Statistics at Square One

VIII—Differences between means

T D V SWINSCOW

British Medical Journal, 1976, **2**, 94-95

We saw in Part VI that the mean of a sample has a standard error, and a mean that departs by more than twice its standard error from the population mean would be expected by chance only in about 5% of samples. Likewise the difference between the means of two samples also has a standard error. We do not normally know the population mean, so we may suppose that the mean of one of our samples estimates it. The sample mean may happen to be identical with the population mean. More likely it lies somewhere above or below the population mean, and there is a 95% chance that it is within 1·96 standard errors above or below it.

Consider now the mean of the second sample. If the sample comes from the same population its mean will also have a 95% change of lying within 1·96 standard errors above or below the population mean. But if we do not know the population mean we have only the means of our samples to guide us. Therefore, if we want to know whether they are likely to have come from the same population, we ask, Do they lie within a certain range, represented by their standard errors, of each other?

Standard error of difference between means

If SD_1 represents the standard deviation of sample 1 and SD_2 the standard deviation of sample 2, and n_1 the number in sample 1 and n_2 the number in sample 2, the formula denoting the standard error of the difference between two means is:

$$\text{SE diff} = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}$$

The computation is straightforward.

Square the standard deviation of sample 1 and divide by the number of observations in the sample	$\frac{SD_1^2}{n_1}$ (1)
Square the standard deviation of sample 2 and divide by the number of observations in the sample	$\frac{SD_2^2}{n_2}$ (2)
Add (1) and (2)	$\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}$
Take the square root	$\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}$

This is the standard error of the difference between the two means.

An example of its calculation with Dr White's figures will be given shortly, but first a note on the so-called null hypothesis is needed.

British Medical Journal
T D V SWINSCOW, MSC, MB, deputy editor

Null hypothesis

In comparing the mean blood pressures of the printers and the farmers we are testing the hypothesis that the two samples came from the same population of blood pressures. The hypothesis that there is no difference between the population from which the printers' blood pressures were drawn and the population from which the farmers' blood pressures were drawn is called the null hypothesis.

But what do we mean by "no difference"? Chance alone will almost certainly ensure that there is some difference between the *sample* means, for they are most unlikely to be identical. Consequently we set limits within which we shall regard the samples as not having any significant difference. If we set the limits at twice the standard error of the difference, and regard a mean outside this range as coming from another population, we shall on average be wrong about once in 20 times if the null hypothesis is in fact true. For we know that, when data are normally distributed, about 5% in a single population will by chance alone be outside the range of two standard deviations from the mean. Likewise if we allow a difference of three times the standard error of the difference, and regard a mean outside this range as coming from another population, we shall on average be wrong once in 370 times.

A range of two standard deviations or standard errors is often taken as implying "no difference." But there is nothing to stop an investigator choosing a range of three standard deviations (or more) if he wants to reduce his chances of rejecting the null hypothesis on the basis of an aberrant observation.

A point to note here is that we try to show that a null hypothesis is *unlikely*, not its converse, that it is likely. So a difference which is greater than the limits we have set, and which we therefore regard as "significant," makes the null hypothesis *unlikely*. A difference within the limits we have set, and which we therefore regard as "non-significant," does not make the hypothesis likely.

Comparison of two means

Dr White wants to compare the mean of the printers' blood pressures with the mean of the farmers' blood pressures. Therefore she erects the null hypothesis that there is no significant difference between them. The figures are set out first as in table 8.1 (which repeats table 6.1).

TABLE 8.1—*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

Analysing these figures in accordance with the formula given above, we have:

$$\text{SE diff} = \sqrt{\frac{4\cdot5^2}{72} + \frac{4\cdot2^2}{48}} = 0\cdot81 \text{ mm Hg.}$$

The difference between the means is 88 − 79 = 9 mm Hg. We now find how many multiples of its standard error this difference represents: 9 ÷ 0·81 = 11·1. Reference to table 7.1 shows that this is far beyond the figure of 3·291 standard deviations representing a probability of 0·001 (or 1 in a thousand). The possibility of a difference of 11·1 standard errors occurring by chance is therefore exceedingly low, and correspondingly the null hypothesis that these two samples came from the same population of observations is exceedingly unlikely. The probability may be written $P \ll 0\cdot001$.

Sometimes a mean may be known from a very large number of observations and the investigator wants to compare the mean of his sample with it. We may not know the standard deviation of the large number of observations or the standard error of their mean. But this need not hinder the comparison if we can assume that the standard error of the mean of the large number of observations is near 0 or at least very small in relation to the standard error of the mean of the small sample.

This is because the formula for calculating the standard error of the difference between the two means—

$\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}$ has n_1 so large that $\frac{SD_1^2}{n_1}$ becomes so small as to be negligible. The formula thus reduces to $\sqrt{\frac{SD_2^2}{n_2}}$, which is the same as that for the standard error of the mean (Part VI), namely $\frac{SD_2}{\sqrt{n_2}}$.

Consequently we find the standard error of the mean of the sample and divide it into the difference between the means.

For example, a large number of observations has shown that the mean count of erythrocytes in men is $5\cdot5 \times 10^{12}/l$. In a sample of 100 men a mean count of 5·35 was found with standard deviation 1·1. The standard error of this mean is $SD/\sqrt{n}$, so that $1\cdot1/\sqrt{100} = 0\cdot11$. The difference between the two means is $5\cdot5 - 5\cdot35 = 0\cdot15$. This difference divided by the standard error is $0\cdot15/0\cdot11 = 1\cdot36$. This figure is well below the 5% level of 1·96 and in fact is below the 10% level of 1·645 (see table 7.1). Consequently we conclude that the difference is of no statistical significance.

Exercise 8. In one group of 62 patients with iron-deficiency anaemia the haemoglobin level was 12·2 g/dl, standard deviation 1·8 g/dl; and in another group of 35 patients it was 10·9 g/dl, standard deviation 2·1 g/dl. What is the standard error of the difference between the two means, and what is the significance of the difference? *Answer:* 0·42 g/dl, $0\cdot01 > P > 0\cdot001$.

If the mean haemoglobin level in the general population is taken as 14·4 g/dl, what is the standard error of the difference between the mean of the first sample and the population mean and what is the significance of the difference? *Answer:* 0·23 g/dl, $P < 0\cdot001$.

Southampton: the first years

III—Epidemiology, psychology, and sociology

W E WATERS, D MARCER, E TOPLISS

British Medical Journal, 1976, **2**, 95-97

The teaching of psychology and sociology, which continues until the end of the fifth term, follows immediately after an introductory course in the first term which has the descriptive title, "Man, Medicine, and Society." Each course lasts about 50 hours. The course in epidemiology and medical statistics is of similar length, but is more heavily concentrated in the second year. The three subjects are examined jointly at the end of the fifth term by two written examinations. Assessment also includes a contribution based on course work, usually essays and reports of projects. The three courses are not presented as an integrated whole. Instead we have tried to make links with disciplines outside the social sciences, particularly those with a clinical content. For example, 11 hours of epidemiology and medical statistics are integrated into the systems courses of the first and second years. Thus during the cardiovascular system course there is an afternoon symposium on ischaemic heart disease, during which the epidemiology of the condition is presented along with the pathology, physiology, and clinical features. There are similar sessions on chronic respiratory disease and also joint lectures on other conditions. In this way the subjects are seen to be relevant to the work of the medical profession, rather than as an isolated set of esoteric academic disciplines.

Southampton Medical School, Southampton
W E WATERS, FFCM, DIH, professor of community medicine

University of Southampton, Southampton
D MARCER, BSC, PHD, lecturer in psychology
E TOPLISS, BSC, senior lecturer in sociology

Man, medicine, and society

During the first six weeks of the first year students have an introductory course on man, medicine, and society. This consists of 12 two-hour sessions and introduces the new students to epidemiology and the social sciences. It shows their relevance to medicine and the importance of populations and groups as units of study in medicine and attempts to put into perspective some of the many factors that influence health, including occupational, environmental, and genetic factors. An integral part of the course is a guided walk in two contrasting areas of the city of Southampton.

Topics, selected for their multidisciplinary appeal, are used for problem-orientated seminars. Students work on their own and meet for three sessions with their tutor, who may be either an epidemiologist, a psychologist, or a sociologist. These seminars consider such problems as "a road traffic accident" or "non-accidental injury in children."