NEQMAP WORKSHOP

Overview of Student Learning Assessment

NEQMAP: Sampling for national assessments

Dr. Kelvin D. Gregory

Education Consultant
Adjunct at Flinders University
kelvin.gregory@flinders.edu.au

4-6 NOVEMBER, 2015 / KATHMANDU, NEPAL

Types of research using sampling

- Experiments
  - Has treament variables
  - Other variables are controlled or randomized
- Surveys
  - Participants are selected from a defined population with known probability
- Investigations
  - Participants are selected not to representative or a population
  - Participants are selected for their ability to give/yield information to build an understanding

Populations

- Desired population
  - The population that ideally would be studied
- Excluded population
  - The population that is excluded from the study
- Target population
  - The population that is actually studied
  - Provides an operational definition which may be used to give the constuction of a list for the population

PISA Population

The desired base PISA target population in each country consisted of 15-year-old students attending educational institutions located within the country, in grades 7 and higher

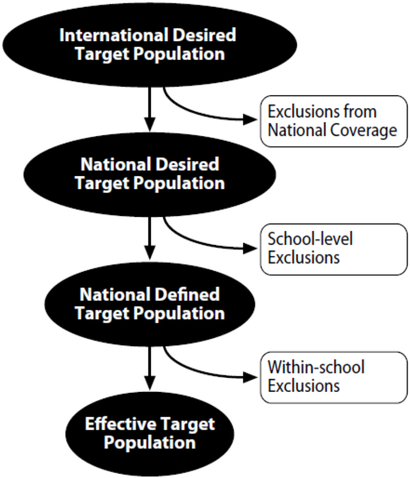
- Countries are to include 15-year-olds enrolled part-time or full-time in educational institutions, students in vocational training types of programmes, or any other related type of educational programmes, and students attending foreign schools within the country (as well as students from other countries attending any of the programmes in the first three categories).
- No testing of persons schooled in the home, workplace or out of the country, these students are not included in the international target population.

PISA Exclusions

- Schools are usually excluded for practical reasons such as increased survey costs, complexity in the sample design, or difficult test conditions.
- They can be excluded, depending on the percentage of 15-year-old students involved, if they were geographically inaccessible (but not part of a region omitted from the national desired target population), or if it was not feasible to administer the PISA assessment.

TIMSS Populations

Exhibit 5.1 Relationship Between the Desired Populations and Exclusions



TIMSS Target Population

Fourth grade population. This includes all students enrolled in the grade that represents 4 years of formal schooling, counting from the first year of ISCED Level 1, provided that the mean age at the time of testing is at least 9.5 years. For most countries, the target grade should be the fourth grade or its national equivalent.

TIMSS School Exclusions

TIMSS participants were expected to ensure that the national defined populations included at least 95 percent of the national desired populations of students.

TIMSS School Exclusions

- Exclusions can occur at the school level, within the sampled schools, or both.
- Countries are expected to do everything possible to maximize coverage of the national desired populations.
- Keeping within the 95 percent limit, school-level exclusions from the sampling frame could be for the following reasons:
  - Schools were geographically remote.
  - They had very few students.
  - The curriculum or structure at the school was different from the mainstream education system.
  - Schools were specifically for students with special needs.

TIMSS Within School Exclusions

- Students with intellectual disabilities. These are students who are considered, in the professional opinion of the school principal or by other qualified staff members, to be intellectually disabled or who have been tested psychologically as such.

TIMSS Within School Exclusions

- Students with functional disabilities. These are students who are physically disabled in such a way that they cannot perform in the TIMSS testing situation. Functionally disabled students who are able to perform should be included in the testing.

TIMSS Within School Exclusions

- Non-native language speakers. These are students who are unable to read or speak the language(s) of the test and would be unable to overcome the language barrier of the test. Typically, a student who has received less than 1 year of instruction in the language(s) of the test should be excluded, but this definition may need to be adapted in different countries.

Sampling

- Often require judgements
  - A balance between the accuracy of the measurements obtained and the reporting requirements
- Before sampling
  - Clarify the reporting requirements
  - What will the report look like?
- What groups and subgroups need to be reported on?
  - Clarify the precision requirements

TIMSS Within school exclusions

- Students with intellectual disabilities who are considered, in the professional opinion of the school principal or by other qualified staff members, to be intellectually or psychologically disabled.
- Students with functional disabilities are students who are physically disabled and cannot perform in the TIMSS testing situation. Functionally disabled students who are able to perform should be included in the testing.
- Non-native language speakers who are unable to read or speak the language(s) of the test and would be unable to overcome the language barrier of the test.

Sampling

- Often require judgements
  - A balance between the accuracy of the measurements obtained and the reporting requirements
- Before sampling
  - Clarify the reporting requirements
  - What will the report look like?
  - What groups and subgroups need to be reported on?
  - Clarify the precision requirements

Representativeness

- Sample characteristics may be compared with known population characteristics
  - Use 'marker variables' that are readily available in both samples and population
    - ★ But there are no rules to say how close the statistics (from the sample) should be to the parameters (from the population)
    - ★ Also agreement between sample and population markers does not necessarily transfer to variables of interest
    - ★ Just because the gender and age markers agree does not mean that the assessment data obtained from the sample is representative of the population
    - ★ Need to use estimates of the standard errors

Common Marker Variables

- In education, the common markers are
  - Gender
  - Year or grade of students
  - Age of students (aggregated or disaggregated)
  - School type
  - Location of school

Types of Samples

- Probability samples
  - Sample is selected so that it should be representative of the population
  - Must use random sampling strategy
  - Each element in population has a known probability of being selected
    - ★ Must be non-zero
    - ★ Can be one
- Nonprobability sample
  - Sample is selected so that it can inform theory
  - Elements in population often have unknown probability of being selected
    - ★ Although some elements have zero chance or are selected with certainty

Sample Size

- Determined by three factors
  - Desired level of precision
    - The level of precision, sometimes called sampling error, is the range in which the true value of the population is estimated to be. This range is often expressed in percentage points, (e.g., $\pm$ 5 percent).
  - Confidence level
    - The confidence or risk level is based on ideas encompassed under the Central Limit Theorem. The key idea encompassed in the Central Limit Theorem is that when a population is repeatedly sampled, the average value of the attribute obtained by those samples is equal to the true population value.

Sample Size

- Determined by three factors
  - Variability
    - ★ The third criterion, the degree of variability in the attributes being measured refers to the distribution of attributes in the population. The more heterogeneous a population, the larger the sample size required to obtain a given level of precision.

Sample Size

To estimate a proportion with a 95% confidence interval within $\pm$ 5%

$$var(p) \geq \frac{p(100-p)}{n^*-1}$$

$$(2.5)^2 \geq \frac{50(100-50)}{n^*-1}$$

$$n^* \geq 400$$

Intraclass Correlation

- There are many forms of intraclass correlations
  - Used to describe how strongly the units within the same group resemble each other
- We basically want one that shows the size of relationship within a school
  - That is the level of shared variance for our subjects of interest. Thus, we really have a one way analysis of variance

Intraclass Correlation Coefficient

The easiest way to calculate it is to use a one-way analysis of variance

$$\rho = \frac{F - 1}{F + b - 1}$$

where F is the ANOVA F-value and b is the average within-school sample size

Design effect

For a typical two-stage survey of students, schools (primary sampling units) are selected at the first stage with a probability proportional to the size of the target population in each school

- At the second stage of sampling, a fixed-size random cluster of students (secondary sampling units) is selected at random from each of the schools.

Design effect

For a typical two-stage survey of students, schools (primary sampling units) are selected at the first stage with a probability proportional to the size of the target population in each school.

- At the second stage of sampling, a fixed-size random cluster of students (secondary sampling units) is selected at random from each of the schools.

Design effect

- This sample is less efficient than a simple random sample (srs) of the same size
- The efficiency of a complex sample is measured by the design effect (deff), which compares the variance errors of sampling for a complex sample and a simple random sample of the same size.
- The variance error of sampling is the square of the standard error of sampling
- $$deff = \rho(b - 1)$$

Design effect

add equation

- deff $= 1$ is equivalent to a simple random sample (srs);
- deff > 1 means that the clustering was having an influence, which we have now allowed for;
- deff < 1 would mean that the design is giving results better than for a srs, possibly as a function of the stratification

Size of simple sample equivalent

This is calculated using the intraclass correlation coefficient, average cluster size and the number of clusters

$$n_{sse} = \frac{n_c b}{deff}$$

Number of schools

- Once you know the simple random sample size needed and you have the design effect
- Calculate the sample size

$$n_c = n^* deff$$

Sampling frame

- Is a list of all the elements of the desired target population
  - This list must be accurate and comprehensive
- The sampling frame must be constructed to aid accuracy
  - One way of doing this is to use multistage sampling
    - Each stage requires an accurate list
    - In these cases, the sampling frame includes a number of important variables useful in planning for sampling

Common large scale sample designs

Two designs are commonly used

- Stratified cluster sampling with random sampling of intact classes
  - Used by TIMSS and PIRLS
  - Easiest to administer
  - Assumes classes are the same, no tracking, streaming or ability grouping
- Stratified cluster sampling with random sampling of students within sampled schools
  - Used by PISA
  - Harder to administer, but has some strengths in secondary data analysis

Sample weights

- Often studies seek to sample students with probability proportional to size
  - In these cases weights are not really necessary
- Weights are important in cases where particular groups are over- or -under sampled
  - Oversampling is often done when there is a small group which needs to be reported upon for example, monitoring disadvantage

- Often studies seek to sample students with probability proportional to size
  - In these cases weights are not really necessary

Calculating sample weights

Essentially you need weights, and weight adjustments, for each sampling stage

- Sampling of schools within each stratum
  - Weight adjustment for sampled schools that did not participate
- Sampling of classrooms within each sampled school
  - Weight adjustment for sampled classes that did not participate
- Sampling of students within each sampled classroom
  - Weight adjustment for sampled students that did not participate

Non-participation

It is important to note that sampling assumes that non-participation is random

- There should be no pattern in non-participation
  - It is important to document the reasons why schools, classrooms and students did not participate
  - These reasons should be analysed for any pattern in non-participation that may affect the validity of the results
    - For example, did only the low proficiency students not participate

Substitution

Sampling theory does not allow for sampled schools, classes, or students to be substituted

- They can be replaced according to sampling theory
  - Often the sampled school is identified with two replacement schools
- Care must be taken to verify and document that the actual sampled schools, classes and students participated

Need to implement quality control procedures to prevent substitution

School sample weight

The school sample weight is the inverse probability of a school being selected

$$BW_{sc}^{i} = \frac{M}{nm_i}$$

- M is the number of students in the stratum
- n is the number the number sampled schools in the stratum (NB: the number of sampled schools, not the number of participating schools)
- m_i is measure of size for the sampled school

School adjustment for non-participation

The school weight is adjusted if a school does not participate

$$A_{sc} = \frac{n_s + n_{r1} + n_{r2}}{n_s + n_{r1}}$$

- n_s is the number of originally sampled schools that participated,
- n_{r1} the number of first replacement schools, that participated,
- n_{rr} is the number of schools that did not participate.

Final school weight

The final school weight is the product of the school weight and the school adjustment

$$F_{sc}^{i} = A_{sc} BW_{sc}^{i}$$

Class sample weight

In designs that sample intact classes within sampled schools, and assuming that the classes within a sample school were sampled with equal probability

$$FW^{i}_{cl1} = \frac{C^{i}}{c^{i}}$$

- C^{i} is the number of classes within the school
- c^{i} is the number of sampled classes within the school

Class adjustment for non-participation

If you are using a classroom sampling strategy then you will need to have a class adjustment for nonparticipation

Student sample weight

Assuming all students in the sample classroom are selected to participate

$$BW_{st1}^{i,j} = 1.0$$

However, if classes are large sub-sampling may take place. And an adjustment for non-participation will be necessary.

The overall sampling weight is simply the product of the final first stage weight, the final second stage weight, and the final third stage weight.

If intact classrooms were tested the overall sampling weight is

$$W^{i,j} = A_{sc}^{i,j} BW_{sc}^{i} FW_{cl1}^{i,j} A_{st}^{i,j} BW_{st1}^{i,j}$$

Total and house weights

- The overall weight gives an estimate of the total population size and is sometimes called the Total Weight (totwgt).
- This weight can be algebraically adjusted back to the sample size. The adjusted weight is called the house weight (houwgt) and gives the weighted contribution of each assessed student to the sample.