

## Almost realistic surveys - simulated practical fieldwork

### Abstract

One of the aims of academic studies in statistics is that students will be able to plan, realize and analyse a survey. It would be advantageous for the students if they had the possibility to conduct some surveys as training during their study time.

The aim with this project is to develop a flexible teaching system which gives the possibility to conduct what we call "almost realistic surveys" during a relatively short time, employing a system to simulate the respondent burden. Some fictitious populations will be constructed, consisting of both community populations and of enterprises. Samples from these populations will act as respondents that will give the students realistic data and, at the same time, a reasonable relationship between the answers to different questions. The fictitious background will include one or two communities and at least one population of enterprises. The population in the community will be of a reasonable size, for example between 2000-10000 inhabitants.

The course in survey methodology in the intermediate course in statistics at the Department of Statistics, Stockholm University may be organised as follows.

- Four lectures about planning of surveys
- Five lectures about sampling methods, estimation formulas for all methods studied, variance formulas for important cases
- Four lectures + one exercise covering analysis, sources of errors and the writing of reports
- Three exercises giving guidance for the simulated survey

A training survey from a student's point of view is:

- Receives a task
- Planning of the survey, prepares a questionnaire
- Hands over the plan for approval
- Draws a sample
- Conducts a testing survey
- Revising the questionnaire, if necessary
- Receives data
- Non-response follow-up
- Processing of data

- Analysis
- Writing a report

We are also planning to examine what modifications will be necessary for the teaching system to be available for distance teaching via Internet.

### Realisation of the project

First, determination of the content and the design of the registers.

Second, programming. The main part of this work is planned to the academic year 2000/2001.

Third, testing the whole system in the courses in survey methodology at the department during the autumn and spring terms 2001/2002.

### Update

The content of university level courses in survey methodology is dominated by sampling theory, while too little is covered regarding planning of surveys and non-sampling errors in surveys. Examination, too, usually consists of solving numerical problems in sampling theory, and even assignments given, as part of fulfilling requirements of the course seldom concern applying various sampling schemes.

It seems impossible to accomplish complete research projects of high quality at this level because of very limited time, and even the amount of work needed and the related costs are the limiting factors. But, the generally most time-consuming phase, data collection, can be substituted by a simulation.

With such a simulation programme available, the assignments have the following general structure:

- a) A group of students, usually three of them, is given a specific (fictive) research topic that is suitable to be investigated by performing a survey.
- b) The group writes a survey plan that includes specification of an appropriate sampling scheme, and constructs a questionnaire. The teaching assistant grades the plan and the questionnaire, each with a "pass" or a "fail". If any of them is deemed unsatisfactory, the group needs to correct prominent shortcomings before being allowed to proceed further.
- c) The teaching assistant fills in the questionnaires using the programme. The completed questionnaires are printed out and handed to the group. (In the case of non-response, some items or even complete forms are left blank.)
- d) The group codes the responses, analyses data statistically and writes a report, which then constitutes a part of the examination.

The research topics that we have given to our students concern a fictive community, Medel (Mean). For this community, we have created a population

register and a map, in order to emphasise the realistic intentions of the task. The students' survey plan must specify all the steps and procedures necessary for conducting a survey, including a questionnaire and a plan for cross-tabulation of variables.

The programme that provides responses to the questionnaires also generates questionnaire forms. There are five types of questions to choose from, which can be considered as limiting with respect to the choice of question types. As of data regarding the population, individuals' values for some of the variables including sex, age, address, family situation and income are found in the register while other variables that the groups need may be constructed from these original variables using the programme. We can in that way accomplish specific relations between different variables and thus achieve realistic cross-tables.

The programme that presently exists comprises a questionnaire generator. More flexibility may be achieved if we implement the ability of the programme to read in practically any kind of questionnaire. Such an extension is seriously contemplated at the moment.

A students' project results in a report. Tables and other statistical analyses are produced using a statistical package of students' preference. The statistical analysis constitutes a part of the results section of the report (and even a methodological part needs to be present in the report).

The programme has thus far been used twice in the intermediate-level survey methodology course. The first occurrence, in the fall of 2001, had a preliminary character. In the spring of 2002 a more complete attempt has been made. Instructions about the use of the questionnaire-generating part of the programme seem to have been understood reasonably well by the students, and the teaching assistant's work with providing answers to the questionnaires using the programme proceeded apparently without problems.

There exist, though, issues that are not yet resolved. What should be the content of that teaching which precedes handing in the survey plan? It is extremely difficult to cover both the topic of planning a survey and a sufficient amount of sampling theory needed for the students' choosing of an appropriate sampling scheme, all in this short time.

### Members in the project

Per Dahmström, project director, Associate Professor (docent och universitetslektor). Teaches in statistics and specialised in statistical computing.

Per.Dahmstrom@stat.su.se.

Håkan Slättman, Research Assistant. Has been working as an Assistant Teacher in statistics. Hakan.Slattman@stat.su.se.

For further information, please contact Per Dahmström or Håkan Slättman

## **Almost realistic surveys - simulated practical fieldwork**

*Per Dahmström*

*Stockholm University*

One of the aims of academic studies in statistics is to provide students with competence to plan, conduct and analyse surveys. It would be advantageous for the students if they had the possibility to conduct some surveys as a training during their study time.

The aim with this project was to develop a flexible teaching system which gives the possibility to conduct what we call "almost realistic surveys" during a relatively short time, employing a system to simulate the respondent burden. A fictitious population has been constructed, consisting of a small community of individuals. Samples from these populations act as respondents that give the students realistic data and, at the same time, a reasonable relationship between the answers to different questions. The population in the community consists of about 3000 inhabitants.

The course in survey methodology in the intermediate course in statistics at the Department of Statistics, Stockholm University, was before this project organised, as in most universities and colleges, around a textbook where the more "practical" job consisted of the exercises in the textbook.

During the project we have tried instead:

- ◆ Four lectures about planning of surveys
- ◆ Five lectures about sampling methods, estimation formulas for all methods studied, variance formulas for important cases

- ◆ Four lectures + one exercise covering analysis, sources of errors and the writing of reports
- ◆ Three exercises giving guidance for the simulated survey

***A training survey, from a student's point of view, consists of performing the following steps:***

- ◆ Receiving a specific survey assignment, individual for each group (consisting of 2 to 3 students)
- ◆ Planning of the survey (choice of sampling design and point and variance estimator, construction of a questionnaire)
- ◆ Handing over the plan for approval
- ◆ Drawing a sample
- ◆ Conducting a pilot survey
- ◆ Revising the questionnaire, if necessary
- ◆ Registering data
- ◆ Performing non-response follow-up
- ◆ Processing data
- ◆ Performing the analysis
- ◆ Writing a report

***Realisation of the project***

First, we determined the content of the population register and designed it. Second, we conducted the programming phase. Third, the whole system was tested in the courses in survey methodology at the department during the years 2002 and 2003.

The content of university level courses in survey methodology is dominated by sampling theory, while too little is covered regarding planning of surveys and non-sampling errors in surveys. Examination, too, usually consists of solving numerical problems in sampling theory, and assignments given as part of fulfilling requirements of the course not seldom concern applying various sampling schemes.

It seems impossible to accomplish complete research projects of high quality at this level because of very limited time, and even the amount of work needed and the related costs are the limiting factors. But, the generally most time-consuming phase, data collection, can be substituted by a simulation.

With such a simulation programme available, the assignments have the following general structure:

a) A group of students, usually three of them, is given a specific (fictive) research topic that is suitable to be investigated by performing a survey.

b) The group writes a survey plan that includes specification of an appropriate sampling scheme, and constructs a questionnaire. A teaching assistant grades the plan and the questionnaire, each with a "pass" or a "fail". If any of them is deemed unsatisfactory, the group needs to correct prominent shortcomings before being allowed to proceed further.

c) A teaching assistant fills in the questionnaires using the programme. The completed questionnaires are printed out and handed to the group. (In the case of non-response, some items or even complete forms are left blank.)

d) The group codes the responses, analyses data statistically and writes a report, which then constitutes a part of the examination.

The research topics that we have given to our students concern a fictive community, Medel (Mean). For this community, we have created a population register and a map, in order to emphasise the realistic intentions of the task. The students' survey plan must specify all the steps and procedures necessary for conducting a survey, including a questionnaire and a plan for cross-tabulation of variables.

The programme that provides responses to the questionnaires also generates questionnaire forms. There are five types of questions to choose from, which can be considered as limiting with respect to the choice of question types. As of data regarding the population, individuals' values for some of the variables including sex, age, address, family situation and income are found in the register while other variables that the groups need may be constructed from these original variables using the programme. We can in that way accomplish specific relations between different variables and thus achieve realistic cross-tables.

The programme that presently exists comprises a questionnaire generator. More flexibility may be achieved if we implement the ability of the programme to read in practically any kind of questionnaire. Such an extension is seriously contemplated at the moment.

A students' project results in a report. Tables and other statistical analyses are produced using a statistical package of students' preference. The statistical analysis constitutes a part of the results section of the report (and even a methodological part needs to be present in the report).

### ***Experiences***

The programme has in its present shape been used four times, the latest occasion being autumn 2003. An evaluation of the programme has been performed after that, with two target groups: students and teaching assistants. The students found the programme in general easy to use, and its graphical user interface intuitive. They have nevertheless had some difficulties in understanding that not all kinds of question types may be generated by the programme. (This difficulty is common for all questionnaire-generating computer programmes, where it is, for instance, necessary to distinguish between single-choice and multiple-choice questions: these cannot be combined.) The students have also given some recommendations for further improvements in the programme. Two teaching assistants gave additional comments on the programme from the teacher's point of view.

Teaching assistants' work with providing answers to the questionnaires using the programme proceeded apparently without problems, but showed to be relatively time-consuming: depending on what is specifically asked in a questionnaire, it takes on average about two hours to fill the answers. The difficulty arises when data necessary to answer a question asked in the questionnaire does not exist in the population frame nor in a data set eventually prepared in advance by the teaching assistant. Such data need to be derived from existing data. The task is performed in some external programme (e.g. Microsoft

Excel, Minitab or Matlab), often using some rudimentary programming constructs.

A crucial problem, from the teaching point of view, has been: what should be the content of the teaching that precedes handing in the survey plan? Namely, in the present set-up of the course, the lectures and the work on the assignment (i.e. the training survey) proceed in parallel. In order to hand in an acceptable survey plan, the students need to acquire knowledge on: (a) planning and conducting a survey, (b) sampling theory needed for making a choice of an appropriate sampling scheme and estimator. It is extremely difficult to cover all this sufficiently fast in order to leave to the students enough time to even perform the study, register and analyse the results, and write the reports.

Thus, the short time used for the course calls for a very careful planning of the content and the disposition.

### ***Members in the project***

Per Dahmström, project director, Associate Professor (docent och universitetslektor). Teaches in statistics and specialised in statistical computing. Per.Dahmstrom@stat.su.se.

Håkan Slättman, Research Assistant, Programmer. Has been working as an Assistant Teacher in statistics. Hakan.Slattman@stat.su.se.

Boris Lorenc, Teaching Assistant, Ph. D. student. Has been engaged on the course since spring 2002. Boris.Lorenc@stat.su.se

Daniel Bruce, Teaching Assistant, Ph. D. student. Has been engaged on the course in autumn 2003. Daniel.Bruce@stat.su.se

099/99

Per Dahmström

Department of Statistics

Stockholm University



Tel: +46 (0)8 16 29 84

Fax:: +46 (0)8 16 75 11

E-post: Per.Dahmstrom@stat.su.se

2004-01-15