COMMUNICATING SAMPLING CONCEPTS TO SOCIAL SCIENTISTS: THE CASS EXPERIENCE

Ray Chambers and C. J. Skinner, Department of Social Statistics,
The University of Southampton.

The Centre for Applied Social Surveys provides courses on survey methodology and technical information to academic social scientists and applied social researchers. In this paper we describe the audience targeted by these survey courses and the approach taken in them, which tends to focus on practical issues and the shared experiences of the course participants. The teaching of theory in such an environment is sometimes a problematic issue. We discuss this problem, as well as the role of computer-based packages for survey design and analysis in courses whose basic aim is the general upgrading of the methodological skills of participants.

INTRODUCTION

The Centre for Applied Social Surveys (CASS) is a Resource Centre of the Economic and Social Research Council (ESRC) run jointly by Social and Community Planning Research (SCPR), an independent survey research institute in London, and the University of Southampton, with the University of Surrey. CASS provides training and information support about social surveys to the UK social science research community and other survey users. CASS's two principal activities are the provision of a programme of short courses and the development of an online survey question bank. This paper will focus CASS's short course programme. Information about the question bank is available at *http://kennedy.soc.surrey.ac.uk/qb*.

Quantitative measurement and analysis based upon sample survey measurement is of growing importance for social research and social policy. Academic researchers in the United Kingdom conduct many kinds of sample surveys, ranging from surveys of no more than 100-200 respondents, run single-handedly by individual PhD students, through to large-scale surveys such as the British Household Panel Survey which involves some 10,000 people, with fieldwork commissioned from a commercial survey agency. An even larger number of academic researchers are engaged in the secondary analysis of survey data collected by others.

Such survey research creates needs for training in survey methods. In large government and commercial survey organisations where there is a concentration of methodological expertise, such training may be organised in-house. In universities, however, expertise in survey methods tends to be spread too thinly to make it viable for

each university to provide adequate training for all its students and staff conducting research involving surveys. In the United Kingdom, even in universities with strong traditions of survey-related research, there may not be the staff with sufficient technical expertise to mount a course in, for example, survey sampling.

In recognition of the importance of sound survey methodology as a basis for quantitative social science research, the ESRC (the primary government agency in the UK funding social science research) awarded a grant in 1995 for the founding of CASS as an ESRC Resource Centre, with the aim of strengthening skills in survey methods in the UK social science research community.

The joint nature of CASS, involving cooperation between academic and non-academic survey methodologists, is no accident. Survey methodology covers a wide range of techniques, from those involved in the planning and implementation of surveys through to those required for survey data analysis. In the UK it seems unlikely that any single organisation would claim to have a concentration of national expertise at all points of this spectrum. No university is the base of major ongoing survey fieldwork operations and so concentrations of expertise in the practical planning and implementation of surveys tend to be outside academia. On the other hand, most of the more sophisticated survey data analysis tends to take place in universities. As a reflection of this diversity, the CASS short course programme is operated jointly by SCPR, an independent survey research institute and the University of Southampton. SCPR runs a wide range of social surveys, with clients from government as well as universities, and has a long history of conducting methodological research. Its particular expertise is in the methods of survey design and implementation. On the other hand, the University of Southampton, through its Department of Social Statistics, brings expertise in survey data analysis to CASS.

THE CASS SHORT COURSE PROGRAMME

CASS offers a programme of short courses on social survey methods, covering all aspects of the quantitative survey process from design to data collection and data analysis. The courses have an applied emphasis, aimed at providing guidance for researchers who face practical decision-making in surveys. CASS is funded by the ESRC and its primary audience thus consists of the constituency which ESRC serves, postgraduate students, researchers and academic staff in the social sciences. This constituency is mainly based at universities but also includes researchers at other organisations working on ESRC-funded

research contracts. CASS courses are advertised publicly and participants from outside the ESRC constituency with experience of survey research are encouraged. The latter participants are charged full fees in contrast to those from the ESRC constituency whose fees are generally reduced via ESRC funding support. A breakdown of the participants attending basic CASS courses in the 1995/96 and 1996/97 academic years is shown below. Although academic participants constitute the great majority, there remains a significant minority from government, where similar training needs arise. Academic participants are primarily associated with social sciences disciplines but there is a significant minority from those health-related disciplines which also make use of the social survey method.

Numbers of Participants on Basic CASS Courses by Area of Work

	Academic				Indept. Res.	Govt.	Private Sector
	Stats	Social Science	Health Science	Other			
1995/96	4	19	9	7	2	17	3
1996/97	4	60	38	18	1	25	5

The needs of the audience for CASS courses so far might broadly be classified into four broad categories:

1. *Junior researchers requiring general training*: needs for skills and knowledge of survey methods to equip participants for use in the future;
2. *Continuing professional education*: needs for updating knowledge and for refreshing understanding of principles;
3. *Researchers with specific needs*: needs for methods to address specific problems faced by participants in their research;
4. *Trainers of survey methods*: needs for updating knowledge to equip participants with knowledge and resources they may incorporate into their own teaching, for example on undergraduate courses in survey methods.

The majority of participants are interested in applying what they learn to some specific survey. Even those with needs classified under (1) above, often come with a certain survey in mind. This audience may thus be contrasted with the usual audience on instructional courses at universities, where the teacher cannot usually assume that each participant will come with detailed knowledge and interest in any specific survey.

Many participants appear to be somewhat isolated in their usual work settings. They are often the only survey statistician or methodologist in a small team charged with

running a particular survey. In large organisations, such as the Office for National Statistics, such staff will usually have other staff they can turn to for advice. In many small organisations, such as university research centres, this may not be the case, however, and CASS courses provide a valuable opportunity for such people to share their experiences, problems and ideas not only with the course teachers but also with fellow participants working in similar settings.

Given the considerable survey experience which many participants bring with them and the practical needs they seek help with, it is desirable to keep numbers on courses low enough to enable participants to share their experiences with other participants and the course presenter(s). Most CASS courses do not have many more than 20 participants.

Beyond the specification of the prior knowledge required for a course in CASS publicity and the setting of an upper limit on course numbers, no selection of participants takes place. Such open enrolment can lead to heterogeneity between students. Varied disciplinary backgrounds and survey experiences can have considerable benefits. But heterogeneity in statistical backgrounds can create difficulties for the more technical courses.

The course programme each academic year consists of a set of individual short courses of varying length. The 1997-98 course programme contains 12 courses, ranging from introductory courses on design and management of surveys to advanced courses on new imputation methods for survey data analysis. The courses currently run for between 1 and 3 days, although courses as long as 5 days have been experimented with. Courses are either basic or advanced. The latter are funded via ESRC's Analysis of Large and Complex Datasets programme, but are presented under CASS. The basic courses are designed so there is progression from introductory courses to more methodologically demanding courses. Many basic CASS courses are approved under a credit accumulation and transfer scheme (CATS), so participants can, provided they satisfactorily complete an assessment module, accumulate CATS points which can be used as "academic credit" in UK postgraduate degree programmes.

APPROACH TO TEACHING SAMPLING CONCEPTS

CASS courses generally involve a mix of lectures and class sessions. In the latter sessions, participants work through exercises, usually in small groups but sometimes individually. This enables active participation and sharing of experiences.

The Role of Practice vs. The Role of Theory

Survey practice is the common denominator linking the experiences of different students. Illustrating methods through examples becomes the principal way to convey both the elements of different methods and their importance. Given the strong practical orientation of most participants, there may be a temptation to deemphasize theory and to present methods in a cook-book style. Some mathematical details can often be omitted without great loss if appropriate references are given. The principles upon which methods are based need emphasis, however, if students are to be able to see the wood for the trees. For example, it is easy to illustrate the problems caused by nonresponse through practical examples. On the other hand, in order for students to understand the implications of these problems it is important to classify types of nonresponse, for example distinguishing between item and unit nonresponse and between noncontact and refusal, and to represent bias and variance effects within a clear theoretical framework which helps clarify the principles guiding the handling of such problems.

Balancing Theory and Practice in an Introductory Course on Survey Sampling

The CASS course “Survey Sampling” is an introduction to some of the basic concepts used in survey design and estimation. The course is over 3 days, with the first two focusing on design, and the last day spent looking at some issues in estimation. The course emphasises 3 key design concepts: randomisation, stratification and clustering. Motivation for the statistical theory comes from consideration of the different sources of uncertainty at the design and analysis stages of a survey. At the design stage the randomisation distribution reflects uncertainty about the particular sample that will be chosen, while population models allow uncertainty about the relationship of sample to population values to be investigated at the analysis stage.

The course attempts to cover methodologies important in both household-based and business establishment surveys. The emphasis throughout is on showing how basic principles relate to important design strategies. Thus, both stratification and clustering are motivated via links to the use of “grouping” models for heterogeneous populations, as

well as via more standard arguments based on cost efficiencies. Since participants tend to have varied statistical and mathematical backgrounds, the level of methodological difficulty is deliberately kept low. Algebraic formulae are kept to a minimum, and real life examples are used to motivate them wherever possible. The format of the course is typically a 60 minute talk (with active audience involvement via questions from the presenter) followed by a 30 minute group discussion and problem solving session based on the material just covered. At the end of days 1 and 2 of the course there is a 60 minute review session which summarises the basic ideas that have been covered and encourages participants to ask clarifying questions. Audience participation is typically good, since virtually all participants have particular survey sampling issues for which they want feedback, both from the course presenters, as well as from other participants.

The course presentation is very traditional - one presenter at the front of the group and the other presenter seated with the group whose role is to be a catalyst for discussion. The small size of the group makes for an informal atmosphere, and so far we have not seen any advantage in trying out more “high tech” methods of presentation. In particular, we have not attempted to integrate some of the emerging software for “teaching” survey sampling. The principal reasons for this are the short “intensive” nature of the course, its wide focus and its emphasis on the link between theory and practice.

The Role of Computers and Survey Software

Computers are primarily used on CASS survey data analysis courses. They allow participants to try out realistic survey analyses, working at their own pace. Participants may bring their own data and thus clarify the relevance of the course to their own specific needs. Some software packages, such as SUDAAN or STATA, offer particular survey-related options and the course may provide the opportunity to learn about these options. Usually, however, the focus of the course will be on survey methodology and it will be desirable for participants not to waste too much time learning new software.

Using Commercial Software in a Course on Regression Analysis

The advanced course “Regression Modelling for Complex Survey Data” was given for the first time in 1997. Its aim was to introduce social statisticians to the

principles and practice of regression modelling (including logistic regression modelling) in the presence of stratification, auxiliary information and clustering. Since dissemination of practical knowledge is a key CASS objective, the theory part of this course was integrated with an application based on a real data set (a subset of the data obtained in the Indian National Family Health Survey) and commercially available software was used to demonstrate the impact of complex sample and population structures on inference. STATA software was chosen because it can deal with sample weights, stratification and clustering both in standard linear regression modelling as well as in logistic regression modelling. Furthermore, STATA was seen as providing an *integrated* statistical modelling environment, where EDA, fitting and confirmatory analysis are all possible within the same package.

As with the "Survey Sampling" course, presentation was a mix of lectures and practical work sessions. In this case, however, the work sessions (3 in all, each of approximately an hour's duration) were in a computer laboratory, with participants executing and interpreting the output from pre-prepared STATA macros.

This course has only been given once, so it is still in a process of development. However, reactions from participants indicate that its "mix" of theory and application still needs work. In particular, participants thought that the concepts demonstrated in the STATA-based analyses of the practice survey application should have been more tightly "linked" to the theory presented in the lectures, making the course more along the lines of "how to use STATA to analyse complex survey data". At present we do not see this orientation as necessarily what the course is about, since there are important issues in regression modelling using complex survey data that cannot be covered using STATA methodology alone. However, in further presentations of this course we will be looking to "tie in" the software and the theory to a greater extent.

THE FUTURE

In terms of growing demand, as well as participant evaluation, the CASS short course programme can be considered successful, and so will continue. In addition, however, there are opportunities for integrating CASS courses with university postgraduate programmes through accreditation. There is also significant demand for extension of the CASS programme to continuing professional education for government statisticians and social researchers, including part-time master's level training in survey

methods for this audience. The courses provided by the Training of European Statisticians (TES) Institute has demonstrated the feasibility and value of pan-European training in this regard, and there are opportunities for CASS to link up with this activity.