

**STATISTICAL METHODS FOR THE  
EARTH SCIENTIST**

STATISTICAL METHODS  
FOR THE  
EARTH SCIENTIST

An Introduction

Roger Till

*Department of Geology, University of Reading*

MACMILLAN

© Roger Till 1974

All rights reserved. No part of this publication  
may be reproduced or transmitted, in any form  
or by any means, without permission

*First published 1974 by*  
THE MACMILLAN PRESS LTD  
*London and Basingstoke*  
*Associated companies in New York Dublin*  
*Melbourne Johannesburg and Madras*

SBN 333 14800 2 (hard cover)  
333 15005 8 (paper cover)

ISBN 978-0-333-15005-4    ISBN 978-1-349-15536-1 (eBook)  
DOI 10.1007/978-1-349-15536-1

Filmset in 'Monophoto' 569 Mathematical Times  
and printed offset litho in Great Britain by  
Page Bros (Norwich) Ltd, Norwich

Distributed in the United States and Canada  
by Halsted Press, a Division of  
John Wiley & Sons, Inc., New York

Library of Congress Catalog Card No. 73-22704

The paperback edition of this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

**To Libby**

# CONTENTS

<i>Preface</i>	(xiii)
<b>1. STATISTICS AND MEASUREMENT IN THE EARTH SCIENCES</b>	<b>1</b>
1.1 The Measurement Process precision and accuracy—scales of measurement—parametric and non-parametric statistics	2
1.2 The Purpose of Measurements in Geology	5
<b>2. PROBABILITY</b>	<b>7</b>
2.1 The Probability Rules probability of compound events—permutations and combinations	8
2.2 Transition Probability Matrices Markov processes	11
2.3 Applications	15
<b>3. SOME DISTRIBUTIONS AND THEIR PROPERTIES</b>	<b>16</b>
3.1 Representing and Describing Distributions graphing a distribution—describing a distribution	17
3.2 The Binomial Distribution	25
3.3 The Normal Distribution and its Standardised Form the standardised normal distribution	30
3.4 Some Other Distributions the circular normal or von Mises distribution—the log-normal distribution—the Poisson distribution	38

4. SAMPLING AND TESTS OF SIGNIFICANCE	47
4.1 Sampling Procedures	48
4.2 Some Important Sample Properties	52
the distribution of sample means—the central limit theorem—the sample variance	
4.3 Student's <i>t</i> -Distribution and an Introduction to Hypothesis Testing	56
Student's <i>t</i> -distribution—testing hypotheses – (comparison of two means)—the power of a test	
4.4 The <i>F</i> -Distribution	66
4.5 The $\chi^2$ -Distribution	68
goodness-of-fit—test for the Markov property	
4.6 Error Theory	75
precision and accuracy—propagation of errors—inter-laboratory experiments	
4.7 Applications	81
5. CORRELATION AND REGRESSION	83
5.1 Correlation	83
Pearson's product moment coefficient of linear correlation—graphing variables—beginning multivariate analysis—closed number systems	
5.2 Classical Regression	91
5.3 The Reduced Major Axis Line	99
data transformations	
5.4 Applications	103
6. ANALYSIS OF VARIANCE	104
6.1 Homogeneity of Variance	105
Hartley's maximum- <i>F</i> test—Cochran's test	
6.2 A Simple (One-Way) Analysis of Variance	108
Fisher's least significant difference test—assumptions and transformations	
6.3 A Two-Way Analysis of Variance	113
6.4 Applications	116
7. NON-PARAMETRIC STATISTICS	117
7.1 Contingency Tables (Comparing Sets of Normal Data)	118
Fisher's exact probability test—the chi-square test	
7.2 Comparing Two Sets of Ordinal Data	124
the Wilcoxon–Mann–Whitney test—the Kolmogorov–Smirnov two-sample test	

7.3 Correlation Methods	130
the Spearman rank correlation coefficient—coefficients of similarity	
7.4 Applications	138
8. CONCLUSION	139
8.1 Partial Correlation and Multiple Regression	140
8.2 Trend Surfaces	140
8.3 Cluster Analysis and Discriminant Analysis	141
8.4 Principal Component Analysis and Factor Analysis	142
8.5 Time-Series Analysis	143
8.6 Postscript	143
<i>References</i>	145
<i>Further Reading</i>	149
<i>Index</i>	151

# PREFACE

This textbook is intended for use by earth science undergraduate and post-graduate students who wish to learn about the application of statistics to their subject. It is also suitable for those earth scientists who having qualified some years ago did not have a statistics course in their degree syllabus, and now feel the need to learn how to apply statistical methods in their work.

The procedures described are relevant to many branches of the earth sciences, although the majority of examples used are taken from the work of geologists. Nonetheless the examples span not only 'traditional' geological study of rocks and fossils, but wider fields including analysis of natural waters, study of the processes of weathering and sedimentation, description of landforms and measurement of engineering properties. As a result this text should be of value to the geophysicist, the physical geographer and the soil scientist as well as the geologist, since all study the same processes and materials, but from different vantage points, and all use the same statistical procedures.

The book has arisen from a course that I have been involved in teaching at Reading for the last three years. I discovered that though lengthy, expensive and advanced tomes on the use of statistics in the earth sciences existed (Miller and Kahn, 1962; Krumbein and Graybill, 1965; Griffiths, 1967; Koch and Link, 1970, 1971; Davis, 1973), no introductory texts suitable for student use were available. My aim is therefore to produce a text which can be used with a one term course of about forty hours.

Throughout the book statistical procedures are described that are needed to answer real problems that occur in the earth sciences and in all cases the procedures are worked through with a realistic data set. A 'cookery-book' approach to the use of statistics is not advocated or employed. I believe, and the literature shows, that serious errors can be made if a statistical test is applied without understanding its background and assumptions. At the same time, only simple algebra is required throughout this book. Many people have tended to shy away from statistics because of the unfamiliar mathematical shorthand it employs. I have attempted to show that if you can understand



the form and nature of a population, the concepts of probability, and the use of sample estimates of population parameters, then all statistical tests become many variations on the same theme.

All the calculations in this book can easily be performed on a small mechanical or electronic calculator. Most students would nowadays progress to using a computer, either through an on-line terminal or a 'cafeteria-type' of service. It is important to start with some form of manual calculation since you learn more about a statistical procedure by working through its computation.

I have not provided an appendix of statistical tables because I think that just as you should have a set of log tables so you should have a pocket set of statistical tables. The comprehensive books of tables by Fisher and Yates (1948) and Pearson and Hartley (1969, generally referred to as the *Biometrika* tables) will be available in all your libraries. In the text I have produced graphs of functions which that are not usually found in the pocket sets of statistical tables.

The first three chapters of the book deal with the basic theory of statistics—measurement, probability and the nature of populations. Building on these, the next four chapters deal with the actual application of statistical tests (to data sets we are likely to collect), sampling, testing samples, *t*-tests,  $\chi^2$  tests, correlation and regression, and *F*-tests. The last of these chapters is devoted to non-parametric statistical methods. The nature of many observations in the earth sciences makes these last methods important although the whole field of non-parametric testing has been neglected in previous texts. The most commonly used non-parametric test for nominal and ordinal data are described, followed by a consideration of correlation methods. The final chapter gives a short, non-mathematical guide to the more advanced statistical procedures. In the relevant chapters a few references are given of examples of good statistical usage in the earth science literature. I hope that readers will go on from this book to read these papers and work through the data they contain and that this book will provide a basis for moving on to the more advanced books I referred to earlier, and also for dipping into some non-earth science statistical texts. I have found those by Fryer (1966), Li (1964), and Yule and Kendall (1953) readable and very useful.

Many friends and colleagues have helped me with this book. John Allen, Roger Anderton, Ken Bailey, Dave Hopkins, Tom Huntingdon, Andrew Parker, Tom Prudence, Robert Sawyer and John Thomas here at Reading kindly provided me with sets of data. So did Paul Bridges (at Derby College of Art and Technology), Charles Curtis (at Sheffield University), Colin Dunn (at Department of Mineral Resources, Regina, Canada), Mike Leeder (at Leeds University) and Pat Wilson (at Kings Langley School, Hertfordshire). These people are all acknowledged at the appropriate places in the text. Perce Allen kindly allowed me to base my derivation of the binomial distri-

bution on one of his former lectures. I have had many useful discussions with John Allen and John Thomas, both of whom have always helped and encouraged me. Clive McCann and Dave Hopkins taught with me on our undergraduate course and we have all learnt a lot from this collaboration and attempt at team-teaching, as I hope did the students!

I am particularly grateful to Clive McCann for his friendship and encouragement. He also helped by reading and offering useful advice on several sections of the book. Don Woodrow (at Hobart and William Smith Colleges, New York) kindly read and commented on an early draft of part of the book. My interest in statistics was fostered by contact with Charles Curtis and Alan Spears when I was a research student at Sheffield University. This early enthusiasm was vital to me.

The facilities and environment in which to write this book have been made readily available to me by Perce Allen here at Reading. I am greatly indebted to him. My thanks go to Linda Barber, Cynthia Daulby, Christine Jones and Ann Fuller, who all dealt uncomplainingly with my handwriting to type the manuscript. My wife has helped and inspired me greatly in this project.

*Reading*

ROGER TILL