Session I: Introduction

Cross-national Studies: Interdisciplinary Research and Training Program - CONSIRT, consirt.osu.edu
the Polish Academy of Sciences (PAN) and The Ohio State University (OSU)
Mershon Center for International Security Studies
OSU Departments of Sociology and Political Science
OSU Polish Study Initiative
Presenters:

Weights – Introduction by Kazimierz M. Slomczynski
- Marcin Zieliński, Polish Academy of Sciences, University of Warsaw

Data Quality Indicators - Introduction by Kazimierz M. Slomczynski
- Matt Schoene, Ohio State University
- Marta Kołczyńska, Ohio State University
- Ilona Wysmułek, Graduate School for Social Research (Warsaw)
- Olena Oleksiyenko, Graduate School for Social Research (Warsaw)
- Przemek Powałko, Polish Academy of Sciences

Multi-level modeling - Introduction by Kazimierz M. Slomczynski
- Robert Kunovich, University of Texas at Arlington
Multi-survey data analysis

Goal of the training event: Bringing weights for units of observation and data-quality indicators into analysis by means of multi-level modeling

Multi-survey data structure:
(1) Survey project (SP): a collection of national surveys under a common administrative umbrella (e.g., European Social Survey, World Values Survey)
   Differentiation of SPs with respect to thematic focus and organizational features

(2) Within SPs are project waves (PW), which contain the surveys conducted at the same time according to the same instrument
   PWs differ within SPs

(3) Within PW are national surveys (NS) conducted in given countries or “national territories”
   NSs differ within PWs

(4) NSs include individuals (ID)
Weights and data-quality indicators

• In the context of multi-survey data, weights are numerical values of the importance of a given unit of observation: an individual (ID), national survey (NS), wave (PW), project (SP).

• Data-quality indicators can deal with records for individuals (ID), documentation for each survey (NS), documentation of data for a given wave (PW) and properties of a survey project (SP).
<table>
<thead>
<tr>
<th>LEVEL</th>
<th>WEIGHTS</th>
<th>DATA QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURVEY PROJECT (SP)</td>
<td>Analytic and importance weights: selection and clustering</td>
<td>Quality of cross-national coordination and control</td>
</tr>
<tr>
<td>WAVE (PW)</td>
<td></td>
<td>Quality of data description</td>
</tr>
<tr>
<td>NATIONAL SURVEY (NS)</td>
<td>Sample-size weights&lt;br&gt;Population-size weights</td>
<td>Quality of survey administration</td>
</tr>
<tr>
<td>INDIVIDUAL (ID)</td>
<td>Design weights&lt;br&gt;Post-stratification weights</td>
<td>Quality of computer records</td>
</tr>
</tbody>
</table>
Using weights and data-quality indicators

Explanation of attitudes and behaviors by variables on individual level and higher-order level(s)

E.g.: DV = trust in parliament ($Y_{ijk}$) by individuals $i$ in surveys $j$ from project-waves $k$

IVs: gender ($X_1$), age ($X_2$), education ($X_3$), type of political regime (measured by political rights, $Z_1$), and economic development (GDP per capita, $Z_2$)

$Y_{ijk} = f(X_{ijk}, Z_{.jk})$

Including Weights ($W$) and Data-Quality Indicators ($Q$)

$Y_{ijk} = f(X_{ijk}, Z_{.jk}, W_{ijk}, W_{.jk}, W_{..k}, Q_{ijk}, Q_{.jk}, Q_{..k})$
Data Recycling of International Survey Projects
Survey Data Recycling (SDR)

- SDR = framework for (re)-processing cross-national survey data;

- SDR concerns survey data quality control & ex-post survey harmonization, to:
  
  (a) account for “messiness” of the original source data,

  (b) provide “comparable” data

Thus, SDR expands the scope of extant projects (time, space, number of observations, types of indicators)

**Unifying thesis:** account for errors & biases in original surveys & harmonization procedures *via* different types of quality control variables, to be included in substantive analyses.
Figure 1. General Schema of Survey Data Recycling

Input
Materials on national surveys from different sources and in different forms

Organized original materials
- General survey documentation
- Specific description of the data
- Computer data files

Transformed materials
- Standardized survey documentation
- Integrated codebook
- Data file of selected source variables

Quality evaluation
- Quality-control variables for general survey documentation, specific data description, & the computer data file
- Weights

Ex-post Harmonized data
- "New" technical variables
- Substantive target variables
- Quality-control variables for target variables
- Cross-national comparability checks

Output
INTEGRATED DATABASE
Standardized documentation and integrated codebook; integrated data file with harmonized technical & substantive variables, quality control indicators
Criteria of selecting survey projects

- containing questions about political attitudes and behaviors
- designed as cross-national, and, preferably, multi-wave;
- with the samples intended as representative of the adult population of given country or territory;
- non-commercial;
- freely available in the public domain;
- with documentation - study description, codebook and/or questionnaire - in English
<table>
<thead>
<tr>
<th>Abbrev.</th>
<th>Survey Project</th>
<th>Time span</th>
<th>Waves</th>
<th>Files</th>
<th>Data Sets</th>
<th>Cases Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFB</td>
<td>Afrobarometer</td>
<td>1999-2009</td>
<td>4</td>
<td>4</td>
<td>66</td>
<td>98942</td>
</tr>
<tr>
<td>AMB 1</td>
<td>Americas Barometer</td>
<td>2004-2012</td>
<td>5</td>
<td>1</td>
<td>92</td>
<td>151341</td>
</tr>
<tr>
<td>ARB</td>
<td>Arab Barometer</td>
<td>2006-2011</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>19684</td>
</tr>
<tr>
<td>ASB</td>
<td>Asian Barometer</td>
<td>2001-2011</td>
<td>3</td>
<td>3</td>
<td>30</td>
<td>43691</td>
</tr>
<tr>
<td>ASES 2</td>
<td>Asia Europe Survey</td>
<td>2000</td>
<td>1</td>
<td>1</td>
<td>18</td>
<td>18253</td>
</tr>
<tr>
<td>CB 3</td>
<td>Caucasus Barometer</td>
<td>2009-2012</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td>24621</td>
</tr>
<tr>
<td>CDCEE 4</td>
<td>Consolidation of Democracy</td>
<td>1990-2001</td>
<td>2</td>
<td>1</td>
<td>27</td>
<td>28926</td>
</tr>
<tr>
<td>CNEP 5</td>
<td>Comparative National Elections Project</td>
<td>2004-2006</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>13372</td>
</tr>
<tr>
<td>EB 6</td>
<td>Eurobarometer</td>
<td>1983-2012</td>
<td>7</td>
<td>7</td>
<td>152</td>
<td>138753</td>
</tr>
<tr>
<td>EQLS 7</td>
<td>European Quality of Life Survey</td>
<td>2003-2012</td>
<td>3</td>
<td>1</td>
<td>93</td>
<td>105527</td>
</tr>
<tr>
<td>ESS 8</td>
<td>European Social Survey</td>
<td>2002-2013</td>
<td>6</td>
<td>2</td>
<td>146</td>
<td>281496</td>
</tr>
<tr>
<td>EVS_9</td>
<td>European Values Study</td>
<td>1981-2009</td>
<td>9</td>
<td>1</td>
<td>312</td>
<td>423084</td>
</tr>
<tr>
<td>ISJP 10</td>
<td>International Social Justice Project</td>
<td>1991-1996</td>
<td>2</td>
<td>1</td>
<td>21</td>
<td>25805</td>
</tr>
<tr>
<td>ISSP 11</td>
<td>International Social Survey Programme</td>
<td>1985-2013</td>
<td>13</td>
<td>13</td>
<td>363</td>
<td>493243</td>
</tr>
<tr>
<td>LB 12</td>
<td>Latinobarometro</td>
<td>1995-2010</td>
<td>15</td>
<td>15</td>
<td>260</td>
<td>294965</td>
</tr>
<tr>
<td>LITS 13</td>
<td>Life in Transition Survey</td>
<td>2006-2010</td>
<td>2</td>
<td>2</td>
<td>64</td>
<td>67866</td>
</tr>
<tr>
<td>NBB 14</td>
<td>New Baltic Barometer</td>
<td>1993-2004</td>
<td>6</td>
<td>1</td>
<td>18</td>
<td>21601</td>
</tr>
<tr>
<td>PA2</td>
<td>Political Action II</td>
<td>1979-1981</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4057</td>
</tr>
<tr>
<td>PA8NS</td>
<td>Political Action – 8 Nation Study</td>
<td>1973-1976</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>12588</td>
</tr>
<tr>
<td>PPE7N</td>
<td>Political Participation in 7 Nations</td>
<td>1966-1971</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>16522</td>
</tr>
<tr>
<td>VPCPCE_15</td>
<td>Values/POLITICAL Change</td>
<td>1993</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>4723</td>
</tr>
<tr>
<td>WVS 16</td>
<td>World Values Survey</td>
<td>1981-2009</td>
<td>9</td>
<td>1</td>
<td>312</td>
<td>423084</td>
</tr>
</tbody>
</table>
Data

• 1,721 national surveys with over 95% of samples from 500 to 3,000 respondents

• National surveys conducted in 142 countries and territories over a period of almost 50 years

• All these surveys contain over 25,000 variables

• From 150 to 200 variables are identical or similar in large majority of 89 waves

• From 25 to 40 variables deal directly with political attitudes and behaviors

Importance of the SDR project for social sciences: opportunities to study various aspects of attitudes and behaviors
Data for training event

- European countries + USA
- Time: 1989-2013
- 16 survey projects (SP), each containing – in our dataset – from 1 to 12 waves, altogether 61 project waves (PW)
- 890 national surveys (NS)
Data for training event
<table>
<thead>
<tr>
<th>LEVEL</th>
<th>Number</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURVEY PROJECT (SP)</td>
<td>16</td>
<td>Different access to data</td>
</tr>
<tr>
<td>WAVE (PW)</td>
<td>61</td>
<td>For some projects only one wave; from 1 to 12 per project</td>
</tr>
<tr>
<td>NATIONAL SURVEY (NS)</td>
<td>890</td>
<td>NSs contain surveys conducted in national territories; average sample = 1,368 cases</td>
</tr>
<tr>
<td>INDIVIDUAL (ID)</td>
<td>1,217,455</td>
<td>-</td>
</tr>
</tbody>
</table>
Formal framework for studying the effect of weights

1. Relationship between target variables T and source variables S
   \[ T = f(S) \] [substantive decision of f]

2. Relationship between T and X and weights W
   \[ X = \text{Substantive independent variables} \]
   \[ W = \text{Weights} \]
   \[
   \begin{align*}
   (1) & \quad T = b_0 + b_1 X + e & \text{without weights} \\
   (2) & \quad T = b_0 + b_1 X + e & \text{with weights}
   \end{align*}
   \]

If e is negligible & \( b_1 \) in (1) = \( b_1 \) in (2) no effect of weights.

The impact of weights on standard errors.