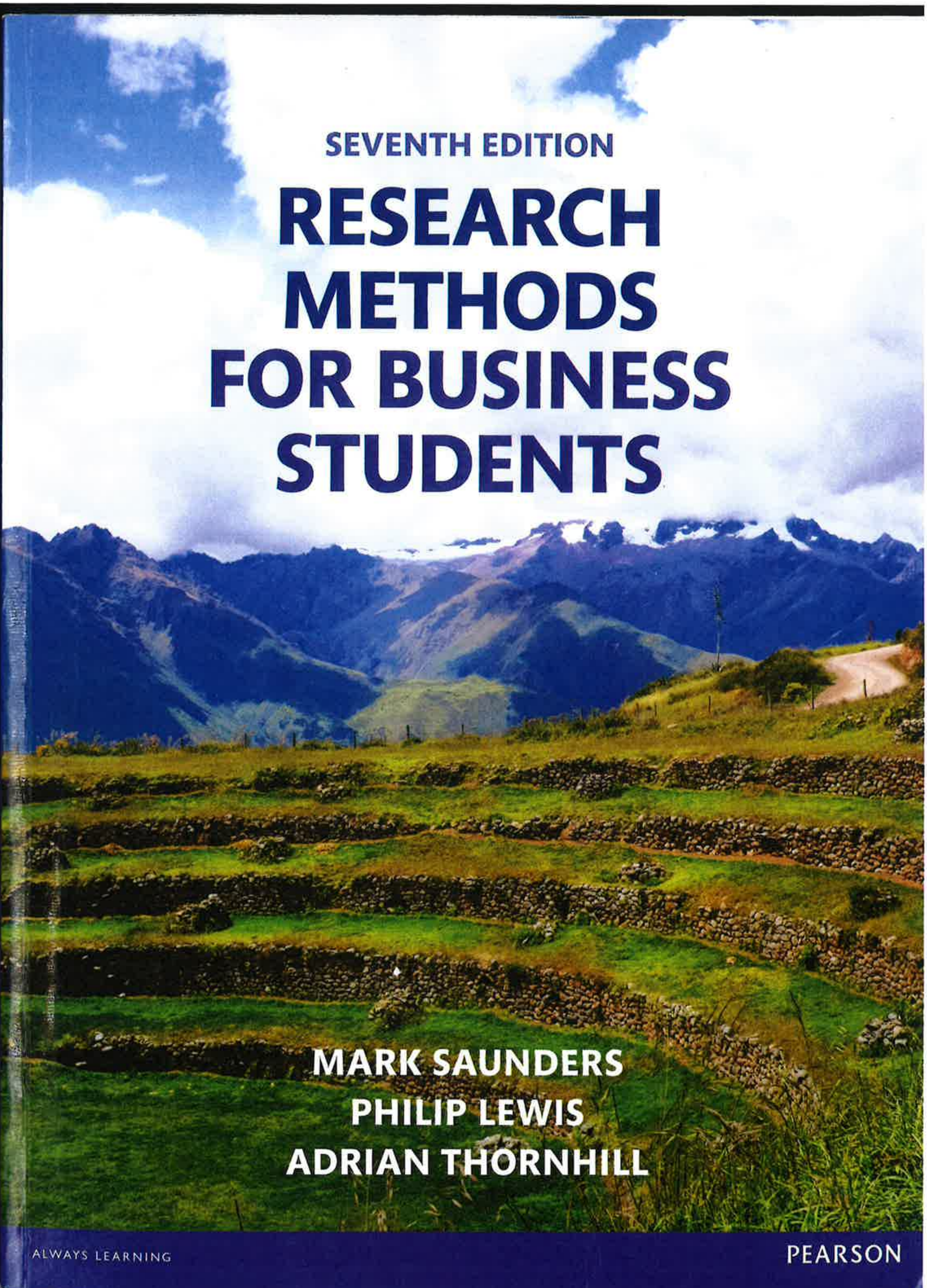

SEVENTH EDITION

RESEARCH METHODS FOR BUSINESS STUDENTS

MARK SAUNDERS
PHILIP LEWIS
ADRIAN THORNHILL

ALWAYS LEARNING

PEARSON

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England
and Associated Companies throughout the world

Visit us on the World Wide Web at:

www.pearson.com/uk

First published under the Pitman Publishing imprint in 1997
Second edition 2000
Third edition 2003
Fourth edition 2007
Fifth edition 2009
Sixth edition 2012
Seventh edition 2016

ISBN: 978-1-292-01662-7

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

Saunders, Mark, 1959-
Research methods for business students / Mark Saunders, Philip Lewis, Adrian Thornhill.
— Seventh edition.
pages cm
Revised edition of Research methods for business students, 2012.
ISBN 978-1-292-01662-7
1. Business--Research. 2. Business--Research--Data processing. I. Lewis, Philip, 1945- II. Thornhill, Adrian. III. Title.
HD30.4.S28 2015
650.072--dc23
2015019451

12 11 10 9 8 7 6 5 4 3
19 18 17 16

Typeset in 9.5/12 ITC Slimbach Std by 76

Printed and bound by L.E.G.O. S.p.A., Italy

The publisher's policy is to use paper manufactured from sustainable forests.

Chapter 7

Selecting samples

Learning outcomes

By the end of this chapter you should:

- understand the need for sampling in business and management 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 select appropriate sampling techniques for a variety of research scenarios and be able to justify their selection;
- be able to use a range of sampling techniques;
- be able to assess the representativeness of the sample selected;
- be able to assess the extent to which it is reasonable to generalise from a sample;
- be able to apply the knowledge, skills and understanding gained to your own research project.

7.1 Introduction

Whatever your research question(s) and objectives, you will need to consider whether you need to use sampling. Occasionally, it may be possible to collect and analyse 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 analyse all the potential data available to you, owing to restrictions of time, money and often access. In the opening vignette you will see that it was not possible to use a list of all outdoor advertising posters from all times, or to obtain opinions of the entire 'general public'. 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 to generalise statistically about all the cases from which your **sample** has been selected. For example, if you asked a sample of consumers what they thought of a new healthy snack and 75 per cent said that they thought it was too expensive, you might infer that 75 per cent of all consumers felt that way. Other research questions may not involve such generalisations. To gain an understanding of how people manage their careers, you may select a sample of company chief executives. For such research your sample selection would be based on the premise that, as these people have reached executive level and have been successful in managing their own careers, they are most likely to be able to offer insights from which you can build understanding. Even if you are adopting a case study strategy using one large organisation and collecting your data using unstructured interviews, you will still need to select your case study (sample) organisation and a group (sample) of employees and managers to interview. For example, in the opening vignette, a report of findings in some UK media inferred they were based on a statistically representative public vote. Consequently, whatever your research question, an understanding of techniques for selecting samples is likely to be very important.

In selecting a sample to study, it should represent the full set of cases in a way that is meaningful and which we can justify (Becker 1998). In the opening vignette, the Outdoor Media Centre provides information regarding how its sample of 228 advertising posters was selected. This allows us to assess whether this was meaningful with regard to establishing the 100 best advertising posters of all time. It also outlines briefly how the people who voted

In February 2011 the UK's Outdoor Media Centre launched its 'Hall of Fame' competition to identify the 100 best advertising posters of all time. Working with the History of Advertising Trust, they generated a list of 500 advertising campaigns. This was reduced to a shortlist of 228 campaigns by a committee of media and creative experts together with the editor of the weekly magazine for the advertising media and communications industry – *Campaign*. These were displayed on a dedicated website, www.outdoorhalloffame.co.uk.

Creative agencies, media planners, advertisers, media owners and the general public were invited in an article in *Campaign* to go to the website, view the advertising campaigns and cast their votes for what they considered to be the best outdoor posters (Bidlake 2011). Each person was able to cast a total of 10 votes, the best advertisements being chosen after more than 10,000 reader votes had been cast (Farey-Jones 2011).

On 31 March 2011 the results were announced, being reported widely in the UK media. TBWA's 1995 'Hello Boys' billboard poster advertising Wonderbra was voted the best poster ever created. In second place was Saatchi & Saatchi's 'Labour Isn't Working' poster created for the Conservative Party in the late 1970s. The 1914 UK First World War army recruiting poster 'Lord Kitchener Wants You' by Caxton Advertising took third place. The highest ranking outdoor poster from the last decade was TBWA's 2002/3 iTunes silhouette poster for Apple.

iTunes silhouette poster for Apple

(creative agencies, media planners, advertisers, media owners and the general public) were selected, providing an indication of who the sample was: predominantly people employed in media industries. This allows us to assess whether the claim that those receiving the most votes were the 100 best advertising posters of all time is justifiable.

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 Indian restaurants throughout a country, the population from which you would select your sample would be all Indian restaurants in that country. Alternatively, you might need to establish the normal 'range' in miles that can be travelled by electric cars in everyday use produced by a particular manufacturer. Here the population would be all the electric cars in everyday use produced by that manufacturer.

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 that represents the entire population. 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.

For all research questions where it would be impracticable for you to collect data from the entire population, you need to select a sample. This is equally important whether you are planning to use interviews, questionnaires, observation or some other data collection technique. You might be able to obtain permission to collect data from only two or three organisations. Alternatively, testing an entire population of products to destruction, such as to establish the actual duration of long-life batteries, would be impractical for any manufacturer.

With other research questions it might be theoretically possible for you to collect data from the entire population but the overall cost would prevent it. It is obviously cheaper for you to collect, prepare for analysis and check data from 250 customers than from 2500, even though the cost per case for your study (in this example, customer) is likely to be higher than with a census. Your costs will be made up of new costs such as sample selection, and the fact that overhead costs such as the questionnaire, interview or observation schedule design and general preparation of data for analysis are spread over a smaller number of cases. Sampling also saves time, an important consideration when you have tight deadlines. The organisation of data collection is more manageable as fewer people are involved. As you have fewer data to prepare for analysis and then to analyse, the results will be available more quickly.

Many researchers, for example Barnett (2002), argue that using sampling makes possible a higher overall accuracy than a census. The smaller number of cases for which you need to collect data means that more time can be spent designing and piloting the means of collecting these data. Collecting data from fewer cases also means that you can collect information that is more detailed. If you are employing people to collect the data (perhaps as interviewers) you can afford higher-quality staff. You can also devote more time to trying to obtain data from more difficult to reach cases. Once your data have been collected, proportionally more time can be devoted to checking and testing the data for accuracy prior to analysis. However, one point remains crucial when selecting a sample: it must enable you to answer your research question!

The importance of defining the research population clearly

The sample selected is related to the population that is highlighted in the research question and objectives. This means that if a research question is about all owners of a particular brand of laptop, then the population is all owners of a particular brand of tablet computer, and the sample selected should be a subset of all those owners. This sample, providing it is selected carefully, will allow conclusions to be drawn about all owners of that brand of tablet. 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** (Figure 7.1). This name indicates it is this population that is the actual focus or target of the research inquiry (Kervin 1999). For example, rather than defining your population as all owners of a particular brand of tablet computers, you may redefine your target population as all owners of a particular brand of tablet who are studying for a business and management degree at one university. However, business and management students at one university are unlikely to be same as all tablet owners and even students from other universities may differ! Consequently, using a sample drawn from this target population of students to find out about all owners of a brand of tablet may result in biased or incorrect conclusions if the research is about all owners. In selecting your sample from this target population, you have narrowed the focus of your research to business and management students at a particular university who own that brand of tablet computer. We discuss this further in Sections 7.2 and 7.3.

An overview of sampling techniques

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

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

Those discussed in this chapter are highlighted in Figure 7.2. With **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

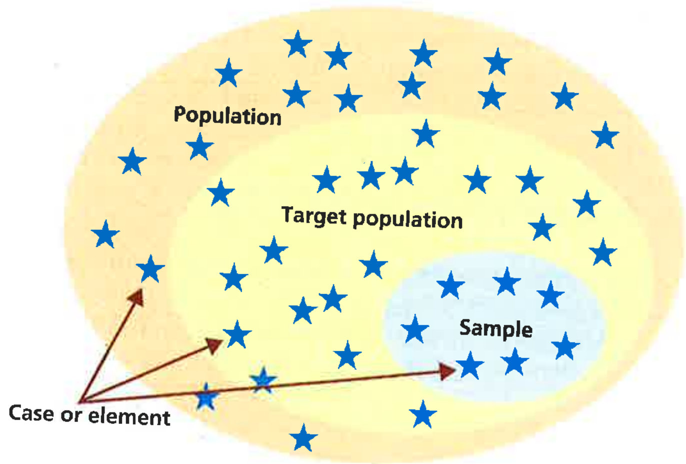

Figure 7.1 Population, target population, sample and individual cases

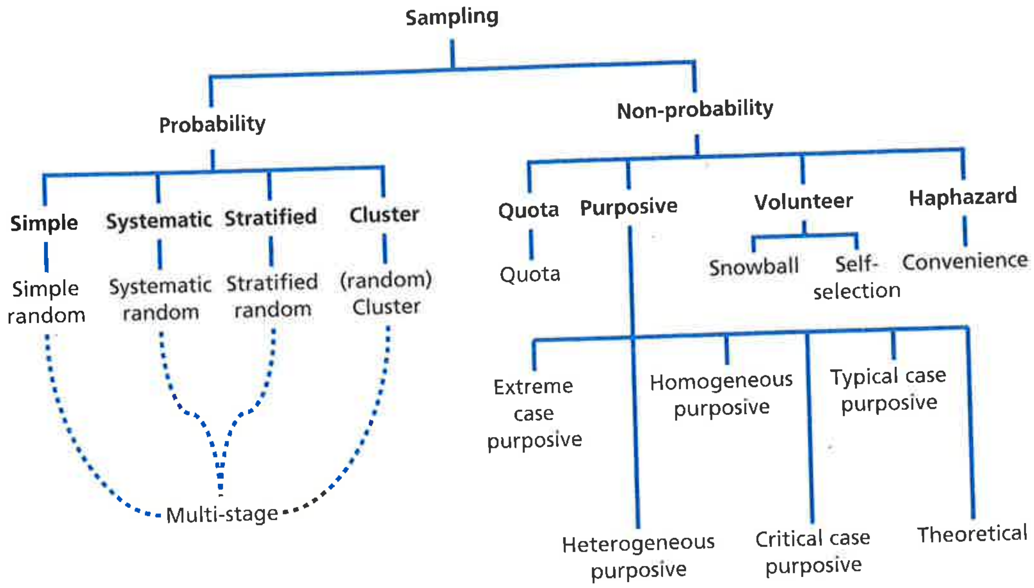

Figure 7.2 Sampling techniques

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 (Section 5.5). 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. You may still be able to generalise from non-probability samples about the target population, but not on statistical grounds. However, with both types of sample you can answer other forms of research question, such as 'What job attributes attract people to jobs?' or 'How are financial services institutions adapting their services in response to the post-2009 crash liquidity rules?'

Subsequent sections of this chapter outline the most frequently used probability (Section 7.2) and non-probability (Section 7.3) sampling techniques, discuss their advantages and disadvantages and give examples of how and when you might use them. Although each technique is discussed separately, for many research projects you will need to use a combination of sampling techniques, some projects involving both probability and non-probability sampling techniques.

7.2 Probability sampling

Probability sampling (or **representative sampling**) is associated most commonly with survey research strategies where you need to make 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.

Each of these stages will be considered in turn. However, for target populations of fewer than 50 cases Henry (1990) advises against probability sampling. He argues that you should collect data on the entire target population, as the influence of a single extreme case on subsequent statistical analyses is more pronounced than for larger samples.

Identifying a suitable sampling frame and the implications for generalisability

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 or objective is concerned with members of a student society, your sampling frame will be the complete membership list for that society. If your research question or objective is concerned with registered child-minders in a local area, your sampling frame will be the directory of all registered child-minders in this area. Alternatively, if your research question is concerned with organisations in a particular sector, you may be thinking of creating a sampling frame from an existing database of companies available at your university, such as Fame or Amadeus. You then select your sample from your list.

Obtaining a sampling frame is therefore essential if you are going to use probability sampling. However, as highlighted in research by Edwards et al. (2007), you need to be aware of the possible problems of using existing databases for your sampling frame. In their work on multinationals in Britain, they found that:

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

This emphasises 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 (Box 7.1).

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. You might decide to use a business directory as the sampling frame from which to select a sample of typical businesses. However, the business directory covers only subscribers who pay to be listed, often in one geographical area. Your sample will therefore be biased towards businesses that have chosen to subscribe. Because the directory is only updated annually, the sampling frame will be out of date ('non-current'). As some businesses choose not to subscribe, it will not be a valid representation as it does not include these businesses. This means that you will be selecting a sample of businesses that choose to subscribe at the date the directory was compiled by a particular company!

The way you define your sampling frame also has implications regarding the extent to which you can generalise from your sample. As we have already discussed, sampling is used when it is impracticable or unnecessary to collect data from the entire population. Within probability sampling, by defining the sampling frame you are defining the target population about which you want to generalise. This means that if your sampling frame is a list of all customers of an organisation, strictly speaking you can only generalise, that is apply the findings based upon your sample, to that target population. Similarly, if your sampling frame is all employees of an organisation (the list being the organisation's payroll) you can only generalise to employees of that organisation. This can create

Box 7.1 Focus on research in the news

The Art Market: Hopes pitched too high for comfort

By Georgina Adam

There is nothing new in seeing art as an asset class but a recent report by three professors of finance claims that returns have been significantly overestimated, and the risk underestimated. "Investors would be wise to be wary," they say.

The authors of the Social Science Research Network paper (available at ssrn.com/abstract=2280099) question the returns that until now have been reported as 10 percent on average annually, over the past four decades, based on the RSR (repeat sales) index. They say that the index does not take into account selection bias, and after studying the repeat sales of more than 20,500 works of art between 1972 and 2010, they conclude that the annual return on art was nearer 6.5 percent.

Selection bias means that a work of art that goes up the most in value tends to be resold more frequently, just as in the real estate market rapidly appreciating properties are traded more often. So only the most desirable works of art resell, and using their good performance to assign a value to others, via the RSR index, is misleading, say the authors, whose conclusion is severe. "When we compared the investment returns and risk of all the styles of art to a portfolio of pure stocks, we found that art investments would not substantially improve the risk/return profile of a portfolio diversified among traditional asset classes, such as stocks and bonds."

Philip Hoffman, chief executive of the Fine Art Fund, rather naturally rejects the findings of the report, as least as far as funds are concerned. "The data is highly inaccurate," he says.

"The sample is far too small. And as far as our fund is concerned, since we buy 95 percent of our works privately, this data isn't relevant." His first fund is due to be wound up at the end of 2015, and he expects its return will be between 6 and 10 percent, "which was always our aim". However, he admits the research gives "insight into what the market does overall".

FT *Source:* Extract from Adam G (2013) 'The Art Market: Hopes pitched too high for comfort', FT.com, 8 November. Available at http://www.ft.com/cms/s/2/ab0ff19a-47a0-11e3-9398-00144feabdc0.html [Accessed 10 Mar. 2015]

problems, as often we hope that our findings have wider applicability than the target population from which our sample was selected. However, even if your probability sample has been selected from one large multinational organisation, you should not claim that what you have found would also occur in similar organisations. In other words, you should not generalise beyond your sampling frame. Despite this, researchers often do make such claims, rather than placing clear limits on the generalisability of the findings.

An increasing number of organisations specialise in selling electronic lists of names, addresses and email addresses. These lists include a wide range of people such as company directors, chief executives, marketing managers, production managers and human resource managers, for public, private and non-profit-making organisations, and can be merged into standard letters such as those included with questionnaires (Section 11.4). Some organisations also specialise in delivering your questionnaire to an online 'survey panel' of potential respondents, guaranteeing you obtain a specified number of completed questionnaires for a sample tailored to your specific requirements. Because you pay for the list or completed questionnaire by the case (named individual), the organisations that provide them usually select your sample. It is therefore important to establish precisely how they will select your sample as well as obtaining an indication of the organisation database's completeness, accuracy and currency. For example, when obtaining a list of email addresses don't forget that some people change their Internet service provider and their email address regularly. This means the sampling frame is likely to under-represent this group. Usage patterns of the Internet both in organisations and at home are changing rapidly; you therefore need to ensure your intended sampling frame is relevant to your target population. If you are intending to use an online survey panel you need to establish whether or not the organisation offers panel members an incentive to encourage response and the likely implications of this for the characteristics of respondents and consequently their responses (Section 11.2). Box 7.2 provides a checklist against which to check your sampling frame.

Deciding on a suitable sample size

Generalisations 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 generalising 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 analysing the data. 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 for each cell (e.g. chi square, Section 12.5); and, to a lesser extent,
- the size of the target population from which your sample is being drawn.

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, as we discuss in Section 12.5, 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 (Lenth 2001). For many research questions and objectives, your need to undertake

Box 7.2 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, in particular is it up to date?

✔ Does the sampling frame include all cases in the target population, in other words is it complete?

✔ Does the sampling frame contain the correct information, in other words is it accurate?

✔ Does the sampling frame exclude irrelevant cases, in other words is it precise?

✔ (For purchased lists and online panels) Can you establish and control precisely how the sample will be selected?

✔ (For an online panel) Can you establish whether incentives will be used to enhance the likely response and provide an assessment of the impact of this on respondent characteristics and consequently responses?

particular statistical analyses (Section 12.5) will determine the threshold sample size for individual categories. In particular, an examination of virtually any statistics textbook (or Sections 12.3 and 12.5 of this book) will highlight that, in order to ensure spurious results do not occur, the data analysed must be normally distributed. While the normal distribution is discussed in Chapter 12, its implications for sample size need to be considered here. 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, Tennent's (2013) advice of a minimum number of 30 for statistical analyses provides a useful rule of thumb for the smallest number in each category within your overall sample. Where the population in the category is less than 30, and you wish to undertake your analysis at this level of detail, you should normally collect data from all cases in that category. Alternatively, you may have access to an expert system such as Ex-Sample™. This software calculates the minimum sample size required for different statistical analyses as well as the maximum possible sample size given resources such as time, money and response rates. In addition, it provides a report justifying the sample size calculated (Idea Works 2012).

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**.

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. Table 7.1 provides a rough 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 (full details of the calculation for minimum sample size and adjusted minimum sample size are given in Appendix 2). For most business and management 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 7.1, 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. De Vaus (2014) argues that it is for this reason that many market research companies limit their samples' sizes to approximately 2000. Unfortunately, from many samples, a 100 per cent response rate is unlikely and so your sample will need to be larger to ensure sufficient responses for the margin of error you require.

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)

	Margin of error			
Target population	**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

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. If 60 per cent of your sample were small service sector companies then, provided the sample was representative, you would expect 60 per cent of the target population to be small service sector companies. You therefore need to obtain as high a response rate as possible to reduce the risk of non-response bias and ensure your sample is representative (Groves and Peytcheva 2008). This is not to say that a low response rate will necessarily result in your sample being biased, just that it is more likely!

In reality, you are likely to have non-responses. Non-respondents are different from the rest of the target population because they have refused 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. 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.

You should therefore collect data on refusals to respond to both individual questions and entire questionnaires or interview schedules to check for bias (Section 12.2) and report this briefly in your project report. For returned questionnaires or structured interviews, the American Association for Public Opinion Research (2011) defines four levels of non-response that can be reported for questionnaires and structured interviews:

- **complete refusal**: none of the questions answered;
- **break-off**: less than 50 per cent of all questions answered other than by a refusal or no answer (this therefore includes complete refusal);

- **partial response**: 50 per cent to 80 per cent of all questions answered other than by a refusal or no answer;
- **complete response**: over 80 per cent of all questions answered other than by a refusal or no answer.

Non-response is due to four interrelated problems:

- refusal to respond;
- ineligibility to respond;
- inability to locate respondent;
- respondent located but unable to make contact.

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 non-response can be minimised by paying careful attention to the methods used to collect your data (Chapters 9, 10 and 11). 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**. Neuman (2014) suggests that when you calculate this you should include all eligible respondents:

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

This he calls the **total response rate**. A more common way of doing this excludes ineligible respondents and those who, despite repeated attempts (Sections 10.3 and 11.5), were unreachable. This is known as the **active response rate**:

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

An example of the calculation of both the total response rate and the active response rate is given in Box 7.3.

Even after ineligible and unreachable respondents have been excluded, it is probable that you will still have some non-responses. You therefore need to be able to assess how

Box 7.3 Focus on student research

Calculation of total and active response rates

Ming had decided to administer a telephone questionnaire to people who had left his company's employment over the past five years. He obtained a list of the 1034 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.

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

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

representative your data are and to allow for the impact of non-response in your calculations of sample size. These issues are explored in subsequent sections.

Estimating response rates and actual sample size required

With all probability samples, it is important that your sample size is large enough to provide you with the necessary confidence in your data. The margin of error must be within acceptable limits, and you must ensure that you will be able to undertake your analysis at the level of detail required. You therefore need to estimate the likely response rate – that is, the proportion of cases from your sample who will respond or from which data will be collected – and increase the sample size accordingly. Once you have an estimate of the likely response rate and the minimum or the adjusted minimum sample size, 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 7.1 or Appendix 2),
$re\%$ is the estimated response rate expressed as a percentage.

This calculation is shown in Box 7.4.

If you are collecting your sample data from a secondary source (Section 8.2) within an organisation 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. Alternatively, you can err on the side

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:

$$n^a = \frac{439 \times 100}{30}$$

$$= \frac{43900}{30}$$

$$= 1463$$

Jan's actual sample, therefore, needed to be 1463 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.

Box 7.5 Focus on management research

Reporting response rates

In their 2008 Human Relations paper 'Survey responses rates: Levels and trends in organizational research', Baruch and Holtom offer useful advice regarding reporting responses rates from questionnaires. Within this they stress that authors should make it clear whether their questionnaire was administered (in other words respondents filled it in as part of their job, role or studies) or truly voluntary. They also offer a checklist for authors (and editors), which covers information that should be included about the sample and the questionnaires returned. In particular:

- ✔ Number of respondents to whom the questionnaire was sent.
- ✔ How the questionnaire was distributed.
- ✔ Whether prior consent was obtained from respondents.
- ✔ The number of questionnaires returned.
- ✔ Of those returned, the numbers that were useable.
- ✔ Reasons (if known) for questionnaires not being useable.
- ✔ Where different populations received a questionnaire, differences (if any) in response rates.
- ✔ Techniques (if any) used to increase response rates.
- ✔ Where response rates differ from likely norms, possible reasons for this.

of caution. For most academic studies involving individuals or organisations' representatives, response rates of approximately 50 per cent and 35 to 40 per cent respectively are reasonable (Baruch and Holtom 2008).

However, beware: response rates can vary considerably when collecting primary data. Neuman (2014) suggests response rates of between 10 and 50 per cent for postal questionnaire surveys and up to 90 per cent for face-to-face interviews. The former rate concurs with a questionnaire survey we undertook for a multinational organisation that had an overall response rate of 52 per cent. In our survey, response rates for individual sites varied from 41 to 100 per cent, again emphasising variability. Our examination of response rates to recent business surveys reveals rates as low as 10–20 per cent for Web and postal questionnaires, an implication being that respondents' questionnaire fatigue was a contributory factor! With regard to telephone questionnaires, response rates have fallen from 36 per cent to less than 9 per cent, due in part to people using answering services to screen calls (Dillman et al. 2014). Fortunately a number of different techniques, depending on your data collection method, can be used to enhance your response rate. These are discussed with the data collection method in the appropriate sections (Sections 10.3 and 11.5).

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. Five main techniques can be used to select a probability sample (Figure 7.3):

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

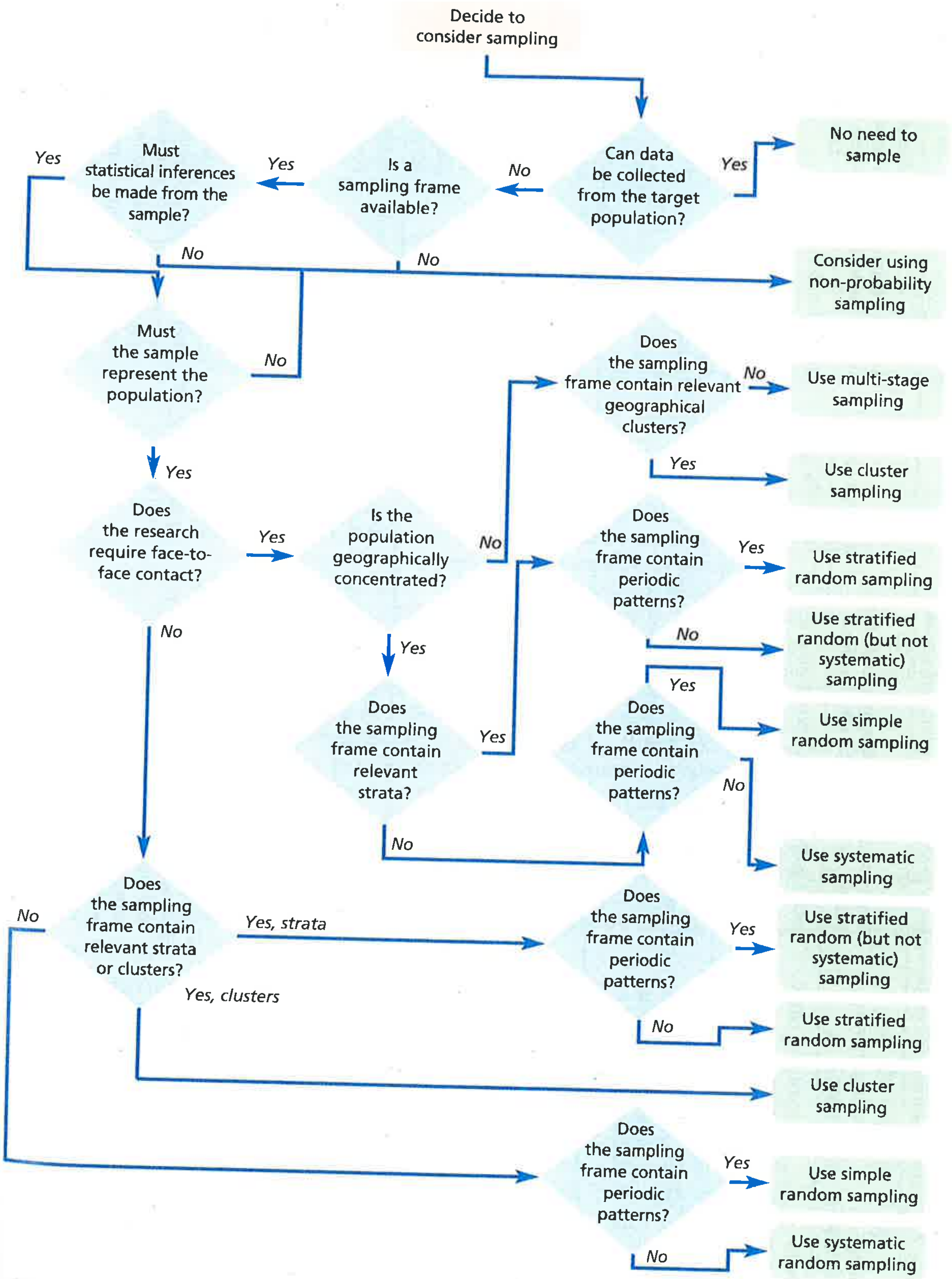

Figure 7.3 Choosing a probability sampling technique

Note: Simple random sampling ideally requires a sample size of over a few hundred

Table 7.2 Impact of various factors on choice of probability sampling techniques

Sample technique	Sampling frame required	Size of sample needed	Geographical area to which suited	Relative cost	Easy to explain to support workers?	Advantages compared with simple random
Simple random	Accurate and easily accessible	Better with over a few hundred	Concentrated if face-to-face contact required, otherwise does not matter	High if large sample size or sampling frame not computerised	Relatively difficult to explain	
Systematic random	Accurate, easily accessible and not containing periodic patterns. Actual list not always needed	Suitable for all sizes	Concentrated if face-to-face contact required, otherwise does not matter	Low	Relatively easy to explain	Normally no difference
Stratified random	Accurate, easily accessible, divisible into relevant strata (see comments for simple random and systematic random as appropriate)	See comments for simple random and systematic random as appropriate	Concentrated if face-to-face contact required, otherwise does not matter	Low, provided that lists of relevant strata available	Relatively difficult to explain (once strata decided, see comments for simple random and systematic random as appropriate)	Better comparison and hence representation across strata. Differential response rates may necessitate reweighting
Cluster	Accurate, easily accessible, relates to relevant clusters, not individual population members	As large as practicable	Dispersed if face-to-face contact required and geographically based clusters used	Low, provided that lists of relevant clusters available	Relatively difficult to explain until clusters selected	Quick but reduced precision
Multi-stage	Initial stages: geographical. Final stage: needed only for geographical areas selected (see comments for simple random and systematic random as appropriate)	Initial stages: as large as practicable. Final stage: see comments for simple random and systematic random as appropriate	Dispersed if face-to-face contact required, otherwise no need to use this technique!	Low, as sampling frame for target population required only for final stage	Initial stages: relatively difficult to explain. Final stage: see comments for simple random and systematic random as appropriate	Difficult to adjust for differential response rates. Substantial errors possible! However, often only practical approach when sampling a large complicated target population

Source: © Mark Saunders, Philip Lewis and Adrian Thornhill 2015

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 (Figure 7.3). The structure of the sampling frame, the size of sample you need and, if you are using support workers, the ease with which the technique may be explained will also influence your decision. The impact of each of these is summarised in Table 7.2.

Simple random sampling

Simple random sampling (sometimes called just **random sampling**) involves you selecting the sample at random from the sampling frame using either a computer or random number tables (Appendix 3). To do this you:

1. Number each of the cases in your sampling frame with a unique number. The first case is numbered 0, the second 1 and so on.
2. Select cases using random numbers (e.g. Table 7.3, Appendix 3) until your actual sample size is reached.

If using random number tables you must select your first random number at random (closing your eyes and pointing with your finger is one way!) as this ensures that the set of random numbers obtained for different samples is unlikely to be the same. If you do not, you will obtain sets of numbers that are random but identical.

Starting with this number, you read off the random numbers (and select the cases or elements) in a regular and systematic manner until your sample size is reached. If the same number is read off a second time it must be disregarded as you need different cases. This means that you are not putting each case's number back into the sampling frame after it has been selected and is termed 'sampling without replacement'. If a number is selected that is outside the range of those in your sampling frame, you simply ignore it and continue reading off numbers until your sample size is reached (Box 7.6).

If you are using an online random number generator or spreadsheet to generate random numbers, you must ensure that the numbers generated are within your range and that if a number is repeated it is ignored and replaced. If details of your sampling frame are stored on the computer it is possible to generate a sample of randomly selected cases. For telephone interviews, many market research companies now use computer-aided telephone interviewing (CATI) software to select and dial telephone numbers at random from an existing database or random digit dialling and to contact each respondent in turn.

Table 7.3 Extract from random number tables

78	41	11	62	72	18	66	69	58	71	31	90	51	36	78	09	41	00
70	50	58	19	68	26	75	69	04	00	25	29	16	72	35	73	55	85
32	78	14	47	**01**	55	10	91	83	21	13	32	59	53	03	38	79	32
71	60	20	53	86	78	50	57	42	30	73	48	68	09	16	35	21	87
35	30	15	57	99	96	33	25	56	43	65	67	51	45	37	99	54	89
09	08	05	41	66	54	01	49	97	34	38	85	85	23	34	62	60	58
02	59	34	51	98	71	31	54	28	85	23	84	49	07	33	71	17	88
20	13	44	15	22	95												

Source: Appendix 3

Box 7.6
Focus on student research

Simple random sampling

Jemma was undertaking her work placement at a large supermarket, where 5011 of the supermarket's customers used the supermarket's Internet purchase and delivery scheme. She was asked to interview customers and find out why they used this scheme. As there was insufficient time to interview all of them she decided to interview a sample using the telephone. Her calculations revealed that to obtain acceptable levels of confidence and accuracy she needed an actual sample size of approximately 360 customers. She decided to select them using simple random sampling.

Having obtained a list of Internet customers and their telephone numbers, Jemma gave each of the cases (customers) in this sampling frame a unique number. In order that each number was made up in exactly the same way she used 5011 four-digit numbers starting with 0000 through to 5010. So customer 677 was given the number 0676.

The first random number she selected was 01 (shown in bold and shaded in Table 7.3). Starting with this number she read off the two-digit random numbers in a regular and systematic manner (in this example continuing along the line):

01 55 10 91 83 21 13 32 59 53 03 38
79 32 71 60 20 . . .

Jemma combined the first two-digit random number (01) with the second (55) to get her first four-digit random number 0155. She continued combining random numbers in this manner until 360 different cases had been selected. These formed her random sample. Numbers selected that were outside the range of those in her sampling frame (such as 8321, 5953 and 7932) were simply ignored.

Random numbers allow you to select your sample without bias. The sample selected, therefore, can be said to be representative of the target population. However, it is not a perfect miniature replica of this population, since it still possesses sampling error. In addition, the selection that simple random sampling provides is more evenly dispersed throughout the target population for samples of more than a few hundred cases. The first few hundred cases selected using simple random sampling normally consist of groups of cases whose numbers are close together followed by a gap and then a further grouping. For more than a few hundred cases, this pattern occurs far less frequently. Because of the technique's random nature it is possible that a chance occurrence of such patterns will result in certain parts of a population being over- or under-represented.

Simple random sampling is best used when you have an accurate and easily accessible sampling frame that lists the target population, preferably in electronic format. While you can often obtain these for employees within organisations or members of clubs or societies, adequate lists are less likely to be available for organisations. If your population covers a large geographical area, random selection means that selected cases are likely to be dispersed throughout the area. Consequently, this form of sampling is not suitable if collecting data over a large geographical area using a method that requires face-to-face contact, owing to the associated high travel costs. Simple random sampling would still be suitable for a geographically dispersed area if you used an alternative technique of collecting data such as web or postal questionnaires or telephone interviewing (Chapter 11).

Sampling frames used for telephone interviewing have been replaced increasingly by random digital dialling. By selecting particular within-country area dialling codes this provides a chance to reach any household within that area represented by that code which has a telephone, regardless of whether or not the number is ex-directory. However, care must be taken as, increasingly, households have more than one telephone number. Consequently there is a higher probability of people in such households being selected as part of the sample. In addition, such a sample would exclude people who use only mobile telephones as their dialling codes are telephone network operator rather than geographical area specific (Tucker and Lepkowski 2008).

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 0, the second 1 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.

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.

On its own, selecting one in every three would not be random as every third case would be bound to be selected, whereas those between would have no chance of selection. To overcome this, a random number is used to decide where to start on the sampling frame. If your sampling fraction is 1/3 the starting point must be one of the first three cases. You therefore select a random number (in this example a one-digit random number between 0 and 2) as described earlier and use this as the starting point. Once you have selected your first case at random you then select, in this example, every third case until you have gone right through your sampling frame (Box 7.7).

In some instances it is not necessary to actually construct a list for your sampling frame. For Internet questionnaires, such as pop-up questionnaires that appear in a window on the computer screen, there is no need to create an actual list if an invitation to participate is triggered at random. For systematic random sampling, a random selection could be triggered by a mechanism such as every tenth visitor to the website over a specified time period (Bradley 1999).

Despite the advantages, you must be careful when using existing lists as sampling frames. You need to ensure that the lists do not contain periodic patterns. Let us assume a high street bank needs you to administer a questionnaire to a sample of individual customers with joint bank accounts. A sampling fraction of 1/2 means that you will need to select every second customer on the list. The names on the customer lists, which you intend to use as the sampling frame, are arranged alphabetically by joint account, with predominantly males followed by females (Table 7.4). If you start with a male customer, the majority of those in your sample will be male. Conversely, if you start with a female customer, the majority of those in your sample will be female. Consequently your sample will be biased (Table 7.4). Systematic random sampling is therefore not suitable without reordering or stratifying the sampling frame (discussed later).

Box 7.7 Focus on student research

Systematic random sampling

Stefan worked as a receptionist in a dental surgery with approximately 1500 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 used the patient files kept in the filing cabinet as a sampling frame, and decided to select 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 0 and 4.

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 filing cabinet). 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.

Unlike simple random sampling, systematic random sampling works equally well with a small or large number of cases. However, if your target population covers a large geographical area, the random selection means that the sample cases are likely to be dispersed throughout the area. Consequently, systematic random sampling is suitable for geographically dispersed cases only if you do not require face-to-face contact when collecting your data.

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.

Table 7.4 The impact of periodic patterns on systematic random sampling

Number	Customer	Sample	Number	Customer	Sample
000	Mr L. Baker	✓	006	Mr A. Saunders	✓
001	Mrs B. Baker	*	007	Mrs C. Saunders	*
002	Mr P. Knight	✓	008	Mr J. Smith	✓
003	Ms J. Farnsworth	*	009	Mrs K. Smith	*
004	Mr J. Lewis	✓	010	Mr R. Dwight	✓
005	Mrs P. Lewis	*	011	Mr D. Furness	*

✓Sample selected if you start with 000. * Sample selected if you start with 001.

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.

In some instances, as pointed out by De Vaus (2014), your sampling frame will already be divided into strata. A sampling frame of employee names that is in alphabetical order will automatically ensure that, if systematic random sampling is used (discussed earlier), 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. 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.

The stratification variable (or variables) chosen should represent the discrete characteristic (or characteristics) for which you want to ensure correct representation within the sample (Box 7.8).

Samples can be stratified using more than one characteristic. You may wish to stratify a sample of an organisation's employees by both department and salary grade. To do this you would:

1 Divide the sampling frame into the discrete departments.
2 Within each department divide the sampling frame into discrete salary grades.
3 Number each of the cases within each salary grade within each department with a unique number, as discussed earlier.
4 Select your sample using either simple random or systematic random sampling, as discussed earlier.

In some instances the relative sizes of different strata mean that, in order to have sufficient data for analysis, you need to select larger samples from the strata with smaller target populations. Here the different sample sizes must be taken into account when aggregating data from each of the strata to obtain an overall picture. More sophisticated statistical analysis software packages enable you to do this by differentially weighting the responses for each stratum (Section 12.2).

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 (Barnett 2002). 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 (Box 7.9).

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,

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 used to select the first case in the public sector (001) and private sector (003) strata. Subsequently, every fourth customer in each stratum was selected.

Public sector stratum			Private sector stratum		
Number	**Customer**	**Selected**	**Number**	**Customer**	**Selected**
000	Anyshire County Council		000	ABC Automotive manufacturer	
001	Anyshire Hospital Trust	✓	001	Anytown printers and bookbinders	
002	Newshire Army Training Barracks		002	Benjamin Toy Company	
003	Newshire Police Force		003	Jane's Internet Flower shop	✓
004	Newshire Housing		004	Multimedia productions	
005	St Peter's Secondary School	✓	005	Roger's Consulting	
006	University of Anytown		006	The Paperless Office	
007	West Anyshire Council		007	U-need-us Ltd	✓

Box 7.9 Focus on student research

Cluster sampling

Ceri needed to select a sample of firms to undertake an interview-based survey 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 interview 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 0, the second 1 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.

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 0, the second 1 and so on.
3 Select your sample of clusters using some form of random sampling, as discussed earlier.

Selecting clusters randomly makes cluster sampling a probability sampling technique. Despite this, the technique normally results in a sample that represents the target population less accurately than stratified random sampling. Restricting the sample to a few relatively compact geographical sub-areas (clusters) maximises the amount of data you can collect using face-to-face methods within the resources available. However, it may also reduce the representativeness of your sample. For this reason you need to maximise the number of sub-areas to allow for variations in the target population within the available resources. Your choice is between a large sample from a few discrete subgroups and a smaller sample distributed over the whole group. It is a trade-off between the amount of precision lost by using a few subgroups and the amount gained from a larger sample size.

Multi-stage sampling

Multi-stage sampling, sometimes called *multi-stage cluster sampling*, is a development of cluster sampling. It is normally used to overcome problems associated with a geographically dispersed population when face-to-face contact is needed or where it is expensive and time consuming to construct a sampling frame for a large geographical area. However, like cluster

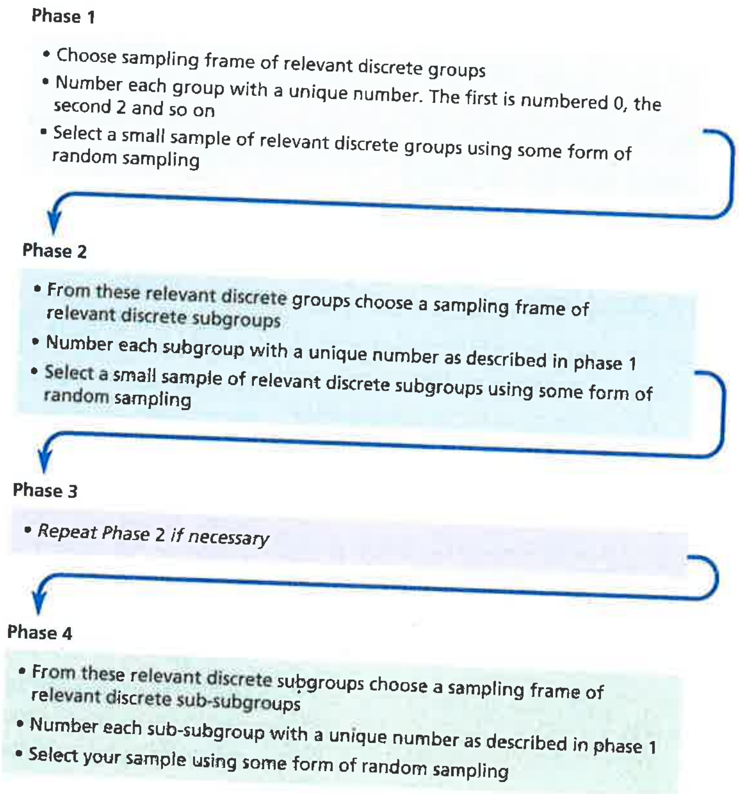

Figure 7.4 Phases of multi-stage sampling

sampling, you can use it for any discrete groups, including those that are not geographically based. The technique involves modifying a cluster sample by adding at least one more stage of sampling that also involves some form of random sampling. This aspect is represented by the dotted lines in Figure 7.2, the drawing of these samples being termed sub-sampling. Multi-stage sampling can be divided into four phases. These are outlined in Figure 7.4.

Because multi-stage sampling relies on a series of different sampling frames, you need to ensure that they are all appropriate and available. In order to minimise the impact of selecting smaller and smaller subgroups on the representativeness of your sample, you can apply stratified random sampling techniques (discussed earlier). This technique can be further refined to take account of the relative size of the subgroups by adjusting the sample size for each subgroup. As you have selected your sub-areas using different sampling frames, you only need a sampling frame that lists all the members of the population for those subgroups you finally select (Box 7.10). This provides considerable savings in time and money.

Checking that the sample is representative

Often it is possible to compare data you collect from your sample with data from another source for the population, such as data contained in an 'archival' database. For example, you can compare data on the age and socioeconomic characteristics of respondents in a marketing survey with these characteristics for the population in that country as recorded by the latest national census of population. If there is no statistically significant difference, then the sample is representative with respect to these characteristics.

When working within an organisation, comparisons can also be made. In a questionnaire Mark sent to a sample of employees in a large UK organisation, he asked closed questions about salary grade, gender, length of service and place of work. Possible responses

Box 7.10 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 using simple random 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 large, each was subdivided into smaller geographically discrete areas (electoral wards). These formed the next sampling frame (stage 2). Laura selected another simple random sample. This time she selected a larger number of wards 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.

to each question were designed to provide sufficient detail to compare the characteristics of the sample with the characteristics of the entire population of employees as recorded by the organisation's personnel database. At the same time he kept the categories sufficiently broad to preserve, and to be seen to preserve, the confidentiality of individual respondents. The two questions on length of service and salary grade from a questionnaire he developed illustrate this:

97 How long have you worked for **organisation's name?**

less than 1 year ❑ 1 year to less than 3 years ❑ 3 or more years ❑

98 Which one of the following best describes your job?

Clerical (grades 1–3)	❑	Management (grades 9–11)	❑
Supervisory (grades 4–5)	❑	Senior management (grades 12–14)	❑
Professional (grades 6–8)	❑	Other (please say)	❑

Using the Kolmogorov test (Section 12.5), Mark found there was no statistically significant difference between the proportions of respondents in each of the length of service groups and the data obtained from the organisation's personnel database for all employees. This meant that the sample of respondents was representative of all employees with respect to length of service. However, those responding were (statistically) significantly more likely to be in professional and managerial grades than in technical, administrative or supervisory grades. He therefore added a note of caution about the representativeness of his findings.

You can also assess the representativeness of samples in a variety of other ways (Rogelberg and Stanton 2007). Those our students have used most often, in order of quality of assessment of possible bias, include:

- replicating your findings using a new sample selected using different sampling techniques, referred to as 'demonstrate generalisability';
- resurveying non-respondents, the 'follow-up approach';
- analysing whether non-response was due to refusal, ineligibility or some other reason through interviews with non-respondents, known as 'active non-response analysis';
- comparing late respondents' responses with those from early respondents, known as 'wave analysis'.

In relation to this list, the quality of the assessment of bias provided by archival analysis, as outlined earlier, is similar to that provided by the follow-up approach and active non-response analysis.

7.3 Non-probability sampling

The techniques for selecting samples discussed earlier have all been based on the assumption that your sample will be chosen at random from a sampling frame. Consequently, it is possible to specify the probability that any case will be included in the sample. However, within business research, such as market surveys and case study research, this may either not be possible (as 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. In the exploratory stages of some research projects, such as a pilot testing a questionnaire, a non-probability sample may be the most practical, although it will not allow the extent of the problem to be determined. Subsequent to this, probability

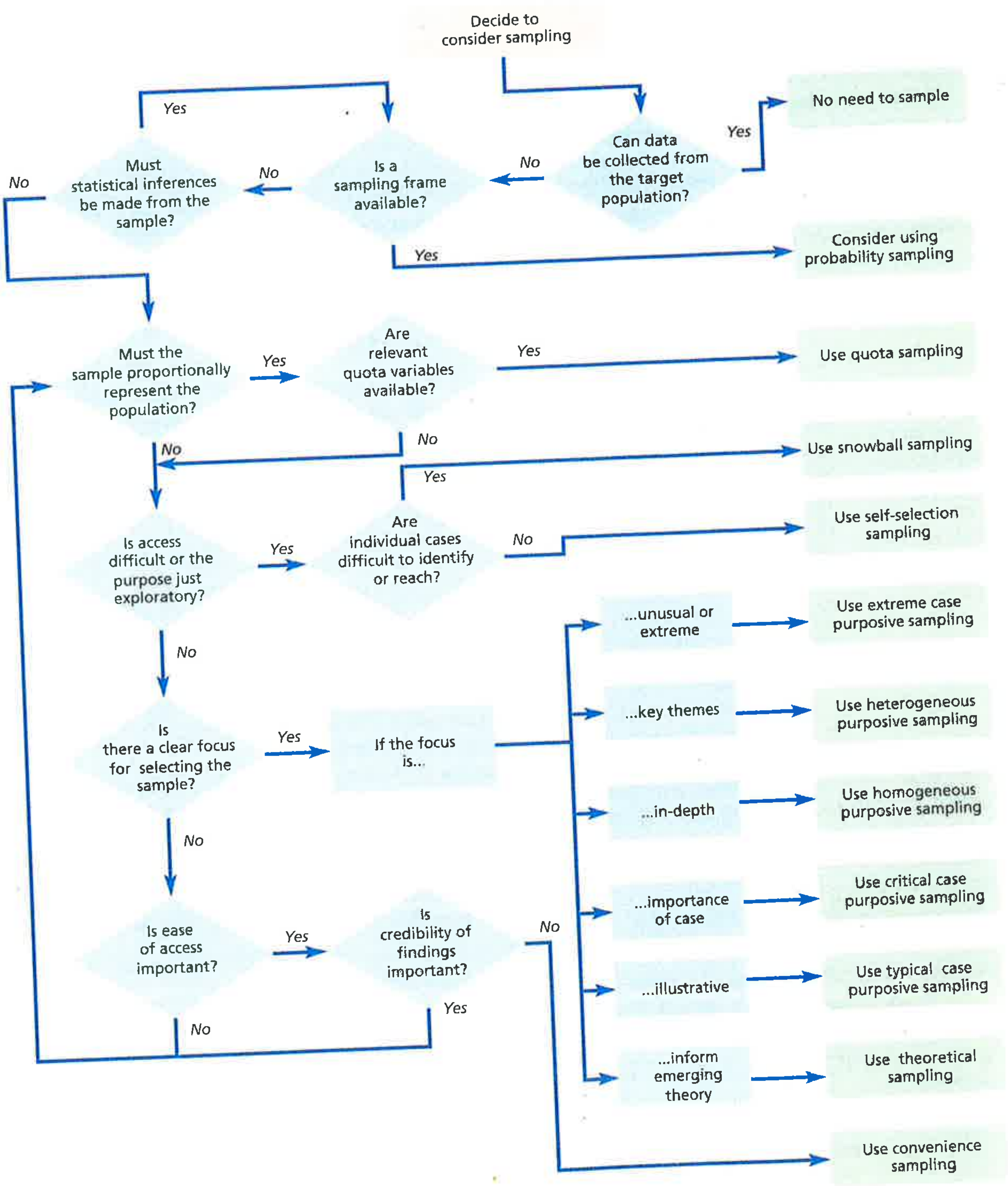

Figure 7.5 Choosing a non-probability sampling technique

sampling techniques may be used. For other business and management research projects your research question(s), objectives and choice of research strategy (Sections 2.4 and 5.5) may dictate non-probability sampling. To answer your research question(s) and to meet your objectives you may need to undertake an in-depth study that focuses on a small number of cases, perhaps one, selected for a particular purpose. This sample would provide you with an information-rich case study in which you explore your research question and gain theoretical insights.

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 (Figure 7.5), generalisations being made to theory rather than about a population. Consequently, your sample size is 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 (Patton 2002). This is particularly so where you are intending to collect qualitative data using semi-structured or unstructured interviews (Chapter 10). Although the validity, understanding and insights that you will gain from your data will be more to do with your data collection and analysis skills than with the size of your sample (Patton 2002), it is possible to offer guidance as to the sample size to ensure you have conducted sufficient interviews or undertaken sufficient observations.

In addressing this issue, many research textbooks simply recommend continuing to collect qualitative data, such as by conducting additional interviews, until **data saturation** is reached: in other words until the additional data collected provide little, if any, new information or suggest new themes. While this is helpful, it does not answer the question of how many participants you are likely to need in your sample. Fortunately, Saunders (2012) summarises the limited guidance available for different types of study (Table 7.5). For research where your aim is to understand commonalities within a fairly homogenous group, 12 in-depth interviews should suffice (Guest et al. 2006). However, Guest et al. also note that 12 interviews are unlikely to be sufficient where the sample is drawn from a heterogeneous population or the focus of the research question is wide ranging. Given this, we would suggest that, for a general study, you should expect to undertake between 5 and 30 interviews (Creswell 2013). Additionally, where your research requires comparison between distinct groups, the sample size will need to be larger, treating each group as a separate homogeneous population.

Table 7.5 Minimum non-probability sample size

Nature of study	Minimum sample size
Semi-structured/In-depth interviews	5–25
Ethnographic	35–36
Grounded Theory	20–35
Considering a homogeneous population	4–12
Considering a heterogeneous population	12–30

Source: Developed from Saunders (2012)

Selecting the most appropriate sampling technique and the sample

Having decided the likely suitable sample size, you need to select the most appropriate sampling technique to enable you to answer your research question from the range of non-probability sampling techniques available (Figure 7.2). At one end of this range is quota sampling, which, like probability samples, tries to represent the total population. At the other end of this range is haphazard sampling, based on the need to obtain a sample as quickly as possible. With this technique you have virtually no control over the cases that will be included in your sample. Purposive sampling and volunteer sampling techniques lie between these extremes (Table 7.6).

Table 7.6 Impact of various factors on choice of non-probability sampling techniques

Group	Technique	Likelihood of sample being representative	Types of research in which useful	Relative costs	Control over sample contents
Quota	Quota	Reasonable to high, although dependent on selection of quota variables	Where costs constrained or data needed very quickly so an alternative to probability sampling needed	Moderately high to reasonable	Specifies quota selection criteria
Purposive	Extreme case	Low	Unusual or special	Reasonable	Specifies selection criteria
	Heterogeneous	Low, although dependent on researcher's choices	Reveal/illuminate key themes	Reasonable	Specifies selection criteria
	Homogeneous	Low	In-depth focus	Reasonable	Specifies selection criteria
	Critical case	Low	Importance	Reasonable	Specifies selection criteria
	Typical case	Low, although dependent on researcher's choices	Illustrative	Reasonable	Specifies selection criteria
	Theoretical	Low	Inform emerging theory	Reasonable	Specifies selection criteria
Volunteer	Snowball	Low, but cases likely to have characteristics desired	Where cases difficult to identify	Reasonable	Selects initial participant
	Self-selection	Low, as cases self-selected	Where access difficult, research exploratory	Reasonable	Offers only general invitation
Haphazard	Convenience	Very low (often lacks credibility)	Ease of access	Low	Haphazard

Sources: Developed from Kervin (1999); Patton (2002); Saunders (2012)

Quota sampling

Quota sampling is entirely non-random and is often used for structured interviews 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. Quota sampling is therefore a type of stratified sample in which selection of cases within strata is entirely non-random (Barnett 2002). Quota sampling has similar requirements for sample size as probabilistic sampling techniques (Section 7.2). To select a quota sample you:

1. Divide the population into specific groups.
2. Calculate a quota for each group based on relevant and available data.
3. Give each interviewer an 'assignment', which states the number of cases in each quota from which they must collect data.
4. Combine the data collected by interviewers to provide the full 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. If, as with television audience research surveys, your data collection needs to be undertaken very quickly then quota sampling may be the only possibility. 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. This often necessitates a sample size of between 2000 and 5000.

Calculations of quotas are based on relevant and available data and are usually relative to the proportions in which they occur in the population (Box 7.11). 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 7.11 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 he needed to interview 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 7.11 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

These were then divided into assignments of between 40 and 50 people for each interviewer.

Where people who are retired are likely to have different opinions from those in work, a quota that does not ensure that these differences are captured may result in the data being biased as it would probably be easier to collect the data from those people who are retired. Quotas used in market research surveys and political opinion polls usually include measures of age, gender and economic activity or social class. These may be supplemented by additional quotas, dictated by the research question(s) and objectives (Box 7.11).

Once you have given each interviewer their particular assignment, they decide whom to interview until they have completed their quota. You then combine the data from this assignment with those collected by other interviewers to provide the full sample. Because the interviewer can choose within quota boundaries whom they interview, your quota sample may be subject to bias. Interviewers tend to choose respondents who are easily accessible and who appear willing to answer the questions. Clear controls may therefore be needed. In addition, it has been known for interviewers to fill in quotas incorrectly. This is not to say that your quota sample will not produce good results; they can and often do! However, you cannot measure the level of certainty or margins of error as the sample is not probability based.

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 **judgemental 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 (Neuman 2005). A particular form of purposive sampling, theoretical sampling, is used by researchers adopting the Grounded Theory strategy (Section 13.8).

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 (Box 7.12). Patton (2002) emphasises this point by contrasting the need to select information-rich cases in purposive sampling with the need to be statistically representative in probability sampling. The more common purposive sampling strategies were outlined in Figure 7.2.

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 (Box 7.12). This is often based on the premise that findings from extreme cases will be relevant in understanding or explaining more typical cases (Patton 2002).

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, Patton (2002) argues that this is in fact a strength. Any patterns that do emerge are likely to be of particular interest and value and represent the key themes. In addition, the data collected should enable you to document uniqueness. To ensure maximum variation within a sample, Patton (2002) suggests you identify your diverse characteristics (sample selection criteria) prior to selecting your sample.

Box 7.12 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 an 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. Sonnenshein 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.

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 organisation's hierarchy. Characteristics of the selected participants are similar, allowing them to be explored in greater depth and minor differences to be more apparent.

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 generalisations can be made. Patton (2002) outlines a number of clues that suggest critical cases. These can be summarised by the questions such as:

- If it happens there, will it happen everywhere?
- If they are having problems, can you be sure that everyone will have problems?
- If they cannot understand the process, is it likely that no one will be able to understand the process?

In contrast, **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.

Theoretical sampling is a special case of purposive sampling, being particularly associated with Grounded Theory and analytic induction (Sections 13.9 and 13.8). 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 are volunteered 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 in the population.
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.

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 snowballs (Box 7.13). 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 (Lee 2000). The next problem is to find these new cases. However, for populations that are difficult to identify, snowball sampling may provide the only possibility.

Self-selection sampling is the second of the volunteer sampling techniques we look at. It 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.

Publicity for convenience samples can take many forms. These include articles and advertisements in magazines that the population are likely to read, postings on appropriate online newsgroups and discussion groups, hyperlinks from other websites as well as letters, emails or tweets of invitation to colleagues and friends (Box 7.13). Cases that self-select often do so because of their strong feelings or opinions about the research

Box 7.13 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 SurveyMethods.com online survey software.

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.

Haphazard sampling

Haphazard sampling occurs when sample cases are selected without any obvious principles of organisation 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 centre for a television programme 'vox pop'. Although convenience sampling is used widely (for example, Facebook polls or questions), it is prone to bias and influences that are beyond your control. Cases appear in the sample only because of the ease of obtaining them; consequently all you can do is make some statement about the people who felt strongly enough about the subject of your question to answer it (and were using Facebook) during the period your poll was available! Not surprisingly, as emphasised in Figure 7.5, findings from convenience samples are often given very little credibility. Despite this, Saunders (2012) points out that samples ostensibly chosen for convenience often meet purposive sample selection criteria that are relevant to the research aim. It may be that an organisation you intend to use as a case study is 'convenient' because you have been able to negotiate access through existing contacts. Where this organisation also represents a 'typical' case, it can also offer an appropriate illustrative scenario, providing justification regarding its purpose when addressing the research aim. Alternatively, whilst a sample of operatives in another division of an organisation for which you work might be easy to obtain and consequently 'convenient', the fact that such participants allow you to address a research aim necessitating an in-depth focus on a particular homogenous group is more crucial.

Where the reasons for using a convenience sample have little, if any, relevance to the research aim, participants appear in the sample only because of the ease of obtaining them. Whilst this may not be problematic if there is little variation in the target population, where the target population is more varied it can result in participants that are of limited use in relation to the research question. Often a sample is intended to represent more than the target population, for example managers taking a part-time MBA course as a surrogate for all managers. In such instances the selection of individual cases may introduce bias to the sample, meaning that subsequent interpretations must be treated with caution.

7.4 Summary

- Your choice of sampling techniques is dependent on the feasibility and sensibility of collecting data to answer your research question(s) and to address your objectives from the target population. For target populations of fewer than 50 it is usually more sensible to collect data from the entire population where you are considering using probability sampling.
- Choice of sampling technique or techniques is dependent on your research question(s) and objectives:
  - Research question(s) and objectives that need you to estimate statistically the characteristics of the target population from a sample require probability samples.
  - Research question(s) and objectives that do not require such generalisations can, alternatively, make use of non-probability sampling techniques.
- Probability sampling techniques all necessitate some form of sampling frame, so they are often more time consuming than non-probability techniques.

- Where it is not possible to construct a sampling frame you will need to use non-probability sampling techniques.
- Factors such as the confidence that is needed in the findings, accuracy required and likely categories for analyses will affect the size of the sample that needs to be collected:
  - Statistical analyses usually require a minimum sample size of 30.
  - Research question(s) and objectives that do not require statistical estimation may need far smaller samples.
- Sample size and the technique used are also influenced by the availability of resources, in particular financial support and time available to select the sample and to collect, input and analyse the data.
- Non-probability sampling techniques also provide you with the opportunity to select your sample purposively and to reach difficult-to-identify members of the target population.
- For many research projects you will need to use a combination of different sampling techniques.
- All your choices will be dependent on your ability to gain access to organisations. The considerations summarised earlier must therefore be tempered with an understanding of what is practically possible.

Self-check questions

Help with these questions is available at the end of the chapter.

7.1 Identify a suitable sampling frame for each of the following research questions.

- **a** How do company directors of manufacturing firms of over 500 employees think a specified piece of legislation will affect their companies?
- **b** Which factors are important in accountants' decisions regarding working in mainland Europe?
- **c** How do employees at Cheltenham Gardens Ltd think the proposed introduction of compulsory Sunday working will affect their working lives?

7.2 Lisa has emailed her tutor with the following query regarding sampling and dealing with non-response. Imagine you are Lisa's tutor. Draft a reply to answer her query.

Help!!! Sampling non response

Draft saved at 09:39

Hi

I interviewed someone yesterday and I (almost) failed to get him to say anything useful for my research project. This was strange as he had what appeared to be an extremely useful background. I was unable to get him to reflect on the issue of inhibitors of spin out companies in the *light of his own experiences or provide actual examples*. He clearly wanted to follow a format he had decided upon prior to the interview 😟. This was that there was a right and a wrong answer and he was guessing at what inhibited people rather than giving me the actual examples I asked for. This did not allow for feedback loops or linkages or ideas, which as you know, is what my research is about. It was a very linear model of innovation. My attempts to get the conversation onto my research were gently, but firmly, put aside. *He had his format on paper when I arrived.*

My question is: Can I just exclude this interview from my sample? He was a super guy and I enjoyed meeting him. But, because I could not get him to answer my questions the interview did not yield any insights. What should I do?

Yours

Lisa

7.3 You have been asked to select a sample of manufacturing firms using the sampling frame below. This also lists the value of their annual output in tens of thousands of pounds over the past year. To help you in selecting your sample the firms have been numbered from 0 to 99.

a Select two simple random samples, each of 20 firms, and mark those firms selected for each sample on the sampling frame.

b Describe and compare the pattern on the sampling frame of each of the samples selected.

c Calculate the average (mean) annual output in tens of thousands of pounds over the past year for each of the samples selected.

d Given that the true average annual output is £6,608,900, is there any bias in either of the samples selected?

	Output		Output		Output		Output		Output
0	1163	20	1072	40	1257	60	1300	80	1034
1	10	21	7	41	29	61	39	81	55
2	57	22	92	42	84	62	73	82	66
3	149	23	105	43	97	63	161	83	165
4	205	24	157	44	265	64	275	84	301
5	163	25	214	45	187	65	170	85	161
6	1359	26	1440	46	1872	66	1598	86	1341
7	330	27	390	47	454	67	378	87	431
8	2097	28	1935	48	1822	68	1634	88	1756
9	1059	29	998	49	1091	69	1101	89	907
10	1037	30	1298	50	1251	70	1070	90	1158
11	59	31	10	51	9	71	37	91	27
12	68	32	70	52	93	72	88	92	66
13	166	33	159	53	103	73	102	93	147
14	302	34	276	54	264	74	157	94	203
15	161	35	215	55	189	75	168	95	163
16	1298	36	1450	56	1862	76	1602	96	1339
17	329	37	387	57	449	77	381	97	429
18	2103	38	1934	58	1799	78	1598	98	1760
19	1061	39	1000	59	1089	79	1099	99	898

7.4 You have been asked to select a 10 per cent sample of firms from the sampling frame used for self-check question 7.3.

a Select a 10 per cent systematic random sample and mark those firms selected for the sample on the sampling frame.

b Calculate the average (mean) annual output in tens of thousands of pounds over the past year for your sample.

c Given that the true average annual output is £6,608,900, why does systematic random provide such a poor estimate of the annual output in this case?

7.5 You need to undertake a series of face-to-face interviews with managing directors of small- to medium-sized organisations. From the data you collect you need to be able to generalise about the attitude of such managing directors to recent changes in government policy towards these firms. Your generalisations need to be accurate to within plus or minus 5 per cent. Unfortunately, you have limited resources to pay for interviewers, travelling and other associated costs.

a How many managing directors will you need to interview?

b You have been given the choice between cluster and multi-stage sampling. Which technique would you choose for this research? You should give reasons for your choice.

7.6 You have been asked to undertake face-to-face interviews with local residents to discover their opinions regarding the siting of a new supermarket in an inner city suburb (estimated catchment population 111,376 at the last census). The age and gender distribution of the catchment population at the last census is listed below.

Gender	Age group							
	0–4	5–15	16–19	20–29	30–44	45–59 /64*	60/65#–74	75+
Males	3498	7106	4884	7656	9812	12892	4972	2684
Females	3461	6923	6952	9460	8152	9152	9284	4488

*59 females, 64 males; #60 females, 65 males.

a Devise a quota for a quota sample using these data.

b What other data would you like to include to overcome likely variations between groups in their availability for interview and replicate the target population more precisely? Give reasons for your answer.

c What problems might you encounter in using interviewers?

7.7 For each of the following research questions it has not been possible for you to obtain a sampling frame. Suggest the most suitable non-probability sampling technique to obtain the necessary data, giving reasons for your choice.

a What support do people sleeping rough believe they require from social services?

b Which television advertisements do people remember watching last weekend?

c How do employers' opinions vary regarding the impact of European Union legislation on age discrimination?

d How are manufacturing companies planning to respond to the introduction of road tolls?

e Would users of the squash club be prepared to pay a 10 per cent increase in subscriptions to help fund two extra courts (answer needed by tomorrow morning!)?

Review and discussion questions

7.8 With a friend or colleague choose one of the following research questions (or one of your own) in which you are interested.

- What attributes attract people to jobs?
- How are financial institutions adapting the services they provide to meet recent legislation?

Use the flow charts for both probability sampling (Figure 7.3) and non-probability sampling (Figure 7.5) to decide how you could use each type of sampling independently to answer the research question.

7.9 Agree with a colleague to watch a particular documentary or consumer rights programme on the television. If possible, choose a documentary with a business or management focus. During the documentary, pay special attention to the samples from which the data for the documentary are drawn. Where possible, note down details of the sample such as who were interviewed, or who responded to questionnaires, and the reasons why these people were chosen. Where this is not possible, make a note of the information you would have liked to have been given. Discuss your findings with your colleague and come to a conclusion regarding the nature of the sample used, its representativeness and the extent to which it was possible for the programme maker to generalise from that sample.

7.10 Obtain or access online a copy of a quality daily newspaper and, within the newspaper, find an article that discusses a 'survey' or 'poll'. Share the article with a friend. Make notes of the process used to select the sample for the 'survey' or 'poll'. As you make your notes, note down any areas where you feel there is insufficient information to fully understand the sampling process. Aspects for which information may be lacking include the target population, size of sample, how the sample was selected, representativeness and so on. Discuss your findings with your friend.

Progressing your research project

Using sampling as part of your research

- Consider your research question(s) and objectives. You need to decide whether you will be able to collect data on the entire population or will need to collect data from a sample.
- If you decide that you need to sample, you must establish whether your research question(s) and objectives require probability sampling. If they do, make sure that a suitable sampling frame is available or can be devised, and calculate the actual sample size required, taking into account likely response rates. If your research question(s) and objectives do not require probability sampling, or you are unable to obtain a suitable sampling frame, you will need to use non-probability sampling.
- Select the most appropriate sampling technique or techniques after considering the advantages and disadvantages of all suitable techniques and undertaking further reading as necessary.
- Select your sample or samples following the technique or techniques as outlined in this chapter.
- Remember to note down the reasons for your choices when you make them, as you will need to justify your choices when you write about your research method.
- Use the questions in Box 1.4 to guide your reflective diary entry.

References

American Association for Public Opinion Research (2011) *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (7th edn). Lenexa, KA: AAPOR. Available at https://www.aapor.org//AM/Template.cfm?Section=Home [Accessed 14 April 2014].

Barnett, V. (2002) *Sample Survey Principles and Methods* (3rd edn). Chichester: Wiley.

Baruch, Y. and Holtom, B.C. (2008) 'Survey response rate levels and trends in organizational research', *Human Relations*, Vol. 61, pp. 1139–60.

Becker, H.S. (1998) *Tricks of the Trade: How to Think About Your Research While You're Doing It*. Chicago: Chicago University Press.

Bidlake, S. (2011) 'Outdoor hall of fame contest opens', *Campaign*, 10 February. Available at http://www.campaignlive.co.uk/news/1054277/Outdoor-Hall-Fame-contest-opens/ [Accessed 14 April 2014].

Bradley, N. (1999) 'Sampling for Internet surveys: An examination of respondent selection for Internet research', *Journal of the Market Research Society*, Vol. 41, No. 4, pp. 387–95.

Creswell, J. (2013) *Qualitative Inquiry and Research Design: Choosing among Five Approaches* (3rd edn). Thousand Oaks, CA: Sage.

De Vaus, D.A. (2014) *Surveys in Social Research* (6th edn). Abingdon: Routledge.

Dillman, D.A., Smyth, J.D. and Christian, J.M. (2014) *Internet, Phone, Mail and Mixed Mode Surveys: The Tailored Design Method* (4th edn). Hoboken, NJ: Wiley.

Edwards, T., Tregaskis, O., Edwards, P., Ferner, A., Marginson, A. with Arrowsmith, J., Adam, D., Meyer, M. and Budjanovcanin, A. (2007) 'Charting the contours of multinationals in Britain: Methodological challenges arising in survey-based research', *Warwick Papers in Industrial Relations*, No. 86. Available at http://www.cbs.dk/files/cbs.dk/charting_the_contours_of_multinationals_in_britain.pdf [Accessed 14 April 2014].

Farey-Jones, D. (2011) '"Hello Boys" voted greatest poster ever created', *Campaign*, 31 March. Available at http://www.campaignlive.co.uk/news/1063405/Hello-Boys-voted-greatest-poster-ever-created/ [Accessed 14 April 2014].

Groves, R.M. and Peytcheva, E. (2008) 'The impact of nonresponse rates on nonresponse bias', *Public Opinion Quarterly*, Vol. 72, No. 2, pp. 167–89.

Guest, G., Bunce, A. and Johnson, L. (2006) 'How many interviews are enough? An experiment with data saturation and validity', *Field Methods*, Vol. 18, No. 1, pp. 59–82.

Henry, G.T. (1990) *Practical Sampling*. Newbury Park, CA: Sage.

Idea Works (2012) *Methodologist's toolchest ex-sample*. Available at http://www.ideaworks.com/mt/exsample.html [Accessed 12 April 2014].

Kervin, J.B. (1999). *Methods for Business Research* (2nd edn). New York: HarperCollins.

Lee, R.M. (2000) *Doing Research on Sensitive Topics*. London: Sage.

Lenth, R.V. (2001) 'Some practical guidelines for effective sample size determination', *The American Statistician*, Vol. 55, No. 3, pp. 187–93.

Neuman, W.L. (2005) *Social Research Methods* (6th edn). London: Pearson.

Neuman, W.L. (2014) *Social Research Methods* (7th edn). Harlow: Pearson.

Patton, M.Q. (2002) *Qualitative Research and Evaluation Methods* (3rd edn). Thousand Oaks, CA: Sage.

Rogelberg, S.G. and Stanton, J.M. (2007) 'Introduction: Understanding and dealing with organizational survey non-response', *Organizational Research Methods*, Vol. 10, No. 2, pp. 195–209.

Saunders, M.N.K. (2012) 'Choosing research participants', in G. Symons and C. Cassell (eds) *The Practice of Qualitative Organizational Research: Core Methods and Current Challenges*. London: Sage, pp. 37–55.

Sonenshein, S., DeCelles, K. and Dutton, J.E. (2014) 'It's not easy being green: The role of self evaluations in explaining the support of environmental issues', *Academy of Management Journal*, Vol. 57, No. 1, pp. 7–37.

Tennent, J. (2013) *The Economist Numbers Guide: The Essentials of Business Numeracy* (6th edn). [Kindle e-book] London: Profile Books.

Tucker, C. and Lepkowski, J.M. (2008) 'Telephone survey methods: Adapting to change', in J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. De Leeuw, L. Japec, P.J. Lavrakas, M.W. Link and R.L. Sangster (eds), *Advances in Telephone Survey Methodology*. Hoboken, NJ: Wiley, pp. 3–28.

Further reading

Baruch, Y. and Holtom, B.C. (2008) 'Survey response rate levels and trends in organizational research', *Human Relations*, Vol. 61, pp. 1139–60. This examines 490 academic studies using surveys published in 2000 and 2005 covering 100,000 organisations and over 400,000 individual respondents. The paper suggests likely response rates for different types of study and offers useful advice for reporting response rates.

De Vaus, D.A. (2014) *Surveys in Social Research* (6th edn). Abingdon: Routledge. Chapter 6 provides a useful overview of both probability and non-probability sampling techniques.

Dillman, D.A., Eltringe, J.L., Groves, J.L. and Little, R.J.A. (eds) (2002) *Survey Nonresponse*. New York: Wiley Interscience. This book contains a wealth of information on survey non-response. Chapter 1 provides a useful overview in relation to the impact of survey design on non-response. This is discussed in more detail in Chapters 7 to 17, Chapter 14 referring specifically to business surveys and Chapter 15 to Internet-based surveys.

Patton, M.Q. (2002) *Qualitative Research and Evaluation Methods* (3rd edn). Thousand Oaks, CA: Sage. Chapter 5, 'Designing qualitative studies', contains a useful discussion of non-probability sampling techniques, with examples.

Saunders, M.N.K. (2012) 'Choosing research participants', in G. Symons and C. Cassell (eds) *The Practice of Qualitative Organizational Research: Core Methods and Current Challenges*. London: Sage, pp. 37–55. Provides a useful discussion of non-probability sampling with examples.

Case 7
Understanding and assessing economic inactivity among Maltese female homemakers

Street scene in Valletta, Malta
Source: ©Mark Saunders, 2015.

Andrew, a final year part-time MBA student employed by a large Maltese market research company, read in a local newspaper that Malta (a small island state situated in the centre of the Mediterranean Sea) has the highest rate of economically inactive women in the European Union. 'Inactivity' is a term that refers to all persons who are not classified as employed or unemployed (Eurostat 2011). This category includes students, retired persons, home makers, and ill or permanently disabled persons. On investigating further, he discovered that inactive women in Malta account for 52.2 per cent of the population of females in the working age group (16 to 64 years), while the female homemakers group represent the single largest dominant group (76.6 per cent) among the inactive population in Malta (Eurostat 2012).

Andrew was curious about these statistics and decided to conduct a study with Maltese inactive female homemakers for his research project. More specifically, he wanted to understand what motivated and what hindered them in seeking paid employment so as to determine which types of policies could facilitate their entry or re-entry into the labour market.

In determining the sample size for the study, Andrew specified a level of confidence in the estimate(s) of 95 per cent (this corresponding to a z score of 1.96) and a margin of error of 5 per cent. Without a pilot study, he assumed a worst-case scenario and hence a response rate of 50 per cent for categorical variables (De Vaus 2014 and Appendix 2). As the total population of female homemakers was 63,290 (Eurostat 2012), the minimum sample size needed to be 383.

Andrew wanted a random sample (i.e. one based on chance) and so all the female homemakers in the sampling frame had to have an equal chance of receiving an invitation. Additionally, to eliminate the threat of loss of participants, he wanted to find random replacements for all the initial invitees who for some reason or other declined to participate in the study. He was aware that his organisation was in possession of a sampling frame of inactive female homemakers, and that this was based on the information gathered during the most recent census carried out in the Maltese Islands. He therefore enquired about the possibility of being provided with a random sample of Maltese inactive female homemakers aged 16–64 whom he could then contact to collect his data. Their initial answer did not appear helpful! His employer was not in a position to provide the list since these inactive persons represent a vulnerable group. Andrew was worried as he did not want to resort to non-probability sampling methods. He believed that if a truly random process was used to select the sample from the population, the resulting sample would be like a miniature replica of the population – one that looks just like the population but smaller. Andrew therefore decided to ask if his employer would distribute a questionnaire on his behalf. Again the answer was no! Fortunately his employer was interested in finding out more about Maltese inactive female homemakers, and so suggested an alternative method with the following terms of reference:

- **a** The target population consisted of inactive female homemakers aged between 16 and 64 years.
- **b** His employer would randomly select 400 female homemakers from their sampling frame. To account for potential non-responses, random replacements would be found for any of the initial invitees who for some reason or another declined to participate in the research. Additionally, if any of the originally selected female homemakers had reached the retirement age (65+), started working, registered themselves as students, or became ill or permanently disabled during the time of study, they would also be replaced at random.
- **c** The questions to be asked would be provided by Andrew, following ethical clearance from his university's Ethics Committee. Demographic information about the respondents concerning age, marital status, highest level of education and past occupation (if any) had already been collected and were included in his employer's sampling frame, and so there was no need to ask for such information in the telephone questionnaire.
- **d** His employer would collect and prepare the data for statistical analysis. Andrew would receive a copy of cleaned and anonymised data in IBM SPSS Statistics format within 15 days of the end of the data collection period.

In total, his employer targeted 486 female homemakers (total number in sample). From these, 402 answered the telephone survey (total number of responses), 57 were unreachable and another 27 were ineligible to respond because at the time of the survey they were either in employment or had reached retirement age. Hence the total response rate was 87.6 per cent while the active response rate was 100 per cent.

Andrew was pleased that the sample was efficient since its size just exceeded 383 (the minimum sample size), and this meant that the sample size met the specified precision and confidence goals he aimed for. Additionally, he was confident that his sample was random and representative – that is, any numerical summaries of the gathered information would be unbiased and any statistical inferences drawn would be valid (Huck 2009). He felt that he could now proceed with the analysis of the data.

References

De Vaus, D.A. (2014) *Surveys in Social Research* (6th edn). Abingdon: Routledge.

Eurostat (2011) Definitions of employment and unemployment. Available at http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/hrst_st_esms_an8.pdf [Accessed 12 December 2013].

Eurostat (2012) *Statistics Database*. Available at http://ec.europa.eu/eurostat/statistics-explained/index.php/Youth_unemployment#Definition_of_unemployment_and_youth_unemployment_indicators [Accessed 15 May 2015].

Huck, S.W. (2009) *Statistical Misconceptions*. New York: Psychology Press.

Questions

1. In what ways is the Census data likely to differ from the actual population when Andrew conducts his research?
2. Is Andrew's random sample a miniature replica of the population? Give reasons for your answer.
3. Does the inactive female homemakers' sample represent a random sample from the population? Give reasons for your answer.
4. What would you recommend to minimise the impact of nonparticipation?

Additional case studies relating to material covered in this chapter are available via the book's companion website: **www.pearsoned.co.uk/saunders**.

They are:

- Change management at Hattersley Electronics.
- Employment networking in the Hollywood film industry.
- Auditor independence and integrity in accounting firms.
- Implementing strategic change initiatives.
- Comparing UK and French perceptions and expectations of online supermarket shopping.

Self-check answers

7.1 **a** A complete list of all directors of large manufacturing firms could be purchased from an organisation that specialised in selling such lists to use as the sampling frame. Alternatively, a list that contained only those selected for the sample could be purchased to reduce costs. These electronic data could be merged into standard letters such as those included with questionnaires.

b A complete list of accountants, or one that contained only those selected for the sample, could be purchased from an organisation that specialised in selling such lists. Care would need to be taken regarding the precise composition of the list to ensure that it included those in private practice as well as those working for organisations. Alternatively, if the research was interested only in qualified accountants then the professional accountancy bodies' yearbooks, which list all their members and their addresses, could be used as the sampling frame.

c Subject to ethical approval, the personnel records or payroll of Cheltenham Gardens Ltd could be used. Either would provide an up-to-date list of all employees with their addresses.

7.2 Your draft of Lisa's tutor's reply is unlikely to be worded the same way as the one below. However, it should contain the same key points:

"tutor's name" <lisas.tutor@anytown.ac.uk>
To: <lisa@anytown.ac.uk>
Sent: today's date 7:06
Subject: Re: Help!!! Sampling non-response?

Hi Lisa

Many thanks for the email. This is not in the least unusual. I reckon to get about 1 in 20 interviews which go this way and you just have to say 'c'est la vie'. This is not a problem from a methods perspective as, in sampling terms, it can be treated as a non-response due to the person refusing to respond to your questions. This would mean you could not use the material. However, if he answered some other questions then you should treat this respondent as a partial non-response and just not use those answers. Hope this helps.

'Tutor's name'

7.3 a Your answer will depend on the random numbers you selected. However, the process you follow to select the samples is likely to be similar to that outlined. Starting at randomly selected points, two sets of 20 two-digit random numbers are read from the random number tables (Appendix 3). If a number is selected twice it is disregarded. Two possible sets are:

Sample 1: 38 41 14 59 53 03 52 86 21 88 55 87 85 90 74 18 89 40 84 71
Sample 2: 28 00 06 70 81 76 36 65 30 27 92 73 20 87 58 15 69 22 77 31

These are then marked on the sampling frame (sample 1 is shaded in blue, sample 2 is shaded in orange) as shown below:

0	1163	*20*	1072	*40*	1257	*60*	1300	*80*	1034
1	10	*21*	7	*41*	29	*61*	39	*81*	55
2	57	*22*	92	*42*	84	*62*	73	*82*	66
3	149	*23*	105	*43*	97	*63*	161	*83*	165
4	205	*24*	157	*44*	265	*64*	275	*84*	301
5	163	*25*	214	*45*	187	*65*	170	*85*	161
6	1359	*26*	1440	*46*	1872	*66*	1598	*86*	1341
7	330	*27*	390	*47*	454	*67*	378	*87*	431
8	2097	*28*	1935	*48*	1822	*68*	1634	*88*	1756
9	1059	*29*	998	*49*	1091	*69*	1101	*89*	907
10	1037	*30*	1298	*50*	1251	*70*	1070	*90*	1158
11	59	*31*	10	*51*	9	*71*	37	*91*	27
12	68	*32*	70	*52*	93	*72*	88	*92*	66
13	166	*33*	159	*53*	103	*73*	102	*93*	147
14	302	*34*	276	*54*	264	*74*	157	*94*	203
15	161	*35*	215	*55*	189	*75*	168	*95*	163
16	1298	*36*	1450	*56*	1862	*76*	1602	*96*	1339
17	329	*37*	387	*57*	449	*77*	381	*97*	429
18	2103	*38*	1934	*58*	1799	*78*	1598	*98*	1760
19	1061	*39*	1000	*59*	1089	*79*	1099	*99*	898

b Your samples will probably produce patterns that cluster around certain numbers in the sampling frame, although the amount of clustering may differ, as illustrated by samples 1 and 2 above.

c The average (mean) annual output in tens of thousands of pounds will depend entirely upon your sample. For the two samples selected the averages are:

Sample 1 (enclosed by a box): £6,752,000
Sample 2 (shaded): £7,853,500

d There is no bias in either of the samples, as both have been selected at random. However, the average annual output calculated from sample 1 represents the target population more closely than that calculated from sample 2, although this has occurred entirely at random.

7.4 a Your answer will depend on the random number you select as the starting point for your systematic sample. However, the process you followed to select your sample is likely to be similar to that outlined. As a 10 per cent sample has been requested, the sampling fraction is 1/10. Your starting point is selected using a random number between 0 and 9, in this case 2. Once the firm numbered 2 has been selected, every tenth firm is selected:

2 12 22 32 42 52 62 72 82 92

These are marked with orange shading on the sampling frame and will result in a regular pattern whatever the starting point:

0	1163	*20*	1072	*40*	1257	*60*	1300	*80*	1034
1	10	*21*	7	*41*	29	*61*	39	*81*	55
2	57	*22*	92	*42*	84	*62*	73	*82*	66
3	149	*23*	105	*43*	97	*63*	161	*83*	165
4	205	*24*	157	*44*	265	*64*	275	*84*	301
5	163	*25*	214	*45*	187	*65*	170	*85*	161
6	1359	*26*	1440	*46*	1872	*66*	1598	*86*	1341
7	330	*27*	390	*47*	454	*67*	378	*87*	431
8	2097	*28*	1935	*48*	1822	*68*	1634	*88*	1756
9	1059	*29*	998	*49*	1091	*69*	1101	*89*	907
10	1037	*30*	1298	*50*	1251	*70*	1070	*90*	1158
11	59	*31*	10	*51*	9	*71*	37	*91*	27
12	68	*32*	70	*52*	93	*72*	88	*92*	66
13	166	*33*	159	*53*	103	*73*	102	*93*	147
14	302	*34*	276	*54*	264	*74*	157	*94*	203
15	161	*35*	215	*55*	189	*75*	168	*95*	163
16	1298	*36*	1450	*56*	1862	*76*	1602	*96*	1339
17	329	*37*	387	*57*	449	*77*	381	*97*	429
18	2103	*38*	1934	*58*	1799	*78*	1598	*98*	1760
19	1061	*39*	1000	*59*	1089	*79*	1099	*99*	898

b The average (mean) annual output of firms for your sample will depend upon where you started your systematic sample. For the sample selected above it is £757,000.

c Systematic sampling has provided a poor estimate of the annual output because there is an underlying pattern in the data, which has resulted in firms with similar levels of output being selected.

7.5 **a** If you assume that there are at least 100,000 managing directors of small- to medium-sized organisations from which to select your sample, you will need to interview approximately 380 to make generalisations that are accurate to within plus or minus 5 per cent (Table 7.1).

b Either cluster or multi-stage sampling could be suitable; what is important is the reasoning behind your choice. This choice between cluster and multi-stage sampling is dependent on the amount of limited resources and time you have available. Using multi-stage sampling will take longer than cluster sampling as more sampling stages will need to be undertaken. However, the results are more likely to be representative of the target population owing to the possibility of stratifying the samples from the sub-areas.

7.6 **a** Prior to deciding on your quota you will need to consider the possible inclusion of residents who are aged under 16 in your quota. Often in such research projects residents aged under 5 (and those aged 5–15) are excluded. You would need a quota of between 2000 and 5000 residents to obtain a reasonable accuracy. These should be divided proportionally between the groupings as illustrated in the possible quota below:

	Age group					
Gender	**16–19**	**20–29**	**30–44**	**45–59/64**	**60/65–74**	**75+**
Males	108	169	217	285	110	59
Females	154	209	180	203	205	99

b Data on social class, employment status, socioeconomic status or car ownership could also be used as further quotas. These data are often available from your national Census and are likely to affect shopping habits.

c Interviewers might choose respondents who were easily accessible or appeared willing to answer the questions. In addition, they might fill in their quota incorrectly or make up the data.

7.7 **a** Either snowball sampling as it would be difficult to identify members of the target population or, possibly, convenience sampling because of initial difficulties in finding members of the target population.

b Quota sampling to ensure that the variability in the target population as a whole is represented.

c Purposive sampling to ensure that the full variety of responses are obtained from a range of respondents from the target population.

d Self-selection sampling as it requires people who are interested in the topic.

e Convenience sampling owing to the very short timescales available and the need to have at least some idea of members' opinions.

Get ahead using resources on the companion website at: **www.pearsoned.co.uk/saunders**.

- Improve your IBM SPSS Statistics and NVivo research analysis with practice tutorials.
- Save time researching on the Internet with the Smarter Online Searching Guide.
- Test your progress using self-assessment questions.
- Follow live links to useful websites.