Slide 1 of 37

Title Slide: Considerations for Analysis of International Activities Program Data

Slide 2 of 37

This module will provide a demonstration of the IEA International Database Analyzer (or IDB Analyzer), both for merging data and analyzing data, as well as describe the analytic considerations for the Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS), Program for International Student Assessment (PISA), and similar international surveys.

As discussed in IAP Module 4 (Sample Designs, Weights, Variance, IRT Scaling, and Plausible Values), studies like PIRLS, TIMSS, and PISA use a complex sample design and IRT scaling of the assessment data. Thus, the formulas needed to calculate the point estimates and standard errors are more complex than what is used for a simple random sample. Many statistical software packages (for example, standard SAS or standard SPSS) assume simple random sampling. NCES complex sample data should NOT be analyzed using a procedure that assumes simple random sampling. This may result in biased estimates and/or underestimated sampling errors that will produce incorrect p values, often indicating that differences are statistically significant when they really are not. Therefore, it is necessary to use special statistical software that accounts for the study design and correctly calculates estimates and standard errors. While several statistical software options have this capability, such as AM, WesVar, and Stata, as well as optional add-ons to SAS and SPSS, this module will focus on teaching you one option that you can use—the IEA IDB Analyzer.

Slide 3 of 37

The IEA IDB Analyzer is special statistical software that produces weighted estimates, correct standard errors, and correct achievement scores using plausible values. The application is designed for use with large-scale international data files, including PIRLS, TIMSS, and PISA, and is available for free download. SPSS for Windows is needed to execute the SPSS code generated by the IDB Analyzer. This enables the user to compute descriptive statistics and conduct statistical hypothesis testing among groups in the population without having to actually write any programming code.

Slide 4 of 37

The IDB Analyzer will work on most IBM-compatible computers using Microsoft Windows XP or any later version. You will need SPSS 15 or higher, .Net Framework 4.0, and Microsoft Excel (2003 or later version) installed on your computer. Also, in order to download the IDB Analyzer, you will need administrator rights to the computer you will be using.

Currently there is no standalone Mac version of the IEA IDB Analyzer. However, the software can be used on Mac through a virtual machine and Windows (2000, XP, Vista or 7) installed on it.

To download the IDB Analyzer and the installation guide, click on the underlined screen text, 'IDB Analyzer'. Also make sure you have downloaded the IAP data files of interest from the appropriate study website.

To run the program, you will need either administrator rights or your SPSS software should "start" in a folder where you have write access. Locate your IBM SPSS Statistics program in your computer and right click on it. Then choose "Properties." You will then see where your SPSS program "starts." In some versions of Windows, the "start" folder is write-protected so that only an administrator can write to it. If you do not meet the requirements, then the IDB Analyzer will not run. If this is the case, you will need to gain administrator rights or have an administrator change the "start" folder to one for which you have write access.

The IDB Analyzer has one common graphical user interface with two modules: the Merge Module and the Analysis Module.

Slide 5 of 37

Once you have installed the IDB Analyzer you will need to run the merge module to merge together specific data files separated by country and data type. In doing this, you select the countries and variables that you wish to analyze. Open the IDB Analyzer from the Start menu (under IEA) and click on "Merge Module."

Slide 6 of 37

The merge module of the IDB Analyzer can be used with the PIRLS and TIMSS databases, in which data files are separated out by country. The merge module cannot be used with the PISA database, where data across countries are already merged into a single data file. To merge PIRLS and TIMSS data files using the IDB Analyzer, you will first select the directory where the data are stored. The IDB Analyzer won't read zipped data files, so make sure they are unzipped in the source folder.

The IDB Analyzer will automatically recognize and complete the Select Study, Select Year, and Select Grade fields located directly under the directory. However, if the source folder contains data files from more than one IEA study, or from more than one grade, you will need to select the study, year, and/or grade from the drop-down menus. The IDB Analyzer will then show the available participants in the left-hand panel. To select individual participating education systems, select one at a time and move them to the right hand panel by clicking on the single right-facing arrow. To select all participating education systems click on the right-facing double arrow to move them all over to the panel on the right. To deselect a participating education system click on the left-facing arrow. Once you have finalized your list of selected participants you can click "Next" at the bottom of the screen.

Slide 7 of 37

The next step is to choose your variables. First you will make a selection from a list of file types in the top left hand corner of the screen. To view the variables within a file type, click on its name, for example school background or student background.

Select your variables of interest from the left panel and click the right-facing arrow in the middle to move variables of interest to the panel on the right. You can move all of the variables at once by clicking on the double arrow, or you can move a single variable by selecting it and clicking on the single arrow. The ID, sampling, and achievement variables are selected by default, so you only need to select background variables. If you wish to analyze variables from more than one file type, you will ultimately select more than one box under Select File Types.

It is preferable to analyze school-level variables as attributes of the students, rather than as elements in their own right. Therefore, analyzing school data should be done by linking the students to their schools. For example, to merge all of the variables from the student and school background data files, select the Student Background file type and click on the right-facing double arrow; then select the School Background file type and click on the right-facing double arrow. Notice in this example that the variables of interest to be included in the merged data file are selected separately by file type.

Next, you 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 as "SchoolStudent_Merged" at "C:\Workshop\Output." Then, click "Save" and then "Start SPSS."

Slide 8 of 37

Before moving on, let's discuss selecting teacher data files.

Analyses with teacher data should be made with students as the units of analysis and reported in terms of students who are taught by teachers with a particular attribute. Therefore, teacher data are analyzed by linking the students to their teachers. The student–teacher linkage data files are used for this purpose and the IDB Analyzer will make use of them automatically. Thus, to analyze teacher data, click the Teacher Background file type and select variables of interest. If you are using TIMSS eighth-grade data, you can choose either the math teacher or science teacher background file type. If you wish to analyze both student and teacher background data simultaneously, you would then click the Student Background file type and select variables of interest.

Slide 9 of 37

After SPSS starts you will get a syntax file that, when run, generates an SPSS data file at the location and with the name you specified. Select Run – All (or use the Ctrl A command to select the entire syntax text followed by Ctrl R, the equivalent of selecting Run – All) to merge your selected files and generate a dataset with your selected countries and variables.

Slide 10 of 37

The merged dataset as well as the syntax file have been saved in the directory you specified. The syntax file, which is the file with the SPS extension, can be used to recreate the file.

Slide 11 of 37

For PISA, you may need to merge data files, such as if you wish to use variables from the school background questionnaire. You will need to manually merge files before being able to use the IDB Analyzer to do your analyses. The school data file by itself is not very useful because the sampling done in PISA in each participating education system was designed to be nationally representative of 15-year-old students, not a particular class of schools. The student is the unit of analysis, so you need to merge the school data file with the student data file. With the merged file in SPSS, you can conduct analyses using the Analysis Module of the IDB Analyzer and make statements about, for example, the percentage of 15-year-old students who attend schools with certain characteristics.

Slide 12 of 37

To merge PISA data using SPSS, follow these steps:

1. Before merging the two data files, they both need to be sorted by the same sorting variables. Open each data file and select Data → Sort Cases. Move the COUNTRY and SCHOOLID variables, in that order, to the 'Sort by' box, and select 'Ascending' for both variables. Click 'OK,' and check that each dataset has been sorted properly.
2. Open the sorted school data file (which will now be your active dataset). Select Data → Merge Files → Add Variables.
3. Select the student data file as the "open dataset" or "external SPSS data file" that will be merged with the active dataset. Click 'Continue.'
4. Check 'Match cases on key variables in sorted files' and 'Active dataset is keyed table.' Enter the COUNTRY and SCHOOLID variables in the 'Key Variables' box. Click 'OK', and SPSS will merge the files.
5. Save the merged data file under a file name. Run some analyses using both student and school variables and check these against corresponding data tables in the PISA international report. Click on the underlined screen text to access the report.

Slide 13 of 37

Once you have merged your data files, if you needed to do so, you are ready to begin conducting analyses using the IDB Analyzer Analysis Module. To begin an analysis using the data files, open the IDB Analyzer from the Start menu of your computer (see folder titled IEA). Then when the Analyzer opens up click on "Analysis Module." Or if you have just used the Merge Module you can return to the Main Menu to access the Analysis Module.

Slide 14 of 37

The Analysis Module is implemented from a single screen where you make various selections to customize your analyses. Shown in this slide are the main steps for using the Analysis Module. In the slides that follow we will go through these steps in detail.

Keep in mind that the advantage of using the IDB Analyzer is that it will automatically produce weighted estimates and it simplifies the computation of the correct standard errors for you. It does this by incorporating the correct sampling weights, jackknife repeated replication or BRR techniques for variance estimation, and plausible values if they are needed.

Slide 15 of 37

The first step to using the analysis module is to select the file for analysis at the top of the module screen. Note that in order for the IDB Analyzer to read in the filename, the characters have to be letters, numbers, the underscore, or pound sign, and there can be no blank spaces. Next, just below the Analysis File selection, you will select the Analysis Type.

In the example shown in the slide, the Analysis File is the “SchoolStudent_Merged” file we created earlier in this module and our Analysis Type is “TIMSS (Using Student Weights)”. The IDB Analyzer automatically selects the appropriate weights to use based on the file types included in the merged data file.

Slide 16 of 37

Next you will select the statistic type from the drop-down menu. The statistics options are: percentages and means, percentages only, regression, correlations, benchmarks, or percentiles. In the example shown, “percentages and means” is selected.

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 “Use PVs”. All statistics types, except percentages only and benchmarks, produce results either with or without achievement scores.

Slide 17 of 37

In the slides that follow, we will go through each one of these statistic types in detail. It is important to note that for all of these analytic procedures the correct standard errors of the statistics are calculated according to the variance estimation procedure required by the design of each study. The statistics and corresponding standard errors are presented in the output files for each analysis and can be used to test whether or not specific results are statistically significant.

Slide 18 of 37

For Percentages and Means, running the analysis without plausible values computes the percentages of students within specified subgroups, and their mean and standard deviation for the selected analysis variables. Selecting to use plausible values computes these statistics the same way but using mean achievement scores based on plausible values as the analysis variable.

Slide 19 of 37

For Percentages Only, percentages will be computed by subgroups based on the grouping variables that you choose. Achievement scores cannot be used with this option.

Slide 20 of 37

If you wish to conduct a regression analysis, you will need to indicate whether you will use plausible values of achievement. Running the analysis without plausible values calculates a multiple linear regression between a dependent variable and a set of independent variables, and computes the standardized regression coefficients. The regression analysis is done by subgroups defined by grouping variables. Selecting to use plausible values calculates a regression analysis in the same manner, but enabling plausible values of achievement to be used as a dependent variable, an independent variable, or both. The regression procedure also calculates descriptive statistics (mean, standard deviation, and variance), Analysis of Variance statistics (sums of square regression, residual, and total), and Model statistics (R Squared and Adjusted R Squared). T-statistics for the regression coefficients and for the standardized regression coefficients are calculated so you can evaluate the statistical significance of each independent variable in the regression model. There are no limits on the number of independent variables that can be entered, and you have the option of pairwise or listwise deletion of cases from the analysis. You also have the option to dummy code or effect code a categorical independent variable.

Slide 21 of 37

Selecting Correlations and indicating that the analysis should be conducted without plausible values will calculate correlation coefficients between selected analysis variables by subgroups defined by grouping variables. Conducting the analysis with plausible values will calculate the correlations between a set of analysis variables that include the plausible values of achievement. Descriptive statistics (mean and standard deviation) are also calculated when doing the correlation analysis. There are no limits on the number of variables that can be included in the correlation matrix, and you have the option of pairwise or listwise deletion of cases from the analysis.

Slide 22 of 37

Selecting Benchmarks will compute percentages of students meeting user-specified benchmarks of achievement. These would typically be based on established international benchmarks, as in the case of PIRLS and TIMSS, or proficiency levels, as in the case of PISA. It computes these percentages in one of three ways: (1) Discrete (that is, the percent of students within given points in the distribution as defined by the cut points entered); (2) Discrete with Analysis Variables. This option computes the mean value of an analysis variable for the groups as defined by the achievement benchmarks. For example, this option allows you to compute the average home educational resource index or the mean age per achievement group within each of the achievement groups; and (3) Cumulative (that is, the percent of students at or above the achievement benchmarks). We will describe in a later slide how to enter individual benchmark cut points.

Slide 23 of 37

The Percentiles statistic type without plausible values will calculate percentiles of any continuous variable by subgroups defined by grouping variables. Choosing Percentiles with plausible values will calculate the percentiles for a set of plausible values of achievement.

Slide 24 of 37

Once you have selected the statistics type, you will select the variables, or parameters, for your analysis. The types of variables include grouping variables, background analysis variables, achievement scores (that is, plausible values) as analysis variables, a dependent variable (which may be a background variable or set of plausible values), and a weight variable. Depending on the statistics type and options you choose, the screen will differ, with certain areas grayed out.

To select a variable for your analysis, first find and select the variable on the left side of the screen. Then select the right-facing arrow next to the parameter type of your choosing. The Analysis module is set up so that the variables are entered into the boxes in order from top to bottom. If you would like to remove a variable at any time, you can select the variable you want to remove and then select 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.

Slide 25 of 37

If you were to analyze benchmarks instead of conducting a regression analysis, the screen would look like this. The arrow on this screen points to where you would enter benchmarks.

Slide 26 of 37

Grouping variables refers to a list of variables that are to be used to define the subgroups for analysis. For example, you may wish to compute means by gender. Any analysis will be computed by education system, so the IDB Analyzer automatically includes IDCNTRY as the first grouping variable for PIRLS and TIMSS, and CNT as the first grouping variable for PISA. 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.

Depending on the statistic type chosen, analysis variables may be those for which means will be computed or those that will be correlated. Analysis variables may also be used for calculating percentiles. For example, you may wish to compute the 50th and 75th percentiles for the wealth variable in PISA. You can specify multiple analysis variables to be computed at the same time.

In addition, analysis variables may be used as predictors in a regression model, and these independent variables may either be categorical or continuous.

Slide 27 of 37

Achievement scores are computed through plausible values. If you want to compute average mathematics scores, select the plausible values area and choose from the left panel the achievement scale that you are interested in. The achievement scores can be used as a dependent or independent variable in a regression analysis.

The dependent variable is the one that will be predicted by the list of analysis variables in a regression model and only one dependent variable can be selected.

The achievement benchmarks or percentiles, or proficiency levels in PISA, are the values that will be used as cut points of the achievement distribution or the values for the percentiles that will be calculated. For example, TIMSS has established an international benchmark of 625 as an advanced level of mathematical and science proficiency. You may wish to enter "625" to compute the percentage of boys and girls for specific education systems that achieved at or above 625. Or you may wish to compute the 75^{th} percentile for specified groups by entering "75." In order to input multiple benchmarks or percentiles, add a space between cut points. For example, 475 (space) 555 (space) 625. If you are using PISA you can compute proficiency levels by entering a proficiency cut point score, such as "669.30" for level 6 in mathematics literacy in 2012. International benchmark values for PIRLS and TIMSS and PISA proficiency level cut point scores can be found in the PIRLS, TIMSS, or PISA international reports. You can access these reports by clicking on the underlined screen text.

Slide 28 of 37

When you have finished selecting your variables or parameters it is time to specify the file name and location for the output files. Note that when naming your output files, the characters have to be letters, numbers, the underscore, or pound sign, and there can be no blank spaces. 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 that says Start SPSS.

Slide 29 of 37

The IDB Analyzer will provide SPSS syntax based on your selections in the Analysis Module. The final step to view the results of your analysis is to run the syntax in SPSS. You can either select Run – All, or use the Ctrl A command to select the entire syntax text followed by Ctrl R, the equivalent of selecting Run – All.

Slide 30 of 37

Now that we have demonstrated the use of the IDB Analyzer, we will describe some analysis considerations for PIRLS, TIMSS, and PISA and similar surveys. First, these studies are surveys. We know what students know and can do in a particular moment in time, and we know the context in which the learning occurs, such as who's teaching them and what the schools are like. However, we collect mostly current background information while learning or its effects might have occurred at a previous time. These are cross-sectional studies with repeated and independent measures over time. Although the data have invaluable descriptive power, we can only make statements about correlations, and not about causation.

There are limitations to the data because these surveys rely on self-reports from participants. It is important to keep in mind that there may be cultural or other biases in survey responses due to self-reporting.

Slide 31 of 37

These studies are not experiments in which we can determine causation. Consider for example that our data show that students with more books in the home show higher achievement. We still do not know if access to books causes higher achievement or if higher achievement causes the parents to buy their children more books. The only way to know this would be to have a random sample of children, assign them to random families, and randomly assign varying amounts of books in the home. PIRLS, TIMSS, and PISA do not control the assignment of students to "treatment" groups. Therefore we cannot establish causality or direct effect. We can only look at relationships and make correlations. The events have already happened and we record what has happened.

Slide 32 of 37

When testing research hypotheses and doing statistical analyses using PIRLS, TIMSS, and PISA data, it is important to correctly word research questions about contextual variables. Avoid making causal relationships between variables when the analysis does not support this. Consider for example, the variable associated with a question about the number of books in the home.

Slide 33 of 37

We could ask: “Is there a statistical relationship between the number of books in the home and mathematics achievement at grade 8?” or “Do students who report having more books in the home tend to do better in mathematics than those who have fewer books?” But we should not ask, “Does having more books in the home affect (increase or decrease) mathematics achievement?”

Slide 34 of 37

We could answer: “Do Grade 8 students who come from homes where there are more books tend to do better in mathematics than those who do not?” or “Do Grade 8 students who do well in mathematics tend to have come from homes where there are more books?” But we should not answer, “Do students do better at mathematics because there are more books in the home?” or “Does high mathematics achievement tend to have an effect on the number of books found in the home?”

Slide 35 of 37

Another thing to consider when conducting analyses using the PIRLS, TIMSS, and PISA datasets is that any difference can be “statistically significant” if the sample is large enough. But the question is: “Does the significance matter?” It’s important to not only examine if differences are statistically significant, but also consider the magnitude of the differences and their practical significance.

Slide 36 of 37

In 2011, the coincidence of the PIRLS and TIMSS assessments offered a unique opportunity for participating countries to assess the same fourth grade students in mathematics, science, and reading.

This combined dataset produced a new set of scores and a new set of background scales and indices using data from both PIRLS and TIMSS. Thirty seven education systems, not including the United States, administered both PIRLS and TIMSS in 2011 to the same set of students and thus are included in this dataset. The combined dataset was released in August 2013 and can be accessed by clicking on the underlined screen text.

Slide 37 of 37

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

Additionally, important resources that have been provided throughout the module are summarized here along with the module's objectives for your reference.

You have now completed this series of modules, click the exit button to return to the landing page.