Statistics at Square One

VI—Variation between samples

T D V SWINSCOW

British Medical Journal, 1976, 1, 1585

Last week we noted the importance, when drawing a sample from a population, of ensuring that every member of the population has a known, and usually an equal, chance of being included in the sample. But naturally it does not follow that a series of samples that are drawn from one population and fulfil this criterion will be identical. They will show chance variations from one to another, and the variation may be slight or considerable. For example, a series of samples of the body temperature of healthy people would show very little variation from one to another, but the variation between samples of the systolic blood pressure would be considerable. Thus the variation between samples depends partly on the amount of variation in the population from which they are drawn.

Furthermore, it is a matter of common observation that a small sample is a much less certain guide to the population from which it was drawn than a large sample. In other words, the more members of a population that are included in a sample the more chance will that sample have of accurately representing the population, provided a random process is used to construct the sample. A consequence of this is that, if two or more samples are drawn from a population, the larger they are the more closely are they likely to resemble each other—again provided the random technique is followed. Thus the variation between samples depends partly also on the size of the sample.

Standard error of the mean

If we draw a series of samples and calculate the mean of the observations in each, we have a series of means. These means themselves generally conform with a "normal" distribution, and they often do so even if the observations from which they are obtained do not. The series of means, like the series of observations in each sample, has a standard deviation. The standard error of the mean of one sample is an estimate of the standard deviation that would be obtained from the means of a large number of samples drawn from that population.

As noted above, if a series of random samples are drawn from a population their means will vary from one to another. The variation depends on the variation in the population and the size of the sample. We do not know the variation in the population so we use as an estimate of it the variation in the sample. This is expressed in the standard deviation. If we now divide the standard deviation by the square root of the number of observations in the sample we have the standard error of the mean. $SEM = SD/\sqrt{n}$.

Dr Louise White is a general practitioner in Yorkshire. Her practice includes part of a town with a large printing works and some of the adjacent sheep farming country. With her patients' informed consent she has been investigating whether the diastolic blood pressure of men aged 20 to 44 differs between the printers and the farm workers. For this purpose she has obtained a random sample of 72 printers and 48 farm workers and calculated the means and standard deviations as shown in table 6.1.

TABLE 6.1—Mean diastolic blood pressures in mm Hg of printers and farmers

	Number	Mean diastolic blood pressure	Standard deviation
Printers	72	88	4.5
Farmers	48	79	4.2

To calculate the standard errors of the two mean blood pressures the standard deviation of each sample is divided by the square root of the number of the observations in the sample.

Printers: $SEM = 4.5/\sqrt{72} = 0.53$.

Farmers: $SEM = 4.2/\sqrt{48} = 0.61$.

These standard errors may be used to study the significance of the difference between the two means, as will be discussed in a subsequent article.

Exercise 6. Dr Green found that the mean urinary lead concentration in 140 children was 2.18 $\mu\text{mol}/24\text{ h}$ (Part II), with standard deviation 0.87 (Part IV). What is the standard error of the mean? *Answer:* 0.074.

How long may one keep tinned food without danger to health and what signs, if any, should one watch out for?

There are several factors that determine how long tinned food will keep, including the temperature and humidity at which the tins are kept, and whether or not the containers are damaged and if so to what extent. Moreover, most of these environmental changes affect the colour, flavour, and texture of the foods but not their safety. Given that the tins are kept in cool dry conditions, milk, and fish in tomato sauce, tend to change soonest, so that the time recommended for these is up to one year. For tinned fruits, it is from one to two years. All other foods—including soups, fish in oil, meat, and vegetables—keep satisfactorily for two years. But the recommended times do not at all imply that the foods will be inedible after those times, though they may be unpalatable. Any tin that, even within these periods, becomes blown—that is, has bulging ends—should be discarded. This is the only way to be sure of safety, since the effect may be due to gas produced by bacterial activity. More frequently it is due to acid from, for example, tinned fruits attacking part of the metal not adequately covered by the lacquer lining, but it is best not to assume this and to discard the tin.