NULL HYPOTHESES

The rather convoluted thinking of null hypotheses is necessary if we are going to set the scene for testing hypotheses using statistical tools. As noted in Chapter 1, theories survive and gain support as a result of not being disproved, rather than being proven conclusively. For sound theories, this does not imply a ticking bomb waiting to explode in the form of some researcher in the future proving it wrong. What it does suggest is that researchers are usually trying out components of a theory in different situations or with different groups; they are looking for the *limits* of applicability or refinements in detail. Hypotheses, as described above, express anticipated outcomes as predicted by a given theory or the expected consequence of an application of principles to a situation, stated in more specific terms than those of a general research question.

When it comes to testing hypotheses, all that statistics can tell us is whether the outcomes we ultimately see could have happened due to some causal relationship *or* simply by chance alone. In other words, the effect has to be big enough, whether it is the difference in average scores on some performance task for two groups, or the size of a correlation coefficient. The null hypothesis simply states that 'no significant difference' is expected between what we obtain and what would happen by chance alone. If the difference observed is greater than some minimum, then it is considered significant and whatever has happened (probably) did not occur by chance alone. It is still up to the researcher to prove through sound design and data collection that nothing could have caused the observed effect other than what is described in the hypothesis.

So the next stage in refining our statement of hypotheses would be to try to express them as null hypotheses related to the data that will be collected. As a consequence of a given study, several types of null hypothesis could be generated – for example, describing differences in scores or frequencies of events between the sample and the population (normative), *or* between two groups or among three or more groups – i.e., they actually belong to the same population, not to separate populations (experimental, quasi-experimental or ex post facto). The statements simply anticipate that any difference(s) will be too small to be attributable to anything but chance.

Alternatively, if one were carrying out a correlational study, the null hypothesis of 'no significant correlation' anticipates correlations that will be so small that they could have happened by chance alone. To illustrate this, the hypotheses of Table 2.2 above are provided in Table 2.3 with corresponding possible null hypotheses.

The process of specifying a null hypothesis is one that focuses the attention on what will happen next, stating the implications of the proposed relationship among variables in terms that can be resolved by statistical instruments (see Figure 2.14). At this stage, it is sometimes possible to identify potential difficulties in carrying out the research. For example, where are we going to find the

TABLE 2.3 Hypotheses from Table 2.2 and potential corresponding null hypotheses

Hypotheses	Null hypotheses
A random sample of assembly-line workers in factories in Birmingham will be found to suffer a greater frequency of sleep interruptions, and a longer amount of time awake after going to bed, than the population as a whole.	(Both of the hypotheses assume that population data exist.) There will be no significant difference between the mean number of times per night that assembly-line workers in Birmingham awaken and the mean for the population of employed adults as a whole, or between the mean number of minutes that these workers are awake per night and that for the population of employed adults.
One of three counselling approaches, A, B or C, will produce a greater reduction in frequency of return to drinking among alcoholics.	There will be no significant difference frequencies of 'dry' and return drinkers across three equivalent sets of alcoholics participating in the three counselling approaches, A, B, C.
It is expected that there will be a negative correlation between social class and drug use, and a negative correlation between educational achievement and drug use for a representative selection of 18–24-year-olds.	There will be no significant correlation between social class and frequency of drug use, or between educational achievement and frequency of drug use for a random selection of 18–24-year-olds (i.e., any correlation will not differ from that which could be expected by chance alone).
For a sample of identical twin boys who are the sons of alcoholic fathers *and* fostered or adopted from infancy separately from each other, one to a family with at least one alcoholic parent, one group will show a greater tendency towards alcoholism than the other.	There will be no significant difference in frequency of alcoholism between groups of separated twins, all sons of alcoholics, when one twin goes to a family with at least one alcoholic parent and the other goes to a family with no alcoholic parents.
In a given hospital, patients on 24-hour prescriptions will be expected to feel more rested if they are awakened for medicines at times that follow REM rather than just at equal time intervals.	There will be no significant difference in the perception of feeling rested, as measured by the Bloggs Restedness Scale completed by patients, between two groups: those whose medication was administered at regular time intervals and those whose medication was administered at times close to times prescribed but following a period of REM.

sample of twins implied by the fourth proposal in Table 2.3? Some of the more interesting questions generate very difficult scenarios for resolving them, compelling researchers to rethink the hypotheses resulting from a question. Obviously, it is better to consider such issues early in the research process before too much is invested in an impossible task.

Testing the null hypothesis

For normally distributed traits, those that produce sample means out in either of the tails of a distribution of sampling means are highly unlikely. Social science researchers commonly accept that events which occur less frequently than 5% of the time are unlikely to have occurred by chance alone and consequently are considered statistically significant. To apply this to a normal distribution would mean that the 5% must be divided between the top and the bottom tails of the distribution, with 2.5% for each (there are occasions when all 5% would occur in one tail, but that is the exception, to be discussed later). Consulting Table B.1 in Appendix B, the top 2.5% is from 47.5% onward, or (interpolating) 1.96 standard deviations (SEMs) or more from the mean. The two ranges of sample means that would be considered *statistically significant*, and result in the rejection of the null hypothesis since they probably did not occur as part of the natural chance variation in the means, are shown shaded in Figure 13.8.

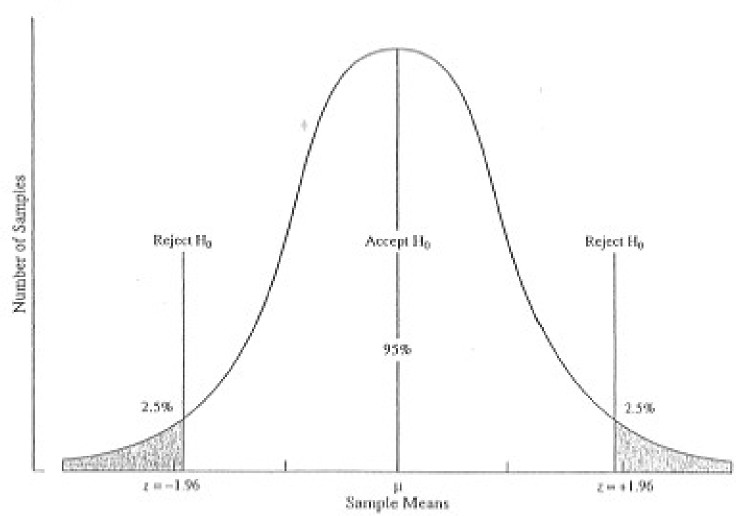

FIGURE 13.8
Normal distribution of sample means with 5% significance levels, where μ is the population mean

Thus for the situation above involving the mean IQ of the sample of 11-year-olds, the null hypothesis and the statement of expected outcomes need an addition:

. . . and, is the probability that the difference between the sample mean and the population mean would occur naturally more or less than 5% (the chosen level of significance that will be used as the test criteria)?

The cut-off point of 1.96 standard deviations (SEMs) would correspond to $1.96 \times 2.5 = 4.9$ points above or below the mean. Thus a sample mean IQ of less than 95.1 or greater than 104.9 would be considered significant and the sample not representative of the population. Therefore, in the example, the group with a mean IQ of 106 would be considered statistically significant and the group not typical, and it is unlikely that they are a representative sample of the whole population, for IQ.

Some researchers present results that are supported by an even lower level of probability, usually designated by the Greek letter α, to support their argument, such as 1% ($\alpha = 0.01$), 0.5% ($\alpha = 0.005$), or even 0.1% ($\alpha = 0.001$). Two problems arise with such a practice. First, for the test to be legitimate, one school of thought says the level of significance should be set *before* the test (or even the study) is conducted. Remember that the hypothesis is a statement of expectation, one that should include what will be expected in terms of statistical outcome. It is not fair to write the rules after the game has begun. Second, there is a feeling that a lower significance level than 5% ($p < 0.05$), such as 1% ($p < 0.01$), provides greater support for the results. In other words, if the probability of the relationship existing is only 1 in 100, that must be a stronger statement than if it were only 1 in 20. This supposition will be challenged in Chapter 14 when the concept of the power of a statistical test is introduced.