Official Statistics – Procedures and Principles

Nick Longford

SNTL, Leicester, England

Abstract

The presentation discusses several technical and organisational issues in the provision of modern statistical service to the national and local government. The key themes are that the agenda of the client (the government) should be taken literally, without its reduction to `what is possible', but at the same time the imperfections of the service provided should me made clear and integrated in the decision making.

These themes are developed on the generic problems of incomplete data, small-area estimation and selective reporting. The well established approach to dealing with missing data is to impute values for them, thus arranging that the database has the originally planned format and analyses prepared for the database can be applied without requiring any adaptation. Such an analysis is 'dishonest', in that the sampling variation and related quantities characterising the quality of the inference are underestimated. The presentation will describe examples of multiple imputation and discuss how it addresses such dishonesty. Examples of survey analysis from beyond the confines of missing data, for which multiple imputation can be applied, will be outlined.

Small-area statistics are an example of successful application of biased estimation. Statistical textbooks emphasise unbiased estimation with minimum variance and this is in practice frequently misinterpreted as rejection of any biased estimators, even if their variances are very small. We should not hesitate to incur bias if it is accompanied by a substantial reduction of the sampling variance. This principle is now well established in small-area estimation, but it is applicable in a much wider arena. First, analysis of a regularly conducted national survey in isolation from its previous runs (years) is grossly inefficient if wealth of information contained in the previous years' data is not exploited. Next, registers and other administrative records can contribute to the analysis of a survey. More generally, we should think about analysis not as applied to surveys but to collections of surveys and other sources of information relevant to a specific agenda. For different inferential agenda a different collection of data sources may be better suited.

Every probability is conditional (De Finetti). We should understand this famous statement as applying also to distributions, and to estimators in particular. As we claim that an estimator is unbiased and has a small variance, the claims is invalid if the estimate is reported conditionally, subject to bringing `good' news, being selected by a procedure from a list of candidates, or when the target of estimation is selected based on the values of the estimates. In brief, estimates and their standard errors are `fragile' quantities, with their properties highly conditional on context. This issue is understood well in the clinical trials, but it is applicable much more universally. In government statistics, this requires a closer integration of statistical analysis with making decisions based on processed information.

We should have on the horizon a set of principles that promote integrity of the statistical profession as a service to all spheres of the society, and disseminate the understanding of the imperfection of the current state of the art, without diminishing our efforts to approach the horizon, and being aware of the distance from it throughout.