Chapter 10. Surveys

Table of Contents

What is a survey?

Surveys involve the collection and analysis of information from a group of people drawn from a specifically targeted group of interest. Surveys of inanimate objects can also be conducted; but where people are the subjects a questionnaire of sorts is usually employed, although interviews are also common. Interviews and sequenced questioning can of course also be employed in action research, case studies and experiments. "Survey research is the method of collecting information by asking a set of pre-formulated questions in a predetermined sequence in a structured questionnaire to a sample of individuals drawn so as to be representative of a defined population" (Hutton 1990, Blaxter book).

A survey is the planned collection of information in a standardized form that aims to give an overall perspective to a field. It permits a number of variables to be examined, which can be used to test several hypotheses and can be analysed in a number of different ways. Surveys are easily replicated when standardized sampling and analysis techniques are used. The accuracy of responses will vary according to how well respondents understand the questions and the issues, how much bias exists, and whether extraneous variables are distorting results.

A survey is typically used in the following standard way: a hypothesis is proposed, a questionnaire and/or interview designed to test the hypothesis, a group of interest ("population") chosen, a sample selected from the target population, the data collected and analysed, the hypothesis tested.

Survey design

To conduct a survey one needs to know what one is testing for and how to obtain measures of this. Having a topic, the characteristic to be studied should be determined first (e.g. knowledge of query optimization); then the indicators to measure this decided upon (e.g. questions relating to specific optimization techniques, questions about the query optimizer manual, etc.) and their relative weightings established. The more indicators one has for measuring a characteristic, the greater the accuracy with which that characteristic can be determined. The weightings help in the case where different indicators for the same characteristic give different measures for that characteristic (e.g. questions relating to optimization techniques would carry more weight than questions relating to knowledge of that part of the manual).

In framing questionnaires and interviews, one should beware of one's own opinions influencing the research. Your personality and views have an impact on the questions you select, how you ask them

and whom you ask; care must be taken to minimize this impact on your work. Usually more than one iteration is needed, with the questions and/or the hypothesis improved each time. The first iteration (or if this goes badly, the first few) serves as a pilot study; a small sample used for the express purpose of finding problems with the survey. Since the results are not of interest, it is best to target as diverse a group of participants as possible in the pilot study, rather than a random sample.

Interviews vs questionnaires

Questionnaires, being pre-determined, limit the scope and amount of feedback compared to interviews which, even if highly structured, offer far more flexibility and typically more lengthy responses. It is useful to follow up the use of questionnaires with some interviews to give more detailed perspective. It can also be helpful to conduct interviews before drawing up a questionnaire, in order to better understand the issues and how to approach them.

A questionnaire is most useful for asking a large number of respondents a highly structured set of unambiguous questions without their being influenced by an interviewer or the sensitivity of the questions. If these are sent by mail or e-mail, they enable a geographically distributed population to be sampled in a limited time, and can be answered at leisure and anonymously by respondents. Interviews are more useful for in-depth probing, for following up on any interesting remarks, for collecting information on more complex issues (since questions and answers can be clarified where necessary), for open-ended questions, for eliciting a large amount of data from individuals, for making certain that only one person's views are given, and for ensuring that respondents do not ignore some or all of the questions, nor answer them in the wrong order. A skilled interviewer can also provide some of the benefits of questionnaires - by adhering to a set question structure, and by taking care not to influence, bias or distort results through their own reactions and comments. However, unlike the use of questionnaires, large numbers of individuals from diverse locations cannot be interviewed without considerable resources; hence a survey may need to incorporate both interviews and questionnaires.

Designing the questions

In a well-designed questionnaire/interview

- only pertinent questions should be asked
- introductory sentences, polite phrasing and non-antagonistic language are needed to maintain a good rapport
- the correct choice between nominal, ordinal, interval or ratio measurements must be made and categories must be appropriately chosen
- for category, multiple choice, scale, LPC scale, ranking and grid/tabular questions, the rule of "seven plus or minus two" should be adhered to - i.e. the number of options that most people can correctly distinguish and choose between is 7; some can only handle 5 choices and 9 is an upper limit
- with LPC scales, consecutive questions should not have all the good extremes on one side and all the bad extremes on the other - this should vary from question to question. For example:

  Place a cross at the point which best describes the query page:

```
  Easy to Use    1   2   3   4   5   6   7    Difficult to Use
    Confusing    1   2   3   4   5   6   7    Clear
Quick to Use     1   2   3   4   5   6   7    Tedious
```

Using clear language

Careful, unambiguous wording of questions and of answers (in the case of category, multiple choice, scale, LPC scale, ranking and grid/tabular questions) is essential for clarity and accuracy in a questionnaire or interview.

A checklist for the phrasing of questions is given below:

- use simple language
- have concise questions
- elicit information that is suitable for analysis
- a question should cover exactly one aspect and no more
- avoid leading questions and beware of your personal opinions and biases inadvertently showing through
- phrase potentially sensitive questions as harmlessly as possible
- use precise questions that are as specific as possible
- clarify whether you want a fact or an opinion
- if time is part of a question (e.g. "currently", "in the past", "regularly") be specific about the period of interest
- use legible, uncluttered questions attractively laid out for easy completion
- leave sufficient space for legible replies to open-ended questions
- do not have too many questions or questions that are difficult to answer
- avoid jargon, buzzwords or any terms subjects may not know or may misinterpret (when it comes to IT in particular, people are loath to admit ignorance)

Questionnaires

Kinds of data gathered from questionnaires

Questionnaires gather some qualitative data (where respondents can comment freely), but the vast majority of information collected is quantitative. This data can be of four kinds: interval, nominal, ordinal and ratio data. Interval data is measured on a scale with a set difference between points but with any base/starting point (e.g. dates). With nominal data a coding system assigns a number to represent a particular response but the number has no significance (e.g. coding gender as 0 or 1). In contrast, ordinal data permits qualitative responses to be coded in numeric sequence to reflect ordering e.g. "useless", "poor", "average", "good", "excellent" could be coded as 1, 2, 3, 4, 5 respectively. Ratio data comprises measurements on a scale where there is a true zero e.g. distances, incomes, etc.

Simple questions

There are 8 main types of survey question. The first four simple kinds of question are illustrated by example below (preceded by the question type in brackets):

1. (quantity or information)

   ```
   In which year did you matriculate? _____________
   ```

2. (category)

```
Have you ever undertaken part-time study?
 [] Yes, currently
 [] Yes, in the past
 [] No, never
```

3. (list or multiple choice)

```
To which of the following Faculties did you apply first?
 [] Commerce
 [] Engineering
 [] Health
 [] Humanities
 [] Science
 [] None of the above
```

4. (scale)

```
How do you feel about your chances of completing this degree?
 [] Very confident
 [] Somewhat/fairly confident
 [] Neutral
 [] Somewhat concerned
 [] Very worried
 [] Not sure how I feel about this
```

Other types of question

The four other types of question are somewhat more difficult to answer. They are illustrated by example below (preceded by the question type in brackets):

1. (LPC scale or "semantic differential scale")

```
Do you prefer command-line interfaces or graphical user interfaces?
 Command-Line  1  2  3  4  5  6  7  Gui
```

2. (ranking)

```
What IT fields are you most interested in? Rank those of interest
in order from 1 downwards:
 [__] Artificial Intelligence
 [__] Computer Architecture and Hardware
 [__] Databases
 [__] Distributed Computing
 [__] Graphics
 [__] Human-computer interaction
 [__] Networks and telecommunications
 [__] Programming Languages
 [__] Other (please state) _____
```

3. (complex grid or table)

```
How would you rank your skills in the following?
```

```
                  Very Good    Average      Weak     Not Sure
      Reading
      Writing
  Programming
Presentations
     Teamwork
```

4. (open-ended)

   Do you have any suggestions for better marketing of the degree?

   _____

   _____

   _____

   _____

Preparing a questionnaire

The layout of a questionnaire should be inviting and the text legible. Questions should be precise and unambiguous; use two simple questions rather than one complex one. Before conducting a survey, try to find different interpretations of the questions and then re-word them to clarify which interpretation is required. Ask friends or colleagues to do likewise. Avoid phrasing that is negative or presumptive of a particular response. Limit the use of open-ended questions as they are time-consuming to answer and to analyse, and end by thanking the respondent and asking for any additional comments. As with most research methods, it is best to conduct a pilot study to detect problems with your questionnaire so that you can improve it; (Blaxter book) reports that as many as 8 attempts at a single question have sometimes been necessary! Once a questionnaire has been drafted, it is advisable to show it to an individual or unit that provides consultancy services for surveys, or else to statisticians who can comment on any potential problems they foresee in analysing responses.

Once a questionnaire has been finalized, researchers must decide which method(s) will be used to distribute it. The post, e-mail, a Web-site and direct visitation are possibilities to consider. Care must be taken in using the Web, e-mail, telephonic or postal surveys if a random sample is required - Web site questionnaires can be completed by almost anyone who is online, including people who are not part of the population of interest; the Internet itself is not accessible to everyone (particularly in developing countries); many young people have only cellphones and are excluded if the telephone directory is used; material posted to a household will not be opened by children; etc.

Maximising response rate

Candidates are more likely to complete a questionnaire if they know that it is part of a study being conducted by a reputable organization and they understand significance and usefulness of the research, so include this in the introduction. Careful timing of the exercise can make a big difference to the response rate, so check when is most and least convenient for your population. Another useful incentive is offering participants access to the results of the survey. This can be relatively easily done via the Web; where resources permit, a hardcopy can be offered to those who have Internet access problems. Response rate is highest if the researcher is present when the questionnaire is completed, or asks the questions over the telephone; otherwise, some chasing up should be scheduled into the project plan.

Analysing questionnaires

Questionnaire analysis is mainly quantitative rather than qualitative in nature, because of the discrete nature of answers to most questions and the volume of questionnaires to be processed. Some form

of statistical analysis is therefore used. This can be fairly simple - e.g. indicating which proportion of respondents gave specific answers, and which answers to specific questions appear to be related. Here straight forward measures of tendencies (e.g. mean (average), median (middle value) and mode (most common response)) and diversity (e.g. range and standard deviation) can be used. To go beyond this, e.g. to compare one sample/population with another in order to determine how similar or how different they are, inferential statistics or multivariate analysis is required. The Chi-square test is used to compare sets of values and the Student's T-test to test means. Using these statistical tests it is important to check that the assumptions under which they apply are true for your data, otherwise they must not be used. For example, the Student's T-test assumes random sampling of interval data with a normal distribution.

If two variables you have measured appear related, this may be due to the effect of some third variable. To demonstrate causality you have to find or suggest a mechanism linking the variables together. Multivariate analysis - e.g. cluster analysis, multiple regression, factor analysis - examines relationships among three or more variables. These can be performed easily by using the appropriate computer package, but it is important to understand their purpose and foundation in order to use such software.

If the number of responses to a questionnaire is small (e.g. less than 10), results showing the actual number of respondents choosing each answer are more honest and clearer than those giving percentages. For medium to large surveys, i.e. where the sample size exceeds 20, percentages are easier to understand and compare.

Interviews

Conducting interviews

An interview can be recorded either by taking notes or by using a tape/video recorder. Audio/video-taping is intimidating and likely to cause anxiety and reticence in the interviewee Note-taking is cheapest and saves considerable time compared with transcribing from tape, but it distracts both interviewer and interviewee (who can become concerned at how much importance is being attached to what they say/do according to when things are written down). One solution is for the interviewer to bring an assistant - someone with knowledge of the project and subject area who is skilled at taking notes; if this person remains in the background, neither interviewer nor interviewee/participant need be distracted by the process.

Analysing interviews

Analysis of interviews should move from the specific (e.g. enumerations) to the more general (e.g. categorizations and detection of common themes). In order to analyse interview transcripts and notes it is best to compare different interviews, looking for similarities and for significant remarks. Since interview statements are made within a specific context, they must not be quoted without knowledge of that context, or in the wrong sequence.