

Lecture Notes in Earth Sciences

Edited by Somdev Bhattacharji, Gerald M. Friedman,
Horst J. Neugebauer and Adolf Seilacher

18

N. M. S. Rock

Numerical Geology

A Source Guide, Glossary and Selective Bibliography
to Geological Uses of Computers and Statistics



Springer-Verlag

Berlin Heidelberg New York London Paris Tokyo

Author

Dr. Nicholas M. S. Rock

Department of Geology, University of Western Australia

Nedlands, 6009, Western Australia

ISBN 3-540-50070-7 Springer-Verlag Berlin Heidelberg New York

ISBN 0-387-50070-7 Springer-Verlag New York Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this publication or parts thereof is only permitted under the provisions of the German Copyright Law of September 9, 1965, in its version of June 24, 1985, and a copyright fee must always be paid. Violations fall under the prosecution act of the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1988

Printed in Germany

Printing and binding Druckhaus Beltz, Hemsbach/Bergstr.

2132/3140-543210

ACKNOWLEDGEMENTS

This book was completed despite the best efforts of Malcolm (aged 6) and Duncan (aged 4). I would like to thank the University of Western Australia (especially my colleagues in the Geology Department) for giving me the opportunity to write the work, and for all those on and off campus who contributed comments on its various stages. Permission to use Fig.3.2 (ERCC), Figs.4.1–4.2 (WARCC), Fig.4.4 (ESRI) and Table 3.3 (Gordon & Breach publishers) was much appreciated. Owners of various trade marks and trade names which are used frequently in this book are also acknowledged below. Other names annotated TM or © in the text are trade names, trade marks or registered trade marks of various software and hardware suppliers, which are too numerous to be listed individually.

Since desktop publishing technology is very new and may be unfamiliar to some readers, brief comments on how this book was produced may be of interest. The text was 'desktop published' entirely by the author on Macintosh microcomputers, using MacWrite for text, Superpaint for figures, MacΣqn for mathematical formulae and ANUGraph for curves of statistical functions. It was output on an Apple Laserwriter Plus laserprinter. The laserwriter output was then used as camera-ready copy for reduction to 80% in final printing by the publishers. The main text is in Times Roman font, and the literature citations in Avant Garde, while Helvetica is used for annotation on figures and Symbol font (Σψμβολ φοντ) for Greek letters.

ANUGraph is a tradename of N.Smythe, M.Ward and the ANU.

Apollo, Apple, CLIRS, Cray, Cyber, DIALOG, ICL, NAG, Oracle, PLATO, Prime, SAS, SPSS, Sun, & Systat are trademarks, tradenames or registered trademarks of the eponymous companies or organizations.

ARC/INFO is a registered trademark of ESRI Ltd.

Avant Garde is a trademark of International Typeface Corporation.

CEEFAX is a trademark of Independent Television (UK).

DEC, UNIX, VAX & VMS are registered trademarks of Digital Equipment Corp.

Helvetica and **Times** are trademarks of Allied (Linotype) Corporation.

IBM, IBM-PC and PC-DOS are registered trademarks of International Business Machines Corp.

Macintosh, MacWrite, MacDraw, MacPaint, MacTerminal, Imagewriter, Laserwriter,

Postscript, are registered or licensed trademarks of Apple Corp. or CLARIS Corp.

MacΣqn is a trade mark of Software for Recognition Technologies.

Microsoft, MS-DOS, PC-DOS and XENIX are registered trademarks of Microsoft Corp.

ORACLE is a trademark of the British Broadcasting Corporation and also of Oracle Corporation.

PRESTEL is a trademark of British Telecom.

Statview 512+ and **Statview II** are trade names of Brainpower Inc. and Abacus Concepts Inc.

Superpaint is a registered trademark of Silicon Beach Software Inc.

VIATEL is a trademark of Australian Telecom.

CONTENTS

List of Symbols and Abbreviations Used.....	1
Introduction — Why this book?	
Why study Numerical Geology?.....	3
Rationale and aims of this book.....	5
How to Use this Book.....	7

SECTION I: INTRODUCTION TO GEOLOGICAL COMPUTER USE

TOPIC 1. UNDERSTANDING THE BASICS ABOUT COMPUTERS

1a. Background history of computer use in the Earth Sciences.....	9
1b. Hardware: computer machinery.....	10
1b1.Types of computers: accessibility, accuracy, speed and storage capacity.....	10
1b2.Hardware for entering new data into a computer.....	13
1b3.Storage media for entering, retrieving, copying and transferring pre-existing data.....	14
1b4.Hardware for interacting with a computer: Terminals.....	15
1b5.Hardware for generating (outputting) hard-copies.....	16
1b6.Modes of interacting with a computer from a terminal.....	17
1b7.The terminology of data-files as stored on computers.....	18
1c. Software: programs and programming languages.....	18
1c1.Types of software.....	18
1c2.Systems software (operating systems and systems utilities).....	18
1c3.Programming languages.....	20
1c4.Graphics software standards.....	23
1d. Mainframes versus micros — which to use?.....	23

TOPIC 2. RUNNING PROGRAMS: MAKING BEST USE OF EXISTING ONES, OR PROGRAMMING YOURSELF

2a. Writing stand-alone programs from scratch.....	25
2b. Sources of software for specialised geological applications.....	26
2c. Using proprietary or published subroutine libraries.....	27
2d. Using 'Everyman' packages.....	29
2e. A comparison of options.....	32

TOPIC 3. COMPUTERS AS SOURCES OF GEOSCIENCE INFORMATION: NETWORKS & DATABASES

3a. Communicating between computer users: Mail and Network Systems.....	34
3b.Archiving and Compiling Large Bodies of Information: Databases and Information systems.....	37
3b1.Progress with Databases and Information Dissemination in the Geoscience Community.....	37
3b2.Implementing and running databases: DataBase Management Systems (DBMS).....	43
3b3.Database architecture — types of database structure.....	45
3b4. Facilitating exchange of data: standard formats and procedures.....	48

TOPIC 4. WRITING, DRAWING AND PUBLISHING BY COMPUTER

4a. Computer-assisted writing (word-processing).....	49
4b. Computer-Assisted (Desktop) Publishing (CAP/DTP).....	50
4c. Producing Maps & Plots: Computer-Assisted Drafting (CAD) and Mapping.....	52
4d. Combining graphics and databases: Geographic Information Systems (GIS).....	54

TOPIC 5. USING COMPUTERS TO BACK UP HUMAN EFFORT: COMPUTER-ASSISTED TEACHING, EXPERT SYSTEMS & ARTIFICIAL INTELLIGENCE

5a. Computers as teachers: computer-aided instruction (CAI).....	57
5b. 'Humanoid computers' in geology: Artificial Intelligence (AI) and Expert Systems.....	59

SECTION II. THE BEHAVIOUR OF NUMBERS: ELEMENTARY STATISTICS

TOPIC 6. SCALES OF MEASUREMENT AND USES OF NUMBERS IN GEOLOGY

6a. Dichotomous (binary, presence/absence, boolean, logical, yes/no) data.....	63
6b. Nominal (multistate, identification, categorical, grouping, coded) data.....	64
6c. Ordinal (ranking) data.....	64
6d. Interval data.....	65
6e. Ratio data.....	66
6f. Angular (orientation) data.....	66
6g. Alternative ways of classifying scales of measurement.....	67

TOPIC 7. SOME CRUCIAL DEFINITIONS AND DISTINCTIONS

7a. Some Distinctions between Important but Vague Terms.....	68
7b. Parametric versus robust, nonparametric and distribution-free methods.....	69
7c. Univariate, Bivariate and Multivariate methods.....	72
7d. Q-mode versus R-mode Techniques.....	72
7e. One-group, Two-group and Many-(multi-)group tests.....	72
7f. Related (paired) and independent (unpaired) data/groups.....	72
7g. Terminology related to hypothesis testing.....	73
7h. Stochastic versus Deterministic Models.....	75

TOPIC 8. DESCRIBING GEOLOGICAL DATA DISTRIBUTIONS

8a. The main types of hypothetical data distribution encountered in geology.....	76
8a1. The Normal (Gaussian) distribution.....	76
8a2. The LogNormal distribution.....	77
8a3. The Gamma (Γ) distribution.....	80
8a4. The Binomial distribution.....	80
8a5. The Multinomial distribution.....	81
8a6. The Hypergeometric distribution.....	81
8a7. The Poisson distribution.....	82
8a8. The Negative Binomial distribution.....	82
8a9. How well are the hypothetical data distributions attained by real geological data?.....	83
8b. The main theoretical sampling distributions encountered in geology.....	84
8b1. χ^2 distribution.....	84
8b2. Student's t distribution.....	84
8b3. Fisher's (Snedecor's) F distribution.....	85
8b4. Relationships between the Normal and statistical distributions.....	85
8c. Calculating summary statistics to describe real geological data distributions.....	86
8c1. Estimating averages (measures of location, centre, central tendency).....	86
8c2. Estimating spread (dispersion, scale, variability).....	91
8c3. Estimating symmetry (skew) and 'peakedness' (kurtosis).....	92
8d. Summarising data graphically: EXPLORATORY DATA ANALYSIS (EDA).....	92
8e. Comparing real with theoretical distributions: GOODNESS-OF-FIT TESTS.....	94
8e1. A rather crude omnibus test: χ^2	95
8e2. A powerful omnibus test: the Kolmogorov ("one-sample Kolmogorov-Smirnov") test.....	95
8e3. Testing goodness-of-fit to a Normal Distribution: specialized NORMALITY TESTS.....	96
8f. Dealing with non-Normal distributions.....	99
8f1. Use nonparametric methods, which are independent of the Normality assumption.....	99
8f2. Transform the data to approximate Normality more closely.....	99
8f3. Separate the distribution into its component parts.....	100
8g. Testing whether a data-set has particular parameters: ONE-SAMPLE TESTS.....	101
8g1. Testing against a population mean μ (population standard deviation σ known): the zM test.....	101
8g2. Testing against a population mean μ (population standard deviation σ known): one-group t -test.....	101

TOPIC 9. ASSESSING VARIABILITY, ERRORS AND EXTREMES IN GEOLOGICAL DATA: SAMPLING, PRECISION AND ACCURACY

9a. Problems of Acquiring Geological Data: Experimental Design and other Dreams.....	102
9b. Sources of Variability & Error in Geological Data, and the Concept of 'Entities'.....	102
9c. The Problems of Geological Sampling.....	105
9d. Separating and Minimizing Sources of Error — Statistically and Graphically.....	107

9e.Expressing errors: Confidence Limits.....	109
9e1.Parametric confidence limits for the arithmetic mean and standard deviation.....	109
9e2.Robust Confidence Intervals for the Mean, based on the Jackknife.....	110
9e3.Robust Confidence Intervals for location estimates, based on Monte Carlo Swindles.....	111
9e4.Nonparametric Confidence Limits for the Median based on the Binomial Model.....	112
9f.Dealing with outliers (extreme values): should they be included or rejected?.....	113
9f1.Types of statistical outliers: true, false and bizarre, statistical and geological.....	113
9f2.Types of geological data: the concept of 'data homogeneity'.....	114
9f3.Tests for identifying statistical outliers manually.....	115
9f4.Avoiding Catastrophes: Extreme Value Statistics.....	116
9f5.Identifying Anomalies: Geochemical Thresholds and Gap Statistics.....	117

SECTION III: INTERPRETING DATA OF ONE VARIABLE: UNIVARIATE STATISTICS

TOPIC 10. COMPARING TWO GROUPS OF UNIVARIATE DATA	118
10a.Comparing Location (mean) and Scale (variance) Parametrically: <i>t</i> - and <i>F</i> -tests.....	120
10a1.Comparing variances parametrically: Fisher's <i>F</i> -test.....	120
10a2.Comparing Two Means Parametrically: Student's <i>t</i> -test (paired and unpaired).....	121
10b.Comparing two small samples: Substitute Tests based on the Range.....	123
10c.Comparing Medians of Two Related (paired) Groups of Data Nonparametrically.....	123
10c1.A crude test for related medians: the Sign Test.....	124
10c2.A test for 'before-and-after' situations: the McNemar Test for the Significance of Changes.....	124
10c3.A more powerful test for related medians: the Wilcoxon (matched-pairs, signed-ranks) Test.....	125
10c4.The most powerful test for related medians, based on Normal scores: the Van Eeden test.....	126
10d.Comparing Locations (medians) of Two Unrelated Groups Nonparametrically.....	126
10d1.A crude test for unrelated medians: the Median Test.....	127
10d2.A quick and easy test for unrelated medians: Tukey's <i>T</i> test.....	127
10d3.A powerful test for unrelated medians: the Mann-Whitney test.....	128
10d4.The Normal scores tests for unrelated medians: the Terry-Hoeffding test.....	129
10e.Comparing the Scale of Two Independent Groups of Data Nonparametrically.....	129
10e1.The Ansari-Bradley, David, Moses, Mood and Siegel-Tukey Tests.....	129
10e2.The Squared Ranks Test.....	130
10e3.The Normal scores approach: the Klotz Test.....	131
10f.Comparing the overall distribution of two unrelated groups nonparametrically.....	132
10f1.A crude test: the Wald-Wolfowitz (two-group) Runs Test.....	132
10f2.A powerful test: the Smirnov (two-group Kolmogorov-Smirnov) Test.....	133
10g. A Brief Comparison of Results of the Two-group Tests in Topic 10.....	134

TOPIC 11. COMPARING THREE OR MORE GROUPS OF UNIVARIATE DATA: One-way Analysis of Variance and Related Tests

11a.Determining parametrically whether several groups have homogeneous variances.....	135
11a1.Hartley's maximum- <i>F</i> test.....	136
11a2.Cochran's <i>C</i> Test.....	136
11a3.Bartlett's <i>M</i> Test.....	136
11b.Determining Parametrically whether Three or more Means are Homogeneous: One-Way ANOVA.....	138
11c.Determining which of several means differ: MULTIPLE COMPARISON TESTS.....	140
11c1.Fisher's PLSD (= protected least significant difference) test.....	141
11c2.Scheffé's <i>F</i> Test.....	142
11c3.Tukey's <i>w</i> (HSD = Honestly Significant Difference) Test.....	142
11c4. The Student-Neuman-Keuls' (S-N-K) Test.....	143
11c5. Duncan's New Multiple Range Test.....	143
11c6. Dunnett's Test.....	144
11d.A quick parametric test for several means: LORD'S RANGE TEST.....	144
11e.Determining nonparametrically whether several groups of data have homogeneous medians.....	145
11e1.The <i>q</i> -group extension of the Median Test.....	145
11e2.A more powerful test: The Kruskal-Wallis One-way ANOVA by Ranks.....	145
11e3.The most powerful nonparametric test based on Normal scores: the Van der Waerden Test.....	147
11f.Determining Nonparametrically whether Several Groups of Data have Homogeneous Scale:	

THE SQUARED RANKS TEST.....	147
11g. Determining Nonparametrically whether Several Groups of Data have the same Distribution Shape.....	148
11g1. The 3-group Smirnov Test (Birnbaum-Hall Test).....	148
11g2. The q -group Smirnov Test.....	149
11h. A brief comparison of the results of multi-group tests in Topic 11.....	150
TOPIC 12. IDENTIFYING CONTROLS OVER DATA VARIATION: MORE SOPHISTICATED FORMS OF ANALYSIS OF VARIANCE	
12a. A General Note on ANOVA and the General Linear Model (GLM).....	151
12b. What determines the range of designs in ANOVA?.....	152
12c. Two-way ANOVA on several groups of data: RANDOMIZED COMPLETE BLOCK DESIGNS and TWO-FACTORIAL DESIGNS WITHOUT REPLICATION.....	155
12c1. The parametric approach.....	156
12c2. A simple nonparametric approach: the Friedman two-way ANOVA test.....	157
12c3. A more complex nonparametric approach: the Quade Test.....	158
12d. Two-way ANOVA on several related but incomplete groups of data: BALANCED INCOMPLETE BLOCK DESIGNS (BIBD).....	159
12d1. The parametric approach.....	159
12d2. The nonparametric approach: the Durbin Test.....	160
12e. Some Simple Crossed Factorial Designs with Replication.....	161
12e1. Two-factor crossed complete design with Replication: Balanced and Unbalanced.....	162
12e2. Three-factor crossed complete design with Replication: Balanced and Unbalanced.....	163
12f. A Simple Repeated Measures Design.....	164
12g. Analyzing data-within-data: HIERARCHICAL (NESTED) ANOVA.....	166

SECTION IV. INTERPRETING DATA WITH TWO VARIABLES: Bivariate Statistics

TOPIC 13. TESTING ASSOCIATION BETWEEN TWO OR MORE VARIABLES: Correlation and concordance	167
13a. Measuring Linear Relationships between two Interval/ratio Variables: PEARSON'S CORRELATION COEFFICIENT, r	168
13b. Measuring Strengths of Relationships between Two Ordinal Variables: RANK CORRELATION COEFFICIENTS.....	170
13b1. Spearman's Rank Correlation Coefficient, ρ	170
13b2. Kendall's Rank Correlation Coefficient, τ	171
13c. Measuring Strengths of Relationships between Dichotomous and Higher-order Variables: POINT-BISERIAL AND BISERIAL COEFFICIENTS.....	172
13d. Testing whether Dichotomous or Nominal Variables are Associated.....	173
13d1. Contingency Tables (cross-tabulation), χ^2 (Chi-squared) tests, and Characteristic Analysis.....	173
13d2. Fisher's Exact Probability Test.....	175
13d3. Correlation coefficients for dichotomous and nominal data: Contingency Coefficients.....	176
13e. Comparing Pearson's Correlation Coefficient with itself: FISHER'S Z TRANSFORMATION.....	177
13f. Measuring Agreement: Tests of Reliability and Concordance.....	179
13f1. Concordance between Several Dichotomous Variables: Cochran's Q test.....	179
13f2. Concordance between ordinal & dichotomous variables: Kendall's coefficient of concordance.....	180
13g. Testing X - Y plots Graphically for Association, with or without Raw Data.....	181
13g1. The Corner (Olmstead-Tukey quadrant sum) Test for Association.....	181
13g2. A test for curved trends: the Correlation Ratio, $\eta(\eta)$	182
13h. Measures of weak trends: Guttman's μ_2 , Goodman & Kruskal's γ	184
13i. Spurious and illusory correlations.....	185
TOPIC 14. QUANTIFYING RELATIONSHIPS BETWEEN TWO VARIABLES: Regression	
14a. Estimating Lines to Predict one Dependent (Response) Variable from another Independent (Explanatory) Variable: CLASSICAL PARAMETRIC REGRESSION.....	187
14a1. Introduction: important concepts.....	187
14a2. Calculating the regression line: Least-squares.....	187
14a3. Assessing the significance of the regression: Coefficient of determination, ANOVA.....	188

14a4. Testing the regression model for defects: Autocorrelation and Heteroscedasticity.....	189
14a5. Assessing the influence of outliers.....	190
14a6. Confidence bands on regression lines.....	191
14a7. Comparing regressions between samples or samples and populations: Confidence Intervals.....	191
14b. Calculating Linear Relationships where Both Variables are Subject to Error: 'STRUCTURAL REGRESSION'.....	192
14c. Avoiding sensitivity to outliers: ROBUST REGRESSION.....	194
14d. Regression with few assumptions: NONPARAMETRIC REGRESSION.....	194
14d1. A method based on median slopes: Theil's Complete Method.....	194
14d2. A quicker nonparametric method: Theil's Incomplete method.....	195
14e. Fitting curves: POLYNOMIAL (CURVILINEAR, NONLINEAR) REGRESSION.....	196
14e1. The parametric approach.....	196

SECTION V: SOME SPECIAL TYPES OF GEOLOGICAL DATA

TOPIC 15. SOME PROBLEMATIC DATA-TYPES IN GEOLOGY

15a. Geological Ratios.....	200
15b. Geological Percentages and Proportions with Constant Sum: CLOSED DATA.....	202
15c. Methods for reducing or overcoming the Closure Problem.....	204
15c1. Data transformations and recalculations.....	204
15c2. Ratio normalising.....	205
15c3. Hypothetical open arrays.....	205
15c4. Remaining space variables.....	206
15c5. A recent breakthrough: log-ratio transformations.....	206
15d. The Problem of Missing Data.....	206
15e. The Problem of Major, Minor and Trace elements.....	208

TOPIC 16. ANALYSING ONE-DIMENSIONAL SEQUENCES IN SPACE OR TIME

16a. Testing whether a single Series is Random or exhibits Trend or Periodicity.....	209
16a1. Testing for trend in ordinal or ratio data: Edgington's nonparametric test.....	210
16a2. Testing for cycles in ordinal or ratio data: Noether's nonparametric test.....	210
16a3. Testing for specified trends: Cox & Stuart's nonparametric test.....	210
16a4. Testing for trend in dichotomous, nominal or ratio data: the one-group Runs Test.....	212
16a5. Testing parametrically for cyclicity in nominal data-sequences: AUTO-ASSOCIATION.....	213
16a6. Looking for periodicity in a sequence of ratio data: AUTO-CORRELATION.....	215
16b. Comparing/correlating two sequences with one another.....	217
16b1. Comparing two sequences of nominal (multistate) data: CROSS-ASSOCIATION.....	217
16b2. Comparing two sequences of ratio data: CROSS CORRELATION.....	218
16b3. Comparing two ordinal or ratio sequences nonparametrically: Burnaby's χ^2 procedure.....	219
16c. Assessing the control of geological events by past events.....	221
16c1. Quantifying the tendency of one state to follow another: transition probability matrices.....	221
16c2. Assessing whether sequences have 'memory': MARKOV CHAINS and PROCESSES.....	222
16c3. Analyzing the tendency of states to occur together: SUBSTITUTABILITY ANALYSIS.....	223
16d. Sequences as combinations of waves: SPECTRAL (FOURIER) ANALYSIS.....	224
16e. Separating 'noise' from 'signal': FILTERING, SPLINES, TIME-TRENDS.....	225

TOPIC 17. ASSESSING GEOLOGICAL ORIENTATION DATA: AZIMUTHS, DIPS AND STRIKES

17a. Special Properties of Orientation Data.....	226
17b. Describing distributions of 2-dimensional (circular) orientation data.....	227
17b1. Graphical display.....	227
17b2. Circular summary statistics.....	228
17b3. Circular data distributions.....	228
17c. Testing for uniformity versus preferred orientation in 2-D orientation data.....	230
17c1. A simple nonparametric test: Hodges-Ajne Test.....	230
17c2. A more powerful nonparametric EDF test: Kuiper's Test.....	231
17c3. A powerful nonparametric test: Watson U^2 Test.....	231
17c4. The standard parametric test: Rayleigh's Test.....	232

17d. One-group tests for mean directions and concentrations.....	233
17e. Comparing two groups of 2-dimensional data.....	234
17e1. A nonparametric test: Mardia's uniform scores.....	234
17e2. An alternative nonparametric test: Watson's U^2	235
17e3. A linear nonparametric test applicable to angular data: the Wald-Wolfowitz Runs test.....	235
17e4. A parametric test for equal concentrations and mean directions: the Watson-Williams Test.....	236
17f. Comparing three or more groups of 2-dimensional data.....	237
17f1. Nonparametric testing: multigroup extension of Mardia's Uniform Scores Test.....	237
17f2. Parametric test for equal concentrations & mean directions: multigroup Watson-Williams Test.....	237
17g. Introduction to 3-dimensional Orientation Data.....	238
17h. Describing distributions of 3-dimensional orientation data.....	240
17h1. Displaying and interpreting spherical data graphically.....	240
17h2. Spherical data distributions.....	240
17h3. Summarising real distributions of spherical data: Vector Means and Confidence Cones.....	241
17i. Testing for uniformity versus preferred orientation in 3-D orientation data.....	241
17j. Comparing two groups of 3-dimensional orientation data.....	243
17j1. Testing for equality of two concentrations.....	243
17j2. Testing for equality of two mean directions: Watson-Williams test.....	243
17k. Comparing three or more groups of 3-dimensional data.....	244
17k1. Testing whether three or more concentrations are homogeneous.....	244
17k2. Testing whether three or more mean directions are homogeneous.....	244

SECTION VI: ADVANCED TECHNIQUES

Introduction.....	246
A Note on Matrix Algebra.....	246
A note on Scales of Measurement and Multivariate Nonparametric Methods.....	246
TOPIC 18. MODELLING GEOLOGICAL PROCESSES NUMERICALLY	
18a. Univariate Modelling of Magmatic Processes.....	247
18b. Multivariate Modelling of Mixing, Reactions, & Parent-Daughter Problems.....	249
TOPIC 19. ANALYZING RELATIONSHIPS BETWEEN MORE THAN TWO VARIABLES	
19a. Homing in on the Correlation of Two Variables among Many: PARTIAL CORRELATION.....	252
19b. Assessing the Effect of Several Independent Ratio Variables on One Dependent Ratio Variable: MULTIPLE REGRESSION.....	253
19b1. Multiple Linear Regression.....	253
19b2. Stepwise Multiple Regression.....	257
19b3. Overcoming problems: Ridge Regression and Generalized Least Squares (GLS).....	257
19b4. Nonparametric multiple regression.....	258
19c. Relationships involving Dichotomous or Nominal Independent Variables: GROUPED REGRESSION.....	258
19d. Relationships involving Dichotomous or Nominal Dependent Variables or Ratios: LOGISTIC REGRESSION.....	259
19e. Correlating two sets of several variables each: CANONICAL CORRELATION.....	261
TOPIC 20. ANALYSING SPATIALLY DISTRIBUTED DATA: Thin Sections, Maps, Mineral deposits and the like	
20a. Analyzing Thin Sections: Petrographic Modal Analysis.....	263
20b. Analyzing Spatial Distributions of Points on Maps.....	263
20b1. Testing for random, uniform and clustered point distributions: QUADRAT ANALYSIS.....	264
20b2. Analyzing distributions via distances between points: NEAREST NEIGHBOUR ANALYSIS.....	266
20b3. Contouring methods.....	267
20c. Fitting Planes and Surfaces to Spatial Data: TREND SURFACE ANALYSIS (TSA) AND TWO-DIMENSIONAL REGRESSION.....	269
20d. Spatial Statistics with Confidence Estimation: GEOSTATISTICS and KRIGING.....	271
20d1. Estimating rates of change of regionalized variables along specific trends: Semi-variograms.....	272
20d2. Kriging.....	273

TOPIC 21. CLASSIFYING OBJECTS FROM FIRST PRINCIPLES

21a.Measuring similarity and dissimilarity between multivariate objects: MATCHING AND DISTANCE COEFFICIENTS..... 275

 21a1.Measuring similarity between dichotomous/nominal data: MATCHING COEFFICIENTS..... 276

 21a2.Measuring similarity between ordinal to ratio data: DISTANCE COEFFICIENTS..... 278

21b.Producing Dendrograms: HIERARCHICAL CLUSTER ANALYSIS..... 281

 21b1.An Overview of Available Methods..... 281

 21b2.Divisive cluster analysis on dichotomous/nominal data: ASSOCIATION ANALYSIS..... 292

21c.Defining Discrete/Fuzzy Clusters: NON-HIERARCHICAL CLUSTER ANALYSIS..... 292

 21c1.K-means Cluster Analysis..... 293

 21c2.The Refined, Iterative K-means Method..... 295

21d.Displaying groupings in as few dimensions as possible: ORDINATION..... 296

 21d1.Metric and Nonmetric MultiDimensional Scaling (MDS)..... 296

 21d2.Principal Coordinates Analysis..... 300

 21d3.Non-linear mapping (NLM)..... 302

 21d4.Quadratic Loss Functions..... 302

TOPIC 22. COMPARING 'KNOWN'S' AND ASSIGNING 'UNKNOWN'S':

Discriminant Analysis and related methods

22a. Introduction..... 304

22b.Comparing Two or more Groups of Multivariate Data: DISCRIMINANT ANALYSIS..... 304

 22b1.Two-group Parametric Discriminant Analysis..... 307

 22b2.Multi-group Discriminant (Canonical Variate) Analysis (MDA)..... 309

22c.Comparing Groups of Multivariate Dichotomous Data: ADAPTIVE PATTERN RECOGNITION..... 312

TOPIC 23. EXAMINING STRUCTURE, RECOGNIZING PATTERNS & REDUCING THE DIMENSIONALITY OF MULTIVARIATE DATA

23a. Graphical Methods of Displaying Multivariate Data in Two-Dimensions..... 314

23b. Finding Structure in Multivariate Data: PRINCIPAL COMPONENTS ANALYSIS..... 315

23c. Fitting Multivariate Data to a Model: FACTOR ANALYSIS..... 318

 23c1.Looking for Geological Control over Data Variations: R-MODE FACTOR ANALYSIS..... 320

 23c2.Analyzing Series and Mixtures in Terms of End-members: Q-MODE FACTOR ANALYSIS..... 326

 23c3.Pattern Recognition by Combining Q- and R-modes: CORRESPONDENCE ANALYSIS..... 327

23d. Epilogue: the Statistical 'Zap' versus the 'Shotgun' Approach..... 328

Selective Bibliography..... 329

Glossary and Index..... 379