field is a point of discussion, but it can lead to bias in the sample that may influence the results.

SAMPLING ERRORS

Errors in this domain are not mistakes, but inaccuracies in resulting data. *Sampling errors* occur simply because data are being collected on a sample and not on the population. The first four sampling techniques considered in the previous section were devised to enhance the probability that any sample would be representative of a population, thus endeavouring to minimize these errors. To illustrate this, consider the distribution of IQ scores for a small population of 3000 shown in curve (a) of Figure 5.3. Beneath this is the distribution for a single sample of 40 subjects, curve (b). Note that its mean, $\bar{x}_A$, is not exactly that of the population, μ, nor would we necessarily expect it to be, due to natural random variability. One way to quantify the collective errors attributable to sampling for interval or ratio data is to consider what would happen if we were to take sample after sample from a population. For each sample, the mean could be found and plotted on a graph, providing us with a *distribution of sample means*. It is expected that for representative samples, though, these means would be very close to that of the population mean. It is also assumed that since this sampling error would be random, the distribution of sampling means would be normal with the mean of the sample means being the population mean, as in the narrow peak shown in curve (c) of Figure 5.3. This natural variability in a set of sample means of scores/measures around the population mean provides an opportunity to describe the outcome mathematically.

Hays (1994) provides a rigorous discussion of the calculation of sampling errors, probabilities and the variance of this distribution, with respect to sampling for surveys, for those who are interested. The derivations given, as well as justification for them, do generate a commonly used and conceptually useful estimation. We use an estimate of the distribution standard deviation simply because it is not reasonable to take all the possible samples from a population and plot a graph of the means. Usually described as the *standard error of the mean* (the square root of this variance), it is estimated from the population standard deviation and the sample size as follows:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{5.1}$$

where σ is the standard deviation of the population and n is the size of the samples. Its dependence on the size of the samples can be seen from the equation: note that the larger the samples, the smaller the variability of the set of sample means. This estimate is based on the assumption that these are simple random samples. Other approaches can produce greater error and consequently a higher value for the standard error of the mean than the one provided by equation (5.1).

Using this value, we can consider some interesting issues. In situations where the population mean is known, it would be possible to determine whether, on the basis of the mean of a sample of size n, it is likely that a particular sample is representative of the population, in other words, whether its mean is within

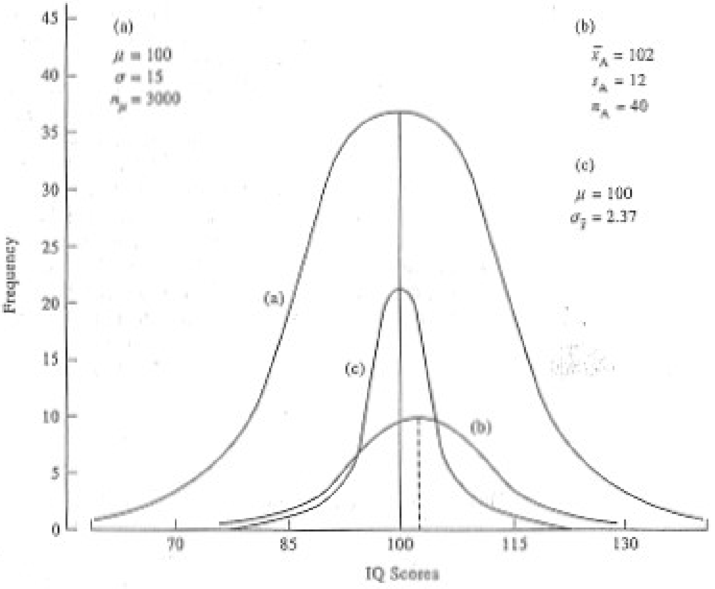

FIGURE 5.3
(a) A population distribution of IQ test scores for all 3000 11-year-olds in a local authority; (b) a single sample distribution for 40 children; (c) the distribution of sampling means for samples size 40

acceptable limits. Obviously, what is 'acceptable' could be an arguable point, but usually it is assumed (as with many statistical tests) that a mean which falls within the range of 95% of all possible means in the normal distribution is representative. When we look at a normal distribution, it is necessary to remember that the area under the curve represents the number of samples. Assuming that the total area represents 100% of all possible samples, then we can identify various percentages. The easiest way to do this is to use the number of standard deviations away from the mean as the x-axis, referred to as a z-score. Thus for a given mean score for a sample and the population mean, we can find a corresponding z-score,

$$z = \frac{\bar{x} - \sigma}{\sigma_{\bar{x}}}$$

where $\bar{x}$ is the sample mean and $\sigma_{\bar{x}}$ is standard error of the mean. Looking at the table showing the areas beyond any given z-score (Table B.1 in Appendix B), we find that 2.5% of the area is beyond 1.96 standard deviations from the mean. Thus, any mean producing a z-score greater than 1.96, is unlikely to be part of the population, as shown in Figure 5.4. This would allow us to reject any null hypothesis that there was no significant difference between the population and sample means. Therefore, there is a difference between the sample and population means greater than what could be attributed to natural variability. The natural expectation is that not all samples will be exactly the same (some possible sources of the extra difference will be considered in the next section). If the difference is too great, it is said to be *statistically* significantly different,

BLACK, Thomas, R. 1999. *Doing Quantitative Research in the Social Sciences. An Integrated Approach to Research Design, Measurement and Statistics*. London: Sage.

because the probability that it belongs to the population is so small. Even if the probability is finite (less than 5%), it is considered to be *so* unlikely that the sample is labelled unrepresentative.

How does one determine this mathematically for a given sample? Let us consider the example of Harry Teacher, who wished to conduct a study on the effectiveness of a new reading programme in primary school. He wanted two equivalent classes of children; he randomly selected two local schools and randomly selected a class in each. In order to try to justify some generalizability of the study based on these two classes of Year 4 children, he decided to see what their mean IQs were and whether they were typical. He found from the teacher that one class of 30 had a mean IQ of $\bar{x}_A = 104$, and the other $\bar{x}_B = 99$. First he had to find the standard error of the mean. Knowing that the population mean for IQ scores was 100 with a standard deviation of 15,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{15}{\sqrt{30}}$$

$$\sigma_{\bar{x}} = 2.74$$

This provides a value for the standard error of the mean for simple random samples of 30 children. Thus the *z*-score for the two means would be

$$z_A = \frac{99 - 100}{2.74} = \frac{-1}{2.74} = -0.36$$

$$z_B = \frac{104 - 100}{2.74} = \frac{4}{2.74} = 1.46$$

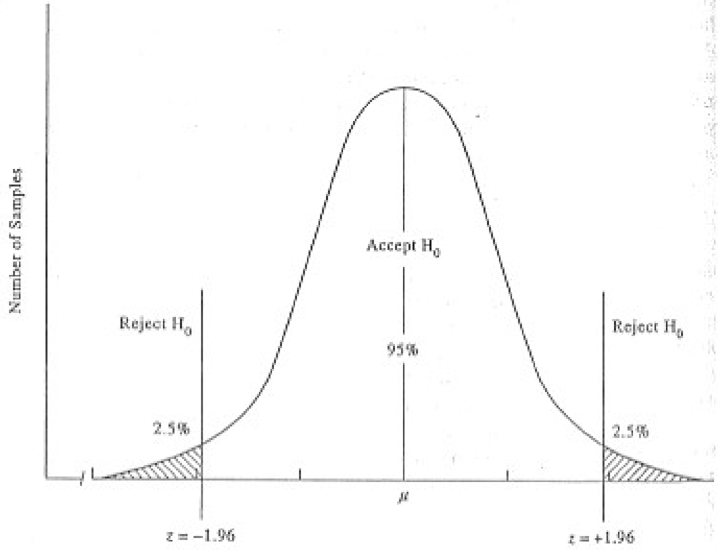

FIGURE 5.4
Normal distribution of sample means with 5% significance levels, where *m* is the population mean and the *z*-scores are the means expressed as number of standard deviations above and below the mean

Since neither of these means was more than 1.96 standard deviations from the population mean, he could assume that the samples were representative of the population, at least for IQ scores. This does not protect him against other traits and experiences for which the two classes might not be typical, but it does provide some support.

Fowler (1993) points out that this estimate of the standard error of the mean is only valid if one has taken a simple random sample. Stratified random samples tend to produce lower values for the standard error of the mean, while cluster samples tend to have higher values than for simple random samples. Thus, Harry Teacher's analysis may produce an underestimate of the standard error of the mean, since his selection of clusters of students may produce a more homogeneous grouping for the trait than that for the population. In some situations, this may be an important factor and may actually influence the results, but here the aim is to test the relative representativeness of the sample.

If population parameters are not available, the standard error of the mean can be estimated using sample statistics,

$$s_{\bar{x}} = \frac{s_A}{\sqrt{n_A}} \tag{5.2}$$

where s_A is the standard deviation of sample group *A* and n_A is the sample size of sample group *A*. It does give one some indication of the error, though strictly it also applies just to simple random sampling. This situation allows us to make a different type of statement about the results: is the sample mean, $\bar{x}_A$, close enough to the population mean, μ? Using the estimate of the standard error of the mean in equation (5.2), it is possible to establish a *confidence interval*, an interval of scores in which we can be reasonably confident that the population mean will fall. For example, if we wish to establish an interval in which we are 95% confident that the population mean occurs, then it will be

$$\bar{x}_A \pm 1.96\, s_{\bar{x}} \tag{5.3}$$

For example, if a simple random sample of 25 children had a mean score on a school-based test in mathematics, $\bar{x}_A = 71.60$, with a standard deviation, $s_A = 12.20$, then

$$s_{\bar{x}} = \frac{12.20}{\sqrt{25}} = 2.44$$

and therefore the 95% confidence interval for the population mean, μ, would be

$$71.60 \pm 1.96 \times 2.44$$

or

$$71.60 \pm 4.78$$

Alternatively, it could be expressed as

$$66.82 < \mu < 76.38$$

This means that for 19 out of 20 samples, we could expect the true population mean to be within the interval. This is illustrated in Figure 5.5 as mean scores with their corresponding confidence intervals for 20 samples with 19 of them overlapping with the population mean, μ, while one does not.

Obviously it would also be possible to calculate 99% confidence intervals as well, if needed. Examining the equation, it becomes apparent that the best way

BLACK, Thomas, R. 1999. *Doing Quantitative Research in the Social Sciences. An Integrated Approach to Research Design, Measurement and Statistics.* London: Sage.

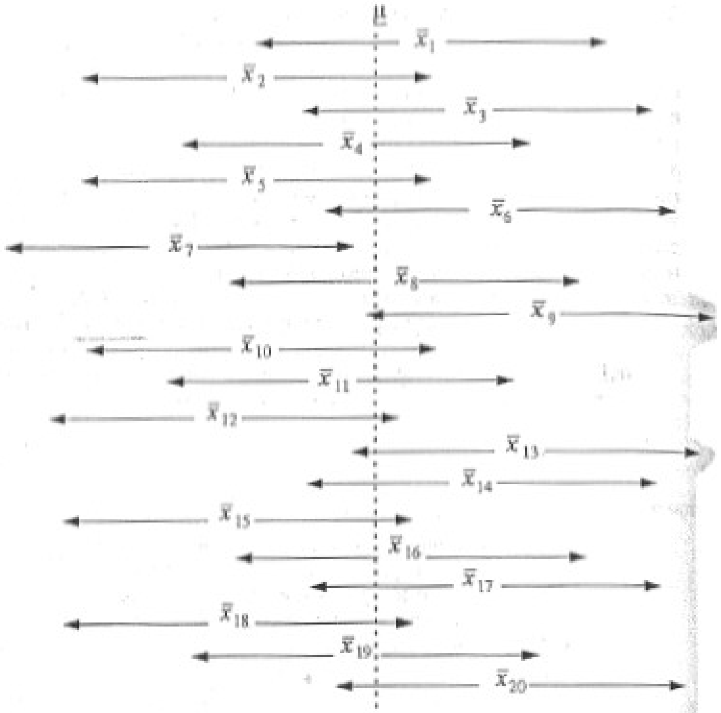

FIGURE 5.5 Twenty samples with 95% confidence intervals and true mean: which one does not include *m*?

to reduce the confidence interval would be to increase the sample size (increase the denominator and the overall value increases). This is not the only consideration of importance and again one must remember that the sampling technique may have a considerable influence on population estimates. More accurate estimates of the standard error of the mean may be lower for stratified samples and higher for cluster samples, thus altering the confidence interval. The importance of this, though, is dependent on the nature of the inferences to be made about the populations.

This is all satisfactory as long as we are considering interval data, I hear you say. So let us consider another example, one where the trait under consideration is nominal (a set of categories) rather than interval or ratio. In a recent study in Uganda in which the author participated (Black *et al.*, 1998), it was necessary to employ a sampling technique that had to be seen as representing secondary schools from each of the 39 districts in the country. The sampling frame consisted of a list of 550 government-supported secondary schools grouped by district. A simple random sample of all schools could have left some sparsely populated districts out of the study, so it was decided there would be at least one school randomly selected from each district (stratum) and districts with over three schools would be proportionally more randomly selected. Thus each district was represented by at least one school and larger districts had a larger representation, providing a sample of 77 schools (14% of the total). Having agreed on this, it was necessary to justify that the sample was still representative of schools in the country for other important traits: day versus boarding, separate sex versus mixed, and foundation body (parents, religious group, etc.). To test the hypotheses that the sample was representative of the whole population of schools for each of these three traits required the use of a non-parametric test: chi-square, χ^2. This involves comparing the observed frequencies of occurrence in each category of the sample with what one would expect from national frequencies. Analogous to normal distributions of interval data for samples which did not match exactly with the population, we would not expect samples to have exactly the same proportions of categories as the population. Again, the question must be asked, how much is too much of a deviation?

Table 5.2 provides the data, with the observed frequencies, O_i, for each type of school foundation body in the first column. The second column has the national percentages and the third contains the frequencies based on what would be expected if the 77 schools were *exactly* like the national distribution, the expected frequency, E_i. Visual inspection of the data does not show much difference, but the χ^2-test provided an indication of the probability that the sample was one that was representative of the whole population. (We will come back to this test later in Chapter 19, and investigate it in greater depth then.)

Roughly speaking, the test sums the absolute differences in frequencies between the two groups for all the characteristics, then determining whether the total is more, or less, than what would be expected by chance. Since the differences between observed and expected frequencies could be either positive or negative, the differences are squared before being added together. Otherwise, the total could approach zero even when there were large differences, the negative cancelling the positive values. The formula for the χ^2-statistic is quite simple, the sum of all the differences squared, each divided by the appropriate expected frequency,

$$\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} \qquad (5.4)$$

where the O_i are the observed frequencies in the sample, the E_i are the expected frequencies based upon population percentages, and i represents the categories

TABLE 5.2 Distributions of Ugandan secondary schools by foundation body: a sample and as expected from national data

Foundation body	Observed O_i (sample)	National percentage	Expected E_i
Church of Uganda	35	42.1	32.4
Catholic Church	15	21.2	16.3
Parents	13	16.4	12.6
Muslim Schools Council	1	1.1	0.8
Government	6	7.7	5.9
United Muslim Schools	3	5.2	4.0
Other	4	6.3	4.9
Totals	77	100.0%	77.0

ranging from 1 to m. Note, when consulting tables for this statistic (see, for example, Table B.9 in Appendix B), that it has $df = m - 1$ degrees of freedom.

The easiest way to process such data is to add another column to the table, in which the differences squared divided by the expected frequency are placed. As we will see later, this is conveniently done on a spreadsheet, but for small amounts of data, the calculation is easily carried out on a calculator. The column values are added as shown in Table 5.3, and the sum compared to the standard table for χ^2. Which value one uses depends on the number of categories minus one: in the example, there are seven foundation bodies, thus six *degrees of freedom*, since six frequencies could vary, but then the seventh would be fixed. Table B.9 in Appendix B provides a value of 14.07, thus a χ^2-ratio greater than this would indicate a significant difference between the sample and the population. In the example in Table 5.3, the final value is 0.74, nowhere near significant ($p \gg 0.05$): thus, at least for this trait, the sample of secondary schools could be said to reflect the national pattern.

Sampling errors are unavoidable, though hopefully minimized through sound sampling procedures. On the other hand, *non-sampling errors* can be attributed to such processes as incorrect sampling frame, poor measuring instruments, incorrect data processing and non-response by subjects in the sample. The definition of sampling frames was considered earlier, and incorrect ones are equivalent to defining the population to be all voters and using a telephone number list as the sampling frame. The skills related to the design of measuring instruments and data collection are the subjects of Chapters 8–11, and the choice of statistical test is covered in Chapters 13–22. This leaves non-response, which is to be considered in the next section with suggestions as how to minimize it and any effects that it may have on the validity of the results. Before going on, carry out Activity 5.3.

BLACK, Thomas, R. 1999. *Doing Quantitative Research in the Social Sciences. An Integrated Approach to Research Design, Measurement and Statistics.* London: Sage.