

INDEX

- Additional discrimination test, 24-26, 30
- Additive
- constant problem, 212, 226
 - interaction, 145
 - model, 200, 292, 293, 297, 300
- Allocation methods for cluster analysis, 324-329, 331, 332, 336
- ALSCAL, 242
- Alternating least squares, 242
- Anderson-Darling test, 63-64, 335, 342
- Analysis of residuals by AID, 298
- ANOVA - relation to AID (see AID, relation to ANOVA)
- Association - measures of, 155, 157, 159, 171-174, 283, 284, 293
- Assumption-free nature of AID, 269, 280, 297
- Asymptotic
- relative efficiency, 23, 51
 - variances and covariances of estimators, 81, 82, 90-94, 105, 109, 111, 114, 118, 154, 170
 - permutation distribution, 281
- Asymptotically distribution free (ADF) methods, 86, 87, 97, 99, 110, 125-130
- AID
- relationship to ANOVA, 270, 271, 278, 279, 283, 284
 - relation to cluster analysis, 276-279, 304
 - relationship to regression, 270, 271, 292, 293, 295, 297, 300
- Bahadur parametrization, 44
- Bartlett's test, 59-63
- Baseline model, 146, 155
- Basic
- structure, 186, 187
 - structure displays, 186-208
 - structure displays, stability of, 207, 208
 - structure, generalized, 186-189
- Bernoulli distribution, multivariate, 39
- Best generalized least squares (BGLS) estimates, 81, 83, 90, 91, 130
- Biased discriminant analysis, 16, 36
- Binary variables in discriminant analysis (see also Categorical variables; Bernoulli distribution), 18, 19, 40-42, 44, 46-48, 57
- Biplot (see also Symmetric biplot, Covariance biplot), 193-198
- Biweighted least squares, 89, 107
- Bonferroni inequality (see also Conservative inference), 7, 8, 28, 282-285, 337-342
- Bootstrap, 15
- Branch and bound, 29
- Canonical
- correlation, 194, 205
 - variables, 8, 9, 17, 346, 347, 351
 - variate analysis, 194, 205
- Categorical variables (see also Chapters 3 and 5), 47
- Categorized interval variables, 57, 280
- Chaining in single linkage cluster analysis, 310, 318, 331
- Change point detection, 282, 307, 345-350
- Chi-squared
- automatic interaction detection (CHAID), 280, 284, 287, 291-299
 - statistic (see also CHAID) 56, 143, 149, 163, 293, 335, 339, 340
- Choice of number of clusters in cluster analysis, 306, 335-340
- Cholesky square root, 107
- Chuprov measure of association, 293
- Classical scaling, 208-212, 245, 253
- Cluster analysis
- allocation methods, 52, 324-332, 336
 - comparisons of methods, 313, 314, 317, 318

Cambridge University Press

978-0-521-09070-4 - Topics in Applied Multivariate Analysis

Edited by Douglas M. Hawkins

Index

[More information](#)

358

- Cluster analysis
 - complete linkage, 318
 - criterion functions, 316, 326-329, 332, 333, 336-349, 345, 346
 - group average linkage, 313, 314, 318
 - hierarchical methods, 306-318
 - invariance under transformation, 314, 334, 339
 - Lance & Williams' flexible method, 317, 318
 - median linkage, 314, 315, 318
 - mixture analysis, 318-322, 331, 332
 - mode analysis, 311-314, 331
 - non-hierarchical methods, 306, 311, 318-331
 - relation to AID, 276-279, 304
 - single linkage, 309, 310, 313, 314, 318, 329, 331, 350
 - use in conjunction with Multidimensional scaling, 254, 255, 306
 - Ward's method, 313, 315-318, 326
- Cluster loops, 247, 255
- Collapsing of contingency tables 160-163
- Comparisons of clustering methods, 313, 314, 317, 318
- Complete linkage cluster analysis, 318
- Complete symmetry model, 112
- Computational procedures (see also Newton-Raphson algorithm), 102-110, 115, 167, 219, 223, 231, 233, 237, 242
- Confidence intervals for parameters, 95-97
- Conjugate gradient method, 103, 223
- Conservative inference - see also Multiple comparisons, Bonferroni inequality, 268-270, 273, 324
- Constrained discriminant analysis, 54, 55
- Constraints in the log-linear model, 147-151
- Contingency tables (see also Chapters 3 and 5), 206-208
- Continuity correction, 164
- Continuous variables in AID, 273, 299
- Convexity property of information, 160
- Cophenetic correlation coefficient, 257
- Correction for the likelihood ratio test statistic, 98, 99, 133
- Corrections for kurtosis, 92, 97, 98, 110, 126, 130
- Correlation structures, 76, 77, 93, 108, 127, 130-132
- Correspondence analysis, 205, 206, 211
- Covariance
 - biplot, 201, 202, 209
 - structures as regression models, 79, 80
 - structures definition, 73
 - structures - estimation of parameters, 78, 80-99
 - structures - invariance under changes of scale, 76, 93, 110, 130
 - structures - invariance under a constant scaling factor, 77, 92, 97, 98, 110, 125, 133
 - structures - linear, 75, 103, 112-114, 124-126
 - structures - nested, 118-120
 - structures - simplex, 114, 120-122, 132, 133
 - structures - time series, 120-122
- Cramer coefficient of contingency, 293
- Criteria for acceptable clustering methods, 310
- Criterion functions in cluster analysis, 316, 326-329, 332, 333, 336-340, 345, 346
- Cross-validation of covariance structures, 101
- Cumulative sum diagram (CSD), 229
- Cyclic predictor in AID, 276
- Decision theory approach to discriminant analysis, 3, 54
- Dendrogram, 270, 271, 273, 278, 285, 291-293, 295, 298, 334
- Departures from model (see also Robustness; Tests of fit), 54

Cambridge University Press

978-0-521-09070-4 - Topics in Applied Multivariate Analysis

Edited by Douglas M. Hawkins

Index

[More information](#)

359

- dichotomous variables in AID, 273,274
- Dimensionality, choice of (see also Eigenvalues - tests of) 212,214,246
- Direct product covariance structures, 76,122,123
- Discrepancy functions, 80,81
- Discriminant
 - analysis and cluster analysis, 325,332,334-336,342
 - function in cluster analysis, 329,333,339,344
- Distance
 - data, 208-261,305-315,331,334,344,349
 - scaling, 186
- Distortion
 - function, 213,214,223,226
 - representation of in scaling, 253
- Distributions of test statistics, 97-99,108,110,114,163-167,281-284,334-340
- Double-centering, 210,211
- Dynamic programming, 50,277,278,347,351
- Eigenstructure, 189,208,339,346
- Eigenvalues - tests of, 207
- Elastic distances multidimensional scaling, 240
- Elbow test, 246,336,337
- Elementary log-linear models, 180
- Elliptical distributions, 85,86,92,97,98,110,125
- Empiric distribution function, 56
- EM algorithm, 52,320-322,329,333,335,346
- Enclosure ratio in cluster analysis, 312,313
- Equal mean discriminant analysis, 55
- Error rate, 7,11-15,24,26,30,43
- Estimation subject to constraints or identifiability conditions, 89,92,93,97,99,108,109,147,148
- Estimative discriminant analysis, 4,10
- Euclidean distance - in cluster analysis, 305,312,314-316,327,352
- Exponential family of distributions, 142,151,152
- Extended automatic interaction detector (XAID), 278-283,286,287,293,295,298,299
- External constraints problem (ECP), 149
- Factor analysis, 74,113-118
 - interbattery, 118
 - restricted, 76,77,118,130-132
 - three mode, 123
 - unrestricted model, 78,94,114-118,127-130
- Factors - Indeterminacy of, 79
- Fisher's method of combining tests, 59
- Fisher scoring algorithm, 104
- Floating predictors in AID, 274,275,282,286
- Free variables in AID, 274-276,281-285
- Furthest Neighbour cluster analysis, 317,318
- General linear model, 90,94,152-154
 - connection to AID, 279,280,295,299
- Generalized
 - automatic interaction detector (GAID), 279,280
 - least squares (GLS), 81,91,110,125-130
- GLIM program, 167,171,324
- Global optimization method (see Dynamic programming)
- Goodness of fit (see also Tests of fit), 189
- Graphical display (see also Chapter 4), 9,171-173,180
- Greatest convex minorant (GCM), 229
- Group average linkage, 313,314,318
- Growth curves, 104
- Hierarchic
 - cluster analysis, 306-318
 - grouping in AID, 277
 - methods in choice of number of clusters, 340
 - methods in segmentation, 348-350
 - model sequences, 154-158
- Homogeneity, marginal, 148,149
- Homoscedasticity, 62,278-280,284,292,294,328,331-336,340-346
- Hotellings T^2 , 351
- Huberized discriminant analysis, 20,58

Cambridge University Press

978-0-521-09070-4 - Topics in Applied Multivariate Analysis

Edited by Douglas M. Hawkins

Index

[More information](#)

360

- Identification keys, 273
- Identified parameters, 78,100
- Independence
 - conditional, 145
 - model, 153
 - mutual, 150,155
 - pairwise, 151
- Information
 - explained criterion, 155
 - statistic, 37
- Informative prior distribution, 35,36
- Initializing distribution, 149
- Interleaved predictor in AID, 275
- Internal constraints problem (ICP), 149,150
- Isotonic regression, 228-234
- Jackknife, 1,15,18,26,47
- Johnson system of distributions, 22
- Kernel function method, 39, 45-48
- Kruskal-Shepard scaling, 234-242, 253
- Kruskal-Wallis statistic, 279,299
- Kurtosis, 18,57,58,61,82, 110
 - assumption of no, 83, 91,92,98,100,105,108, 115,118,126,133
 - relative, 85,86,97, 124-126
- Lance & Williams' flexible clustering method, 317, 318
- Latent
 - class analysis, 323, 324
 - profile analysis, 322,324,328
 - variables, 74,79,119
- Least squares scaling, 212, 219-228
- Likelihood ratio test, 25, 27,30,154,164
- Linear
 - discriminant function (LDF) 1,6,10,17,19,41, 44,48,49,50,54
 - structural relations model, 119,120
- Linking function, 152
- Local
 - minima, 245,261
 - order nonmetric scaling, 243-245
- Location model (see also Reversal) 42,43,57
- Logistic regression, 39,168-171
- Logit analysis, 142
- Log-linear model
 - (see also Chapter 3), 39, 43,57
 - in latent class analysis, 324
- Log-odds ratio, 142,168
- Loss function, 315-317
- Magnitude criterion, 213,214,219, 247
- Mahalanobis distance, 6,30,54, 201,305,314,335,351
- Marginal
 - association, 155,157,171-174
 - homogeneity, 148,149
- Marginality of tests, 159
- Matching coefficients, 305
- Matrix
 - approximation, 186-189
 - norm, 187,188
- Maximal predictive classification, 307,331
- Maximum Wishart likelihood (MWL) estimation, 84,94,98,99, 110,111,125-132
- Mean-square contingency coefficient (MSCC), 293-295
- Median linkage, 314,315,318
- Minimal adequate subset, 29,158
- Minimum
 - discrimination information (MDI) statistic, 146-151, 154,266
 - spanning tree, 254,309-311
- MINRES estimation, 88,91,99,110
- Misclassification - see Error rate,
- Missing
 - data, 53,88,208,224,246
 - information in AID, 274,275, 280,292
- Mixture analysis, 51,318-322,331, 332
- Mode analysis, 311-314,331
- Model
 - marginal association, 172
 - partial association, 172,173
- Modelling using AID, 287-297
- Models - baseline, 155,156

Cambridge University Press

978-0-521-09070-4 - Topics in Applied Multivariate Analysis

Edited by Douglas M. Hawkins

Index

[More information](#)

361

Models

- elementary, 180
- generalized linear, 152-154
- hierarchic, 154,157,158, 180
- independence, 171
- minimal adequate, 158
- saturated, 145,153,158, 168,170,171

Monotonically increasing polynomials, 225,226

Monotonic

- predictor in AID, 273-275, 281-283,285,288
- regression, 228-235,239, 241-243

Multidimensional unfolding, 253, 259-261

Multinomial distribution, 43, 153,164

Multiple

- comparisons, 157,283, 284,286,287
- regression, 11,53,56, 79,80,169

Multiplicative interaction, 144

Manova - connection to AID, 279

Multivariate automatic interaction detector (MAID), 279

Nearest neighbour cluster analysis see Single linkage

Newton-Raphson algorithm (see also Quasi-Newton algorithms) 40,104,153

Non-hierarchic clustering methods, 306,311,318-331

Nonlinear mapping, 223-226

Non-metric distances in cluster analysis, 315

Ordinal (or nominal) variables in AID, 269,271,273,274,280, 299

Outlier, 58,63,88,160,161,254, 299,300

Parallel batteries, 113

Parametric mapping, 225

Parsimony - principle of, 156,158 270,293

Partial association, 155,157,159, 171-174

Partially constrained predictors in AID, 275,276

Partitioning of test statistics, 150

Patterned

- correlation matrices, 73, 77,114
- covariance matrices, 73, 75,112,113,124-126

Pearson coefficient of contingency, 293

Percentage of variance - See Eigenvalues - tests of

Pool-adjacent-violators algorithm, 231

Posterior

- distribution, 4,32-37
- probability, 3,32,37,39

Predictive discriminant analysis, 4,31-39

Principal

- components analysis, 191, 192,194-201,209,211,327, 339,346
- components of standardized data, 198
- components - use in cluster analysis, 328,330,339, 346,351
- components with weights, 200,201
- co-ordinates analysis, 210,211
- plane, 192

Prior

- distribution, 4,32-37
- probabilities, 3,12,42

Procrustes statistic, 257,258

Profile data, 184-208,305-308,316, 318,331,334-336,344

Proportional covariance matrices, 33,34

Q-mode factor analysis, 330,331

Quadratic logistic discriminant function, 42,45

Quadratic discriminant function (QDF), 6,10,49,54,55

Quasi-Newton algorithms, 103

Random effects model, 3,16,34,35

Ranks, 21,48,49

Rates and Proportions, 167-171

Reallocation methods (see Allocation methods)

Regression

- models for ordinal data, 146,171,269
- relation to AID (see AID, relation to regression)

Cambridge University Press

978-0-521-09070-4 - Topics in Applied Multivariate Analysis

Edited by Douglas M. Hawkins

Index

[More information](#)

362

- Replicated data, 246
- Rescaling estimates of parameters, 93,94,111,128
- Reversal, 19,44,57
- Ridging, 16
- Robustness, 15,17-21,58,111,352
 - of classical scaling, 212-214,219,224
- Rotation of the factor matrix, 78,116,117
- Sample
 - re-use, 15
 - size requirements of AID, 281,297,298
- Scalar products, 188,192,193,197,207,209,210
- Scaling and cluster analysis (see Cluster analysis)
- Segmentation, 307,344-350
 - problem and AID, 276, 282
- Selection of variables, 2, 24-30,159
- Sequential discriminant analysis, 49-51
- Significance of principal axes (see Eigenvalues - tests of)
- Sign test, 26
- Simplex structure, 114,120-122
- Simpson's paradox, 161,162
- Simulation, 14,44,48,53,286,287,340-344
 - studies in scaling, 240,245,255,256
- Simultaneous test procedure (STP), 29,158
- Single linkage cluster analysis, 309,310,313,314,318,329,331,350
- Singular value decomposition, 186
- Skewness, 18,57,58,61
- Smoothing parameter, 45-47
- Spherical
 - Cauchy distribution, 46
 - normal distribution, 46
- Split moving window method of segmentation, 347,348,351
- Stability of results (see also Robustness), 207,252
- Starting configurations for scaling, 219,245,246
- Steepest descent algorithm, 237,242,245
- Stepwise
 - discriminant analysis, 25,27-30
 - procedures, 60-62,157,277,278,349
- Stress, 234,239-242,244,306
- Symmetric biplot, 198-200
- Tests of fit, 53,55-64,97-100,156,173,207,333-335,340,344
- Theta automatic interaction detection (THAID), 280
- Ties, in scaling, 234,237,239
- Time series
 - and cluster analysis, 307,344-352
 - covariance structures (see Covariance structures - time series)
- Tolerance region, 37
- Trace criterion (see also Cluster analysis - criterion functions), 212,214,219,247,278
- Transformation
 - of data, 169,170,328,331
 - of variables, 1,18
- Unclassified observations, 51,55
- Unsampled source populations, 6,53,54
- Up-and-down-blocks algorithm, 233
- Vector scaling, (see Profile data)
- Ward's method, 313,315-318,326
- Weighted data, 200,201,224,242,243
- Wilcoxon-Mann-Whitney test, 279
- Young-Householder scaling, 210