BOOK REVIEW

STATISTICS FOR GEOSCIENTISTS: TECHNIQUES AND APPLICATIONS

by Saroj K. Pal, Concept Publishing Company, New Delhi, 610p.

Statistics is a vital branch of, to use a cliche, applied mathematics. Therefore, books on this topic are welcome. The author Saroj K. Pal is a Professor of physical geography. This book is a revised and enlarged version of his earlier book entitled *Statistical techniques: a basic approach to geography*. The enlargement of his perspective and interests is reflected in the new title. Saroj Pal says that he is not a statistician, but as statistical techniques are not a prerogative of anybody, he has written these books. That bold attitude is very welcome.

Statistics is a disadvantaged discipline. Under the influence of Laplacian determinism, which has outstretched its life, many scientists are averse to entertain chance and probability. There is also a popular perception of statistics being worse than not just lies, but even damned lies. Therefore, a new book on statistics, particularly one motivated and illustrated with applications, is very welcome. This welcome anticipates that the author will do justice to this disadvantaged field, and help make statistical techniques accessible, enhance their comprehension, and enlarge their applicability.

The opening chapter discusses role of statistics in development of models in geosciences. Model-building is portrayed as a recursive engagement in which statistical criticism can lead to rejection of the earlier model. After that, the concepts of statistical data analysis are introduced. More specifically, measures of central tendency and dispersion are introduced over two chapters. These are not merely the starting points of statistical analysis. T.S. Eliot would suggest that progress consists in coming back to the starting point again and again. Little more than three decades ago the entire statistical thinking has witnessed such a home-coming in the form of *robust statistics*, whose concern has been to critically reassess the by then classical measures of central tendency and dispersion, and propose better ones. *This entire revolution has not touched this book.*

The notions of Lorenz curve and Gini coefficient have been introduced and their application to quantification of inequality has been suggested. The definition of Gini coefficient can be put in many equivalent forms, and each of them gives a different insight into its meaning and the possibility of generalization. Including such a discussion would have been welcome.

Three classical distributions, viz. the normal, the binomial and the Poisson have been dealt with in separate chapters. The related distributions such as the negative binomial, hypergeometric, gamma, etc. have also been introduced. Then there is a chapter on extreme value distributions, but the underlying notion of order statistics does not find a place.

Geoscience has a special problem of having to deal with spatial data and statistics. A chapter discusses the related issues. The treatment of nearest neighbour analysis is particularly good. Sampling techniques are discussed in a separate chapter. Then a note says that the subsequent statistical tests discussed in the book assume that the simple random sampling has been adopted. The methodology required for other sampling plans such as the stratified and cluster sampling is regarded as beyond the scope of the book. It would have been better if specific references where such methodologies are discussed were given. The select bibliography at the end of the book does not serve the role of specific indicators. Moreover, the author is not averse to giving references within the text. For example, on p. 486, full references to W.S.Robinson (1950) and H.R. Alker (1969) are given as running text.

Parametric statistical methods are discussed over three chapters. One of them deals with large-sized samples, the next with the small-sized samples, and the last one with the analysis of variance. While dealing with Student's t test, the author is tempted to tell us who this by now immortal Student was. William S. Gosett was working with Guinness Breweries. The employer did not allow him to publish his work, hence the *nom de plume*. Here is a good hint for many, rather than let the BP rise. There could have been more such anecdotal information.

Next seven chapters deal with nonparametric statistics, individual chapters dealing with chi-squared test, Kolmogorov-Smirnov test, Mann-Whitney test, Wilcoxon ranked pair test, Kruksal-Wallis test, Spearman's rank correlation and Kendall's rank correlation. The distribution of chapters between parametric and nonparametric indicates author's interest in and preference for nonparametric methods. Parametric statistics makes assumptions about the underlying density functions. Most frequently, it is the normal assumption. Unwarranted specificity of parametric statistics, as it has developed, is really problematic. But then, one does not have to fly to the other extreme of nonparametric statistics, which shuns assumptions about the density functions. Now there is a lot of middle ground under robust statistics, which makes nonspecific and elastic assumptions about the underlying density functions. The book should have covered some of this middle path.

The last five chapters deal with regression analysis. Product-moment correlation, linear regression, spatial correlation, linear multiple regression and higher order methods get a chapter each of discussion. There are statistical tables at the end. A book of statistics has to have them. But there should also have been specific indicators to find more comprehensive tables.

The concept of this book is in line with many books written by Indian authors for Indian students. As many concepts as possible are illustrated with numerical examples. Almost all the intermediate steps of arithmetic, including substitution of values of the variables, opening the brackets, results of multiplication and division, etc are explicitly and elaborately shown. There are exercises at the end of every chapter, and worked answers at the end. But most exercises require direct application of formulae and procedures, and not the extension of the concepts introduced or the derivation of new formulae. Formulae are listed at the end of every chapter. There is no doubt that most students would find these features useful. Collection of useful formulae could be left as a good exercise for a student. Explicit arithmetic details prompt a student to learn by arithmetic mimicry, rather than by understanding basic concepts. Given a set of formulae, it may be possible to substitute some of them in others, to get a simpler problem-specific formula, before numerical substitutions are made. Answers at the end prompt a student to look at them for clues whenever he is stuck, and the clue he gets may be fortuitous.

These points become more serious when a book lacks linguistic precision, notational clarity, has typographical errors generously scattered, and has many nontypographical and graphical errors too. Unfortunately, this book has all these features. Here is a sample:

The name Mahalanobis is misspelt on p. 305. Eq. (20.3) on p.438 is wrong. It should have been $(\Sigma Y)^2$ in the denominator. Eq. (6.8) on p. 179 is wrong. The correct version appears on p. 185. Eq. (23.8) in the matrix form on p.506 is wrong. It does not correspond to the scalar Eqs. (23.5) to (23.7). The error has apparently arisen because a similar block was copied from elsewhere and subsequently all the requisite changes necessitated by a different notation were not entered. This is a case of a DTP typo! The last line of the first table on p.508 has undue summation signs.

In the discussion on the geometric mean (p. 52), it is not stated at all that the concept is applicable only when all the quantities averaged are positive (not just nonnegative). Similarly, it is not mentioned (pp. 100-101) that the coefficient of variation is used only for positive-valued

variables. The definition of median on p. 57 (that it is (N+1)/2) is wrong. It would mean that as long as the size of the sample is known, the sample need not even be available to calculate the median. This error could be compared with discussion on p. 536, where the author suggests different transformations for odd and even number of observations, rather than for odd- and even-numbered observations. The problem is that of linguistic precision. On p. 598, where the author is being helpful to explain the notions of mantissa and characteristic of Log, there are statements such as *6254 is Log 3.7959*, instead of *Log 6254 is 3.7959*. Indeed, God should only help those who help themselves. Some of the Greek alphabets on p. 602 are wrongly introduced. On p. 598, there is an unwarranted statement that the number *2.71828..... is designated as e*. There is a similar statement or at least an implication about π. Students frequently believe that π is 22/7. They should be helped out of such misunderstandings.

Fig. (2.2) on p. 61 is wrong. The plot on the right can not have the mean as shown. A similar error is repeated in Fig.(4.2 d) on p. 123. A student will be misguided rather than helped by such plots. Fig. (3.3) on p. 106 shows three normal curves, but areas under them are not equal. If a student is to understand something more than a *bell-shaped curve* by the phrase *normal curve*, he should be brought up on better curves. Fig. (10.1) on p. 263 has a caption *One-tailed and Two-tailed Tests*, but the two parts of the figure are in reverse sequence. In the figure on the left, one more portion should have been hashed. In Fig. (24.1) on p. 527, the plots, the caption and the legends do not go well with each other. Fig. (11.1) on p. 291 is wrong. It shows standard normal curve and Student's density on the same plot. Both of them are even functions. Therefore, their crossovers also must be symmetrical. In the figure, there is only one crossover on the left, the corresponding one on the right is absent.

The use of statistics is recommended in areas where the observations are in error, scant or even missing. It can deal with uncertainties and internal inconsistencies, and yet help draw the best conclusions under the circumstances. Here is a book on statistics, which can be used by those trained to face such situations stoically.

National Geophysical Research Institute,
Hyderabad

P.S. MOHARIR

Obituary

K.V. Venkatesh, Fellow of the Geological Society of India died of heart failure on 3-10-1998 at Bangalore. He was holding the office of Director in Geological Survey of India and was attached to the Operation West Coast - I at Mangalore. He was a pioneer in the field of Marine Geological Studies in India.