Teaching Sampling in a Government Statistical Agency: The Canadian Experience

Jack Gambino
Statistics Canada, Social Survey Methods Division
15-O R.H. Coats Building
Tunney's Pasture
Ottawa, Canada K1A 0T6
gambino@statcan.ca

Hew Gough
Statistics Canada, Social Survey Methods Division
15-G R.H. Coats Building
Tunney's Pasture
Ottawa, Canada K1A 0T6
hew.gough@statcan.ca

1. Introduction

Statistics Canada has a staff of approximately 300 survey statisticians (known internally as methodologists) with a diversity of educational backgrounds. To help them do their work, these statisticians have a wide variety of courses available to them. In this paper we describe the types of courses that are offered and how they are geared to the target audience. We begin in section two by describing the target audience. In section three, we give an overview of the courses that are offered and mention some related documents. In section four we discuss some of the challenges we face in developing and presenting suitable course material.

It should be emphasised that courses are just one aspect of the learning that survey statisticians pursue throughout their careers. On-the-job training and learning from colleagues are two other important aspects of their career development. There are several ways of helping survey statisticians at Statistics Canada plan the learning component of their careers. Each employee benefits from the guidance of a supervisor and possibly a mentor. There is also a list of suggested courses to be taken at various stages in one's career. Employees, and supervisors, are also encouraged to consult the document *Methodologists' Training and Development Handbook*, which covers both statistical and other courses (e.g., computer-related, project management, effective communication) as well as other aspects of career development. Finally, Statistics Canada recently started to offer each employee a Personal Learning Plan, which is used by the employee and his supervisor to establish a commitment to learning objectives over the next one to two years.

2. Who we teach

Survey statisticians from Statistics Canada give courses to a variety of audiences. About half the courses are developed for, and given to, fellow survey statisticians in the agency. Some courses are developed specifically for non-statisticians at Statistics Canada. Typically, these latter courses are less technical, with little or no explicit mathematical content. Some courses are also offered to audiences outside Statistics Canada. These are usually given on a cost-recovery basis. Finally, courses and workshops are given internationally, typically either in association with a conference or as part of an agreement with a foreign government statistical agency. In this paper, we will focus on courses given within Statistics Canada to our own survey statisticians.

Perhaps the greatest challenge we face in developing suitable courses for our staff is their wide diversity of educational backgrounds. They range from people with a bachelor's degree who

have never taken a course in sampling to people who have written a doctoral dissertation on a topic in survey sampling. As a result, we offer courses that range from the very basic to those that cover the latest research in the field. We describe these in the next section and return to the challenges in section four.

3. What we teach

A key course taken by all new professional employees of Statistics Canada, including statisticians, computer scientists and economists, is the Survey Skills Development Course (SSDC). Over a period of six weeks, the students in this course are exposed to all the steps in a survey, from conception to the dissemination of survey results. The students actually put into practice what they learn by developing a real survey. The last two weeks of the course are devoted to conducting interviews, processing the data and writing a report, including the results (estimates, analysis).

Because of its intended audience, the survey sampling component of SSDC conveys the important ideas of the subject with minimal use of mathematical notation. Important mathematical formulas, such as those for variances and confidence intervals, are presented but not derived. The sampling component of the course covers the following topics.

- populations, sampling frames and sampling units (making a distinction between the target population and the survey population)
- probability sampling and non-probability sampling
- simple random sampling, systematic sampling, PPS sampling (three PPS methods are illustrated using small examples)
- stratification
- multi-stage sampling
- multi-phase sampling
- estimation, including adjustment for nonresponse
- adjustment of weights for known totals (sometimes referred to as calibration)
- variances and variance estimation

At the end of the SSDC, each student will have a good idea of the steps a survey statistician follows when selecting a sample and processing the results, including the many choices that must be made along the way. The technical reasons underlying these choices are left to more advanced courses, which have a narrower intended audience, namely, survey statisticians.

The SSDC, which is in its fifteenth year, has been extremely successful. Since it is taken by all new professionals at Statistics Canada, it provides a common foundation for these employees, including an understanding of what colleagues in other parts of the agency do. Because of the SSDC's success, it has also been adapted by, or has greatly influenced, training programs in several other countries.

While the SSDC is designed for a broad audience, most other statistics courses offered to survey statisticians are designed specifically for them. A key course is entitled *Statistical Sampling Theory*. It covers much of the same material given in point form above, but is geared toward people with a statistics background. It is more or less equivalent to the first two courses in sampling offered at some universities. Of course, employees who have taken a comparable course in a university do not need to take the Statistics Canada course.

The SSDC and *Statistical Sampling Theory* courses bring new employees with little or no background in sampling to a level where they can follow more specialized courses. Before describing these, we note that occasionally, a prominent survey statistician from a university is hired to give an advanced course that covers the "state of the art" in sampling. The most recent one,

offered in the fall of 2004 and entitled *Recent Developments in Sample Survey Theory,* was given by Prof. J.N.K. Rao of Carleton University in Ottawa. Such courses give our staff a broad perspective on the latest developments in the field.

Other courses developed for and by our staff cover topics that are directly related to our work. Although there are no restrictions, most survey statisticians take only a subset of these courses, depending on the nature of their work. The courses can be classified into categories such as data analysis and time series. Under data analysis, there is a series of courses on the analysis of data from complex surveys, as well as basic courses in exploratory data analysis and a more specialized course on survival analysis. Under time series, there is a suite of courses covering the basics as well as seasonal adjustment, ARIMA modelling and forecasting. There are courses in quality assurance and control, with versions for statisticians and non-statisticians. Finally, there are several courses on more specialized topics: longitudinal surveys, nonresponse and imputation, questionnaire design, and record linkage.

In addition to courses, staff have a variety of seminars and workshops available to them. Each of Statistics Canada's three methodology divisions has its own series of seminars, which are open to all staff. There are also seminars with a strong statistical component, particularly analysis-related ones, offered in other parts of the agency that survey statisticians are free to attend. They are also encouraged to attend workshops, both in Ottawa and elsewhere. Each year, Statistics Canada holds a methodology symposium, which usually has associated workshops. At the most recent symposium last November, there were three workshops offered: Data Mining and Text Mining, The Use of Probabilistic Networks for Imputation, and Indirect Sampling: A Practical Solution for Populations Difficult to Reach. The latter was given by one of our own staff and the first two were given by experts from France and Italy, respectively.

A recent innovation is the development of a series of introductory seminars aimed at recently hired survey statisticians. Each seminar introduces a topic such as collection methods, estimation and confidentiality. These seminars give new staff a quick taste of a variety of topics which they can then delve into by taking courses or reading some of the reference material highlighted by each presenter.

Course material has been used as a source for creating major documents. Two recent highlights are the publication of the book *Survey methods and practices* (Statistics Canada 2003) and the development of an *Analysis Handbook*. The former is a practical guide to survey planning, design and implementation. The latter provides practical information to analysts of Statistics Canada's data.

The book *Survey methods and practices* can be viewed as the textbook equivalent of the Survey Skills Development Course discussed in section two, and is, in fact, now used in the course. This book originated as a document prepared by Statistics Canada for the National Bureau of Statistics of China as part of an international technical assistance program funded by the Canadian International Development Agency. The published book is a revision of this document. In addition to covering survey design (i.e., the topics itemized above), it also covers survey operations, data processing and survey management. It also includes a realistic case study that is used throughout the book to illustrate and reinforce ideas.

The *Analysis Handbook* was inspired by the many consultations done by Statistics Canada's Data Analysis Resource Centre (DARC). Analysts from both within and outside the agency turn to DARC when they have questions about the proper analysis of data from complex surveys. The *Handbook* is a distillation of these experiences and provides practical information to users of some of Statistics Canada's microdata. It includes information on the use of various methods as well as tips on the proper use of popular software packages. In addition, for specific surveys, it contains a description of the design of each survey, its microdata files and the proper use of its variables.

4. The challenges

We have already noted that one of the challenges we face in teaching sampling is the diversity of academic backgrounds of our staff. We deal with this in part by offering the SSDC course to give everyone the same foundation. For individuals who took no sampling courses in their university days, the *Statistical Sampling Theory* course is sufficient to get them started doing useful work. Their further development is achieved by a combination of courses and on-the-job experience, including participation in research with fellow statisticians.

Another major challenge is finding individuals who have the knowledge, the ability and, most importantly, the time to prepare course materials and teach courses. In most cases, teaching is just one of many duties performed by an individual, so finding a sufficient number of people who have the time to get involved in teaching can be difficult. This difficulty is exacerbated by the fact that all recurring courses are offered in English and French, which often implies that two instructors (or two sets of instructors in some cases) have to be found for each course.

Even when a course is presented in a format similar to a university course, as is the case for the *Statistical Sampling Theory* course mentioned above, students are not tested, and any "homework" assignments they may get are not evaluated. As a result, there is no formal way of assessing whether the course material has been understood. There is also no direct information on whether the material is retained and used in practice. There is only anecdotal and observational evidence that course participation benefits the work done by survey statisticians. And, of course, one of the benefits of course attendance for the individual is that he can turn to the course material and references at a future date when he is faced with a problem addressed in the course.

5. Conclusion

The survey sampling curriculum for survey statisticians at Statistics Canada was developed over many years and is now mature. It addresses most of our training needs as far as our current work is concerned. However, there is always room for improvement and, furthermore, as the nature of our work evolves, so does the need to modify existing courses and add new ones. This is done on an ongoing basis in an effort to face our final challenge: ensuring that our suite of courses remains relevant in a rapidly evolving environment.

REFERENCES

Statistics Canada (2003). *Survey methods and practices.* Statistics Canada catalogue number 12-587-XPE.

RÉSUMÉ

Statistique Canada compte environ 300 statisticiens d'enquête. En plus d'avoir accès à de la formation en cours d'emploi, ces statisticiens ont également accès à un large éventail de cours internes en statistiques. Dans le présent document, nous décrirons les types de cours offerts et la façon dont ils sont adaptés au public cible, qui comprend des personnes possédant des acquis scolaires très variés. Nous discuterons aussi de certains des défis reliés à l'élaboration de matériel de cours approprié et à la présentation opportune et efficace de ce matériel. Bien que la présentation soit axée sur les cours internes offerts à nos employés, nous discuterons aussi brièvement du fait d'offrir nos cours à des personnes de l'extérieur et de la disponibilité de cours externes pour nos employés.