5.1.1 Sample, population, and representativeness

The *sample* is the group of participants whom the researcher actually examines in an empirical investigation and the *population* is the group of people whom the study is about. For example, the population in a study might be EFL learners in Taiwanese secondary schools and the actual sample might involve three Taiwanese secondary classes. That is, the target population of a study consists of all the people to whom the survey's findings are to be applied or generalized.

A good sample is very similar to the target population in its most important general characteristics (for example, age, gender, ethnicity, educational background, academic capability, social class, or socioeconomic status) as well as all the more specific features that are known to be related to the variables that the study focuses on (for example, L2 learning background or the amount and type of L2 instruction received). That is, the sample is a subset of the population that is *representative* of the whole population. The issue of representativeness is crucial, because as Milroy and Gordon (2003) along with many other scholars point out, the strength of the conclusions we can draw from the results obtained from a selected small group depends on how accurately the particular sample represents the larger population.

Why don't we include every member of the population in a study? This is a valid question and, indeed, there is one particular survey type which does just that: the 'census'. In most other cases, however, investigating the whole population is not necessary and would in fact be a waste of resources. By adopting appropriate *sampling procedures* to select a smaller number of people to be investigated we can save a considerable amount of time, cost, and effort and can still come up with accurate results—opinion polls, for example, succeed in providing national projections based on as few as 1,000–3,000 respondents. The key question, then, is what exactly we mean by 'appropriate sampling procedures'?

5.1.2 Sampling procedures

Broadly speaking, sampling strategies can be divided into two groups: (a) scientifically sound 'probability sampling', which involves complex and expensive procedures that are usually well beyond the means of applied linguists, and (b) 'non-probability sampling', which consists of a number of strategies that try to achieve a trade-off, that is, a reasonably representative sample using resources that are within the means of the ordinary researcher.

Probability sampling

Probability sampling is a generic term used for a number of scientific procedures, the most important of which are the following:

- *Random sampling* The key component of probability sampling is 'random sampling'. This involves selecting members of the population to be included in the sample on a completely random basis, a bit like drawing numbers from a hat (for example, by numbering each member and then asking the computer to generate random numbers). The assumption underlying this procedure is that the selection is based entirely on probability and chance, thus minimizing the effects of any extraneous or subjective factors. As a result, a sufficiently large sample should contain subjects with characteristics similar to the population as a whole. Although this is rarely fully achieved, the rule of thumb is that random samples are almost always more representative than non-random samples.

- *Stratified random sampling* Combining random sampling with some form of rational grouping is a particularly effective method for research with a specific focus. In 'stratified random sampling' the population is divided into groups, or 'strata', and a random sample of a proportionate size is selected from each group. Thus, if we want to apply this strategy, first we need to identify a number of parameters of the wider population that are important from the point of view of the research in a 'sampling frame'—an obvious example would be a division of males and females—and then select participants for each category on a random basis. A stratified random sample is, therefore, a combination of randomization and categorization. In studies following this method, the population is usually stratified on more than one variable and random samples are selected from all the groups defined by the intersections of the various strata (for example, we would sample female learners of Spanish, aged 13–14, who attend a particular type of instructional programme in a particular location).

- *Systematic sampling* In anonymous surveys it can be difficult to make a random selection because we may have no means of identifying the participants in advance and thus their names cannot be 'put in the hat' (Cohen *et al.* 2000). A useful technical shortcut is in such cases to apply 'systematic sampling', which involves selecting every *n*th member of the target group.

- *Cluster sampling* One way of making random sampling more practical, especially when the target population is widely dispersed, is to randomly select some larger groupings or units of the populations (for example, schools) and then examine all the students in those selected units.

It is clear from these brief descriptions that selecting a truly representative sample is a painstaking and costly process, and several highly technical monographs have been written about the topic (for example, Cochran 1977; Levy and Lemeshow 1999). However, it needs to be reiterated that in most applied linguistic research it is unrealistic or simply not feasible to aim for perfect representativeness in the psychometric sense.

Non-probability sampling

Most actual research in applied linguistics employs 'non-probability samples'. In qualitative research such purposive, non-representative samples may not be seen as a problem (see Section 6.2), but in quantitative research, which always aims at representativeness, non-probability samples are regarded as less-than-perfect compromises that reality forces upon the researcher. We can distinguish three main non-probabilistic sampling strategies:

- *Quota sampling* and *dimensional sampling* 'Quota sampling' is similar to proportional stratified random sampling without the 'random' element. That is, we start off with a sampling frame and then determine the main proportions of the subgroups defined by the parameters included in the frame. The actual sample, then, is selected in a way as to reflect these proportions, but within the weighted subgroups no random sampling is used but rather the researcher meets the quotas by selecting participants he/she can have access to. For example, if the sampling frame in a study of 300 language learners specifies that 50 per cent of the participants should come from bilingual and the other 50 per cent from monolingual families, the researcher needs to recruit 150 participants from each group but the selection does not have to be random. 'Dimensional sampling' is a variation of quota sampling: the researcher makes sure that at least one representative of every combination of the various parameters in the sampling frame is included in the sample.
- *Snowball sampling* This involves a 'chain reaction' whereby the researcher identifies a few people who meet the criteria of the particular study and then asks these participants to identify further appropriate members of the population. This technique is useful when studying groups whose membership is not readily identifiable (for example, teenage gang members) or when access to suitable group members is difficult for some reason.
- *Convenience or opportunity sampling* The most common sample type in L2 research is the 'convenience' or 'opportunity sample', where an important criterion of sample selection is the convenience of the researcher: members of the target population are selected for the purpose of the study if they

meet certain practical criteria, such as geographical proximity, availability at a certain time, easy accessibility, or the willingness to volunteer. Captive audiences such as students in the researcher's own institution are prime examples of convenience samples. To be fair, convenience samples are rarely completely convenience-based but are usually partially purposeful, which means that besides the relative ease of accessibility, participants also have to possess certain key characteristics that are related to the purpose of the investigation. Thus, convenience sampling often constitutes a less controlled version of the quota sampling strategy described above.

No matter how principled a non-probability sample strives to be, the extent of generalizability in this type of sample is often negligible. It is therefore surprising that the majority of empirical research in the social sciences is not based on random samples. However, Kemper *et al.*'s (2003: 273–4) conclusion is very true:

> Sampling issues are inherently practical. Scholarly decisions may be driven in part by theoretical concerns, but it is in sampling, perhaps more than anywhere else in research, that theory meets the hard realities of time and resources. ... Sampling issues almost invariably force pragmatic choices.

In any case, because of the compromised nature of non-probability sampling we need to describe in sufficient detail the limitations of such samples when we report the results, while also highlighting the characteristics that the particular sample shares with the defined target population. In a similar vein, we also have to be particularly careful about the claims we make about the more general relevance of our findings.

5.1.3 How large should the sample be?

When researchers ask the question, 'How large should the sample be?' what they really mean is 'How small a sample can I get away with?' Therefore, the often quoted 'the larger, the better' principle is singularly unhelpful for them. Unfortunately, there are no hard and fast rules in setting the optimal sample size; the final answer to the 'how large/small?' question should be the outcome of the researcher considering several broad guidelines:

- *Rules of thumb* In the survey research literature a range of between one per cent to ten per cent of the population is usually mentioned as the magic sampling fraction, with a minimum of about 100 participants. However, the more scientific the sampling procedures, the smaller the sample size can be, which is why opinion polls can produce accurate predictions from samples as small as 0.1 per cent of the population. The following rough estimates of sample sizes for specific types of quantitative methods have also been agreed on by several scholars: correlational research—at least 30 participants; comparative and experimental procedures—at least 15

participants in each group; factor analytic and other multivariate procedures—at least 100 participants.

- *Statistical consideration* A basic requirement in quantitative research is that the sample should have a 'normal distribution', and Hatch and Lazaraton (1991) argue that to achieve this the sample needs to include 30 or more people. However, Hatch and Lazaraton also emphasize that smaller sample sizes can be compensated for by using certain special statistical procedures (for example, non-parametric tests—see Section 9.9).
- *Sample composition* A further consideration is whether there are any distinct subgroups within the sample that may be expected to behave differently from the others. If we can identify such subgroups in advance (for example, in most L2 studies of school children, girls have been found to perform differently from boys), we should set the sample size so that the minimum size applies to the *smallest subgroup* in the sample.
- *Safety margin* When setting the final sample size, it is advisable to leave a decent 'margin' to provide for unforeseen or unplanned circumstances. For example, some participants are likely to drop out of at least some phases of the project; some questionnaires will always have to be disqualified for one reason or another; and we may also detect unexpected subgroups that need to be treated separately.
- *Reverse approach* Because statistical significance (see Section 9.4.4) depends on the sample size, our principle concern should be to sample enough learners for the expected results to be able to reach statistical significance. That is, we can take a 'reverse approach': first we approximate the expected magnitude or power of the expected results and then determine the sample size that is necessary to detect this effect if it actually exists in the population. For example, at a $p < .05$ significance level an expected correlation of .40 requires at least 25 participants. (These figures can be looked up in correlational tables available in most texts on statistics; see, for example, http://www.uwe.ac.uk/hlss/llas/statistics-in-linguistics/Appenix1.pdf.) Researchers can also use a more complex procedure, 'power analysis', to decide on the necessary sample size. (For a free, Internet-based power analysis software see Buchner *et al.* 1997.)