

Additional samples with overlapping and balancing conditions: theoretical aspects and application to students' assessment data

Marc Christine
INSEE, Paris, France marc.christine@insee.fr

Thierry Rocher*
DEPP, Paris, France thierry.rocher@education.gouv.fr

This paper provides a theoretical frame and methods to solve a problem which occurs as soon as a first sample has been drawn at a given time and that one intends later to draw a 2nd sample in an updated sampling frame, linked in a way with the 1st one, but without any possibility of changing the conditions or results of the drawing of the former sample. The origin of this issue lies in PISA (Programme for International Student Assessment) surveys: the cycle 2012 was on the same main topic as in 2003 and it was necessary to make comparisons between both surveys. In this context, some countries wish to build the 2012 sample of schools with overlapping conditions with the 2003 sample. But it is also necessary to have the best representativeness for the new sample. This one can be met introducing balancing conditions when the new sample is drawn. Other constraints should be prescribed (fixed size, given inclusion probabilities...). The main tools used are first conditional successive samples and, secondly, balancing techniques. But it will be shown that only approached solutions can be reached, not only from a statistical point of view, but also from a computational one, to obtain numerical solutions. After developing the theoretical approach, results on French sampling frames of schools will be given.

Key words: balancing conditions, overlapping, inclusion probabilities