Sample coordination methods and systems

Alina Matei [1] & Paul A. Smith [2]

[1] *University of Neuchâtel, Switzerland, alina.matei@unine.ch*
[2] *University of Southampton, UK, p.a.smith@soton.ac.uk*

Résumé. La coordination d'échantillons fait référence aux méthodes qui permettent de créer une dépendance probabiliste entre les sélections d'échantillons aléatoires afin d'optimiser la taille de leur chevauchement. Si l'objectif d'une enquête est d'estimer les changements dans le temps, ou de réduire les coûts associés au recrutement d'une nouvelle unité d'échantillonnage, la taille du chevauchement des échantillons doit être maximisé. Parfois, on souhaite contrôler le risque que la même unité d'échantillonnage soit sélectionnée dans différentes enquêtes, et donc limiter la charge de réponse pour cette unité particulière dans une période donnée. Dans ce cas, l'objectif est de minimiser la taille du chevauchement des échantillons. Plusieurs méthodes sont utilisées pour la coordination d'échantillons. Des systèmes de coordination d'échantillons sont également utilisés dans les offices statistiques nationaux, et ils sont basés sur des méthodes de coordination d'échantillons. La littérature existante ne fait pas de distinction entre 'méthodes' et 'systèmes'. Mais, à notre avis, les deux représentent des concepts différents. Pour distinguer les deux concepts, nous fournissons une définition d'un système de coordination d'échantillons, et classifions les mesures qui sont actuellement utilisées dans de tels systèmes dans la statistique officielle.

Mots-clés. statistique officielle, échantillonnage, coordination des échantillons, nombres aléatoires permanents, programmation mathématique.

Abstract. Sample coordination refers to methods that allow the creation of a probabilistic dependence between sample selections in order to optimize their overlap size. If a survey's objective is to estimate change across time, or to reduce the costs associated with recruiting a new sample unit, the overlap size between the samples should be maximized. Sometimes one would like to control the risk of having the same sampling unit selected across different surveys, and therefore to limit the response burden for that particular unit in a given period. In such cases the objective is to minimize the overlap size between samples. Several methods are in use for sample coordination. Sample coordination systems are also used in national statistical offices, and they are based on sample coordination methods. The existing literature doesn't distinguish between 'methods' and 'systems'. But in our opinion the two represent different concepts. To distiguish the two concepts, we provide a definition of a sample coordination system, and classify measures which are currently used in such systems in official statistics.

Keywords. official statistics, sampling, sample coordination, permanent random numbers, mathematical programming.

1 Sample coordination methods and systems

Sample coordination refers to methods that allow the creation of a probabilistic dependence between sample selections in order to optimize their overlap. Positive coordination refers to the case where the overlap is larger than under independent sampling, and is generally used to reduce the variance of measures of change between successive periods of repeating surveys, though it can also be used to link together information from two separate surveys. Negative coordination is when the overlap is smaller than under independent sampling, and is used particularly to reduce the number of surveys in which a particular unit is selected in a given period, and therefore to control the response burden of units. Various methods have been proposed for sample coordination (such as permanent random number procedures, methods based on mathematical programming, etc.). For an overview on sample coordination methods, one can see, for instance, Ernst (1999), Mach et al. (2006), and the references therein.

We consider that the existing sample coordination methods do not provide enough structure to deal with a range of different surveys with various designs and rotation patterns, and with the resulting range of positive and negative coordination characteristics. In order to manage this wider problem, we need a *sample coordination system*, and there has been a range of implementations, based on sample coordination methods.

The national statistical offices of different countries currently use 'sample coordination systems'. 'Sample coordination methods' and 'sample coordination systems' represent, in our opinion, two different concepts. The existing literature does not distinguish between them. Moreover, a definition of a sample coordination system has not yet been provided, while the term is widely used. It is challenging to define precisely what a sample coordination system consists of. Our contribution refers to this challenge, as well as to the introduction of a typology of systems currently used by the national statistical offices (p-coordination, s-coordination and b-coordination). The presentation is based on Matei and Smith (2023).

Bibliographie

Ernst, L. R. (1999), The maximization and minimization of sample overlap problems: a half century of results, In *Proceedings of the International Statistical Institute, 52nd Session*, pp. 168–182, Helsinki, Finland.

Mach, L. and Reiss, P. T. and Ṣchiopu-Kratina, I. (2006), Optimizing the Expected Overlap of Survey Samples via the Northwest Corner Rule, *Journal of the American Statistical Association*, 101 (476), pp. 1671–1679.

Matei, A. and Smith, P. A. (2023), Sample coordination methods and systems for establishment surveys, book chapter in *Advances in Business Statistics, Methods and Data Collection*, editors Snijkers, G., Bavdaž, M., Bender, S., Jones, J., MacFeely, S., Sakshaug, J., Thompson, K.J. and van Delden, A., Wiley, New York, USA.