KnowHow

Learning Resources for the KnowFife Partnership Hub

Knowledge

Learning

Skills

Best Practice

Researchers'Alliance

Survey Sampling

Sampling allows you to draw conclusions about a particular population by examining a part of it. When carrying out a survey, or other information gathering exercise, it is not usually possible to ask questions of all the people that you would want to, unless the numbers involved are very small. Instead, you usually collect information from only a proportion or sub-set of your population, which is usually referred to as a **sample**. This briefing covers some sampling methods that you can use when carrying out surveys.

Who is it for?

Anyone that needs to carry out some form of survey, consultation, feedback or other information gathering work, e.g. customer or employee surveys, evaluations, reviews, etc. involving a sub-set of a target group.

What do you need to know?

The purpose of selecting a sample is to identify a group of people from a population that reflects the same characteristics as that overall population. You can then use the information from the sample to draw conclusions or inference about the population.

It is essential that you clearly define the population at the outset. This is often referred to as the "target population". For example, deciding to survey a sample of males in Fife, you would have to identify whether this would include children as well as adults, people only resident in Fife, or people that work here, people in hospital or in care homes, etc. In many cases, your sample will depend on the availability of the population that you want to include.

1

The sample size

The level of accuracy that you need for your survey will determine the size of your sample. In general, the larger your sample the more accurate your results will be, up to a point. A number of other factors will also determine your sample size –

- Your available budget or any time constraints you may be working towards

- The variability of the characteristics of your target population - if everyone in your target population had the same age, then you would only need to ask one person in your sample in order to calculate the average age of your target population

- Target population size – in most cases the larger your target population, the larger your sample will need to be. However, once your population reaches a certain level, an increase in the sample size will not increase the accuracy of your survey by any appreciable amount. For example, the accuracy of a sample of 1,000 people out of a target population of 1 million will be about the same as for a target population of 10 million.

Statistical accuracy of the sample size

There are two main aspects to ensuring that your sample is statistically accurate.

1. **Confidence interval**
   This is the plus or minus figure that you often see in media and newspaper opinion polls. For example, if the poll showed a confidence interval of +/-3% and 45% of respondents answered in a certain way, then you can be fairly certain that if you had asked the entire target population, then between 42% (45 – 3) and 48% (45 + 3) would have answered in the same way.

2. **Confidence level**.
   This indicates how sure you are of your results. It is usually expressed as a percentage and signifies how often the percentage of the target population would respond in a certain way within your confidence interval. A 95% confidence level means that you would be 95% certain, a 99% confidence level would mean you would be 99% certain. For social research surveys, a 95% confidence level is sufficiently accurate.

As a rough guide, for target populations of under 1,000, you would need a sample of around 30% (300). For populations of up to 10,000, you would need a sample of about 10%. Above 10,000 population, a sample size above 1,000 does not usually change the accuracy of your survey significantly.

An important point to bear in mind with sample sizes in surveys, is the anticipated response rate. The sample size you decide on is the actual minimum number of returns you need in order to ensure that your survey is statistically accurate. However, particularly for postal surveys, response rates can often be fairly low, e.g. 30% or less response rate. If you anticipate that your survey will only provide a 30% response rate, this would mean that if you wanted to get a sample of 1,000 people, you would need to send out a minimum of 3,000 survey questionnaires.

Sampling Methods

There are two main approaches to selecting samples, probability and non-probability sampling:

1.   Probability sampling

In probability sampling, each person in the target population has an equal chance of being included in the sample. The main advantage of probability sampling is that you can calculate the "sampling error". The sampling error is the extent to which a sample might be different to the target population. When drawing conclusions about the sample compared with the target population, you can report the results as a +/- sampling error. In non-probability sampling, the extent to which the sample differs from the population will be unknown.

There are several probability sampling methods available including:

Random sampling

This is sometimes called **simple random sampling**. In this method, each person in the target population has an equal chance of being included in the sample. For example, you could select 1,000 names from a list of 50,000 names, by generating 1,000 random numbers, and allocating these numbers to your list of names. The easiest method to carry out simple random sampling is "without replacement". This means that the computer package or other method of generating random numbers, does not generate the same number more than once, otherwise the same number may come up several times and you would then have to generate a different set of random numbers.

Advantage	It is easy to carry out, and does not need additional information, such as, geographic location, gender, age, or other variables.
Disadvantage	Because each person has an equal chance of being selected for the sample, very small proportions of the target population may be missed altogether in the sample. For example, obtaining the views of ethnic minority groups, or people living in small towns may not be included in the sample due to their low numbers.

Systematic sampling

This is sometimes called **interval sampling**. In this method, you leave a gap or interval between each person included in the sample. For example, to select 100 people from a list of 400 people, you would select every fourth person, starting from a random number from 1-4. If the random number you started with was say 2, the next number would be 6, then 10, and so on up to 400.

Advantage	The sample is distributed evenly over the target population listing.
Disadvantage	You need to be careful that your target population listing does not coincide with the interval you have chosen for the sample. For example, if you have a list of employees listed by grade you may find that the more senior employees have a higher chance of being selected than lower graded employees do. To overcome this, you can jumble up or randomise the listing into alphabetical order or some other random listing.

Stratified sampling

This method tries to ensure that you include all important views in your sample. You divide your target population into mutually exclusive groups called "**strata**", for example, by gender, age, location, job type, etc. You then take a sample of each group or stratum depending on their representation within your target population.

Advantage	It reduces sampling error, and ensures that you include views from all groups
Disadvantage	It is more time consuming than simple random sampling, and may over-represent certain groups if you don't apply an appropriate weighting to the results.

Cluster sampling

In some cases it is too costly to spread a sample across the entire target population, for example, where there is a large geographical area to cover. In order to reduce costs you can use a cluster sampling technique. Cluster sampling involves dividing your target population into groups or clusters, depending on the characteristics you want to investigate. You then randomly select a number of clusters to represent your target population. Depending on the size of your clusters, you can either include everyone in the selected clusters to form your sample or randomly select people from the selected clusters, often known as "multi-stage sampling". This method differs from stratified sampling where you select people from each group.

Advantage	It reduces costs in terms of survey coverage. While a list of all groups in your population may not always be available, a list of clusters may be relatively easy to create.
Disadvantage	It is less efficient than simple random sampling. It is usually better to survey a large number of small clusters, rather than a small number of large clusters. This is because neighbouring groups can often have similar characteristics, and this can result in a sample that does not represent the range of views of the target population.

2.   **<u>Non-probability sampling</u>**

In non-probability sampling, people are selected for the sample in some non-random manner. In non-probability sampling there is an assumption that your target population has an equal distribution of characteristics, for example, same age category. Non-probability sampling is often used in the early stages of a survey, for example, to test survey questions.

The most common methods used in non-probability sampling include:

Convenience sampling

Often called haphazard or accidental sampling. Here the sample is not usually representative of the target population, but is collected on the basis of ease or convenience. For example, television reporters often select "people on the street" style interviews to find out what the person on the street thinks about a particular issue.

Advantage	It is easy to use, and where the target population is homogeneous or has similar characteristics, then it is reasonably accurate.
Disadvantage	From a statistical viewpoint, there is no way to calculate the probability of a person being included in the sample, and it is impossible to estimate the variability of the sample, or to identify possible bias.

Volunteer sampling

Where people volunteer to take part in a survey.

Advantage	It is usually cheap and easy to administer
Disadvantage	It can sometimes result in bias, as people that volunteer do not always exhibit the same views or attitudes as the overall target population.

Judgement sampling

This method relies on the judgement of the person carrying out the survey to select a sample that reflects the general characteristics of the target population.

Advantage	It is less costly and time consuming to select a sample than other methods.
Disadvantage	The objectivity of the person carrying out the survey may be less robust, and biases may be more pronounced based on any preconceived views that they may have.

Quota sampling

This is where you sample a specific number of people (or quotas) from groups that you identify in your target population. These groups can include people of a certain age, gender, ethnicity, etc.

Advantage	This non-probability sampling method is more accurate than other forms, as it includes people from identified groups within your target population.
Disadvantage	Unlike in stratified sampling, where people are selected randomly, quota sampling is left up to the person carrying out the survey to select the people for the quotas, and this may introduce bias into the sample.

Snowball sampling

This is where research participants select other participants for the study. This is a good technique for reaching specific population groups, or subsets of the community for which not a lot is known, ie homeless people. For example, in a study of migrant workers in Fife, interviewees asked participants to identify other people who met the sample target characteristics who might be approached for interview.

Advantage	It is a less costly method, and relatively easy to do
Disadvantage	It is much more subjective, it can lead to bias. Because of the ease it can lose focus on reaching the key target group. It is important to know where to draw the line.

Summary

Survey sampling is about choosing a representative group from a target population, and drawing conclusions from that sample that will be applicable to the target population. The two basic approaches to sampling are probability and non-probability sampling. The approach that you choose will depend on the purpose and accuracy of the information that you want to collect.

KnowFife
PartnershipHub
know.fife.scot

Follow on Twitter
@KnowFife

KnowFife
Dataset
knowfife.fife.gov.uk
6