SAMPLING



Sampling is about choosing who or what we wish to study in order to answer our research question. Sometimes these units of study are chosen randomly (more typically for quantitative research) and some-times they are purposively selected (more typically for qualitative research). In any case the way we choose the units of study will have a major impact on our results and should, therefore, be done thought-fully and with a clear rationale.

**Q**UANTITATIVE AND QUALITATIVE RESEARCHSAMPLES

It is important to note that quantitative researchers use probability sampling while qualitative researchers will more often opt for non-probability sampling. Quantitative researchers require a random or probability sample for the statistical processes they undertake (see chapter 14). Qualitative researchers, on the other hand, use non-probability sampling because of the nature of their research which is largely exploratory. Researcher bias is unavoidable but the process allows a significant sample to be generated quickly.

Qualitative researchers often seek typical cases. Sometimes, how-ever, they deliberately seek atypical, extreme or deviant cases, as a means of shedding light on the typical or for their own intrinsic

interest. Often sampling in qualitative research is controlled not by a need for statistical rigour but by the developing theoretical argument, and so the sampling paradigm is sometimes referred to as *theoretical* *sampling*. In other words, the sample is chosen to assist the researcherto understand the phenomena under study and to illuminate the researcher’s emerging theory. Because the type of research under-taken by quantitative and qualitative researchers is different, neither probability nor non-probability sampling is superior to the others, or necessarily more effective.

 **Quantitative versus qualitative sampling**

*Quantitative* *Qualitative*

Probability sampling  Non-probability sampling

Objectivity  Subjectivity

Representative  Non-representative

Results generalisable  Results not generalisable

Statistically rigorous  Not statistically rigorous

Random sampling  Theoretical sampling

Claims no researcher bias  Researcher is integral to

sample selection

All units equal or known  No attempt to give units an

chance of selection equal chance of selection

**W**HAT IS A SAMPLE?

Thus once we have decided on our research question and other aspects of our research design, some thought must be given to the subjects, cases or events we wish to study. In order to be completely accurate, every person (or case or event) in the population under study would need to be surveyed. This is called a *saturation sample* or *census*. In practice, saturation sampling is rarely possible because of the nature of the study question, the size of the population and because of time and resource constraints. Instead, we select a *sample* from the total population under study. In quantitative research, this sample is ideally chosen at random. As a result, researchers can claim that results are generalisable because inferences can be drawn from the research about the wider population (that is, the results from the sample are representative of results in the population).

For a sample to be representative, it must be chosen in such a way that subjects or cases have beliefs, attitudes or experiences which are similar to the population being investigated. Conventional social science wisdom suggests that a random sample drawn from the study population and selected according to sampling theory will reflect the

characteristics of the entire population. In this way, the researcher is able to make informed statements about the group under study.

Let us look at an example to allow us to better understand the concept of sampling. Imagine yourself an employee of a government income-support organisation that provides payments for people who are unemployed, ill or unable to work due to disabilities. It also pays pensions to people of retirement age and income supplements to people caring for others full time, or for children with disabilities.

You are interested in examining the attitudes and experiences of clients coming to your agency for the first time. You are concerned that clients may find the experience humiliating and confronting. Your ultimate aim with such a study might be to make this experience easier for clients. To study such an issue affecting a very large group you must draw a sample. First-time clients will represent at least several thousand people each week. Ideally you would study all first-time clients in every office of the income-support organisation across the country over a specific period of time. Pragmatically, you might choose a sample that is localised (first-time clients in your own office) and time-defined (in a one-week period). Such a sample may allow you to generalise beyond the subjects of the study but with a large degree of caution.

No doubt you have noted that drawing such a sample will have serious limitations. Can you generalise beyond the local area? Do your findings reflect the experiences of clients in branches in other locality areas? Will clients in small communities find the experience more difficult than those in larger, more impersonal offices? By confining your time frame to one week, are you picking up on seasonal variations in the workforce? Are recent school leavers, for example, a part of your sample?

As you can see from this example, drawing a sample that accu-rately reflects the study population requires a great deal of thought. Because of pragmatic concerns such as time and resource constraints, you may have to make trade-offs that reduce the generalisability of your findings. This is not to say that you should not undertake the proposed study. Your findings, and particularly any strategies you develop which improve client service, will be useful to workers in offices across the country.

Let us examine some of the most common types of sampling. First, you should note that sampling can be categorised as either *probability* or *non-probability* sampling.

**P**ROBABILITY SAMPLING

Probability sampling refers to sampling in which each unit of the population has an equal (as in the case of simple random sampling)

or known chance of being selected for study. The units of the popu-lation under study are referred to as the *sampling frame*. The sample is chosen from the sampling frame in an unbiased and rigorous way, allowing a small sample to be used to assess or predict the studied behaviour in the larger population. Probability sampling is favoured by quantitative researchers as it allows a high degree of repre-sentativeness from which results can be generalised. The four main types of probability sampling are:

1. Simple random sampling
2. Systematic random sampling
3. Stratified random sampling
4. Cluster random sampling

**Probability sampling**

Each population unit has an equal, or known, chance of selection

High degree of representativeness

Allows researchers to generalise results

Favoured by quantitative researchers

Four main types:

Simple random sampling

Systematic random sampling

Stratified random sampling

Cluster random sampling

**Simple random sampling**

Simple random sampling is the most common form of probability sampling. There are several methods we might choose to select a simple random sample.

Let us use another example to illustrate the process of simple random sampling. Imagine you are working on a government policy research team and you are interested in assessing the effects of a new policy to move previously institutionalised people with psychiatric problems into group homes in the community. Your study is moti-vated by a number of violent incidents reported by police and causing community concern. Your sampling frame represents the entire pop-ulation of people who have moved from institutions to group homes in the five years since the policy was adopted. Suppose that you find that there are 10 000 people in your sampling frame and you choose to sample 1000 of them. We might first allot a number to each person on the list from 1 to 10 000. Numbers from 1 to 10 000 are then placed in a container and are selected (without replacement) until 1000 numbers have been drawn. Referring to your list, the numbers are

Research for Social Workers

related to the people in the sampling frame. These 1000 are then used as subjects for your study.

Alternatively, you might use a table of random numbers; these are readily available in many research methods textbooks or can be generated by a computer. Begin at a randomly selected point in the table and choose the first number on the table between 1 and 10 000. Work your way up (or down) the column selecting numbers and relating them to your list of names. In this way you can generate your sample of 1000.

**Table of random numbers**

|  |  |  |
| --- | --- | --- |
| 1986 | 3067 | 1309 |
| 2254 | 5321 | 0532 |
| 4763 | 9854 | 5643 |
| 1589 | 8623 | 2875 |
| 7415 | 9792 | 3261 |

A simple random sample may be chosen in a number of other ways, including by dates of birth or by initials. Any randomised technique is suitable.

**Simple random sampling**

Sampling frame is identified

Desired sample number identified

Numbers assigned to subjects in sampling frame

Random numbers selected in some way

Numbers related to list of subjects

Sample generated

**Systematic random sampling**

Systematic random sampling varies from simple random sampling in that the chosen units are not independent of each other. For instance, you might decide to select every tenth person in order to generate your sample.

The size of the interval between chosen units is decided by dividing the total population, or sampling frame, by the desired sample size. Thus, using our previous example of previously insti-tutionalised people (10 000) and our desired sample (1000) we decide the interval (*x*) in the following way:

1.  total sampling frame desired sampling
	1. 000
2.  1000  10

The first name on our list is chosen at random from those numbered from 1 to 10. Thereafter, we choose every tenth person. We will generate a sample list of 1000 using this method.

When using this method, we need to be careful that no unin-tended bias creeps into the sample. For example, by choosing every tenth name, we might be missing residents in smaller communities and in certain types of group housing. (If the initial list is randomly generated this should not be an issue.) Once your sample is selected, carefully check that you have generated a sample that is not biased on certain variables that may be important to your study.

**Systematic random sampling**

Sampling frame identified

Desired sample number identified

Numbers assigned to each subject in sampling frame

Sampling interval (*x*) identified

First subject randomly chosen

Every *x*th subject chosen

Sample generated

**Stratified random sampling**

Stratified random sampling allows us to divide our sampling frame into various strata or groups before selecting our sample. This allows us to ensure that each group is represented proportionately or dis-proportionately to their numbers in the overall population. For example, we might decide that it is important to assess how the effects of group home living vary by gender. We divide our frame by gender and select a random sample of 500 from each group.

Alternatively we might decide that age is a critical factor. We thus divide our sampling frame by age and select from each group. If we are to select a proportionate sample then we choose a sample from the subgroup that reflects their numbers in the total sampling frame. If we choose to sample disproportionately, then we choose equal numbers from each group regardless of their relative proportions. Table 5.1 illustrates the way we might select a stratified random sample by age.

Whether you choose a proportionate or disproportionate sample depends on how valuable you feel the information from each group might be and what it is you wish to find out. For example, reported incidents of violence among deinstitutionalised people might be related disproportionately to the older age groups, in which case you might feel a disproportionate sample will be more valuable. To choose

Research for Social Workers

**Table 5.1** **Selection of a stratified random sample by age**

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Number in sampling* | *1/10 proportionate* | *Disproportionate* |
| *Strata* | *frame* | *sample* | *sample* |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Age 20–29 | 4 000 | 400 |  | 250 |  | (1/16) |
| Age 30–39 | 2 500 | 250 |  | 250 |  | (1/10) |
| Age 40–49 | 2 000 | 200 |  | 250 |  | (1/8) |
| Age 50 + | 1 500 |  | 150 |  | 250 |  | (1/6) |
|  |  |  |  |  |  |  |  |  |  |  |
|  | 10 000 | 1 000 | 1 000 |  |
|  |  |  |  |  |  |  |  |  |  |  |

a disproportionate sample the sampling frame is divided into age groups. The sampling fraction is used to decide the interval between chosen subjects. For example, in the 20–29-year age group every 16th person is chosen, in the 30–39-year age group every 10th, in the 40–49-age group every 8th and in the 50+ age group every 6th.

A proportionate sample, on the other hand, is selected by choosing every tenth person in each group.

**Stratified random sampling**

Sample frame identified

Desired sample identified

Strata or groups identified

Proportionate or disproportionate sample numbers identified

Sampling interval (*x*) identified for each strata or group

First name in each group randomly selected

Every *x*th person chosen from each group

Sample generated

**Cluster random sampling**

The final type of random sampling to be discussed here is cluster random sampling. Cluster random sampling is generally used when there is no sampling frame available; that is, we do not know who is in the group from which we are sampling as there is no readily available list of subjects. We also use this type of sampling when we are limited by resource constraints. Suppose, for example, that we wish to survey homeless street kids. We do not have access to a convenient list of names from which to draw our sample. What we can do is randomly select certain areas or *clusters* that are relevant to our research problem. For example, we might choose to survey kids in youth refuges in both city and country areas. While we cannot hope to survey all homeless youth in all areas we can choose the areas or *clusters* that represent the sample under study. To reduce the possibility of sampling bias, it is a good idea to increase the number of clusters surveyed. In our homeless youth study we might reduce

bias by surveying in inner-city areas and in outer suburbs with high ethnic populations. We might also survey in regional cities and a selection of country areas, making sure we sample coastal as well as inland regions. Once our clusters or areas are identified, we then randomly choose our desired sample from the current population in the identified youth refuges.

Cluster random sampling allows us to systematically sample a population which is not readily identified. While it allows us to work within budgetary limits and other constraints, it does increase the possibility of sampling error. The only way to minimise this problem is to increase the number of clusters surveyed.

**Cluster random sampling**

Sampling frame unknown

Desired sample identified

Clusters identified

Random sample drawn from clusters

Sample generated

**N**ON-PROBABILITY SAMPLING

The second major category of sampling is non-probability sampling. Non-probability sampling is generally used in exploratory research and by qualitative researchers. It does not make any claims to be representative of the population under study and therefore the generalisability of results is limited. This, however, is not the point of the research.

Non-probability sampling is very useful and justifiable when the researcher is seeking information in a new area and targets subjects or cases who typify the issue to be studied. Suppose, for example, you are working in a respite unit and you wish to examine the pressures facing carers of AIDS patients. Because AIDS is a relatively new disease, little is known about the unique issues facing carers. To explore this new area of investigation you select a small sample (for example, ten) of carers known to you through your position in the respite unit. You will note that these ten cases will not be repre-sentative of all carers of AIDS patients because of the small sample number and because they are limited by locality. However, this sample will give insights into a previously unexplored area and will provide a qualitative researcher with a rich source of data.

The four most common types of non-probability sampling are:

1 Accidental (sometimes called convenience or availability) sampling

1. Quota sampling
2. Purposive sampling
3. Snowball sampling

**Non-probability sampling**

Each population unit does not have an equal chance of selection

No claim to be representative

Does not necessarily allow the researcher to generalise results

Favoured by qualitative researchers

Four main types:

Accidental sampling

Quota sampling

Purposive sampling

Snowball sampling

**Accidental sampling**

Accidental sampling is, as the name implies, a sample you chance upon by accident. The sample is convenient or available to you for some reason. The most common form of accidental sampling is standing in a public place such as a supermarket or railway station for a certain period of time and interviewing people who walk by. For workers in the welfare industry an accidental sample might be drawn from a worker’s caseload or from clients coming to the agency. Suppose you are working in a women’s refuge and you wish to understand more about the difficulties women have seeking court protection orders. You might choose a sample from your case records or, alternatively, you might sample all new residents over a two-week period. In either case, your sample is not representative of the entire population of women seeking protection orders but it will give you valuable insights into the legal and court process and allow you to explore the problems facing women in this situation.

**Accidental sampling**

Sample drawn from available or convenient group

Sample reflects the problem being investigated

Number of subjects determined by access and availability

Sample generated

**Quota sampling**

Quota sampling allows us to set quotas for subgroups of our sample. Suppose, for instance, that we believe from our experience in the women’s refuge that women with ethnic backgrounds find it more difficult to obtain protection orders because they are reluctant to approach the court system. We might then choose to study women from a non-English speaking background as well as English speaking women. We might also decide that it is important to survey women with children under five as well as those with no children or older children. We can draw up a matrix that will allow us to categorise women in the study.

|  |  |  |
| --- | --- | --- |
| Children under 5 | Non-English speaking | English speaking |
|  | Background | background |
| Yes | A | B |
| No | C | D |
|  |  |  |

We must now decide how many women in each category will be surveyed. If we decide all categories are important, we may seek equal numbers of subjects for each category. For example, should we seek a sample of 20, our matrix will look like this:

|  |  |  |
| --- | --- | --- |
| Children under 5 | Non-English speaking | English speaking |
|  | Background | background |
| Yes | 5 | 5 |
| No | 5 | 5 |
|  |  |  |

Quota sampling allows us to target certain characteristics that are important to our research problem. Because it is a type of non-probability sampling technique, it does not claim to be representative of the population being studied. While it does allow us to make observations about particular subgroups of the population, these results cannot be generalised with any degree of certainty.

**Quota sampling**

Significant categories determined

Quota determined for each category

Quota selected

Sample selected for each category

Sample generated

**Purposive sampling**

This sampling technique allows us to select the sample for our study for a *purpose*. We may have prior knowledge that indicates that a particular group is important to our study or we select those subjects who we feel are ‘typical’ examples of the issue we wish to study. In our study of women seeking protection orders, we might decide that women from rural areas or outlying metropolitan areas appear to have more serious problems and so we choose a sample from these clients in order to determine why the system appears to be letting them down. Alternatively, we might choose a sample of experienced women who work in refuges in these areas to allow us a different perspective to aid our understanding of the issues involved.

**Purposive sampling**

Sample is chosen for a particular purpose

Sample gives insights into a particular issue related to the study area

Number determined by the research topic, availability Sample generated

Research for Social Workers

**Snowball sampling**

Snowball sampling is used when we have no knowledge of the sampling frame and limited access to subjects who may meet the criteria for our research. Suppose we do not have access to women seeking protection orders as we do not have access to a refuge. However, we do know a woman who has taken out an order. We might approach her for an interview and ask her to nominate other women she might know in the same circumstances. We contact these women for interviews and they nominate further women. We continue collecting our sample in this way until we feel we have reached the stage where our sample is *saturated*. In other words, no new information is emerging from our research and so we determine our sample is complete.

**Snowball sampling**

Contact a ‘typical’ case

Ask this person to recommend further cases Continue until sample is complete and saturated

**H**OW BIG SHOULD MY SAMPLE BE?

One of the most often asked questions is about sample size. Students and beginning researchers often beg for a magic sample size number.

Research, however, is not as simple as that. Sample size depends on what it is we wish to know, how certain we want to be about our findings, the resources we have available, the research design and its purpose, the type of statistical analysis required and the degree of rep-resentativeness we consider desirable (Yegidis, Weinbach & Morrison-Rodriguez 1999). Naturally, to be entirely accurate we would need to survey the entire sampling frame. As this is rarely possible, you should note that, in general, the larger the sample the more accurate will be your findings. Of course, you should also note that a small repre-sentative sample may be more accurate than a large unrepresentative one. It depends on the accuracy of the sampling technique. For small populations you should choose proportionately more for your sample than for samples drawn from larger populations.

With quantitative research, sample size is related to the type of statistical analysis you may wish to undertake. A minimum size for adequate statistical analysis would be 30, although many texts sug-gest your sample should be at least 100 or 120 (see, for example, Williams, Tutty & Grinnell, 1995, p. 233).

It is important that you understand the concept of *sampling error* in relation to sampling in quantitative research. We will explain this concept in detail in chapter 14, once we have explained some basic statistical concepts. For now, it is important to know that sampling error is an estimate of the amount of error you could make if you use just one sample of a certain size from a population to estimate the results for the whole population. In chapter 14 we show you that standard error or sampling error is inversely proportional to the size of your sample—hence you will have more confidence that your sample statistics will be closer to those of the population if you use a larger sample. The general principle is to take the largest repre-sentative sample you can.

Experienced statisticians have developed ways of estimating sample sizes that will give you a fairly accurate result based on the size of the population. One such table is reproduced here (Table 5.2) for your consideration. Note that the ideal sample size varies accord-ing to the confidence level required.

*Confidence level* (or confidence interval) refers to the level ofconfidence we have that the results accurately reflect the views of the population. A 95 per cent confidence level means that our results could occur by chance only 5 times in 100 trials. You should aim to have a 95 per cent confidence level with a 5 per cent standard error, meaning that the results will be out by 5 per cent only 5 times in 100 (Royse 1999).We will discuss confidence intervals in more detail in chapter 14.

Using Table 5.2, you can see that should you wish to sample a population group of 100 with a 95 per cent confidence level and a 5 per cent margin of error, you would need to draw 79 of the

Research for Social Workers

**Table 5.2** **Appropriate sizes of simple random samples for specific permissible errors expressed as absolute proportions when the confidence level is 95 per cent**

*Sample size for permissible error*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Population size | 0.05 | 0.04 | 0.03 | 0.02 | 0.01 |
|  |  |  |  |  |  |
| 100 | 79 | 86 | 91 | 96 | 99 |
| 200 | 132 | 150 | 168 | 185 | 196 |
| 300 | 168 | 200 | 234 | 267 | 291 |
| 400 | 196 | 240 | 291 | 343 | 384 |
| 500 | 217 | 273 | 340 | 414 | 475 |
| 600 | 234 | 300 | 384 | 480 | 565 |
| 700 | 248 | 323 | 423 | 542 | 652 |
| 800 | 260 | 343 | 457 | 600 | 738 |
| 900 | 269 | 360 | 488 | 655 | 823 |
| 1000 | 278 | 375 | 516 | 706 | 906 |
| 2000 | 322 | 462 | 696 | 1091 | 1655 |
| 3000 | 341 | 500 | 787 | 1334 | 2286 |
| 4000 | 350 | 522 | 842 | 1500 | 2824 |
| 5000 | 357 | 536 | 879 | 1622 | 3288 |
| 6000 | 361 | 546 | 906 | 1715 | 3693 |
| 7000 | 364 | 553 | 926 | 1788 | 4049 |
| 8000 | 367 | 556 | 942 | 1847 | 4364 |
| 9000 | 368 | 563 | 954 | 1895 | 4646 |
| 10000 | 370 | 566 | 964 | 1936 | 4899 |
| 15000 | 375 | 577 | 996 | 2070 | 5855 |
| 20000 | 377 | 583 | 1013 | 2144 | 6488 |
| 25000 | 378 | 586 | 1023 | 2191 | 6938 |
| 30000 | 379 | 588 | 1030 | 2223 | 7275 |
| 40000 | 381 | 591 | 1039 | 2265 | 7745 |
| 50000 | 381 | 593 | 1045 | 2291 | 8056 |
| 75000 | 382 | 595 | 1052 | 2327 | 8514 |
| 100000 | 383 | 597 | 1056 | 2345 | 8762 |
| 500000 | 384 | 600 | 1065 | 2390 | 9423 |
| 1000000 | 384 | 600 | 1066 | 2395 | 9513 |
| 2000000 | 384 | 600 | 1067 | 2398 | 9558 |
|  |  |  |  |  |  |

*Note:* This table was calculated for binomial distributions.

*Source:* *Research Methods in Social Work,* 3rd edn*,* David Royse, Nelson-Hall Publishers, Chicago.

members of that group into your research sample. If the population is 1 000 000, your sample size should be 384.

Social work students frequently do qualitative research where sample size is not such a big issue and relates more to convenience and availability. With qualitative research you tend to continue to sample until no new information is emerging. Once you get to the point where you feel you’ve heard it all before you know your sample size is complete.

Size of sample is also guided by the diversity of the population you are studying. If your target group is heterogeneous you will need a larger sample size than if it were homogeneous. For example, if the target group is white, anglo-saxon and middle class you will achieve a high level of accuracy with a small sample. If the target group includes several ethnic groups, and diverse income levels, you will need a much larger sample.

**S**UMMARY

When we conduct a piece of research we are interested in a particular target group or population. It is rarely possible to survey the entire population so, using sampling techniques, we select a sample of this target group. This chapter has examined the types of sampling tech-niques we can use to generate an adequate sample. We have seen how a sample can be developed using either probability or non-probability techniques. Probability sampling is a more precise or non-biased method which allows a representative sample to be selected, and results to be generalised. In order to use probability sampling we should have access to the entire target group—this is called the sampling frame. Four types of probability sampling have been discussed in this chapter. They are simple random sampling, systematic random sampling, stratified random sampling and cluster random sampling.

Non-probability sampling, on the other hand, is less precise and involves researcher bias. It allows the researcher to carefully select the sample for a particular, usually theoretical, purpose. Non-probability sampling techniques are used for exploratory, qualitative and often feminist research. They include accidental sampling, purposive sampling, quota sampling and snowball sampling.

Decisions on sample size, which are particularly critical in quan-titative sampling, are governed by the accuracy we desire and the degree of homogeneity of the group. In general, the larger the sample the more confident we can feel about the generalisability of our findings. However, if the group is homogeneous we can feel confident about limiting sample size.

Non-probability sample sizes are not as critical as they are for probability sampling, as sample size is governed by the emerging data such that a researcher should continue sampling until no new data is emerging.

Research for Social Workers