Chapter 4.

Statements of Probability and Confidence Intervals

We have seen that when a set of observations have a Normal distribution multiples of the standard deviation mark certain limits on the scatter of the observations. For instance, 1.96 (or approximately 2) standard deviations above and 1.96 standard deviations below the mean ($\pm 1.96 SD$) mark the points within which 95% of the observations lie.

Reference Ranges

We noted in Chapter 1 that 140 children had a mean urinary lead concentration of 2.18 $\mu mol/24hr$, with standard deviation 0.87. The points that include 95% of the observations are 2.18 ± (1.96 × 0.87), giving a range of 0.48 to 3.89. One of the children had a urinary lead concentration of just over 4.0 $\mu mol/24hr$. This observation is greater than 3.89 and so falls in the 5% beyond the 95% probability limits. We can say that the probability of each of such observations occurring is 5% or less. Another way of looking at this is to see that if one chose one child at random out of the 140, the chance that their urinary lead concentration exceeded 3.89 or was less than 0.48 is 5%. This probability is usually used expressed as a fraction of 1 rather than of 100, and written $P < 0.05$

Standard deviations thus set limits about which probability statements can be made. Some of these are set out in Table A (appendix). To use to estimate the probability of finding an observed value, say a urinary lead concentration of 4 $\mu mol/24hr$, in sampling from the same population of observations as the 140 children provided, we proceed as follows. The distance of the new observation from the mean is 4.8 - 2.18 = 2.62. How many standard deviations does this represent? Dividing the difference by the standard deviation gives 2.62/0.87 = 3.01. This number is greater than 2.576 but less than 3.291 in , so the probability of finding a deviation as large or more extreme than this lies between 0.01 and 0.001, which maybe expressed as $0.001P < 0.01$ In fact Table A shows that the probability is very close to 0.0027. This probability is small, so the observation probably did not come from the same population as the 140 other children.

To take another example, the mean diastolic blood pressure of printers was found to be 88 mmHg and the standard deviation 4.5 mmHg. One of the printers had a diastolic blood pressure of 100 mmHg. The mean plus or minus 1.96 times its standard deviation gives the following two figures:

88 + (1.96 x 4.5) = 96.8 mmHg

88 - (1.96 x 4.5) = 79.2 mmHg.

We can say therefore that only 1 in 20 (or 5%) of printers in the population from which the sample is drawn would be expected to have a diastolic blood pressure below 79 or above about 97 mmHg. These are the 95% limits. The 99.73% limits lie three standard deviations below and three above the mean. The blood pressure of 100 mmHg noted in one printer thus lies beyond the 95% limit of 97 but within the 99.73% limit of 101.5 (= 88 + (3 x 4.5)).

The 95% limits are often referred to as a "reference range". For many biological variables, they define what is regarded as the normal (meaning standard or typical) range. Anything outside the range is regarded as abnormal. Given a sample of disease free subjects, an alternative method of defining a normal range would be simply to define points that exclude 2.5% of subjects at the top end and 2.5% of subjects at the lower end. This would give an empirical normal range. Thus in the 140 children we might choose to exclude the three highest and three lowest values. However, it is much more efficient to use the mean ± 2 SD, unless the data set is quite large (say >400).

Confidence Intervals (CI)

The means and their standard errors can be treated in a similar fashion. If a series of samples are drawn and the mean of each calculated, 95% of the means would be expected to fall within the range of two standard errors above and two below the mean of these means. This common mean would be expected to lie very close to the mean of the population. So the standard error of a mean provides a statement of probability about the difference between the mean of the population and the mean of the sample.

In our sample of 72 printers, the standard error of the mean was 0.53 mmHg. The sample mean plus or minus 196 times its standard error gives the following two figures:

88 + (1.96 x 0.53) = 89.04 mmHg

88 - (1.96 x 0.53) = 86.96 mmHg.

This is called the ***95% confidence interval***, and we can say that there is only a 5% chance that the range 86.96 to 89.04 mmHg excludes the mean of the population. If we take the mean plus or minus three times its standard error, the range would be 86.41 to 89.59. This is the 99.73% confidence interval, and the chance of this range excluding the population mean is 1 in 370. Confidence intervals provide the key to a useful device for arguing from a sample back to the population from which it came.

The standard error for the percentage of male patients with appendicitis, described in Chapter 3, was 4.46. This is also the standard error of the percentage of female patients with appendicitis, since the formula remains the same if p is replaced by 100 - p. With this standard error we can get 95% confidence intervals on the two percentages:

60.8 $\pm$(1.96 x 4.46) = 52.1 and 69.5

39.2 $\pm$(1.96 x 4.46) = 30.5 and 47.9.

These confidence intervals exclude 50%. Can we conclude that males are more likely to get appendicitis? This is the subject of the rest of the book, namely *inference* .

With small samples - say under 30 observations - larger multiples of the standard error are needed to set confidence limits. This subject is discussed under the *t* distribution (Chapter 7).

There is much confusion over the interpretation of the probability attached to confidence intervals. To understand it we have to resort to the concept of repeated sampling. Imagine taking repeated samples of the same size from the same population. For each sample calculate a 95% confidence interval. Since the samples are different, so are the confidence intervals. We know that 95% of these intervals will include the population parameter. However, without any additional information we cannot say which ones! Thus with only one sample, and no other information about the population parameter, we can say there is a 95% chance of including the parameter in our interval. Note that this does not mean that we would expect with 95% probability that the mean from another sample is in this interval. In this case we are considering differences between two sample means, which is the subject of the next chapter.

Common questions

What is the difference between a reference range and a confidence interval?

There is precisely the same relationship between a reference range and a confidence interval as between the standard deviation and the standard error. The reference range refers to individuals and the confidence intervals to *estimates* . It is important to realize that samples are not unique. Different investigators taking samples from the same population will obtain different estimates, and have different 95% confidence intervals. However, we know that for 95 of every 100 investigators the confidence interval will include the population mean interval.

When should one quote a confidence interval?

There is now a great emphasis on confidence intervals in the literature, and some authors attach them to every estimate they make. In general, unless the main purpose of a study is to actually estimate a mean or a percentage, confidence intervals are best restricted to the main outcome of a study, which is usually a *contrast* (that is, a difference) between means or percentages. This is the topic for the next two chapters.

Exercises

Exercise 4.1 A count of malaria parasites in 100 fields with a 2mm oil immersion lens gave a mean of 35 parasites per field, standard deviation 11.6 (note that, although the counts are quantitative discrete, the counts can be assumed to follow a Normal distribution because the average is large). On counting one more field the pathologist found 52 parasites. Does this number lie outside the 95% reference range? What is the reference range?

Exercise 4.2 What is the 95% confidence interval for the mean of the population from which this sample count of parasites was drawn?