Slide 1 of 35

Title Slide: Considerations for Analysis of PIAAC Data

Slide 2 of 35

This module provides a demonstration of how to use the IEA IDB Analyzer for merging and analyzing data. It also describes considerations for analyzing PIAAC data.

Many statistical packages assume simple random sampling. PIAAC data should not be analyzed using such procedures, as they may result in biased estimates. They can also result in underestimated sampling errors that will produce incorrect p values, which can incorrectly indicate differences that are statistically significant when they are not. Therefore, it is necessary to use special statistical software that accounts for the PIAAC study design. While several statistical programs such as AM, WesVar, R, and Stata have this capability, and while SAS and SPSS offer the capability through optional add-ons, this module will focus on teaching you one particular tool —the IEA IDB Analyzer.

Slide 3 of 35

The IEA IDB Analyzer is statistical software developed by the International Association for the Evaluation of Education Achievement Data Process and Research Center, or IEA-DPC. It is designed for use with large-scale international assessments. The IDB Analyzer creates SPSS syntax that can be used to merge data files and perform statistical analyses with PIAAC and other international assessment data files. The software generates code that appropriately uses plausible values, and takes into account complex sampling design in the computation of sampling variance.

Slide 4 of 35

In order to use the IDB Analyzer, you will need a PC that has Windows XP or higher, .Net 4, SPSS 15 or higher, and Microsoft Excel version 2003 or higher installed on it. In order to download and install the program, you will need administrator rights to the computer you will be using.

The IDB Analyzer and its installation guide can be downloaded by clicking on the corresponding underlined screen text.

The IDB Analyzer has one common graphical user interface with two modules: the Merge Module and the Analysis Module. The Merge module will be described first before going into detail about the Analysis module.

Slide 5 of 35

Once you have installed the IDB Analyzer, the first step in your PIAAC data analysis will be to run the “Merge” module to create an analysis dataset. The Merge module allows you to select the countries and subsets of variables that you wish to analyze.

To do this, open the IDB Analyzer and click on “Merge Module.”

Slide 6 of 35

To merge PIAAC data files using the IDB Analyzer, first, select the directory where the PIAAC data files are stored. Make sure you have downloaded the data files you wish to use in SPSS format from OECD or NCES. Links to public-use data files are available on their respective websites, as described in the Module titled “Getting Started with the PIAAC Data”.

Select your study from the drop down menu if it is not automatically selected for you. The IDB Analyzer will then show the available countries in the left-hand panel.

Slide 7 of 35

To select a country dataset you wish to use, click on that country in the “Available Participants” list on the left, and click on the *right facing arrow* to add it to the “Selected Participants” list on the right.

To select all participating countries, click on the *right bar facing arrow* to add them to the “Selected Participants” list.

To remove a participating country, select it from the “Selected Participants” list, then click on the *left facing arrow*.

To remove all participating countries selected, click on the *left bar facing arrow* to remove them from the “Selected Participants” list.

To select a limited list of countries, click on the first country in the list, hold down the Shift key, click on the last country in the list, and click on the *right facing arrow*.

To select multiple countries in any order, click on any country in the list, hold down the Control key, select all the countries you wish to include, then click on the *right facing arrow*.

Once your list of selected participants is complete for your purposes, click “Next” at the bottom of the screen.

Slide 8 of 35

The next step is to choose your variables. First, make sure that “General Response File” has been selected in the top-left corner of your screen.

All PIAAC background variables, plausible values, and cognitive assessment items will be listed in “Available Variables” on the left side of your screen under “Background Variables and Scores”. The variables you select from this list for your analysis dataset will appear in the “Selected Variables” list on the right. The “ID and Sampling Variables”

necessary for correctly running your data analysis will automatically be added to the "Selected Variables" list on the right side.

To add variables to your dataset, select them from the list on the left and click the right-facing arrow in the middle to move them to the list on the right. To select all of the variables at once, click on the first variable in the list, hold down the Shift key, click on the last variable in the list, and click on the right-facing arrow. To remove variables from your dataset, select them from the list on the right and click the left-facing arrow.

In this example, two variables and three plausible values have been selected and moved to the analysis dataset.

Slide 9 of 35

Next, define the output file's location and provide a file name at the bottom of the left panel. This example shows that the output file is being saved at "C:\temp\PIAAC\" with the file name "analysis".

Once you've established the location, click "Start SPSS." This will load the syntax file that IDB Analyzer has created into SPSS.

Slide 10 of 35

After SPSS starts, you will see a syntax file that, when run, generates an SPSS data file with your chosen variables and countries. Use the Ctrl+A *(pronounced "Control A")* command to select the entire syntax text, then use Ctrl+R *(pronounced "Control R")* or select Run – All from the menu at the top to run the syntax and generate your dataset.

The merged dataset and syntax file will be saved with the location and name you specified. The syntax file, which is the file with the SPS extension, can be used to recreate the file.

Slide 11 of 35

If you are using only one country's data, you can start off right away with the analysis module. If you are using more than one country's data, once you have merged your data files, you are ready to begin using the analysis module. Analyses conducted with the IDB Analyzer account for a complex sample design with plausible values methodology, including weighted estimates and the correct calculation of standard errors. The Analysis Module can be used to calculate estimates of percentages, means, regression coefficients, proficiency levels, and percentiles.

To begin an analysis, open the IDB Analyzer and click on "Analysis Module". If you just used the Merge Module, you can return to the Main Menu to access the Analysis Module.

Slide 12 of 35

The Analysis Module consists of one screen where you make various selections to customize your analyses. We will go through the steps for using the Analysis Module in the next few slides.

Slide 13 of 35

The first step to using the Analysis module is to select a file for analysis at the top of the module screen.

Next, just below the Analysis file selection, select your Analysis Type as “PIAAC” if it is not already selected.

Slide 14 of 35

Next you will select the statistic type from the drop-down menu. You can also select whether you want achievement scores included or not by making a selection under the “Plausible Value Option” drop-down menu. The default under the plausible value option is “None Used”. If you wish to include the plausible values for achievement scores, select the “Use PVs” option.

Slide 15 of 35

The IDB Analyzer offers the following options for statistic types: percentages and means; percentages only; regression; correlations; percentiles; and benchmarks with discrete or cumulative options. Correct standard errors are calculated for all analytic procedures. Together with the statistics, these are presented in output files so that you may evaluate the statistical significance of your results.

We will now go through each statistic type in detail.

Slide 16 of 35

For percentages and means, running the analysis “Without Plausible Values” computes percentages, means, and standard deviations for the selected variable or variables, by subgroups defined by the grouping variable or variables. Run the analysis “Without Plausible Values” when only background variables are included in the analysis.

Selecting “With Plausible Values” computes percentages, **mean achievement scores**, and standard deviations **based on plausible values**, by subgroups defined by grouping variable or variables. Run the analysis “With Plausible Values” when assessment domains are included in the analysis.

Slide 17 of 35

When running the analysis with “Percentages only,” percentages will be computed with standard errors by subgroups that are defined by the grouping variable or variables that you chose. When using this statistic type, you will not have access to plausible values, and thus cannot compute achievement scores.

Slide 18 of 35

If you wish to conduct a regression analysis, you will need to indicate whether you will use plausible values. Selecting the “Without Plausible Values” option calculates standardized regression coefficients for selected independent variables in order to predict a dependent variable, by subgroups defined by the grouping variable or variables.

Selecting regression “With Plausible Values” computes a regression model in which plausible values can be used as either the dependent or independent variable.

More than one Plausible Values domain can be specified as the independent variable in order to study the relationship between the domains.

Slide 19 of 35

Selecting correlations as your statistics type and indicating that the analysis should be conducted without plausible values will result in calculating correlation coefficients for selected variables by subgroups that are defined by the grouping variable or variables.

Conducting the analysis with plausible values will allow you to include plausible values when computing correlation coefficients.

Slide 20 of 35

Using the “Percentiles” option without plausible values will calculate averages for defined percentile groups for any continuous variable, by subgroups that are defined by the grouping variable or variables.

Choosing the percentiles statistics type with plausible values will calculate mean scores for defined percentile groups for a specified cognitive assessment domain, by subgroups that are defined by the grouping variable or variables.

Slide 21 of 35

Selecting benchmarks as your statistics type will compute percentages of the population within, at, above, or below benchmarks of performance defined by cut-point scores, by subgroups defined by the grouping variable or variables. You have the option to choose “Cumulative” or “Discrete” benchmarks, depending on your preference.

The cut-point scores for pre-defined PIAAC benchmarks are provided for your reference. The process for entering individual benchmarks will be illustrated on the next slide.

Slide 22 of 35

Once you have selected the statistics type, you will select the variables, or parameters, for your analysis. Depending on the statistics type and options you choose, this screen will differ. Areas on the screen that are not used for a particular type of analysis are grayed out. The example shown here is a benchmark analysis, in which all variables may be used. The types of variables for this type of analysis include grouping variables, plausible values, a weight variable, and achievement benchmarks.

You will need to enter the benchmarks in the box indicated on the screen, separated by spaces without commas or semicolons. If you were conducting a percentiles analysis, then you would enter percentiles in this box.

Next, find and select the variable on the left of the screen and then select the right-facing arrow next to the parameter type of your choosing. If you would like to remove a variable at any time, you can select the variable you want to remove, then click on the left-facing arrow next to the box.

You will notice that the IDB Analyzer Analysis Module automatically selects the appropriate Weight Variable for your convenience.

If you were to conduct a regression analysis, the parameters would include grouping variables, background analysis variables, plausible values as analysis variables, a dependent variable (which may be a background variable or plausible value), and a weight variable.

Slide 23 of 35

"Grouping variables" refers to a list of the independent variables that you will use to define the subgroups for your analysis. For example, you may wish to compute achievement scores by gender. The list can consist of one or more variables, and the IDB Analyzer automatically includes country IDs as the first grouping variable. In addition to conducting your analysis by country, IDB Analyzer will also include estimates for the overall sample in all countries in your analysis output.

If the option "Exclude Missing from analysis" is checked, only cases that have non-missing values in the grouping variables will be used in the analysis.

Slide 24 of 35

The Analysis Variables, or independent variables, can be either continuous or categorical variables (excluding plausible values). Depending on the statistic type chosen, analysis variables may be those for which means will be computed, those that

will be correlated, those for which percentiles will be calculated, or those that will be used as predictors in a regression model.

Slide 25 of 35

Achievement scores are computed through plausible values. If you want to use plausible values, select the plausible values parameter area and choose from the left panel the achievement scale that you are interested in.

Slide 26 of 35

The dependent variable is the one that will be predicted by the list of analysis variables in a regression model. The dependent variable may be plausible values, or a variable from the background questionnaire. Only one dependent variable can be selected.

Slide 27 of 35

The achievement benchmarks or percentiles are the values that will be used as cut points of the achievement scales and other continuous variables. For example, there is a benchmark of 376 for Level 5 in literacy. You may wish to enter "376" to compute the percentage of respondents that achieved a score at or above 376 on the literacy scale (THAT IS, scored in Level 5). Or, you may wish to compute the average score for the 75^{th} percentile of a specified group by entering "75." In order to input multiple benchmarks or percentiles, enter the values with spaces in between; do NOT enter commas or other separators.

To find PIAAC benchmark cut point scores for proficiency levels, click the underlined screen text "Statistic Types – Benchmarks"

Slide 28 of 35

Once you have finished selecting your statistics and variables, it is time to specify the file name and location for the output files. The output files will include the SPSS syntax used for the analysis that, when run, will provide an SPSS data file and output with the results from the analysis.

Once you have defined your Output files, click the button at the bottom of the page that says "Start SPSS."

Slide 29 of 35

The IDB Analyzer will provide a customized SPSS syntax based on your selections in the Analysis Module. Selecting "Start SPSS" will open this syntax automatically in SPSS.

The final step is to run the syntax in SPSS. You can run the syntax by using the Ctrl+A *(pronounced "Control A")* command to select the entire syntax text, followed by hitting

Ctrl+R *(pronounced "Control R")* or selecting Run – All from the drop down menu at the top.

Slide 30 of 35

Now that you have seen a demonstration of the IDB Analyzer, we will describe some considerations for analyzing the PIAAC data.

First, the data are collected from samples of people who are representative of groups, not individuals. It is not appropriate to perform analyses on the responses of any one respondent. Furthermore, United States data cannot be disaggregated to make state or county level estimates due to sampling.

The PIAAC data are cross-sectional and only capture one point in time. However, several measures, including those for literacy and numeracy, can be compared with the data collected in the 1994 IALS and 2003 ALL studies. Trend analyses can provide estimates for specified groups of individuals, but the studies did not follow the same individuals from survey to survey so causal interpretations of trend data are not possible.

Although the data have invaluable descriptive power, the dataset was derived from non-experimental research. Researchers are therefore advised to analyze the data in terms of non-causal relationships. In other words, we can only make statements about correlations, and not about causation.

Slide 31 of 35

When doing statistical analyses using PIAAC data, it is important to correctly word research questions so that causal relationships are not inferred. Consider, for example, an analysis of earnings and the use of literacy skills at work. We could ask, "Is there a statistical relationship between earnings and the use of literacy skills at work?" We could also ask, "Does the group who earns more money also have a higher use of literacy skills at work than other groups?" However, we should NOT ask, "Does the use of literacy skills at work affect earnings?"

Slide 32 of 35

Another consideration for conducting analyses with PIAAC data is evaluating the statistical significance of your results. T-tests are used to assess whether point estimates found for two groups are *statistically* different from each other. Conventionally, the alpha level is set at 0.05. This means that to be a *statistically* significant difference, the difference in point estimates found for two groups could occur by chance 5% of the time.

In PIAAC, the t-test for independent groups is used to compare population statistics where there is no overlap between the sampled adults who represent the two groups being compared; for example, males and females.

The t-test for dependent groups is used when parts of a group are being compared to the group as a whole; for example, when comparing a country or territory to the international average.

Slide 33 of 35

The formula for the t-test for independent groups is shown here, where A_i is the statistic being tested for group i; for example, the mean for group i.

In this example, A_j would be the mean for group j. $S(A_i)$ is the standard error of the statistic; for example, the standard error of the mean for group i. Statistical significance of the difference between the statistics for groups i and j is calculated with the formula shown.

The absolute difference in estimates of the two groups divided by the square root of the sum of the squared standard error for the estimates needs to be greater than or equal to the T-statistic at a certain alpha level, commonly 0.05, for the estimates of the two groups to be statistically significantly different.

Slide 34 of 35

The formula for the t-test for dependent groups is shown here.

As in the formula for independent t-tests, A_i is the statistic being tested for group I, and $S(A_i)$ is the standard error of the statistic. Group i is the whole, for example the international average, and group j is the part, for example, the United States.

Statistical significance of the difference between the statistics for groups i and j is calculated using the formula shown, where p is the proportion of adults from group j within group i.

The absolute difference in estimates of the two groups, divided by the square root of the sum of the squared standard error of the estimate for the “whole” group, and the multiple of 1 minus twice the proportion of the “part” group in the “whole” group and squared standard error of the estimates for the “part” group, needs to be greater than or equal to the T-statistic at a certain alpha level, commonly 0.05, for the estimates of the two groups to be statistically significantly different.

Slide 35 of 35

This module has demonstrated the use of the IDB Analyzer and described the analytic considerations that should be kept in mind when using data from PIAAC. Particularly, considerations for interpreting background variables, and the difference between statistical significance and substantive significance were discussed.

The module’s objectives are summarized here for your reference. The resources provided throughout the module are also listed and can be accessed by clicking on the corresponding underlined screen text.

You have now completed this series of modules on PIAAC. Click the exit button to return to the landing page.