

## COMBINING INDEPENDENT TESTS OF SIGNIFICANCE

I have made several tests of significance on independent data. The tests I have made concern the difference between mean scores under two different treatments. Retaining the same order of differencing in the numerator of my  $t$ -test, I get  $t$  values of  $-.68$ ,  $+1.53$ ,  $+2.21$ ,  $+1.85$  with 6, 18, 22, and 25 degrees of freedom respectively. I would like to make a combined test of significance to test whether the difference between the means is positive. I do not want to try to form a single value of  $t$  from the original raw

**Answer.** If the test based on the product of probabilities from different trials is adopted, allowance can be made for the occasional occurrence of negative values of  $t$  by calculating the corresponding probability for a single tail. Negative values of  $t$  will then supply probabilities greater than .5, which do not greatly reduce the product and, since they introduce two additional degrees of freedom, will tend to lower its significance.

Thus with the illustrative values given in the problem, I obtain the following table:

$n$	$t$	$P$ (single tail)	$\log_e 1/P$
6	$-.68$	.739	.3025
18	1.53	.0717	2.6353
22	2.21	.01932	3.9466
25	1.85	.03809	3.2698
			<hr/> 10.1542

Doubling the total of the natural logarithms we have chi-square equal to 20.31 for 8 degrees of freedom, corresponding to a probability rather less than 1%. If, as seems here to be the case, consistent deviations in either direction would be deemed relevant, we may say that the probability is less than 2% for the

data for various experimental reasons. In *Statistical Methods for Research Workers* R. A. Fisher presents a method of combining tests of significance, but it seems not to apply to the present situation since one of my differences is in the wrong direction. Also I fail to see how Fisher's method takes account of the varying degrees of freedom in the four sets of data. Can you tell me an appropriate procedure for combining these data? If you have any other advice about combining independent tests of significance, I would be glad to get it.

observation of deviations so large and so consistent in direction as those observed.

The differences in the degrees of freedom of the several trials, and also differences in their respective precisions, are, of course, reflected in the values of  $P$  to be multiplied together. The inclusion of the first trial lowers the significance of the test, and it may well be that this first trial was on so small a scale that it was likely *a priori* to do so. It would have been open to the experimenter to decide in advance not to conduct any trials with less than 15 degrees of freedom, or, equally legitimately, not to include such trials in his test. Such a decision must, however, be taken in advance and not after inspection of the result of the trials. Generally, therefore, if combination by the product method is intended, it will be worth while to aim at a succession of trials of approximately equal sensitiveness. More usually, however, the test will be employed to give provisional guidance based on all the evidence to hand, and this will often be of very variable precision.

It may be noticed that the same body of data may give a significant combined test in whichever direction the effect is tested. This would, of course, indicate reality of departure from the null hypothesis, but in discrepant directions in different cases, so that the results of the several trials would be heterogeneous.

R. A. Fisher