8

SAMPLING

In addition, many of the video and web links on the Companion Website (study.sagepub.com/brotherton) provide both basic and more advanced material on sampling and the sampling issues and techniques covered in this chapter.

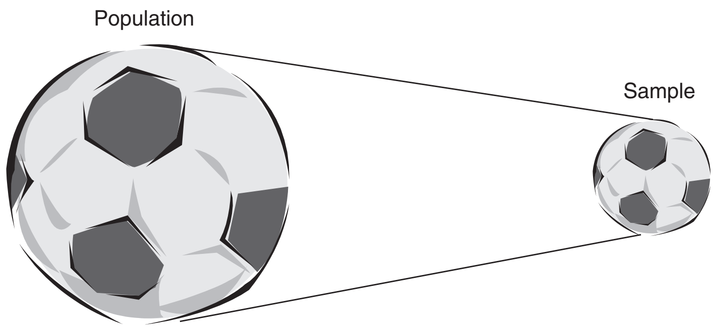


FIGURE 8.1 A representative sample

Key Concept	**Validity and Generalisability**

It is extremely important to understand the difference between the validity of sample data and its suitability for generalisation. How valid or accurate the sample data are, in terms of the extent to which they provide a true picture of the views of the respondents, is largely a function of the quality of the data instrument (questionnaire) used to obtain their responses.

The use of a valid instrument should produce valid responses regardless of the size, composition or representativeness of any sample these are derived from. Therefore, even very small and/or biased samples can provide valid data. However, certainly biased and, depending on the size of the population, very small samples will not provide data that are suitable to be used for further generalisations.

Key Concept	Population

Although the concept of ‘population’ is generally defined as a set or collection of elements that share some common characteristics it can be defined in different ways.

The ‘conceptual’ population is one likely to be relevant to your study. For example, all budget hotels in the UK or all front-line service staff in UK five-star hotels would constitute populations that would be appropriate and relevant to research studies focusing on these contexts.

The ‘study’ population is a part, or sub-set, of the conceptual population. The study population is the one that you are going to use to select your sample, e.g. taking the idea of all budget hotels in the UK as the conceptual population we may, often for very practical reasons, decide to reduce this to one that is more specific and manageable. For example, you might decide that only budget hotels operated by the main budget hotel brands or companies will be used to define the population for your study. Alternatively, you may decide to limit the study population geographically by defining it as all budget hotels within the city of Manchester. These decisions are not random ones, they are taken for good reasons relating either to the nature of the research questions/context defined earlier in the study and/or for more practical, i.e. access, issues.

Key Concept	The Sample

The sample you decide to select must be related to the purpose/s or objectives of your research.

Choosing a sample from which you intend to collect data is essentially a ‘means to an end’ decision. Or, put another way, the sample is a vehicle that you are going to use to help you collect the type and volume of data you need to complete your research.

Although this will be partly a ‘theoretical’ decision related to the nature of your research objectives, it is also invariably a practical or pragmatic one as well. All research projects have parameters within which they have to operate, even those with large budgets, teams of researchers and extended timescales, so we need to think about the ‘practicalities’ we face.

This means you have to consider the time you have available, the ease of access you may have to your potential sample respondents and the costs involved.

All sampling is essentially a compromise between the ideal and the achievable and the bottom-line is that all you can do is seek to make your sample as good as it can be given the constraints you face.

Key Concept	**Sampling Terminology**

All the elements contained within your chosen population are known as the **‘sample frame’**, from which the actual sample can be drawn or selected.

Those selected to form part of the sample become known as **‘sample subjects’** and the actual people who respond to the questions put to them are the **‘sample respondents’**.

The data or values generated from the respondents’ responses are known as the **‘sample statistics’** and the corresponding values in the population are referred to as the **‘population parameters’**.

Key Concept	**Randomness**

One issue that causes a lot of confusion is that of random sampling or selecting a sample on a random basis.

In everyday life we tend to think that something described as being random is something that happens over which we have no prior knowledge and no control and, because it is random, we cannot predict when, and possibly how, it will occur. Well, of course, this is true but there is something of a paradox when we talk about randomising the process of selecting a sample because to create a randomly selected sample we have to do this in a deliberately random way.

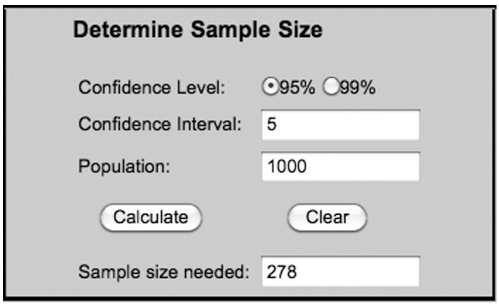
In other words, it is purposeful and planned and we know the likelihood, or probability, of each member of the sampling frame being selected to be included in the sample. We deliberately create a random process to ensure that the selection takes place randomly. So, a random sample is not simply created without prior planning.

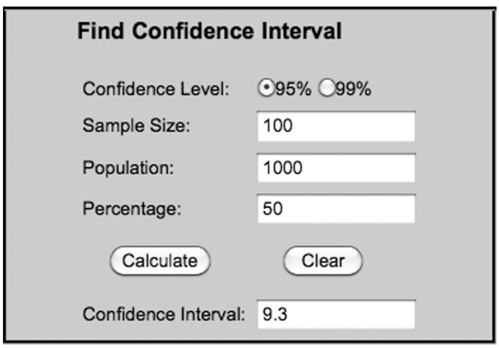
Technique Tip

Calculating Sample Size

The various sample size calculators available on the internet (see the Companion Website) make this issue relatively straightforward. Using one of these – www.surveysystem.com/sscalc.htm – the following example shows how you can do this.

As you can see you can set the confidence level at the 95% or 99% levels and enter a value for the desired confidence interval. In this case I entered one of 5 and specified the population as 1,000. Hitting the calculate button then delivers the result for sample size needed which, in this case, is 278. In the following box – Find Confidence Interval – you can see the effect that reducing the sample size to 100 has. Now, with this smaller sample, the confidence level has declined to 9.3. It has declined because plus or minus 9.3 provides a wider margin for error, or less precision, than plus or minus 5.





Reproduced with permission of http://www.surveysystem.com/sscale.htm

Males					Females				
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀
♂	♂	♂	♂	♂	♀	♀	♀	♀	♀

FIGURE 8.2 Symmetrical population distribution

Males							Females		
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀
♂	♂	♂	♂	♂	♂	♂	♀	♀	♀

FIGURE 8.3 Asymmetrical population distribution

Research in Action

Critical Success Factors and Budget Hotels

The questionnaire package was mailed to the general manager of each budget hotel in the sample. **The population** for the **sample selection** was defined as the leading budget hotel brands in this sector of the UK hotel industry … The **sampling frame** was derived from the literature, i.e. the annual Deloitte & Touche UK budget hotel surveys referred to earlier and the recent Mintel (1999) report on budget hotels. The names and addresses of the constituent hotels were obtained from the companies' budget hotel directories and/or their web sites. **A sample of 549 was selected** from this information. This procedure generated **an initial return** of 209 **completed and useable questionnaires.** To address the **validity** issues associated with **non-response** appropriate follow-up action was taken. A reminder letter, with another copy of the questionnaire, was posted to all the non-responding hotels approximately one week after the date given for the return of the original questionnaire. This resulted in a further 30 questionnaires being completed and returned. Thus, **the final useable sample** comprised 239 questionnaires, comprising a very satisfactory **final response rate** of 44 per cent for this type of mailed survey.

The size of the realised sample (n = 239) was very encouraging in terms of providing a representative data set from the budget hotel sector … the sample was, not unsurprisingly, dominated by the two leading brands, Premier Inn (originally branded Travelinn, then Premier Travel Inn) and Travelodge. Though this did skew the sample in favour of these brands **it nevertheless reflects the population distribution of budget hotel brands in the UK**. The sample was also dominated by budget hotels in motorway and A (trunk road) road locations, with these accounting for almost two-thirds of the respondent hotels. However, this **again reflects the nature of the population distribution for budget hotel locations**. Interestingly, the more recent growth locations of suburban and city centre sites also feature quite strongly, accounting for nearly a further 30 per cent of the sample.

The size distribution shows the 31–40 bedroom range to be the largest single category, followed by the over 60 bedroom group. Cumulatively these two size categories account for 74 per cent of the total. If the 41–50 category were to be added to these this would account for some 90 per cent of the total. **Once again, this is strongly representative of the budget hotel population distribution by size**. Table II [not included here] indicates further characteristics of the sample. This suggests that the sample is very representative of the breadth of budget hotel operations, as it comprises a considerable range of responses in relation to average room occupancy, number of full- or part-time staff and the business mix. Given all of these characteristics it is reasonable to claim that the sample as a whole is highly representative of branded budget hotel operations in the UK.

Source: Brotherton (2004: 949–50)

Key Concept ☞

Target and Realised Samples

Once you have selected the members from the sampling frame that you wish to include in your sample this can be referred to as your ‘target’ sample or, put another way, all the potential respondents, individuals or companies, that you would ideally like to collect data from.

However, some of these ‘target’ respondents may refuse, for a variety of reasons, to take part in your study. Even when individuals or companies have initially indicated that they would be prepared to cooperate and provide you with an opportunity to obtain information from them it is not unknown for some of these to change their minds at a later date!

Therefore, there is likely to be a difference between the selected ‘target’ sample and the ‘actual’ or ‘realised’ sample that you eventually achieve. In short, your actual sample will be comprised of respondents, whereas your target sample is likely to turn out to be a combination of those who did respond (respondents) and those who did not (non-respondents). The relative proportions of these two can be a significant issue (see Non-Response Error).

Key Concept ☞

Non-Response Error

Using the budget hotels scenario we used previously let’s say that you identified 50 budget hotels as the ‘target’ sample, for both theoretical and practical reasons, and that this number and type of hotel were regarded as appropriate to give you the data you need to answer your research question/s. If, however, you manage to interview only 25 of the 50 hotels’ general managers then the question arises: how do I know that the responses provided by my realised sample, which is only half of my target sample, are representative of the other half who did not respond? The answer is, unfortunately, you cannot know this with any great certainty either way. In this sense there is an obvious problem.

While it is perfectly reasonable to say that you can be pretty confident that the nature of the responses provided by your actual respondents are accurate and a true reflection of their opinions, views, perceptions, practices etc., it becomes increasingly difficult to extrapolate or generalise this to a wider context the greater the gap between the target and realised samples. This is known as the degree of ‘non-response’ error that exists.

A simple example can illustrate this. If you had managed to interview 45 of the 50 general managers in your target sample then, both numerically and qualitatively, it would not be unreasonable to suggest that the views of the vast majority, i.e. 45 (90 per cent), of your target sample are more likely than not to be representative of the small minority, i.e. 5 (10 per cent), who did not respond. Hence, any threat to the validity of your sample data would be quite small because the extent of any error in this regard due to a very small number of non-respondents would be negligible.

Alternatively, if the degree of non-response was significantly higher, say 20 (40 per cent) of the 50 general managers refused to be interviewed, I think you may agree that it would now be more difficult to be confident that the opinions of the 30 managers interviewed could be regarded as highly likely to be the same as those held by the 20 who were not. At this point it should be obvious that there is an inverse relationship between the response and non-response rates: the higher the response rate, the lower the non-response rate, and non-response error, will be, and vice versa.

Furthermore, while it is obvious that non-response error has a size or numerical dimension, i.e. the larger the percentage of non-response the greater the threat to overall validity, it may also have a qualitative dimension where the pattern of non-response is uneven. This is an issue related to the composition of the target sample.

Continuing with our budget hotel example, let's say that we designed our original target sample to include 25 female and 25 male general managers, the reasons for which may be quite obvious, but that the 20 non-responding general managers were comprised of 15 females and 5 males. Now, not only do we have the numerical non-response threat discussed earlier but we also have a final, realised sample that has a much higher proportion of male respondents than females. It is clear to see that this presents a potential problem as the views of male general managers are now over-represented compared to those of females in our realised sample, whereas the intention was to obtain a much more balanced set of views by gender in our original sample design.

See also the Technique Tip box – Dealing with Non-Response Error.

Technique Tip

Dealing with Non-Response Error

There are two basic types of non-response: unit or item. The former refers to a nil-return, or non-completion, of the questionnaire, while the latter is concerned with missing responses to some question items. Unit non-response frequently occurs due to an inability to contact the potential respondent, a lack of willingness, ability or refusal to respond. Item non-response can also occur for similar reasons and also be due to errors in the construction of the questionnaire.

The best strategy for dealing with potential non-response error is to try to avoid or reduce it happening in the first place. This can be achieved by ensuring that the questionnaire being used is clear, concise and easy to complete to encourage more target respondents to become actual respondents. Put simply, the more interesting the questions and the easier they are to understand and answer the more likely it is that people will willingly cooperate. Ensuring that the target respondents have been properly and accurately identified in the first place, both in terms of who they are and where/how they can be contacted, can help to increase questionnaire return rates. If you are intending to conduct a survey within an organisation, recruiting a supporter, or champion, can help. Such a person will encourage his/her colleagues to cooperate and may even help you with part of the distribution/collection process.

In the case of item non-response this is often relatively insignificant. If there are just a few missing responses to a few questions on your questionnaire, which is quite common, then you can effectively ignore these because the degree of error they are likely to generate will be insignificant. Where this is more widespread it is possible to use a range of weighting, or re-weighting, procedures to deal with these issues but these are really beyond the scope of this text and are not likely to be relevant to the vast majority of student research work, particularly at undergraduate level.

Finally, when you come to write up your results do not ignore or try to hide any non-response, either unit or item, that has occurred because the people who are supervising and/or marking your work will be smart enough to recognise it and will assume that you have not if you don't deal with it. Be honest and record where it has arisen and explain what you believe the effects of this may have been on your results and the conclusions you can draw from these.

Key Decisions

Dealing with Your Sampling Demons

Most sensible researchers recognise that sampling is hardly ever a perfect process and, frequently, is one that is driven by a combination of ideal and realistic decisions, and by various compromises and trade-offs. Any supervisor or marker of your work should be aware of this and will understand the constraints you are working under. The key point here is that you will not be expected to create the 'perfect' sample for your work but the best one you can manage given the constraints you face. So, although in many cases random sample selection is ideal, being able to achieve this may be difficult, if not impossible. Not being able to achieve this is not the key issue. What is of vital importance is to explain why this has been the case and what would have been required to do this. Explain and justify why you were not able to achieve such a sample and be clear and honest about what your alternative was. If you didn't manage to achieve as large a sample as you intended explain why this happened and recognise what limitations this has created for the analysis of the data and the generalisability of the results.

If you make sampling decisions that are illogical and cannot be justified in any reasonable manner then you might rightly expect to be criticised for this. A more common sampling problem encountered by students, for which they are likely to be correctly penalised, is what is claimed on the basis of the sample. I cannot recall the number of occasions when I have read student dissertation work that claims the results derived from a convenience sample can be generalised! In the general pecking order of sampling, convenience sampling is regarded as distinctly inferior to random sampling but this is only the case if the aim is to generalise the results from the sample back to the population it was drawn from.

There is nothing inherently wrong with convenience sampling *per se* and, as this is often the most feasible choice for student researchers, you should not feel that this will somehow be regarded as such by your supervisor or marker. As long as you provide a sensible explanation for why you made, or had to make, this choice and do not make unjustifiable claims about being able to generalise the results from this you should not be penalised. In fact, you are likely to receive credit for recognising the limitations of convenience sampling because this shows you are aware of them.

Technique Tip	Generating Random Numbers

Using Excel to generate random numbers

[1] Open a new worksheet, go to the Tools menu and select 'Data Analysis'.

[2] Select 'Random Number Generation', click on 'OK' and this will open a dialogue box so that you can set the requirements.

[3] For the 'Number of variables', enter '1'. This will provide a list of random numbers in one column of the spreadsheet.

[4] In the 'Number of Random' option, enter the number of random numbers you need. For example, if you want to select a sample of 50 from a sampling frame of 300, then enter '50' here.

[5] In the 'Distribution' option, select 'Uniform'.

[6] In the 'Between' option, enter the lowest and highest values between which you want the random selection to be made. So, in our example above, this would have been between 1 and 300.

[7] Click on the 'OK' button and you should have a list of 50 random numbers in the column on your spreadsheet.

[8] You may want to tidy up the output as it will be displayed to a number of decimal places, unless you have previously specified the cell output to be different to this. To do this, go to the Format menu and select 'Cells' to open the cell formatting options box. In the 'Number' option, simply change the output to one with no decimal places and you will get whole numbers in the column.

Using SPSS to generate random numbers

Let's say you want 200 random numbers between 1 and 1000.

[1] Start up SPSS and go to the Variable View, choose a variable, and type in a name for it (e.g. 'Rand'). Now format the 'Decimals' to 0 to make SPSS produce the random numbers to the nearest whole number.

[2] Remaining in the Variable View, set the column labelled 'Measure' to 'Scale'.

[3] Now return to the Data View. Go to your 'Rand' column (variable) and enter any number in the 200th row of this column to inform SPSS you want 200 random numbers.

[4] In the Transform menu, click 'Compute variable' and in the 'Target Variable' box, type the name of the column ('Rand') where your random numbers are to go.

[5] Go to the ‘Function group’ box, and select ‘Random Numbers’. A list of types of random numbers will appear. Double-click ‘Rv.Uniform’ in the list of types of random numbers and RV.UNIFORM(?,?) will appear in the ‘Numeric Expression’ box.

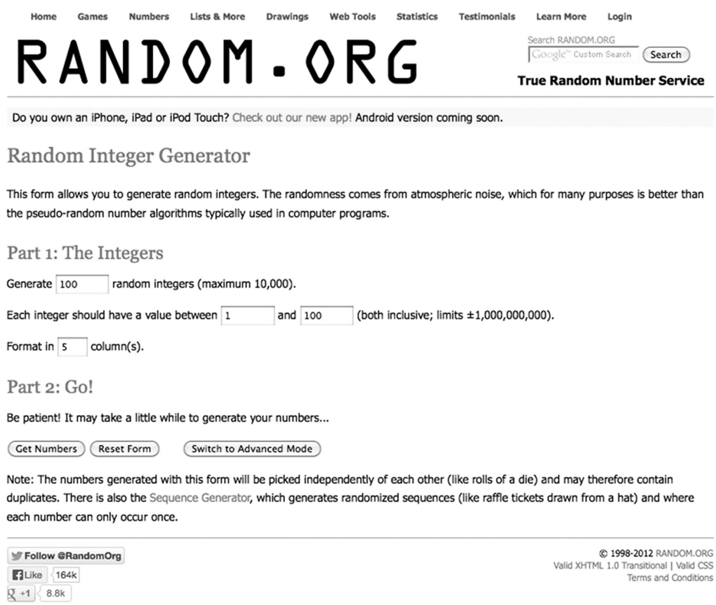
[6] The (?,?), is designed to specify the range of random numbers, so for 200 random numbers between 1 and 1,000 you simply replace the two question marks with a 1 and a 1,000, thus (1,1,000).

[7] Click ‘OK’ and a message pop-up will appear, asking if you want to ‘Change the existing variable’.

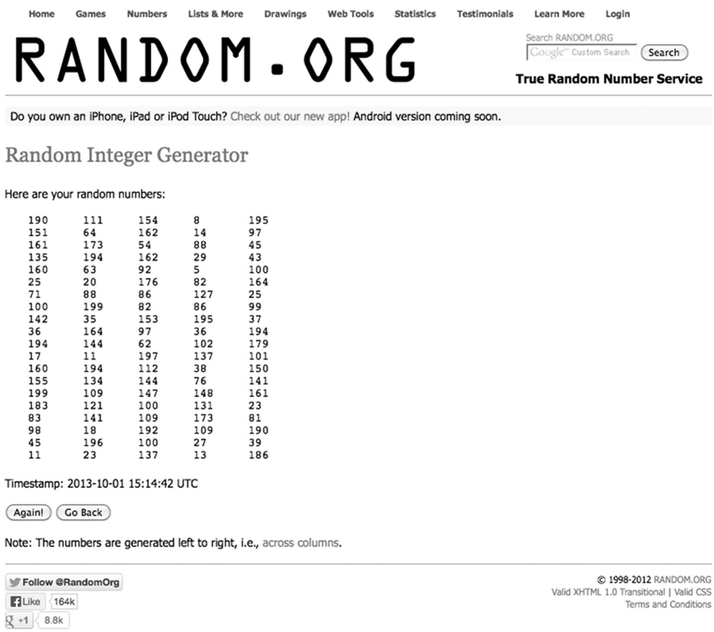
[8] Click ‘OK’ and you will see that your ‘Rand’ column now has 200 numbers in it, between 1 and 1,000.

Using websites to generate random numbers

As you might expect, there are a range of online random number generators available. A simple search using the search term ‘online random number generator’ will deliver many hits, such as that to be found at www.random.org/integers/. The opening page for this is shown in the following screenshot:



The results page (which appeared in a flash!) of a request for 100 random numbers between 1 and 200 is shown below.



Home Games Numbers Lists & More Drawings Web Tools Statistics Testimonials Learn More Login

RANDOM.ORG

Search RANDOM.ORG

Google™ Custom Search Search

True Random Number Service

Do you own an iPhone, iPad or iPod Touch? Check out our new app! Android version coming soon.

Random Integer Generator

Here are your random numbers:

190	111	154	8	195
151	64	162	14	97
161	173	54	88	45
135	194	162	29	43
160	63	92	5	100
25	20	176	82	164
71	88	86	127	25
100	199	82	86	99
142	35	153	195	37
36	164	97	36	194
194	144	62	102	179
17	11	197	137	101
160	194	112	38	150
155	134	144	76	141
199	109	147	148	161
183	121	100	131	23
83	141	109	173	81
98	18	192	109	190
45	196	100	27	39
11	23	137	13	186

Timestamp: 2013-10-01 15:14:42 UTC

Again! Go Back

Note: The numbers are generated left to right, i.e., across columns.

Follow @RandomOrg

Like 164k

+1 8.8k

© 1998-2012 RANDOM.ORG
Valid XHTML 1.0 Transitional | Valid CSS
Terms and Conditions

Screenshots with permission of www.random.org/integers

Key Concept	Key Informants

In qualitative, or indeed mixed-method, research it is frequently recognised that not all potential respondents are equal. Some are likely to be more knowledgeable, experienced and/or be in positions that enable them to have insights that others do not. Such people may be referred to as 'Key Informants' and are likely to be primary data collection targets for the qualitative researcher.

Although, by definition, these people are likely to be limited numerically, there may be situations where a relatively high number of these could be identified. In such circumstances it may be necessary to select a sample of these rather than try to collect data from all of them. Of course, this raises the question of how this sample should be determined. One way of addressing this issue can be to select them at random to avoid potential criticisms of bias. Hence, although the overall sampling strategy is non-random, certain components of this might be selected by using random procedures.