



WILEY

APPLIED MANOVA AND DISCRIMINANT ANALYSIS

Second Edition

CARL J. HUBERTY
STEPHEN OLEJNIK

WWW.
LINE AVAILABLE

Wiley Series in Probability and Statistics

Contents

List of Figures	xix
List of Tables	xxi
Preface to Second Edition	xxv
Acknowledgments	xxvii
Preface to First Edition	xxix
Notation	xxxii
I INTRODUCTION	1
1 Discriminant Analysis in Research	3
1.1 A Little History, 3	
1.2 Overview, 5	
1.3 Descriptive Discriminant Analysis, 5	
1.4 Predictive Discriminant Analysis, 7	
1.5 Design in Discriminant Analysis, 9	
1.5.1 Grouping Variables, 9	
1.5.2 Response Variables, 9	
Exercises, 13	
2 Preliminaries	15
2.1 Introduction, 15	
2.2 Research Context, 15	
2.3 Data, Analysis Units, Variables, and Constructs, 16	
2.4 Summarizing Data, 18	
2.5 Matrix Operations, 21	
2.5.1 SSCP Matrix, 22	

- 2.5.2 Determinant, 23
- 2.5.3 Inverse, 24
- 2.5.4 Eigenanalysis, 25
- 2.6 Distance, 26
- 2.7 Linear Composite, 28
- 2.8 Probability, 28
- 2.9 Statistical Testing, 29
- 2.10 Judgment in Data Analysis, 30
- 2.11 Summary, 31
 - Further Reading, 31
 - Exercises, 32

II ONE-FACTOR MANOVA/DDA 33

3 Group Separation 35

- 3.1 Introduction, 35
- 3.2 Two-Group Analyses, 35
 - 3.2.1 Univariate Analysis, 35
 - 3.2.2 Multivariate Analysis, 39
- 3.3 Test for Covariance Matrix Equality, 41
- 3.4 Yao Test, 43
- 3.5 Multiple-Group Analyses—Single Factor, 44
 - 3.5.1 Univariate Analysis, 44
 - 3.5.2 Multivariate Analysis, 47
- 3.6 Computer Application, 52
- 3.7 Summary, 56
 - Exercises, 57

4 Assessing MANOVA Effects 61

- 4.1 Introduction, 61
- 4.2 Strength of Association, 62
 - 4.2.1 Univariate, 62
 - 4.2.2 Multivariate, 62
 - 4.2.3 Bias, 65
- 4.3 Computer Application I, 66
- 4.4 Group Contrasts, 67
 - 4.4.1 Univariate, 67
 - 4.4.2 Multivariate, 68
- 4.5 Computer Application II, 72
- 4.6 Covariance Matrix Heterogeneity, 74
- 4.7 Sample Size, 74

4.8	Summary, 75 Technical Notes, 76 Exercises, 77	
5	Describing MANOVA Effects	81
5.1	Introduction, 81	
5.2	Omnibus Effects, 82	
	5.2.1 An Eigenanalysis, 82	
	5.2.2 Linear Discriminant Functions, 83	
5.3	Computer Application I, 85	
5.4	Standardized LDF Weights, 87	
5.5	LDF Space Dimension, 88	
	5.5.1 Statistical Tests, 89	
	5.5.2 Proportion of Variance, 91	
	5.5.3 LDF Plots, 91	
5.6	Computer Application II, 93	
5.7	Computer Application III, 94	
5.8	Contrast Effects, 96	
5.9	Computer Application IV, 96	
5.10	Summary, 98 Technical Note, 99 Further Reading, 100 Exercises, 100	
6	Deleting and Ordering Variables	103
6.1	Introduction, 103	
6.2	Variable Deletion, 103	
	6.2.1 Purposes of Deletion, 103	
	6.2.2 McCabe Analysis, 104	
	6.2.3 Computer Application, 105	
6.3	Variable Ordering, 106	
	6.3.1 Meaning of Importance, 106	
	6.3.2 Computer Application I, 108	
	6.3.3 Variable Ranking, 110	
6.4	Contrast Analyses, 110	
6.5	Computer Application II, 111	
6.6	Comments, 113 Further Reading, 114 Exercises, 115	
7	Reporting DDA Results	117
7.1	Introduction, 117	
7.2	Example of Reporting DDA Results, 117	

7.3	Computer Package Information, 122	
7.4	Reporting Terms, 123	
7.5	MANOVA/DDA Applications, 124	
7.6	Concerns, 124	
7.7	Overview, 126	
	Further Reading, 127	
	Exercises, 127	
III	FACTORIAL MANOVA, MANCOVA, AND REPEATED MEASURES	129
8	Factorial MANOVA	131
8.1	Introduction, 131	
8.2	Research Context, 131	
8.3	Univariate Analysis, 134	
8.4	Multivariate Analysis, 136	
	8.4.1 Omnibus Tests, 136	
	8.4.2 Distribution Assumptions, 138	
8.5	Computer Application I, 139	
8.6	Computer Application II, 146	
8.7	Nonorthogonal Design, 150	
8.8	Outcome Variable Ordering and Deletion, 151	
8.9	Summary, 152	
	Technical Notes, 152	
	Exercises, 159	
9	Analysis of Covariance	163
9.1	Introduction, 163	
9.2	Research Context, 164	
9.3	Univariate ANCOVA, 166	
	9.3.1 Testing for Equality of Regression Slopes, 166	
	9.3.2 Omnibus Test of Adjusted Means, 168	
9.4	Multivariate ANCOVA (MANCOVA), 170	
	9.4.1 Matrix Calculations, 170	
	9.4.2 Testing for Equal Slopes, 171	
9.5	Computer Application I, 173	
9.6	Comparing Adjusted Means—Omnibus Test, 174	
9.7	Computer Application II, 175	
9.8	Contrast Analysis, 180	
9.9	Computer Application III, 180	

9.10	Summary, 184 Technical Note, 184 Exercises, 190	
10	Repeated-Measures Analysis	193
10.1	Introduction, 193	
10.2	Research Context, 195	
10.3	Univariate Analyses, 196	
10.3.1	Omnibus Test, 196	
10.3.2	Contrast Analysis, 197	
10.4	Multivariate Analysis, 199	
10.5	Computer Application I, 202	
10.6	Univariate and Multivariate Analyses, 204	
10.7	Testing for Sphericity, 207	
10.8	Computer Application II, 210	
10.9	Contrast Analysis, 212	
10.10	Computer Application III, 214	
10.11	Summary, 216 Technical Notes, 217 Exercises, 223	
11	Mixed-Model Analysis	227
11.1	Introduction, 227	
11.2	Research Context, 228	
11.3	Univariate Analysis, 229	
11.4	Multivariate Analysis, 231	
11.4.1	Group-by-Time Interaction, 232	
11.4.2	Repeated-Measures Variable Main Effect, 235	
11.5	Computer Application I, 237	
11.6	Contrast Analysis, 240	
11.7	Computer Application II, 243	
11.8	Summary, 246 Technical Note, 247 Exercises, 249	
IV	GROUP MEMBERSHIP PREDICTION	253
12	Classification Basics	255
12.1	Introduction, 255	
12.2	Notion of Distance, 256	

- 12.3 Distance and Classification, 259
- 12.4 Classification Rules in General, 260
 - 12.4.1 Maximum Likelihood, 260
 - 12.4.2 Typicality Probability, 261
 - 12.4.3 Posterior Probability, 262
 - 12.4.4 Prior Probability, 263
- 12.5 Comments, 264
 - Technical Note, 265
 - Further Reading, 265
 - Exercises, 266

13 Multivariate Normal Rules

269

- 13.1 Introduction, 269
- 13.2 Normal Density Functions, 269
- 13.3 Classification Rules Based on Normality, 271
- 13.4 Classification Functions, 273
 - 13.4.1 Quadratic Functions, 273
 - 13.4.2 Linear Functions, 274
 - 13.4.3 Distance-Based Classification, 275
- 13.5 Summary of Classification Statistics, 277
- 13.6 Choice of Rule Form, 278
 - 13.6.1 Normal-Based Rule, 278
 - 13.6.2 Covariance Matrix Equality, 279
 - 13.6.3 Rule Choice, 280
 - 13.6.4 Priors, 281
- 13.7 Comments, 281
 - Technical Notes, 283
 - Further Reading, 283
 - Exercises, 284

14 Classification Results

285

- 14.1 Introduction, 285
- 14.2 Research Context, 285
- 14.3 Computer Application, 286
- 14.4 Individual Unit Results, 287
 - 14.4.1 In-Doubt Units, 288
 - 14.4.2 Outliers, 289
- 14.5 Group Results, 290

14.6	Comments, 291	
	Technical Note, 291	
	Exercises, 292	
15	Hit Rate Estimation	295
15.1	Introduction, 295	
15.2	True Hit Rates, 296	
15.3	Hit Rate Estimators, 297	
	15.3.1 Formula Estimators, 297	
	15.3.2 Internal Analysis, 299	
	15.3.3 External Analysis, 300	
	15.3.4 Maximum-Posterior-Probability Method, 302	
15.4	Computer Application, 304	
15.5	Choice of Hit Rate Estimator, 306	
15.6	Outliers and In-Doubt Units, 306	
	15.6.1 Outliers, 307	
	15.6.2 In-Doubt Units, 307	
15.7	Sample Size, 309	
15.8	Comments, 310	
	Further Reading, 311	
	Exercises, 312	
16	Effectiveness of Classification Rules	315
16.1	Introduction, 315	
16.2	Proportional Chance Criterion, 316	
	16.2.1 Definition, 316	
	16.2.2 Statistical Test, 317	
16.3	Maximum-Chance Criterion, 319	
16.4	Improvement over Chance, 320	
16.5	Comparison of Rules, 320	
16.6	Computer Application I, 321	
16.7	Effect of Unequal Priors, 323	
16.8	PDA Validity/Reliability, 325	
16.9	Applying a Classification Rule to New Units, 325	
	16.9.1 Computer Application II, 326	
	16.9.2 Computer Application III, 327	
16.10	Comments, 330	
	Technical Notes, 330	
	Further Reading, 331	
	Exercises, 332	

17	Deleting and Ordering Predictors	335
17.1	Introduction, 335	
17.2	Predictor Deletion, 336	
17.2.1	Purposes of Deletion, 336	
17.2.2	Deletion Methods, 336	
17.2.3	Package Analyses, 337	
17.2.4	All Possible Subsets, 337	
17.3	Computer Application, 337	
17.4	Predictor Ordering, 340	
17.4.1	Meaning of Importance, 340	
17.4.2	Variable Ranking, 340	
17.5	Reanalysis, 343	
17.6	Comments, 343	
17.7	Side Note, 345	
	Further Reading, 346	
	Exercises, 347	
18	Two-Group Classification	349
18.1	Introduction, 349	
18.2	Two-Group Rule, 349	
18.3	Regression Analogy, 351	
18.4	MRA–PDA Relationship, 353	
18.5	Necessary Sample Size, 355	
18.6	Univariate Classification, 356	
	Further Reading, 357	
	Exercises, 359	
19	Nonnormal Rules	361
19.1	Introduction, 361	
19.2	Continuous Variables, 362	
19.2.1	Rank Transformation Analysis, 362	
19.2.2	Nearest-Neighbor Analyses, 363	
19.2.3	Another Density Estimation Analysis, 366	
19.2.4	Other Analyses, 366	
19.3	Categorical Variables, 366	
19.3.1	Direct Probability Estimation Analysis, 367	
19.3.2	Dummy Variable Analysis, 367	
19.3.3	Overall–Woodward Analysis, 368	
19.3.4	Fisher–Lancaster Analysis, 368	
19.3.5	Other Analyses, 369	
19.4	Predictor Mixtures, 369	

19.5	Comments, 370	
	Further Reading, 371	
	Exercises, 373	
20	Reporting PDA Results	375
20.1	Introduction, 375	
20.2	Example of Reporting PDA Results, 375	
20.3	Some Additional Specific PDA Information, 378	
20.4	Computer Package Information, 379	
20.5	Reporting Terms, 379	
20.6	Sources of PDA Applications, 381	
20.7	Concerns, 381	
20.8	Overview, 382	
	Further Reading, 383	
	Exercises, 383	
21	PDA-Related Analyses	385
21.1	Introduction, 385	
21.2	Nonlinear Methods, 385	
	21.2.1 Classification and Regression Trees (CART), 385	
	21.2.2 Logistic Regression, 385	
	21.2.3 Neural Networks, 386	
21.3	Other Methods, 386	
	21.3.1 Cluster Analysis, 386	
	21.3.2 Image Analysis, 387	
	21.3.3 Optimal Allocation, 387	
	21.3.4 Pattern Recognition, 387	
	Further Reading, 388	
V	ISSUES AND PROBLEMS	391
22	Issues in PDA and DDA	393
22.1	Introduction, 393	
22.2	Five Choices in PDA, 393	
	22.2.1 Linear Versus Quadratic Rules, 393	
	22.2.2 Nonnormal Classification Rules, 394	
	22.2.3 Prior Probabilities, 394	
	22.2.4 Misclassification Costs, 394	
	22.2.5 Hit-Rate Estimation, 395	
22.3	Stepwise Analyses, 395	
22.4	Standardized Weights Versus Structure r 's, 396	

22.5	Data-Based Structure, 398	
	Further Reading, 400	
23	Problems in PDA and DDA	401
23.1	Introduction, 401	
23.2	Missing Data, 401	
23.2.1	Data Inspection, 401	
23.2.2	Data Imputation, 402	
23.2.3	Missing <i>G</i> Values, 404	
23.2.4	Ad Hoc Strategy, 404	
23.3	Outliers and Influential Observations, 405	
23.3.1	Outlier Identification, 405	
23.3.2	Influential Observations, 406	
23.4	Initial Group Misclassification, 406	
23.5	Misclassification Costs, 407	
23.6	Statistical Versus Clinical Prediction, 407	
23.7	Other Problems, 409	
	Further Reading, 409	
Appendix A	Data Set Descriptions	411
Appendix B	Some DA-Related Originators	415
Appendix C	List of Computer Syntax	419
Appendix D	Contents of Wiley Website	421
	References	425
	Answers to Exercises	449
	Index	481