9 Gathering Quantitative Data

Babak Taheri, Liang Lu and Nikolas Valantasis

In this chapter ...

The role of managers and researchers is concerned with analysing and solving problems. These problems come in many forms with common features and normally include some numerical information. Both managers and researchers need to understand a range of quantitative methods analyses. This chapter focuses on sampling and measurement issues, surveys and experimental research, and the principles of decision making studies.

The nature of quantitative research

Social science research has tended to be influenced by the hypothetico-deductive paradigm (a research approach that starts with a theory about how things work and derives testable hypotheses from it), and how this informs a quantitative methodology. Quantitative studies are defined as: quantifying the problem and understanding how widespread it is by seeking projectable outcomes for a larger population, and which are associated with the interpretation and presentation of numerical information. The following phrases are linked with a quantitative methodology: a deductive approach, an etic view, objective epistemology, a structured approach, systematic approach, numerically-based data collection, statistical analyses, and replicable research design. In other words, quantitative studies have four main characteristics: systematic/reconstructed logic and linear path (step-by-step straight line); hard data in nature (e.g. numbers); they rely on positivist principles, they have an emphasis on measuring variables and testing hypotheses; finally, they usually verify or falsify a relationship or hypothesis which the researchers already have in mind.

Sampling and measurement

Population, sample size and type of sampling

Considerable efforts should be made to ensure that the sample obtained is representative of the population under investigation (i.e. 'the generalizability issue'). A population is normally a collection of all the concerned units that researchers would like to study within a particular problem space. It is vital to define a research population clearly before beginning to collect our sample. Normally, it is impossible to survey all members of a given group. Therefore, researchers use two broad types of sampling techniques: probability/random and non-probability/non-random. Probability sampling is a way of achieving samples that are representative of the whole population and involve random selection. Non-probability sampling involves a specifically chosen sample based on particular characteristics or similar differentiating feature relevant to the study, therefore it cannot be determined whether the results of the study are representative of the entire population. There are two main assumptions for probability sampling (Walter, 2013). First, the sample frame, which is a list of the population of interest from which we choose the individuals for inclusion in our study. For example, in order to survey university students studying in the city of Glasgow, we need a list of all university students in Glasgow, however such a list could be hard to come by because there are many different types of people who could be defined as a 'student' and Glasgow is a city which has four large universities, as well as a range of students studying on-line or via distance learning with universities located elsewhere. The second assumption is that researchers are able to randomly select features from their sampling frame in relation to a mathematical justification in probability sampling including sample statistics, population parameters and the degree of sampling error. Table 9.1 shows a summary of probability and non-probability sampling.

Table 9.1: Types of sampling

Sampling sub-groups	Description
Probability sampling	
Simple random	Every unit has an equal chance of being selected; it requires a good sampling frame; the population is geographically concentrated
Systematic	It is similar to simple random sampling except that it is simpler; researcher should consider periodicity of sampling frame; the selection of one unit is dependent on the previous unit
Stratified random	The population is divided into strata, and these strata make up the final sample in the study; it is mainly based on homogeneous subgroups, e.g. gender and age
Multistage cluster	It involves several different samples; researcher mainly wants to study clusters in geographical areas
Non-probability sampling	
Convenience	Accidental, haphazard, chunk and grab sampling; selection of participants for a study is based on their proximity to the scholar
Purposive	Researcher makes decision about who/what study units will be involved in the research
Snowball	Researcher does not know about formal/informal network connection at the start of study, but when he/she begins the research process by identifying someone who meets the criteria for involvement in the study, this person acts as the link to the next participant through their own network, hence the sample 'snowballs'
Expert	Researcher identifies some people as expert with demonstrable experience in some particular area
Quota	Researcher specifies the minimum number of sampled units he/she wants in each category; researchers give organised quotas in terms of characteristics in order to find out the distribution of the variable in the population; e.g. street sampling

No sample is likely to produce results that accurately represent the entire population, as is easily seen by the inaccuracy of opinion polls during elections. Debatably, it is always possible to pick, strictly by chance, a group whose members happen to be different in some attribute from the population as a whole (i.e. the sampling error). This brings attention to the systematic bias (i.e. extraneous sampling factors) which affects survey results and reduces data validity, including frame bias (a wrongly chosen population), selection bias (under-representing certain types of population members), non-response bias (data skewed based on who from the chosen sample actually engaged with the study), interviewer bias, questionnaire bias, respondent bias and processing bias (interviewer writes down the wrong answers); see De Vaus (2007) for more information.

There is considerable debate over what constitutes an acceptable sample size for the results to be statistically valid, with there being no accepted rule to describe a suitable sample size. It is generally acceptable that the bigger the sample is, the more the results of the study are generalizable. Different authors recommend diverse sample sizes as appropriate for quantitative research, including an absolute sample from 200 to 300 (De Vaus, 2007; Hair, Black, Babin, & Anderson, 2010). It is arguable that sample size depends on a number of considerations and parameters and there is no one ultimate answer, however for large populations small samples are less problematic.

Previous studies suggest a general guideline for using confidence intervals (an estimated range of values which include an unknown population parameter with certain probability) to configure sample size. For example, for a sample of 500, confidence levels range from 1.9 percent for homogeneous populations to 4.4 percent where there is a 50/50 split on variables. There are three factors that determine the ideal sample size, namely: confidence level (corresponds to a Z-score), population size, and the margin of error which is beyond the scope of this chapter (see De Vaus (2007) for more information). Finally, in the ideal world, some respondents may decide not to participate in the survey. This may become an issue when non-respondents differ from respondents in a non-random manner, which consequently introduces sampling bias (Walter, 2013). For example, the responding members of the sample may be more educated and older than the actual education and age distribution represented in the population from which the sample is drawn. Researchers generally use a formula to calculate response rate:

$$\text{Response rate} = \frac{\text{number of surveys completed}}{\text{number of sample} - \text{unreachable}} \times 100\%$$

Main measurement types

Data analysis and design involves measuring variables which could be dependent and independent. We define dependent and independent variables as: