UNECE

UNIVERSITÀ
DI SIENA
1240

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Measuring Poverty and Inequality: SDGs 1 and 10

Assessing and Improving Survey Methods

Gianni Betti

University of Siena, Italy, gianni.betti@unisi.it

Geneva, December the 6th, 2019

UNIVERSITÀ DI SIENA 1240

Table of contents

- Survey errors and quality
- A typology of survey errors
- Errors in measurement
- Errors in estimation
- Measurement bias and sampling error for **disaggregations**
- Recommendations
- References

UNIVERSITÀ DI SIENA 1240

Survey errors and quality

To establish trust in poverty measurement and prevent misguided policies, Statistical Offices have to regularly assess and continuously improve the quality of their processes and accuracy of their data.

Quality reports which describe the quality criteria and explain any instances in which these criteria could not met, will not only assist the correct interpretation, but can also provide the basis for future improvements.

UNIVERSITÀ DI SIENA 1240

A typology of survey errors

Verma (1981), Hussmanns *et al.* (1990) and more recently Verma *et al.* (2010) propose a typology of survey errors, as follows:

(1) Errors in measurement

(2) Errors in estimation

(3) Item non-response (mixed category)

UNIVERSITÀ DI SIENA 1240

Errors in measurement

A) conceptual errors; these include: i) errors in basic concepts, definitions and classifications; ii) errors in putting them into practice (questionnaire design, preparation of survey manuals, training and supervision of interviewers and other survey workers).

B) response (or 'data collection') errors; these include: i) response bias; ii) simple response variance; iii) correlated response variance.

C) processing errors; these include: i) recording, data entry and coding errors; ii) editing errors; iii) errors in constructing target variables; iv) other programming errors.

Mixed category

- *D)* item non-response*;* this includes:

- i) only approximate or partial information sought in the survey;
- ii) respondents unable to provide the information sought (‘don’t knows’);
- iii) respondents not willing to provide the information (‘refusals’);
- iv) information suppressed (for confidentiality or whatever reason).

UNIVERSITÀ DI SIENA 1240

Errors in estimation

E) coverage and related errors; these include: i) under-coverage; ii) over-coverage; iii) sample selection errors

F) unit non-response; this includes: i) unit not found or inaccessible; ii) not-at-home; iii) unable to respond; iv) refusal (potentially 'convertible'); v) 'hard core' refusal

G) sampling errors; these include: i) sampling variance; ii) estimation bias

Recalling the ***classical classification*** into sampling and non-sampling errors, the latter category is comprised of errors of types A) to F) above.

UNIVERSITÀ DI SIENA 1240

Measurement bias and sampling error for disaggregation - 1

Measurement bias

Bias arises from shortcomings which affect the whole survey operation: basic conceptual errors in defining and implementing the survey content; incorrect instructions for interviewers; errors in the coding frame or programs for processing the data; etc.

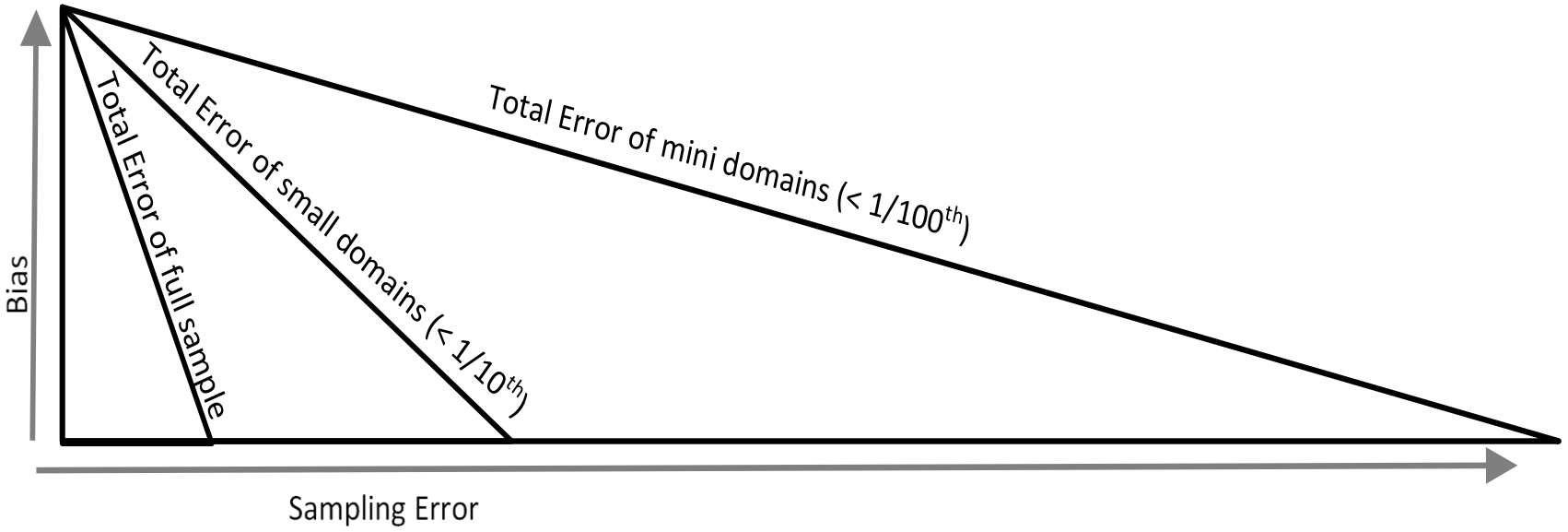
Sampling error

Sampling error is a measure of the variability that would be observed between estimates from different samples drawn using the same sample design as the survey, disregarding any variable errors and biases resulting from the process of measurement and sample implementation.

Measurement bias and sampling error for disaggregation - 2

The relative importance of sampling errors for disaggregation is portrayed by the right-angled triangles in

Figure 1. Increasing Sampling Error/Bias Ratios for disaggregations



Source: Kish (1988)

UNIVERSITÀ DI SIENA 1240

Measurement bias and sampling error for disaggregation - 3

Total error is often denoted as the root mean squared error and defined by taking the square root of variance and squared bias. It can thus be represented by the hypotenuses of a right-angled triangle.

The smallest triangle of this figure depicts large survey samples where total error is dominated by the bias. In such situations the precision of poverty measures cannot be much increased by increases of sample size. Instead, it is worthwhile to focus on reducing non-sampling errors.

UNIVERSITÀ DI SIENA 1240

Measurement bias and sampling error for disaggregation - 4

- For subpopulations below the national level, the magnitude of bias is, however, often very similar, whereas the sampling error drastically gains in relative importance.
- Following the classification of Purcell and Kish (1980), groups which comprise less than 1/10th of the population may be considered as **minor domains**.
- The standard error for estimates for such domains is more than 3 times higher than for the full sample.
- For these domains, sampling error will overtake bias in many surveys

UNIVERSITÀ DI SIENA 1240

Measurement bias and sampling error for disaggregation - 5

- Finally, for **mini domains**, which Purcell and Kish categorised as groups which comprise between 1/100th and 1/10000th of the population, sampling error will often be the dominant factor and controlling the bias may not substantially improve total error.

UNIVERSITÀ DI SIENA 1240

Recommendations - 1

1 - Document coverage, precision and any departure from international standards in quality reports to be issued with each data release.

2 - Strictly adhere to probability sampling and refrain from substitution.

3 - Adjust sample design and allocation for the required disaggregation. As a minimum the sample size for each group should never be below 50 units. Ideally estimation domains will be used as strata so that their size may be controlled.

UNIVERSITÀ DI SIENA 1240

Recommendations - 2

4 - Ensure a sufficient number of contact attempts (e.g.), use group translations, and interviewer's addresses. Field guidelines to be further elaborated.

5 - Compensate for non-response and random variation by adjustments to external control distributions (calibration). Ensure that these adjustments (g-weights) remain within reasonable limits. (e.g. 1/3 to 3).

6 - Ensure that vulnerable participate in the design of the study and are protected from any harm and can benefit from its results.

UNIVERSITÀ DI SIENA 1240

Recommendations - 3

7 - Minimize the amount of missing information. Because poverty status is usually derived from several variables that have to be observed simultaneously it is essential that group representation is preserved by adequate and transparent imputation rules.

8 - Carefully assess the variance of all key estimates. Publish coefficients of variation in quality reports so that precision can be compared to other surveys and over time. As a minimum present an estimates of the design effect of the poverty rate.

9 - Establish filtering rules for publications that eliminate unreliable estimates. never publish results based on less than 20 observations or with a correspondingly high coefficient of variation.

UNIVERSITÀ DI SIENA 1240

References

Hussmanns, R., Mehran, F. and Verma, V. (1990), *Surveys of the Economically Active Population, Employment, Unemployment and Underemployment*, International Labour Organisation, Geneva.

Kish, L. (1988). Multipurpose sample designs. *Survey Methodology*, 14, 19-32.

Purcell, N. J. and Kish, L. (1980). "Postcensal Estimates for Local Areas (Small Domains)", *International Statistical Review*, 48, pp. 3-18.

Verma, V. (1981), 'Assessment of errors in household surveys', *Bulletin of the International Statistical Institute*, 49(2), pp. 905-919.

Verma V., Betti G. and Gagliardi F. (2010), *An assessment of survey errors in EU-SILC*, Eurostat Methodologies and Working papers, Luxembourg: Publications Office of the European Union.

UNIVERSITÀ
DI SIENA
1240

THANK YOU FOR YOUR ATTENTION !!!