

CONTENTS

PREFACE.....	vii
--------------	-----

I

PHILOSOPHICAL CONSIDERATIONS AND COMPUTER PACKAGES

I.1. PHILOSOPHICAL CONSIDERATIONS	1
Introduction	1
Data Considerations.....	1
Pattern Recognition Approach	2
Pattern Recognition Questions	3
Pattern Recognition Nomenclature	4
Variable Coding.....	5
Categorization of Data	6
Considerations to Keep in Mind	7
Programs to Be Discussed	7
Possible Approaches.....	8
I.2. BIOMEDICAL COMPUTER PROGRAM (BMDP).....	8
Introduction	8
Program Groups	8
BMDP Control Language	9
I.3. STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES (SPSS)	10
Introduction	10
Package Structure	10
SPSS Programs	11
SPSS Control Language	11
I.4. ARTHUR	11
Introduction	11
ARTHUR Control Language	12
ARTHUR Programs	12
I.5. CLUSTAN	13
Introduction	13
CLUSTAN Format	13
I.6. SAS	13
Introduction	13
SAS Programming	14
Procedures Available	14

II

PATTERN RECOGNITION APPROACH TO DATA ANALYSIS

	Introduction	17
II.1. PRELIMINARY DATA EXAMINATION	Introduction	18
	Step II.1a. Computer Compatibility	18
	Step II.1b. Eliminating Format Errors	20
	Step II.1c. Data Listing	22
	Step II.1d. Variable Distributions	22
	Step II.1e. Identifying Unique Cases	25
	Step II.1f. Variable Distributions After Elimination	26
	Step II.1g. Introductory Statistics	26
	Standard Scores	30
II.2. DATA STRATIFICATION	Introduction	30
	<i>t</i> -Tests	30
	Step II.2a. Comparison of Two Groups	32
	Analysis of Variance	34
	Step II.2b. Comparisons of All Groups Simultaneously	35
	Multiple Way Analysis of Variance	36
	Step II.2c. Variable Subgroups	38
	Chi-Square Test	40
	Other Similar Programs	41
II.3. INTERVARIABLE RELATIONSHIPS	Introduction	41
	Correlation Coefficients	41
	Step II.3a. Calculation of Correlation Coefficients	42
	Correlation Results	44
	Step II.3b. Bivariate Plots	44
	Partial Correlations	48
	Step II.3c. Partial Correlations	49
	Step II.3d. Clustering of Variables	50
	Results of the Clustering	51
	Canonical Correlations	53
	Step II.3e. Application of Canonical Correlations	54
	Regression Analysis	57
	Step II.3f. Regression Analysis	58
	Stepwise Regression	62
	Summary	63
	PLS-2	63
	Nonparametric Statistics	64
II.4. UNSUPERVISED LEARNING TECHNIQUES	Introduction	64
	Cluster Analysis	65
	Data Set Considered	65
	Step II.4a. Minimal Spanning Tree	66
	Drawing of the Tree	66
	Cluster Definition	70
	Minimal Spanning Tree Results	70
	Step II.4b. Hierarchical Clustering	71

	Description of the Dendrogram	71
	Other Hierarchical Clustering Programs	74
	Step II.4c. Nonlinear Mapping	74
	Summary	76
	Step II.4d. CLUSTAN	76
	Hierarchical Techniques in CLUSTAN	77
	Other Clustering Techniques	78
	Summary	79
II.5.	SUPERVISED LEARNING TECHNIQUES	79
	Introduction	79
	Step II.5a. <i>k</i> -Nearest Neighbors	80
	Testing Supervised Results	82
	Step II.5b. Discriminant Analysis	83
	Results from Discriminant Analysis	86
	Other Discriminant Programs	88
	Comparison of Discriminant Programs	91
II.6.	VARIABLE REDUCTION	91
	Introduction	91
	Tools from Previous Steps	92
	Step II.6a. Selection of Variables	92
	Principal Component and Factor Analysis	94
	Step II.6b. Principal Component Analysis	95
	Classical Factor Analysis	98
	Data Considerations	99
	Factor Choices	99
	Step II.6c. Classical Factor Analysis	100
	Factor Rotations	104
	Underlying Variable Factor Analysis	105
II.7.	DATA MANIPULATIONS	105
	Introduction	105
	SPSS Capabilities	106
	BMDP and ARTHUR Alterations	107

III

IMPLEMENTATION

	Introduction	109
III.1.	TYPICAL SPSS RUNS	110
	Introduction	110
	Card Deck	111
	Data Format Statement	115
	Missing Data	115
	Statistical Procedures	116
	Category Definitions	117
	File Saving	118
	File Saving	119
III.2.	TYPICAL ARTHUR RUNS	119
	Introduction	119
	Input Card	122
	Task Definition Cards	123
	Test Data	124

III.3. TYPICAL BMDP RUNS	124
Introduction	124
Card Deck	125
Variable Information	127
Data Grouping	128
III.4. SPSS IMPLEMENTATIONS	129
Introduction	129
CANCORR	130
CONDESCRIPTIVE	131
CROSSTABS	131
DISCRIMINANT	133
FACTOR	134
FREQUENCIES	135
NONPAR CORR	136
PARTIAL CORR	136
PEARSON CORR	137
REGRESSION	137
SCATTERGRAM	139
T-TEST	139
III.5. ARTHUR IMPLEMENTATIONS	140
Introduction	140
CORREL	140
DISTANCE	141
HIER	141
KARLOV	142
KNN	142
MULTI	142
NLM	143
PLANE	143
SCALE	143
SELECT	144
TREE	145
VARVAR	145
WEIGHT	145
III.6. BMDP PROGRAMS	146
Introduction	146
BMDP1D	146
BMDP2D	147
BMDP3D	148
BMDP4D	148
BMDP5D	149
BMDP6D	150
BMDP1M	151
BMDP2M	151
BMDP4M	152
BMDP6M	153
BMDP7M	154
BMDP1R	155
BMDP2R	156

BMDP4R	157
BMDP6R	158
BMDP3S	159

IV

NATURAL SCIENCE APPLICATIONS

Introduction	161
Biological Applications	161
Medical Applications	164
Geological and Earth Science Applications	166
Environmental Applications	168
Physics Applications	169
Chemical Applications	170
Summary	172
REFERENCES	173
APPENDIX I. PATTERN RECOGNITION DEFINITIONS AND REFERENCE BOOKS	177
APPENDIX II. THE MULTIVARIATE NORMAL DISTRIBUTION	181
APPENDIX III. DATA BASE DESCRIPTION	183
APPENDIX IV. INDICES OF BMDP, SPSS, AND ARTHUR PACKAGES	195
APPENDIX V. PROGRAMS, MANUALS, AND REFERENCE INFORMATION	199
APPENDIX VI. SUMMARY OF ANALYSES AND PROGRAM CROSS-REFERENCE FOR CHAPTER II	203
APPENDIX VII. NONPARAMETRIC STATISTICS	209
APPENDIX VIII. MISSING VALUES	213
APPENDIX IX. STANDARD SCORES AND WEIGHTINGS	217
SUBJECT INDEX	221
COMPUTER PROGRAM INDEX	223