PERMUTATION TESTS FOR COMPLEX DATA
Theory, Applications and Software

Fortunato Pesarin • Luigi Salmaso
University of Padua, Italy
Contents

Preface xv

Notation and Abbreviations xix

1 Introduction 1

1.1 On Permutation Analysis 1

1.2 The Permutation Testing Principle 4

1.2.1 Nonparametric Family of Distributions 4

1.2.2 The Permutation Testing Principle 5

1.3 Permutation Approaches 7

1.4 When and Why Conditioning is Appropriate 7

1.5 Randomization and Permutation 9

1.6 Computational Aspects 10

1.7 Basic Notation 11

1.8 A Problem with Paired Observations 13

1.8.1 Modelling Responses 13

1.8.2 Symmetry Induced by Exchangeability 15

1.8.3 Further Aspects 15

1.8.4 The Student's t-Paired Solution 16

1.8.5 The Signed Rank Test Solution 17

1.8.6 The McNemar Solution 18

1.9 The Permutation Solution 18

1.9.1 General Aspects 18

1.9.2 The Permutation Sample Space 19

1.9.3 The Conditional Monte Carlo Method 20

1.9.4 Approximating the Permutation Distribution 22

1.9.5 Problems and Exercises 23

1.10 A Two-Sample Problem 23

1.10.1 Modelling Responses 24

1.10.2 The Student t Solution 25

1.10.3 The Permutation Solution 25

1.10.4 Rank Solutions 28

1.10.5 Problems and Exercises 28

1.11 One-Way ANOVA 29

1.11.1 Modelling Responses 29

1.11.2 Permutation Solutions 30

1.11.3 Problems and Exercises 32
4 The Nonparametric Combination Methodology
4.1 Introduction
4.1.1 General Aspects
4.1.2 Bibliographic Notes
4.1.3 Main Assumptions and Notation
4.1.4 Some Comments
4.2 The Nonparametric Combination Methodology
4.2.1 Assumptions on Partial Tests
4.2.2 Desirable Properties of Combining Functions
4.2.3 A Two-Phase Algorithm for Nonparametric Combination
4.2.4 Some Useful Combining Functions
4.2.5 Why Combination is Nonparametric
4.2.6 On Admissible Combining Functions
4.2.7 Problems and Exercises
4.3 Consistency, Unbiasedness and Power of Combined Tests
4.3.1 Consistency
4.3.2 Unbiasedness
4.3.3 A Non-consistent Combining Function
4.3.4 Power of Combined Tests
4.3.5 Conditional Multivariate Confidence Region for $\delta$
4.3.6 Problems and Exercises
4.4 Some Further Asymptotic Properties
4.4.1 General Conditions
4.4.2 Asymptotic Properties
4.5 Finite-Sample Consistency
4.5.1 Introduction
4.5.2 Finite-Sample Consistency
4.5.3 Some Applications of Finite-Sample Consistency
4.6 Some Examples of Nonparametric Combination
4.6.1 Problems and Exercises
4.7 Comments on the Nonparametric Combination
4.7.1 General Comments
4.7.2 Final Remarks
5 Multiplicity Control and Closed Testing
5.1 Defining Raw and Adjusted $p$-Values
5.2 Controlling for Multiplicity
5.2.1 Multiple Comparison and Multiple Testing
5.2.2 Some Definitions of the Global Type I Error
5.3 Multiple Testing
5.4 The Closed Testing Approach
5.4.1 Closed Testing for Multiple Testing
5.4.2 Closed Testing Using the MinP Bonferroni–Holm Procedure
5.5 Mult Data Example
5.5.1 Analysis Using MATLAB
5.5.2 Analysis Using R
5.6 Washing Test Data
5.6.1 Analysis Using MATLAB
5.6.2 Analysis Using R
5.7 Weighted Methods for Controlling FWE and FDR
Contents

7.6 General Aspects of Permutation Testing with Missing Data 232
7.6.1 Bibliographic Notes 232

7.7 On Missing Data Processes 233
7.7.1 Data Missing Completely at Random 233
7.7.2 Data Missing Not at Random 234

7.8 The Permutation Approach 234
7.8.1 Deletion, Imputation and Intention to Treat Strategies 235
7.8.2 Breaking Down the Hypotheses 236

7.9 The Structure of Testing Problems 237
7.9.1 Hypotheses for MNAR Models 237
7.9.2 Hypotheses for MCAR Models 238
7.9.3 Permutation Structure with Missing Values 239

7.10 Permutation Analysis of Missing Values 240
7.10.1 Partitioning the Permutation Sample Space 240
7.10.2 Solution for Two-Sample MCAR Problems 241
7.10.3 Extensions to Multivariate C-Sample Problems 242
7.10.4 Extension to MNAR Models 243

7.11 Germina Data: An Example of an MNAR Model 244
7.11.1 Problem Description 245
7.11.2 The Permutation Solution 245
7.11.3 Analysis Using MATLAB 248
7.11.4 Analysis Using R 248

7.12 Multivariate Paired Observations 251

7.13 Repeated Measures and Missing Data 252
7.13.1 An Example 253

7.14 Botulinum Data 254
7.14.1 Analysis Using MATLAB 256
7.14.2 Analysis Using R 258

7.15 Waterfalls Data 260
7.15.1 Analysis Using MATLAB 260
7.15.2 Analysis Using R 264

8 Some Stochastic Ordering Problems 267
8.1 Multivariate Ordered Alternatives 267
8.2 Testing for Umbrella Alternatives 269
8.2.1 Hypotheses and Tests in Simple Stochastic Ordering 270
8.2.2 Permutation Tests for Umbrella Alternatives 271
8.3 Analysis of Experimental Tumour Growth Curves 273
8.4 Analysis of PERC Data 276
8.4.1 Introduction 276
8.4.2 A Permutation Solution 278
8.4.3 Analysis Using MATLAB 279
8.4.4 Analysis Using R 286

9 NPC Tests for Survival Analysis 289
9.1 Introduction and Main Notation 289
9.1.1 Failure Time Distributions 289
9.1.2 Data Structure 290
9.2 Comparison of Survival Curves 291
9.3 An Overview of the Literature 292
9.3.1 Permutation Tests in Survival Analysis 294
9.4 Two NPC Tests 295
  9.4.1 Breaking Down the Hypotheses 295
  9.4.2 The Test Structure 296
  9.4.3 NPC Test for Treatment-Independent Censoring 297
  9.4.4 NPC Test for Treatment-Dependent Censoring 298
9.5 An Application to a Biomedical Study 300

10 NPC Tests in Shape Analysis 303
  10.1 Introduction 303
  10.2 A Brief Overview of Statistical Shape Analysis
      10.2.1 How to Describe Shapes 304
      10.2.2 Multivariate Morphometrics 306
  10.3 Inference with Shape Data 308
  10.4 NPC Approach to Shape Analysis
      10.4.1 Notation 309
      10.4.2 Comparative Simulation Study 311
  10.5 NPC Analysis with Correlated Landmarks 312
  10.6 An Application to Mediterranean Monk Seal Skulls
      10.6.1 The Case Study 316
      10.6.2 Some Remarks 319
      10.6.3 Shape Analysis Using MATLAB 319
      10.6.4 Shape Analysis Using R 321

11 Multivariate Correlation Analysis and Two-Way ANOVA 325
  11.1 Autofluorescence Case Study
      11.1.1 A Permutation Solution 326
      11.1.2 Analysis Using MATLAB 329
      11.1.3 Analysis Using R 329
  11.2 Confocal Case Study
      11.2.1 A Permutation Solution 333
      11.2.2 MATLAB and R Codes 335
  11.3 Two-Way (M)ANOVA
      11.3.1 Brief Overview of Permutation Tests in Two-Way ANOVA 344
      11.3.2 MANOVA Using MATLAB and R Codes 346

12 Some Case Studies Using NPC Test R10 and SAS Macros 351
  12.1 An Integrated Approach to Survival Analysis in Observational Studies
      12.1.1 A Case Study on Oesophageal Cancer 351
      12.1.2 A Permutation Solution 353
      12.1.3 Survival Analysis with Stratification by Propensity Score 353
  12.2 Integrating Propensity Score and NPC Testing
      12.2.1 Analysis Using MATLAB 358
  12.3 Further Applications with NPC Test R10 and SAS Macros
      12.3.1 A Two-Sample Epidemiological Survey: Problem Description 359
      12.3.2 Analysing SETIG Data Using MATLAB 360
      12.3.3 Analysing the SETIG Data Using R 362
      12.3.4 Analysing the SETIG Data Using NPC Test 365
      12.3.5 Analysis of the SETIG Data Using SAS 369
  12.4 A Comparison of Three Survival Curves 370
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.4.1</td>
<td>Unstratified Survival Analysis</td>
<td>371</td>
</tr>
<tr>
<td>12.4.2</td>
<td>Survival Analysis with Stratification by Propensity Score</td>
<td>371</td>
</tr>
<tr>
<td>12.5</td>
<td>Survival Analysis Using NPC Test and SAS</td>
<td>375</td>
</tr>
<tr>
<td>12.5.1</td>
<td>Survival Analysis Using NPC Test</td>
<td>375</td>
</tr>
<tr>
<td>12.5.2</td>
<td>Survival Analysis Using SAS</td>
<td>377</td>
</tr>
<tr>
<td>12.5.3</td>
<td>Survival Analysis Using MATLAB</td>
<td>378</td>
</tr>
<tr>
<td>12.6</td>
<td>Logistic Regression and NPC Test for Multivariate Analysis</td>
<td>378</td>
</tr>
<tr>
<td>12.6.1</td>
<td>Application to Lymph Node Metastases</td>
<td>378</td>
</tr>
<tr>
<td>12.6.2</td>
<td>Application to Bladder Cancer</td>
<td>380</td>
</tr>
<tr>
<td>12.6.3</td>
<td>NPC Results</td>
<td>382</td>
</tr>
<tr>
<td>12.6.4</td>
<td>Analysis by Logistic Regression</td>
<td>384</td>
</tr>
<tr>
<td>12.6.5</td>
<td>Some Comments</td>
<td>385</td>
</tr>
</tbody>
</table>

References 387

Index 409