values as dependent or independent variables in the regression equation
  - ✓ When more than one set of plausible values is selected, analysis is carried out: First with first, second with second, etc.
  - ✓ Results are summarized across the plausible values
- Sample questions
  - ✓ Predict salary from education and computer experience
  - ✓ Using contrast coding, calculate effect and significance of group membership



## Analysis Types - Benchmarks

---

- Computes percent of people meeting a set of user-specified performance “benchmarks”
- Two modes
  - ✓ Cumulative: Percent of people at or above given points in the distribution
  - ✓ Discrete: Percent of people given points of the distribution
    - As an option, it can compute the mean of a variable for those at a particular achievement level
- PIAAC refers to these as “proficiency levels”
  - ✓ Cutscores literacy and numeracy (5 levels): 176, 226, 276, 326, 376
  - ✓ Cutscores problem-solving (3 levels): 241, 291, 341



## Analysis Types – Benchmarks (cont'd)

---

- Option to report those “not classified” (usually literacy-related non-response)
- Sample questions
  - ✓ What percent of the population is at each of the proficiency levels?
  - ✓ What is the average income for each of the proficiency levels?
  - ✓ What percent of the population is at or above proficiency level 3?



## Analysis Types – Correlations

---

- Computes correlation for selected variables
  - ✓ Listwise or pairwise deletion
- Compute correlation with and between plausible values
  - ✓ When more than one set of plausible values is selected, analysis is carried out: First with first, second with second, etc.
- Sample questions
  - ✓ What is the correlation between education level and income?
  - ✓ What is the correlation between income & numeracy?
  - ✓ What is the correlation between numeracy & literacy?





## Analysis Types - Percentiles

---

- Computes the score point(s) in the distribution that separate a given proportion of the population from the rest
- Sample questions:
  - ✓ What is the income that separates the top 5% of earners from the bottom 95%? (95<sup>th</sup> percentile)
  - ✓ Which numeracy score separates the top 25% of the distribution from the bottom 75%? (75<sup>th</sup> percentile)



## Some Generalities

---

- Results are ALWAYS computed by “country”
  - ✓ A “country” is defined as a unique code in variable COUNTRYID
  - ✓ You can specify other grouping variables within country
  - ✓ Separate reporting for the UK will require use of COUNTRYID\_E yet mind Canada (French/English)
- “International average” is the mean of the country estimates
  - ✓ Each country contributes equally regardless of sample or population size
  - ✓ Has its own formula for calculating the standard error



## Running the Analysis Module

---

- Follow the numbers...
  - ✓ 1) Select the file for the analysis
  - ✓ 2) Select the analysis type
    - Indicate whether:
      - a) You will use achievement scores (PVs)
      - b) You want to include cases in the analysis with missing values in the grouping variable
  - ✓ 3) Select the variables for your analysis
  - ✓ 4) Specify the name of your output file
    - a) This name is used to save the syntax, and an SPSS and Excel file with your results
  - ✓ 5) "Start SPSS"



## Running the Analysis Module

---

- Variable selection
  - ✓ Grouping Variables
    - List of variables that will be used to define the subgroups for the analysis
    - CNTRYID is required as the first grouping variable (dictated not least by variance estimation)
  - ✓ Analysis Variables
    - Depending on the analysis chosen, could be:
      - Variables for which means will be computed
      - Variables that will be correlated
      - Variables used in the regression model
      - Variables for which percentiles will be calculated
    - Multiple analysis variables can be specified



## Running the Analysis Module

---

- Variable selection
  - ✓ Plausible Values
    - Plausible values that will be used in the analysis
  - ✓ Dependent Variable
    - The variable treated as dependent when a regression is calculated
  - ✓ Achievement Benchmarks/Percentiles
    - Values that will be used as cut points of the achievement distribution
    - Values for the percentiles that will be calculated
- Information not required for the analysis appears grayed out



## Running the Analysis Module

---

- When running the analysis module, you get the following output:
  - ✓ The syntax used for the analysis
  - ✓ One or more SPSS data file with the results from the analysis
  - ✓ One or more Excel spreadsheet with the results from the analysis
- All these will be in the same directory, using same filename, but in different formats
- Once you create an SPSS syntax, you can reuse it with newer data files or tweak it, copy it ...



## Analysis Module Demonstrations

---

Means, Regression, Correlations, and Benchmarks  
from Chapter 2 of the international report



## A Few Loose Thoughts...

---

- After you create the syntax using the IDB Analyzer, you can add comments to the syntax file
  - ✓ This will be the log of your work
  - ✓ You can pass this along to someone else and they will be able to use it provided files are in the SAME location (change is also easy)
- When you receive updated data, simply replace it and use the same syntax
- While you can tinker with the parameters in the syntax, this is not recommended for beginners
- Not all parameters have a GUI switch



## A Few Loose Thoughts...

---

- Analysis not included in the IDB Analyzer are also possible
- Remember the principles:
  - ✓ Take into account sampling uncertainty
    - Repeat according to sample design using replicate weights and summarize according to formula
  - ✓ Take into account measurement uncertainty
    - Repeat analysis x 10, then combine results according to formulae



## Computing 'International Averages'

---

- Taken to mean the "mean of the country estimates"
  - ✓ Each country contributes equally regardless of sample or population size
  - ✓ Has its own formula for calculating the standard error



## Some Key Points to Keep in Mind About PIAAC ...

---

- Surveys have invaluable descriptive power
- It susceptible to self-reporting, social desirability etc. though
- We know what people know and can do now, and the context in which this occurs
  - ✓ At the population/domain level, not at the level of individuals
- We mostly have current background information, while learning or effect might have occurred a while back
- Cross-sectional data
- Can make statements about correlations, not causation
- Antecedent and outcomes interrelated

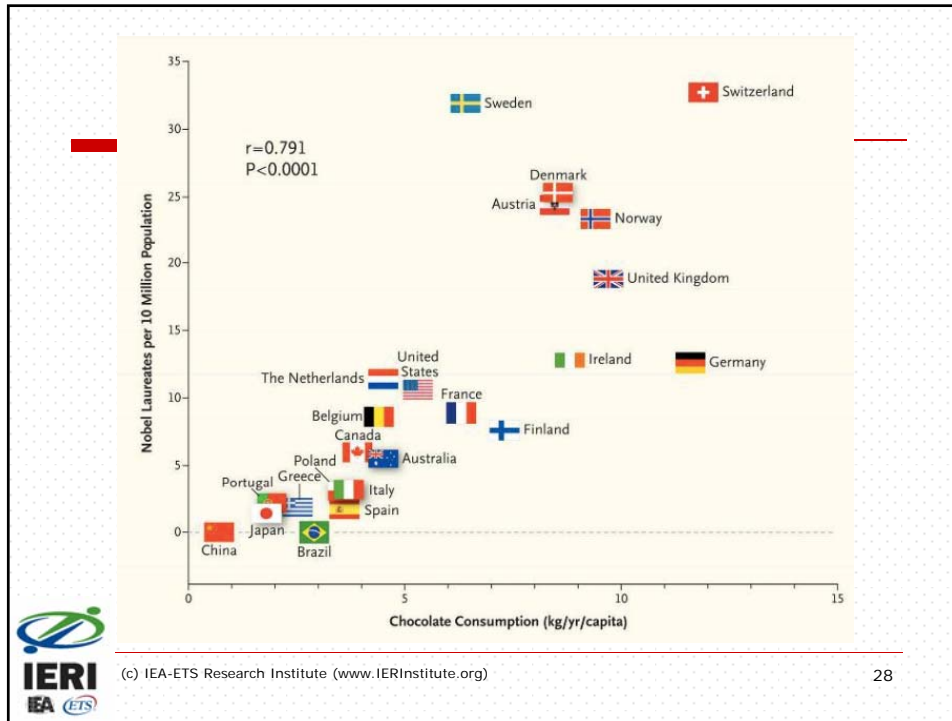


## Interpreting Background Variables

---

- PIAAC is not an experiment
  - ✓ We do not control assignment of people to “treatment” groups
  - ✓ We can not establish causality or direct effect
  - ✓ Events have already happened; all we do is record what has happened





## Interpreting Background Variables

- Important to know how to word statements about contextual variables correctly
- For example (J\_Q08):
  - ✓ About how many books are there in your home?
    - 1 = 10 books or less
    - 2 = 11 to 25 books
    - 3 = 26 to 100 books
    - 4 = 101 to 200 books
    - 5 = 201 to 500 books
    - 6 = More than 500 books

## Interpreting Background Variables

---

- We could ask ...
  - ✓ Is there a (statistical) relationship between the number of books in the home and literacy?
  - ✓ Are those who report having more books in the home more likely to do better in literacy than those who report fewer books in the home?
- We should not ask ...
  - ✓ Does having more books in the home have an effect (increase/decrease) on reading literacy?



## Interpreting Background Variables

---

- We could answer...
  - ✓ Those who report having more books in the home tend to have higher reading skills than those who report fewer books in the home
  - ✓ Those who have higher reading skills are more likely to report living in homes where there are many books
- We should not (and cannot empirically) answer ...
  - ✓ People have higher reading skills because there are more books in the home
  - ✓ High reading skills increase the number of books found in the home





## Statistically Significant...?

---

- Any difference can be “statistically significant” if the sample is large and precise enough
- But a “statistically significant” difference is not necessarily “substantively significant”



(c) IEA-ETS Research Institute ([www.IERInstitute.org](http://www.IERInstitute.org))

32

## Thank you for your attention!

---



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ