Block 9: Sampling - design and procedures

(Activity solutions can be found at the end of the document.)

Sampling is a key component of any research design. **Sampling design** involves several basic questions.

- Should a sample be taken?
- If so, what process should be followed?
- What kind of sample should be taken?
- How large should it be?
- What can be done to control and adjust for non-response errors?

We now consider how to answer these questions.

Learning Objectives

- discuss the sampling design process: definition of the target population, determination of the sampling frame and selection of sampling technique(s)
- classify sampling techniques as non-probability and probability sampling techniques
- describe the non-probability sampling techniques of convenience, judgemental, quota and snowball sampling
- describe the probability sampling techniques of simple random, systematic, stratified and cluster sampling
- identify the conditions that favour the use of non-probability sampling versus probability sampling.

Reading List

Malhotra, N.K., D. Nunan and D.F. Birks. Marketing Research: An Applied Approach. (Pearson, 2017) 5th edition [ISBN 9781292103129] Chapter 14.

9.1 Sampling - design and procedures

For each section of *Sampling - design and procedures*, use the LSE ELearning resources to test your knowledge with the Key terms and concepts flip cards.

Sample or census?

We define the following terms:

Population - the aggregate of all the elements, sharing some common set of characteristics, which comprise the universe for the purpose of the market research problem.

Census - a complete enumeration of the elements of a population or study objects.

Sample - a subgroup of the elements of the population selected for participation in the study.

Conditions favouring the use of a sample versus a census

Factor	Sample	Census
Budget	Small	Large
Time available	Short	Long
Population size	Large	Small
Variance in the characteristic	Small	Large
Cost of sampling errors	Low	High
Cost of non-sampling errors	High	Low
Nature of measurement	Destructive	Non-destructive
Attention to individual cases	Yes	No

The sampling design process

The sampling design process is:

Define the target population

⇓

Determine the sampling frame

⇓

Select a sampling technique

⇓

Determine the sample size

⇓

Execute the sampling process

⇓

Validate the sample

The sampling design process

The target population is the collection of elements or objects which possess the information sought by the researcher and about which inferences are to be made. The target population should be defined in terms of elements, sampling units, extent and time.

- An element is the object about which, or from which, the information is desired - for example, the participant.
- A sampling unit is an element, or a unit containing the element, which is available for selection at some stage of the sampling process.
- Extent refers to the geographical boundaries.
- Time is the time period under consideration.

Sample size determination

The next block will consider sample size determination from a *statistical perspective*. However, for now we consider important *qualitative factors* in determining the sample size. These are:

- The importance of the decision
- the nature of the research
- the number of variables
- the nature of the analysis
- sample sizes used in similar studies
- incidence rates
- completion rates
- Resource constraints.

Type of study	**Minimum size**	**Typical range**
Problem identification research	500	1,000-2,500 (e.g. market potential)
Problem-solving research	200	300-500 (e.g. pricing)
Product tests	200	300-500
Test marketing studies	200	300-500
TV, radio, print or online advertising	150	200-300 (per advertisement tested)
Test-market audits	10 stores	10-20 stores
Focus groups	6 groups	6-12 groups

Usual sample sizes used in market research studies

Classification of sampling techniques

Figure 14.2 of the textbook provides a classification of sampling techniques.
Features of **non-probability sampling** include the following.

- Some units in the population have a *zero probability* of selection.
- Individual units in populations have an *unknown probability* of being selected.

- Inability to measure *sampling error*.
- Features of **probability sampling** include the following.
- Every population element has a *known, non-zero probability* of being selected in the sample.
- Probability sampling makes it possible to estimate the *margins of sampling error*.

All statistical techniques can be applied.
To perform probability sampling we need a **sampling frame** (a list of all population elements), but we should assess the sampling frame's:

- **adequacy** (fit for purpose?)
- **completeness** (missing units? duplication?)
- **accuracy** (dynamic populations?)
- **convenience** (accessibility?).
- Examples include the following.
- Register of electors.
- Postcode Address File (PAF).
- Directories.

Non-probability sampling techniques

Convenience sampling attempts to obtain a sample of convenient elements. Often, participants are selected because they happen to be *in the right place at the right time*. Examples include:

- the use of students and members of social organisations
- street interviews without qualifying the participants
- 'people-on-the-street' interviews.

Judgemental sampling is a form of convenience sampling in which the population elements are selected based on the *judgement of the researcher*. Examples include:

- test markets
- purchase engineers selected in business-to-business (B2B) market research
- expert witnesses used in court.

Quota sampling may be viewed as two-stage restricted judgemental sampling. The first stage consists of developing *control categories*, or *quota controls*, of population elements. In the second stage, sample elements are selected based on convenience or judgement.

In **snowball sampling**, an initial group of participants is selected, usually at random. After being interviewed, these participants are asked to identify others who belong to the target population of interest. Subsequent participants are selected based on the *referrals*.

Figure 14.3 of the texbook provides a graphical illustration of non-probability sampling techniques.

Probability sampling techniques

In **simple random sampling (SRS)**, each element in the population has a *known and equal probability of selection*. Each possible sample of a given size, nn, has a known and equal probability of being the sample actually selected. This implies that every element is selected *independently* of every other element.

In **systematic sampling**, the sample is chosen by selecting a *random starting point* and then picking every iith element in succession from the sampling frame. The **sampling interval**, ii, is determined by dividing the population size, NN, by the sample size, nn, and rounding to the nearest integer. When the *ordering of the elements is related to the characteristic of interest*, systematic sampling *increases the representativeness* of the sample. If the ordering of the elements produces a *cyclical pattern*, systematic sampling may *decrease the representativeness* of the sample.

For example, there are N=100000N=100000 elements in the population and a sample of n=1000n=1000 is desired. In this case the sampling interval, ii, is:

i=Nn=1000001000=100.i=Nn=1000001000=100.

A random number between 1 and 100 is selected. If, for example, this number is 23, the sample consists of elements 23, 123, 223, 323, 423, 523 and so on.

Stratified sampling is a two-step process in which the population is partitioned into subpopulations, or *strata*. The strata should be **mutually exclusive and collectively exhaustive** in that every population element should be assigned to one, and only one, stratum and no population elements should be omitted. Next, elements are selected from each stratum by a random procedure, usually SRS. A major objective of stratified sampling is to *increase precision without increasing cost*.

The elements *within a stratum* should be as *homogeneous* as possible, but the elements *in different strata* should be as *heterogeneous* as possible. The **stratification factors** should also be closely related to the characteristic of interest. Finally, the factors (variables) should decrease the cost of the stratification process by being *easy to measure and apply*.

In **proportionate stratified sampling**, the size of the sample drawn from each stratum is *proportionate to the relative size of that stratum* in the total population. In **disproportionate (optimal) stratified sampling**, the size of the sample from each stratum is proportionate to the relative size of that stratum *and* to the standard deviation of the distribution of the characteristic of interest among all the elements in that stratum.

In **cluster sampling**, the target population is first divided into mutually exclusive and collectively exhaustive subpopulations, or *clusters*. Next, a *random sample of clusters* is selected, based on a probability sampling technique such as SRS. For each selected cluster, either *all the elements are included* in the sample (**one-stage**) or a *sample of elements* is drawn probabilistically (**two-stage**).

Elements *within a cluster* should be as *heterogeneous* as possible, but clusters themselves should be as *homogeneous*as possible. Ideally, each cluster should be a *small-scale representation* of the population. In **probability proportionate to size** sampling, the clusters are sampled with *probability proportional to size*. In the second stage, the probability of selecting a sampling unit in a selected cluster varies inversely with the size of the cluster.

Differences between stratified and cluster sampling

Factor	**Stratified sampling**	**Cluster sampling (one-stage)**
Objective	Increase precision	Decrease cost
Subpopulations	All strata are included	A sample of clusters is chosen

Differences between stratified and cluster sampling

Factor	Stratified sampling	Cluster sampling (one-stage)
Within subpopulations	East stratum should be homogeneous	Each cluster should be heterogeneous
Across subpopulations	Strata should be heterogeneous	Clusters should be homogeneous
Sampling frame	Needed for the entire population	Needed only for the selected clusters
Selection of elements	Elements selected from each stratum randomly	All elements from each selected cluster are included

Figure 14.4 of the textbook provides a graphical illustration of probability sampling techniques.

In **multistage sampling**, sample selection is performed at two or more successive stages. This technique is often adopted in large surveys. At the first stage, large compound units are sampled (*primary units*). At the second stage, smaller units are sampled from these (*secondary units*), and several sampling stages of this type may be performed until we at last sample the basic units.

The technique is commonly used in cluster sampling so that we are at first sampling main clusters, and then clusters within clusters etc. We can also use multistage sampling with *mixed techniques*, i.e. cluster sampling at stage 1 and stratified sampling at stage 2 etc. Example: In a national survey of salespeople in a company, sales areas could be identified and a random selection taken of these. Instead of interviewing every person in the cluster, only randomly selected people within the cluster would be interviewed.

Procedures for drawing probability samples

Simple random sampling:

1. Select a suitable sampling frame.
2. Each element is assigned a number from 1 to NN (population size).
3. Generate nn (sample size) different random numbers between 1 and NN.
4. The numbers generated denote the elements which should be included in the sample.

Systematic sampling:

1. Select a suitable sampling frame.
2. Each element is assigned a number from 1 to NN (population size).
3. Determine the sampling interval ii, where i=N/ni=N/n. If ii is a fraction, round to the nearest integer.
4. Select a random number, rr, between 1 and ii, as explained in simple random sampling.
5. The elements with the following numbers will comprise the systematic random sample: r,r+i,r+2i,r+3i,r+4i,…,r+(n−1)i.r,r+i,r+2i,r+3i,r+4i,…,r+(n−1)i.

Stratified sampling:

1. Select a suitable sampling frame.
2. Select the stratification factor(s) and the number of strata, HH.
3. Divide the entire population into HH strata. Based on the classification variable, each element of the population is assigned to one of the HH strata.
4. In each stratum, number the elements from 1 to NhNh (the population size of stratum hh).
5. Determine the sample size of each stratum, nhnh, based on proportionate or disproportionate stratified sampling, where:∑h=1Hnh=n.∑h=1Hnh=n.
6. In each stratum, select a simple random sample of size nhnh.

Cluster sampling:

1. Assign a number from 1 to NN to each element in the population.
2. Divide the population into CC clusters of which cc will be included in the sample.
3. Calculate the sampling interval ii, where i=N/ci=N/c (round to the nearest integer).
4. Select a random number rr between 1 and ii, as explained in simple random sampling.
5. Identify elements with the following numbers: r, r+i, r+2i, …,r+(c−1)i.r, r+i, r+2i, …,r+(c−1)i.
6. Select the clusters which contain the identified elements.
7. Select sampling units within each selected cluster based on SRS or systematic sampling.

Comparison of sampling techniques

Conditions favouring the use of non-probability sampling versus probability sampling

Factor	**Non-probability sampling**	**Probability sampling**
Nature of research	Exploratory	Conclusive
Relative magnitude of sampling and non-sampling errors	Non-sampling errors are larger	Sampling errors are larger
Variability in the population	Homogeneous (low)	Heterogeneous (high)
Statistical considerations	Unfavourable	Favourable
Operational considerations	Favourable	Unfavourable

Strengths and weaknesses of non-probability sampling techniques

Technique	Strengths	Weaknesses
Convenience sampling	Least expensive, least time-consuming, most convenient	Selection bias, sample not representative, not recommended for descriptive or causal research
Judgemental sampling	Low cost, convenient, not time-consuming, ideal for exploratory research designs	Does not allow generalisation, subjective
Quota sampling	Sample can be controlled for certain characteristics	Selection bias, no assurance of representativeness
Snowball sampling	Can estimate rare characteristics	Time-consuming

Strengths and weaknesses of probability sampling techniques

Technique	Strengths	Weaknesses
Simple random sampling	Easily understood, results projectable	Difficult to construct samplign frame, expensive, lower precision, no assurance of representativeness
Systematic sampling	Can increase representativeness, easier to implement than SRS, sampling frame not always necessary	Can decrease representativeness depending upon 'order' in the sampling frame
Stratified sampling	Includes all important subpopulations, precision	Difficult to select relevant stratification factors, not feasible to stratify on many variables, expensive
Cluster sampling	Easy to implement, cost effective	Imprecise, difficult to compute and interpret results

Online sampling

Online sampling is the fastest growth area in market research. There are already many research companies which deal exclusively in online research. The most high profile in the UK is YouGov. Internet penetration in the UK is very high (users both at home, work, internet cafes etc.), similar to landline penetration (of households). So, in terms of coverage, there is not too much difference between internet users and households which have a landline.

Advantages:

- It is *cheaper*.
- It is potentially *very quick*.
- Participants tend to give *more honest answers*.
- Tend to get *more young males* online - traditionally this is a hard-to-reach group for market research.

Disadvantages:

- Internet penetration is still not as high as landline telephone penetration and participants who are not online may have a different profile.
- Participants to online surveys tend to be heavy internet users. Light internet users may have a different profile, and will take longer to respond!
- Participants to online surveys are much more inclined to be *technology savvy* and early adopters (in the UK). Surveys measuring technology usage are likely to provide *biased results*.
- It is difficult to know what the online universe is. This is especially a problem when it comes to *weighting*.
- One tends to get many more *'professional panellists'* - those who sign up simply to receive incentives and answer as many surveys as they can with as little effort as possible (= poor data quality).
- There is *no recognised sampling frame* for the online population.

There are three ways of sampling online.

- Use an **'access panel'**.
- Use **web intercept** (a pop-up).
- Use **client lists** (email addresses).

Access panels are the most common approach. Many research companies, including YouGov, maintain their own panel of participants who are signed up. There are also other panels available from which to choose.

Advantages include the following.

- Panels are relatively cheap and quick - most panels have hundreds of thousands, if not millions, of participants 'signed up'.
- Demographic information (age, gender, location etc.) are already known about the participants, so a representative sample can be instantly selected.

Disadvantages include the following.

- Clearly, we are reliant on participants telling the truth about themselves (age, gender etc.). This is also a problem with telephone interviewing, but there is more to hide behind online.
- Different access panels have, in practice, been shown to give different results. Hence there is a need to select from different panels.

- Many participants are members of multiple panels (some are even members of the same panel more than once). This can be difficult to identify and *de-duplicate*.
- It is still necessary to *set quotas for completion*, as the panel is not an accurate sampling frame.

Web intercept advantages include the following.

- It is possible to *target users* of one or more particular websites.
- You can ensure that you are picking up participants who are accessing at different times (daytime, evenings, weekends etc.).
- By by-passing panels, you get some (hopefully many) non-panellists, so the overall *representativeness should be improved*.

Web intercept disadvantages include the following.

- It is still essentially a *self-selecting sample* so will still only pick up participants who are keen participants.
- It is difficult to define your population - how do you know those responding are, for instance, just in the UK?
- Many internet
- It is not very useful for getting a widespread sample of users of the internet, as it is just limited to users of the websites where the intercepts are used.

Client list advantages include the following.

- We have a *meaningful sampling frame*.
- There are no concerns about the vagaries of an internet population.

Client list disadvantages include the following.

- Client lists tend to be *badly maintained* as they will invariably contain omissions, duplications or invalid/out-of-date/incorrect contact information.
- It is still necessary to be careful of *differential response rates*.

Weighting data from online surveys to be representative of a given population may be difficult. Demographic weighting may not suffice as it may be necessary to use more sophisticated techniques such as propensity weighting. Online *data quality* is a real problem. Many participants (especially ‘professional panellists’) complete surveys too quickly for their responses to be reliable. The longer an online survey takes place, the poorer the data quality. Research has shown this effect is more marked than in traditional methods of data collection. Some of the time you save in conducting an online survey may be eaten up by checking data quality and removing poor quality participants.

Discussion forum and case studies

To access the solutions to these questions and case study, click here to access the printable Word document or click here to go to LSE’s Elearning resources.

Discussion points

- Given that many governments use sampling to check the accuracy of various censuses and that non-response rates to censuses are growing, national decennial censuses should be abolished in favour of the use of existing databases and sample surveys.
- Because non-sampling errors are greater in magnitude than sampling errors, it really does not matter which sampling method is used.

Learning outcomes checklist

Use this to assess your own understanding of the chapter. You can always go back and amend the checklist when it comes to revision!

- ○ Discuss the sampling design process: definition of the target population, determination of the sampling frame and selection of sampling technique(s)
- ○ Classify sampling techniques as non-probability and probability sampling techniques
- ○ Describe the non-probability sampling techniques of convenience, judgemental, quota and snowball sampling
- ○ Describe the probability sampling techniques of simple random, systematic, stratified and cluster sampling
- ○ Identify the conditions that favour the use of non-probability sampling versus probability sampling.

Block 9: Sampling - design and procedures

Commentary on Discussion points

- The issue here is the difference between a census and a sample, and the associated errors which arise from the sampling process. A census, while more expensive to conduct, is also more reliable. Sampling simply serves as a check on the census results. If only a sample was taken, non-response error would be a critical factor undermining the reliability of the findings.
- Issues which can be generated from this discussion include the nature of the various sampling techniques. While each technique will have a degree of error, each one is appropriate for a particular sampling problem. Therefore, the researcher can reduce both non-sampling and sampling errors when the techniques are applied in the appropriate situations. The point is that non-sampling error, like sampling error, could be affected by the technique chosen.