Sampling

The world of politics and international relations is vast, made up of many countries, organizations and societies, and it is often the case that it is physically impossible to research every individual member (in terms of time, cost and access). Therefore, much political research relies on the sampling of research populations – we study some cases in order to make generalizations about the group as a whole. In relation to measuring public opinion, sampling theory tells us that a carefully selected sample of 1000 people will give a very good indication of the views of the population. As Fielding and Gilbert (2000: 226) state, 'It is quite rare for it to be sensible to conduct a census, that is, to obtain data from everyone in the population. By measuring just those in a sample, time and money can be saved that are better used on other aspects of the research, while still obtaining sufficient accuracy for valid conclusions to be drawn.'

The first stage in applying sampling is to identify the research population – that is, the collective group of people who share a common characteristic. This may be those who are registered to vote, those who are elected representatives or those who are members of certain campaigning organizations. Those who are members of the research population provide the sampling frame (Burnham et al., 2004: 86):

The sampling frame is the list of units from which the sample is to be drawn. The choice of sampling frame will depend on what lists of the population are available and how accurate they are. The choice will also depend on the subject of the research.

Clearly, there is going to be a relationship between the accuracy of the sampling frame and the ability to choose a representative sample. Unsurprisingly, the first and often significant challenge that a researcher will need to overcome is locating population lists, particularly if the research population is large and likely to change on a frequent basis. For example, electoral registers are updated at regular points, but if you are researching an area with a very mobile population, this source will quickly become unreliable. A sampling error may occur if the proportion in the sample is different from its true value in the population as a whole.

Depending on the sampling techniques used, we may or may not be able to make confident inferences about the research population (see causality and correlations). Being able to make inferences based on the relationship between independent variables and dependent variables can be a complicated process. There is no 'set figure' for how large or small a sample should be and, as Payne and Payne (2004: 201) state, it is likely to be determined by three factors:

- the resources available, including financial and access issues
- planned analysis method
- the homogeneity or heterogeneity of the research population.

In the end the sample size must take into account the degree of diversity in the population on key variables, the level of sampling error that is tolerable and the reliability required of the sample. Decisions about one factor have implications for other factors. (de Vaus, 2004: 81)

Once we progress beyond research populations that are relatively small in size and easy to identify, our opportunities for probabilistic sampling are reduced. A common misunderstanding among students untrained in appropriate research methods is that standing in the middle of a town centre and asking passers by to respond to a questionnaire is the equivalent of probabilistic, or random, sampling. It is not, as you will find very different people in the centre of town depending on the day of the week and even the time of day. The ‘typical person in the street’ will vary, for example, by age and occupation depending on whether you are carrying out your survey at 9 a.m. or 9 p.m.

Probabilistic sampling can take the following forms:

- A random sample, in which every member of the research population has an equal chance of being selected. de Vaus (2004: 71) stipulates that random sampling is, to a large extent, reliant on the existence of a good sampling frame. A key advantage of random sampling is that ‘it is free of the systematic bias that might stem from choices made by the researcher’ (Gorard, 2003: 67). However, many research studies you may come across will include ‘case studies’ of particular social or geographic groups, the justification being that certain groups are easier to access in practice than others.
- A systematic sample is a frequently used technique when the sampling frame takes the form of an address, telephone or e-mail
directory. The sample is created by selecting every nth case – for example, you could send a questionnaire to every tenth e-mail address in a long list. Yet, this can be problematic for, as de Vaus (2004: 73) claims, ‘a certain type of person may recur at regular intervals within the sampling frame. If the sampling fraction is such that it matches this interval, the sample will include only certain types of people and systematically exclude others.’

- A stratified sample entails dividing the sampling frame into subgroups or strata and then taking random samples from each. This may be useful if we believe that particular relevant variables are underrepresented in our research population and this scarcity does not lead to a lack of representation in the sample. We could, for example, stratify university students by age categories to ensure older students (such as those over 30) are not absent from the sample.

- A cluster sample is used if we wish to study a research population that is widely dispersed, particularly when we have limited resources for completing the data-collection process. As Gorard (2003: 70) notes, ‘Using a clustered sample implies not so much a difference in selection procedures as a difference in defining population units. The cases we are interested in often occur in natural clusters such as institutions’. We might, for example, construct a study that is interested in black and ethnic minority (BME) political participation and the impact of critical mass – that is, patterns of participation in areas of high BME density different from that in areas with low BME density? Electoral wards may form the source of our clusters (or what we would term cases), with potential participants then being selected by random or stratified sampling.

You may also see reference to ‘booster samples’, but this is not a sampling method per se, it is a strategy adopted by researchers to ensure the representation of a particular research population subgroup is large enough to make confident claims without having to substantially increase the sample size for the entire study. Booster samples have been applied to Scottish, Welsh and ethnic minority strata of the British Election Studies at various times.

There are often very practical reasons for adopting non-probabilistic sampling and this is often underpinned by an inability to locate a reliable sampling frame. Gorard (2003: 72) suggests that it is most appropriate to use non-probabilistic sampling in research design, particularly in pilot studies for trialling the research design rather than gathering data per se. Non-probabilistic sampling can take the following forms:

- A snowball sample, in which the researcher uses a small number of contacts to generate a larger sample group. For Devine (1992), this was a useful technique as her study of car workers took place during a period of poor industrial relations, but she found that speaking to a small group of respondents opened up the opportunity to expand her list of contacts. The challenge of such a sampling approach lies with the fact that ‘the quality of the sample depends on the starting point, and strength, of the network’ (Payne and Payne, 2004: 210). This was a problem encountered by Lupton and Tulloch (2002: 321) in their Australian study of notions of risk in society. Snowball sampling was used to recruit participants, but ‘this strategy, however, was only partly successful in achieving a heterogeneous group of interviewees, as the group was dominated by well-educated, young and middle-aged adults of British ancestry.’

- A quota sample relies on the selection of characteristics to reflect their distribution within the research population. So, if we know that 10 per cent of university students are over 30, we select a sample that reflects this quota. However, the researchers may be selecting a sample that is not altogether typical – that is, while the required characteristics may be matched, the sample may be substantially different in other ways. As Payne and Payne (2004: 212) say, ‘The danger is that interviewers select those easiest to contact, or who seem friendly and approachable, whose answers may not be representative.’

- A volunteer, or convenience, sample is commonly employed in ‘popularity polls’ – the kind commonly conducted by television and radio programmes in which audiences are invited to phone or text in their choice of winner or ‘evictee’. Political researchers are unlikely to use this approach, however, and certainly would not attempt to make claims about representativeness, as those who choose to express their views often have a motive for doing so.

**FURTHER READING**


