

Springer Series in Statistics

Advisors:

P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg,

I. Olkin, N. Wermuth, S. Zeger

Springer Science+Business Media, LLC

Springer Series in Statistics

- Andersen/Borgan/Gill/Keiding*: Statistical Models Based on Counting Processes.
- Berger*: Statistical Decision Theory and Bayesian Analysis, 2nd edition.
- Bolfarine/Zacks*: Prediction Theory for Finite Populations.
- Borg/Groenen*: Modern Multidimensional Scaling: Theory and Applications
- Brockwell/Davis*: Time Series: Theory and Methods, 2nd edition.
- Chen/Shao/Ibrahim*: Monte Carlo Methods in Bayesian Computation.
- Efromovich*: Nonparametric Curve Estimation: Methods, Theory, and Applications.
- Fahrmeir/Tutz*: Multivariate Statistical Modelling Based on Generalized Linear Models.
- Farebrother*: Fitting Linear Relationships: A History of the Calculus of Observations 1750-1900.
- Federer*: Statistical Design and Analysis for Intercropping Experiments, Volume I: Two Crops.
- Federer*: Statistical Design and Analysis for Intercropping Experiments, Volume II: Three or More Crops.
- Fienberg/Hoaglin/Kruskal/Tanur (Eds.)*: A Statistical Model: Frederick Mosteller's Contributions to Statistics, Science and Public Policy.
- Fisher/Sen*: The Collected Works of Wassily Hoeffding.
- Good*: Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses, 2nd edition.
- Gouriéroux*: ARCH Models and Financial Applications.
- Grandell*: Aspects of Risk Theory.
- Haberman*: Advanced Statistics, Volume I: Description of Populations.
- Hall*: The Bootstrap and Edgeworth Expansion.
- Härdle*: Smoothing Techniques: With Implementation in S.
- Hart*: Nonparametric Smoothing and Lack-of-Fit Tests.
- Hartigan*: Bayes Theory.
- Hedayat/Sloane/Stufken*: Orthogonal Arrays: Theory and Applications.
- Heyde*: Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation.
- Huet/Bouvier/Gruet/Jolivet*: Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS Examples.
- Kolen/Brennan*: Test Equating: Methods and Practices.
- Kotz/Johnson (Eds.)*: Breakthroughs in Statistics Volume I.
- Kotz/Johnson (Eds.)*: Breakthroughs in Statistics Volume II.
- Kotz/Johnson (Eds.)*: Breakthroughs in Statistics Volume III.
- Küchler/Sørensen*: Exponential Families of Stochastic Processes.
- Le Cam*: Asymptotic Methods in Statistical Decision Theory.
- Le Cam/Yang*: Asymptotics in Statistics: Some Basic Concepts.
- Longford*: Models for Uncertainty in Educational Testing.
- Miller, Jr.*: Simultaneous Statistical Inference, 2nd edition.
- Mosteller/Wallace*: Applied Bayesian and Classical Inference: The Case of the Federalist Papers.
- Parzen/Tanabe/Kitagawa*: Selected Papers of Hirotugu Akaike.
- Politis/Romano/Wolf*: Subsampling.

(continued after index)

Phillip Good

Permutation Tests

A Practical Guide to Resampling Methods
for Testing Hypotheses

Second Edition

With 14 Figures



Springer

Phillip Good
205 W. Utica Avenue
Huntington Beach, CA 92648
USA
brother_unknown@yahoo.com

Library of Congress Cataloging-in-Publication Data
Good, Phillip I.

Permutation tests : a practical guide to resampling methods for
testing hypotheses / Phillip Good.—2nd ed.

p. cm.—(Springer series in statistics)

Includes bibliographical references and index.

ISBN 978-1-4757-3237-5 ISBN 978-1-4757-3235-1 (eBook)

DOI 10.1007/978-1-4757-3235-1

1. Statistical hypothesis testing. 2. Resampling (Statistics)

I. Title. II. Series.

QA277.G643 2000

519.5'6—dc21

99-16557

Printed on acid-free paper.

© 2000, 1994 Springer Science+Business Media New York

Originally published by Springer-Verlag New York, Inc. in 2000

Softcover reprint of the hardcover 2nd edition 2000

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher Springer Science+Business Media, LLC, except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Timothy Taylor; manufacturing supervised by Jerome Basma.
Typeset by TechBooks, Fairfax, VA.

9 8 7 6 5 4 3 2 1

ISBN 978-1-4757-3237-5

SPIN 10735885

Preface to the Second Edition

In 1982, I published several issues of a *samdizat* scholarly journal called *Randomization* with the aid of an 8-bit, 1-MH personal computer with 48K of memory (upgraded to 64K later that year) and floppy disks that held 400 Kbytes. A decade later, working on the first edition of this text, I used a 16-bit, 33-MH computer with 1 Mb of memory and a 20-Mb hard disk. This preface to the second edition comes to you via a 32-bit, 300-MH computer with 64-Mb memory and a 4-Gb hard disk. And, yes, I paid a tenth of what I paid for my first computer.

This relationship between low-cost readily available computing power and the rising popularity of permutation tests is no coincidence. Simply put, it is faster today to compute an exact p-value than to look up an approximation in a table of the not-quite-appropriate statistic. As a result, more and more researchers are using Permutation Tests to analyze their data.

Of course, some of the increased usage has also come about through the increased availability of and improvements in off-the-shelf software, as can be seen in the revisions in this edition to Chapter 12 (Publishing Your Results) and Chapter 13 (Increasing Computation Efficiency).

These improvements helped persuade me it was the time to publish a first course in statistics based entirely on resampling methods (an idea first proposed by the late F.N. David). As a result, *Permutation Tests* has become two texts: one, *Resampling Methods*, designed as a first course, and this second edition aimed at upper division graduate students and practitioners who may already be familiar with the application of other statistical procedures. The popular question section at the end of each chapter now contains a number of thesis-level questions, which may or may not be solvable in their present form. While the wide applicability of permutation tests continues to be emphasized here, their limitations are also revealed. Examples include expanded sections on comparing variances (Chapter 3, Testing Hypotheses), testing interactions in balanced designs (Chapter 4, Experimental Design), and multiple regression (Chapter 7, Dependence).

Sections on Sequential Analysis (Chapter 4) and comparing spatial distributions (Chapter 8) are also new. Recent major advances in the analysis of multiple dependent tests are recorded in Chapter 5 on Multivariate Analysis.

My thanks to the many individuals who previewed chapters for this edition, including, in alphabetical order, Brian Cade, Mike Ernst, Barbara Heller, John Kimmel, Patrick Onghena, Fortunato Pesarin, and John Thaden.

Phillip Good
Huntington Beach, California

Preface to the First Edition

Permutation tests permit us to choose the test statistic best suited to the task at hand. This freedom of choice opens up a thousand practical applications, including many which are beyond the reach of conventional parametric statistics. Flexible, robust in the face of missing data and violations of assumptions, the permutation test is among the most powerful of statistical procedures. Through sample size reduction, permutation tests can reduce the costs of experiments and surveys.

This text on the application of permutation tests in biology, medicine, science, and engineering may be used as a step-by-step self-guiding reference manual by research workers and as an intermediate text for undergraduates and graduates in statistics and the applied sciences with a first course in statistics and probability under their belts.

Research workers in the applied sciences are advised to read through Chapters 1 and 2 once quickly before proceeding to Chapters 3 through 8 which cover the principal applications they are likely to encounter in practice.

Chapter 9 is a must for the practitioner, with advice for coping with real-life emergencies such as missing or censored data, after-the-fact covariates, and outliers.

Chapter 10 uses practical applications in archeology, biology, climatology, education and social science to show the research worker how to develop new permutation statistics to meet the needs of specific applications. The practitioner will find Chapter 10 a source of inspiration as well as a practical guide to the development of new and novel statistics.

The expert system in Chapter 11 will guide you to the correct statistic for your application. Chapter 12, more “must” reading, provides practical advice on experimental design and shows how to document the results of permutation tests for publication.

Chapter 13 describes techniques for reducing computation time; and a guide to off-the-shelf statistical software is provided in an appendix.

The sequence of recommended readings is somewhat different for the student and will depend on whether he or she is studying the permutation tests by

themselves or as part of a larger course on resampling methods encompassing both the permutation test and the bootstrap resampling method.

This book can replace a senior-level text on testing hypotheses. I have also found it of value in introducing students who are primarily mathematicians to the applications which make statistics a unique mathematical science. Chapters 1, 2, and 14 provide a comprehensive introduction to the theory. Despite its placement in the latter part of the text, Chapter 14, on the theory of permutation tests, is self-standing. Chapter 3 on applications also deserves a careful reading. Here in detail are the basic testing situations and the basic tests to be applied to them. Chapters 4, 5, and 6 may be used to supplement Chapter 3, time permitting (the first part of Chapter 6 describing the Fisher exact test is a must). Rather than skipping from section to section, it might be best for the student to consider one of these latter chapters in depth—supplementing his or her study with original research articles.

My own preference is to parallel discussions of permutation methods with discussion of a second resampling method, the bootstrap. Again, Chapters 1, 2, and 3—supplemented with portions of Chapter 14—are musts. Chapter 7, on tests of dependence, is a natural sequel. Students in statistical computing also are asked to program and test at least one of the advanced algorithms in Chapter 12.

For the reader's convenience, the bibliography is divided into four parts: the first consists of 34 seminal articles; the second of two dozen background articles referred to in the text that are not directly concerned with permutation methods; the third of 111 articles on increasing computational efficiency; and a fourth, principal bibliography of 574 articles and books on the theory and application of permutation techniques.

Exercises are included at the end of each chapter to enhance and reinforce your understanding. But the best exercise of all is to substitute your own data for the examples in the text.

My thanks to Symantek, TSSI, and Perceptronics without whose Grand-View® outliner, Exact® equation generator, and Einstein Writer® word processor this text would not have been possible.

I am deeply indebted to Mike Chernick for our frequent conversations and his many invaluable insights, to Mike Ernst, Alan Forsythe, Karim Hiriji, John Ludbrook, Reza Modarres, and William Schucany for reading and commenting on portions of this compuscript and to my instructors at Berkeley including E. Fix, J. Hodges, E. Lehmann, and J. Neyman.

Phillip Good
Huntington Beach, California

Contents

Preface to the Second Edition	v
Preface to the First Edition	vii
1. A Wide Range of Applications	1
1.1. Permutation Tests	1
1.1.1. Applications	1
1.2. “I Lost the Labels”	4
1.3. Five Steps to a Permutation Test	6
1.3.1. Analyze the Problem	6
1.3.2. Choose a Test Statistic	6
1.3.3. Compute the Test Statistic	7
1.3.4. Rearrange the Observations	7
1.3.5. Make a Decision	8
1.4. What’s in a Name?	8
1.4.1. Comparison with Other Tests	9
1.4.2. Sampling from the Data at Hand	9
1.5. History	10
1.6. To Learn More	12
1.7. Questions	12
2. A Simple Test	13
2.1. Properties of the Test	13
2.2. Fundamental Concepts	14
2.2.1. Population and Sample Distributions	14
2.2.2. Two Types of Error	16
2.2.2.1. Losses and Risk	18
2.2.3. Significance Level and Power	19
2.2.4. Power and Sample Size	20
2.2.5. Power and the Alternative	22

2.2.6. Exact, Unbiased Tests	23
2.2.7. Exchangeable Observations	24
2.3. Which Test?	25
2.4. World Views	27
2.4.1. Generalized Permutations	28
2.5. Questions	29
3. Testing Hypotheses	31
3.1. One-Sample Tests	31
3.1.1. Tests for a Location Parameter	31
3.1.2. Properties of the Test	33
3.1.3. Exact Significance Levels: A Digression	33
3.2. Confidence Intervals	34
3.2.1. Comparison with Other Tests	35
3.3. Comparison of Locations	36
3.3.1. An Example	36
3.3.2. Violation of Assumptions	37
3.4. Comparing Dispersions	38
3.4.1. The Permutation Approach	38
3.4.2. The Bootstrap Approach	42
3.5. k -Sample Comparisons	42
3.5.1. Comparing Location Parameters	42
3.5.2. Pitman Correlation	45
3.5.3. Effect of Ties	47
3.5.4. Cochran–Armitage Test	48
3.5.5. Linear Estimation	49
3.5.6. A Unifying Theory	49
3.6. Blocking	49
3.6.1. Extending the Range of Applications	51
3.7. Matched Pairs	51
3.8. Questions	52
4. Experimental Designs	54
4.1. Introduction	54
4.2. Balanced Designs	54
4.2.1. Main Effects	55
4.2.2. An Example	57
4.2.3. Testing for Interactions	59
4.2.3.1. Blocking	60
4.2.3.2. Synchronized Permutations	61
4.2.3.3. To Learn More	62
4.2.4. Designing an Experiment	62
4.2.5. Latin Squares	63
4.2.6. Other Designs	65

- 4.3. Analysis of Covariance 65
 - 4.3.1. Covariates Defined 65
 - 4.3.2. Eliminate the Functional Relationship 65
 - 4.3.3. Selecting Variables 66
 - 4.3.4. Restricted Randomization 66
- 4.4. Unbalanced Designs 67
 - 4.4.1. Missing Combinations 68
 - 4.4.2. The Boot-Perm Test 70
- 4.5. Clinical Trials 70
 - 4.5.1. Avoiding Unbalanced Designs 70
 - 4.5.2. Missing Data 71
- 4.6. Sequential Analysis 72
- 4.7. Very Large and Very Small Samples 73
- 4.8. Questions 74

- 5. Multivariate Analysis 76
 - 5.1. Introduction 76
 - 5.2. One- and Two-Sample Comparisons 76
 - 5.2.1. Hotelling's T^2 76
 - 5.2.2. An Example 79
 - 5.2.3. Doing the Computations 80
 - 5.2.4. Weighting the Variables 81
 - 5.2.5. Interpreting the Results 82
 - 5.2.6. Alternative Statistics 82
 - 5.3. Runs Test 83
 - 5.3.1. Which Statistic? 84
 - 5.4. Experimental Designs 86
 - 5.4.1. Matched Pairs 86
 - 5.4.2. Block Effects 86
 - 5.5. Multiple Comparisons 87
 - 5.5.1. Step Up or Step Down? 87
 - 5.5.1.1. Standardized Statistics 87
 - 5.5.1.2. Paired Sample Tests 88
 - 5.5.2. Fisher's Omnibus Statistic 89
 - 5.6. Repeated Measures 90
 - 5.6.1. Missing Data 91
 - 5.6.2. Bioequivalence 91
 - 5.6.3. Omnibus Test 92
 - 5.7. Questions 92

- 6. Categorical Data 94
 - 6.1. Fisher's Exact Test 94
 - 6.1.1. One-Tailed and Two-Tailed Tests 96
 - 6.1.2. The Two-Tailed Test 97

6.1.3. Determining the p -Value	97
6.1.4. What Is the Alternative?	98
6.1.5. Increasing the Power	99
6.2. Odds Ratio	99
6.2.1. Stratified 2×2 's	100
6.3. Exact Significance Levels	102
6.4. Unordered $r \times c$ Contingency Tables	103
6.4.1. Agreement Between Observers	106
6.4.2. What Should We Randomize?	106
6.4.3. Underlying Assumptions	107
6.5. Ordered Contingency Tables	109
6.5.1. Ordered $2 \times c$ Tables	109
6.5.2. More than Two Rows and Two Columns	110
6.5.2.1. Singly Ordered Tables	110
6.5.2.2. Doubly Ordered Tables	111
6.5.2.3. An Example	112
6.6. Covariates	113
6.7. Higher Dimensional Tables	114
6.8. To Learn More	115
6.9. Questions	115
7. Dependence	118
7.1. The Models	118
7.2. Testing for Independence	119
7.2.1. Independence	119
7.2.2. Quadrant Dependence	119
7.3. Testing for Trend	120
7.4. Serial Correlation	121
7.4.1. An Example	122
7.4.2. Trend	124
7.4.3. First-Order Dependence	124
7.4.4. An Economic Model	125
7.5. Known Models	126
7.5.1. Testing a Specific Hypothesis	126
7.5.2. Testing a General Hypothesis	126
7.5.3. Confidence Intervals	127
7.6. Multiple Regression	128
7.6.1. Eliminating Covariate Effects	128
7.6.1.1. Permute the Raw Data	129
7.6.1.2. Permute Residuals Under the Full Model	129
7.6.1.3. Permute Residuals Under the Reduced Model	129
7.6.2. LAD or LSD?	130
7.6.3. An Exact Solution	130
7.6.4. Testing All Coefficients	131

7.7. Single-Case Phase Designs	131
7.8. Questions	131
8. Clustering in Time and Space	134
8.1. The Generalized Quadratic Form	134
8.1.1. Mantel's U	134
8.1.2. An Example	135
8.2. Applications	135
8.2.1. The MRPP Statistic	136
8.2.2. The BW Statistic of Cliff and Ord [1973]	137
8.2.3. Equivalances	137
8.2.4. Extensions	137
8.2.5. Another Dimension	137
8.3. Alternate Approaches	138
8.3.1. Quadrant Density	138
8.3.2. Nearest-Neighbor Analysis	138
8.3.3. Comparing Two Spatial Distributions	139
8.4. Questions	139
9. Coping with Disaster	140
9.1. Missing Data	140
9.2. Covariates After the Fact	141
9.2.1. Observational Studies	142
9.3. Outliers	143
9.3.1. Original Data	144
9.3.2. Ranks	144
9.3.3. Scores	145
9.3.4. Robust Transformations	146
9.3.5. Use an L_1 Test	146
9.3.6. Censoring	146
9.3.7. Discarding	147
9.4. Censored Data	147
9.4.1. GAMP Tests	147
9.5. Censored Matched Pairs	149
9.5.1. GAMP Test	149
9.5.2. Ranks	150
9.5.3. One-Sample: Bootstrap Estimates	151
9.6. Adaptive Tests	151
9.7. Questions	152
10. Which Statistic? Solving the Insolvable	154
10.1. The Permutation Distribution	154

10.2. New Statistics	154
10.2.1. Nonresponders	154
10.2.1.1. Extension to K -Samples	155
10.2.2. Animal Movement	155
10.2.3. The Building Blocks of Life	156
10.2.4. Model Validation	157
10.2.5. Structured Exploratory Data Analysis	158
10.2.6. Comparing Multiple Methods of Assessment	159
10.3. Going Beyond	159
10.3.1. Sufficiency	162
10.3.2. Invariance	163
10.3.3. Applying the Principles	163
10.3.4. Losses	164
10.4. Likelihood Ratio	165
10.4.1. Goodness of Fit and the Restricted Chi-Square	166
10.4.2. Censored Data	167
10.4.3. Logistic Regression	167
10.5. Questions	168
11. Which Test Should You Use?	170
11.1. Parameters and Parametric Tests	170
11.2. Parametric Tests, Permutations, and the Bootstrap	171
11.3. What Significance Level Should I Use?	172
11.4. A Guide to Selection	173
11.4.1. The Data Are in Categories	173
11.4.2. The Data Take Discrete Values	174
11.4.3. The Data Are Continuous	174
11.5. Quick Key	176
12. Publishing Your Results	179
12.1. Design Methodology	179
12.1.1. Randomization in Assignment	179
12.1.2. Choosing the Experimental Unit	180
12.1.3. Determine Sample Size	181
12.2. Statistical Software for Exact Distribution-Free Inference	181
12.2.1. Freeware and Shareware	181
12.2.2. Commercial Software	182
12.3. Preparing Manuscripts for Publication	183
13. Increasing Computational Efficiency	185
13.1. Seven Techniques	185

13.2. Monte Carlo	185
13.2.1. Stopping Rules	186
13.2.2. Variance of the Result	186
13.2.3. Cutting the Computation Time	187
13.3. Rapid Enumeration and Selection Algorithms	187
13.3.1. Matched Pairs	188
13.4. Recursive Relationships	189
13.5. Focus on the Tails	190
13.5.1. Contingency Tables	191
13.5.1.1. Network Representation	191
13.5.1.2. The Network Algorithm	193
13.6. Gibbs Sampling and a Drunkard's Walk	194
13.7. Characteristic Functions	195
13.8. Asymptotic Approximations	195
13.8.1. A Central Limit Theorem	195
13.8.2. Edgeworth Expansions	196
13.8.3. Generalized Correlation	196
13.9. Confidence Intervals	196
13.10. Sample Size and Power	197
13.10.1. Simulations	197
13.10.2. Network Algorithms	198
13.11. Some Conclusions	199
13.12. Questions	200
14. Theory of Permutation Tests	201
14.1. Fundamental Concepts	201
14.1.1. Dollars and Decisions	201
14.1.2. Tests	202
14.1.3. Distribution Functions, Power, Exact, and Unbiased Tests	203
14.1.4. Exchangeable Observations	203
14.2. Maximizing the Power	204
14.2.1. Uniformly Most Powerful Unbiased Tests	204
14.2.2. The Fundamental Lemma	206
14.2.3. Samples from a Normal Distribution	207
14.2.4. Testing the Equality of Variances	208
14.2.5. Testing for Bivariate Correlation	208
14.3. Confidence Intervals	209
14.4. Asymptotic Behavior	210
14.4.1. A Theorem on Linear Forms	211
14.4.2. Asymptotic Efficiency	211
14.4.3. Exchangeability	212
14.5. Questions	213

Bibliography	215
B.1. Permutation Test Articles	216
B.2. Computational Methods	250
B.3. Seminal Articles	255
Author Index	257
Subject Index	265