Statistics in Action - Learning and Comprehending by Doing

Karin Winqvist - Orebro, Sweden

1. The Statistics in Action Course (STAC)

Training is an important component in the Swedish programme for developing countries. Statistics Sweden has bilateral agreements of assistance with statistical offices in several developing countries.

Statistics in Action (STAC) is a course which is designed to give practical knowledge in statistics. The course has been given at the Central Statistical Office in Zimbabwe during the period March 7 - April 28, 1988, and similar courses have been given at Statistics Sweden since 1975.

The purpose of the course is to present and carry out all phases of a survey in a proper order. In particular, it demonstrates how the different phases of a survey are connected. The staff of a statistical office easily becomes specialised in a particular phase of a statistical survey, for example, questionnaire design, data collection, or data processing. Then, quite frequently, they have little knowledge of other phases of the survey and of how their own work is affected by the work of others. In order to present the full picture, a small-scale survey is performed during the course and all the statistical theory and practical problems are discussed against the background of the actual survey.

The STAC courses are organised for groups of 16-20 persons. They may be any staff above clerical level at a statistical office. The number is limited since working in small groups is the main method of learning. It is a great advantage if the participants have practical experiences in various parts of producing statistics, for example, there may be both subject-matter statisticians, field supervisors, and programmers. Also, it is desirable that the knowledge of statistics does not vary too much in the group. The course may be more or less theoretical and it is held at a theoretical level appropriate to the participants.

The topic of the survey is unique for each STAC course. It may concern any topic that someone actually needs information about. The population of the survey

must be geographically concentrated since the time for data collection is limited. It is essential that the participants know something about the subject and that they are motivated by the fact that they collect valuable information. For example, the topic may concern "Personal experiences of computers", "Lunch habits", "Means of getting to work" of some group of employees. The STAC course is a very good opportunity for some management person, for example the staff manager or the manager of the staff dining-room, to get statistical information about people of his concern. In the beginning of the course the assignment is introduced by the person who is responsible for the subject. He has to be aware of the purpose of the course and of the tight schedule. He also has to be available for clarification during the first phases of the survey.

The course is quite extensive since there is a lot of practical work to be done. The schedule of the course runs over a period of about four months but the actual working time is about a hundred hours. Some of the days are full days (7 hours) and some are part-time days (4 hours). The listing of the frame and the data collection are not included in the hundred hours.

2. Contents of the course

The contents of the course reflect all the phases in planning, making and reporting a statistical survey (see Figure 1). It is essential that all the details of the survey are as realistic as possible.

The planning of the survey starts with the analysis of the survey topic. The participants get a brief written description of the assignment. The text should be fairly vague, since this is often the case when starting up a real survey. The first part of the course deals with the translation of the assignment into a statistical problem. A useful way is to break down the assignment into logically smaller subject areas which are easier to demarcate. The population is defined and very strictly demarcated. There is a discussion to choose the appropriate domains of study. A list of measurable variables is formulated and at this early stage it is important that the participants do not think in terms of questions. The overall picture of the topic is easily lost if one starts to formulate questions before having had a thorough discussion of the topic. In this first stage an overview of future tables is constructed, mainly to set a limit on the tables and to discover if any of the variables are redundant. Here it is essential to have several contacts with the person in charge of the assignment.

The participants choose the method of measurement: some type of questionnaire which is completed by the respondent or some type of interview. Of course the choice depends on the topic and on the time available for data collection. If interviews are chosen, a short module on interviewer training is added to the course.

The measurement instrument, that is, the questionnaire, is worked out by the course participants. The previous list of variables is now transformed into questions. First the wording of all the questions is thoroughly discussed and then a questionnaire is constructed. In doing this one considers both the requirements put up by the respondents and the data processing aspects of the questionnaire. A minor pilot survey is performed to test and review the questionnaire.

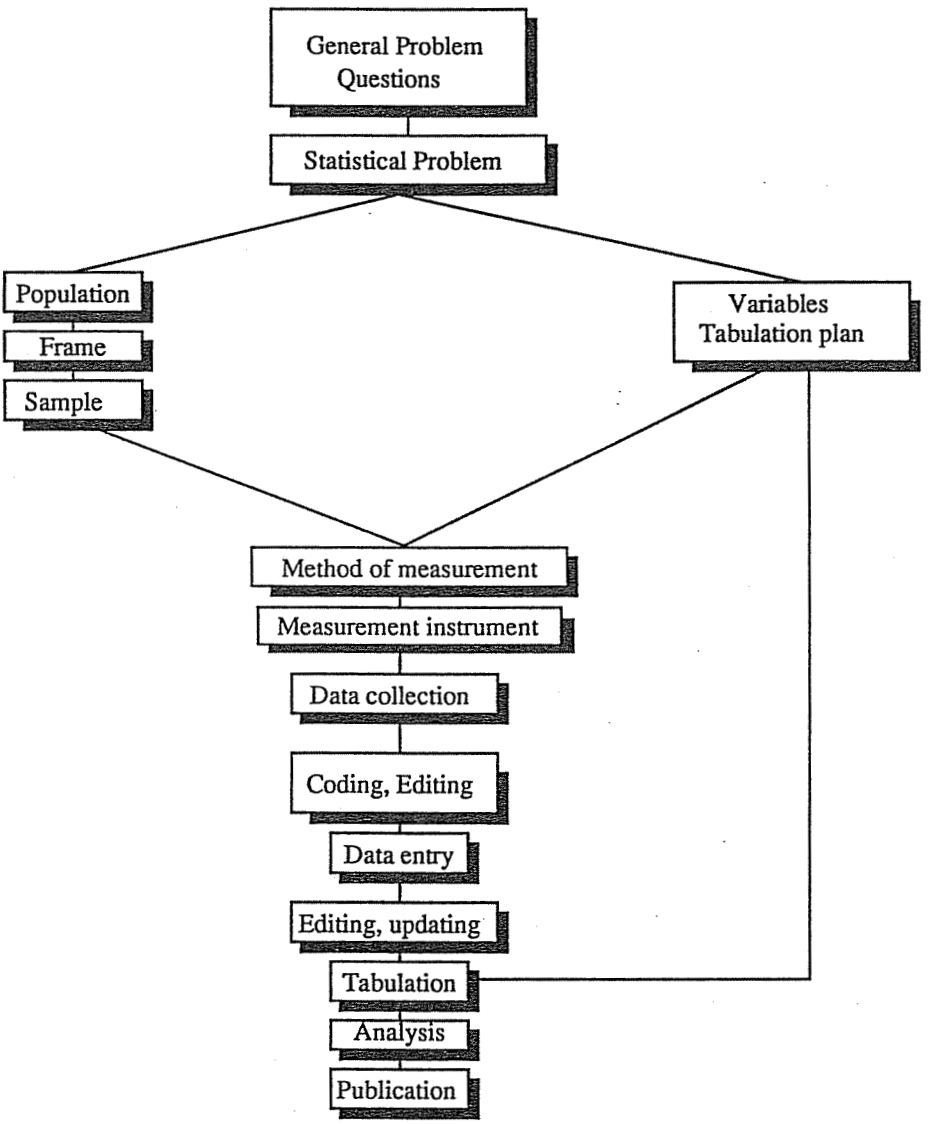

FIGURE 1
Phases in a statistical survey

Listing of the frame is often a tedious part of a survey in developing countries, since there are few registers and files which are up to date. This is the first part of the real field work of the survey, but, since the population should be fairly small, it can be done in a couple of hours.

The sample size mainly depends on the time available for data collection and the number of enumerators (the participants in the course) and the method of measurement. Every enumerator can make about twenty personal interviews or collect data from about thirty respondents.

The sample design may be a simple random sample within each stratum, probably equal to the domain of study, but it may also be more sophisticated if necessary. The sample is selected manually from the frame using a table of random numbers. Even if it is possible to use an EDP-procedure for the sample selection, it is worthwhile to do at least part of the selection manually just to illustrate the method.

The data collection is done during quite a short measurement period and of course it is important that there are no holidays during this period. Every enumerator is responsible for collecting data from twenty to thirty respondents. He or she must be aware that the result of the data collection is really very crucial to the outcome of the survey. If there is a non-response it is important to collect information about the reasons for this non-response. A couple of meetings are held during the period of data collection to check that everyone is successful and, if necessary, to clarify the difference between over-coverage and non-response. At the end of the measurement period every enumerator presents the results of the data collection in terms of numbers of respondents, non-respondents, and over-coverage, for every stratum.

Manual coding may be necessary for some of the answers and there is probably a need for some manual editing before the data entry.

The data is entered into a computer, personal computer (PC), or mainframe, using any available standard program. If there are usable routines for data entry and other EDP-procedures at the particular statistical office, it is important to use these. Then the participants get to know these procedures and their requirements. If there are programmers or other persons with knowledge of EDP among the participants, they can be of great help, otherwise the data entry program has to be prepared by someone else at the statistical office. Hopefully the data can also be checked and edited on line, otherwise one has to work with listings of errors and updating in several rounds.

When the data is edited and updated, tables are run according to detailed specifications from the participants of the course. The tabulation is made with a standard program for statistics. In most statistical programs one can work with weighted data, which is essential if the sample is stratified and the objects are selected with unequal probabilities. The tables that are produced are working tables to be used for the analyses.

The analysis and the report writing are important parts of every statistical survey, since this is an obvious opportunity to become familiar with the users of the results. About one third of the total time is devoted to these phases. In STAC every table is discussed in group sessions, and suggestions for minor tables, diagrams, or simply text related to the table, are forwarded to classroom discussions. Here it is particularly useful to make at least two groups work with the same task, since presenting data can be done in several good ways. All these simplified tables, diagrams, and texts are typed out before they are put together into a report. The report also contains the background of the

survey and a description of how the survey was carried out. There is also a discussion of the quality of the results. The report writing involves discussions about layout and other graphical aspects to make the report attractive.

The publication of the results are made in a report which is handed over to the person who gave the assignment. He has to receive the report some time before the handover so that he can study the results. At this final meeting the report is presented verbally by the participants and there is time for a discussion. At this point the person in charge should comment on the results and perhaps ask for clarification on the text.

3. Action learning

The method of learning is by doing. Classroom lectures and groupwork sessions are mixed. The theoretical background to the tasks is presented in short classroom lectures. Then most of the practical work is done in groups of three to four persons. It is often useful to give the same task to at least two groups. Then when the groups present the results of their discussions on large sheets of paper in a classroom session there is usually a lively discussion. The point is to hand over the responsibility of carrying out the survey to the participants of the course. Usually there are several possibilities to solve the statistical and practical problems. Presenting a couple of quick ideas is a good way of getting an open-minded discussion. There are a lot of decisions to be made and both the teachers and the participants must be aware that some will be good and others will be less so. Having a time limit is a very common predicament when making a statistical survey and then it is better to make a fairly good decision than to make no decision at all.

Since this course is based on action learning, only a few regular lectures can be prepared ahead. On the other hand, there is a large amount of planning that has to be done to secure the actual survey. Therefore there had to be two teachers during the whole course. These teachers, who should be highly experienced statisticians, must work well together and be good improvisers. They mainly serve as coordinators and consultants. There might also be expert teachers, for example on measurement methods or on EDP, but it is very important to keep continuity in the course, so it is necessary to have at least one main teacher. If there are experts, they must be very well aware that they may have to make compromises with the participants of the course. The method of teaching and the time limit of the course make it impossible to go into details in every aspect of the survey.

4. Prerequisites

Planning STAC involves several practical considerations. The time schedule is very tight and there is no room for time lags. There has to be both some technical equipment and some personal assistance during the course. Here the prerequisites are mentioned in the order that they are needed during the course.

The classroom has to be fairly large, since people move around quite often, and there should be an ordinary blackboard. The frequent groupwork requires three to four preferably small rooms with desks that can hold large sheets of paper. Since the results

of the group discussions are presented to the full group, there is a big need for large sheets of white paper and board markers.

The work with the survey is documented during the course, and a copying machine is necessary in order to keep everyone well informed. One also needs to make copies for the person in charge of the assignment. The copying machine is also used as a printing machine when duplicating the questionnaire and the final report.

The questionnaire has to be typed out over a couple of days, and this may be done by the teachers on a personal computer. It may also be typed by a typist using a word processing program. The first version of the questionnaire will surely be changed and word processing saves a lot of time when making alterations.

EDP is required for the tabulation in STAC, otherwise it would be possible to make very few tables. Any type of computer can be used if a statistical program that can be used for tabulation is installed. There is no time or money for programming the tables in an ordinary programming language. A personal computer with a graphical program may also be used when drawing the diagrams for the report, but the participants or the teachers can also do this manually.

The final report is typed out in several parts. First the simplified tables and the diagrams with brief analysis are written. Then the rest of the text, the description of the survey, etc. are added. In reading and rereading the report, the participants of the course often make several alterations. This typing task is preferably done by a professional typist.

5. Conclusions

STAC is unique in the sense that a real statistical survey is performed mainly by the participants of the course. The course illustrates how the different phases of the survey are connected and that the very tight time schedule of the survey makes detailed planning a necessity. The most important thing in planning the course is to engage a fairly important person who has an interesting topic for the survey. Teaching STAC requires a good theoretical background and a large amount of practical experience in statistics. There should be two very well coordinated teachers who realise that their roles in guiding and keeping an overview of the survey are just as important as teaching the statistical theory needed to solve the problems. Preparing the participants of the course is likewise important. Since action learning differs very much from ordinary classroom lectures, a letter of introduction gives information about the course and about the method of working, and hopefully creates the appropriate expectations. The participants have a large influence on the contents of the course and they must take most of the responsibility for planning, carrying out, and reporting the survey. The groupwork sessions create an atmosphere of cooperation that is very valuable for this course as well as for future work.