Considerations for Analysis of PIAAC Data

Module Objectives

- Describe the use of the IEA IDB Analyzer
- Describe considerations for analyzing PIAAC data
IDB Analyzer

- Statistical software developed by the International Association for the Evaluation of Education Achievement Data Process and Research Center (IEA-DPC)
- Designed for use with large-scale international assessments
- Creates SPSS or SAS syntax that can be used to merge data files and perform statistical analyses with PIAAC and other international data files
- Generates code that appropriately uses Plausible Values and takes into account complex sampling design in computation of sampling variance
- For additional information on the latest updates or additions to the IEA IDB Analyzer, refer to the help guide
- This module focuses primarily on using the IDB Analyzer with SPSS
  - Using it with SAS is similar with only minor differences

IDB Analyzer (Continued)

- System requirements
  - Windows Vista, 7, or 8.Net 4
  - SPSS 18 or higher
  - SAS 9 or higher
  - Microsoft Excel 2003 or higher
  - Administrator rights required for installation but not for use
- [IDB Analyzer installation package](#) can be downloaded from the IEA website
- Interface has two modules: Merge and Analysis
Using the IDB Analyzer Merge Module

Merge Module allows for creation of analysis datasets by
- Combining specific country data files
- Selecting specific subsets of variables

Using the IDB Analyzer Merge Module (Continued)
Selecting Countries

![Image of Selecting Countries](image1)

Selecting Variables

![Image of Selecting Variables](image2)
Saving Your Dataset

- Select the syntax in SPSS by hitting Ctrl+A
- Run by hitting Ctrl+R, or by going to the Run menu and selecting "All"
Combining U.S. national prison and household files

- For combining the U.S. national prison Public Use File (PUF)/Restricted Use File (RUF) with household PUF/RUF, it is necessary to first combine the two files outside of the IDB Analyzer, in SPSS.
- To combine the files, assuming prison file is opened first:
  - Use the SPSS menu options to add cases from the combined household U.S. national file: DATA -> Merge Files -> Add Cases
  - Unpaired variables are prison-only variables and one could add them to have a complete set of both prison and household variables, although the data will be missing for prison-only variables for household cases.
  - Save the merged file.
- Then, proceed to use the IDB Analyzer Analysis module:
  - Results for household sample and prison sample computed separately.

Using the IDB Analyzer Analysis Module

- Accounts for complex sample design with plausible values methodology, including correct calculation of standard errors.
- Can be used to calculate estimates of percentages, means, regression coefficients, proficiency levels, and percentiles.
Considerations for Analysis of PIAAC Data

Slide 13 of 39

Overview of Analysis Steps

- Step 1: Select the analysis file and analysis type
- Step 2: Select the statistic type
- Step 3: Select variables and parameters
- Step 4: Specify output files

Slide 14 of 39

Step 1: Selecting Analysis File and Type

Select your PIAAC data

Selecting Analysis File and Type
Step 2: Selecting the Statistic Type

Statistic Types

- Percentages and means
- Percentages only
- Linear Regression
- Logistic Regression
- Correlations
- Percentiles
- Benchmarks (discrete or cumulative options)
- Group Differences by Performance

*Correct standard errors calculated for all analytic procedures, and presented with statistics in output files to evaluate statistical significance
### Statistic Types - Percentages and Means

- **Without Plausible Values (PVs)**
  - Computes percentages, means, and standard deviations for selected variable(s), by subgroups defined by grouping variable(s)
  - Used when only background variables are included in the analysis
- **With Plausible Values (PVs)**
  - Computes percentages, mean achievement scores, and standard deviations based on plausible values, by subgroups defined by grouping variable(s)
  - Used when assessment domains are included in the analysis

### Statistic Types - Percentages Only

- Computes percentages by subgroups defined by grouping variable(s)
- No access to Plausible Values
- Achievement scores cannot be computed with this option
Statistic Types - Linear Regression

- Without Plausible Values
  - Computes standardized regression coefficients for selected independent variables to predict a dependent variable, by subgroups defined by grouping variable(s)
- With Plausible Values
  - Computes standardized regression coefficients for selected independent variables to predict a dependent variable, whereby plausible values can be used as dependent or independent variable, by subgroups defined by grouping variable(s)
  - More than one Plausible Values domain can be specified as the independent variable in order to study the relationship between the domains

Statistic Types – Logistic Regression

- Without Plausible Values
  - Computes logistic regression coefficients for selected independent variables to predict a dichotomous dependent variable, by subgroups defined by grouping variable(s)
- With Plausible Values
  - Computes logistic regression coefficients for selected independent variables to predict a dichotomous dependent variable, whereby plausible values can be used as the independent variable, by subgroups defined by grouping variable(s)
  - More than one Plausible Values domain can be specified as the independent variable in order to study the relationship between the domains
  - Plausible values cannot be used as a dependent variable
### Statistic Types - Correlations

- **Without Plausible Values**
  - Computes correlation coefficients for selected variables by subgroups defined by grouping variable(s)

- **With Plausible Values**
  - Computes correlation coefficients for selected variables that include the plausible values, by subgroups defined by grouping variable(s)

### Statistic Types - Percentiles

- **Without Plausible Values**
  - Calculates averages for defined percentile groups for any continuous variable by subgroups defined by grouping variable(s)

- **With Plausible Values**
  - Computes mean scores for defined percentile groups for a specified domain, by subgroups defined by grouping variable(s)
Considerations for Analysis of PIAAC Data

**Slide 23 of 39**

**Statistic Types - Benchmarks**

Computes percentages of the population within, at, above, or below benchmarks of performance defined by cut-point scores, by subgroups defined by grouping variable(s)

- Cumulative percentages (e.g., Level 2 and below)
- Discrete percentages (e.g., at Level 2)
- Discrete with Analysis Variable(s) (e.g. variable mean of those at Level 2)

NOTE: The proficiency score ranges for Literacy and Numeracy are as follows: Below Level 1 = 0-175; Level 1 =176-225; Level 2 = 226-275; Level 3 = 276-325; Level 4 = 326-375; and Level 5 = 376-500. The proficiency score ranges for PS-TRE are as follows: Below Level 1 = 0-240; Level 1 =241-290; Level 2 = 291-340; Level 3 = 341-500.

**Statistic Types - Group Differences by Performance**

- Without Plausible Values
  - Computes differences in an analysis variable, between groups, across different achievement or performance levels by subgroups defined by grouping variable(s)
- With Plausible Values
  - Computes differences in an analysis variable, between groups, across different achievement or performance levels based on plausible values by subgroups defined by grouping variable(s)
Step 3: Selecting the Variables and Parameters

First select grouping variables, then analysis variables

If you are using percentiles or benchmarks, enter them here separated by spaces (no commas or semicolons)

IDB Analyzer Parameters

Grouping Variables

- Variables used to define your subgroups
- List can consist of one or more variables
- As all PIAAC analyses are computed by country, IDB Analyzer always includes CNTRYID as the first grouping variable
- If "Exclude Missing from Analysis" is checked, only cases with non-missing values in the grouping variables will be used in the analysis
IDB Analyzer Parameters (Continued)

Analysis Variables
- Independent variables can be continuous or categorical
- Depending on statistic type, analysis variables may be
  - Those for which means will be computed
  - Those that will be correlated
  - Those for which percentiles will be calculated
  - Those that will be used as predictors in regression model

IDB Analyzer Parameters (Continued)

Achievement Scores (Plausible Values)
- Computed through plausible values
  - Select the "Plausible Values" parameter area
  - Choose the achievement scale you are interested in
### IDB Analyzer Parameters (Continued)

**Dependent Variable (Regression only)**
- Predicted by the list of analysis variables in regression model
- May be plausible values (for linear regression) or a variable from the background questionnaire (for logistic or linear regression)
- Only one dependent variable can be listed

**Analysis Group Variable (Group Differences by Performance Group only)**
- Will be used to classify the cases into 2 or more groups that will be compared within each of the levels of performance or achievement

**Performance Variable (Group Differences by Performance Group only)**
- Will be used to classify the cases by achievement or performance levels
- May be plausible values or a variable from the background questionnaire
IDB Analyzer Parameters (Continued)

Achievement Benchmarks/Percentiles/Cutpoints

- Values that will be used as cut points of the achievement scales and other selected continuous variables
- PIAAC benchmark cut point scores can be found on the slide titled "Statistic Types - Benchmarks"

Step 4: Specifying Output Files

- Specify the file name and location (e.g. C:\temp\PIAAC\analysis.*)
- Click "Start SPSS" to save your code and open it in SPSS
Running Your Analysis

Run the created syntax file in SPSS by selecting all (Ctrl+A) and then run (Ctrl+R), or by going to the Run menu and selecting "All"

Further Considerations for Data Analysis

- Data collected from samples representative of groups, not individuals
- U.S. data cannot be disaggregated to state or county level
- PIAAC data are cross-sectional and only capture one point in time
  - Several measures (including literacy and numeracy) can be compared with the 1994 IALS and 2003 ALL
- Because dataset was derived from non-experimental research, data should only be analyzed in terms of non-causal relationships (correlation rather than causation)
Correctly Wording Research Questions

Consider 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?
  - Are those who earn more money more likely to use literacy skills at work?

- We should not ask...
  - Does the use of literacy skills at work affect (increase/decrease) earnings?

Statistical Significance Testing

- T-tests are used to examine and report the results
- Conventionally, the difference in the point estimates is said to be statistically significant at an alpha level of 0.05
- Two T-tests are used
  - T-test for Independent groups
    - Used to compare population statistics where there is no overlap in terms of sampled adults representing these populations (e.g. male vs. female)
  - T-test for Dependent groups (Part-Whole comparison)
    - Used when part-whole comparisons are being made (e.g., comparing a country/territory to the International average)
Statistical Significance Testing (Continued)

- T-test for Independent Groups
- Let $A_i =$ the statistic in question (e.g., the mean for group i)
- $S(A_i) =$ the standard error of the statistic
- The estimates for groups i and j are said to be significantly different, if

$$\frac{|A_i - A_j|}{\sqrt{S(A_i)^2 + S(A_j)^2}} \geq T_{\alpha/2}$$

Where $T_{\alpha/2}$ is the $\alpha$ equal to 0.05 percentile of the t distribution with df degrees of freedom

---

Statistical Significance Testing (Continued)

T-test for Dependent groups (Part-Whole comparison)

- $A_i$ - the statistic in question (e.g., the mean for group i)
- $S(A_i)$ - the standard error of the statistic
- Group i is the whole (e.g., International Average) and group j is the part (e.g., United States)
- Groups i and j are said to be significantly different if

$$\frac{|A_i - A_j|}{\sqrt{S(A_i)^2 + (1-2p)S(A_j)^2}} \geq T_{0.05}$$

Where $p$ is the proportion of adults from group j within group i
Module Summary and Resources

Summary
- Described the use of the IEA IDB Analyzer
- Described considerations for analyzing PIAAC data

Resources
- [IEA IDB Analyzer installation package](#)
- [Statistic Types - Benchmarks](#)