Sage Research Methods

Survey Research and Sampling

Author: Jan Eichhorn
Pub. Date: 2024
Product: Sage Research Methods
DOI: https://doi.org/10.4135/9781529682793
Methods: Survey research, Sampling, Sampling distribution
Keywords: population
Disciplines: Sociology, Criminology and Criminal Justice, Business and Management, Communication and Media Studies, Education, Psychology, Health, Social Work, Political Science and International Relations
Access Date: August 29, 2024
Publisher: SAGE Publications Ltd
City: 55 City Road
Online ISBN: 9781529682793

Sampling Approaches: How to Achieve Representativeness

Introduction

The whole goal of survey samples is to enable us, using quantitative methods, to analyse social scientific questions in relation to large groups of people – thousands or even millions of them. If we wanted to gather data about each person, for example, who is registered to vote in the national elections of a particular country, say Japan, we would have to talk to an unfeasibly large amount of people, in this case more than 100 million (International Foundation for Electoral Systems, 2018). The amount of time and resources this would require is prohibitive. So, what is the next best thing that we can do instead? We can try to identify a smaller group of all the registered voters that is in its characteristics close to identical to the characteristics of the group we are interested in. **Sampling** then allows us to find a compromise between the degree of accuracy we aim to achieve with regard to a question we have about a large group of people and the resources that we have available to address the question (Stephan & McCarthy, 1958, p. 12). Sampling is complicated and requires a lot of thinking and the making of difficult decisions between a great variety of possible approaches. When considering all aspects that we have to take into account when engaging with sampling, it can seem quite daunting, in particular when our theoretical ideal cannot be achieved in practice, because of limitations that may hinder us to access those respondents we would ideally like to reach. However, the effort is worth it, because sampling allows us to do something very powerful. Using robust approaches to selecting a sample and developing meaningful surveys, we are able to make statements about characteristics of groups much larger than our sample.

If we are confident that the composition of our smaller *sample* is then approximately similar to that of the larger group (which is referred to as the *population*), we can conduct analyses with only our sample, but we are able to make statements about the population as a whole, with a degree of certainty. There will be some margin of error around the estimates from our sample, but within that margin we can be confident that we can talk about our population as a whole. For this to work, it is crucial though that our sample indeed reflects the characteristics of the population. If it does not, any analyses of the sample only reveal to us what the sample looks like, but we could not generalise to our population. So, it is important to 'get the sampling right', if we want to make statements about our target population. That is why we need to pay close attention and understand how the decisions about our approach to sampling affect the results thereof. In this chapter, we will introduce the most important approaches to sampling. However, before we begin to discuss the samples, we first need to be clear about what constitutes our target population.

Box 2.1: Case Study

Predicting the US presidential elections 1936: Why sampling is so important

(Based on Squire, 1988)

In 1936, the *Literary Digest*, a popular magazine in the USA, conducted a poll to predict who would win the presidential elections that year – the incumbent, Franklin Roosevelt, or the main challenger, Alf Landon. The magazine had correctly predicted who would win the elections on all previous occasions since 1920 and prepared a large-scale effort to do this again. They sent out more than 10 million ballots and received more than 2.2 million returns that they counted. The selection of their sample was based on entries in car registration lists and phone books. Such a large sample would initially instil a fair deal of confidence in anyone reading about it without any knowledge of sampling methods. However, the result was not marginally but categorically different from the real results. While the magazine's poll suggested Landon would win with 55% of the vote, followed by Roosevelt with 41%, the actual results on election day saw Roosevelt re-elected with a massive 61%, while Landon received only 37% of all votes cast. At the same time, based on a competitor survey conducted by Gallup, Roper and Crossley predicted Roosevelt's win correctly (albeit not with the perfect percentages). That survey, however, had only a few thousand respondents. How could it be that such a small survey could outperform such a large sample?

The crucial answer lies both within the design and the actual undertaking of the sampling process. Good surveys measuring national public opinion can commonly achieve meaningful results with rather small samples of 1000 to 2000 respondents, while surveys with much larger samples, but with much poorer sampling designs, may produce results that are not representative of the real population at all. The 1936 *Literary Digest* survey suffered from two important shortcomings. First, the initial sample was biased and included a disproportionate amount of Landon voters. In the period following the Great Depression, one suggested explanation was that basing a sample on car and telephone registration lists biased the sample against people who were hard hit by the depression and may not have had access to either good. That effect may have been stronger than the effects in previous elections, before and in the early phase of the economic crisis, where fewer people had experienced strong negative impacts that may have distorted the sample. Second, Landon supporters were more likely to respond to the questionnaire, and therefore, when looking at the returned straw ballots, an overestimation of the support for Landon was registered.

Sample size is important when conducting survey research, but the quality of the composition of the sample is even more fundamental. Regardless of the sample size, a poorly constituted sample will lead to potentially heavily biased results. Understanding the different approaches to sampling and their advantages and shortcomings is therefore crucial to the design of useful surveys.

Populations and sampling frames

At first, it may sound strange and nearly trivial to say that identifying a target population carefully is of paramount importance. Is it not usually rather obvious what we want to study? Actually, it can be quite tricky. Let us imagine that your overarching goal is to investigate public attitudes in a recent election within a country. Who comprises the set of people that should form part of your investigation? Is it everyone who lives in a particular country, because they are all affected by what is being decided by the politicians elected to office? Or should it only be about those people who are actually eligible to vote, as you may want to focus on the behaviour of the electorate specifically, rather than the perceptions of people in the country more broadly? If that is the case, do you need to include people who have the right to vote but live outside the country of interest and take part through external voting (e.g. in embassies or by international mail)? Could it be that you are actually only interested in the behaviour of people who took part in the election and, therefore, you should exclude people who are eligible to vote but who decided not to participate?

Something that might sound like a straightforward theme, such as an election, can be looked at in many ways, even when focussed on public attitudes. It reminds us that, as for any good research, we must start with a well-defined understanding of our goals and, ideally, a clear research question. Once we have developed it, we can actively make a decision about who needs to be included in our conception of the population for our study and who should be excluded. The population should include everyone who matches the criteria required by our question. In order to do this well, Sudman (1976) suggests a two-step process. First, we should identify the units of your analysis, and second, define the characteristics that need to apply to them.

With regard to the former, survey research is often interested in individual persons. But that does not always have to be the case. Our unit of analysis could not only be families or households, for example, but also organisations (e.g. firms), organised groups of individuals (e.g. initiatives or community groups) or transactions (e.g. exchanges in a particular market place). After identifying that clearly, common characteristics that may further describe the structure of our sample may include geography (e.g. if we are only interested in the electoral results in a particular part of the country), the age of persons (e.g. if we are investigating first-time voters only), other demographic variables (e.g. gender, marital status or education) or a range of measures that could reflect personal background or household composition (Sudman, 1976, pp. 12–13).

When we try to emulate the population in our samples, it is important that we only formulate insights that speak to the population our sample is designed to represent. This is very important and often a challenge. The broader and more complex the population is, the harder it can be to derive a sample that genuinely matches its composition. Conversely, the more specific and narrow the parameters defining our population are, the more difficult it can be to actually identify potential respondents for a survey. We see the former concern, for example, often prominently raised in experimental studies based on samples composed of students, who can easily be accessed by researchers at universities (Kam et al., 2007, p. 416). But, obviously, there are many reasons to question whether a finding based on a group of university students is reflective of a population that is conceived of as encompassing more groups than university students in a particular place (Sears, 1986) – though some research suggests that such student-based samples may often provide equivalent results to non-student samples (Druckman & Kam, 2011). These considerations also act as an important reminder that findings from respondents in one cultural context may not be translatable directly to people elsewhere in the world, so even if a sample is representative for the population of a country more broadly, we should be cautious to generalise with regard to people more widely (Henrich et al., 2010). This is an issue we will discuss more in Chapter 5.

The latter problem may be particularly prominent if we specify a population that is very small, but not linked together through a simple structure of comprehensively identifiable institutions or organisations (e.g. female vegans aged 45–60 years). These considerations highlight the difficulty in identifying a suitable so-called *sampling frame.* After defining a target population, we obviously want to create a useful sample thereof. In order to achieve this, we need to have access to all people who make up that population in the first place. That, however, is often easier said than done. While it may be fairly straightforward to identify a list of all participants in a programme at a given university, for example, it could be very hard to identify a list of contacts for all university students in a country overall. While official databases exist for certain populations in some countries, such as electoral registers, properties of other registers may be less well known. Even when we have a register, it may be incomplete and particularly under-represent certain groups of the population (see Volume 1 in this series for a discussion of non-coverage and other survey errors). Even if full sampling frames exist in certain registers, they may not always be accessible to everyone straightforwardly, highlighting several of the practical issues that may arise in finding a suitable sampling frame that corresponds to our target population (Moser & Kalton, 1971, p. 48).

Rachel Ormston, one of the expert interviewees for this volume, highlights the difficulties in sampling very specific groups, looking at surveys of young people on the one hand and minority ethnic groups on the other hand in the following discussion.

Box 2.2: Ask an Expert

Rachel Ormston | Sampling narrowly defined groups of respondents

What approaches work best to reach very specific, smaller groups of the population, like young people or ethnic minority groups?

With young people you can obviously do school-based surveys if you’re looking at school-age young people, but then you have to take into account the fact that that will miss out young people who are not in education. . . . Also I think you have to think about subject matter, because there’s been some quite interesting studies around asking young people about crime. If you ask them in a survey at school you get different responses because they feel different about answering that in different contexts. With minority ethnic groups that’s really difficult actually, and there are different things that have been done, but some raise ethical issues. The most robust way you could do it generally would be focused enumeration, where you randomly select addresses but you get the interviewer to ask about the addresses on either side. So, they’re basically screening say five addresses for every one address they visit to try and identify whether there’s anyone from a minority ethnic background who lives there. But obviously that then raises questions, because the interviewer is asking questions about your neighbour's ethnicity. So that has to be done quite carefully. There are studies that have done it based on surname, which raises similar kinds of ethical issues because you're making assumptions about people's ethnicity based on their surname, which is obviously quite a crude tool. You can use existing sample frames, such as online panels that will have data about people's ethnicity, to sample based on ethnicity, but obviously that's not going to be a completely pure probability sample, because they're opt-in panels basically but they are ethical given they've provided that information voluntarily. You can argue that that might actually be a more ethical way of doing it than screening a surname. If you want to do it well – it's really expensive and quite ethically difficult.

If you had unlimited resources, what would be your ideal approach?

I would do focused enumeration, because if you really want to get an estimate of a sort of population level, I think all of the other approaches will just leave out too many people. But I would do focused enumeration where you probably start by sampling say census output areas that are known to have a higher proportion. So you oversample areas where you know there are more people from minority ethnic backgrounds living there. But do it so that you still have some areas where actually there aren't that many people and you have to try a bit harder to find them.

Are there any approaches for sampling small, specific groups that you'd say are absolute no-gos?

While you might use it in qualitative sampling . . ., but if you're snowballing from people, asking respondents, 'Do you know anyone else who fits in this group', you're going to get just a very homogeneous sample which probably isn't particularly representative. That said, a lot of charities will use that kind of approach. They will email out to their database of people who support them or are engaged with them in some way. Say a young persons' advocacy group might do that to all the people on their database and then that would be their survey of kind of 16- to 24-year-olds, which I don't think I would say that it was a no go, I think I would say it's fine, but you just have to present it as this is a survey of 16- to 24-year-olds who are in contact with your organisation, which means that they are likely to be slightly more kind of activist, slightly more concerned about the sorts of issues that they're being asked about already. So that's not necessarily representative of all 16- to 24-year-olds, some of whom might be much less bothered about some of the issues that they're being asked about . . . [These organisations often] say, 'we asked 16- to 24-year-olds'. But I would say, 'no you didn't, you asked 16- to 24-year-olds on your contact database who are likely to be quite different from the population of all 16- to 24-year-olds in ways that might be quite significant.'

Fundamentally, defining our target population carefully and identifying an appropriately corresponding sampling frame are crucial initial steps for any survey research project before we can begin the process of sampling. While it may appear straightforward initially, these decisions are not trivial and require careful attention in order to assure that the analyses conducted actually allow the researcher to make statements about the groups of people, institutions, organisations or transactions that they are interested in.

Probability and non-probability sampling

We distinguish two main approaches to sampling, probability and non-probability. In survey research, probability approaches are classically seen as the most desirable way of achieving a representative sample. Fundamentally, they provide a sampling mechanism in which each member of the population has a specific and known probability of being selected into the sample. This allows us to design sampling processes in which the likelihoods of any particular sample composition can be estimated, and therefore, we can calculate the probability of our sample results being representative of the population as a whole, although in the sampling process, we do not need to take account of any characteristics of the particular individuals being selected. The most direct applications of this approach are random sampling techniques, which will be discussed below and which will enable us to understand this rather abstract principle more clearly. However, such techniques cannot always be applied directly, mainly because of the lack of suitable sampling frames, which is why we will also consider other techniques (specifically multistage, stratified and cluster-sampling approaches). There are also sampling techniques that do not follow a probability-based design and instead rely on other techniques. In particular, in certain areas of polling work, non-probability sampling is often applied, usually through some form of quota sampling. In that approach, a researcher aims to actively mimic the population by measuring a range of characteristics within the sample to try to maximise their similarity to the population overall. Below, we will discuss this in more detail and review their advantages and disadvantages.

Approaches to probability sampling

Simple random sampling

Kalton (1983) usefully notes,

> Simple random sampling (SRS) provides a natural starting point for a discussion of probability sampling methods, not because it is widely used – it is not – but because it is the simplest method and it underlies many of the more complex methods. (p. 9)

In other words, simple random sampling represents the conceptually most straightforward approach to using probability-based approaches in obtaining a representative sample. However, as we will see, it is very difficult to execute in practice in many instances, which is why we use its logic as the foundation for other techniques.

Simple random sampling is rather straightforward indeed. Starting with our target population, we define a certain sample size, and we will select that number of respondents for our survey from the population through a fully randomised procedure. In practice, this means crucially that each individual within the target population has the exact same chance of being selected into our sample. If that can be achieved, we are in a very strong position. By virtue of randomisation, we could expect that our sample, as it increases in size, would become more and more similar in its distribution to that of the population. Think about a simple example of a coin toss, where you have two possible outcomes, heads or tails. Each has an equal chance of 50% to occur. If you toss the coin only a few times, it is fairly likely that the distribution may be skewed towards more heads or more tails, but if you flip the coin very often (and nobody tampered with the coin), say 1000 times, you would expect that overall the imbalances would roughly even out and you would get a result overall where just less than or more than 500 tosses would result in heads and tails, respectively. So without actively manipulating your sample of coin tosses (your target population here being the infinitely many coin tosses that could be undertaken) in any other way, merely by randomisation, you would be able to achieve a sample that would likely be close to the distribution of the population overall (50/50). While it might not be perfect, because of the known probabilities, we are actually able to calculate the likelihood of our sample results being equivalent to the results in the population. The logic applied is that of inference – however, we are not dealing with that in this volume. It is covered extensively in its own volume (Volume 3) as part of this series.

So why then do we not simply use simple random sampling methods all the time, if they seem so ideal and straightforward? They can even be easy to implement: imagine you wanted to generate a random sample of 100 of all the members of the UK Parliament (House of Commons and House of Lords combined). All you would have to do was to get the list of all their names (which is publicly available) and create a random selection mechanism. You could do this (quite tediously) in a manual fashion, for example, by writing each name on an equally sized sheet of paper, mixing those up well in a closed box and picking out 100 of the sheets of paper. More commonly nowadays, you would probably use a computer program that would replicate this process of random selection from your list. But the only reason you can do this in the first place is that the list of all members of that target population is actually easily available. For it to work, you need to have a complete sampling frame. That, however, is unfortunately often not the case when we want to undertake social surveys. If you wanted to conduct an attitudes survey of all adults in a particular country, you would already begin to struggle. Even if there is a register of every person in a country with their home address and that register was kept up to date consistently, it is very unlikely that the state institution holding that register would permit anyone to access it for their own research needs. But without a sampling frame, you cannot apply any randomised selection mechanism in practice, because you simply do not even know the names of all the individuals who form your target population (so you could not make a set of sheets to put in a, in this case very large, box to randomly pick from).

This has an important practical implication for implementing random sampling-based techniques. We need to make sure that the respondents who were randomly selected are indeed the ones that are asked the questions of a particular survey. Imagine, for example, that we had a sampling frame with the telephone numbers of the target population. We would call a randomly selected number of people on the list. But not all respondents will be available the moment we first dial the number. Should we then simply move on and call someone else? No, we should not! If we did that, we may create biases in our sample (maybe some groups of the population are more available to be reached over the phone, for example). Instead, we should make repeated attempts to contact the originally selected person to make sure that our random sampling approach properly works (we will discuss how to properly conduct the data collection in more detail in Chapter 3). So even when we have a clear sampling frame, random sampling requires a lot of effort in practice.

Often, however, we do not have a straightforward sampling frame at all. In those instances, we need to apply other methods to still obtain probability-based samples. We will discuss some of the most prominent approaches below.

Cluster and multistage sampling

Often, we may not have the information for all potential respondents in our target population for a survey; however, we may have information about aggregations of these respondents that we can utilise. In multistage and cluster sampling approaches, we may not be able to create a sample from the full list of our population, but we may be able to generate a random sample of groups of our respondents as a first step. A very common example where this approach is applied to is in relation to studying school students. Even if there are registers for all school students, researchers would be unlikely to be granted access to those because of data protection. However, a list of all schools in a particular region or country of interest may very well exist, thus providing a researcher with a route to comprehensively identifying all possible locations where the ultimate members of the target population (school students in this case) could be found.

What we effectively do is divide our total target population in several defined sub-clusters. These clusters have to fulfil certain conditions (Arnab, 2017, p. 409). They should be comprehensive (i.e. they should cover all respondents from the population – in our case, all students in a particular country) and they have to be mutually exclusive (i.e. each respondent can only be part of one cluster at a given time – in our case, this means each student should only be enrolled in one school). If these conditions are fulfilled, we can then draw a sample, following the approach outlined above from random sampling techniques, of all schools in the country in the first instance. Subsequently, we could then include all the students within each cluster in our sample. In a sense, we have moved the randomisation one aggregation level up. We obviously need to sample a sufficiently large number of schools, as schools themselves will differ with regard to the composition of their students, but the general logic of randomisation still applies and we can continue to work within a probabilistic framework that permits us to use inferential logics – albeit having to take into account some caveats regarding the calculation of estimates following the different sampling process (Kalton, 1983, pp. 29–38) and being able to explicitly investigate not just individual-level effects but also differences between clusters (which can, in the first instance, form part of the research question).

However, it may not always be feasible to include all respondents within all clusters within our sample in the final survey. If we have selected a large number of schools, for example, the total number of students may be very large and beyond the scope required for the analyses in a project. It may also be organisationally difficult to organise written consent from parents as part of the process for everyone, while it could be more feasible to arrange for these things to be done at the class level rather than overall school level. So we may have more stages of identifying sub-clusters or indeed only select a random sample of respondents within the final clusters we have identified. While the terms are sometimes mixed in the literature and generally referred to as cluster sampling (Kalton, 1983, p. 29), to be precise, we can differentiate between cluster sampling in the narrow sense and multistage sampling. While *cluster sampling* specifically refers to the sampling of clusters and then the selection of all individuals within those clusters selected, *multistage sampling* permits that within selected clusters a further sampling process takes place that will result in only subsamples of the initial clusters being selected (Arnab, 2017, p. 423). Crucially, at each stage the sampling processes should follow randomised processes as closely as possible, so that the probabilistic approach can be retained and inferential logic applied.

Subsequent analyses that utilise data from multistage sampling designs need to take the complex structure of the data into account carefully. As stated above, there may be systematic differences between our clusters in the first place. For example, some schools may have more students from higher income families than others. Therefore, the individual respondents (here the students) are not fully independent of each other. This applies at all levels of the sampling. The performance of students in a standardised test may, for example, be affected by the quality of the teaching they have been exposed to, so there may be group effects when comparing one class to another, because they were taught by different teachers. So in analyses of survey data originating from cluster or multistage sampling, we need to take into account these complexities and potential clustering patterns that are a consequence of subsamples of respondents being somehow connected to each other through shared experiences. One common approach that utilises the complexity of information originating from such samples is multilevel modelling – a technique discussed extensively in Volume 8 of this series.

Stratified sampling

Cluster sampling approaches can be particularly suitable when we try to recruit respondents within a clearly defined group of the population. But even when we are trying to engage with much broader populations (e.g. all adults in a country), it can be very helpful to break down our analysis into certain subgroups. Quite often, we may actually have some knowledge about the composition of our target population and different constituent parts. It may be easier, for example, to undertake sampling processes within each of a set of administrative regions or distinguish the ethnicities that people may have, in particular if our research is about comparisons between different groups of a certain characteristic (e.g. geography or ethnicity). Stratification allows the researcher to define the subpopulations or *strata* to then draw a sample within each of these strata separately. This can be very important when we expect, for example, that a whole-population approach to random sampling may actually result in biased samples, because response rates may not be uniform across different groups. People in a particular region or who have a particular ethnicity may be more or less likely to take part in surveys, for example. Stratification can ensure that the sample size for each relevant group is therefore completely reached. The most crucial point then is that sample sizes for subgroups are not left to randomisation but are controlled by the researcher (Kalton, 1983, p. 20).

Researchers can choose to match the sample sizes proportionately to the distribution of the different strata in the target population (referred to as proportionate stratification). However, sometimes, researchers may intentionally deviate from this to increase the sample size of a particular strata beyond the number you would expect from whole-population random sampling (disproportionate stratification). This is commonly done when the research question asks for the engagement with a group that is relatively small in terms of the whole population and if left to random sampling, we would only get a very small sample of them that we could analyse – potentially too small for the investigations we would want to undertake. Therefore, we may oversample such groups intentionally, and stratification-based approaches allow us to do this. Apart from this strength, the ability to estimate characteristics of distinct subpopulations, Arnab (2017, p. 214) identifies several other advantages of stratification: administrative convenience, in making the sampling more manageable through applications to subgroups; improvements in the representativeness of the sample, in particular if certain subgroups may otherwise be underrepresented, due to, for example, differences in response rates, as mentioned above; efficiency in the estimation of group characteristics under scrutiny; and improved data quality if, for example, different investigators can carry out the data collection for different subgroups based on the language they speak.

Approaches to non-probability sampling

While probability sampling methods have many advantages, as outlined above, in particular as they are underpinned by a theoretical approach that focusses on minimising biases in the selection of cases and thus permitting the estimation of our confidence in the strength of the relationship between the sample and our target population, these approaches also have a major shortcoming: when focussed on large populations, such as people residing in a city or even a whole country, they are very expensive to administer. We will look at the different modes of practically collecting the data in Chapter 3, but to implement randomisation procedures for thousands or even millions of people is very labour-intensive, onerous and can take a long time. However, researchers do not always have such extensive resources or the time to undertake the work. For example, if a newspaper wants to conduct a quick opinion poll to find out whether people like a newly elected party leader, they need a response within a few days to report on it in a timely fashion and cannot wait for, what could sometimes be several months to undertake the data collection for certain types of extensive probability sample–based surveys (as we will see in Chapter 5). Non-probability sampling, in particular quota sampling, is therefore a rather commonly used method, especially in market and certain types of polling research. We will discuss below how it works and engage with the particular problems that need to be considered when employing it, before also briefly considering further alternative forms of non-probability sampling.

Quota sampling

The basic idea behind quota sampling is explained fairly straightforwardly. It takes our general starting point that we want to generate a sample that mirrors our target population well literally and tries to proactively recruit respondents that match the characteristics of the population. The process begins by deciding what characteristics most crucially describe the population we want to study. For example, in any sample of the adult population of the country, we would all quickly agree that it would have to contain both men and women, and it would also have to cover the different age groups of the population. Commonly (unless, again, one wanted to actively oversample a group), a researcher would try to design the structure of the sample according to the proportions in the population (Kalton, 1983, p. 92). So say, for example, that 51% of the adult population were female and 49% were male, the researcher would decide to use the same proportions as the targets for their sample as well. So if the sample had 1000 respondents, the aim would be to recruit 510 women and 490 men into the sample. Similarly, quotas for age would be set as well and the complexity of the quotas could be increased further through linkage. If they do not get linked, there would be a danger that, for example, we could get the right number of men and women, respectively, and the right number for each age group, but nearly all the women were in younger age groups and nearly all the men in older age groups – which would make the sample very unrepresentative in a substantive sense, of course.

Consider the following example. In Table 2.1 you see the distribution for sex and age in Poland based on 2016 Eurostat data. If we simply recruited respondents and made sure that each quota was filled, we would not be able to control in any way whether the sex distribution applied actually within each age group. So, assuming that we did not apply any other measures to address this, we could theoretically end up with a heavily distorted set of recruited respondents that would nevertheless be in compliance with our initial quotas. Therefore, what we would need to do is calculate so-called cross-quotas, which in this case, for simplicity, apply the sex ratio to each of the age groups, thus giving us much more fine-grained quotas. This would enable us to create a sample that in relation to sex and age would look very similar to the population overall. A researcher recruiting respondents would begin any interview by screening respondents for their age and sex, and if they fell into a category where the target was reached already, the survey would not be administered for them any more.

Demographic group	National population distribution	Raw quota for the sample (total = 1000)	Sex	Age	One possible scenario based on raw quotas only	Estimation of cross-quotas
Male	48%	480	Male	18–25	20	62
Female	52%	520	26–34	30	86	
Age (years)						
18–25	13%	130	35–44	90	82	
26–34	18%	180	45–54	100	77	
35–44	17%	170	55–64	110	86	

45–54	16%	160	65+	130	86	
55–64	18%	180	Female	18–25	110	68
65+	18%	180	26–34	150	94	
	35–44	80	88			
45–54	60	83				
55–64	70	94				
65+	50	94				

In many ways, quota sampling is quite intuitive and is very straightforward in its administration. So why do we not simply use it all the time, when probability sampling is so resource-intensive? It is because we do not benefit from the main advantage of probability sampling. Probabilistic methods, as shown earlier, allow us to obtain a sample that is representative of the population, because of the randomisation techniques we apply. We do not design the sample to have particular characteristics, the sample develops those characteristics (which are similar to the population), because of the technique (unless there are biases, e.g., in response rates, which we return to in Chapter 3, but which could affect any sampling method). When we apply quota samples, we only mirror the population distribution according to the quota characteristics, which we have decided upon as being relevant. However, we do not know whether there are other relevant distributional factors, which we are not taking into account.

Consider our example above again. While the distribution of sex and age would be identical to that in the population overall, we do not know what sorts of men and women, respectively, at each age group we recruited into our sample. Societies are stratified according to income and socio-economic class, for example. It is possible that our sample may include respondents who are disproportionately wealthy across all categories, or it may be unevenly biased, including men who are on average less wealthy than men in the population and women who tend to be more wealthy than the average we would expect to see. In order to ensure that this was not the case, we would also have to add quotas for these characteristics and supplement them potentially with further cross-quotas. But where should we stop? Do we need quotas for education, housing tenure, marital status, the number of children and religious affiliation, to name just a few? Crucially, many of these characteristics are related to each other. If we know a person's age, sex, income, education and housing tenure, we actually have a lot of information that permits us to predict a wide range of other socio-economic variables fairly well. But it fundamentally depends on the variables of interest we want to analyse in our survey. If there are particular characteristics that are strongly linked to those variables that are at the core of our analyses, it would be particularly pertinent to ensure that those are reflected well in the sampling distribution. However, this can be very difficult. If, for example, you are interested in analysing vote choice in an election, there are many factors that influence the outcome and how they influence it may change over time, as well.

It means quota sampling always has to rely on a degree of 'subjective evaluation' (Kalton, 1983, p. 92) that is not required in probability sampling methods. The problem could be observed, for example, in the 2015 UK general election, where most polls predicted a hung parliament requiring coalition government to continue, while in fact the Conservative Party ended up winning a majority of seats in the House of Commons. While several issues contributed to the incorrect estimations by most polling companies before the election, an inquiry by the British Polling Council and the Market Research Society found that the dominant reason leading to the failure was 'unrepresentative samples' (Sturgis et al., 2016, p. 4). They found that, crucially, support for the Labour Party was systematically overestimated, because groups that supported the Labour Party were over-represented following the application of the quotas used. It demonstrates the difficulty in applying quota sampling designs to topics of interest that are influenced by many factors, but it is unknown how precisely those factors actually are distributed in relation to the variable of interest in the population.

Importantly, this does not mean that quota sampling will always lead to problematic outcomes. Several researchers have demonstrated that indeed quota samples can be useful in particular contexts and if a lot of detailed work goes into the construction of the quotas and their interrelations as well as the design of the actual data collection (which we discuss in Chapter 3). Cumming (1990), for example, found, when comparing results from the administration of a survey through both quota and probability sampling, that differences between the two samples were either insignificant or substantially fairly small with regard to indicators on health promotion campaigns it engaged with. Vidal Díaz de Rada and Valentín Martínez Martín (2014) found that non-probability samples performed well or even slightly better than probability samples on socio-demographic indicators that quotas would typically be built around (e.g. as age and education), as those are specifically targeted. However, they also noted that probability samples achieved better results in secondary concepts, such as unemployment, which appears to be in line with the discussion presented above: when concepts cannot be easily attached to a small number of easily defined quota indicators, it becomes very hard to know whether the quotas chosen will actually result in a representative sample. This does not only apply to political attitudes but also other comprehensive concepts in the social sciences. Yang and Banamah (2014), for example, demonstrate the issues of using non-probability sampling in surveys engaging with social capital concepts. So while quota sampling can be useful, a lot of caution needs to be used in its administration and careful consideration applied to the specific topic and context regarding its suitability for such an approach that is prone to more biases than probability-based techniques.

Common alternative non-probability sampling methods

While quota sampling is indeed utilised quite extensively in survey work, other forms of non-probability sampling commonly are not, as they do not tend to aim at obtaining representative samples in the same way quota samples do. Most of these techniques are used more commonly for other research methods, where representativeness in a strict sense is less of a concern. However, it is worth briefly reviewing them, partially to also understand the contrast between them and the more suitable approaches we discussed above.

The simplest approach we could consider is called a *convenience sampling*. It is very easy to undertake and straightforward with a focus on quick delivery and, as the name suggests, convenience. The most illustrative image that this approach invokes is that of the interviewer standing on some street with a clipboard in their hand, simply interviewing whichever person comes up next to them, before moving on to the next person that comes along. Unless, of course, the population is that of all people who walk along that street specifically, the data collected from such a process would permit for very little in terms of generalisability or representativeness, as the sort of people who walk down a particular street will have certain distinct characteristics (e.g. they may work or live specifically in that area). There are analogous forms online, where websites may ask users on that site to fill in a survey. This might be a suitable idea for a company that wants to find out something about the actual users of the website, but if those questions were, for example, on political attitudes, while easy to collect, there would surely be a bias of what users would even look at that particular website, and therefore, we could not generalise to any meaningful population. Convenience samples therefore are rarely suitable for proper survey research; effectively they simply include the respondents a researcher can get easily.

A slightly more targeted approach would be *purposive sampling*. In this case, the researcher would have a clear target group or population in mind but would recruit respondents without the use of a probability or quota-based approach. This could often be the case, because there is no explicit sampling frame that could be utilised or the group under consideration is fairly narrow (e.g. executives in a particular type of industry in a certain region). The approach may be chosen, because researchers do not plan to generalise beyond the sample of individuals surveyed. Indeed, the respondents surveyed may be the target population, if it is small. In its own right, the findings based on such samples only have a rather narrow scope. However, they could be meaningful if, for example, used in conjunction with qualitative interviews of the individuals surveyed to obtain additional information from them or to gain basic data on a larger sample of individuals within the target group from whom interviewees are selected based on certain criteria that information was gathered for in the survey.

Another approach that is commonly used for small, narrowly defined groups is *snowball sampling*. This is a way to undertake sampling that is often used when it is very difficult to reach subjects, for example, because they are part of a small and vulnerable group or because they undertake illegal activities. Snowball sampling implies that the researcher will utilise their respondents in order to gain further respondents. The assumption is that those who fulfil certain characteristics that make them part of the target group for the sample might be able to generate contacts with others who also fulfil those characteristics. So, the sample grows continuously with the help of the research participants. Obviously, the nature of this approach means that, again, we are not able to generalise from the findings in this sample to any larger population in a statistical sense. Also, typically snowball samples tend to be rather smaller, which is why the approach is more commonly used in qualitative methods.

Respondent-driven sampling

Because of the limitations posed by non-probability methods that are not based on quota designs, surveying hard-to-reach groups in a way that allows us to make meaningful statements with the aim to generalise can be very difficult and is often seen as prohibitively difficult. However, researchers have developed techniques that aim to overcome some of these problems. In particular, respondent-driven sampling (RDS), initially developed by Heckathorn (1997), warrants some discussion. Typical snowball sampling, as discussed above, is not suitable for survey research at a larger scale if the aim is to ultimately make statements about a specific target population rather than just the group of specific interviewees. The two key problems encountered are the biases emerging from the initial selection of informants (usually a non-probability purposive or convenience sample) and their referrals to other potential participants. In hard-to-reach groups, there may be significant reluctance to pass on the details of contacts to a researcher.

Heckathorn (1997) developed an approach to address these issues practically. RDS begins similar to other non-probability approaches with a relatively small selection of initial respondents ('seeds') who are recruited on a convenience basis. They are offered some financial compensation for their participation in the survey interview. Instead of then simply asking them about other possible participations though, they are then asked to actively help recruit further participants and they are rewarded for those efforts additionally. This dual approach to the sampling and the focussed enumeration are key to RDS approaches (and reflect some of the suggestions raised by Rachel Ormston earlier in this chapter). Initial respondents are typically given a limited number of 'coupons' they are asked to pass on to other potential respondents inviting them to take part in an interview as well. When those respondents then take part in the interview themselves, not only do they receive a financial reward, but also the person who gave them the coupon does. So participants are compensated both for their participation in the interview and successful further referrals. Each new participant is presented with the same model and therefore recruits further potential participants. That way the bias inherent in the original selection of seed respondents is reduced, as the networks in the target population are broadened and the connections between initial and final participants become more remote.

Using RDS, Salganik and Heckathorn (2004) develop techniques for how to calculate proper estimates about the target population in a meaningful way. Furthermore, they show that under certain conditions the estimates are unbiased regardless of how the initial respondents were selected. However, while RDS provides an important methodological innovation, the conditions made have been shown to be rather strong in terms of their potential impact on the results. Gile and Handcock (2010) show that the positive evaluations of the researchers who developed RDS heavily depend on the applicability of the assumptions made in their models and that those assumptions often are not realistic. Nevertheless, they consider the approach important and useful but suggest that it could be developed further to reduce existing biases. In particular, they emphasise that the selection or 'seeds' should be carefully designed and the behaviour of further recruitment be carefully monitored and adjusted through a significant number of waves of recruitment (for which they make specific suggestions for the implementation). So while RDS requires a lot of careful considerations to be addressed, it can provide a potential avenue for survey researchers who want to overcome problems of studying hard-to-reach groups for whom classic probability or quota designs would not be feasible.

Box 2.3: Ask an Expert

John Curtice | Surveys, polls and exit polls

We often hear about surveys and polls. How would you distinguish between the two?

Well in truth, the distinction between a poll and a survey is partly in the eyes of the beholder. They're both attempts to try to interview a sample of people with a view to getting them to represent a wider population. A poll will usually tend to have, certainly in the UK, two characteristics that would distinguish it from what we might call a survey. The first is the sampling design. The word survey is usually applied to projects that in some way or other can be regarded as using a probability-based sampling design. So in some way or other there is a known probability of each member of the relevant population to be included in the sample, and indeed for all or at least virtual members of the population that known probability is greater than zero. The second distinction I think I would make is between exercises which are conducted over a long period of time and which therefore are able to try to maximise the response rates. And indeed, there is something where response rates matter. So there is a preselected sample of at least addresses if not individuals which you're going to interview. And it's only those that you interview and that's the end of it. A poll in contrast tends to be done over a short period of time, and it isn't always the case that there is some predefined population of people you're going to contact. That is, for example, if a poll is done by telephone, a pollster will often keep on polling until they've eventually got the thousand-person sample of what it is they want to do. But in any event, it's usually done over a pretty short period of time. And so therefore, it tends to be confined to those people who are (a) willing and (b) available in a relatively short period of time. So those are probably the two crucial distinctions that one might make. But in truth, these are both animals of the same kind. And in part, academic survey researchers like to call what they do – surveys. And commercial pollsters are happy for what they're called to be polls but some people will indeed attempt to call what I would call a poll a survey. And maybe even occasionally vice versa.

On election nights, we encounter so-called exit polls conducted on the day. How do they work?

All exit polls are incredibly geographically clustered samples, because they have to be done at (a sample of) polling stations. The first crucial insight on which the exit poll operates is to say, hang on, it may be true, and we know it's true that support for Conservative, Labour and Liberal Democrat, SNP whatever, varies dramatically from one constituency to another. But the change in party support doesn't vary as much. Therefore, any sample of polling stations, however selected, has a better chance of estimating the change in support than the level of support. So the exit poll in the UK doesn't attempt to estimate levels, it tends to estimate change. Now that still leaves you with a question, 'but hang on, how can you estimate change if you don't know what happened last time?' The answer is the exit poll is conducted, wherever possible, at the same polling stations as last time. We pray the polling station boundaries have not been changed and you have to adjust on occasion. And therefore, we . . . get 140 estimates (which is the number of polling stations we sample) of change for each party, well 10 in the case of SNP. Some other advantages of this approach are that if Conservative voters, for example, are less willing to talk to an exit poll than Labour voters it doesn't matter so long as it's constant between the two elections. If postal voters vote differently it doesn't matter so long as the difference is constant across both elections, and so on. So although there is bias, as it is constant this method works. So that's what we do, and the other crucial bit of the exercise is you're then modelling those 140 estimates looking to see if you can identify how the change in support varies according to the known characteristics of constituencies. And you're trying to do that in order to improve the predicted seat outcomes, which are not one/zero estimates. So if, for example, a model ends up saying, 'well we think Labour is going to get 45 and the Tories get 44', we don't simply say that the Labour party is going to win that seat. We see, 'well there's probably about a 52% chance that Labour will win that seat and about a 48% chance the Tories will win', and we sum the probabilities. Now that matters if you've got a skewed distribution. If you've got a situation where maybe one party has got a rather high number of seats which we think it's going to win by a small majority and the other party has got a rather small proportion [of such seats], it's almost undoubtedly the case if you go for a one/zero calculation you would overestimate the number of seats that the party that's got lots of 51, 52% chances of winning would actually win [compared to adding up the probabilities across seats]. And that, on occasion, has made a difference and that's the secret.

Chapter Summary

- Survey research is featured frequently in media outlets, employed by politicians to back up arguments and cited in scientific enquiries on a range of issues. We encounter surveys, polls and the peculiar exit polls on election nights (and John Curtice discusses the differences between those terms).
- To ensure that the results from these investigations are meaningful and representative of the groups they are meant to study, it is crucial that we can appraise the quality of the samples that underpin those surveys.
- To maximise representativeness, probability sampling techniques theoretically are the best approaches we could choose. However, their feasibility depends on the availability of good sampling frames and sufficient resources. While we can adapt perfect random sampling techniques through cluster, multilevel and stratified sampling approaches to overcome some problems, there are situations in which administering them can be prohibitively difficult.
- Non-probability sampling techniques face several problems and difficulties, but in particular quota-sampling approaches can at times help us to gain meaningful insights indeed, and respondent-driven designs may help in studying hard-to-reach groups. However, they require very careful consideration and analyses, as the researcher has to make a wide range of choices about the sampling design that can have a very significant impact on the results.
- Fundamentally, any approach chosen should always be transparently described alongside the results of a study, so that the potential limitations can be understood well and readers can examine how substantial the scope of the investigation really is and what group of people may or may not be represented by those who have been included in a particular survey.

Further Reading

Arnab, R. (2017). *Survey sampling theory and applications.* Academic Press.

This is a text for anyone interested in more advanced sampling theory and more formal ways of engaging with it mathematically.

Squire, P. (1988). Why the 1936 *Literary Digest* poll failed. *Public Opinion Quarterly, 52*(1), 125–133. https://doi.org/10.1086/269085

This is an article that provides more details about the case study used in the chapter and discusses why the classic, non-probability-based poll in the *Literary Digest* got the 1936 US presidential election wrong, although it had such a large sample size.

Sturgis, P., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Kuha, J., Lauderdale, B., & Smith, P. (2016). *Report of the inquiry into the 2015 British general election opinion polls.* British Polling Council and Market Research Society.

The 2015 general election in the UK found many people surprised by the results, as the results on the election day differed substantially from many pre-election polls. This report discusses in detail and in a rather accessible manner what happened and acts as a good case study.

https://doi.org/10.4135/9781529682793