LETTERS, NOTES AND NEWS

Dear Sir

I wish to comment on the review of the book 'Statistics for Ornitholgists' (*Safring News* 16 [2] 1987). Not having seen the book, I am in no position to assess the validity of the reviewer's comments on the work in question. I do, however, wish to discuss the following misleading sentence: "When most researchers, ornithologists included, go into statistician mode, the starting point is the data set to hand and they are looking for appropriate statistical methods". If this is a reflection of much of the research in this country, then I am disheartened.

Any research, particularly if it goes under the name and, therefore, the philosophy of science, starts with a purpose, not a data set. The purpose dictates the objectives, the key questions addressed, the sampling strategy and the means of data analysis. The use of hypothesis testing and inferential statistical techniques is intended before data collection, and the sampling strategy is designed with this intention. This is true whether the data are to be collected afresh from the field, or from existing records. *Ad hoc* application of most of the statistical procedures which are apparently covered in this book to *ad hoc* data sets is likely to produce spurious conclusions. This practice should be discouraged, for far-reaching management decisions are all too frequently based upon such conclusions.

Ad hoc data sets are important, and may supply valuable information. I have no problem with those who wish to collate such data, for that is their purpose. For such a purpose, frequency distributions, means, modes and so on, are merely intuitive ways of presenting data. Anybody is free to make of these presentations what they will. It is the ignorant manipulation of data sets with more powerful statistical techniques, without a purpose and an understanding of the philosophy behind the approach, which is to be avoided.

Yours sincerely

A. C. Carter, Department of Botany, University of the Witwatersrand, P. O. WITS, 2050.

(See overleaf for Professor Underhill's reply)

Reply by Professor L. G. Underhill

I wish that all reseachers had collected their data knowing precisely what statistical procedures they were going to use afterwards. However, for most of the data that come my way (both from local and overseas researchers), detailed decisions about how the data will be analyzed are made after the data are collected. I don't believe that Dr. Carter need feel too disheartened about this. Most of the researchers I work with know the questions they want to ask, they know what data they need to collect to answer these questions, and they are reasonably certain that they will be able to analyse the data collected, even if they are not sure of precisely which statistical methods are most appropriate. And very often, data collected in this way, stand up to much more sophisticated analyses than the researcher had in mind.

I agree with Dr. Carter that this is not the ideal way to conduct research, and that lack of design in research is deplorable. In fact, the greatest contribution a statistician can make to any research project is ensuring beforehand that the right data are collected and in the right volume. However, in many of the observational disciplines, experimental design (in the strict statistical sense of the term) is notoriously difficult. For example, it is ludicrous to tell an ornitholgist that to investigate some hypothesis, he must trap 100 birds - 50 males and 50 females - each month, for three years. For many researchers, observational data sets (a term I prefer to the rather disdainful *ad hoc* data sets) are the norm. Also, even the best experimental designs get wrecked by missing data (for example, when a data-logger breaks down), by variables that behave unexpectedly, by assistants who misinterpret the protocol, and by many other unforeseen problems. Continuing the metaphor, the statistician then takes on the role of salvor.

Finally, let me add that the sentence from my review has been quoted somewhat out of context: the point I was making, and it remains valid whether you do your statistical thinking before or after you collect your data, is that 'Statistics for Ornithologists' would be more useful if it were arranged by data types rather than by methods. I apologise if I have mislead anyone into believing that statistics only plays its role after data collection.

L. G. Underhill, Department of Mathematical Statistics, University of Cape Town, RONDEBOSCH, 7700.