In Defence of Inference

H. Goldstein, University of London Institute of Education

Summary
It is argued that the process of inference from a set of sample data is an important part of educational research. While there may be some abuse of statistical significance testing, it does not follow that educational research should become concerned mainly with so called 'descriptive' statistics.

In a short review entitled 'The Criticism of Inferential Statistics' Derrick (1976) argues for a greater use of 'descriptive' statistics and a corresponding downgrading of 'inferential' statistics in education. In an even shorter note I would like to enter a general defence of inference, although I do agree with some of Derrick's detailed points.

It appears that Derrick's main objection is to the over-indulgent use of probability levels resulting from tests of significance. He also seems to be complaining about generalizations from single research studies to the wider area of, for example, educational policy. This latter point, however, is concerned with scientific rather than statistical inference, and although an interesting topic, it is not especially relevant to his main argument and I will ignore it.

Many statisticians are concerned about the use of significance tests as a seal of scientific approval, and many of us involved in refereeing journal articles have long despaired over this. To say that a technique has been abused, however, is not to say that it is of little use if properly applied. In the social sciences we nearly always make some kind of inference from sample data, since we generally wish to make statements about values in the eventual aim is to provide 'good' estimates of population parameters, such as the difference between two mean values or a regression coefficient. In the process of doing this we might carry out a significance test in order to establish, say, that the population difference between two means was not consistent with being zero. We would also usually go on to indicate the accuracy of the estimate by calculating, say, a confidence interval. When we come to judge the real life significance of such an estimated difference, then its actual size will be the focus of interest, and this might be evaluated in comparison with other differences etc.

Derrick's use of the word 'descriptive' appears to refer to 'population' or 'official' statistics. Such statistics, however, can often be used inferentially. Furthermore, he must be aware of the limitations of many such statistics, including those concerned with education, which are often collected solely for administrative purposes and which rarely can be a substitute for properly conducted research studies.

While agreeing that all is not well with the use of statistics in research, I do not believe that the situation is helped by an indiscriminate condemnation of 'inferential statistics'. The problem is a complex one and can only properly be dealt with by a careful process of statistical education, both of the present and the future research community.

Reference

DERRICK, T. (1976). 'The criticism of inferential statistics', *Educ. Res.*, **19**, 1, 35–40.