GEOSTATISTICS

a colloquium

Proceedings of a Colloquium on Geostatistics held on campus at The University of Kansas, Lawrence on 7-9 June 1970. Sponsored by the Kansas Geological Survey, International Association for Mathematical Geology, and University of Kansas Extension.

edited by Daniel F. Merriam

Chief of Geologic Research
Kansas Geological Survey, The University of Kansas

PLENUM PRESS • NEW YORK-LONDON • 1970

Library of Congress Catalog Card Number 71-142040
SBN 306-30519-4

A Division of Plenum Publishing Corporation
227 West 17th Street, New York, N. Y. 10011

United Kingdom edition published by Plenum Press, London
A Division of Plenum Publishing Company, Ltd.
Donington House, 30 Norfolk Street, London W.C. 2, England

Printed in the United States of America

computer applications in the earth sciences

GEOSTATISTICS

a colloquium

COMPUTER APPLICATIONS IN THE EARTH SCIENCES

A series edited by Daniel F. Merriam

1969—Computer Applications in the Earth Sciences

1970—Geostatistics

To all geostatisticians and statistical geologists

with permission of © King Features Syndicate, Inc.
and courtesy of Bud Blake

PREFACE

Although statistics have been used by geologists for many years, only recently has the subject received the attention needed and deserved. Geologists and other earth scientists have a use for summary statistics of large data bases, knowledge of frequency distributions, understanding of sampling designs and problems, and application of stochastic models, but in general they are unaware of the many aspects of help available through the statistician. It seemed warranted at this time to get the two disciplines together and to find a common meeting ground for further collaboration.

Thus the subject of the 8th Colloquium was proposed as GEOSTATISTICS. Statisticians with interests in applications in the earth sciences were asked to participate with earth scientists interested in applying statistics to problems. This volume records the proceedings of the meeting.

The Kansas Geological Survey, the International Association for Mathematical Geology and the University Extension were hosts to 120 participants on campus at The University of Kansas during 7-9 June 1970. The Colloquium was the 8th in a series on "Computer Applications in the Earth Sciences." Previous subjects were classification, trend analysis, time-series analysis, simulation, sampling, computer applications, and optical-data processing. The stated purpose of the meeting was to explore some assumptions, limitations, and applications for statistical geology and geostatistics.

Participants in the Colloquium were from many parts of the world. Countries represented were Canada, Czechoslovakia, France, Great Britain, India, Italy, Japan, The Netherlands, South Africa, United States (20 states), and West Germany. Approximately 44 percent were from academia, 40 percent from industry, and 16 percent from governmental agencies.

Many people helped with preparations for the meeting. Mr. R. F. Treece of University Extension again made all the logistical arrangements. Drs. J. C. Davis and C. D. Conley, of the Geological Survey, helped with technical arrangements. Chancellor E. L. Chalmers, Jr., Vice Chancellor F. H. Heller, and Associate Dean of Faculties W. J. Argersinger of the University and Drs. F. C. Foley and W. W. Hambleton of the Geological Survey welcomed the group. Technical assistance with the many problems was ably given by Messrs. N. Minoura and G. S. Srivastava of the Geological Survey. Typing of the program and proceedings was by Mrs. Jo Anne Crossfield and Miss Cheryl Morgison of the Geological Survey. Drafting assistance was provided by Miss S. K. Hagen, Mr. Charles Barksdale, Mrs. Renate Hensiek, and Mrs. K. S. Mumford of the Geological Survey. Mrs. Kathy Remark and Mrs. Joan Combes of the Geological Survey helped in many ways preparing material and assisting with the arrangements. Dr. A. B. Vistelius, President of the International Association for Mathematical Geology was represented by Vice Presidents Prof. W. C. Krumbein and Prof. G. S. Watson. Others taking part in the meeting were Dr. V. Nemec, Treasurer, and Council Members Dr. F. P. Agterberg, Prof. G. Matheron, and Prof. E. H. T. Whitten, and Editor-in-Chief Dr. D. F. Merriam.

It is hoped that this volume will serve as an introduction to those interested in the subject and as a catalyst to those working in this area. It was obvious to those attending the oral presentations that statisticians and earth scientists have much to offer each other; a theme of cooperation also appears in the written presentations. To all geostatisticians and statistical geologists may these proceedings serve as a basis for inspiration and innovation!

Fontainebleau, France
September 1970

Daniel F. Merriam

INTRODUCTION

It occurs to me that Dan Merriam really needs to introduce some variety into his programming: in looking back over the years I find that I have had the pleasure of welcoming no less than 6 of these 8 colloquia. This means that there is a number of people whom I have seen before. In fact, one can assume that the occasion for my being here is to have a reunion with Bill Krumbein.

It occurs to me also that just a week ago I found myself talking to a large assemblage of our graduating seniors and their parents when it was my pleasure to single out 6 members of our faculty who had received awards for distinguished teaching. In summarizing their achievements I noted that they represented in their preparations contributions of the taxpayers of the states of Washington, Wisconsin, Iowa, Indiana, and New Jersey and the foreign countries of Germany, Spain, and Yugoslavia. As I look at the program of this 8th Colloquium, I find that one could easily do the same sort of thing, and point out that this is indeed an international spread of talents assembled here. This, of course, is not a new development; it has been true of previous colloquia. But it is indicative, I would suggest, of the role which these colloquia have come to play.

There are a number of things that can happen whenever people begin to develop ideas and notions by which scholarly disciplines can be moved forward. But, both here and abroad, when such movements gain real momentum, institutional resistance develops.

There is, for better or worse, a departmental structure which makes it a little difficult, in most universities, for new ideas to take root. I see a great many people who come to the university interested in associating themselves with us and whenever we talk about the fact that a sizeable number of members of this university's faculty are associated not with one department but with two or three or sometimes four of them, the eyebrows go up and they say, "But that's an impossible situation."

They say this because they know people to whom this happened and who had found themselves at a disadvantage because no one department really concerned itself with them. I find myself pointing to the fact that, for one reason or another, the pattern at this university has worked exactly the opposite way. What has happened here is a continuing emphasis upon the interdisciplinary contact, the interaction between different disciplines, and a sense of openness that enables a member of the faculty or staff to affiliate himself and to associate himself almost as he pleases. In fact, I was slightly stunned when it became necessary for me last year to ask for a complete list of people who were carried on more than one budget, and this list actually included 28 percent of the faculty.

It is in this type of environment that an operation such as the Kansas Geological Survey can develop as it has developed. I am sure that all of you are aware of the fact that there are other geological surveys around the country that do not offer or concern themselves with the same type of things as this Survey. I think the reason for this is that it happens to find itself in an intellectual and academic setting in which its type of interactions are not only being encouraged but supported and enhanced.

But I did not come here to talk about The University of Kansas, although I am here to extend to you the official welcome of the University. I came here in order to talk briefly about the implications of the type of efforts in which you ladies and gentlemen are involved.

A few years ago Dan Merriam came to my office and talked about the ideas and aspirations of his research group. What he thought was needed was a new institute, particularly because this would enable a name to be attached to the type of things that he--and most of you--are interested in developing. The name that he and his associates had come up with at that time was "geometrics." He was proud of that name, although he was concerned also that perhaps the mathematicians would feel that he was encroaching upon their territory. He was even more concerned when I reached behind my desk and pulled out a symposium issue of a legal journal published at Duke University with the intriguing title of "jurimetrics." When he discovered that the lawyers had resorted to the same kind of word-snitching that he had indulged in, the geometrics notion lost some of its appeal. I relate this little tale not to show that lawyers are also in the game but to illustrate the need people have to see their efforts and activities clearly identified.

This matter of identification is, of course, one that dominates much of the American academic profession. It is even more dominant in places where the professor is as sacrosanct as he is in the university systems in Europe, particularly in Germany and France. And yet if we are to look forward to and talk about the development of our knowledge and our sense of perception, I suggest that we are going to have to be more enterprising and think in terms that do not necessarily fit departmental structures. The direction of the future, for the American university at least, rests in the charting of courses that take us, if not through the departmental structure, at least around it.

At this university we are not afraid to have a research unit in the Geological Survey. We are assuming that it will discover to whom it should relate, to whom it ought to talk, both as individuals and as a group, and if in the process people have their small jealousies, I for one, am not going to be terribly concerned. If the jealousies turn to intellectual and academic issues, that is a lot better than if they turn on personality conflicts or on such major issues as who happens to have a parking sticker closer to the building.

But all of you have a lot of work to do and it behooves me to get out of the way so that you may get started. Whether you are in academia, in industry, or in a research organization, I am confident that in the aura of interdisciplinary effort which The University of Kansas has sought to nourish, you will see your own contributions maximized through and in this Colloquium on Geostatistics.

University of Kansas F. H. Heller
June 1970

LIST OF PARTICIPANTS

F. P. Agterberg, Geological Survey of Canada

F. Chayes, Carnegie Institution of Washington

W. R. James, Coastal Engineering Research Center

G. S. Koch, Jr., Artronic Information Systems, Inc.

W. C. Krumbein, Northwestern University

R. F. Link, Artronic Information Systems, Inc.

R. B. McCammon, University of Illinois at Chicago Circle

A. Marechal, Centre de Morphologie Mathematique (France)

G. Matheron, Centre de Morphologie Mathematique (France)

V. Nemec, Geoindustria (Czechoslovakia)

J. Serra, Centre de Morphologie Mathematique (France)

P. Switzer, Stanford University

J. W. Tukey, Princeton University

G. S. Watson, Johns Hopkins University

CONTENTS

THE STATISTICAL TREATMENT OF ORIENTATION DATA

Geoffrey S. Watson

The Johns Hopkins University

ABSTRACT

A short sketch will be given of the types of orientation data that have been described statistically. Some specific problems will be mentioned. An effort will be made to elicit new situations requiring statistical treatment. The discussion of past work indicates what can be done when a problem is formulated.

INTRODUCTION

The writer has given several surveys of orientation statistics with special reference to the earth sciences (Watson, 1966, 1970). The second paper is more comprehensive and in particular, has a complete bibliography. At a more elementary level, and especially for two-dimensional problems, the monograph by Batschelet (1965) will be found useful.

The word "orientation" is used here broadly. In geometric terms, one may be interested in the orientation of directed lines, undirected lines or axes, or a solid body. Most existing literature deals with the first two cases; in the last case, the orientation may be specified by a set of angles or rotation matrix (i.e., orthogonal matrix). In each case, of course, there is some basic set of coordinate axes with respect to which the object is oriented. The orientation of three lines fixed in a solid body defines its orientation. Thomas D. Downs (personal communication) has given an interesting treatment of a bundle of lines.

The simplest problems deal with the description of, and comparison of, populations of orientations. For the most part we assume our data are a random sample from a population. It then may be possible to answer questions such as "are the two populations the same?" with no further assumptions. Such methods are called distribution-free (or more usually and more confusingly as non-parametric). In other situations the data may be summarized, without loss of information and for more specific comparisons, by fitting a plausible mathematical formula for the distribution of the population. Statistical methods that work reasonably well when the assumed formula is not exactly correct are called robust. Because such methods are clearly desirable, much research effort is expended in either deriving robust methods, or testing (usually by simulation) whether or not a given method is robust.

The approach outlined in the last paragraph is that of classical statistics, and is not special for orientation data. Therefore, it is not surprising that almost all the literature has been written in this spirit. Some details in the next section give the flavor. It is both useful and of theoretical interest. For although much theory has the format "(i) consider random sample $x_1, \ldots, x_n$ where the x_i's are points in some abstract set X on which a probability density is defined, $f(x,\theta)$ say, where θ belongs to some other set Θ, (ii) test the null hypothesis $\theta = \theta_o$," the sets X and Θ are usually Euclidean.

There are, however, other directions and styles of statistical inquiries. The mathematical formality of classical statistics is partly a legacy of the days of small data collections and of low computing power. Less data - more assumptions is a sensible doctrine, and simple models usually mean simple calculations. To use J. W. Tukey's expressions, classical statistical theory is more concerned with confirmatory data analysis than with exploratory data analysis. The computer can arrange, transform and display data with speed and ease. With small bodies of simple data, graphical techniques are invaluable in suggesting regularities and peculiarities. Large and complex data collections now are common and demand new and automated exploratory devices.

Whatever the approach, there must be new and different problems awaiting either treatment or application. In the section on new possibilities, some notions are proposed for discussion.

EXAMPLES OF STATISTICAL METHODS FOR ORIENTATIONS

Let us consider a direction in three dimensions. It is equivalent to a vector r of unit length or to a point on a sphere of unit radius. Thus data will be shown by projection, for example, on a Schmidt net. The points mapping a homogeneous sample may (1) form a roughly circular cluster with density falling off as one goes away from its center, (2) concentrate in two circular clusters at either end of a diameter, (3) fall on a great circle, or (4) make some more complex pattern.

In case (1), the direction of the vector resultant Σr_i will be a sensible measure of the "mean" direction, whatever the configuration of the cluster. It is not impossible therefore that methods using vector means will be robust--it has been shown that they are (Watson, 1967). If the points are highly clustered, we would expect that the estimate of the mean direction is accurate, i.e., has a small "dispersion" ("variance" is inappropriate because it is defined only for real valued random variables). As a measure of dispersion of our N observations it is natural to consider N - R when R is the length of the sample resultant Σr_i. If the cluster is circular, one such number should be enough. The average dispersion is thus (N - R)/N so the "accuracy" of the sample might be taken as the inverse of the dispersion, that is N/(N - R). Its mean direction is the unit vector $\Sigma r_i/R$. Most readers will be familiar with the "contouring" methods that result in what corresponds to the histogram of linear statistics. Thus two samples of this type could be contrasted by comparing their mean directions and their accuracies. The usual method to proceed now is to introduce Fisher's distribution

$$f(r;K,\mu) = \frac{K}{4K \sinh K} \exp(Kr.\mu), \qquad \text{(Kappa)}$$

where K is clearly an accuracy constant and μ the mean direction. It can be shown that the maximum likelihood estimators are

$$\hat{\mu} = \Sigma r_i / R, \qquad \text{and}$$

$$\hat{K} \equiv k = (N-1)/(N-R).$$

If μ were known and only K required estimation, we find

$$\hat{K} = N/(N-X).$$

Thus N - X = dispersion of sample about the true mean, and N - R = dispersion of sample about the sample mean. We need an analog of the algebraic identity, so useful for the normal distribution,

$$\Sigma(x-\mu)^2 = \Sigma(x-\bar{x})^2 + n(\bar{x}-\mu)^2.$$

Because this has the interpretation

(Dispersion of sample about the true mean) = (Dispersion of sample about sample mean) + (Dispersion of sample mean about true mean)

and required identity is

$$N - X = (N - R) + (R - X).$$

It may be shown that with this understanding many problems may be solved by analogy with normal theory. The methods are widely used in rock magnetism.

If one does not wish to assume that the Fisher distribution describes the population, methods may be obtained either by assuming the samples to be large or by finding nonparametric procedures. The former method recognizes that Σr_i has three components Σx_i, Σy_i, Σz_i, and that these become jointly normal as N increases. Thus one is returned to trivariate normal methodology. In particular, the classic test for a purely random, or uniform, distribution might be mentioned. In this case Σx_i, Σy_i, Σz_i are shown to have zero means, to have variance 1/3 and to be uncorrelated. Thus

$$\frac{(\Sigma x_i)^2 + (\Sigma y_i)^2 + (\Sigma z_i)^2}{1/3} \approx \chi_3^2$$

or

$$R^2 \approx \frac{\chi_3^2}{3}.$$