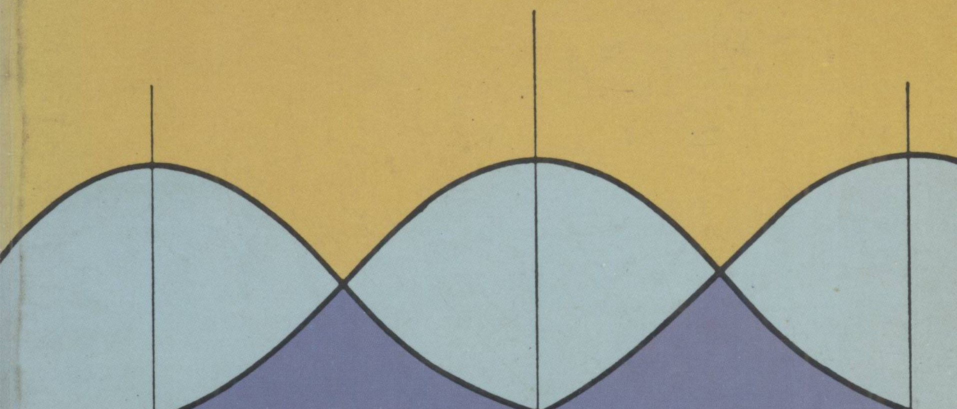
STATISTICAL ANALYSIS OF GEOLOGICAL DATA



George S. Koch, Jr. and Richard F. Link

Two Volumes Bound as One

P5-32
K76

STATISTICAL ANALYSIS OF GEOLOGICAL DATA

George S. Koch, Jr.
and
Richard F. Link

Two Volumes Bound as One
Volume I

Dover Publications, Inc.
New York

Copyright © 1970, 1971 by George S. Koch, Jr., and Richard F. Link.
All rights reserved under Pan American and International Copyright Conventions.

Published in Canada by General Publishing Company, Ltd., 30 Lesmill Road, Don Mills, Toronto, Ontario.
Published in the United Kingdom by Constable and Company, Ltd., 10 Orange Street, London WC2H 7EG.

This Dover edition, first published in 1980, is an unabridged and corrected republication of the work originally published as two separate volumes in 1970 (Volume I) and 1971 (Volume II) by John Wiley & Sons, Inc.

International Standard Book Number: 0-486-64040-X
Library of Congress Catalog Card Number: 80-66468

Manufactured in the United States of America
Dover Publications, Inc.
180 Varick Street
New York, N.Y. 10014

Preface

All geologists who work with numbers use statistical methods, whether or not formally, and can profit from a knowledge of applied statistics. The purpose of this book is to explain some effective statistical procedures for the analysis of geological data and to discuss methods to obtain reliable data that are worth analyzing.

We write for the person who has numerical data from which he wants to draw conclusions or who has a problem whose solution will require obtaining and interpreting numerical data. We stress basic statistical methods and emphasize that thoughtful application of these methods will yield valid results. No statistical methods are introduced for their own sake but only because they have proved useful for data analysis. Because geology is a complicated and diverse science, we have purposely included some involved geological arguments. The mathematics is relatively simple, however, and requires for its understanding only elementary algebra and geometry.

This book is for readers with some geological training, who may or may not have professional experience. Although most readers may be geologists, the book will also interest mining and petroleum engineers, geochemists and geophysicists, mineral economists, and others.

The scope and arrangement of the book, outlined in Chapter 1, reflect our selection of the statistical methods that are most useful now, given the present state of geology, statistics, and electronic computing. The book is divided into two volumes. The first chapters, constituting most of Volume 1, place statistics first and geology second because the same statistical methods serve for varied geological problems. Most of the example analyses in this volume are chosen for numerical properties that illustrate specific statistical methods and are fictitious. Volume 2 places geology first and statistics second, and most of the examples are real ones for which the statistical principles introduced in Volume 1 are put to work, extended, and refined if necessary. Nowhere in the book is a large body of geological data analyzed completely because to do so would require a monographic presentation, too long for a textbook.

Although data for examples come from many fields of geology, most are from economic geology, and many are data that we ourselves have analyzed. We think that the reader will profit more from our experience with real data than from analyses of fictitious data or the second-hand data of others. This opportunity to learn from real analyses should offset the two disadvantages that readers relatively unfamiliar with economic geology may need to invest a small amount of thought in order to make a transfer to the fields of geology that interest them and that our own work is mentioned more often than we would like.

We intend this book for the practicing geologist as well as for the student in a formal course. We have closely tied it to J. C. R. Li's book on applied statistics, which is excellent for self-instruction, and have adopted Li's notation. A page of notation follows the table of contents; we hope that readers familiar with other mathematical symbols will not be seriously inconvenienced.

Our collaboration on the statistical analysis of geological data began at Oregon State University in 1956 and has been continued since 1962 in the Mine Systems Engineering Group of the U.S. Bureau of Mines. This book was outlined and partly written in the spring of 1966, when we were visiting research statisticians at Princeton University.

We are grateful for financial support to Oregon State and Princeton universities, the American Philosophical Society, and the National Science Foundation.

Several mining companies, cited in the text, have generously supplied data. We thank in particular two Mexican companies which have, at some expense and inconvenience, furnished data. Henry B. Hanson, General Manager, Minera Frisco, S.A., has permitted us access to assay records, and his staff has been most cooperative in assembling the data. J. B. Stone, formerly General Manager, Compañia Fresnillo, S.A., and his successor Ing. Luis Villaseñor, also allowed us to copy assay records; G. K. Lowther, Chief Geologist, discussed Fresnillo geology with Koch on many occasions.

Many of the statistical analyses in the book were first made for the U.S. Bureau of Mines and published in the Bureau reports cited in the text. S. W. Hazen, Jr., formerly chief of the Bureau's Mine Systems Engineering Group, encouraged and aided us in this work. Many past and present members of this group have also helped us, particularly J. H. Schuenemeyer, G. W. Gladfelter, and Velma Sturgis in computer programing. This book is not an official statement or representation of the U.S. Bureau of Mines, however.

Many of the geologists and statisticians cited in the text have helped us by discussions as well as by their published works. We have also profited from the criticisms of the following men who read all or part of Volume 1: J. C. Davis, R. C. Flemal, David Hoaglin, D. F. Merriam, A. T. Miesch, D. B. Morris, C. W. Ondrick, G. S. Watson, and Alfred Weiss. The errors and misjudgments that remain are ours alone.

We have also been helped by librarians of the U.S. Geological Survey, Denver, Colorado; the Colorado School of Mines, the Denver Public Library, and especially by Samuel Shephard, librarian of the U.S. Bureau of Mines

此为试读，需要完整PDF请访问：www.ertongbook.co

library in Denver. G. W. Johnson and Sally Konnak improved the English expression and organization of the book by careful editing. The book was typed by Vicky Yen Contreras and Verna Bertrand.

Finally, GSK would like to acknowledge a personal debt of gratitude to L. C. Graton, who taught him the importance of careful observation and verification of scientific data, to Ernst Cloos, who taught him to draw conclusions from numerical measurements, and to the late H. E. McKinstry, who taught him that information of fundamental geologic interest resides in the numerical data of the mining industry and urged him, in 1952, to begin the investigations that eventually led to this book.

Similarly RFL would like to acknowledge the help and inspiration of W. J. Dixon, who introduced him to statistics and computing, and to J. W. Tukey, who contributed materially to his development in statistics and data analysis.

George S. Koch, Jr.
Richard F. Link

Denver, Colorado
New York, New York
February 1970

Contents

PART I. INTRODUCTION

PART II. UNIVARIATE STATISTICAL METHODS

PART I

INTRODUCTION

此为试读，需要完整PDF请访问：www.ertongbook.com

Chapter 1

Introduction

Every geologist obtains numerical data—the number of hand specimens collected from a formation; strike and dip of a bed or fault; chemical rock analyses; porosity measurements; assays of ore, coal, or oil; etc. And every geologist summarizes these numerical data when he prepares the report that, as an essential part of the scientific process, communicates his findings to others. He summarizes because it is usual in the discipline of geology to obtain many more data than can be reported verbatim; for instance, a geologist may omit strike and dip measurements on a quadrangle map because they are too crowded, or he may average chemical or modal rock analyses. Exactly what he summarizes, and how, will determine much of the value of the report.

The geologist who has substantial amounts of data must use statistics—whether formally or informally. Our thesis is that he can sharpen his thinking and improve the reliability of his conclusions through purposefully devised statistical methods that yield incisive results. He may thus avoid the end effect too often produced today—tables of uninterpreted or misinterpreted numerical data, attached only as ornaments to his report.

Although an increasingly large number of geologists study statistics, a gap remains between academic study and useful practical application. Most text-books on statistics stress such problems as drawing black and white marbles from urns, measuring the life of light bulbs, and counting beer bottles that leak. Problems of time and space relations, problems that are at the heart of geological thinking, are seldom found. Therefore it is difficult for geologists to make the translation from statistical theory to their immediate problems.

In this book we emphasize data analysis rather than the application of formal statistics. Although the geologist interested in applying statistics to geology must learn some formal techniques, it is even more important for him to develop taste and judgment. We would rather explain the effective allocation of effort to real problems than develop complicated analyses that are likely to be mathematically unstable and computationally unreliable. Our point of view owes much to J. W. Tukey (1962, p. 2), who has written the following:

> I have come to feel that my central interest is in *data analysis*, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data. Large parts of data analysis are inferential . . . but these are only parts, not the whole. Large parts of data analysis are incisive, laying bare indications which we could not perceive by simple and direct examination of the raw data, but these too are only parts, not the whole. Some parts of data analysis, as the term is here stretched beyond its philology, are allocation, in the sense that they guide us in the distribution of effort and other valuable considerations in observation, experimentation, or analysis. Data analysis is a larger and more varied field than inference, or incisive procedures, or allocation.

1.1 SCOPE AND ARRANGEMENT OF THE BOOK

Because we write for those who want to apply statistics to geology, we assume that our readers know some geology, but not necessarily any statistics. The book starts from elementary statistical principles that are sufficiently developed to be put to work on geological data. No attempt is made to survey applied statistics or to review the literature on statistics applied to geology.

Although the discussion of statistical methods starts with first principles, it sometimes ends in specialized procedures not treated in elementary statistics books. Only those statistical methods are introduced that advance interpretation of geological data and are needed to explain a geological problem. Many numerical examples, most of them based on real geological data, are calculated. Simple algebraic arguments are given but complicated proofs requiring calculus are omitted. Even the reader who knows statistics will find it helpful to scan the first, elementary chapters to become acquainted with our viewpoint and notation.

All of the above is within the scope of the book. In the next paragraphs, some related subjects that are not treated in the book are mentioned, and works that explain them are suggested.

Because we are concerned with selected statistical subjects, a broad exposition of statistics, even on an elementary level, is not attempted. Of the many excellent textbooks on statistics, two of the best for geologists are by J. C. R. Li (the two volumes are referenced as 1964, I; and 1964, II) and Dixon and Massey (1969). Li's book (especially vol. I) is well suited to self-instruction. It is written for scientists and engineers and explains statistics in great detail for readers whose mathematics is limited or rusty. Anyone who reads one or both of these books and works the problems will obtain a good grounding in the basic statistics useful to a geologist.

We do not discuss mathematical statistics, a subject explained in many books, for example, the introductory books by Hoel (1962 and 1966) and an advanced book by Wilks (1962). Probability, a central subject in theoretical statistics, is discussed here only briefly; the reader interested in this subject is referred to an elementary text by Mosteller and others (1961) and an advanced book by Feller (vol. I, 1957 and vol. II, 1966).

Five excellent books on statistics in geology cover some additional topics not touched upon in this book. Miller and Kahn (1962) provide an excellent summary of the literature up to 1962, after which so many papers have appeared that a full review volume could contain little else. Besides reviewing the literature, Miller and Kahn discuss such subjects as probability, probability density functions, and paleobiometrics, to all of which we devote little attention. Krumbein and Graybill (1965) stress model formulation and draw most of their examples from sedimentation and oil geology. In *U.S. Bureau of Mines Bulletin 621*, Hazen (1967) covers mining technology, special kinds of sampling, and size distributions of particles in ores and rocks. Griffiths (1967) discusses statistical methods of studying sediments, but his book has wider application, contains statistical methods not mentioned in our book, and offers perceptive comments on the science of geology. Finally, Harbaugh and Merriam (1968) emphasize statistical methods implemented by electronic computers for studying stratigraphy.

This book is divided into two volumes and six parts, three in each volume. Part I is this first introductory chapter. In Part II, which covers Chapters 2 to 6, univariate statistical methods for the analysis of single variables are explained and illustrated through examples; only enough geological data are introduced to help explain the statistics. Part III, which is Chapters 7 and 8, takes up geological sampling and variability. In Part III the formidable problems of data interpretation as it pertains to geology are stressed, and attention is given, therefore, to geological as well as statistical problems of data analysis. Part IV, comprising Chapters 9 to 11 in Volume 2, develops multivariate statistical methods for the analysis of two or more variables. Part V, which includes Chapters 12 to 16, is about the statistical analysis of data from applied geology, mainly mining geology; the chapters pertain to exploration for natural resources and their valuation, to decision making through operations research, and to some specialized methods of sampling and data analysis. Finally, Part VI, or Chapter 17, is about the use of electronic computers to implement the statistical methods.

1.2 PROBLEM SOLVING

The geological investigation that statistics is likely to benefit is that which focuses on solving a specific problem. Good advice, simply summarized, is as follows: have a goal, plan ahead, and use any valid geological and statistical methods, but only those methods that will lead to that goal.

First and foremost, the investigator should think carefully about the problem he is about to pursue: the scope, limitations, ramifications, and objectives. Then through appropriate methodology he may construct that type of hypothesis known as a model (sec. 1.5), purposefully collect data to test the hypothesis, and solve the problem within stated limits of reliability—perhaps in the process revising his model in the light of the preliminary results and then, if necessary, collecting additional data to refine and to verify his conclusions. Thus, he will neither become overwhelmed with enormous amounts of data, nor fail to collect that which is pertinent. He will be able to reach an effective solution to a specific problem while keeping within limits of cost—whether reckoned in money or in time, or in both.

An example of the success to be achieved by striving for one objective is the exploration program that resulted in the discovery in the United States after World War II of large new uranium deposits by government and industry geologists. Much of the new uranium was found by systematic search in environments deemed unfavorable by theoretical geology. It is interesting to note that later, when the data collected primarily for finding ore were used for secondary purposes, such as stratigraphic and geochemical studies, they were found to be not wholly suitable. Thinking in retrospect, some of the investigators wished that, for the new studies directed at different problems, they had collected new data.

In problem solving, rather than being tied only to traditional geology, one should accept the imaginative, integrated use of any and all sciences and engineering fields that have anything to contribute. It is in this spirit that this book includes approaches from a number of disciplines. Cloud (1964) writes in a similar vein:

> What is most characteristic philosophically, and most gratifying to me personally, about the earth sciences today is their blending of the useful parts of classical

science with the most exciting aspects of advancing science. . . . All of the systematic sciences, and I use *systematic* in the broad sense of classifying *and* explaining, are in a state of ferment as new equipment, new measurements, and improved computer facilities provide different and in some instances more fundamental bases for classification and rapid quantitative methods of evaluation—this is true, not only of mineralogy but also of paleontology and petrology. . . .

If we were to tabulate the things that most generally characterize the earth sciences in the modern world we might include:

1. A growing restiveness with traditional methods of investigation.
2. An increasing tendency to express observations and conclusions quantitatively wherever it is possible to do so.
3. An increasing degree of interaction with other sciences.
4. The assumption or requirement of an increasing degree of familiarity with mathematics as a form of communication, and with physics, chemistry, and biology. . . .
5. A high degree of sophistication of instrumentation that is increasing our resolving power in all fields and permitting us to make new observations and discoveries, both in new fields and in fields that once appeared on the verge of foreclosure.

William Hambleton (1966), associate director of the Kansas Geological Survey, takes a similar view toward the work of state geological surveys in particular. He writes as follows:

As to the program of a modern survey, one might say that it is characterized by change, urgency, and involvement in the social, economic, and political problems of the state and region which it influences. . . . One hundred years ago, the purpose of our organization was to survey the mineral resources of the state. Today, it is an instrument of economic development; its research serves to catalyze activity in the mineral industries; and its plans, purposes, and programs are characterized by innovation. For a number of years we have engaged in projects involving computer techniques for fundamental geologic problems. . . . For the past several years we have involved people from geophysics, statistics, petroleum engineering, mining engineering, geology, and econometrics in studies to develop methodologies for regional economic analysis.

In other words, we are dealing increasingly with the whole field of systems analysis and operations research in the economics of the mineral industries. We are launching urban development and environmental geology studies that relate to the problems of environmental health, transportation, land use, and urban and regional planning. . . .

Most of the studies that I have mentioned emphasize change and new directions. The methods used involve transference of ideas from one discipline to another. The systems approach of the engineer is evident; we have drawn heavily upon management science, biometrics, the correlation techniques of psychology, mathematics and statistics, as well as chemistry and physics.

1.3 THE NATURE OF GEOLOGICAL DATA

Geology differs from the experimental sciences in that most geological data are fragmentary and are derived from the surface manifestations of natural processes that are uncontrolled by the investigator. When a geologist inspects a particular place on the earth, he finds a unique situation developed over a long time through more processes than he can take into account. He cannot erase the natural processes that produced this environment and do it over with a simpler, controlled laboratory experiment. Nor can he observe the natural processes; most of them are finished, and the results are fixed. Furthermore, he finds that the natural processes worked to destroy or remove part of the evidence. And most of the remaining evidence is buried inaccessibly deep in the ground, while the surface outcrops are contaminated by water, weather, and the works of mankind.

So the geologist must make do with the data available, which are seldom those with which he would prefer to work. It is here that statistics may enable him to plan data collection and deduce inferences that are not readily discernible from the raw fragmentary observations that he collects.

Despite these difficulties the investigator will find before him ten thousand times more potential data than he can collect, most of it useless to his purpose. He must, first, be aware that this is the situation and then be selective in choosing the data he will collect. Sampling guided by statistical design will serve him well in choosing these data. For this reason, designed sampling and deliberate selection of data are the major concerns of this book. The available material and the purpose govern the design, the design governs the sampling, and sampling provides the data from which valid conclusions may be inferred.

The sampling results—raw data, derived observations, and conclusions—should be verified before being accepted as valid information on which to base an analysis. Verification, an essential requirement of the scientific method, is too often done informally or not at all in geology. The reader has undoubtedly visited areas whose geology does not correspond to the published maps and has searched for fossil or mineral localities whose positions are inadequately described. Even "quantitative" data such as rock analyses are often suspect, as the well-known silicate rock studies of the U.S. Geological Survey showed ("B-1" and "G-1" rocks, R. E. Stevens and others, 1960). Yet with reason-

此为试读，需要完整PDF请访问：www.ertongbook.com