Analysing and Understanding Learning Assessment for Evidence-based Policy Making

Introduction to survey data structure

Dr Alvin Vista, ACER
Bangkok, 14-18, Sept. 2015
Why do we survey?

• One reason is to use the survey results to explore things about the whole population
  – What is the mathematics achievement of students in country X?
  – What background factors appear to be connected with outcomes in mathematics?
Sample versus Census - 1

• Surveys usually draw a sample because a census:
  – Requires substantially increased financial and logistical resources
  – Time consuming and thus incompatible with deadlines
  – Does not necessarily bring additional and required information
Sample versus Census - 2

• Several ways of drawing a sample

• Two major principles:
  – Avoiding bias in the selection
    • Randomness
    • Sampling frame covers the population adequately
    • Ensuring the varying probabilities of selection are reflected ⇒ weights
  – Achieving maximum precision, given a set of constraints (time, funds, and so on)
Advantages of sampling

• Reduced costs
  – Data collection
  – Data analysis
  – Reduced training of personnel to conduct fieldwork

• Tighter control of field operations

• Improved speed in data analysis and reporting
Advantages of sampling

• The population divided into mutually exclusive units called clusters
• A sample of these is selected to participate in the survey
  – (not all clusters are selected)
  – (compare with stratification: a sample taken from every stratum)
• Clusters for educational surveys
  – Schools
  – Sections
• With area sampling at the first stage
  – (well defined) regions are clusters
Advantages of sampling

- Easier and cheaper to administer to groups of students within schools
- We don’t need a full list of the student population
- Limits burden to the set of sampled schools
- Better control of field operations
- Analyses of school or class influences possible
Advantages of sampling

Multi-stage approaches
  - 1) a sample of schools
  - 2) a sample of students from those schools

• TIMSS
  - 1) Sample schools
  - 2) Sample intact classes from those schools

• PISA
  - 1: Sample schools
  - 2: Sample 15 YOs from those schools
Multistage sampling
Field Trial Samples

• Non-probability samples used in ‘field trials’ or pilots
  – Aim is to field-test the instruments (e.g., test questions, questionnaire items), as well as procedures
  – No intention to make statistical inferences, report estimates of population characteristics, etc.
  – Estimates of precision impossible without further modelling/assumptions
  – A convenience or judgment sample is usually quite adequate

• Examples of non-probability sampling
  – Judgement sampling
  – Convenience sampling
  – Quota sampling
Main survey sampling

• ‘Probability’ sampling
  – The common characteristic of these sampling methods is that each element in the population has a known, non-zero probability of selection into the sample
  – This quality allows for measures of sampling precision to be calculated directly
Key components

1. The definition of the student population to be surveyed.
2. The development of a list of schools in which eligible students are enrolled.
3. The selection of the sample of schools from this list.
4. The development of a list of eligible students within each sampled school.
5. The selection of the sample of students from within each sampled school.
6. The documentation of the sampling process, and the calculation of indicators of the effectiveness of the sampling and survey operations.
Field trial vs Main survey

- More rigorous requirements on the sampling procedures compared to the field trial, to enable valid statistical inferences about the student population.
- Definition of the student population is much more rigorously applied for the main study than in the field trial.
- The development of a list of schools, and sampling from this list, requires much more care and preparation for the main study.
- A probability sampling approach is usually preferred
  - Robust estimates of precision for reporting
  - Efficient
  - Methods can be standardised across subpopulations and over time
- The procedures for listing and sampling students within schools will be very similar to the field trial.
Definitions
Distinction between the Desired Population and the Defined Population

- Desired Population = the one to which inferences from the sample outcomes will be made.
- Defined Population = the one that is actually studied.
Population definition structure
SEA-PLM example

• Sampling standards will ensure that students tested come from the same target population in every country, to enable comparability across countries.

• Regional Desired Target Population
  – Students in grade 5
  – Schools in SEAMEO countries
  – Schools following the mainstream curriculum
  – Must cover eligible students as much as possible

• Country-level Defined Target Population
  – Desired population minus exclusions
  – At least 95% of desired target population
  – Total exclusions should not exceed 5% of the desired target population
Exclusions

• Important that exclusions are kept to a minimum and documented
• Variations across countries will affect comparability
• School-level and within-school exclusions should not exceed 5% of the desired target population
• Technical Standards for acceptable limits on exclusions may be provided depending on need.
Exclusions

• School-level exclusions (examples):
  – Schools with curriculum, or school structure, radically different from the mainstream educational system
  – Very small schools
  – Geographically inaccessible
  – Schools exclusive to students in the within-schools exclusion categories

• Within-schools exclusions (examples):
  – Functional disability
  – Limited proficiency in assessment language
Field trial sampling

• School level variables related to educational outcomes for each country should be considered when determining the list of schools
• Additional backup schools across the identified strata to be used as replacement schools
• Sampled schools should not be used as replacement school for another sampled school
• Good to aim for 85% participation rate or greater among sampled schools
• Communication with selected schools is important. Make sure they are engaged and cooperative!
Field trial sampling
SEA-PLM example

• Expected sample sizes, number of schools (Field Trial)
  – Required sample size for the field trial = 1800, ~35 schools w/ students at Grade 5 (minimum 30 schools)
  – Most schools selected should have at least two classes (target cluster size ~60-70 students per school)
  – Acceptable to include more than 2 classes with fewer students

• Broadly representative -- stratification variables include:
  – urbanisation (urban/rural)
  – funding (e.g. public/private)
  – management (e.g. secular/religious)
  – school type (e.g. primary / middle)
  – socio-economic background (e.g. wealthier or poorer locations)
Replacement schools

• Example situation: 20 schools are sampled from a stratum, but only 18 participate
  – The two schools that did not participate might lead to some survey bias
  – E.g., the non-respondents are both metropolitan schools, and so the participant data is biased away from metropolitan school students

• Find a school very similar to each non-responding school and use those schools as substitutes

• Care must be taken to ensure that the adjacent schools have not crossed implicit stratum boundaries.
  – For example, sampled school is the last urban school in the sorted stratum, and the next school is a rural school.
Summary

- Many large scale surveys involve sampling
- Sampling purpose is different between field trial and main survey:
  - Convenience sample for the pilot phase
  - Probability sample for the main survey
- Development of population definitions, exclusions and response rates across the region. Trial provides information for the main survey.
- Field trial prepares and tests the procedures for listing and sampling students within schools.
Checklist for Sampling

- Is the **national** Target Population defined?
- What limitations to full coverage of this population are expected?
- Do you expect some States/Provinces to have particular coverage issues?
- Do **nationally consistent** definitions exist that can be applied to school and student level exclusion categories?
- Will schools be excluded on the basis of size? What is the minimum size?
- Will school level exclusions be able to be applied consistently across States/Provinces?
- Will student level exclusions be applied consistently?
- Do you expect some States/Provinces to have difficulties meeting the standards with respect to exclusions?
- Other concerns about population and coverage?
Technical details of Systematic Sampling
Simple random sample

• An SRS gives an equal probability of selection for each \( n \) selected from population \( N \) which is

\[
p_i = \frac{n}{N}
\]

• and each selected \( n \) has a weight of

\[
w_i = \frac{1}{p_i} = \frac{N}{n}
\]

for example \( w_i = \frac{400}{40} = 10 \)
Using weights with SRS

- Since the weights are the same for all selected members of the sample, weighted and unweighted means are identical:
Using weights with SRS

- The sum of the weights always rebuilds the size of the population

\[ \sum_{i=1}^{n} w_i = \sum_{i=1}^{n} \frac{N}{n} = N \]
Student surveys

• Simple random sampling generally too expensive and often impossible.
  – Often no central database for sampling of individual students
  – Use of SRS is impractical for survey administration.
  – No link of students to classrooms or schools available.

• Therefore, two-stage designs common, where schools are selected at first stage and students or intact classrooms at the second.
Two stage sample

- $n$ schools are selected from a population of $N$ schools
  - $\Rightarrow$ what is the probability of a particular school to be drawn?

- Within school $i$, $n_i$ students or $n_i$ classrooms are drawn from a population of $N_i$ students or $N_i$ classroom
  - $\Rightarrow$ what is the probability of a particular student in school $i$ to be drawn?
Example of a two stage sample

• Let us imagine:
  – a population of 10 schools
  – Within any school, there are 40 students
  – So the educational system counts 400 students

• 4 schools are drawn according to a SRS procedure

• 10 students within the 4 selected schools are selected according to a SRS
Two-stage probabilities

- Four schools selected out of 10
  \[ p_{1_i} = \frac{n_{sc}}{N_{sc}} = \frac{4}{10} = 0.4 \]

- 10 students selected within each school
  \[ p_{2_{ij}} = \frac{n_i}{N_i} = \frac{10}{40} = 0.25 \]
Final weights (SRS two-stage)

- The final selection probability is

\[ p_{ij} = p_{1\_i} p_{2\_ij} = \frac{n_{sc} n_i}{N_{sc} N_i} \]

- and the final total weight is the product of school and student weight or the inverse of the selection probability

\[ w_{ij} = \frac{1}{p_{ij}} = \frac{1}{0.1} = 10 \]
Weights summary

<table>
<thead>
<tr>
<th>School label</th>
<th>School size</th>
<th>School prob.</th>
<th>School weight</th>
<th>Within school prob.</th>
<th>Within school weight</th>
<th>Final student prob.</th>
<th>Final student weight</th>
<th>Sum of final weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>0.4</td>
<td>2.5</td>
<td>0.25</td>
<td>4</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0.4</td>
<td>2.5</td>
<td>0.25</td>
<td>4</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>0.4</td>
<td>2.5</td>
<td>0.25</td>
<td>4</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>0.4</td>
<td>2.5</td>
<td>0.25</td>
<td>4</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>0.4</td>
<td>2.5</td>
<td>0.25</td>
<td>4</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400</td>
</tr>
</tbody>
</table>
Different school sizes

<table>
<thead>
<tr>
<th>School label</th>
<th>School size</th>
<th>School prob.</th>
<th>School weight</th>
<th>Within school prob.</th>
<th>Within school weight</th>
<th>Final student prob.</th>
<th>Final student weight</th>
<th>Sum of final weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>0.4</td>
<td>2.5</td>
<td>0.66</td>
<td>1.5</td>
<td>0.27</td>
<td>3.75</td>
<td>37.5</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>0.4</td>
<td>2.5</td>
<td>0.33</td>
<td>3</td>
<td>0.13</td>
<td>7.5</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>0.4</td>
<td>2.5</td>
<td>0.25</td>
<td>4</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>0.4</td>
<td>2.5</td>
<td>0.1</td>
<td>10</td>
<td>0.04</td>
<td>25</td>
<td>250</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>462.5</td>
</tr>
</tbody>
</table>
From SRS to PPS Sampling

• As school sizes differ across schools, then:
  – student weights will vary across schools.
    • In smaller schools, probability of selection is higher.
    • In larger schools, probability of selection is lower.
  – Very large uncertainty on the student population estimates ⇒ population estimates will have a large sampling variance
Sampling with school *probabilities proportional to size* (PPS)

- Larger schools get a higher probability of selection than smaller schools...
- ...but students in larger schools have a smaller within-school probability of being selected than students in small schools, and...
- the school and within school weights differ between schools, but final student weights do not vary (‘self-weighted sample’).
School weights with PPS sampling

- Probability of a school to be selected is equal to the ratio of the school size multiplied by the number of schools to be sampled and divided by the total number of students in the population:

\[ p_{1_i} = \frac{N_i \times n_{sc}}{N} \]
# PPS sample of schools

<table>
<thead>
<tr>
<th>School label</th>
<th>School size</th>
<th>School prob.</th>
<th>School weight</th>
<th>Within school prob.</th>
<th>Within school weight</th>
<th>Final student prob.</th>
<th>Final student weight</th>
<th>Sum of final weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>0.2</td>
<td>5.00</td>
<td>0.500</td>
<td>2.0</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>40</td>
<td>0.4</td>
<td>2.50</td>
<td>0.250</td>
<td>4.0</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>80</td>
<td>0.8</td>
<td>1.25</td>
<td>0.125</td>
<td>8.0</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>1</td>
<td>1.00</td>
<td>0.100</td>
<td>10.0</td>
<td>0.1</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>400</td>
<td>9.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400</td>
</tr>
</tbody>
</table>
Replacement schools

• From a conceptual point questionable but in practice usually applied
  – Attempt to increase sampling coverage

• Selection of adjacent two schools as possible replacements in case of refusals/non-participation:
  – 1\textsuperscript{st} replacement next in sampling frame
  – 2\textsuperscript{nd} replacement preceding school

• In case of school closures NO replacement is not appropriate!
Explicit stratification

• Explicit stratification is used to:
  – increase sampling precision
  – over- or under-sample certain sub-groups of schools
  – to deal with small or large schools

• Explicit strata are treated as separate sampling frames

• Samples are drawn separately for each explicit stratum
Implicit stratification

• Implicit stratification is applied when selecting schools systematically from a sampling frame
• Implicit stratification ensures that schools appear in the sampling frame next to schools with similar characteristics
• Typical variables: Rural/Urban, SES deciles, regions/
Sampling Frame

• Include ALL schools with students in desired target population (except possible school-level exclusions)

• List appropriate enrolment statistics indicating number of students in desired target population

• Sort by
  – Implicit stratification variables (alternate ascending/descending order)
  – by enrolment - in alternate Ascending/Descending order within implicit strata
PPS Sampling - summary

• More efficient way of sampling
  – Reduction of schools
  – Reduction of variation in weights

• Weights may still differ because of ...
  – Over- or under-sampling of some strata in the population
  – Lack of accuracy in measures used for selection of sample
  – Adjustments to weights for school and student non-response
Practical exercise!