

Selecting samples

Prepared by Associate Professor:
Ahmed Rageh Ismail

ENT 636
Research Project

Lecture 5



Chapter

05

Learning outcomes

By the end of this chapter you should:

- understand the need to select samples in entrepreneurship research;
- be aware of a range of probability and non-probability sampling techniques and the possible need to combine techniques within a research project;
- be able to choose appropriate sampling techniques for a variety of research scenarios and justify your choices;
- be able to use a range of sampling techniques;
- be able (where appropriate) to assess the representativeness of the sample selected;
- be able to assess the extent to which it is reasonable to generalize from a sample;
- be able to apply the knowledge, skills and understanding gained to your own research project.

Whatever your research question(s) and objectives, you will need to consider whether you need to select one or more samples. Occasionally, it may be possible to collect and analyze data from every possible case or group member; this is termed a **census**.

However, for many research questions and objectives it will be impossible for you either to collect or to analyze all the potential data available to you, owing to restrictions of time, money and often access.

This means you will need to select data for a subgroup or **sample** of all possible cases.

Sampling techniques enable you to reduce the amount of data you need to collect by considering only data from a subgroup rather than all possible cases or **elements**.

Some research questions will require sample data that can allow you to generalize statistically about all the cases from which your sample has been selected.

Introduction

To gain an understanding of how entrepreneurs manage their businesses, you may select a small sample of entrepreneurs. Consequently, whatever your research question, an understanding of techniques for selecting samples is likely to be very important.

The full set of cases or elements from which a sample is taken is called the **population**. In sampling, the term 'population' is not used in its normal sense, as the full set of cases need not necessarily be people. For research to discover the level of service at Japanese restaurants throughout a country, the population from which you would select your sample would be all Japanese restaurants in that country.

Introduction

When selecting a sample to study, it should represent the population from which it is taken in a way that is meaningful and which we can justify in relation to answering our research question and meeting our objectives.

If we are using our sample data to infer statistically something about a population, it is important that our sample is sufficiently large to allow such statistical inferences to be made with an acceptable margin of error.

The need to sample

For some research questions it is possible to collect data from an entire population as it is of a manageable size. However, you should not assume that a census would necessarily provide more useful results than collecting data from a sample.

Sampling provides a valid alternative to a census when:

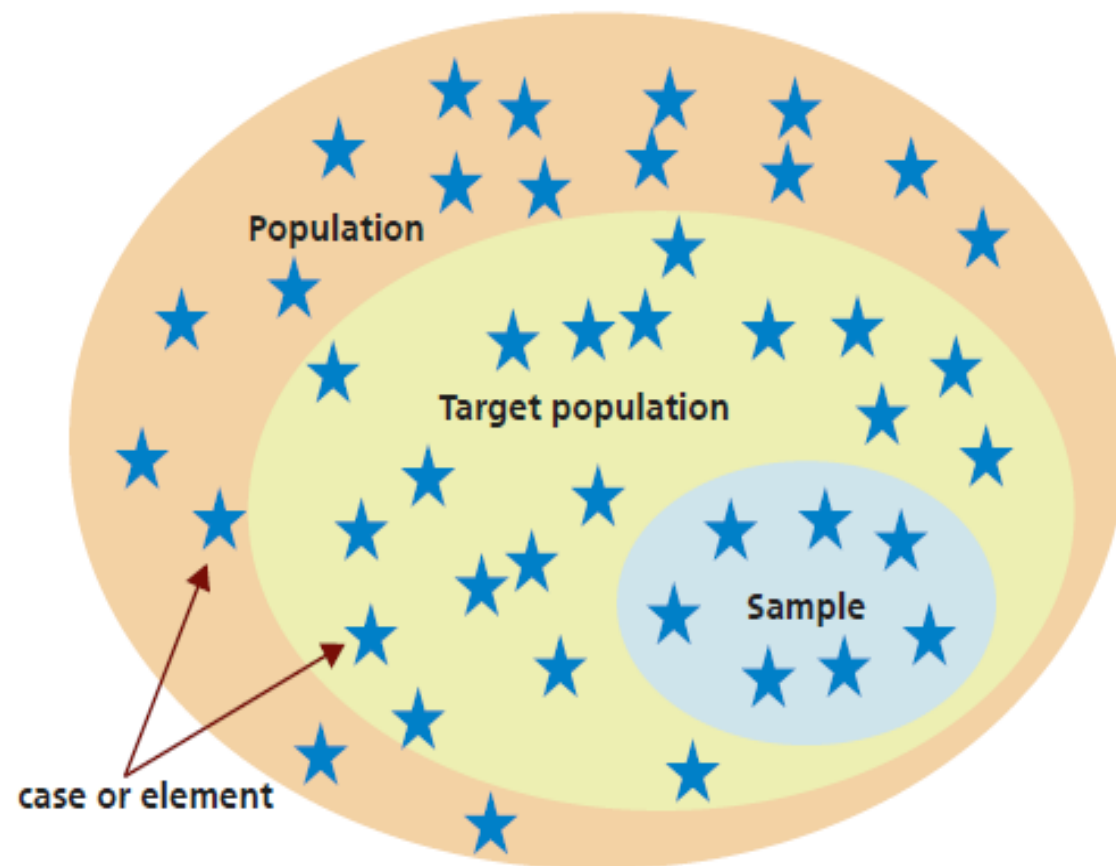
- it would be impracticable for you to survey the entire population;
- your budget constraints prevent you from surveying the entire population;
- your time constraints prevent you from surveying the entire population.

The importance of defining the research population clearly

The sample selected should be related to the population that is highlighted in the research question and objectives. This sample, providing it is selected carefully, will allow conclusions to be drawn about the entire population.

However, such a population may be difficult to research as not all elements or cases may be known to the researcher or easy to access. Consequently, the researcher may redefine the population as something more manageable.

This is often a subset of the population and is called the **target population** and is the actual focus or target of the research inquiry.



Population, target population, sample and individual cases

An overview of sampling techniques

Sampling techniques available to you can be divided into two types:

- probability or representative sampling;
- non-probability sampling.

The **probability samples** the chance, or probability, of each case being selected from the target population is known and is usually equal for all cases. This means it is possible to answer research questions and to achieve objectives that require you to estimate statistically the characteristics of the target population from the sample. Consequently, probability sampling is often associated with survey and experiment research strategies.

An overview of sampling techniques

For **non-probability samples**, the probability of each case being selected from the target population is not known and it is impossible to answer research questions or to address objectives that require you to make statistical inferences about the characteristics of the population.

Although each technique is discussed separately, for many research projects researchers will need to use a combination of sampling techniques, some projects involving both probability and non-probability sampling techniques. Sampling designs that have two or more successive stages using either probability, non-probability, or both types of sample selection techniques are known as **multi-stage sampling**

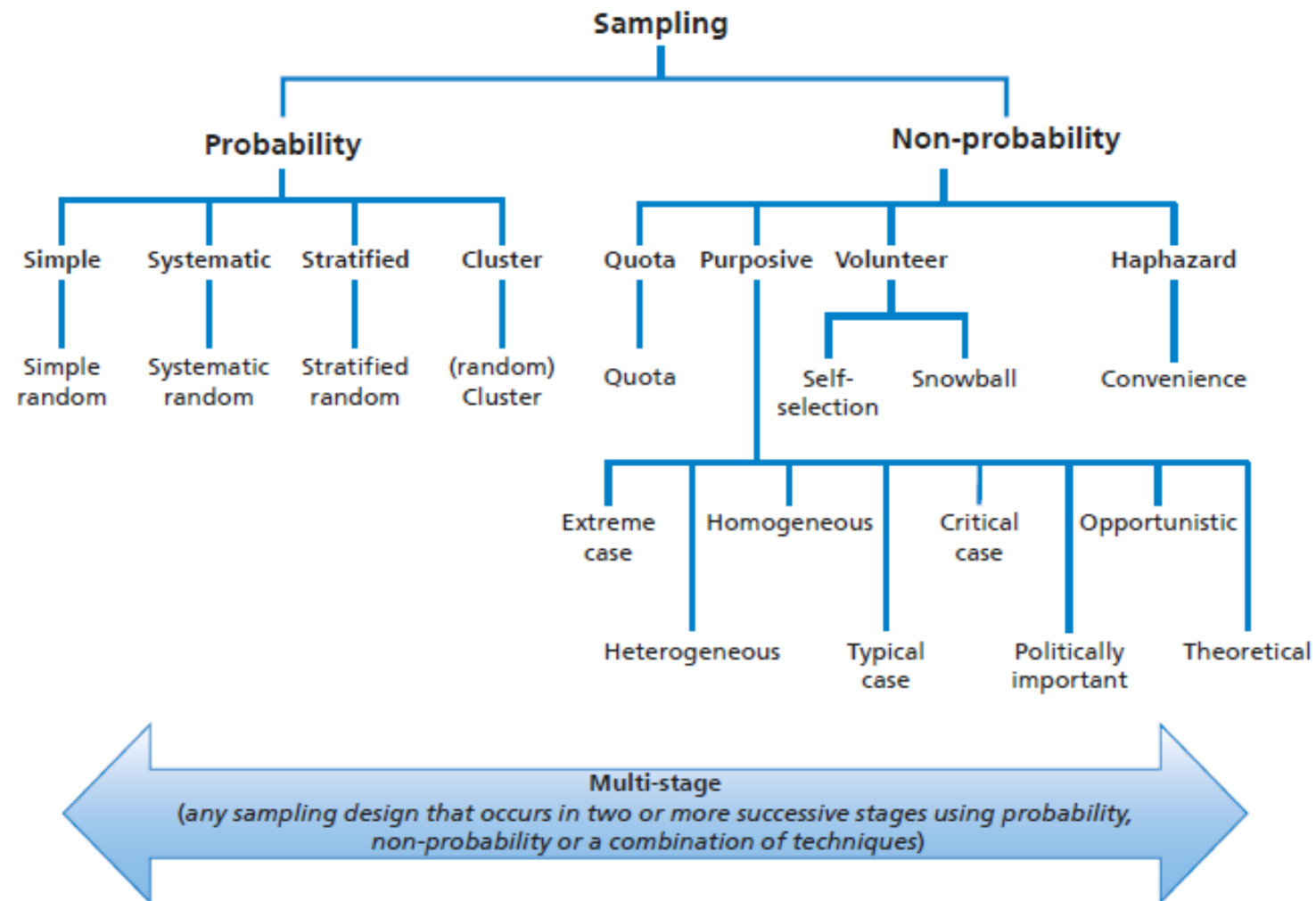


Figure 7.2 Sampling techniques

Probability sampling

Probability sampling (or **representative sampling**) is associated most commonly with survey research strategies where you need to make statistical inferences from your sample about a population to answer your research question(s) and to meet your objectives.

The process of probability sampling can be divided into four stages:

- 1 Identify a suitable sampling frame based on your research question(s) and objectives.
- 2 Decide on a suitable sample size.
- 3 Select the most appropriate sampling technique and select the sample.
- 4 Check that the sample is representative of the target population.

Identifying a suitable sampling frame and the implications for generalizability

The **sampling frame** for any probability sample is a complete list of all the cases in the target population from which your sample will be drawn. Without a sampling frame, you will not be able to select a probability sample and so will have to consider using non-probability sampling.

If your research question is concerned with organizations in a particular sector, you may be thinking of creating a sampling frame from an existing databases of companies. You then select your sample from your list.

Identifying a suitable sampling frame and the implications for generalizability

Obtaining a sampling frame is therefore essential if you are going to use probability sampling. However, you need to be aware of the possible problems of using existing databases for your sampling frame. Because,

- individual databases are often incomplete;
- the information held about organizations in databases is sometimes inaccurate;
- the information held in databases soon becomes out of date.

Identifying a suitable sampling frame and the implications for generalizability

This emphasizes the importance of ensuring your sampling frame is as complete, accurate and up to date as possible. An incomplete or inaccurate list means that some cases will have been excluded and so it will be impossible for every case in the target population to have a chance of selection.

If this is the case you need to state it clearly. Where no suitable list exists, and you wish to use probability sampling, you will have to compile your own sampling frame (perhaps drawing upon existing lists). It is important to ensure that your sampling frame is valid



Box 7.1 Checklist

Selecting your sampling frame

- ✓ Are cases listed in the sampling frame relevant to your research topic, in other words does your target population enable you to answer your research question and meet your objectives?
- ✓ How recently was the sampling frame compiled; is it up to date?
- ✓ Does the sampling frame include all cases in the target population; is it complete?
- ✓ Does the sampling frame contain the correct information; is it accurate?
- ✓ Does the sampling frame exclude irrelevant cases; is it precise?
- ✓ For purchased lists, can you establish and control precisely how the sample will be selected?

Deciding on a suitable sample size

Generalizations about target populations from data collected using any probability samples are based on statistical probability. The larger your sample's size the lower the likely error in generalizing to the target population.

Probability sampling is therefore a compromise between the accuracy of your findings and the amount of time and money you invest in collecting, checking and analyzing the data.

Deciding on a suitable sample size

Your choice of sample size within this compromise is governed by:

- the confidence you need to have in your data – that is, the level of certainty that the characteristics of the data collected will represent the characteristics of the target population;
- the margin of error that you can tolerate – that is, the accuracy you require for any estimates made from your sample;
- the types of analyses you are going to undertake – in particular, the number of categories into which you wish to subdivide your data, as many statistical techniques have a minimum threshold of data cases (e.g. chi square); and, to a lesser extent,
- the size of the target population from which your sample is being drawn..

Deciding on a suitable sample size

Given these competing influences, it is not surprising that the final sample size is almost always a matter of judgement as well as of calculation. However, if your sample is extremely large you may find that while relationships are statistically significant, the practical implications (effect size) of this difference are small.

For many research questions and objectives, the specific statistical analyses you need to undertake will determine the threshold sample size for individual categories. In particular, an examination of virtually any statistics textbook will highlight that, in order to ensure spurious results do not occur, the data analyzed must be normally distributed. While the normal distribution is discussed later, its implications for sample size need to be considered here.

Deciding on a suitable sample size

Statisticians have proved that the larger the absolute size of a sample, the closer its distribution will be to the normal distribution and thus the more robust it will be. This relationship, known as the **central limit theorem**, occurs even if the population from which the sample is drawn is not normally distributed.

Statisticians have also shown that a sample size of 30 or more will usually result in a sampling distribution for the mean that is very close to a normal distribution. For this reason, it is advised a minimum number of 30 for statistical analyses a rule of thumb for the smallest number of cases in each category within the overall sample.

Deciding on a suitable sample size

It is likely that, if you are undertaking statistical analyses on your sample, you will be drawing conclusions from these analyses about the target population from which your sample was selected. This process of coming up with conclusions about a population on the basis of data describing the sample is called **statistical inference** and allows you to calculate how probable it is that your result, given your sample size, could have been obtained by chance. Such probabilities are usually calculated automatically by statistical analysis software.

However, it is worth remembering that, providing they are not biased, samples of larger absolute size are more likely to be representative of the target population from which they are drawn than smaller samples and, in particular, the mean (average) calculated for the sample is more likely to equal the mean for the target population. This is known as the **law of large numbers**.

Deciding on a suitable sample size

Researchers normally work to a 95 per cent level of certainty. This means that if your sample was selected 100 times, at least 95 of these samples would be certain to represent the characteristics of the target population. The confidence level states the precision of your estimates of the target population as the percentage that is within a certain range or margin of error.

The table below provides a guide to the different minimum sample sizes required from different sizes of target population given a 95 per cent confidence level for different margins of error. It assumes that data are collected from all cases in the sample.

Table 7.1 Sample sizes for different sizes of target population at a 95 per cent confidence level (assuming data are collected from all cases in the sample)

Target population	Margin of error			
	5%	3%	2%	1%
50	44	48	49	50
100	79	91	96	99
150	108	132	141	148
200	132	168	185	196
250	151	203	226	244
300	168	234	267	291
400	196	291	343	384
500	217	340	414	475
750	254	440	571	696
1 000	278	516	706	906
2 000	322	696	1091	1655
5 000	357	879	1622	3288
10 000	370	964	1936	4899
100 000	383	1056	2345	8762
1 000 000	384	1066	2395	9513
10 000 000	384	1067	2400	9595

Chapter

05

Deciding on a suitable sample size

For most entrepreneurship research, researchers are content to estimate the target population's characteristics at 95 per cent certainty to within plus or minus 3 to 5 per cent of its true values. This means that if 45 per cent of your sample are in a particular category then you will be 95 per cent certain that your estimate for the target population, within the same category, will be 45 per cent plus or minus the margin of error – somewhere between 42 and 48 per cent for a 3 per cent margin of error.

As you can see from table, the smaller the absolute size of the sample and, to a far lesser extent, the smaller the relative proportion of the target population sampled, the greater the margin of error. Within this, the impact of absolute sample size on the margin of error decreases for larger sample sizes.

The importance of a high response rate

The most important aspect of a probability sample is that it represents the target population. A perfect representative sample is one that exactly represents the target population from which it is taken.

Researchers therefore need to obtain as high a response rate as possible to reduce the risk of non-response bias and ensure the sample is representative. This is not to say that a low response rate will necessarily result in your sample being biased, just that it is more likely!

The importance of a high response rate

In reality, you are likely to have non-responses. Non-respondents are different from the rest of the target population because they are unable or unwilling to be involved in your research for whatever reason.

Consequently, your respondents will not be representative of the target population and the data you collect may be biased. Bias resulting from respondents differing in meaningful ways from non-respondents is known as **nonresponse bias**. In addition, each non-response will necessitate an extra respondent being found to reach the required sample size, increasing the cost of your data collection.

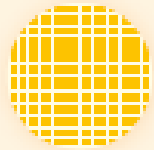
The importance of a high response rate

The most common reason for non-response is that your respondent refuses to answer all the questions or be involved in your research but does not give a reason. Such nonresponse can be minimized by paying careful attention to the methods used to collect your data.

Alternatively, some of your selected respondents may not meet your research requirements and so will be ineligible to respond. Non-location and non-contact create further problems; the fact that these respondents are unreachable means they will not be represented in the data you collect. As part of your research report, you will need to include your response rate. When you calculate this you should include all eligible respondents:

$$\text{total response rate} = \frac{\text{total number of responses}}{\text{total number in sample} - \text{ineligible}}$$

$$\text{active response rate} = \frac{\text{total number of responses}}{\text{total number in sample} - (\text{ineligible} + \text{unreachable})}$$



Box 7.3

Focus on student research

Calculation of total and active response rates

Ming had decided to collect data from people who had left his company's employment over the past five years by using a telephone questionnaire. He obtained a list of the 1,034 people who had left over this period (the total population) and selected a 50 per cent sample. Unfortunately, he could obtain

current telephone numbers for only 311 of the 517 ex-employees who made up his total sample. Of these 311 people who were potentially reachable, he obtained a response from 147. In addition, his list of people who had left his company was inaccurate, and nine of those he contacted were ineligible to respond, having left the company over five years earlier.

$$\text{His total response rate} = \frac{147}{517 - 9} = \frac{147}{508} = 28.9\%$$

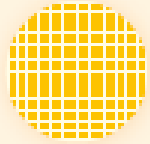
$$\text{His active response rate} = \frac{147}{311 - 9} = \frac{147}{302} = 48.7\%$$

Estimating response rates and actual sample size required

The actual sample size you require can be calculated using the following formula:

$$n^a = \frac{n \times 100}{re\%}$$

where n^a is the actual sample size required,
 n is the minimum (or adjusted minimum) sample size (see Table
 $re\%$ is the estimated response rate expressed as a percentage.



Box 7.4

Focus on student research

Calculation of actual sample size

Jan was a part-time student employed by a large manufacturing company. He had decided to send a questionnaire to the company's customers and calculated that an adjusted minimum sample size of 439 was required. From previous questionnaires that his company had used to collect data from customers, Jan knew the likely response rate would be

approximately 30 per cent. Using these data he could calculate his actual sample size:

$$\begin{aligned}n^a &= \frac{439 \times 100}{30} \\&= \frac{43900}{30} \\&= 1463\end{aligned}$$

Jan's actual sample, therefore, needed to be 1,463 customers. The likelihood of 70 per cent non-response meant that Jan needed to include a means of checking that his sample was representative when he designed his questionnaire.

Chapter

05

Estimating response rates and actual sample size required

If you are collecting your sample data from a secondary source within an organization that has already granted you access, for example a database recording customer complaints, your response rate should be virtually 100 per cent. Your actual sample size will therefore be the same as your minimum sample size.

In contrast, estimating the likely response rate from a sample to which you will be sending a questionnaire or interviewing is more difficult. One way of obtaining this estimate is to consider the response rates achieved for similar surveys that have already been undertaken and base your estimate on these.

Estimating response rates and actual sample size required

For most academic studies involving individuals or organizations' representatives, response rates of approximately 50 per cent and 35 to 40 per cent respectively are reasonable.

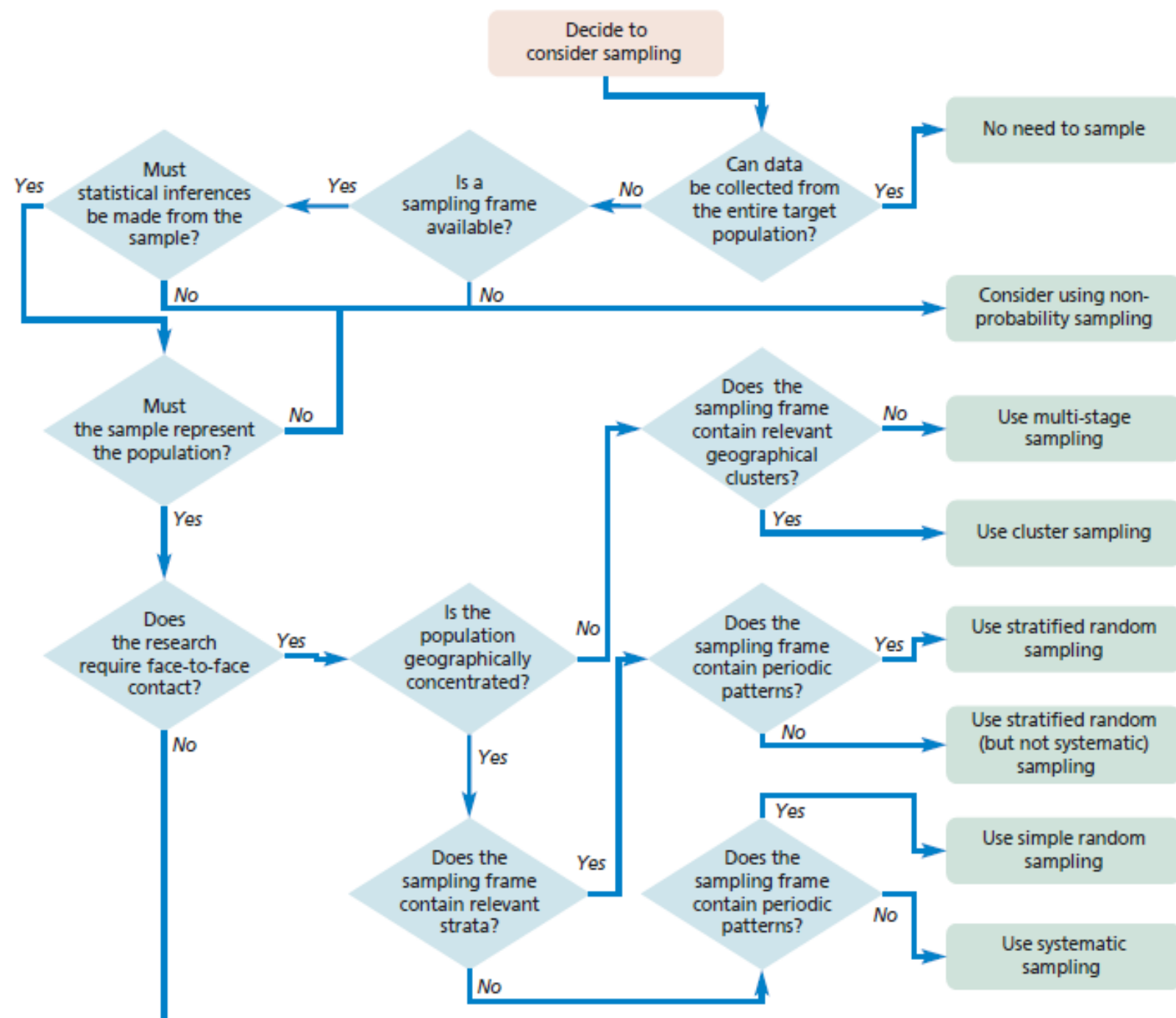
However, beware: response rates can vary considerably when collecting primary data. Reviewing literature on response rates for questionnaires noted wide variation and no consensus as to what was acceptable. Noting response rates of between 1 per cent and 100 per cent in published Business and Management research they offer general guidelines dependent upon discipline suggesting mean response rates of 35 per cent for International Business and Marketing, while Human Resource Management and General Management typically achieve 50 per cent.

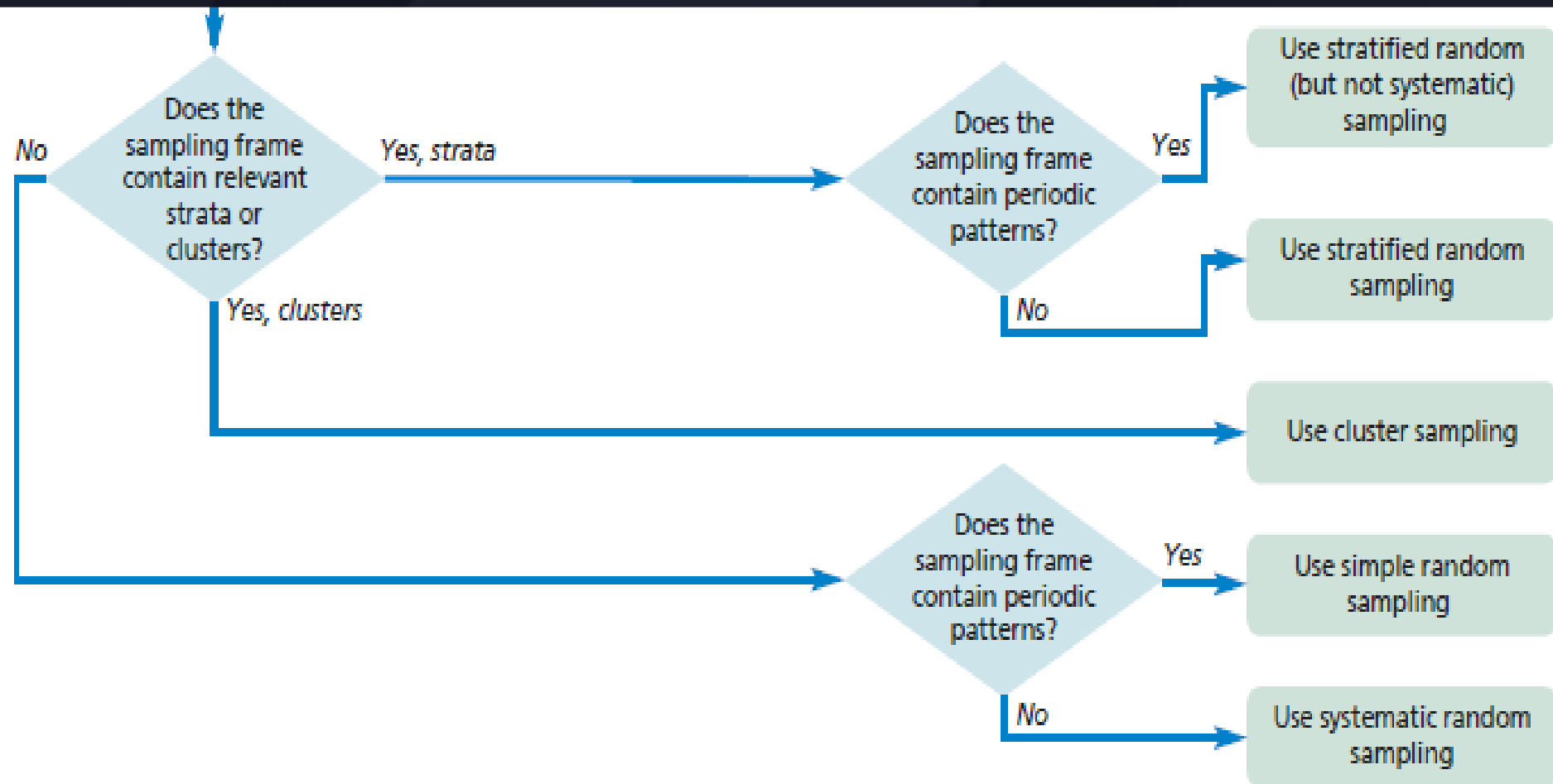
Selecting the most appropriate sampling technique and the sample

Having chosen a suitable sampling frame and establishing the actual sample size required, you need to select the most appropriate sampling technique to obtain a representative sample. Four main techniques can be used when selecting a probability sample (See the figure below):

- simple random;
- systematic random;
- stratified random;
- cluster.

Your choice of probability sampling technique depends on your research question(s) and your objectives. Subsequently, your need for face-to-face contact with respondents, and the geographical area over which the population is spread, further influence your choice of probability sampling technique.





Simple random sampling

Simple random sampling (sometimes called just random sampling) involves you selecting the sample at random from the sampling frame using a spreadsheet's random number generator function or random number tables. To do this you:

- 1 Number each of the cases in your sampling frame with a unique number. The first case is numbered 1, the second 2 and so on.
- 2 Select cases using random numbers such as those generated by a spreadsheet until your actual sample size is reached.

Starting with your first random number, you use this and subsequent random numbers in the order they were generated to select the cases (elements) until your sample size is reached

Systematic random sampling

Systematic random sampling (often called just systematic sampling) involves you selecting the sample at regular intervals from the sampling frame. To do this you:

- 1 Number each of the cases in your sampling frame with a unique number. The first case is numbered 1, the second 2 and so on.
- 2 Select the first case using a random number.
- 3 Calculate the sampling fraction.
- 4 Select subsequent cases systematically using the sampling fraction to determine the frequency of selection.

Systematic random sampling

To calculate the sampling fraction – that is, the proportion of the target population that you need to select – you use the formula:

$$\text{Sampling fraction} = \frac{\text{actual sample size}}{\text{total population}}$$

If your sampling fraction is 1/3 you need to select one in every three cases – that is, every third case from the sampling frame. Unfortunately, your calculation will usually result in a more complicated fraction. In these instances it is normally acceptable to round your population down to the nearest 10 (or 100) and to increase your minimum sample size until a simpler sampling fraction can be calculated.



Box 7.7

Focus on student research

Systematic random sampling

Stefan worked as a receptionist in a dental surgery with approximately 1,500 patients. He wished to find out their attitudes to the new automated appointments scheme. As there was insufficient time and money to collect data from all patients using a questionnaire he decided to send the questionnaire to a sample. The calculation of sample size revealed that to obtain acceptable levels of confidence and accuracy he needed an actual sample size of approximately 300 patients. Having obtained ethical approval he generated an alphabetical list of all registered patients from the patient record system for his sampling frame and selected his sample systematically.

First, he calculated the sampling fraction:

$$\frac{300}{1500} = \frac{1}{5}$$

This meant that he needed to select every fifth patient from the sampling frame. Next, he used a random number to decide where to start on his sampling frame. As the sampling fraction was $1/5$, the starting point had to be one of the first five patients. He therefore selected a one-digit random number between 1 and 5.

Once he had selected his first patient at random he continued to select every fifth patient until he had gone right through his sampling frame (the list of patients). If the random number Stefan had selected was 2, then he would have selected the following patient numbers:

2 7 12 17 22 27 32 37 . . .

and so on until 300 patients had been selected.

Stratified random sampling

Stratified random sampling

Stratified random sampling is a modification of random sampling in which you divide the target population into two or more relevant and significant strata based on one or a number of attributes.

In effect, your sampling frame is divided into a number of subsets.
A random sample (simple or systematic) is then drawn from each of the strata.

Consequently, stratified random sampling shares many of the advantages and disadvantages of simple random or systematic random sampling.

Stratified random sampling

Dividing the population into a series of relevant strata means that the sample is more likely to be representative, as you can ensure that each of the strata is represented proportionally within your sample.

However, it is only possible to do this if you are aware of, and can easily distinguish, significant strata in your sampling frame. In addition, the extra stage in the sampling procedure means that it is likely to take longer, to be more expensive, and to be more difficult to explain than simple random or systematic random sampling.

Stratified random sampling

In some instances, your sampling frame will already be divided into strata. A sampling frame of employee names that is in alphabetical order will automatically ensure that employees will be sampled in the correct proportion to the letter with which their name begins.

Similarly, membership lists that are ordered by date of joining will automatically result in stratification by length of membership if systematic random sampling is used.

Stratified random sampling

However, if you are using simple random sampling or your sampling frame contains periodic patterns, you will need to stratify it. To do this you:

- 1 Choose the stratification variable or variables.
- 2 Divide the sampling frame into the discrete strata.
- 3 Number each of the cases within each stratum with a unique number, as discussed earlier.
- 4 Select your sample using either simple random or systematic random sampling, as discussed earlier.



Box 7.8 Focus on student research

Stratified random sampling

Dilek worked for a major supplier of office supplies to public and private organisations. As part of her research into her organisation's customers, she needed to ensure that both public- and private-sector organisations were represented correctly. An important

stratum was, therefore, the sector of the organisation. Her sampling frame was therefore divided into two discrete strata: public sector and private sector. Within each stratum, the individual cases were then numbered (see below).

She decided to select a systematic random sample. A sampling fraction of 1/4 meant that she needed to select every fourth customer on the list. As indicated by the ticks (✓), random numbers were generated to select the first case in the public sector (2) and private sector (4) strata. Subsequently, every fourth customer in each stratum was selected.

Public sector stratum			Private sector stratum		
Number	Customer	Selected	Number	Customer	Selected
1	Anyshire County Council		1	ABC Automotive manufacturer	
2	Anyshire Hospital Trust	✓	2	Anytown printers and bookbinders	
3	Newshire Army Training Barracks		3	Benjamin Toy Company	
4	Newshire Police Force		4	Jane's Internet Flower shop	✓
5	Newshire Housing		5	Multimedia productions	
6	St Peter's Secondary School	✓	6	Roger's Consulting	
7	University of Anytown		7	The Paperless Office	
8	West Anyshire Council		8	U-need-us Ltd	✓

Cluster sampling

Cluster sampling (sometimes known as one-stage cluster sampling) is, on the surface, similar to stratified random sampling as you need to divide the target population into discrete groups prior to sampling.

The groups are termed clusters in this form of sampling and can be based on any naturally occurring grouping. For example, you could group your data by type of manufacturing firm or geographical area.

Cluster sampling

For cluster sampling, your sampling frame is the complete list of clusters rather than a complete list of individual cases within the population. You then select a few clusters, normally using simple random sampling. Data are then collected from every case within the selected clusters. The technique has three main stages:

- 1 Choose the cluster grouping for your sampling frame.
- 2 Number each of the clusters with a unique number. The first cluster is numbered 1, the second 2 and so on.
- 3 Select your sample of clusters using some form of random sampling, as discussed earlier.



Box 7.9 **Focus on student research**

Cluster sampling

Ceri needed to select a sample of firms from which to collect data using an interviewer completed face-to-face questionnaire about the use of large multi-purpose digital printer copiers. As she had limited resources with which to pay for travel and other associated data collection costs, she decided to collect

data from firms in four geographical areas selected from a cluster grouping of local administrative areas. A list of all local administrative areas formed her sampling frame. Each of the local administrative areas (clusters) was given a unique number, the first being 1, the second 2 and so on. The four sample clusters were selected from this sampling frame of local administrative areas using simple random sampling.

Ceri's sample was all firms within the selected clusters. She decided that the appropriate directories could probably provide a suitable list of all firms in each cluster.

Non-probability sampling

Within entrepreneurship research, selecting probability sample may either not be possible (where you do not have a sampling frame) or not be appropriate to answering your research question. This means your sample must be selected some other way.

Non-probability sampling (or non-random sampling) provides a range of alternative techniques to select samples, the majority of which include an element of subjective judgement.

Deciding on a suitable sample size

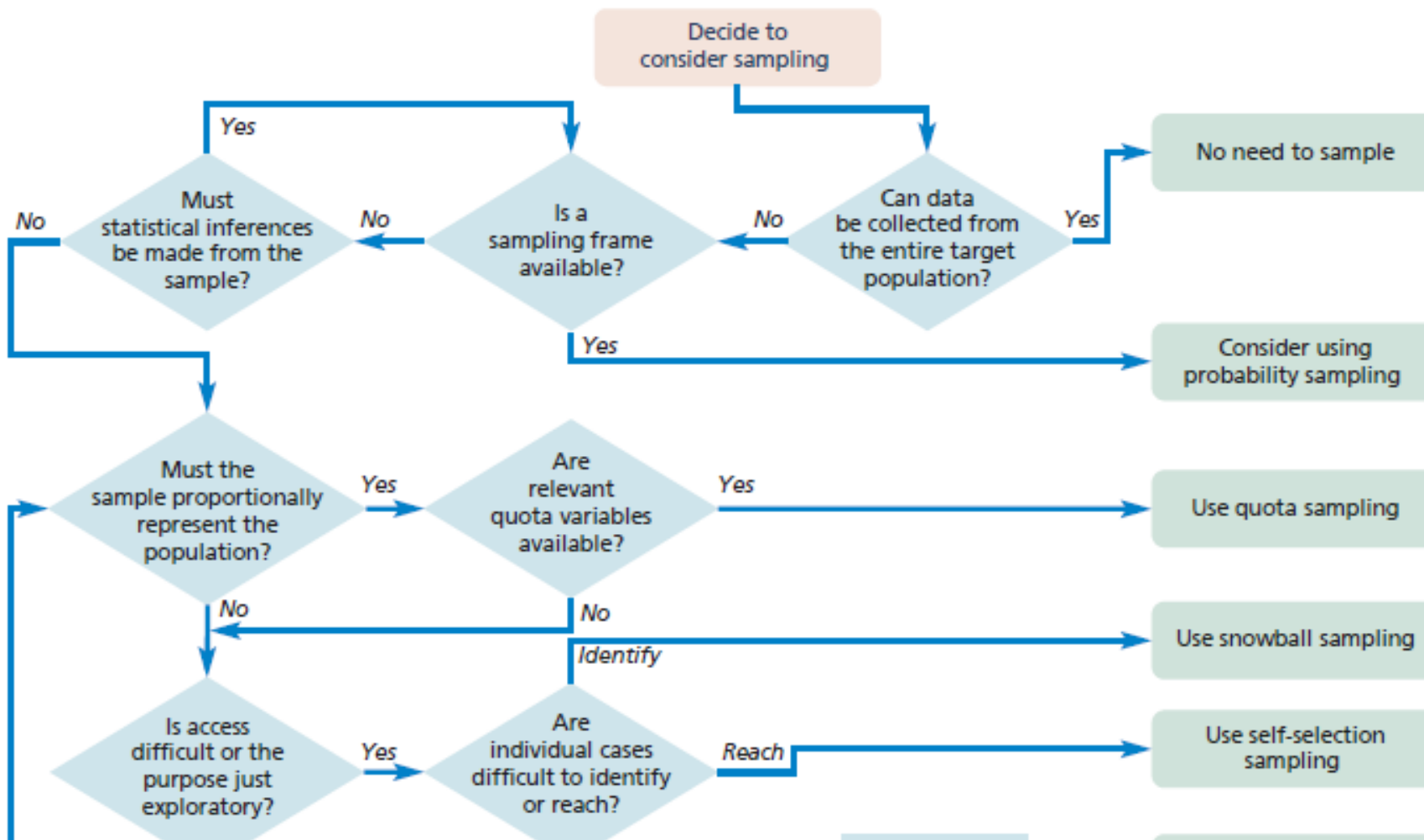
For all non-probability sampling techniques, other than for quota samples (which we discuss later), the issue of sample size is ambiguous and, unlike probability sampling, there are no rules. Rather the logical relationship between your sample selection technique and the purpose and focus of your research is important.

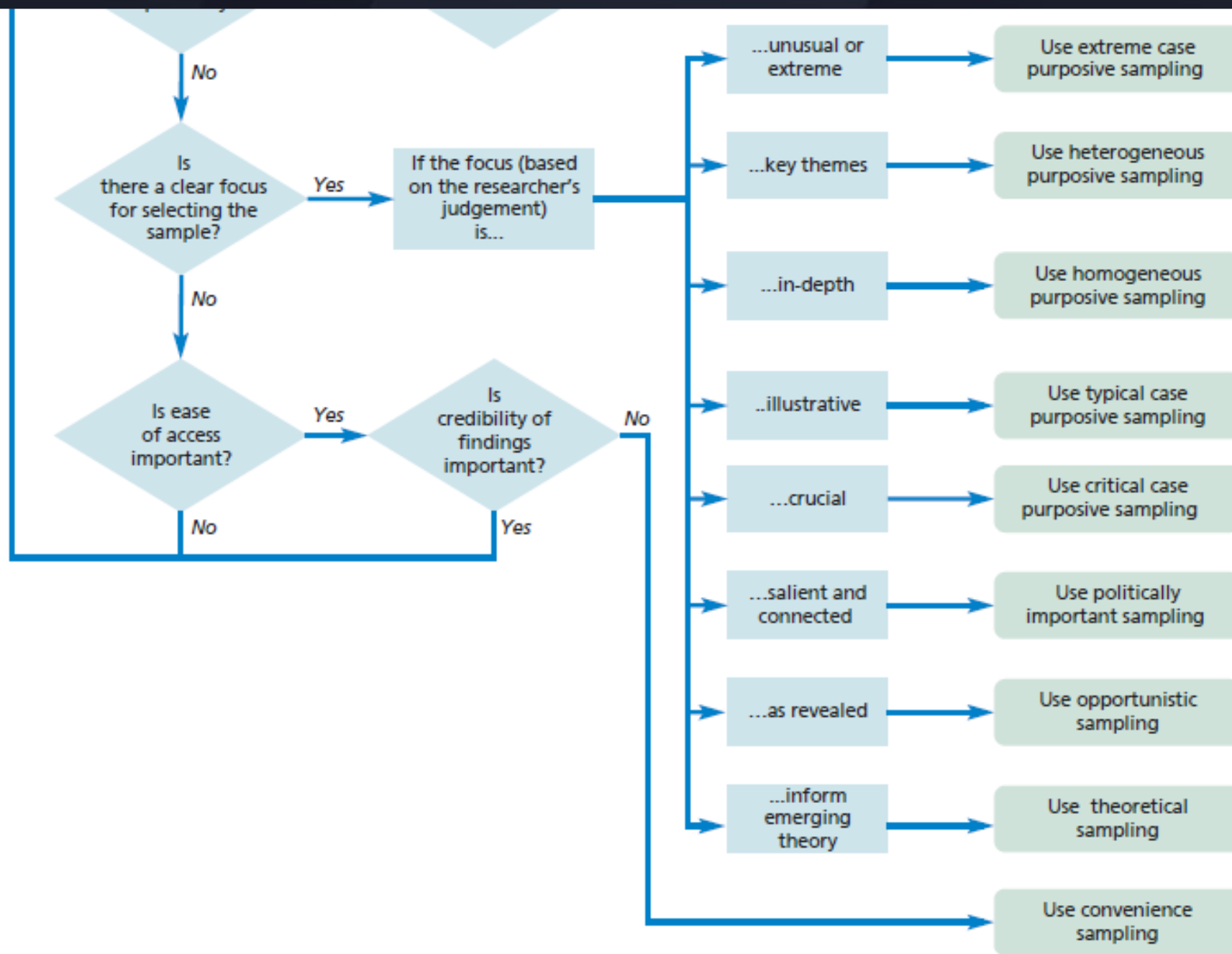
The sample selected being used to illustrate a particular aspect or to make generalizations to theory rather than about a population.

Deciding on a suitable sample size

Your sample size is therefore dependent on your research question(s) and objectives – in particular, what you need to find out, what will be useful, what will have credibility and what can be done within your available resources.

This is particularly so where you are intending to collect qualitative data using participant observation, semi-structured or unstructured interviews.





Selecting the most appropriate sampling technique and the sample

Quota sampling

Quota sampling is entirely non-random and is often used as an alternative to probability sampling for Internet and interviewer completed questionnaires as part of a survey strategy.

It is based on the premise that your sample will represent the target population as the variability in your sample for various quota variables is the same as that in the target population. However, this depends on the appropriateness of the assumptions on which the quota are based and having high quality data.

Selecting the most appropriate sampling technique and the sample

Quota sampling has a number of advantages over the probability sampling techniques. In particular, it is less costly and can be set up very quickly. In addition, it does not require a sampling frame and therefore may be the only technique you can use if one is not available.

Quota sampling is normally used for large target populations. Decisions on sample size are governed by the need to have sufficient responses in each quota to enable subsequent statistical analyses to be undertaken.

Selecting the most appropriate sampling technique and the sample

This often necessitates a sample size of between 2,000 and 5,000. Calculations of quotas are based on relevant and available data and are usually relative to the proportions in which they occur in the population.

Without sensible and relevant quotas, data collected may be biased. For many market research projects, quotas are derived from census data. Your choice of quota is dependent on two main factors:

- usefulness as a means of stratifying the data;
- ability to overcome likely variations between groups in their availability for interview.



Box 7.10 Focus on student research

Devising a quota sample

Paolo was undertaking the data collection for his dissertation as part of his full-time employment. For his research his employer had agreed to pay an online panel company to distribute his questionnaire to a sample of people representing those aged 16–74 who were either economically active or inactive. No sampling frame was available. Once the data had been collected, he was going to disaggregate his findings into subgroups dependent on gender and whether they were economically active or economically inactive. Previous research had suggested that whether or not people were retired would also have an impact

on responses and so he needed to make sure that those interviewed in each group also reflected these people. Fortunately, his country's national census of population contained a breakdown of the number of people who were economically active and inactive, their employment status and gender. These formed the basis of the categories for his quotas:

Gender	×	economic activity	×	employment status
male, female		active, inactive		part-time employee, full-time employee, self-employed, unemployed, full-time student, retired, student, looking after home or family, long-term sick or disabled, other





Box 7.10 Focus on student research (*continued*)

As he was going to analyse the data for economic activity and gender, it was important that each of

these four groups (male and economically active, male and economically inactive, female and economically active, female and economically inactive) had sufficient respondents (at least 30) to enable meaningful statistical analyses. Paolo calculated that a 0.00001 per cent quota (1 in 100,000) would provide sufficient numbers in each of these four groups. This gave him the following quotas:

Gender	Economic activity	Employment status	Population	Quota
Male	Active	Part-time employee	1 175 518	12
		Full-time employee	9 013 615	90
		Self-employed	2 670 662	27
		Unemployed	1 015 551	10
		Full-time student	619 267	6
	Inactive	Retired	2 270 916	22
		Student	1 148 356	11
		Looking after home or family	156 757	2
		Long-term sick or disabled	823 553	8
		Other	385 357	4
Female	Active	Part-time employee	4 158 750	42
		Full-time employee	6 002 949	60
		Self-employed	1 122 970	11
		Unemployed	687 296	7
		Full-time student	717 556	7
	Inactive	Retired	3 049 775	30
		Student	1 107 475	11
		Looking after home or family	1 538 377	15
		Long-term sick or disabled	750 581	8
		Other	467 093	5
Total			38 882 374	388

Purposive sampling

With **purposive sampling** you need to use your judgement to select cases that will best enable you to answer your research question(s) and to meet your objectives. For this reason it is sometimes known as judgmental sampling. You therefore need to think carefully about the impact of your decision to include or exclude cases on the research when selecting a sample in this way.

Purposive sampling is often used when working with very small samples such as in case study research and when you wish to select cases that are particularly informative. A particular form of purposive sampling, theoretical sampling, is used by researchers adopting the Grounded Theory strategy.

Purposive sampling

Purposive samples cannot be considered to be statistically representative of the target population. The logic on which you base your strategy for selecting cases for a purposive sample should be dependent on your research question(s) and objectives.

Extreme case or deviant sampling focuses on unusual or special cases on the basis that the data collected about these unusual or extreme outcomes will enable you to learn the most and to answer your research question(s) and meet your objectives most effectively. This is often based on the premise that findings from extreme cases will be relevant in understanding or explaining more typical cases.



Box 7.11 Focus on management research

Extreme case sampling

In their 2014 *Academy of Management Journal* article 'It's not easy being green: The role of self evaluations in explaining the support of environmental issues', Sonenshein, DeCelles and Dutton outline their mixed methods approach comprising an initial inductive qualitative study followed by a quantitative observational study. In the first study they develop theory regarding how environmental supporters evaluate themselves both positively and negatively and how these evaluations are shaped on an ongoing basis by work, home and other contexts. In the second study, using observational data, they derive three distinct profiles of environmental supporters and relate these profiles to environmental issue supportive behaviours.

The sample for the first qualitative study was drawn from a degree programme at a North American university called the 'Environment and Business Program'. This programme was designed to develop sustainability-orientated leaders who could also act as change agents. Sonenshein et al. (2014) argue that

although these people were clearly different from the population, this sample was important for developing theory as it allowed the researchers to learn from a non-typical group of people who had taken steps to learn about how to address climate change.

Using contact details provided by the degree programme for 25 current and 25 past students selected at random, individuals specifically interested in climate change were asked if they would be willing to take part in the research. Twenty-nine (14 current students and 15 past students) agreed to participate, identifying themselves as climate change issue supporters. Each of these people were interviewed for approximately an hour, each interview being transcribed in full.

Following analysis of data from the qualitative study, the second quantitative study collected data from two independent samples of environmental issues supporters in a large North American city. Participants were recruited by contacting the leaders of 21 groups that described themselves as active in environmental issues. Nineteen of these groups' leaders agreed to forward information about the research to their members with a link to a secure website through which they could sign up to take part in the research. In all, 91 people who were active members of environmental groups agreed to take part, comprising a second extreme case sample.

Purposive sampling

Heterogeneous or maximum variation sampling uses your judgement to choose participants with sufficiently diverse characteristics to provide the maximum variation possible in the data collected. It enables you to collect data to describe and explain the key themes that can be observed. Although this might appear a contradiction, as a small sample may contain cases that are completely different, it is argued that this is in fact a strength.

To ensure maximum variation within a sample, it is suggested that you identify your diverse characteristics (sample selection criteria) prior to selecting your sample. In direct contrast to heterogeneous sampling, homogeneous sampling focuses on one particular subgroup in which all the sample members are similar, such as a particular occupation or level in an organization's hierarchy.

Purposive sampling

Typical case sampling is usually used as part of a research project to provide an illustrative profile using a representative case. Such a sample enables you to provide an illustration of what is 'typical' to those who will be reading your research report and may be unfamiliar with the subject matter. It is not intended to be definitive.

In contrast, **critical case sampling** selects critical cases on the basis that they can make a point dramatically or because they are important. The focus of data collection is to understand what is happening in each critical case so that logical generalizations can be made.

Purposive sampling

Politically important sampling relies on your judgement regarding anticipated politically sensitive issues and associated outcomes when deciding whether to include one or a number of prominent potential participants. Consequently you choose to include (or exclude) participants on the basis of their connections with politically sensitive issues.

Opportunistic sampling acknowledges how, particularly within qualitative research involving inductive theory building, unforeseen opportunities can occur. For example, new potential research participants may emerge requiring an on-the-spot decision about their fit with the research and their inclusion. As such it relies on you using your judgment as to recognize such opportunities and assess whether or not to take them

Purposive sampling

Theoretical sampling is a special case of purposive sampling, being particularly associated with Grounded Theory and analytic induction. Initially, you need to have some idea of where to sample, although not necessarily what to sample for, participants being chosen as they are needed.

Subsequent sample selection is dictated by the needs of the emerging theory and the evolving storyline, your participants being chosen purposively to inform this. A theoretical sample is therefore cumulatively chosen according to developing categories and emerging theory based upon your simultaneous collecting, coding and analysis of the data.

Volunteer sampling

Snowball sampling is the first of two techniques we look at where participants volunteer to be part of the research rather than being chosen. It is used commonly when it is difficult to identify members of the desired population; for example, people who are working while claiming unemployment benefit. You, therefore, need to:

- 1 Make contact with one or two cases.
- 2 Ask these cases to identify further cases.
- 3 Ask these new cases to identify further new cases (and so on).
- 4 Stop when either no new cases are given or the sample is as large as is manageable or data saturation has been reached.

Volunteer sampling

The main problem is making initial contact. Once you have done this, these cases identify further members of the population, who then identify further members, and so the sample grows like a snowball being rolled in snow. For such samples the problems of bias are huge, as respondents are most likely to identify other potential respondents who are similar to themselves, resulting in a homogeneous sample.

The next problem is to find these new cases. However, for populations that are difficult to identify, snowball sampling may provide the only possibility. A development of snowball sampling is **respondent driven sampling** (RDS). This combines snowball sampling with the use of coupons or some other method to track the identification of further cases and statistical modelling to compensate for the sample being collected in a non-random way. This can enable researchers to make unbiased estimates of their target population.

Volunteer sampling

Self-selection sampling occurs when you allow each case, usually individuals, to identify their desire to take part in the research. You therefore:

- 1 Publicise your need for cases, either by advertising through appropriate media or by asking them to take part.
- 2 Collect data from those who respond.

Cases that self-select often do so because of their strong feelings or opinions about the research question(s) or stated objectives. In some instances, this is exactly what the researcher requires to answer her or his research question and meet the objectives.



Box 7.12

Focus on student research

Self-selection sampling

Siân's research was concerned with the impact of student loans on studying habits. She had decided

to distribute her questionnaire using the Internet. She publicised her research on Facebook in a number of groups' pages, using the associated description to invite people to self-select and click on the link to the questionnaire. Those who self-selected by clicking on the hyperlink were automatically taken to the Web questionnaire she had developed using the SurveyMonkey.com online survey software.

Haphazard sampling

Haphazard sampling occurs when sample cases are selected without any obvious principles of organization in relation to your research question, the most common form being **convenience sampling** (also known as **availability sampling**). This involves selecting cases haphazardly only because they are easily available (or most convenient) to obtain for your sample, such as the person interviewed at random in a shopping center.

Although convenience sampling is used widely, it is prone to many sources of bias and influences that are beyond your control. Findings from convenience samples may be given very little credibility

Multi-stage sampling

Multi-stage sampling refers to any sampling design that occurs in two or more successive stages using either probability, non-probability, or both types of sample selection techniques.

For example, in the first stage you might select two organizations using critical case purposive sampling. Subsequently, in the second stage, you might select a sample of employees from each organization using stratified random sampling, thereby combining non-probability with probability sampling.

Multi-stage sampling

Alternatively, you may first select a large sample of customers or organizations using quota sampling. Subsequently, based on analysis of the data collected from these customers or organizations, you may select a smaller heterogeneous purposive sample to illustrate the key themes.

Multi-stage sampling can also use cluster sampling to overcome problems associated with a geographically dispersed population when face-to-face contact is needed, or when it is expensive and time consuming to construct a sampling frame for a large geographical area.



Box 7.13

Focus on student research

Multi-stage sampling

Laura worked for a market research organisation that needed her to interview a sample of 400 households in England and Wales. She decided to use the electoral register as a sampling frame. Laura knew that selecting 400 households using either systematic or simple random sampling was likely to result in these 400 households being dispersed throughout England and Wales, resulting in considerable amounts of time spent travelling between interviewees as well as high travel costs. By using multi-stage sampling Laura felt these problems could be overcome.

In her first stage the geographical area (England and Wales) was split into discrete sub-areas (counties). These formed her sampling frame. After numbering

all the counties, Laura selected a small number of counties at random using cluster sampling. Since each case (household) was located in a county, each had an equal chance of being selected for the final sample.

As the counties selected were still too geographically large, each was subdivided into smaller geographically discrete areas (electoral wards). These formed the next sampling frame (stage 2). Laura selected another sample at random. This time she selected a larger number of wards using simple random sampling to allow for likely important variations in the nature of households between wards.

A sampling frame of the households in each of these wards was then generated. Laura purchased copies of the edited electoral register from the relevant local authorities. These contained the names and addresses of people who had registered to vote and had not 'opted out' of allowing their details to be made widely available for others to use. Laura finally selected the actual cases (households) that she would interview using systematic random sampling.



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY
BY EVERYONE'S SMART UNIVERSITY



Thank You

