Probing Interactions in Repeated Measures Designs: Applications in Clothing and Textiles Research

by

36

Lisa M. Lix

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Interdisciplinary Program
University of Manitoba
Winnipeg, MB

ISBN 0-612-13309-5

Name ____________________

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

STATISTICS | 0 4 6 3 | U·M·I

SUBJECT TERM | SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES

Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences

Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences

Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

PROBING INTERACTIONS IN REPEATED MEASURES DESIGNS: APPLICATIONS IN CLOTHING AND TEXTILES RESEARCH

BY

LISA M. LIX

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

Mixed designs, which contain one or more repeated measures factors in addition to one or more independent groups factors, are used in a variety of disciplines, including the clothing and textiles discipline. While many researchers may adopt the conventional analysis of variance (ANOVA) procedure to test repeated measures hypotheses in such designs this approach is not recommended, particularly for omnibus tests of interactions, as it is known to be highly sensitive to departures from the derivational assumption of multisample sphericity. Furthermore, omnibus tests of interactions in mixed designs are not useful in providing specific information on the localized sources of these effects.

A content analysis of clothing and textiles literature published between 1987 and 1993 revealed that the conventional ANOVA approach is popular for testing repeated measures hypotheses. However in using mixed designs, clothing and textiles researchers do not take full advantage of the factorial structure of the data, either by not testing for the presence of interactions or by following omnibus tests of interactions with tests of simple effects which do not provide relevant information about the specific nature of variable interactions.

It is shown that in two-factor designs, tetrad contrasts are the only viable way to probe interactions. Monte Carlo simulation techniques were used to collect empirical familywise Type I error and power rates for ten procedures for testing multiple tetrad contrast hypotheses in mixed designs when the multisample sphericity assumption was violated. Only three procedures provided acceptable control of error

rates; these relied on a test statistic formed using an estimate of the standard error of the tetrad contrast based on only those data used in defining the contrast (i.e., a nonpooled test statistic), in combination with either a Studentized maximum modulus, Hochberg (1988) step-up Bonferroni, or Shaffer (1986) modified sequentially rejective Bonferroni critical value. Minimal power differences between these three procedures were observed.

The application of these nonpooled tetrad contrast procedures to data from a hypothetical clothing and textiles data set was made with a computer program based on a general linear model approach to hypothesis testing using a nonpooled statistic.

ACKNOWLEDGEMENTS

The completion of this research owes much to the hard work, time, and energy generously given by committee chair, Dr. H. J. Keselman of the Department of Psychology, University of Manitoba. Sincere thanks are also extended to committee members, Dr. Carl J. Schwarz of the Department of Mathematics and Statistics, Simon Fraser University and Dr. Nelma I. Fetterman of the Department of Clothing and Textiles, University of Manitoba, who provided helpful input on this work, as well as other projects undertaken during my doctoral program. Constructive comments were also provided by external examiner Dr. Stephen Olejnik, Department of Educational Psychology, Research and Measurement, University of Georgia.

The financial support of the Social Sciences and Humanities Research Council of Canada (Award #752-92-1628) and the Canadian Home Economics Association is gratefully acknowledged.

My parents, sister, and parents-in-laws deserve special recognition for the support given over the course of my studies.

Finally, the greatest appreciation is owed to my husband, Lorne, who endured much during my years of post-graduate education, but provided unfailing support and always managed to help me keep my work in perspective and my life in order.

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS v

LIST OF TABLES viii

LIST OF FIGURES xi

CHAPTER

1. INTRODUCTION 1
   - Preamble 1
   - Introductory Remarks 3
   - Statement of the Problem 6
   - Purpose of Proposed Research 8
   - Study Rationale 9
   - Study Limitations 11
   - Definition of Terms 12

2. REPEATED MEASURES APPLICATIONS IN THE CLOTHING AND TEXTILES LITERATURE 14
   - An Overview of Repeated Measures Designs 14
   - Analysis Procedures and Associated Derivational Assumptions 15
     - Single-Group Designs 15
       - Simple Repeated Measures Designs 15
       - Factorial Repeated Measures Designs 19
       - Selecting an Analysis Procedure 20
     - Mixed Designs 20
     - Multivariate Repeated Measures Designs 24
     - Probing Correlated Effects 24
   - Repeated Measures Designs and the Applied Researcher 26
   - Assessing Repeated Measures Methodologies in Other Disciplines . . . 27
   - Assessing Repeated Measures Methodologies in the Clothing and Textiles Literature 29
     - Literature Search 30
     - Data Coding 34
       - Citation Information 34
       - Subject Area 34

Research Design 38
Data Analysis 39
Independent and Dependent Variables 42
Additional Information 42

3. RESULTS OF A CONTENT ANALYSIS OF THE CLOTHING AND TEXTILES REPEATED MEASURES LITERATURE 43
Citation Information 43
Subject Area 46
Repeated Measures Design Characteristics and Analysis Procedures . . 48
Simple Repeated Measures Designs 50
Factorial Repeated Measures Designs 55
Mixed Designs 62
Synopsis of the Content Analysis 70

4. TESTING INTERACTIONS IN MIXED DESIGNS 76
Univariate and Multivariate Tests of Interactions 76
Robust Tests of Interactions 79
Interaction Contrasts 84
Tetrad Contrast Test Statistics in the Mixed Design 88
Controlling the Familywise Error Rate for Multiple Tetrad Contrasts 91

5. SIMULATION STUDY METHODOLOGY 94
Overview of a Monte Carlo Study 94
Description of Tetrad Contrast Procedures 94
Monte Carlo Study Variables 95
Number of Repeated Measures Factor Levels 96
Sphericity Pattern 98
Heterogeneity of Group Variance-Covariance Matrices 98
Total Sample Size 101
Pairing of Group Covariance Matrices and Group Sizes 101
Population Shape 102
Nature of the Null Hypothesis 103
Effect Size 104
Verification of the Simulation Program 105
Design of the Simulation Study 105

6. SIMULATION STUDY RESULTS 108
Type I Error Rates 108
Power Rates 119
Synopsis of Simulation Results 123

7. APPLICATIONS, SUMMARY, AND CONCLUSIONS 125
Applications of Results 125
Summary of Results 134
Conclusions 137
Recommendations for Future Research 138

REFERENCES 141

APPENDICES 152
A. Statistical References Cited in Clothing and Textiles Repeated Measures Articles 152
B. Simulation Results for the Complete Null Hypothesis When N = 45 155
C. Simulation Results for a Partial Null Hypothesis 160
D. Example Data Set and Computer Programming Statements for Tetrad Contrasts Example 169

LIST OF TABLES

Table | Page

1. Frequency of RM Research Reports by Journal 44
2. RM Research Reports by Subject Area and Major Journal 47
3. RM Research Reports by Subject Area and Research Design 49
4. Profile of Simple RM Designs (f = 22) 51
5. Profile of Factorial RM Designs: Design Variables (f = 38) 56
6. Profile of Factorial RM Designs: Analysis Variables (f = 38) 59
7. Profile of Mixed Designs: Design Variables (f = 43) 63
8. Profile of Mixed Designs: Design and Analysis Variables (f = 43) 65
9. Methods for Testing RM Effects by Year of Publication (N = 100) . . . 73
10. Monte Carlo Study Variables 97
11. Pooled Variance-Covariance Matrix Element Values (K = 4) 99
12. Pooled Variance-Covariance Matrix Element Values (K = 8) 100
13. Population Means (μ_{jk}s) for Interaction Effect (J = 3; K = 4) 106
14. Population Means (μ_{jk}s) for Interaction Effect (J = 3; K = 8) 107
15. Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Complete Null Hypothesis; K = 4; N = 30) 110
16. Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Complete Null Hypothesis; K = 8; N = 30) 111
17. Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis; K = 4; N = 30) 113
18. Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis; K = 8; N = 30) 115

19. Average Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis) 117

20. Empirical Average Per-Comparison Power Rates for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Averaged Across Effect Sizes, Sphericity Values, and Population Shapes) 120

21. Empirical All-Comparison Power Rates for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Averaged Across Effect Sizes, Sphericity Values, and Population Shapes) 122

22. Means ($\bar{Y}$) and Variance-Covariance Matrices ($\hat{\Sigma}$) for Hypothetical Data Set 128

23. Tetrad Contrast Results for Hypothetical Data Set 130

24. Significance Criterion for Tetrad Contrasts 131

B1. Emprical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Complete Null Hypothesis; K = 4; N = 45) 156

B2. Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Complete Null Hypothesis; K = 8; N = 45) 157

B3. Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis; K = 4; N = 45) 158

B4. Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis; K = 8; N = 45) 159

C1. Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Partial Null Hypothesis; K = 4; N = 30) 161

C2. Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Partial Null Hypothesis; K = 4; N = 45) 162

C3. Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Partial Null Hypothesis; K = 8; N = 30) 163

C4. Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Partial Null Hypothesis; K = 8; N = 45) 164

C5. Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Partial Null Hypothesis; K = 4; N = 30) 165

C6. Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Partial Null Hypothesis; K = 4; N = 45) 166

C7. Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Partial Null Hypothesis; K = 8; N = 30) 167

C8. Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Partial Null Hypothesis; K = 8; N = 45) 168

LIST OF FIGURES

Figure Page

1. Graphical Representation of Mean Support Scores by Cup Size and Bra Condition 129

CHAPTER 1

INTRODUCTION

Preamble

Researchers in a variety of disciplines conduct investigations in which serial measurements are made on the same unit of analysis on one or more dependent variables. These observations typically occur as a function of time or as a result of experimental manipulations. Regardless of the manner in which the data arise, the measurements obtained for each unit of analysis are correlated, and the independent variable under investigation is known as a correlated groups or repeated measures (RM) factor.

As Lovie (1981) notes, interest in the application of analysis of variance (ANOVA) techniques to data from RM designs dates from the 1940s. Much of this work was conducted by psychologists, who drew on the writings of such notable statisticians as Fisher (1935) and Snedecor (1937). This early research focussed on the formulation of an ANOVA F statistic that could be used to test for the presence of a RM effect.

In the post-1950s era, interest among both statisticians and psychologists centred on statistical validity problems relevant to RM analyses incorporating the ANOVA F statistic. This research was prompted in large part by the knowledge that the mathematical assumptions which underlie traditional procedures for testing hypotheses in RM designs will, in practice, rarely be satisfied. Therefore, concern existed that the results of a RM analysis which incorporated conventional data analytic techniques would be invalid, and therefore provide misleading information.

As Lindsey (1993) notes, RM designs today enjoy a high degree of popularity in many different disciplines, from biology, to economics, to zoology. Contemporary investigations of methods for RM analyses have not been the exclusive domain of either statisticians or psychologists, but instead have been conducted by researchers from various backgrounds. As a result, this field of inquiry has a strong interdisciplinary focus.

The current project is a continuation of the examination of the appropriateness of statistical procedures for RM analysis, and accordingly, portions of this investigation should be of interest to researchers from a diverse range of disciplines. At the same time, this study also narrows its focus, and considers applications of RM methodologies (i.e., combinations of research design and analysis procedures) in the clothing and textiles (C&T) discipline.

The C&T orientation of the current study was selected for two reasons. First, in comparison to other disciplines, the C&T discipline is relatively new, and, as a result, lacks a well-defined base of both theoretical and methodological knowledge which researchers can use in formulating research problems, planning study designs, and choosing methods of data analysis (Nagasawa, Kaiser, & Hutton, 1989). Second, while RM designs find usage by researchers in the C&T field, there is always the potential for such designs to find a wider range of applications. Thus, this study will serve in part to introduce RM designs to some C&T researchers, and reinforce the use of such designs and appropriate analysis strategies for others. The following discussion provides a more detailed discussion of the objectives and rationale for the current research project.

Introductory Remarks

In the C&T field, RM designs have a number of different applications. One example is where individuals are asked to rate the overall acceptability of a garment during repeated wear trials; a second is where study participants are asked to record their perceptions of the personal traits of stimulus figures dressed in different clothing styles in which some detail, such as level of fashionability, has been manipulated; a third is where fabric properties are evaluated during multiple cycles of laundering.

One research question typically of interest to researchers who use RM designs is: Do differences among the RM sample means provide sufficient evidence to infer differences in the study population means? Investigations in other disciplines have demonstrated that applied researchers routinely adopt the conventional ANOVA *F* test to obtain an answer to this question (Brigham, 1974; Ekstrom, Quade, & Golden, 1990; LaTour & Miniard, 1983). This approach is generally considered inappropriate, since it rests on an assumption known as sphericity, which is unlikely to be satisfied in the majority of data-analytic situations. For the sphericity assumption to be met, the population variance-covariance matrix of the repeated measurements must be structured so that the variances of the differences between scores at all pairs of RM factor levels are equivalent. Furthermore, for mixed designs, in which repeated measurements are made for each of several independent groups of units of analysis, the more stringent assumption of multisample sphericity must be satisfied; this requires homogeneity of the common variance of the pairwise RM differences across groups. Because of the restrictive nature of these assumptions, investigations of the use of RM methodologies in various disciplines have concluded that applied researchers seldom adopt correct strategies for testing correlated effects in RM designs.

As with research data from other disciplines, it is likely that C&T data will often not satisfy the mathematical assumptions underlying the conventional ANOVA method for analyzing correlated effects in RM designs. Consequently, C&T researchers may not be using data-analytic strategies that will produce valid results, which in turn, may lead them to conclude that differences exist in their study populations, when in fact none are present. Such erroneous conclusions are known as Type I errors.

These errors have, on several occasions, been noted as a cause for concern among C&T researchers. Because much of the research in the C&T field is exploratory in nature and replication studies are rarely published (Turnbull & Lix, 1991), false positive results may be given undue importance in directing the course of future research. Moran (1986) suggests that

> we, as consumers of research, are often not as careful about evaluating findings as we should be (after all, if it gets published, the findings should be trustworthy). We tend to overgeneralize the findings of a single study on a single sample and suggest that these findings (and often even the more

> speculative interpretations that go along with these findings), are facts applicable to much broader groups. This tendency will probably always be the case and thus it becomes incumbent upon the researcher to do what he or she can to avoid the publication of misleading data, i.e., Type I errors. (p. 380)

It would, however, be unjustified to discontinue using RM designs for the investigation of research problems in the C&T field, or, for that matter, in other fields of scientific inquiry. Given an equal sample size, a correlated groups design usually offers greater power to detect a treatment effect than an independent groups design in which each unit of analysis is evaluated at only one point in time or under a single experimental condition (Kirk, 1982, p. 240). The former design is generally more efficient than the latter because the degree of variability among observations made on a single unit of analysis will typically be smaller than the degree of variability among observations made on several different units of analysis. Since less error variability is likely to be present in a correlated groups design than in an independent groups design, the former provides for a greater probability of detecting a true difference among the population means. Consequently, researchers who make use of RM designs may require smaller sample sizes to achieve results comparable to those that could be obtained using an independent groups design. Thus, the adoption of a RM design is one way to make better use of scarce research resources.

Instead of abandoning RM designs, the solution lies in the identification of tests that are insensitive, or in other words, robust to mathematical assumption violations in RM designs. Monte Carlo simulation techniques are valuable in this respect, because they can be used to examine the behaviour of statistical procedures when assumptions are violated to various known degrees in the simulated data. Based on the results obtained from manipulation of the simulation model, it is possible to make recommendations to researchers on the appropriate use of these statistical procedures. In other words, the purpose of simulation studies is to provide information on the conditions under which a statistical test will produce valid results.

Statement of the Problem

Both Damhorst (1990) and Turnbull and Lix (1991) have noted the trend towards the use of more complex research designs and analysis procedures among C&T researchers, which is due in large part to the recognition that clothing as a form of human behaviour is a multifaceted process. For example, because it is well known that the "relevance of [clothing] information is dependent upon the setting and background cues" (Damhorst, p. 6), C&T researchers are more likely to be interested in examining the joint effect of two or more independent variables on a dependent variable, than in studying the predictive ability of only a single independent variable. As a consequence, tests of interactions may be of greater importance to C&T researchers than those of individual or main effects.

Mixed designs may be particularly useful to C&T researchers, as they offer the opportunity to investigate differences in responses of independent groups of units of analysis under exposure to all levels of one or more RM factors; these groups may be formed, for example, on the basis of age or sex. C&T researchers who use mixed designs may be interested in determining whether the pattern of differences among

RM sample means are the same across levels of an independent groups factor, and consequently will test for an interaction. Past research has demonstrated, however, that conventional procedures for analyzing interactions in mixed designs may be highly sensitive to departures from derivational assumptions. In recognition of this, Algina and Oshima (1994) and Keselman, Carriere, and Lix (1993) examined alternative solutions for testing omnibus interaction effects in mixed designs under departures from the multisample sphericity assumption; both sets of authors concluded that the investigated procedures could produce valid results under assumption violations.

While the procedures described by Algina and Oshima (1994) and Keselman et al. (1993) enable researchers to obtain valid tests of omnibus interaction effects in mixed designs, interest generally extends beyond identification of a significant omnibus result to the determination of the localized source of this result. For example, Lix (1995) found that in both correlated and independent groups designs, C&T researchers routinely follow a significant omnibus main effect result with multiple pairwise comparisons of treatment means to tease out the specific source of the effect. In factorial designs, Kaufman, Dudley-Marling, and Serlin (1986) and Rosnow and Rosenthal (1989a) have confirmed that many researchers conduct tests on simple effects when a significant interaction is obtained, including simple main effect tests and simple pairwise comparisons. However, as these authors note, such procedures are inappropriate for examining interaction effects and researchers who adopt such an approach are likely to misinterpret a significant result. Specific procedures for probing interactions, known as interaction contrasts, are available, but as Boik (1993) notes, most applied researchers are not familiar with these techniques.

To summarize, C&T researchers may either already use mixed designs in their research or find the adoption of such designs beneficial, the use of mixed designs implies interest in examining one or more variable interactions, and research has demonstrated that applied researchers may frequently misinterpret the nature and source of an interaction in their data by failing to adopt techniques which allow for the exploration of this effect. Furthermore, because procedures for testing interactions in mixed designs are known to be highly sensitive to departures from the multisample sphericity assumption, it is likely that procedures for probing interactions will suffer from this same weakness. Yet at present, no studies have investigated potentially robust solutions for conducting tests which are designed specifically for probing interactions within the context of mixed designs. The results of simulation studies which investigate the robustness of statistical techniques for probing interaction effects in RM designs should therefore provide relevant information for C&T researchers as well as researchers in other fields where mixed designs find a high degree of usage.

Purpose of Proposed Research

Recommendations on improvements in data-analytic strategies are often most meaningful when they are accompanied by documentation of existing research practices. This documentation is useful for directing recommendations to appropriate audiences and for providing empirical evidence of inadequacies in currently adopted methodologies. Accordingly, one purpose of the present research is to investigate the use of RM designs by researchers in the C&T field, particularly the strategies adopted for testing interaction effects in mixed designs. This investigation is undertaken via a content analysis of published C&T research.

A second purpose is to use simulation techniques to examine the efficacy of several multiple comparison procedures (MCPs) for probing interactions via interaction contrasts in mixed designs when the multisample sphericity assumption is not satisfied. Lix (1995) found that while a number of different MCPs are used by C&T researchers, procedures which do not provide control of the overall probability of committing at least one Type I error among the overall set of comparisons, otherwise known as the familywise Type I error rate (FWR) are routinely selected. Furthermore, the issue of violations of derivational assumptions is rarely considered when selecting a MCP. The MCPs considered in this project utilize an approximate degrees of freedom (df) test statistic which is based on the solutions of Welch (1947, 1951) and James (1951, 1954) for testing the equality of means in the presence of variance heterogeneity.

Study Rationale

Green (1984) notes that the home economics discipline has been criticized for its failure to produce research papers which address methodological issues in data analysis. This is particularly evident in the C&T field. At present, only a single study has addressed specific inadequacies in data analysis conducted by C&T researchers (Lix, 1995). This represents a significant gap in the development of the field, because, as Schumm (1982) argues, an integrated understanding of theory, variable measurement, and statistical analysis is necessary for researchers to competently conduct scientific investigations (see also Nagasawa et al., 1989). Furthermore, since new developments in data-analytic techniques are continually appearing in the literature, it is essential for applied researchers to be exposed to these innovations so that they are aware of the range of choices available, and accordingly, select statistical techniques which are most appropriate for achieving the stated objectives of their investigations.

Because C&T researchers may find it advantageous to adopt RM designs in their investigations it is important that researchers understand how to proceed in the analysis of correlated data. This involves an awareness of: (a) the types of research questions that can be addressed with such designs, (b) why and how data may not conform to the assumptions underlying traditional methods of RM analysis, (c) problems with existing data analysis strategies, (d) alternative methods of analysis, and (e) how to implement these alternative procedures. The present research will address these issues within the context of techniques for examining variable interactions.

However, as noted at the outset of this introduction, the findings of this study will be applicable to other disciplines. C&T research tends to be interdisciplinary in nature (Turnbull & Lix, 1991), hence it is likely to be exhibit characteristics representative of research in other fields, such as marketing and psychology. As well, this study will add to the growing body of knowledge of RM analysis procedures found in such fields as statistics and psychology.

Study Limitations

The following limitations on the scope of this research project are noted:

1. The investigation of the use of RM methodologies in the C&T field is limited to published research. This restriction was applied for two reasons. First, published works are likely to have the greatest influence on the way in which researchers conduct their own studies. Researchers may tend to emulate the approaches towards design and analysis taken in published works in the hopes that this will improve the chances of having their own works accepted for publication. Secondly, unpublished works are often more difficult to identify and their existence is rarely documented in published indices. Without a systematic approach available to identify unpublished works, their inclusion in this investigation might allow for bias to enter into the study selection process. However, the exclusion of unpublished works may, in and of itself, create bias, as unpublished works are often systematically different from their unpublished counterparts; they may contain fewer statistically significant result, or a greater number of methodological flaws, or both (Greenwald, 1975; Moran, 1986).

2. This study only considers methods for probing interactions within the context of two-factor designs that contain only a single dependent variable. These methods may be generalized to higher-order designs and designs that contain multiple dependent variables which are being investigated simultaneously.

However, before considering their use in more complex designs, it is important to consider the issue of robustness in a simple design.

Definition of Terms

Throughout this study, the following terms and concepts will find frequent usage, and are defined here in detail for purposes of clarity:

Type I Error Rate: The probability of erroneously rejecting a true null hypothesis.

Familywise Error Rate (FWR): The overall or joint probability of committing at least one Type I error in a complete family of hypothesis tests.

Per Comparison Error Rate (PCR): The probability of committing a Type I error when testing a given hypothesis.

Balanced Design: A research design in which group sizes are equal. When group sizes are unequal the design is unbalanced.

Contrast (Comparison): A set of coefficients which specifies a comparison among a set of population means, and defines a hypothesis of interest to the researcher. At least two contrast coefficients in the set must be nonzero, and the coefficients must sum to zero.

Orthogonal Contrasts: A set of unrelated (uncorrelated) contrasts. The sum of the crossproducts of the coefficients of each pair of orthogonal contrasts is equal to zero when the design is balanced. For unbalanced designs, a pair of contrasts is orthogonal if the weighted sum of the crossproducts of a set of coefficients are equal to zero, where the weights are equal to the inverse values of the group sizes.

Linearly Independent Contrasts: A set of contrasts in which no individual contrast can be formed from a linear combination of any other contrast(s) in the set.

Pairwise Comparison: A contrast in which two means are compared.

Complex Contrast: A contrast in which more than two means are compared.

Marginal (Main) Effect: The effect of one treatment (experimental) factor that is obtained when the population values are averaged across the levels of all other factors. This averaging may be conducted using either weighted or unweighted values when the design is unbalanced.

Simple Effect: The effect of one treatment factor at a particular level of another factor, or at a combination of levels of two or more factors.

Omnibus Test: A test that is used to evaluate differences among more than two different groups of units of analyses, and has degrees of freedom equal to those available for the effect under investigation.

CHAPTER 2

REPEATED MEASURES APPLICATIONS IN THE CLOTHING AND TEXTILES LITERATURE

The first objective of this study was to assess the use of RM methodologies in the C&T literature, particularly with respect to mixed designs. This chapter begins with a discussion of: (a) the various types of RM research designs that may be adopted by the applied researcher, (b) methods for analyzing correlated effects, and (c) the derivational assumptions on which these procedures rest. An examination of the literature on the use of RM methodologies in other disciplines follows. A description of the method used to conduct a search of the C&T RM literature and to define the characteristics of this literature concludes the chapter.

An Overview of Repeated Measures Designs

In a simple RM design, a single group of units of analysis (e.g., study participants) is evaluated at each level of one RM factor. A single group of units of analysis may also be evaluated at each combination of levels of two or more RM factors; such a design is referred to as a factorial RM design. In contrast, in a mixed design, units of analysis are classified on the basis of one or more independent groups factors and are evaluated at each level of a single RM factor, or at each combination of levels of two or more RM factors. Such designs are denoted as mixed because both independent groups and correlated groups factors are involved. All of these RM designs may be univariate in nature, such that each unit of analysis is evaluated on only a single dependent variable, or they may have a multivariate structure, such that each unit of analysis is measured at each level of the RM factor(s) on multiple dependent variables.

Analysis Procedures and Associated Derivational Assumptions

Single-Group Designs

Simple repeated measures designs. Let Y_{ik} denote the *k*th score for the *i*th unit of analysis ($i = 1, \ldots, N$) in a simple RM design in which only one dependent variable is under investigation. The observations $\mathbf{Y}_i = [Y_{i1} \ldots Y_{iK}]$ are assumed to be independently and normally distributed random variables with mean vector $\mu = [\mu_1 \ldots \mu_K]$ of dimension 1 x K and variance-covariance matrix Σ of dimension K x K. The omnibus null hypothesis under consideration is H_o: $\mu_1 = \cdots = \mu_K$. The conventional univariate statistic used to test H_o is $F = MS_K/MS_{KS}$, where MS_K is the mean square for the RM effect, and MS_{KS} is the mean square error (MSE).

When the variances of the K repeated measurements are equal, and the covariances among them are equal, the variance-covariance matrix, Σ, is said to possess compound symmetry (Kirk, 1995, p. 275). Compound symmetry is a sufficient condition for the conventional ANOVA F statistic to follow an F distribution under the null hypothesis. However it is not a necessary condition.

The necessary condition underlying the conventional ANOVA approach for testing H_o is sphericity or circularity (Huynh & Feldt, 1970; Rouanet & Lepine, 1970), which, in matrix notation, is defined as

$$\mathbf{C}^{T}\Sigma\,\mathbf{C} = \lambda\mathbf{I}_{(K-1)}\,, \tag{2.1}$$

where $\mathbf{C}$ is a K x (K - 1) orthonormal contrast matrix for the K RM, λ is a scalar > 0, $\mathbf{I}$ is an identity matrix, and superscript T is the transpose operator. For $\mathbf{C}$ to be orthonormal, the columns of $\mathbf{C}$ must form a set of (K - 1) orthogonal contrasts among the levels of the RM factor. Furthermore, each column must have a length of one, or in other words, $\mathbf{C}^{T}\mathbf{C} = \mathbf{I}_{(K-1)}$.

Another way of stating this sphericity assumption is based on the algebra of expectations. It may be shown that

$$\sigma^2_{(k-k')} = \sigma^2_{k} + \sigma^2_{k'} - 2\rho_{kk'}\sigma_{k}\sigma_{k'}\,. \tag{2.2}$$

Equation 2.2 illustrates that the variance of the difference between scores at any pair of levels of the RM factor may be expressed as a function of the corresponding variances and covariance. The latter is a function of the correlation among the repeated measurements and the standard deviations. When variances are equal for all possible pairwise difference variables, the data are spherical. As is apparent from Equation 2.2, when there are only two levels of the RM factor, the sphericity assumption is trivially satisfied, since only a single difference variable can be formed.

One way in which violations of the sphericity assumption may arise is because of variations in the degree of correlation among dependent variable scores at all pairs of levels of the RM factor, a situation that may be encountered in many data-analytic problems. For example, in the C&T discipline, in studies of the acceptability of a garment during repeated wear trials, successive trials may be evaluated in a similar manner and therefore may be highly correlated, whereas trials separated by a greater length of time may give rise to dissimilar evaluations, and will not be as highly correlated. It is rarely possible to limit the existence of such serial correlation patterns when the units of analysis are studied as a function of time. However, in situations where the RM factor is an experimental variable, it is important for units of analysis to be randomly exposed to the factor levels to limit the existence of deleterious carry-over effects that may give rise to nonspherical data. Unfortunately, many researchers may not attend to the issue of randomization in designing an experiment, which may have serious implications for the validity of the statistical procedures that are used in data analysis.

Box (1954) proved that under the null hypothesis, the usual F statistic is approximately distributed as an F variate with $(K - 1)\epsilon$ and $(K - 1)(N - 1)\epsilon$ df. The parameter ϵ is an index of the degree of sphericity in the population covariance matrix. In matrix notation,

$$\epsilon = \frac{[\mathrm{tr}(\mathbf{C}^T \Sigma \mathbf{C})]^2}{(K-1)\,\mathrm{tr}[\mathbf{C}^T \Sigma \mathbf{C}]^2}, \tag{2.3}$$

where tr is the trace operator, and the remaining elements are as previously defined. The parameter ϵ may assume a range of values, from an upper bound of 1.0 when sphericity is present in the data, to a minimum of $(K-1)^{-1}$.

For testing the hypothesis of no RM effect, one could compare the computed F statistic to a critical F value with numerator and denominator df adjusted by ϵ. However, it is unlikely for researchers to have information about the population parameter ϵ. Therefore, one approach suggested for testing the null hypothesis is to compare the computed F statistic to a critical value (CV) for which the numerator and denominator df have been adjusted by a factor equal to the lower bound of ϵ (Geisser & Greenhouse, 1958). Accordingly, the F statistic is compared to a CV with 1 and (N - 1) df. This approach, often referred to as a conservative F test, is not widely recommended, as it may result in a test that is insensitive to differences among the RM population means (Rogan, Keselman, & Mendoza, 1979).

Greenhouse and Geisser (1959) suggested using a sample estimate, $\hat{\epsilon}$, for df adjustment, which is computed by replacing Σ in Equation 2.3 with the sample covariance matrix, $\hat{\Sigma}$. However, Collier, Baker, Mandeville, and Hayes (1967) found $\hat{\epsilon}$ to be a biased estimate of the population parameter for values of ϵ greater than 0.75, particularly when sample sizes were small. Hence, Huynh and Feldt (1976) recommended $\tilde{\epsilon}$, where

$$\tilde{\epsilon} = \min\left[1\,,\,\frac{N(K-1)\hat{\epsilon}-2}{(K-1)[N-1-(K-1)]\hat{\epsilon}}\right]. \tag{2.4}$$

Finally, Quintana and Maxwell (1985) proposed $\bar{\epsilon}$, due to the findings of Maxwell and Arvey (1982) that $\tilde{\epsilon}$ tends to produce an excessive number of Type I errors under conditions in which sample size is small and/or there are a large number of repeated measurements (see also Quintana & Maxwell, 1994). This estimate is defined as

$$\bar{\epsilon} = \min\left[1\,,\,1/2(\hat{\epsilon} + \tilde{\epsilon})\right]. \tag{2.5}$$

A multivariate procedure may also be adopted to test the null hypothesis. Due to the existence of correlations among the repeated measurements, Cole and Grizzle (1966) argue that multivariate analysis of variance (MANOVA) is the appropriate procedure to adopt. Under this analysis strategy, the K RM are transformed into a set of (K - 1) linearly independent difference variables. For the simple RM design, Hotelling's (1931) T^2 statistic is used to test the null hypothesis of equality of the means of these difference variables. This approach makes no specific assumptions regarding the structure of Σ, although the assumptions of independence of observations and multivariate normality must be satisfied.

Factorial repeated measures designs. When testing correlated effects in factorial RM designs using the conventional ANOVA approach, both overall and local sphericity assumptions must be considered (Mendoza, Toothaker, & Crain, 1976). The overall sphericity assumption is satisfied when the **C** matrix in Equation 2.1 defines a set of orthonormal contrasts for all of the levels of the RM factors, while a local sphericity assumption is satisfied when **C** defines an orthonormal contrast matrix for a particular effect of interest (e.g., a main or interaction effect). Overall and local sphericity need not be simultaneously satisfied in a single set of data, although the former, which is the more stringent assumption, implies that the latter will be met. The conventional ANOVA F statistic can be computed in different ways depending on whether or not the researcher is willing to assume that the data conform to the overall sphericity assumption. In the case where neither assumption is likely to be met, either a univariate df-adjusted procedure, or Hotelling's (1931) T^2 procedure may be adopted. However, with respect to the former, it is important to note that a different value of $\hat{\epsilon}$, $\tilde{\epsilon}$, or $\bar{\epsilon}$ can be computed for df adjustment for each effect under investigation.

Selecting an analysis procedure. For both simple and factorial RM designs, many researchers favour the multivariate approach over univariate df-adjusted procedures, as the former is exact, while the latter are only approximate tests of the null hypothesis (Keselman & Keselman, 1993; O'Brien & Kaiser, 1985). However, it is known that Hotelling's (1931) T^2 test can be sensitive to departures from the multivariate normality assumption when total sample size is less than 30 (Lix, Keselman, & Keselman, 1995). Furthermore, the multivariate procedure is not uniformly more powerful than a univariate one. The relative power advantage of either approach is a function of the alternative hypothesis, the structure of Σ, and the relationship between these two factors (Barcikowski & Robey, 1990; Davidson, 1972).

Mixed Designs

The simplest mixed design contains a single independent groups factor, A, with $j = 1, \ldots, J$ levels and n_j units of analysis within each level of A ($\sum_j n_j = N$), and a single correlated groups factor, B, with $k = 1, \ldots, K$ levels. Let Y_{ijk} represent the kth measure on the ith unit of analysis ($i = 1,\ldots,n_j$) in the jth group. The observations $\mathbf{Y}_{ij} = [Y_{ij1} \ldots Y_{ijK}]$ are assumed to be independently and normally distributed, with mean vector $\boldsymbol{\mu}_j = [\mu_{j1} \ldots \mu_{jK}]$ and covariance matrix Σ_j.

Under the conventional approach to RM analysis, the null hypothesis of no RM main effect, H_o: $\mu_{.1} = \cdots = \mu_{.K}$, where $\mu_{.k}$ is the *k*th marginal RM mean, is tested via $F_K = MS_K/MS_{KS/J}$. Here, MS_K and $MS_{KS/J}$ are the mean square for the RM effect and the MSE, respectively. In the latter term, the forward slash is used to indicate that units of analysis are nested with groups. The null hypothesis of no interaction effect is H_o: $\mu_{jk} - \mu_{.k} - \mu_{j.} + \mu_{..} = 0$ for all j and k, where $\mu_{.k}$, $\mu_{j.}$, and $\mu_{..}$ are respectively the RM, group, and grand means. This hypothesis is tested via $F_{JxK} = MS_{JxK}/MS_{KS/J}$, where MS_{JxK} is the interaction mean square and $MS_{KS/J}$ is as previously defined. Finally, the null hypothesis of no group main effect, H_o: $\mu_{1.} = \cdots \mu_{J.}$, is tested via $F_J = MS_J/MS_{S/J}$, were MS_J is the group mean square, and $MS_{S/J}$ is the MSE.

In order for the ANOVA F test to provide a valid test of either the RM main or interaction effect in a mixed design, the assumptions of multivariate normality and independence must be satisfied. In addition, it is assumed that the sphericity assumption holds, that is,

$$\mathbf{C}^T \Sigma_p \mathbf{C} = \lambda \mathbf{I}_{(K-1)} \, , \tag{2.6}$$

where Σ_p is the population covariance matrix that has been averaged (pooled) across the levels of the grouping factor, and the remaining elements are as defined for the simple RM design. Furthermore, for multisample sphericity to exist (Huynh, 1978), the covariance matrix of the orthonormal contrast variables must be equal across all levels of the grouping factor. In matrix notation, this assumption is represented as

$$C^T\Sigma_1 C = C^T\Sigma_2 C = \cdots = C^T\Sigma_J C = \lambda I_{(K-1)} . \tag{2.7}$$

For the ANOVA F test to provide a valid result for a test of the independent groups main effect, the multisample sphericity assumption need not be satisfied. However, it is assumed that homogeneity of variances exists across the levels of the grouping factor for the average value of the repeated measurements.

For the mixed design, Box (1954) showed that under the hypothesis of no RM main effect, F_K is approximately distributed as an F variable with $(K - 1)\epsilon$ and $(K - 1)(N - J)\epsilon$ df. Similarly, F_{JxK} is approximately distributed as an F variate with $(J - 1)(K - 1)\epsilon$ and $(K - 1)(N - J)\epsilon$ df under the hypothesis of no RM interaction. The population parameter, ϵ, is computed in the same manner as for the simple RM design, with the exception that Σ is replaced with Σ_p in Equation 2.3.

Any one of the $\hat{\epsilon}$, $\tilde{\epsilon}$, or $\bar{\epsilon}$ df-adjusted F tests may be adopted for testing correlated effects in mixed designs when is it unlikely that the data will conform to the sphericity assumption. The $\hat{\epsilon}$ statistic is defined in the same manner as for the simple RM design, with the exception that $\hat{\Sigma}_p$ replaces Σ in Equation 2.3. However, the $\tilde{\epsilon}$ statistic (Huynh & Feldt, 1976; Lecoutre, 1991) is defined as,

$$\tilde{\epsilon} = \min\left[1 , \frac{(N-J+1)(K-1)\hat{\epsilon}-2}{(K-1)[N-J-(K-1)]\hat{\epsilon}}\right] . \tag{2.8}$$

Finally, the $\bar{\epsilon}$ statistic is computed using Equation 2.5.

Data from mixed designs may also be analyzed using a multivariate procedure. In order to test the RM main effect, the grouping factor is ignored, and the set of linearly independent difference variables are pooled across the levels of the grouping factor. Hotelling's (1931) T^2 statistic is used to assess whether the vector of means for these difference variables is equal to the null vector. For the interaction effect, the hypothesis of interest is equality of the vector of mean difference variables across the levels of the grouping factor. Hence the problem is reduced to a one-way independent groups MANOVA on the difference scores, and the null hypothesis may be evaluated using Hotelling's T^2 when the number of groups is equal to two, or one of: (a) Hotelling-Lawley (Hotelling, 1951; Lawley, 1938) trace, (b) Pillai-Bartlett (Bartlett, 1939; Pillai, 1955) trace, (c) Roy's (1953) largest root criterion, or (d) Wilks' (1932) likelihood ratio, for multi-group mixed designs.

In order for a multivariate procedure to provide a valid test of the data, multivariate normality is assumed, as is equality of the group orthonormal covariance matrices. However, the data need not be spherical.

In mixed designs, the multivariate approach is generally favoured over either of the $\hat{\epsilon}$, $\tilde{\epsilon}$, or $\bar{\epsilon}$ tests when the design is balanced, provided that the degree of inequality of the group orthonormal covariance matrices is not large (Keselman, Lix, & Keselman, 1994). However neither the univariate or multivariate approach is considered to be appropriate when the design is balanced but the degree of inequality of the group covariance matrices is large, or when the design is unbalanced, regardless of the degree of covariance heterogeneity. Specific information concerning the operating characteristics of the univariate and multivariate approaches under such conditions is discussed in a subsequent chapter.

Multivariate Repeated Measures Designs

All of the methods for analyzing correlated effects have been described within the context of a design in which only a single dependent variable is under consideration, or in which each of several independent variables is being evaluated in isolation. In those instances where the data structure is multivariate in nature and the researcher is interested in conducting tests of hypotheses for the set of dependent variables, a multivariate MANOVA procedure may be adopted. Hotelling's (1931) test is again used to test multivariate RM hypotheses in either simple or factorial RM designs as well as to test main effect hypotheses in the mixed design. Any one of the four multivariate procedures described previously for analyzing interactions in the mixed univariate design can be applied to test interactions in a mixed multivariate experiment. However, the statistics for testing multivariate RM main or interaction hypotheses may be computed in different ways, depending on whether or not the researcher is willing to assume that multivariate (multisample) sphericity is satisfied. Boik (1991) provides specific details of the methods available for computing multivariate statistics under these two approaches.

Probing Correlated Effects

In many instances, the researcher will follow a significant test of a RM effect with contrasts to probe that effect. Alternatively, the researcher may elect to bypass a test of the omnibus hypothesis altogether in favour of a series of contrasts to aid in the identification of the localized source of the effect. In both instances, the most common approach is to conduct pairwise comparisons of the RM means (Jaccard, Becker, & Wood, 1984; Lix, 1995).

In the simple RM design, the test statistic for performing these comparisons may be computed in two different ways; the choice of one approach over the other is a function of the assumptions the researcher is willing to make about the data. In the first, the test statistic incorporates the MSE term used to test the omnibus hypothesis. This statistic is known as a pooled statistic because the error term is based on the data from all levels of the RM factor. The sphericity assumption must be satisfied for such an approach to provide valid tests of pairwise comparisons (Keselman, 1982). The alternative, a nonpooled test statistic, uses an error term based on only that data associated with the particular levels of the RM factor that are being compared and is equivalent to a paired t statistic (Maxwell, 1980). Thus, each pairwise comparison statistic has a separate error term. Since only two levels of the RM factor are used to derive the test statistic, the sphericity assumption is trivially satisfied. However, it is still assumed that the data follow a multivariate normal distribution.

The same concepts of pooled and nonpooled statistics apply to pairwise comparisons in factorial RM designs and in mixed designs. However, pooling may be conducted in different ways, and is a function of whether the researcher is probing a marginal effect or a simple effect. For example, in the A x B mixed design described previously, pairwise comparisons of the Factor B marginal means may be conducted using a test statistic which employs an error term based on the usual MSE for the omnibus test, which is pooled across all of the data. The use of such a statistic necessarily assumes that multisample sphericity is satisfied (Keselman & Keselman, 1988). Alternatively, the test statistic may incorporate an error term based on only that data at the two levels of Factor B which are a part of the comparison and thus only requires homogeneity of the group orthonormal covariance matrices, not sphericity. Pairwise comparisons of the Factor B simple main effect means at a particular level of Factor A may be made using an error term based on the usual MSE, which assumes that the data conform to the multisample sphericity assumption. These tests may also be conducted using the MSE computed at a particular level of Factor A (Keselman & Keselman, 1993); this approach is only dependent on sphericity of the variance-covariance matrix at the chosen level of the grouping factor. Alternatively, the test statistic may employ an error term which is pooled over neither the Factor B or Factor A levels, and is therefore based on only that data used in defining the comparison of interest. This nonpooled statistic is not dependent on either component part of the multisample sphericity assumption.

Repeated Measures Designs and the Applied Researcher

The information provided to this point on methods available for RM analysis may be too technical in nature for the applied researcher. Individuals requiring a less complex discussion have a number of sources at their disposal. Most statistical textbooks include a section on RM analyses; Maxwell and Delaney (1990) provide an excellent treatment of this topic. O'Brien and Kaiser (1985) give a simplistic discussion of MANOVA methods for RM analysis which includes a basic introduction to matrix algebra. Barcikowski and Robey (1984) and Looney and Stanley (1989) consider data analysis procedures for single-group and mixed designs, respectively. Selected authors in the home economics discipline have also dealt with RM analyses within the context of family studies research problems (see Ball, McKenry, & Price-Bonham, 1983; Sanik, 1983; Schumm, Barnes, Bollman, Jurich, & Milliken, 1985; Schumm, Bugaighis, & Jurich, 1985).

Assessing Repeated Measures Methodologies in Other Disciplines

As a result of the variety of procedures available for the analysis of RM data, a number of studies in various fields of scientific inquiry have investigated applications of RM methodologies by applied researchers and used this information to formulate recommendations on appropriate methods of RM analysis. This section provides a summary of both the findings and recommendations of these content analyses.

Brigham (1974) surveyed research reports published between 1969 and 1971 in Ergonomics to examine the popularity of a variety of different statistical procedures. RM analyses were adopted in 27 of the 108 studies which the author identified and in all of these, the conventional ANOVA F procedure was adopted. Brigham criticized this approach, and suggested that a conservative F procedure should have been used instead. Furthermore, the author noted that if this strategy had been used, 13 of the 27 articles would no longer have reported significant results.

LaTour and Miniard (1983) evaluated published articles in two marketing journals for the period 1974 to 1979 to identify studies employing correlated data. The authors included studies using simple, factorial, and mixed designs. Of the 55 research reports which the authors identified, 24 reported the use of the conventional ANOVA approach for testing hypotheses on RM effects, and therefore the consequences associated with possible violations of the (multisample) sphericity assumption were not considered. None of the studies reported the use of a df-adjusted procedure, but in one article the authors relied on a conservative F test. In only two of the 55 studies was MANOVA adopted. LaTour and Miniard also identified two papers in which the RM factor was erroneously treated as an independent groups factor in the computation of ANOVA F statistics. Finally, in several papers in which a mixed design was used, tests of simple independent group effects were conducted at each level of the RM factor(s), thereby bypassing tests of an omnibus RM effect. In their conclusions, LaTour and Miniard recommended a multivariate approach to RM analysis since it is "the most versatile of the analytic methods" (p. 55). Furthermore, they suggested that a multivariate analysis is likely to afford statistical power which is comparable to that of a df-adjusted procedure.

In a more recent study, Ekstrom et al. (1990) evaluated analysis procedures applied to test correlated effects in mixed designs in the psychiatric RM literature for a six-month period in 1988. The authors' most significant finding was that more than one third of the 63 articles they identified did not include sufficient information to conclude what type of analysis had been performed, although in many of these articles it appeared that the traditional ANOVA approach had been adopted. In seven studies MANOVA was used to test RM main or interaction effects and in another four studies a df-adjusted test was used. A further 16% only reported the results of tests of simple independent group effects conducted at each level of a single RM factor or at each combination of levels of two or more factors. Ekstrom et al. also recommended a multivariate approach and, like LaTour and Miniard (1983), noted that it is likely to perform well in comparison to a df-adjusted procedure for detecting a false null hypothesis.

Assessing Repeated Measures Methodologies in the Clothing and Textiles Literature

The three studies which investigated applications of RM methodologies in other disciplines all found that applied researchers routinely adopt a conventional ANOVA approach for testing correlated effects despite its reliance on the stringent (multisample) sphericity assumption. The popularity of this approach does not seem to have faded over time, despite the repeated admonishments against its use that have appeared in the literature and the favour directed towards the use of MANOVA.

While the authors of these papers provide important insights into the ways in which applied researchers test RM effects, there are a variety of issues which they did not consider. Specifically, methods for probing omnibus effects were not evaluated in any detail, despite the known popularity of MCPs (Jaccard et al., 1984; Lix, 1995). Furthermore, while many of the studies included in these investigations employed either factorial RM designs or mixed designs, a detailed report of the manner in which interaction effects were analyzed was not given and neither was the choice of an error term for conducting follow-up tests.

The present assessment of the use of RM designs in the C&T literature was designed to provide a more thorough evaluation of the use of RM methodologies than has previously been conducted. Although the primary goal was to examine methods of interaction analysis in mixed designs, procedures used for testing correlated effects in other types of RM designs were not excluded. Thus, as a whole, the current study provides a timely successor to the works of previous authors, as it extends the knowledge of methods used by applied researchers for testing RM effects into the 1990s.

Literature Search

Four journals may be regarded as primary sites for publication of C&T research reports that are read by North American researchers. These are: (a) Canadian Home Economics Journal, (b) Clothing and Textiles Research Journal, (c) Home Economics Research Journal, and (d) Journal of Consumer Studies and Home Economics. Although the Journal of Home Economics may also be viewed as central to the C&T field, it does not currently publish reports of original research. Articles dealing with textile science topics may also be found in the Textile Research Journal and Journal of the Textile Institute. However, these two journals are not considered central to the C&T discipline due to their more specialized focus.

The four major journals of the discipline could provide a representation of applications of RM methodologies in the C&T discipline. However, it is well known that C&T research has a strong interdisciplinary focus (Oliver & Mahoney, 1991; Turnbull & Lix, 1991) and for this reason, researchers frequently publish in journals of other disciplines (Hutton, 1984). For example, in a citation analysis of three volumes of the Clothing and Textiles Research Journal published between 1982 and

1990, Oliver and Mahoney identified 165 different journals referenced in 72 articles. It was therefore deemed important to include journals of other disciplines in the content analysis in order to gain a comprehensive view of the treatment of correlated data in C&T research.

Reports of original research published between 1987 and 1993 were considered. This seven-year period was selected in order to allow for the identification of possible trends in RM analyses adopted by researchers in the C&T field.

All articles in the four journals considered central to the C&T discipline were individually reviewed for their relevance to the current research project. Computerized literature searches were conducted to identify C&T RM research reports published in other journals. The primary tool for these searches was the Clothing and Textile Arts CD-ROM. This index encompasses English-language serial literature published between 1970 and 1992 that deals with clothing as a form of human behaviour, as well as textile and apparel arts. With respect to the latter, only applied textile science, as opposed to pure textile science topics, are included since the focus of the C&T discipline is the relationship between humans and either apparel or textile products (Kaiser & Damhorst, 1991). Each entry in the database is accompanied by a brief summary, which aided in the identification of articles employing RM methodologies.

Other CD-ROM data bases were searched, as the Clothing and Textile Arts CD-ROM does not cover the most recent year of the designated time period and may not comprehensively cover all published literature relevant to the current study. These data bases included PsychLit, Medline, and the Science Citation Index. All entries in the first source are accompanied by abstracts, as are the majority of entries in the second; the Science Citation Index does not provide abstracts.

Where a summary or abstract was not available, or where details of the methodology were not clearly defined in a summary or abstract, the original article was consulted to determine whether a RM design had been used in conducting the research. Articles from journals not found in the University of Manitoba library system were excluded unless a summary or abstract specifically indicated that a RM design had been used. In the latter case, the articles were obtained through the interlibrary loan system.

Sixteen major content areas are used to categorize entries in the Clothing and Textile Arts CD-ROM: consumer, social-psychological, clothing selection, functional, energy, industry, textile science, historical, merchandising, clothing fabrication, handicapped, cultural, costume design, medical, professional issues, and textile design. Because each entry may be classified using more than one subject identifier, research articles contained in each of the 16 subject areas were reviewed for the period 1987 to 1992. The PsychLit, Medline, and Science Citation indices were searched for selected time period using a variety of key terms, including dress, clothing, apparel, fashion, uniforms, textile(s), and fabric.

The following criteria were used to identify relevant articles in these literature searches:

1. Only research encompassing the study of clothing or textile products (i.e., bedding, window coverings, carpets) was considered. For the purpose of this research, clothing is defined as any textile item which covers all or part of the body and contains as least one seam. Schlick (1991) defines two categories of clothing: (a) that which covers the torso, including undergarments, outergarments, and overgarments; and (b) that which covers body extremities (leg/foot, head/neck, hand/arm). No RM research articles which considered footwear were identified in the literature search. Several which focussed on headgear were retrieved, but most were excluded as they dealt with helmets, and therefore were not contained within the boundaries of the definition of clothing adopted for this study. A small number of research reports on gloves were retrieved, and most of these were included in the investigation. Finally, no RM research which focussed solely on accessories, where an accessory is defined as an item worn as decoration or carried in addition to the garment (Schlick, 1991) was identified.

2. Research articles not contained in the four core journals were not included unless at least one clothing variable was manipulated. This resulted in the exclusion of most of the studies on physical attractiveness, self-concept, and body perceptions, as well as many articles found in medical, agricultural, and environmental journals. For example, entomological research dealing with protection of humans from insects frequently mentions clothing, but generally

is not concerned with the effect that clothing has on insect repellency. Such articles were not considered to fall within the realm of the C&T discipline.

3. If the data base entry identified the citation as an abstract, or as coming from the popular serial literature (i.e., Psychology Today, Consumer Reports), the citation was excluded.

4. Articles in which all of the RM factors had only two levels were excluded. In such cases, the sphericity assumption is trivially satisfied, and the researcher need not make a decision among the various univariate and multivariate procedures for data analysis.

Data Coding

A data base of C&T RM research articles was established using the Pro-Cite software package (Personal Bibliographic Software, Inc., 1992). Each article was perused to identify specific information considered essential to characterizing the features of the literature. Information obtained for each article was recorded in a single data base entry.

Citation information. Citation information, including article title, author name(s), year of publication, journal, volume and issue numbers, and page numbers was recorded. As well, the education unit, research unit, or company affiliation of the first author was recorded; if this was not provided, the affiliation of the author responsible for requests for reprints was recorded.

Subject area. The introduction of each article was perused to identify the subject area of the research. Initially, the retrieved articles were classified on the basis of content using a scheme developed by Oliver and Mahoney (1991), which contained the following categories: apparel design and manufacturing, consumer issues, cultural/historical, educational, merchandising, social/psychological, and textile science. However, the classifications created by these authors were too broad, as most of the C&T RM articles fell within the boundaries of the categories of consumer issues, social/psychological, and textile science. Not all of the content areas defined by Oliver and Mahoney were represented among the retrieved articles since RM methodologies are not appropriate for addressing all C&T research problems (e.g., cultural/historical research). A more detailed categorization scheme was needed to provide specific information on the research areas in which RM methodologies are used.

A revised scheme was developed from the work of Kaiser and Damhorst (1991). These researchers identified three global content areas of C&T research through a survey of the membership of the International Textile and Apparel Association: (a) textile product evaluation, which "emphasize[s] the connections between product attributes or properties and human responses to these tangible characteristics" (p. 4); (b) appearance and social realities, which "connects human use of textiles, clothing, and related artifacts with human perceptions of the social order--how everyday life is defined, shaped, and organized on the basis of social relationships and meanings" (p. 5); and (c) textile and apparel production/distribution systems, which is "concern[ed] with [the] relationship of one product (e.g., fiber) to another (e.g., fabric) throughout the product pipeline, culminating with the purchase of apparel or other textile-related end-products by consumers" (p. 5).

Kaiser and Damhorst (1991) identified common research topics that fall within the boundaries of each of these content areas. Using these topics as a basis, a detailed method of categorizing research topics was created specifically for use in the current context. The content areas, along with their definitions, which were created by the author, were:

1. Textile Product Evaluation

(a) Quality: Objective or subjective evaluations of the structural or visual integrity of clothing or textile products; Use of brand name as a cue to quality.

(b) Performance: Objective or subjective evaluations of mechanical, physical, chemical or biological properties of clothing or textile products.

(c) Care/Maintenance: Responses of clothing or textile products to laundering or drycleaning; Care labelling; Evaluations of detergent properties.

(d) Comfort: Physical sensations of clothing or textile properties, including tactile, thermal, moisture, and motion sensations; Psychological comfort.

(e) Protective Clothing: Evaluations of performance or comfort of functional clothing intended for personal protection in specialized work environments.

2. Appearance and Social Realities

(a) Fashion: Fashion awareness and acceptance; Fashion opinion leadership; Evaluations of clothing fashionability.

(b) Aesthetics: Subjective evaluations of liking or attractiveness of clothing or textile products.

(c) Social Judgments: Perceptions of gender orientation, age, social class, or group/organizational membership via clothing cues.

(d) Character Judgments: Assessments of personal attitudes, beliefs, or values via clothing; Effect of clothing on self-concept.

(e) Occupational Perceptions: Evaluations of employment characteristics and job suitability via clothing cues; Use of uniforms to identify occupational status; Effect of clothing on perceptions of occupational skill and ability.

3. Textile and Apparel Production/Distribution Systems

(a) Retail Operations: Store buying and selling operations; Retail personnel; Consumer perceptions of stores and/or store brands.

(b) Marketing: Consumer perceptions of advertising and promotional strategies; Effect of advertising campaigns on consumer buying practices; Perceptions of the marketability of new product innovations.

Each article was classified according to the major subject area. Additionally, where the research topic overlapped content areas, a second subject classification was used.

Research design. Three categories were used to define the type of research design: simple, factorial, and mixed. The design was further categorized as either univariate (i.e., involving a single dependent variable) or multivariate (i.e., involving multiple dependent variables). In articles where the details of more than one study, or more than one phase of a research project were reported, each type of research design was noted. For both factorial and mixed designs, the number of RM factors was recorded. As well, the number of levels of each RM factor was identified for all three types of designs. Finally, for mixed designs, both the number of independent groups factors, and the number of levels of each such factor were recorded.

Information was obtained regarding the total number of units of analyses for which data was collected. Every attempt was made to identify the final size of the sample, as this number sometimes differed from initial sample size due to the presence of missing data. This was particularly evident in research articles which employed a survey format.

Mixed designs were further classified as either balanced or unbalanced. Wherever possible, the number of units of analysis in each group (cell) was also recorded, and this information was used to quantify the degree of group (cell) size imbalance, using a coefficient of variation (e.g., see Box, 1954, p. 300). For example, in the A x B mixed design described previously in this chapter, this coefficient is given by

$$\Delta n_j = \frac{\sqrt{\sum_{j=1}^{J} (n_j - \bar{n})^2 / J}}{\bar{n}}, \tag{2.9}$$

where $\bar{n}$ is the average group size.

Data analysis. Information concerning methods for testing hypotheses involving RM main, interaction, and simple effects was obtained through examination of the Method and Results sections of each research article. As Latour and Miniard (1983) and Ekstrom et al. (1990) observed, most applied researchers adopt the conventional ANOVA, conservative ANOVA, df-adjusted ANOVA, or MANOVA procedures for testing RM effects. Additionally, some researchers will employ RM designs containing quantitative covariates, and the data from such designs may be analyzed using conventional analysis of covariance (ANCOVA), conservative ANOCVA, df-adjusted ANCOVA, or multivariate analysis of covariance (MANCOVA) techniques. Where a df-adjusted ANOVA or ANCOVA F test was used, it was noted whether the $\hat{\epsilon}$, $\tilde{\epsilon}$, or $\bar{\epsilon}$ adjustment factor was adopted. For RM designs which are multivariate in nature, and which are analyzed as such, multivariate MANOVA or MANCOVA procedures may be used.

Nonparametric and trend analysis procedures may also be used to test RM effects (Maxwell & Delaney, 1990). The former approach is often adopted when the researcher is unwilling to assume that the data satisfy the assumption of multivariate normality, or when the data are comprised of ranks. The latter may be used when the levels of the RM factor(s) represent quantitative, rather than qualitative, differences in the presence of an experimental treatment.

All of the previously described analysis procedures are appropriate when the dependent variables are continuous in nature or are treated as continuous, or where rank scores are obtained in the case of a nonparametric procedure. In some situations, responses for a particular dependent variable may represent frequencies. Methods for testing hypotheses involving proportions might include z tests or chi-square tests of independence or association (Glass & Hopkins, 1984).

Information pertaining to the use of MCPs for testing hypotheses concerning pairs of means was recorded. The specific strategy adopted to control either the FWR or the PCR was noted, as was the use of either a pooled or nonpooled test statistic.

Based on previous research (Jaccard et al., 1984; Lix, 1995) the following procedures were deemed most likely to be represented among the RM articles, and are briefly described here for purposes of clarity:

1. Multiple t tests: Each pairwise comparison t statistic is evaluated at the α level of significance using the CV, $t[1 - \alpha/2; \nu]$, the $1 - \alpha/2$ centile of Student's t distribution with ν df, where ν is the error df.

2. Scheffe (1953): Each pairwise comparison t statistic is evaluated with the CV, $\{(K - 1)F[1 - \alpha; K - 1, \nu]\}^{1/2}$, where K is the number of means in the family, and $F[1 - \alpha; K - 1, \nu]$ is the $1 - \alpha$ centile of the F distribution with $(K - 1)$ and ν df.

3. Bonferroni (Dunn, 1961): The t statistics are evaluated for significance using $t[1 - \alpha/(2C), \nu]$, where C represents the number of pairwise comparisons in the family and ν is as previously defined.

4. Fisher's (1935) Least Significant Difference (LSD): This procedure begins with an omnibus test of the null hypothesis. If this hypothesis is rejected, multiple t tests are conducted for all possible pairs of means (see #1); otherwise testing stops.

5. Tukey's (1953) Honestly Significant Difference (HSD): The CV used in hypothesis testing is $q[1 - \alpha, K, \nu]/\sqrt{2}$, the $1 - \alpha$ centile of the Studentized range distribution, where K and ν are as previously defined.

6. Duncan's (1955) Multiple Range: This method involves a stepwise approach to hypothesis testing. In a set of K means, one begins by ranking the means in ascending order. The CV used in assessing whether two means are significantly different is $q[(1 - \alpha)^{p-1}, p, \nu]/\sqrt{2}$, where p represents the number of steps between ordered means and ν is as previously defined. Thus, the significance level varies as a function of p. The hypothesis associated with the largest pairwise difference, which corresponds to means that are said to be $p = K$ steps apart, is tested first. Successive pairs of ordered means are tested for statistical significance only if they are contained within the range of a previously rejected hypothesis, otherwise they are declared nonsignificant.

7. Newman-Keuls (Keuls, 1952; Newman, 1939): This method also involves a stepwise approach to hypothesis testing, but unlike Duncan's (1955) method,

the level of significance is not modified according to the value of p. Consequently, the CV used is $q[\alpha; p, \nu]/\sqrt{2}$. The sequence of hypothesis testing is the same as that described for Duncan's method.

Multiple t tests allow for control of the PCR whereas the Bonferroni (Dunn, 1961), Scheffe (1953), and Tukey (1953) HSD MCPs control the FWR. While the Fisher (1935) LSD, Duncan (1955) Multiple Range, and Newman-Keuls (Keuls, 1952; Newman, 1939) procedures are popular among researchers, it is known that none of these can limit the FWR to α (Lix, 1995).

Independent and dependent variables. Information concerning the independent and dependent variables investigated in each study was collected as a means of providing more detailed information on the types of C&T research problems addressed using RM methodologies. As well, the manner in which the dependent variable was operationalized (i.e., continuous versus Likert scale) was noted, as were the number of response points in the case of Likert scales.

Additional information. Problems of assumption violations noted by the researchers, preliminary tests for violations of derivational assumptions, and additional comments pertaining to the data analysis were also recorded, including citations of specific statistical reference materials.

CHAPTER 3

RESULTS OF A CONTENT ANALYSIS OF THE CLOTHING AND TEXTILES REPEATED MEASURES LITERATURE

A total of 101 C&T research reports which employed RM methodologies were retrieved from the literature through a search of various C&T journals and CD-ROM data bases. The findings of the content analysis are presented in this chapter.

Citation Information

The search of the C&T RM literature extended from 1987 to 1993 inclusive. An average of 14 research reports were obtained for each year in this period (SD = 2), with a range from 18 in 1988 to 11 in 1992. In each of these articles, at least one RM factor had more than two levels.

Table 1 contains information pertaining to the journals in which the C&T RM articles were published. One third were found in three journals which are central to the C&T discipline: Clothing and Textiles Research Journal, Home Economics Research Journal, and Journal of Consumer Studies and Home Economics. The Canadian Home Economics Research Journal was not represented among the identified research reports. As expected, the Clothing and Textiles Research Journal contained the greatest number of C&T RM articles (f = 24).

Two thirds of the articles were published in 21 different journals that are peripheral to the discipline. The largest number of these reports (f = 13) were

Table 1

Frequency of RM Research Reports by Journal

Journal	f
C&T Core Journals	
Clothing and Textile Research Journal	24
Journal of Consumer Studies and Home Economics	5
Home Economics Research Journal	4
Peripheral Journals	
Aviation, Space, and Environmental Medicine	13
Ergonomics	12
Perceptual and Motor Skills	12
American Industrial Hygiene Association Journal	6
Textile Research Journal	6
European Journal of Applied Physiology	2
Journal of Applied Social Psychology	2
Journal of the Textile Institute	2
Archives of Environmental Contamination and Toxicology	1
ASHRAE Transactions	1
Empirical Studies of the Arts	1
Hospital and Community Psychiatry	1
International Archives of Occupational and Environmental Health	1
Journal of Early Adolescence	1
Journal of Interdisciplinary Cycle Research	1
Journal of Police Sciences and Administration	1
Journal of Social Behavior and Personality	1
Journal of Sports Sciences	1
Medicine and Science in Sports and Exercise	1
Psychotherapy	1
The Physician and Sportsmedicine	1
TOTAL	101

published in Aviation, Space, and Environmental Medicine, however, almost an equal number were found in Ergonomics and Perceptual and Motor Skills (f = 12, respectively).

The affiliation of the first or primary author was recorded for each of the research articles contained in the data base. This information was rather difficult to categorize given the variety of education and research units which were listed. However, on the basis of the author address provided with each article, it was possible to discern that 49 of the 101 articles had a first or primary author associated with a home economics/human ecology, consumer studies, or C&T education or research unit. An additional 15 of the articles were associated with a primary or first author from a health, recreation, physiology, or kinesiology unit, and 13 were associated with a defense, aviation, naval, or army research unit. Smaller numbers of researchers were affiliated with psychology, psychiatry, communications, occupational health and safety, engineering, human development, marketing, and statistics. For three articles, the first or primary author was affiliated with a private company. Finally, it should be noted that while slightly less than half of the articles had a first or primary author with a home economics/human ecology, consumer studies, or C&T background, several of the studies were collaborative, and therefore one or more individuals with such an affiliation may have been a part of the research team.

Subject Area

Table 2 contains information pertaining to the classification of the C&T RM articles by subject area. As well, a cross-classification by subject for the four journals containing the greatest number of articles is included.

The information contained in Table 2 reveals that the majority of the articles (i.e., f = 70) dealt with the general topic of textile product evaluation. Within this category, the greatest attention was directed at RM research problems dealing with protective clothing (f = 25), although comfort topics were also frequently studied (f = 20). Many of the articles pertaining to textile product evaluation were contained in Aviation, Space, and Environmental Medicine and Ergonomics. However, neither comfort nor performance research was concentrated in any single journal.

Another 34 articles dealt with appearance and social realities as they relate to clothing. The greatest number of these articles (f = 10) focussed on the use of clothing in making occupational perceptions. The occupational perceptions articles tended to be scattered across a variety of different journals, while the remainder of the research reports in the appearance and social realities category were contained in either the Clothing and Textiles Research Journal or Perceptual and Motor Skills. Finally, only ten of the 101 research reports focussed on textile and apparel production and distribution systems and the majority of these were published in the Clothing and Textiles Research Journal. Table 2 reveals that every subject area but care/maintenance was represented in the primary journal of the field, that is, the Clothing and Textiles Research Journal. Most of the RM articles (f = 5) pertaining to

Table 2

RM Research Reports by Subject Area and Major Journal

Subject Area	TOTAL (N = 101)	CTRJ (f = 24) f	CTRJ %	ASEM (f = 13) f	ASEM %	ERG (f = 12) f	ERG %	PMS (f = 12) f	PMS %
Textile Product Evaluation	70	8	33	14	108	14	117	--	--
Protective Clothing	25	1	4	9	69	6	50	--	--
Comfort	20	2	8	3	23	5	4	--	--
Performance	15	2	8	2	15	3	25	--	--
Care/Maintenance	6	--	--	--	--	--	--	--	--
Quality	4	3	12	--	--	--	--	--	--
Appearance and Social Realities	34	16	67	--	--	--	--	9	75
Occupational Perceptions	10	2	8	--	--	--	--	2	17
Character Judgments	9	6	25	--	--	--	--	2	17
Aesthetics	8	3	12	--	--	--	--	4	33
Fashion	4	2	8	--	--	--	--	1	8
Social Judgments	3	3	12	--	--	--	--	--	--
Production/Distribution Systems	10	6	25	--	--	--	--	3	25
Marketing	6	2	8	--	--	--	--	3	25
Retail Operations	4	4	17	--	--	--	--	--	--

Note: CTRJ = Clothing and Textiles Research Journal; ASEM = Aviation, Space, and Environmental Medicine; ERG = Ergonomics; PMS = Perceptual and Motor Skills; Column totals may exceed the specified N or f and percentages may exceed 100 because each article could be classified in up to two categories.

this one topic were found in either the Home Economics Research Journal or Journal of Consumer Studies and Home Economics.

Repeated Measures Design Characteristics and Analysis Procedures

An investigation of the types of RM research designs which were adopted to investigate C&T research problems revealed that in 22 of the articles one or more simple RM designs was used, another 38 articles incorporated at least one factorial RM design, and 43 articles reported analyses associated with at least one mixed design. Since some articles reported the results of more than a single study or more than one phase of a project, the total number of articles classified by research design exceeds the total of 101. In addition, it is important to note that one article in which the authors stated that a simple RM design had been used, revealed, upon closer inspection, to involve a one-way independent groups design with multiple dependent variables; this article was excluded from further analysis.

Table 3 contains a cross-classification of articles by subject area for the three different types of RM designs. This table reveals that simple and factorial RM designs were used proportionately more often to investigate textile product evaluation research problems than were mixed designs. The latter were more often used for addressing research questions in the area of appearance and social realities, particularly with respect to aesthetics and character judgments. A higher percentage of articles which focussed on apparel production and distribution systems reported the use of simple RM designs rather than factorial or mixed designs. Finally, the majority of the occupational perceptions articles (60%) reported the use of factorial RM designs.

Table 3

RM Research Reports by Subject Area and Research Design

	Simple (f = 22)		Factorial (f = 38)		Mixed (f = 43)	
Subject Area	f	%	f	%	f	%
Textile Product Evaluation	16	73	29	76	26	60
Protective Clothing	8	36	12	31	6	14
Comfort	1	5	12	31	8	19
Performance	3	14	4	11	8	19
Care/Maintenance	2	9	1	3	4	9
Quality	2	9	--	--	--	--
Appearance and Social Realities	7	32	9	24	18	42
Occupational Perceptions	1	4	6	16	3	7
Character Judgments	3	14	1	3	5	12
Aesthetics	1	4	1	3	6	14
Fashion	2	9	1	3	1	2
Social Judgments	--	--	--	--	3	7
Production/Distribution Systems	4	18	1	3	4	9
Marketing	2	9	1	3	3	7
Retail Operations[a]	2	9	--	--	1	2

Note: See the note from Table 2.
[a]One article in this category was deleted from the analysis because it used an independent groups design, not a repeated measures design as indicated by the authors.

Simple Repeated Measures Designs

Table 4 contains information pertaining to the characteristics of the research reports which incorporated at least one simple RM design. For this subset of 22 articles, the number of levels of the RM factor ranged from three to 15, with four being the most common number. In eight of the research articles, all of which focussed on protective clothing, the functional qualities of multiple garments were evaluated in wear trials. In another five articles, all of which were contained in the appearance and social realities category, participants assessed the overall image projected by a stimulus model in various clothing styles as a means of studying the nonverbal cues provided by dress. A variety of other RM factors were used in the remaining studies, including fabric type (i.e., variation in fibre content) and laundry detergent type. It is also interesting to note that only one study specifically indicated that the order of presentation of treatments to units of analysis had been randomized to remove possible sequence effects; in this instance, study participants were asked to evaluate mutiple types of fabrics in a clothing comfort study.

In all 22 simple RM articles the units of analysis were human subjects. Total sample size ranged from four to 604; the latter value was associated with a study in which the RM factor had 14 levels. This sample size seems excessively high, given the power advantage that may be achieved by adopting a RM design instead of an independent groups design. Not surprisingly, almost all of the statistical tests reported in this study were significant at the selected criterion of significance (i.e., $\alpha = .05$).

Table 4

Profile of Simple RM Designs (f = 22)

Variable	f	%	Variable	f	%
Number of RM Factor Levels			Analysis Procedure		
3	5	23	Conventional ANOVA	4	18
4	8	36	Conventional ANCOVA	1	4
5	6	27	Conservative ANOVA	1	4
6	3	14	MANOVA	3	14
> 6	3	14			
			Descriptive Analysis Only	4	18
Total Sample Size			Incorrect Analysis	3	14
≤ 10	8	36	Not Stated	3	14
11 - 25	2	9	No Omnibus Analysis	2	9
26 - 50	4	18			
51 - 100	4	18	Other	3	14
> 100	5	23			
			Pairwise MCP		
Nature of the Design			Multiple t Tests	3	14
Univariate	7	32	Bonferroni	2	9
Multivariate	16	73	Tukey HSD	2	4
			Duncan's Multiple Range	1	4
			Fisher's LSD	1	4
			Newman-Keuls	1	4
			Not Stated	2	9

Note: Frequencies for each variable may exceed the total frequency and percentages may exceed 100 because some research articles reported the results of more than one study, or more than one phase of an analysis.

However, the author did not consider that the power to detect an effect, regardless of the size of the effect, would essentially be equal to one. In the one study in which total size was equal to four, no statistical tests were performed, perhaps because the authors felt that the results of such tests would not be meaningful given the small number of observations on which they would be based. Instead, graphical plots of the data were used to descriptively analyze the results.

The majority of the simple RM research reports involved the investigation of more than one dependent variable. For example, in the wear studies, participants typically provided physiological and psychological response data. The former includes such variables as skin and rectal temperature while the latter includes variables such as thermal and sweat sensation. In the image evaluation studies, multiple personal traits (i.e., intellect, sociability, professionalism) were usually assessed for each style of dress, typically using 5-point Likert scales.

While multivariate designs were used in the majority of the articles, in none of these was a multivariate approach to data analysis adopted. Instead, researchers conducted separate analyses for each of the dependent variables. While a wide variety of analytic techniques were adopted in the 22 simple RM articles, four of these used no inferential statistics, and instead only reported results associated with a descriptive analysis, such as means and standard deviations or frequencies and percentages. Graphical plots were also used in some of these articles to describe the data.

The conventional ANOVA approach, which assumes that the data conform to the sphericity assumption, was used to test at least one omnibus hypothesis in four articles; in one of these articles the conventional ANCOVA F test was also used to test selected hypotheses. The three articles which did not clearly specify a method of analysis also appeared to incorporate the conventional approach to testing the RM effect, as all three studies reported that an ANOVA method had been used. However, this could not be confirmed due to insufficient details of the methodology (i.e., df were not reported). In three articles the authors adopted Hotelling's (1931) T^2 procedure for testing an omnibus hypothesis.

For two of the three articles in which the correlated data was incorrectly analyzed, the RM factor was erroneously treated as an independent groups factor in the computation of the omnibus F statistic. Furthermore, in one of these articles the design was also flawed, as some, but not all, subjects provided more than one set of data (i.e., more than one replication). In the third article which incorporated an incorrect analysis, the authors acknowledged that the data were obtained from a RM design, but they chose to ignore this in the analysis and conducted statistical tests appropriate for an independent groups factorial design.

Of the three articles which were classified in the other category, two reported the results of correlations among several dependent variables at each level of the RM factor. One reported the results of a select number of complex contrasts, which were conducted in addition to a test of the omnibus hypothesis using MANOVA.

In two articles, no omnibus analysis was conducted; the authors proceeded directly to pairwise mean comparisons using multiple t tests. However, in 56% (f = 10) of the 18 articles in which inferential analyses were performed, the omnibus test was followed by pairwise comparisons; the most popular MCPs for conducting these tests were the Bonferroni (Dunn, 1961) and Tukey (1953) HSD methods. In two articles the authors indicated that the means were probed using a MCP, however, the procedure adopted for controlling the Type I error rate was not specified.

Three quarters of the 12 articles in which a MCP was used did not contain sufficient detail to determine whether a pooled or nonpooled t statistic had been adopted. However, of the remaining studies, two used a nonpooled statistic, and only one used a pooled statistic.

It should be noted that in two of the 22 simple RM research reports, the omnibus analysis was preceded by a test of the sphericity assumption. In one case, the authors reported the use of Bartlett's (1937) technique and a significant result was obtained; the author elected to use MANOVA to test the omnibus hypothesis. However, in this article it is not clear that the author understood the rationale for this test, as it was reported that although a significant result was obtained, the ANOVA F test is known to be robust to violations of the variance homogeneity assumption if group sizes are equal. Since only a single group of units of analysis is under investigation in a simple RM design, this explanation is not relevant. In the second article, the preliminary test of the sphericity assumption was nonsignificant and the authors elected to use the conventional ANOVA approach. Finally, in another research report, the authors' concern over possible violations of the sphericity and multivariate normality assumptions led them to transform the data prior to analysis.

Factorial Repeated Measures Designs

Tables 5 and 6 contain information pertaining to the characteristics of articles incorporating at least one factorial RM design. As Table 5 reveals, of the 38 research articles contained in this subset, the individual or joint effects of two RM factors were studied in 71% of the articles, three RM factors were considered in nine reports, and more than three RM factors were investigated in only two articles. A count of the number of RM factor levels revealed that the factorial studies typically involved at least one RM factor with only two levels, in addition to one or more factors with three or more levels. Most often, the levels of one RM factor represented styles of clothing, particularly protective clothing, which were evaluated in wear trials ($f = 15$). Often, at least one additional RM factor was time ($f = 16$), yet in the majority of these articles, time was treated as a fixed effects factor rather than a random effects factor. Also popular in the wear studies was the investigation of various styles of protective gear under different environmental conditions (e.g., variations in temperature and/or humidity) or different levels of intensity of human activity. In six of the articles study participants were involved in making image evaluations of the impressions conveyed by multiple styles of dress within an occupational context; all of these studies contained only two factors, and the second one typically related to the characteristics of the stimulus model (e.g., sex, age, body type). Overall, in only two articles did the authors indicate that use of random

Table 5

Profile of Factorial RM Designs: Design Variables (f = 38)

Variable	f	%
Number of RM Factors		
2	27	71
3	9	24
4	1	3
> 4	1	3
Number of RM Factor Levels		
2	29	76
3	20	53
4	15	39
5	5	13
6	4	10
> 6	13	34
Total Sample Size		
≤ 10	25	66
11 - 25	3	8
26 - 50	4	10
51-100	3	8
> 100	3	8
Nature of the Design		
Univariate	12	32
Multivariate	26	68

Note: See the note from Table 4.

assignment of treatments to study participants. In both cases, different types of protective gear were evaluated.

In all but three articles, human subjects were the units of analysis. Textile laboratories were the unit of analysis in two articles and individual fabric samples were studied in the third. In the majority of the articles (66%) a total sample size of no more than ten units of analysis was reported, but total sample size varied considerably, from one to 300. However, while more than one quarter of the simple RM articles reported a total sample size greater than 100, less than 10% of those which incorporated a factorial RM design did so.

Like the simple RM articles, the majority of the articles which incorporated a factorial design reported the investigation of more than a single dependent variable. For example, in the wear studies, both physiological and psychological response data were collected from study participants. In the occupational perceptions research articles, a variety of skill-related and personality characteristics were considered (i.e., professionalism, competence, trustworthiness). Where Likert scales were used, the number of response points varied considerably, from three to 20. In the former case, a colour preference scale was involved. In the latter case, several 20-point scales were used to evaluate subjective responses on a number of dependent variables (e.g., clothing comfort, clothing temperature) in a wear study; the authors noted that neither validity nor reliability of the scales had been previously established and they did not conduct such assessments as part of their investigation.

As Table 6 reveals, unlike the simple RM articles, three of the factorial RM articles incorporated a multivariate approach to RM analysis. In all cases this analysis was conducted for a main effect factor with only two levels, which means that the local multivariate multisample sphericity assumption was trivially satisfied. For all three studies a significant main effect test was followed by univariate analyses (i.e., paired t tests) for each dependent variable.

A univariate approach to data analysis was used in the majority of the factorial RM articles and, as anticipated, the conventional ANOVA F test was the most popular among all of the methods (24%). Additionally, of the nine research reports in which adequate details of the analysis strategy were not provided, all appeared to incorporate the conventional ANOVA approach, but again, this information could not be clearly determined due to a lack of details.

Unlike the simple RM articles, none of those incorporating a factorial design reported the use of MANOVA for testing correlated effects. However, in four articles, df-adjusted tests were conducted; the $\hat{\varepsilon}$ correction factor was adopted in three and in one, the $\tilde{\varepsilon}$ statistic was used.

The RM data from four articles were analyzed using other methods. In two of these, due to extremely small total sample sizes (i.e., two or less), the authors reported the use of small sample F and t statistics to test main effects; however, due to a lack of details of these statistics, it is not clear how they were computed. In another research report, frequency data were collected, and Fisher's (1935) z test was

Table 6

Profile of Factorial RM Design: Analysis Variables (f = 38)

Variable	f	%
Analysis Procedure		
Conventional ANOVA	9	24
DF-Adjusted ANOVA	4	10
DF-Adjusted ANCOVA	1	3
Multivariate MANOVA	3	8
Not Stated	9	24
Descriptive Analysis Only	5	13
No Omnibus/Simple Effect Analysis	4	10
Incorrect Analysis	1	3
Trend Analysis	1	3
Other	4	10
Effect Tested		
Marginal Main	21	55
Interaction	10	26
Simple Main	15	39
Simple Interaction	2	5
Cell Means	1	3
Pairwise MCP		
Tukey HSD	6	16
Multiple t Tests	4	10
Fisher's LSD	4	10
Bonferroni	1	3
Duncan's Multiple Range	1	3
Scheffe	1	3
Nonparametric	1	3
Not Stated	4	10
Effect Probed		
Marginal Main	13	34
Simple Main	14	37

Note: See the note from Table 4.

used to test for differences in frequencies between the two levels of one factor at each level of the second factor. In the fourth study, a single complex contrast was applied to test a hypothesis for the data associated with one RM factor following tests of main and interaction effects using the conventional ANOVA approach.

It is surprising to note that of the 38 factorial RM articles, only ten specifically reported that tests of interaction effects had been conducted. However, it may be that tests of interaction effects were conducted, but due to nonsignificance, were not reported. Marginal main effect test results were provided in 55% of the articles. Seven of the 22 studies (18%) provided the results of simple main effect tests, but no results for omnibus tests.

As expected, tests of simple main effects were often conducted following a significant interaction (f = 6). In two studies simple interaction effect tests were conducted; in one, these tests were used to probe a significant three-way interaction and in the other, tests of simple interaction effects were conducted instead of a test of the three-way interaction. Finally, in one article, after the authors performed tests of simple effects, the factorial structure of the data was reduced to a one-way model, and a test for an overall effect was conducted on the cell means.

The data were incorrectly analyzed in only a single research report. In this case, the researchers erroneously treated the data as though it were obtained from a mixed design, rather than a factorial RM design, in the computation of main and interaction effect test statistics.

In half of the RM research articles in which a factorial design was used, pairwise comparisons were used to probe the source of a marginal or simple effect. In four articles (18%), however, the authors bypassed both omnibus and simple effect tests in favour of multiple t tests on either marginal or simple pairs of means. In another study the nonparametric Wilcoxon sign rank test was used to evaluate pairwise differences among the RM factor levels (Marascuilo & McSweeney, 1977), and was not preceded by an omnibus nonparametric analysis. The PCR was controlled for these nonparametric comparisons at $\alpha = .05$. Tukey's (1953) HSD procedure was most frequently adopted to control the FWR (27%). Fisher's (1935) LSD was also a popular procedure (18%). Finally, in four research reports the specific procedure used to control the error rate for multiple pairwise comparisons was not given.

The authors of the great majority of research reports in which pairs of means were probed to identify the localized source of an effect did not indicate whether a pooled or nonpooled error term had been adopted (i.e., f = 19). Four of the studies, however, did clearly indicate that an error term which pooled across none of the factors was selected; such an approach does not assume that either local or overall sphericity assumptions are satisfied.

One set of authors acknowledged the potential for an inflated FWR when conducting multiple omnibus tests, as a result of performing separate univariate analyses for each of several dependent variables in a multivariate design. Consequently, Scheffe's (1953) method was adopted to control the error rate for the entire family of main effect tests. However, it is surprising to note that these same authors proceeded to conduct multiple t tests following a significant omnibus result (i.e., Duncan's LSD method), adopting an $\alpha = .05$ significance level for each t test.

Mixed Designs

Tables 7 and 8 contain information pertaining to the characteristics of the 43 research reports in which at least one mixed design was used. As Table 7 reveals, one third of the mixed designs were two-way designs involving a single grouping factor and a single repeated measures factor and 35% were four-way designs. Half of the articles reported the use of a mixed design with only a single grouping factor and this factor typically had only two levels. In two thirds of the articles a mixed design which had only one RM factor was used; this factor frequently had three levels. However, one third of the articles incorporated a mixed design in which at least one RM factor had more than six levels; the maximum value was 14.

The variable most frequently used for classifying study participants into independent groups was sex (f = 9), although a variety of demographic and personal attributes were used, including geographic location, age, income, college major, height, and weight. In four articles, participants were assigned to groups on the basis of scores obtained on indices designed to measure fashion leadership, self-monitoring style, fashion type, or psychological type. In 17 articles, participants were randomly assigned to independent experimental treatment groups.

In 11 of the 43 mixed articles, the levels of the RM factor represented styles of dress designed to portray different images. Time was rarely used as a RM factor in

Table 7

Profile of Mixed Designs: Design Variables (f = 43)

Variable	f	%	Variable	f	%
Number of Study Factors			Number of RM Factor Levels		
2	14	33	2	10	23
3	10	23	3	22	51
4	15	35	4	10	23
> 4	7	16	5	3	7
			6	6	14
Number of Grouping Factors			>6	12	28
1	22	51			
2	15	35	Total Sample Size		
> 2	9	21	≤ 10	6	14
			11 - 25	7	16
Number of Grouping Factor Levels			26 - 50	4	9
2	33	77	51 - 75	7	16
3	10	23	76 - 100	3	7
4	8	19	101-150	2	5
> 4	7	16	> 150	11	26
Not Stated	2	5	Not Stated	6	14
Number of RM Factors			Nature of Group Size Equality/Inequality		
1	29	67	Balanced	14	32
2	12	28	Unbalanced	25	58
> 2	5	12	Not Stated	7	16

Note: See the note from Table 4.

the mixed designs (f = 4). In six articles the authors noted that randomization was used in assigning RM factor levels to study participants; three of the studies were concerned with image evaluations of multiple dress styles, two dealt with comfort of different clothing styles, and one focussed on protective gear.

A noteworthy point is that in seven of the research articles the units of analysis were fabric samples. In all cases, the performance of different fabrics were evaluated during multiple laundering or weathering cycles.

Total sample size varied considerably, from four units of analysis to more than 1000. As is evident from Table 6, the majority of the mixed designs were unbalanced. In several instances, this imbalance was implied from the study format (i.e., survey), and specific values of group (cell) sizes were not given. For those studies in which this information was provided, the coefficient of variation of group (cell) size inequality ranged in value from .056 to .862. For the study in which the former value was obtained, the ratio of the largest to the smallest group size was 16 to 15, while for the latter value, the ratio for cell sizes was 74 to 1.

Like the simple RM research articles, the majority of those which incorporated a mixed design were multivariate in nature (see Table 8). A much greater variety of dependent variables were investigated in mixed designs than in either the simple or factorial RM designs, and included clothing quality, attractiveness, and perceived intelligence in the image evaluation studies, and perceived comfort, thermal sensation, heart rate, and oxygen consumption in wear trials. Fabric samples were evaluated on

Table 8

Profile of Mixed Designs: Design and Analysis Variables (f = 43)

Variable	f	%	Variable	f	%
Nature of the Design			Pairwise MCP		
Univariate	17	40	Newman-Keuls	5	12
Multivariate	26	60	Fisher's LSD	4	9
			Scheffe	4	9
RM Analysis Procedure			Multiple t Tests	2	5
Conventional ANOVA	19	44	Tukey HSD	3	7
Conventional ANCOVA	2	5	Nonparametric	2	5
DF-Adjusted ANOVA	4	9	Duncan	1	2
MANOVA	2	5			
Multivariate MANOVA	1	2	Not Stated	5	12
Descriptive Analysis Only	8	19	RM Effect Probed		
Not Stated	4	9	Marginal Main	18	42
Nonparametric	3	7	Simple Main	13	30
Incorrect Analysis	2	5	Cell Means	2	5
No Omnibus Analysis	2	5			
Trend Analysis	2	5	Interaction MCP		
			Multiple t Tests	1	2
Other	1	2	Cicchetti	1	2
RM Effect Tested					
Marginal Main	25	58			
Interaction	15	35			
Simple Main	17	40			
Simple Interaction	3	7			

Note: See the note from Table 4.

such attributes as colour change, weight loss, and wrinkle resistance. Where Likert scales were used the number of scale points ranged from three to 14.

As Table 8 reveals, only one of the articles reported the adoption of a multivariate approach to data analysis. In this article, a significant test of a RM marginal main effect was followed by separate univariate analyses for each dependent variable using the conventional ANOVA F test. In keeping with the results reported for the simple and factorial RM designs, the most common method of analyzing correlated effects in mixed designs was the conventional approach, which was represented in 44% of the 43 research reports. In four articles, insufficient details were available to determine the method of analysis adopted, but in all cases the information provided by the author(s) would suggest that a conventional ANOVA approach had been selected. In four of the research reports a df-adjusted procedure was adopted, but the correction factor was not specified in any of these. MANOVA was applied to the analysis of the RM effect(s) in only two of the publications.

In the simple and factorial RM articles only parametric analyses of marginal or simple main or interaction effects were performed. This was not so for the mixed research reports, as three studies reported the use of Friedman's (1937) nonparametric test. In one of these studies the authors noted that a nonparametric analysis was adopted due to concern over possible violations of the assumptions associated with a parametric analysis; in another, the data collected were rank values and consequently a nonparametric analysis was necessary. At the same time, in this particular study the authors also performed parametric analyses on mean ranks, and thus were not consistent in their approach to the data. The last of the three articles also reported the results of an inconsistent analysis, as the authors noted that Friedman's test was adopted because the dependent variable was measured using a Likert scale, which constituted ordinal, rather than interval level data. However, the authors proceeded to descriptively analyze the study results using means.

Two of the mixed studies reported an incorrect analysis of the RM effect. In one, the RM factor was treated as an independent groups factor in the computation of ANOVA F statistics. In the other, the authors began by testing the RM main effect in an appropriate manner, using the conventional ANOVA approach. However, they then proceeded to treat each level of the RM factor as a separate dependent variable, and applied an independent groups MANOVA to test each of several independent groups effects. This was followed by independent groups ANOVA F tests.

Trend analysis was used in two studies, and was applied to test for a main effect trend or a simple main effect trend or both. The one study which was classified in the other category reported the use of z tests for analyzing frequency data.

Finally, it is important to note that four articles did not report any tests associated with a RM effect. In three of these, tests of simple independent group effects were conducted at each level of the RM factor or at each combination of levels of two or more RM factors. In the fourth, chi-square tests of independence were performed at each level of one RM factor.

The content analysis also revealed that for mixed designs, researchers typically were either only interested in testing marginal main effects, or only reported results associated with RM main effects. However, while only slightly more than one third of the research reports clearly identified that a RM interaction effect hypothesis had been tested, this was a greater percentage than for the single-group factorial studies, in which only 26% reported tests of interaction effects.

Among those articles in which the authors indicated that a significant interaction had been obtained (i.e., f = 11), seven reported follow-up tests of simple main effects, and in another one, simple interaction effects were used because the model contained more than two factors. In another article, significant two-way interactions in a four-way model were followed by pairwise comparisons among the cell means (i.e., the data were reduced to a one-way model).

It is surprising to note that the authors of four articles reported testing both main and simple main effects, but not interaction effects; another three reported tests of simple RM main effects only.

Significant main effect test results were routinely followed by pairwise mean comparisons. In only two research reports were marginal or simple main effect tests bypassed in favour of multiple t tests.

Among all of the procedures identified for conducting pairwise comparisons, the Newman-Keuls (Keuls, 1955; Newman, 1939) MCP was used with the greatest frequency (12%). However, in an equivalent number of articles the authors did not report sufficient details to identify the method for controlling the Type I error rate. The Scheffe (1953) and Fisher (1935) LSD procedures were also popular (9%, respectively).

Two of the mixed research reports reported specific tests to probe the interaction effect. In one, multiple t tests were applied to test interaction contrasts, this approach will be discussed in detail in the following chapter. However, this was also the study in which the data represented rank scores, and thus a parametric analysis can not be considered appropriate. Furthermore, the authors did not indicate that interaction contrasts were being conducted to probe an interaction. In another study, a method for probing interactions suggested by Cicchetti (1972) was adopted. Under Cicchetti's approach, unconfounded cell mean comparisons are conducted, and either a modified Tukey (1953) HSD or Scheffe (1953) approach is recommended for controlling the FWR. Unconfounded cell mean comparisons are those that involve pairs of means in the same row or column of a matrix of the individual factor level combination means. The authors of the study in which Cicchetti's approach was adopted did not describe which approach to FWR control was selected.

Finally, as with the single-group factorial research reports, it was rarely possible to determine whether a pooled or nonpooled multiple comparison test statistic had been adopted. Of the 18 studies in which a marginal main effect was probed, only one study clearly indicated that an error term which did pool across the levels of the RM factor was used. Furthermore, of the 13 articles in which a simple main effect was probed, only one study reported using a MCP in which the test statistic employed an error term that was not pooled across the levels of the independent groups factors.

As a final comment, the authors of one article reported the application of a test of the sphericity assumption to the data prior to testing any RM effects. Because this test was nonsignificant, the authors elected to adopt the conventional ANOVA approach to test RM main and interaction effects.

Synopsis of the Content Analysis

The results of this content analysis of the C&T RM literature communicate four distinct messages: (a) there is a great deal of diversity in the use of RM designs in the C&T literature, (b) researchers do not report sufficient details of their research methodology, (c) researchers continue to cling to traditional methods of RM analysis, and (d) there is a great deal of inconsistency in the analysis of RM effects, particularly with respect to the analysis of interaction effects.

The content analysis did reveal that researchers who investigate C&T research problems typically adopt RM designs which contain more than one experimental factor, as less than one quarter of all articles reported the results of an analysis associated with a simple RM design. Furthermore, the majority of the articles dealt with research questions related to the evaluation of textile or apparel products, particularly protective clothing. Beyond this, however, few similarities emerged. RM designs were not always used in the analysis of human subject data, and may be adopted, for example, in studies of fabric performance. With respect to the design of the studies, there was great variation in such characteristics as total sample size and the number of RM factor levels, and the types of independent and dependent variables under investigation. A diverse range of statistical analysis procedures were also used.

An additional noteworthy point is that it was not surprising to find that a large number of C&T RM articles were published in journals of other disciplines.

Given the interdisciplinary nature of C&T research, it frequently deals with subject matter appropriate for publication in many different types of journals. As well, investigators from other disciplines may often consider research problems which directly or indirectly focus on clothing or textiles, even though they may not have specific training in these areas.

Similar to the findings of Ekstrom et al. (1990), many of the articles lacked information about key aspects of the research projects. As Lavori (1990) notes, this is a severe problem because "clear exposition of design and analysis conveys a reassuring sense of mastery by the investigator, disarms critics, makes work useful and repeatable, and keeps us all from error" (p. 775). Several of the research reports contained insufficient details to accurately conclude what method of RM analysis had been adopted for testing a marginal or simple effect. Beyond this, there was a lack of information concerning other aspects of the methodology, such as the number of units of analysis represented at each level of one or more independent groups factors and whether randomization techniques were used in applying RM treatments to units of analysis. One of the more important details that was not available in the majority of the research reports was the choice of a test statistic to probe marginal or simple effects. This information is critical to the reader's understanding of the assumptions the researcher is making about the data and has important implications for the validity of the data analysis.

With respect to the third message communicated by this content analysis, Table 9 contains a summary of the findings pertaining to methods of testing correlated effects in all types of RM designs. Separate frequency analyses are provided for the period from 1987 to 1989 and from 1990 to 1993, as a means of assessing differences in the methods adopted over time. Consistent with the results reported by Ekstrom et al. (1990) and LaTour and Miniard (1983), the conventional ANOVA approach was most popular, and was represented in approximately one third of all of the articles. Moreover, use of this approach changed very little over the two time periods considered in Table 9. However, a dramatic reduction in the number of articles which did not clearly state what method of analysis had been adopted was observed. If, in fact, these articles incorporated the conventional ANOVA approach, then this would suggest a decline in popularity of conventional methods for RM analysis. Twice as many articles reported the use of either a df-adjusted test or MANOVA in the latter time period as compared to the former, but the extent to which either approach was adopted was, overall, very small.

It must be questioned why this trend continues, despite the great volume of literature that urges researchers to consider MANOVA and df-adjusted procedures. Perhaps this can be explained in part by examination of the statistical references which were cited in the research reports included in this data base. The most popular texts cited were by Winer (1962, 1971), which, while being excellent texts on research design, do not incorporate recent information on RM analysis procedures

Table 9

Methods for Testing RM Effects by Year of Publication (N = 100)

	1987 - 1989 (n = 47)		1990 - 1993 (n = 53)	
RM Analysis Procedure	f	%	f	%
Conventional ANOVA	13	28	18	34
Conventional ANCOVA	2	4	--	--
Conservative ANOVA	--	--	1	2
DF Adjusted ANOVA	2	4	6	11
DF Adjusted ANCOVA	1	2	--	--
MANOVA	2	4	3	6
Multivariate MANOVA	1	2	3	6
Not Stated	13	28	3	6
Descriptive Analysis Only	7	15	8	15
No Omnibus/Simple Effect Analysis	3	6	1	2
Nonparametric	3	6	--	--
Incorrect Analysis	2	4	5	9
Trend Analysis	1	2	2	4
Other	2	4	5	9

Note: See the note from Table 4.

(see Appendix A for a complete listing of the statistical references cited in this content analysis). It was surprising to note that of the 25 different statistical references cited in the research reports, 17 were published prior to 1980. The most current text was by Vercruyssen and Hendrick (1989). Moreover, the majority of the articles analyzed in this content analysis did not contain any statistical references. This may imply that many researchers are not aware of the problems associated with adopting the conventional approach to RM analysis, or of alternative procedures discussed in current statistical sources.

Furthermore, while a small number of articles reported the use of a preliminary test for the sphericity assumption, this test does not provide a sound basis for a decision regarding the method of analysis to adopt. It is known that sphericity tests are sensitive to departures from multivariate normality. For this reason, failure to reject the null hypothesis does not necessarily mean that the data are spherical (Cornell, Young, Seaman, & Kirk, 1992). Furthermore, even if the null hypothesis is not rejected, the data may not be spherical, again because of the known sensitivity of these tests to violations of the normality assumption.

Finally, it appears that while researchers are electing to use factorial designs, they are not taking full advantage of the factorial structure of the data, by failing to investigate joint variable effects of RM factors. This should not be surprising, as Rosnow and Rosenthal (1989b) describe the results of tests of interaction effects as "the most misinterpreted empirical results in psychology" (p. 1282). Perhaps this statement needs to be expanded to include the C&T field as well.

This content analysis suggests that researchers most often consider only main effects in their analysis of factorial designs. Where interaction effects were tested, these effects were typically probed in a manner that is inconsistent with the omnibus hypothesis, through the use of simple main effect tests and simple pairwise comparisons. In other instances, authors only reported the results of simple main effect tests, perhaps because interaction effect tests were bypassed in favour of the former analyses. It is impossible to know the reason why researchers choose to conduct main effect tests more often than interaction effect tests in factorial designs. This may be due to a lack of understanding of the meaning of an interaction effect or of the information that can be gleaned from an interaction test.

While it is a simple matter to change the manner in which research results are reported so that more details are forthcoming to the reader, it is a more crucial concern that analysis procedures which will produce valid results are selected by researchers, and that these procedures are applied correctly to a set of data in a way that will produce meaningful results. The remainder of this research project is devoted to examining robust methods for interaction analysis, particularly as they apply to probing variable interactions. The mixed design was the most popular of the RM designs used by C&T researchers and it will form the basis for the subsequent investigation.

CHAPTER 4

TESTING INTERACTIONS IN MIXED DESIGNS

This chapter reviews the research on the operating characteristics of conventional univariate, df-adjusted univariate, and multivariate procedures for testing interactions in mixed designs. A discussion of alternative methods of interaction analysis is also provided.

Univariate and Multivariate Tests of Interactions

Both theoretical and empirical studies have been used to study the behaviour of univariate and multivariate procedures for analyzing interactions in mixed designs. Huynh and Feldt (1980) computed exact Type I error rates for the conventional ANOVA F test in a mixed design with three groups and five RM and a total sample size of either 18 or 33. For balanced designs in which equality of the group covariance matrices existed, the number of Type I errors consistently exceeded the nominal value (a liberal test) when ϵ was less than .75. Error rates for the F test were as high as .12 for $\alpha = .05$ when $\epsilon = .39$. Increasing total sample size had the effect of decreasing the liberalness of the F test, but error rates for the F test were never less than α (a conservative test).

When group sizes were unequal and covariance matrices were heterogeneous, but sphericity was present in the data, Huynh and Feldt (1980) found that the F test was very sensitive. In one case where group sizes were in the ratio of 1:5:5, and the first group exhibited an average correlation (ρ) of .10 among the RM, while the remaining two groups exhibited average correlations of .90, the actual rate of Type I errors was .65 for a nominal value of .05. Conversely, when a large group size was paired with small values of ρ, the rate of Type I errors was considerably less than the nominal value (e.g., less than .0001 for $\alpha = .05$).

Huynh and Feldt (1980) also considered the effects of violating the multisample sphericity assumption (i.e., $\epsilon < 1.00$ and unequal Σ_js), but only when group sizes were equal. Under such conditions, the F test again proved to be liberal when a high degree of nonsphericity existed in the data, with a maximum value of .16. Based on the results of the study, Huynh and Feldt concluded that "in all situations under investigation, the test for interaction proved to be more vulnerable than the one for treatment [main] effects, especially when the plot [group] sizes are not equal" (p. 71).

Belli (1988) employed Monte Carlo techniques to examine the robustness of multivariate tests in a mixed design containing two groups and five levels of the RM factor when the data were spherical. The tests examined, Hotelling-Lawley (Hotelling, 1951; Lawley, 1938) trace, Pillai-Bartlett (Bartlett, 1939; Pillai, 1955) trace, Roy's (1953) largest root, and Wilks' (1932) likelihood ratio, could not, in general, provide Type I error control under conditions of group size imbalance when heterogeneity of group covariance matrices existed. Belli noted that even when group sizes were equal and existed in combination with heterogenous covariance matrices, only the Pillai-Bartlett trace could control the rate of false positives, and then only when the degree of heterogeneity was small.

Keselman and Keselman (1990) considered a mixed design where the number of groups was set at three and the number of RM was set at either four or eight, in a Monte Carlo study. In the case of four RM, the $\hat{\epsilon}$, $\tilde{\epsilon}$, and $\bar{\epsilon}$ df-adjusted F tests and the Pillai-Bartlett (Bartlett, 1939; Pillai, 1955) trace criterion provided robust tests of the interaction when multisample sphericity was violated but group sizes were equal, and also when group covariance matrices were homogeneous but sphericity was not present in the data.

However, the df-adjusted F tests and the MANOVA test were sensitive to violations of multisample sphericity when the design was unbalanced. In situations where group sizes and covariance matrices were positively paired, so that the group with the largest sample size also exhibited a covariance matrix with the largest element values, all statistical procedures were conservative. This degree of conservatism increased with increases in group size inequality and covariance matrix heterogeneity. Holding all else constant for this condition of positive pairings, the conservatism of the univariate tests increased as ϵ approached its upper bound of 1.0.

In the situation of a negative pairing, where the group with the largest sample size also exhibited a covariance matrix with the smallest element values, the number of false positives exceeded the nominal alpha level under both univariate and multivariate testing. The liberalness of these statistical procedures increased as the degree of group size inequality and covariance heterogeneity increased. Paralleling the findings for conditions of positive pairings, the liberalness of the univariate tests increased as the degree of nonsphericity in the data decreased.

When the RM factor had eight levels, Keselman and Keselman (1990) found neither the univariate nor multivariate procedures could offer robust tests of the interaction effect when the data were nonspherical, even when the covariance matrices were homogeneous. When multisample sphericity was violated and group sizes were unequal, the same pattern of positive pairings resulting in conservative tests, and negative pairings leading to liberal tests, was identified, albeit the results were more extreme.

Finally, Keselman, Keselman, and Lix (in press) considered whether the use of both a univariate and multivariate procedure in a combined testing strategy could offer Type I error control for tests of the interaction in a mixed design with either four or eight levels of the RM factor, when total sample size ranged from 30 to 191. The $\hat{\varepsilon}$ F and Pillai-Bartlett (Bartlett, 1939; Pillai, 1955) trace statistics were each computed, and if either were significant at the .025 level, then the null hypothesis was rejected. The authors found that this combined approach could not limit the rate of Type I errors to the nominal .05 level when the multisample sphericity assumption was violated. Furthermore, the error rate remained consistently high when unequal covariance matrices and unequal group sizes were negatively paired, even for the largest total sample size condition investigated.

Robust Tests of Interactions

While researchers have long been advised to avoid the conventional ANOVA procedure for testing correlated effects, the studies described in the previous section illustrate that even df-adjusted ANOVA and MANOVA procedures should not be adopted to test for the presence of an RM interaction in mixed designs if the assumption of multisample sphericity is untenable, particularly if the design is unbalanced. However, two alternative approaches have been shown to provide Type I error control for tests of both RM main and interaction effects in the majority of situations in which the multisample sphericity assumption is violated and the design is unbalanced.

Algina and Oshima (1994) demonstrated that robust tests of RM effects may be obtained using Huynh's (1978) improved general approximate univariate statistic. This procedure involves calculation of the usual F statistic for a test of the interaction. However, a modified CV, which reflects the degree of violation of the multisample sphericity assumption, is used in assessing statistical significance. While the procedure developed by Huynh can provide effective Type I error control in a variety of situations, it has the disadvantage of being computationally complex, and is not currently available in any statistical software package.

Keselman et al. (1993) found that an approximate df Welch-James (James, 1951, 1954; Welch, 1947, 1951) multivariate procedure described by Johansen (1980) can control the Type I error rate for tests of interactions in unbalanced mixed designs. Welch developed a statistical test for equality of means in the one-way independent groups design when the assumption of equal variances across groups can not be considered tenable; this statistic uses a nonpooled estimate of error variance. The multivariate analog of Welch's procedure was developed by Johansen for designs with more than two groups.

A general linear model (GLM) approach can be used to illustrate the application of Johansen's (1980) approximate df solution for mixed designs. Let

$$\mathbf{Y} = \mathbf{X}\beta + \xi \,, \tag{4.1}$$

where $\mathbf{Y}$ is an N x p matrix of scores on p repeated measurements, N is the total sample size, $\mathbf{X}$ is an N x r design matrix with rank($\mathbf{X}$) = r, β is an r x p matrix of nonrandom parameters (i.e., population means), and ξ is an N x p matrix of random error components. Denote $\mathbf{Y}_j = \mathbf{Y} \cdot (\mathbf{X}_j \mathbf{1}_p^T)$ as a Hadamard product (Searle, 1987, p. 49), where $\mathbf{X}_j$ is the *j*th column of $\mathbf{X}$ (j = 1 ,..., r) and consists entirely of zeros and ones, $\mathbf{1}_p$ is a p x 1 vector of ones, and $\cdot$ is the dot product function, such that $\mathbf{Y}_j$ is an element-by-element product matrix. The model assumes that the observations in $\mathbf{Y}_j$ are independently distributed normal variates with mean vector β_j and variance-covariance matrix Σ_j [i.e., i.d. N(β_j, Σ_j)], where β_j is the *j*th row of β and $\Sigma_j \neq \Sigma_{j'}$ (j $\neq$ j').

Let

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \,, \tag{4.2}$$

represent an estimate of the matrix of population means and

$$\hat{\Sigma}_j = \frac{(\mathbf{Y}_j - \mathbf{X}_j\hat{\beta}_j)^T(\mathbf{Y}_j - \mathbf{X}_j\hat{\beta}_j)}{n_j - 1} \,, \tag{4.3}$$

estimate Σ_j, where $n_j = \mathbf{X}_j^T\mathbf{X}_j$, and $\hat{\beta}_j$ estimates β_j.

The general linear hypothesis is

$$H_o: \mathbf{R}\mu = \mathbf{0}\ , \tag{4.4}$$

where $\mathbf{R} = \mathbf{C} \otimes \mathbf{U}^T$, $\mathbf{C}$ is a df_C x r matrix which controls contrasts on the independent groups factor(s), with rank($\mathbf{C}$) = $df_C \leq r$, $\mathbf{U}$ is a p x df_U matrix which controls contrasts on the correlated factor(s), with rank($\mathbf{U}$) = $df_U \leq p$, and $\otimes$ is the Kronecker or direct product function. Furthermore, $\mu = \text{vec}(\beta^T) = [\beta_1 \ldots \beta_r]^T$, where $\beta_j = [\mu_{j1} \ldots \mu_{jp}]$. In Equation 4.4, μ is the column vector with rp elements obtained by stacking the columns of β^T. The $\mathbf{0}$ column vector is of order df_C x df_U.

The generalized test statistic given by Johansen (1980) is

$$T_{WJ} = (\mathbf{R}\hat{\mu})^T(\mathbf{R}\hat{\Sigma}\mathbf{R}^T)^{-1}(\mathbf{R}\hat{\mu})\ , \tag{4.5}$$

where $\hat{\mu}$ estimates μ, and $\hat{\Sigma} = \text{diag}[\hat{\Sigma}_1/n_1 \ldots \hat{\Sigma}_r/n_r]$, a block matrix with diagonal elements $\hat{\Sigma}_j/n_j$. This test statistic divided by a constant, c, approximately follows an F distribution with $\upsilon_1 = df_C$ x df_U, and $\upsilon_2 = \upsilon_1(\upsilon_1 + 2)/(3A)$, where $c = \upsilon_1 + 2A - (6A)/(\upsilon_1 + 2)$. The formula for the statistic A is

$$A = \frac{1}{2}\sum_{j=1}^{r}\left[\text{tr}\left\{\hat{\Sigma}\mathbf{R}^T(\mathbf{R}\hat{\Sigma}\mathbf{R}^T)^{-1}\mathbf{R}\mathbf{Q}_j\right\}^2 + \left\{\text{tr}(\hat{\Sigma}\mathbf{R}^T(\mathbf{R}\hat{\Sigma}\mathbf{R}^T)^{-1}\mathbf{R}\mathbf{Q}_j)\right\}^2\right]/(n_j - 1)\ . \tag{4.6}$$

The matrix $\mathbf{Q}_j$ is a symmetric block matrix of dimension r x p associated with $\mathbf{X}_j$, such that the (s,t)-th diagonal block of $\mathbf{Q}_j = \mathbf{I}_p$ if s = t = j and is $\mathbf{0}$ otherwise.

In order to test the interaction in a two-way mixed design, $\mathbf{C} = \mathbf{C}_j$ and $\mathbf{U} = \mathbf{U}_k$, so that $\mathbf{R} = \mathbf{C}_j \otimes \mathbf{U}_k^T$, where $\mathbf{C}_j$ is a (J - 1) x J matrix which defines a set of (J - 1) linearly independent contrasts for the grouping factor, and $\mathbf{U}_k$ is a K x (K - 1) matrix which defines a set of (K - 1) linearly independent contrasts for the RM factor. For example, in a mixed design containing a grouping factor with four levels and a RM factor with three levels, **C** and **U** matrices that may be used to obtain a test of the interaction are:

$$\mathbf{C} = \mathbf{C}_4 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}, \text{ and } \mathbf{U} = \mathbf{U}_3 = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Here, the rows of **C** represent a set of three linearly independent contrasts among the levels of the grouping factor, while the columns of **U** form a set of two linearly independent contrasts among the levels of the RM factor. The Kronecker product, $\mathbf{C}_4 \otimes \mathbf{U}_3^T$, is

$$\mathbf{R} = \begin{bmatrix} (1)\mathbf{U}^T & (-1)\mathbf{U}^T & (0)\mathbf{U}^T & (0)\mathbf{U}^T \\ (1)\mathbf{U}^T & (0)\mathbf{U}^T & (-1)\mathbf{U}^T & (0)\mathbf{U}^T \\ (1)\mathbf{U}^T & (0)\mathbf{U}^T & (0)\mathbf{U}^T & (-1)\mathbf{U}^T \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix},$$

and has six linearly independent rows.

Tang and Algina (1993) considered the robustness of Johansen's (1980) statistic in a multivariate independent groups design with more than two groups, but only for normal data, when group covariance matrices were heterogeneous and group sizes were unequal. They observed that Johansen's solution was generally robust, except when the ratio of total sample size to the number of dependent variables was small, the homogeneity assumption was violated, and group sizes and covariance matrices were negatively paired, in which case it produced liberal results.

Similar findings were reported by Keselman et al. (1993) for the A x B mixed design. Moreover, they found that the effect of nonnormality was to inflate Type I error rates. Accordingly, the authors suggested that in order to obtain a robust test of the RM interaction hypothesis using an approximate df solution, the ratio of the number of observations in the smallest group (i.e., n_{min}) to (K - 1) should be at least 3 or 4 to one, and preferably higher if the validity of the multivariate normality assumption is questionable. Where there are an insufficient number of units of analysis to achieve this requirement, the authors suggested adopting a .01 significance criterion in order to maintain the rate of Type I errors below five percent.

Interaction Contrasts

In a mixed design, researchers are typically most interested in testing for the presence of an interaction before examining main effects. However, the omnibus test procedures described by Algina and Oshima (1994) and Keselman et al. (1993) offer no insight into the nature of the interaction and researchers will routinely follow a significant result with statistical tests to probe this effect (Boik, 1993; Olejnik & Huberty, 1993; Rosnow & Rosenthal, 1989b).

Two techniques that may be used to examine interactions are simple main effect tests and interaction contrasts. The former approach, which, as evidenced by the results reported in the previous chapter, is favoured by many applied researchers, involves examining the effects of one factor at a particular level of the second factor.

For example, for a mixed design, the researcher may test for equality of the K RM means at each of the J levels of the grouping factor. By focusing on a particular level of the grouping factor, this factor is eliminated from the analysis, and the researcher is left to conduct a series of simple RM univariate or multivariate tests. The procedure used to conduct these simple main effect tests might include an df-adjusted ANOVA or MANOVA test, or one of the two alternative procedures described previously. The choice among these approaches depends on the assumptions the researcher is willing to make about the data.

However, this procedure of analyzing simple main effects has been criticized because

> it is assumed that the interaction can be interpreted by determining which simple effects are significant. It must be remembered that a significant interaction does not indicate that one or more simple main effects are significantly different from zero but rather that at least one contrast of one treatment [factor] is different at two or more levels of the second treatment [factor]. (Boik, 1975, p. 32)

The use of interaction contrasts is favoured from a theoretical standpoint since the null hypothesis under consideration for a particular contrast is consistent with the hypothesis associated an omnibus test of the interaction (Boik, 1993; Marascuilo & Levin, 1970, 1976; Timm, 1994). However, in practice, interaction contrasts are rarely used by applied researchers (Kaufman et al., 1986; Rosenow & Rosenthal, 1989a). Boik suggests that this is because most researchers are not familiar with "specialized multiple comparisons for interactions" (p. 2), or methods for applying these techniques with statistical software packages.

While several different types of interaction contrasts have been discussed in the literature (Bradu & Gabriel, 1974), product contrasts have been given the most attention (Boik, 1993; Gabriel, Putter, & Wax, 1973; Johnson, 1976; Timm, 1994). In a two-way design, a product contrast is a Kronecker product of two vectors, each of which forms a contrast among the levels of one main effect factor. Gabriel et al. (1973) discuss several types of product contrasts which may be used to probe interactions. To understand these product contrasts, it is helpful to consider some techniques for probing effects in one-way RM designs.

Let $\mu_1 \ldots \mu_K$ denote the population means for a simple RM design with $k = 1,\ldots, K$ levels, where $\bar{Y}_k$ estimates μ_k. As Gabriel et al. (1973) note, to probe the main effect, one may use deviations from the mean, of the form $\delta_k = \mu_k - \mu_.$, where $\mu_.$ is the grand mean. The second approach, which is the most popular for probing main effects, is the pairwise difference, $\psi = \mu_k - \mu_{k'}$, where $k \neq k'$.

Interaction residuals are extensions of deviations from the mean in one-way designs, and are defined as $\gamma_{jk} = \mu_{jk} - \mu_{j.} - \mu_{.k} + \mu_{..}$, where μ_{jk}, $\mu_{j.}$, $\mu_{.k}$, and $\mu_{..}$ respectively represent cell, row, column and grand population means, and μ_{jk} is estimated by $\bar{Y}_{jk}$. A tetrad contrast is defined as $\psi = (\mu_{jk} - \mu_{jk'}) - (\mu_{j'k} - \mu_{j'k'})$, where $j \neq j'$, $k \neq k'$. Interaction residuals and tetrad contrasts are related, as the latter can be expressed in terms of the former, so that

$$\psi = (\gamma_{jk} - \gamma_{jk'}) - (\gamma_{j'k} - \gamma_{j'k'}).$$

To illustrate why interaction contrasts should be favoured over simple effect tests for probing an interaction, consider a test of a simple pairwise comparison at one level of the grouping factor that involves two different levels of the RM factor. This simple pairwise comparison tests the hypothesis, H_0: $\mu_{jk} - \mu_{jk'} = 0$, where $k \neq k'$. Using the previously defined notation, $\mu_{jk} = \gamma_{jk} + \mu_{j.} + \mu_{.k} - \mu_{..}$. By substituting the appropriate elements of this equality into the simple pairwise hypothesis, the following solution is obtained:

$$\mu_{jk} - \mu_{jk'} = 0,$$
$$(\gamma_{jk} + \mu_{j.} + \mu_{.k} - \mu_{..}) - (\gamma_{jk} + \mu_{j.} + \mu_{.k'} - \mu_{..}) = 0,$$
$$(\gamma_{jk} - \gamma_{jk'}) + (\mu_{.k} - \mu_{.k'}) = 0.$$

This simple effect is only partially comprised of interaction components (i.e., the γ_{jk}s). Thus, if this hypothesis were rejected, one could not determine if this was due to the difference in the marginal means, the difference in the interaction residuals, or both. With a tetrad contrast, however, the hypothesis that is tested only involves interaction residuals.

Since tetrad contrasts are direct extensions of the popular pairwise contrasts for probing main effects, they are perhaps easiest for the applied researcher to understand and interpret. In a two-way design, a tetrad contrast essentially involves testing for the presence of an interaction between rows and columns in a 2 x 2 submatrix of the A x B data matrix, and represents a test for a difference in two pairwise differences.

Two final points must be made regarding methods for probing interactions. First, it should be recognized that Cicchetti's (1972) method of testing unconfounded cell mean comparisons is not an appropriate approach for probing an interaction, as such comparisons are equivalent to simple pairwise comparisons. As was demonstrated previously, such comparisons are in fact confounded by either row or column main effects. Second, Rosnow and Rosenthal (1989a; 1991) recommend that the only appropriate methods for probing interactions involve comparisons on what they call corrected cell means. However as Boik (1993) notes, corrected cell means are equivalent to interaction residuals. Since a tetrad contrast on the cell means can be expressed as a contrast on the interaction residuals, the approach presented in this research is equivalent to that advocated by Rosnow and Rosenthal.

Tetrad Contrast Test Statistics in the Mixed Design

Two choices of a test statistic exist for performing tetrad contrasts in mixed RM designs (Keselman & Keselman, 1993, pp. 125-126). One statistic that can be used to test H_o: $\psi = 0$ employs an estimate of the standard error of the contrast which uses $MS_{KS/J}$, the error mean square for the usual omnibus F test of the interaction. The test statistic is

$$t = \frac{(\bar{Y}_{jk} - \bar{Y}_{jk'}) - (\bar{Y}_{j'k} - \bar{Y}_{j'k'})}{\sqrt{MS_{KS/J}\left[\frac{2}{n_j} + \frac{2}{n_{j'}}\right]}}, \tag{4.7}$$

where, as noted previously, $j \neq j'$ and $k \neq k'$. This statistic is distributed as Student's t with df $\nu_p = (K - 1)(N - J)$. Conducting tetrad contrasts with this statistic may be appealing because results may be obtained using statistical software packages such as SAS (SAS Institute Inc., 1989b). As well, because this statistic has error df equal to those available for a test of the omnibus interaction hypothesis, it should provide greater power to reject the null hypothesis than other statistics. However, if the data do not satisfy the multisample sphericity assumption, the error term for this statistic, which involves pooling over both the RM and grouping factors, will result in a biased estimate of the standard error of the contrast, and, as a consequence, this statistic will produce invalid results.

An alternate statistic employs a standard error derived only from that data used in forming the contrast and is defined as

$$t = \frac{(\bar{Y}_{jk} - \bar{Y}_{jk'}) - (\bar{Y}_{j'k} - \bar{Y}_{j'k'})}{\sqrt{\dfrac{\mathbf{c}^T \hat{\Sigma}_j \mathbf{c}}{n_j} + \dfrac{\mathbf{c}^T \hat{\Sigma}_{j'} \mathbf{c}}{n_{j'}}}}, \tag{4.8}$$

where $\mathbf{c}$ is a K x 1 vector of coefficients which contrasts the kth and k'th levels of the RM factor. In other words, the standard error of the tetrad contrast is formed using data from only four cells of the A x B data matrix and does not rest on the multisample sphericity assumption. The nonpooled statistic does not follow a t distribution, but can be approximated by Student's t with Satterthwaite (1941, 1946) estimated df

$$\nu_s = \frac{\left[\dfrac{\mathbf{c}^T \hat{\Sigma}_j \mathbf{c}}{n_j} + \dfrac{\mathbf{c}^T \hat{\Sigma}_{j'} \mathbf{c}}{n_{j'}}\right]^2}{\dfrac{[\mathbf{c}^T \hat{\Sigma}_j \mathbf{c}/n_j]^2}{n_j - 1} + \dfrac{[\mathbf{c}^T \hat{\Sigma}_{j'} \mathbf{c}/n_{j'}]^2}{n_{j'} - 1}}. \tag{4.9}$$

This approximate df solution for conducting interaction contrasts may also be conceptualized from a multivariate perspective. By forming all possible pairwise differences among the K repeated measures and denoting these variables as $D_{ij} = Y_{ijk} - Y_{ijk'}$, the test statistic may be defined as

$$t = \frac{\bar{D}_j - \bar{D}_{j'}}{\sqrt{\sum_{j=1}^{J} \frac{c_j^2 s_{j(D)}^2}{n_j}}}, \tag{4.10}$$

where $\bar{D}_j$ is the mean difference at level j of the grouping factor, $s_{j(D)}^2$ is the corresponding variance of a D variable and c_j is the contrast coefficient at level j of the grouping factor. The error df are then expressed as

$$\nu_s = \frac{\left[\sum_{j=1}^{J} c_j^2 s_{j(D)}^2 / n_j\right]^2}{\sum_{j=1}^{J} \frac{\left[c_j^2 s_{j(D)}^2 / n_j\right]^2}{n_j - 1}}. \tag{4.11}$$

Interaction contrasts that employ a nonpooled test statistic may also be conceptualized from a GLM perspective using Johansen's (1980) approximate df solution. If **C** and **U** respectively denote contrast vectors on factors A and B in a mixed design, then $\mathbf{R} = \mathbf{C} \otimes \mathbf{U}^T$ represents a product contrast. Furthermore, if $\mathbf{C} = \mathbf{c}_{jj'}$ forms a contrast among two levels of Factor A, and $\mathbf{U} = \mathbf{u}_{kk'}$ forms a contrast among two levels of Factor B, then the Kronecker product of these two vectors is a tetrad contrast. For example, in a mixed design with four levels of the grouping factor and three levels of the RM factor, a tetrad contrast involving the first and third levels of the grouping factor and the first and second levels of the RM factor would require the formation of the following **C** and **U** vectors

$$\mathbf{C} = \mathbf{c}_{13} = \begin{bmatrix} 1 & 0 & -1 & 0 \end{bmatrix}, \text{and } \mathbf{U} = \mathbf{u}_{12} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

The resulting **R** matrix has the form

$$\mathbf{R} = \mathbf{c}_{13} \otimes \mathbf{u}_{12}^{T} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Controlling the Familywise Error Rate for Multiple Tetrad Contrasts

A number of CVs have been proposed for limiting the FWR for the set of all possible tetrad contrasts on the data. A Scheffe (1953) CV may be adopted (Boik, 1993; Gabriel et al., 1973; Marascuilo & Levin, 1970), $\{\nu_1 F[1 - \alpha; \nu_1, \nu_2]\}^{1/2}$, where $\nu_1 = (J - 1)(K - 1)$ and ν_2 is the error df. However, Scheffe's method controls the FWR across all possible interaction contrasts on the data, including the subset of tetrad contrasts. Therefore it is likely to be less powerful than competing alternatives. Consequently, a Studentized maximum root CV, $R[1 - \alpha, p, q, \nu_2]$, where $p = \min(J - 1, K - 1)$, $q = \max(J - 1, K - 1)$, and $R[1 - \alpha, p, q, \nu_2]$ is the $1 - \alpha$ centile of the Studentized maximum root distribution, is considered to be a better choice, as it is intended to provide FWR control across the set of all possible product contrasts (Boik, 1993; Bradu & Gabriel, 1974; Gabriel et al., 1973; Johnson, 1976). Furthermore, Gabriel et al. (1973) recommend a Bonferroni CV (Dunn, 1961), $t[1 - \alpha/(2C); \nu_2]$, where $C = J^*K^*$, $J^* = J(J - 1)/2$, and $K^* = K(K - 1)/2$. Finally, Hochberg and Tamhane (1987, p. 299) suggest that a Studentized maximum modulus

CV, M[1 - α; C, ν_2], where M[1 - α; C, ν_2] is the 1 - α centile of the Studentized maximum modulus distribution, is more likely to maintain the FWR at α than a Bonferroni CV.

An alternative to adopting one of these simultaneous MCPs is to select a stepwise procedure. A number of stepwise procedures based on the Bonferroni inequality have been developed, which will necessarily control the FWR for tetrad contrasts when derivational assumptions are satisfied. However, since these methods rely on a different CV at each stage of hypothesis testing, they may provide greater power to detect tetrad interactions than Dunn's (1961) method. Two procedures, derived by Hochberg (1988) and Shaffer (1986), are particularly promising.

Hochberg's (1988) step-up Bonferroni procedure is an attractive choice because it is one of the simpler stepwise procedures available. Hommel (1988) and Rom (1990) have proposed Bonferroni procedures which are known to be more powerful than Hochberg's method. However, as Dunnett and Tamhane (1992) note, marginal power differences exist among these three procedures and Hochberg's procedure is much easier to use than the other two.

With Hochberg's (1988) method, one begins by rank ordering the p values corresponding to the statistics used for testing the hypotheses $H_{(1)}, \ldots, H_{(C)}$ [i.e. $\psi = (\mu_{jk} - \mu_{jk'}) - (\mu_{j'k} - \mu_{j'k'}) = 0$], so that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(C)}$ represent the ordered p values. The decision rule is to reject $H_{(m')}$ ($m' \leq m$; $m = C, \ldots, 1$) if $p_{(m)} \leq \alpha/(C - m + 1)$. Testing begins with the hypothesis corresponding to the largest p value, $p_{(C)}$. If $p_{(C)} \leq \alpha$, all C hypotheses are rejected; if not, $H_{(C)}$ is retained and testing moves to $H_{(C-1)}$. If $p_{(C-1)} \leq \alpha/2$, $H_{(C-1)}$ is rejected, as are all remaining hypotheses; if not $H_{(C-1)}$ is also retained, and $p_{(C-2)}$ is compared to $\alpha/3$, and so on. This continues, if all previous hypotheses have been retained, until $p_{(1)}$ is compared to α/C.

Shaffer (1986) proposed a more powerful version of Holm's (1979) sequentially rejective Bonferroni procedure, and has demonstrated its use for testing tetrad contrast hypotheses. Under Shaffer's method, one proceeds in the same manner as for the Hochberg (1988) procedure by rank ordering the p values. However, testing begins by comparing the smallest p value, $p_{(1)}$ to α/C, and if $p_{(1)} \leq \alpha/C$, the corresponding hypothesis, $H_{(1)}$, is rejected, otherwise it is retained. If $H_{(1)}$ is rejected, one goes on to compare the next largest p value, $p_{(2)}$, to α/C_2^*, where C_2^* represents the maximum number of remaining hypotheses that could be true, given rejection of the previous hypothesis. One proceeds in this manner by rejecting $H_{(m)}$ ($m = 1, \ldots, C$) if $p_{(m)} \leq \alpha/C_m^*$. Shaffer has tabled values of C_m^* for selected A x B designs. However, since the value of C_m^* may not be readily apparent for other factorial designs (i.e., higher-order designs), Shaffer recommends the following approach to hypothesis testing for tetrad contrasts in these cases: If $H_{(1)}$ is rejected using an α/C criterion, one proceeds to test $H_{(2)}$ by letting $C_2^* = C - (J - 1)(K - 1)$. If $H_{(2)}$ is rejected, C_m^* is set equal to C_2^* for all $2 \leq m \leq C - C_2^* + 1$, and is assigned a value of $C - m + 1$ for all $m > C - C_2^* + 1$.

CHAPTER 5

SIMULATION STUDY METHODOLOGY

Overview of a Monte Carlo Study

The purpose of a Monte Carlo simulation study is to empirically evaluate the operating characteristics of a statistic test under a range of derivational assumption violations. Pseudorandom sets of numbers are generated using a computer algorithm and are sampled from populations with known characteristics. The simulation experiment is designed so that the null hypothesis is either true or false. For each replication of an experiment, a test statistic is computed from the generated data and based on this result, the null hypothesis is either rejected or retained. Thus, depending on the nature of the null hypothesis, empirical estimates of either Type I error or power are obtained.

Description of Tetrad Contrast Procedures

Ten procedures were selected to investigate the viability of conducting tetrad contrasts in mixed designs using Monte Carlo methods. A test statistic employing a pooled error term (i.e., Equation 4.7) was used, in addition to one based on a nonpooled error term (i.e., Equation 4.8), even though it was anticipated that only the latter would provide control of the FWR. The former was investigated to obtain some empirical evidence of the extent of bias in error rates that may result when using a MCP which incorporates a pooled test statistic when the multisample sphericity assumption is violated.

The pooled and nonpooled statistics were considered in combination with either a Scheffe (1953), Studentized maximum root, Hochberg (1988) step-up Bonferroni, Shaffer (1986) modified sequentially rejective Bonferroni, or Studentized maximum modulus CV. For the procedures employing a pooled test statistic, the CVs were defined by setting $\nu_2 = \nu_p$; when a nonpooled test statistic was used, $\nu_2 = \nu_s$.

Although a Scheffe CV will be larger than either of the other four CVs, this method was included in case the other procedures could not limit the number of Type I errors under violations of the multisample sphericity assumption. Furthermore, both Jaccard et al. (1984) and Lix (1995) found that applied researchers routinely adopted Scheffe's method for pairwise mean comparisons. Thus, it was considered desirable to determine if this is an acceptable method of controlling the FWR for tetrad contrasts. Finally, the Hochberg (1988) and Shaffer (1986) procedures were selected over the Bonferroni (Dunn, 1961) method for the reasons enumerated in the previous chapter.

Monte Carlo Study Variables

The ten procedures for testing tetrad contrast hypotheses were compared for the simplest mixed design, that is, a design containing a single independent groups factor and a single RM factor. In addition, the design was univariate in nature.

One aspect of the study was held constant, that being the number of levels of the grouping factor, which was set at three. From the C&T content analysis it would appear that in mixed designs, many researchers elect to use a grouping factor with only two levels. However, this would not provide sufficient diversity in the present study to investigate the behaviour of the selected tetrad contrast procedures, as data from both levels of the grouping factor would necessarily be involved in forming each tetrad contrast.

Nine variables were selected to investigate the behaviour of the selected statistical procedures with respect to Type I error and power rates. These were the: (a) number of levels of the RM factor, (b) sphericity pattern, (c) equality/inequality of the group variance-covariance matrices, (d) total sample size, (e) group size equality/inequality, (f) nature of the pairing of unequal covariance matrices and unequal group sizes, (g) population shape, (h) nature of the null hypothesis, and (i) population effect size. Table 10 provides summary information concerning the values of the variables which were investigated and the following discussion deals with each of these variables in turn.

Number of Repeated Measures Factor Levels

Keselman and Keselman (1990) found that df-adjusted univariate and multivariate omnibus procedures for interaction tests in mixed designs became increasingly sensitive to violations of the multisample sphericity assumption as the number of levels of the RM factor increased. Hence, the ten procedures were studied when the number of factor levels was set at four and eight. As well, given that more than one quarter of the 43 mixed C&T studies included a RM factor with more than 6 levels, it would seem important to investigate the procedures for a large number of RM factor levels.

Table 10

Monte Carlo Study Variables

Variable	Values
RM Factor Levels (K)	4; 8
Sphericity Pattern (ϵ)	1.0; .75; .40
Equality/Inequality of Σ_js	1:1:1; 1:3:5
Total Sample Size (N)	30; 45
Equality/Inequality of n_js	N = 30: 8, 10, 12 (Δn_j = .163) 6, 10, 14 (Δn_j = .327) N = 45: 12, 15, 18 (Δn_j = .163) 9, 15, 21 (Δn_j = .327)
Pairing of Σ_js and n_js	(a) equal n_js, equal Σ_js (b) equal n_js, unequal Σ_js (c) unequal n_js (Δn_j = .163), unequal Σ_js, positive pairing (c') unequal n_js (Δn_j = .327), unequal Σ_js, positive pairing (d) unequal n_js (Δn_j = .163), unequal Σ_js, negative pairing (d') unequal n_js (Δn_j = .327), unequal Σ_js, negative pairing
Population Shape	Normal; χ^2_3
Null Hypothesis	Complete; Partial
Effect Size (f)	.50; 1.00; 1.50

Sphericity Pattern

Box's (1954) correction factor, ϵ, was used to quantify the degree of departure from the assumption of sphericity (i.e., Equation 2.3). Without loss of generality, the pooled variance-covariance matrix (i.e., Σ_p) contained element values of ten and five on the diagonal and off-diagonal, respectively, when sphericity was satisfied. Matrices with ϵ values of .75 and .40 were chosen to investigate nonspherical conditions and represent moderate and extreme departures from sphericity, respectively. For $K = 4$, the minimum value that ϵ may attain is .33, while for $K = 8$, the lower bound is $\epsilon = .14$.

The elements of the pooled variance-covariance matrices were chosen such that the average variance and covariance were equal to ten and five, respectively, in order to achieve comparability across the simulation conditions. The pooled matrices values for the $K = 4$ and $K = 8$ conditions can be found in Tables 11 and 12, respectively.

Heterogeneity of Group Variance-Covariance Matrices

The effects of heterogeneity of the group orthonormal variance-covariance matrices was investigated by creating two sets of matrices. For one set, a given element in a covariance matrix for a particular group was equal to the corresponding element in each of the matrices for the other two groups, so that the elements of the group covariance matrices were in a 1:1:1 ratio. For the second set, corresponding elements in the group covariance matrices were not equal to one another. Each element in the covariance matrix for the second group was three times that of the

Table 11

Pooled Variance-Covariance Matrix Element Values (K = 4)

$\epsilon = .75$

18.0	8.0	6.0	4.0
	8.0	5.0	4.0
		7.0	3.0
			7.0

$\epsilon = .40$

23.8	11.9	6.4	0.9
	9.5	5.7	2.6
		3.9	2.5
			2.8

Table 12

Pooled Variance-Covariance Matrix Element Values (K = 8)

$\epsilon = .75$

18.0	8.0	7.0	7.0	6.0	5.0	5.0	5.0
	12.0	8.0	7.0	6.0	5.0	5.0	2.0
		10.0	6.0	6.0	5.0	5.0	2.0
			10.0	5.0	5.0	4.0	4.0
				9.0	5.0	5.0	3.0
					8.0	4.0	4.0
						7.0	1.0
							6.0

$\epsilon = .40$

28.8	12.8	10.1	9.8	8.3	7.3	6.0	1.8
	17.4	8.1	7.4	6.9	4.1	3.4	-1.0
		9.9	7.7	6.5	5.7	3.4	1.4
			8.3	5.6	4.3	3.9	2.4
				5.6	4.4	2.6	1.4
					4.3	2.4	1.4
						3.2	1.9
							2.5

matrix for the first group. As well, each element in the matrix of the third group was five times that of the matrix for the first group. Thus the elements in the group variance-covariance matrices were in a 1:3:5 ratio. This degree of covariance heterogeneity was chosen because Keselman and Keselman (1990) found it to have the greatest effect, among the heterogeneity conditions they investigated, on Type I error rates for omnibus tests of the interaction in mixed designs.

Total Sample Size

Total sample size (N) was set at either 30 or 45, to allow investigation of the effects of both a small and moderate sample size. The ten procedures were investigated when the design was balanced, and also when it was unbalanced. When the design was balanced (i.e., group sizes were equal), there were either ten or fifteen observations per group. Two cases of group size imbalance were considered for each total sample size. For $N = 30$, $n_j = 8, 10, 12$ and $n_j = 6, 10, 14$, while for $N=45$, $n_j = 12, 15, 18$ and $n_j = 9, 15, 21$. For both values of N, the coefficient of group size variation (i.e., Δn_j; see Equation 2.9) is .163 for the former condition, and .327 for the latter. Thus, both mild and moderate degrees of imbalance were considered.

Pairing of Group Covariance Matrices and Group Sizes

For those conditions involving both unequal group sizes and unequal group orthonormalized variance-covariance matrices, both positive and negative pairings of these group sizes and covariance matrices were investigated, since these pairings have been associated with conservative and liberal results, respectively, in tests of the omnibus interaction effect in mixed designs. In the former case, the largest n_j was associated with the covariance matrix containing the largest element values; while in the latter, the largest n_j was associated with the covariance matrix containing the smallest element values.

In summary, six pairings of variance-covariance matrices and group sizes were investigated: (a) equal n_j, equal Σ_j; (b) equal n_j, unequal Σ_j; (c/c′) unequal n_j, unequal Σ_j (positively paired); and (d/d′) unequal n_j, unequal Σ_j (negatively paired). The c′/d′ conditions denotes the more disparate unequal group sizes cases, while the c/d conditions designates the less disparate unequal group sizes cases.

Population Shape

The test statistics which are the basis for the various tetrad contrast procedures rest on the assumption of multivariate normality. Although no information on the extent to which this assumption may or may not be satisfied was collected in the content analysis, it would seem unlikely that it would be satisfied in all cases, particularly given the results obtained by other researchers, such as Micceri (1989), who analyzed the distributional characteristics of 440 educational and psychological data sets, and found that few of these could be characterized as normal in form. Thus, it was deemed important to examine the operating characteristics of the selected procedures when the underlying population distribution was normal and nonnormal.

For the normal distribution, pseudorandom vectors of observations $\mathbf{Y}_{ij} = [Y_{ij1}, Y_{ij2}, \ldots, Y_{ijK}]$ with mean vector $\boldsymbol{\mu}_j = [\mu_{j1}, \mu_{j2}, \ldots, \mu_{jK}]$ and variance-covariance matrix Σ_j were generated using the International Mathematical and

Statistical Library (IMSL) subroutine GGNSM (International Mathematical and Statistical Library, 1987).

Sawilowsky and Blair (1992) investigated the robustness of Student's t statistic, for both independent and correlated samples, using eight nonnormal distributions identified by Micceri (1989) as representative of those found in educational and psychological data. They found that the Type I error rates for the t statistics were affected only under conditions of skewness where $\gamma_1 = 1.64$. Therefore, the nonnormal data for the current study were obtained from a χ^2 distribution with three df, for which skewness and kurtosis values are $\gamma_1 = 1.63$ and $\gamma_2 = 4.00$, respectively. This distribution is skewed to the right. The IMSL subroutine GGCHS (International Mathematical and Statistical Library, 1987) was used to generate deviates following a univariate χ^2 distribution, which were then standardized to have a mean of 0 and a variance of 1. The corresponding multivariate observations were obtained by a triangular decomposition of Σ_j, which is often referred to as the Cholesky factorization or the square root method (Harman, 1976), that is,

$$\mathbf{Y}_{ij} = \mu_j + \mathbf{L}\,\mathbf{Z}_{ij} \quad ,$$

where $\mathbf{L}$ is a lower triangular matrix satisfying the equality $\Sigma_j = \mathbf{L}\mathbf{L}^T$ and $\mathbf{Z}_{ij}$ is a K x 1 vector of χ^2 variates.

<u>Nature of the Null Hypothesis</u>

Empirical FWRs for the ten procedures were obtained under a complete null hypothesis, when all of the γ_{jk}s were equal, and under a partial null hypothesis, when not all γ_{jk}s were equal. The FWR was defined as the probability that at least one tetrad contrast was statistically significant when the corresponding population contrast was null. Seaman, Levin, and Serlin (1991) investigated the FWRs for a number of independent sample MCPs, and found that the error rates were generally lower under a partial null than under a complete null hypothesis. Keselman (1993, 1994) reported similar findings for RM marginal mean comparisons in a mixed design. Since a researcher can never know the nature of the null hypothesis under investigation for a given set of data, it is advisable to select a procedure which can maintain the FWR at α across all population mean configurations.

Effect Size

While power is generally defined as the probability of rejecting a false null hypothesis, as Ramsey (1978) notes, in multiple comparison testing situations there is more than one definition which may be adopted. In the current study, both all-comparison and per-comparison power were investigated. Ramsey defined all-comparison power as the probability of correctly rejecting all nonnull contrasts, which corresponds to the probability of making no Type II errors (i.e., no false acceptances). Per-comparison power is defined as the probability of rejecting a particular nonnull contrast.

Both per-comparison and all-comparison power rates were collected for various values of the effect size for the omnibus test of the interaction. The effect size is given by $f = \sigma_m/\sigma_e$, where $\sigma_m = \{\Sigma_j\Sigma_k\gamma_{jk}^2/(J - 1)(K - 1) + 1\}^{½}$ and σ_e is the positive square root of the difference between the average variance and average covariance for the pooled variance-covariance matrix (Cohen, 1988). The effect size was arbitrarily set equal to 0.50, 1.00 and 1.50. The population cell means used to obtain these three effects, for K = 4 and K = 8, can be found in Tables 13 and 14, respectively.

Verfication of the Simulation Program

The simulation program was written in the FORTRAN program language. To verify the accuracy of the random number generation process, a set of 50,000 observations was generated, first with the GGNSM subroutine and then with GGCHS subroutine, for a specified population variance-covariance matrix and mean vector. Variances, covariances, and means were computed for this set of data. In both cases, the computed statistics were close to the population parameters, indicating satisfactory performance of the program.

Design of the Simulation Study

The ten tetrad contrast procedures were evaluated for the six pairings of group sizes and group variance-covariance matrices under each possible combination of RM factor levels, total sample size, sphericity pattern, and degree of normality. Five thousand replications of each condition were performed using a .05 significance level. For each replication, all possible tetrad contrasts were computed on the data and each MCP was applied to the calculated test statistics

Table 13

Population Means (μ_{jk}s) for Interaction Effect (J = 3; K = 4)

f[a] = 0.50				
	.301904	.301904	1.50952	1.50952
	.301904	.301904	1.50952	1.50952
	3.01904	3.01904	.603808	.603808
f = 1.00				
	.60381	.60381	3.01904	3.01904
	.60381	.60381	3.01904	3.01904
	6.03808	6.03808	1.20762	1.20762
f = 1.50				
	.90571	.90571	4.52856	4.52856
	.90571	.90571	4.52856	4.52856
	9.05712	9.05712	1.81142	1.81142

[a]f = Effect size

Table 14

Population Means (μ_{jk}s) for Interaction Effect (J = 3; K = 8)

f[a] = 0.50							
.3125	.3125	.3125	.3125	1.5625	1.5625	1.5625	1.5625
.3125	.3125	.3125	.3125	1.5625	1.5625	1.5625	1.5625
3.125	3.125	3.125	3.125	.625	.625	.625	.625
f = 1.00							
.625	.625	.625	.625	3.125	3.125	3.125	3.125
.625	.625	.625	.625	3.125	3.125	3.125	3.125
6.25	6.25	6.25	6.25	1.25	1.25	1.25	1.25
f = 1.5							
.9375	.9375	.9375	.9375	4.6875	4.6875	4.6875	4.6875
.9375	.9375	.9375	.9375	4.6875	4.6875	4.6875	4.6875
9.375	9.375	9.375	9.375	1.875	1.875	1.875	1.875

[a]f = Effect size

CHAPTER 6

SIMULATION STUDY RESULTS

Type I Error Rates

A quantitative measure of robustness developed by Bradley (1978) was used to evaluate the Type I error performance for the ten tetrad contrast procedures. According to Bradley's liberal criterion, a test may be considered robust if its empirical Type I error rate ($\hat{\alpha}$) falls within the range $.5\alpha \leq \hat{\alpha} \leq 1.5\alpha$. Bradley also suggested a more stringent criterion of $.9\alpha \leq \hat{\alpha} \leq 1.1\alpha$. However, the latter was deemed to be too conservative as only slightly inflated (or deflated) error rates would result in a test procedure being considered nonrobust. Hence, for the nominal .05 significance level selected for this study, a robust procedure was defined as one having an empirical FWR between .025 and .075. In the tables of Type I error results reported in this chapter, daggers (†) are used to denote values which exceed the upper limit of this bound, and asterisks (*) are used to denote values which are less than the lower limit of this bound.

Type I error rates for the five procedures which use a test statistic based on a pooled estimate of error variance are given in Tables 15 and 16 for $K = 4$ and $K = 8$, respectively. In these and subsequent tables, the abbreviations S, R, H, MSB, and M are used to denote the Scheffe (1953), Studentized maximum root, Hochberg (1988), Shaffer (1986), and Studentized maximum modulus procedures, respectively. The data associated with a complete null hypothesis when $N = 30$ is provided. Since the FWRs for the H and MSB procedures were identical under a complete null hypothesis, these values have not been reported separately.

What is immediately apparent from the tabled values is that none of the procedures could control the FWR under violation of either part of the multisample sphericity assumption. This finding holds for balanced (conditions a and b), as well as unbalanced designs (conditions c, c′, d, and d′), with the discrepancy between the nominal and empirical values being greatest for unbalanced designs.

It is also important to note that even when the multisample sphericity assumption was satisfied (i.e., $\epsilon = 1.00$; condition a), the S and R procedures produced conservative results. This was not unexpected, given findings that have been reported for Scheffe's (1953) method for pairwise comparisons in independent groups designs (e.g., Carmer & Swanson, 1973). Furthermore, given that both the S and R procedures are designed to control the error rate for much larger families of contrasts than the set of all possible tetrad contrasts, they will tend to perform less optimally than other procedures.

The M and H/MSB procedures which used a pooled test statistic were very liberal under large departures from the assumption of sphericity of the pooled covariance matrix (i.e., $\epsilon = .40$), attaining FWRs as high as .24 for K = 4 (see Table 15). The empirical values became even more extreme when the number of levels of the RM factor was increased to eight; the largest value attained was .31 (see Table 16). The nonnormal values were not consistently larger or smaller than their

Table 15

Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Complete Null Hypothesis; K = 4; N = 30)

		Normal				χ^2			
		S	R	H/MSB	M	S	R	H/MSB	M
ϵ									
	a	.008*	.016*	.037	.038	.006*	.020*	.041	.042
	b	.014*	.026	.054	.057	.013*	.026	.052	.055
	c	.007*	.015*	.031	.031	.006*	.016*	.031	.033
1.00	c′	.003*	.010*	.023*	.025*	.003*	.007*	.015*	.016*
	d	.028	.054	.091†	.094†	.025*	.046	.083†	.085†
	d′	.055	.094†	.151†	.154†	.052	.089†	.140†	.144†
	a	.024*	.041	.071	.072	.021*	.040	.071	.074
	b	.036	.061	.094†	.097†	.028	.048	.079†	.081†
	c	.018*	.034	.058	.059	.017*	.032	.058	.060
.75	c′	.012*	.019*	.036	.037	.010*	.020*	.035	.036
	d	.058	.085†	.128†	.132†	.048	.077†	.120†	.124†
	d′	.073	.111†	.166†	.170†	.077†	.116†	.171†	.174†
	a	.070	.100†	.151†	.154†	.061	.090†	.151†	.154†
	b	.073	.106†	.146†	.147†	.070	.098†	.142†	.145†
	c	.054	.077†	.110†	.112†	.052	.078†	.108†	.111†
.40	c′	.040	.057	.083†	.085†	.039	.056	.083†	.085†
	d	.111†	.144†	.190†	.192†	.106†	.147†	.192†	.196†
	d′	.139†	.183†	.237†	.241†	.138†	.178†	.228†	.231†

Note: S = Scheffe (1953); R = Studentized maximum root; H/MSB = Hochberg (1988) step-up Bonferroni/Shaffer (1986) modified sequentially rejective Bonferroni; M = Studentized maximum modulus; a = pairings of equal covariance matrices and equal group sizes; b = pairings of unequal covariance matrices and equal group sizes; c/c′ = positive pairings of covariance matrices and group sizes [c: n_j = 8, 10, 12; c′: n_j = 6, 10, 14]; d/d′ = negative pairings of covariance matrices and group sizes [d: n_j = 12, 10, 8; d′: n_j = 14, 10, 6]; * = empirical value < .025; † = empirical value > .075.

Table 16

Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Complete Null Hypothesis; K = 8; N = 30)

		Normal				χ^2			
		S	R	H/MSB	M	S	R	H/MSB	M
ϵ									
	a	.000*	.002*	.036	.038	.001*	.002*	.038	.040
	b	.001*	.008*	.068	.070	.001*	.010*	.069	.070
	c	.000*	.003*	.038	.039	.001*	.004*	.035	.036
1.00	c′	.000*	.001*	.020*	.021*	.000*	.001*	.024*	.025*
	d	.003*	.019*	.115†	.118†	.004*	.020*	.120†	.123†
	d′	.007*	.042	.208†	.211†	.014*	.046	.208†	.214†
	a	.002*	.009*	.066	.067	.002*	.009*	.078†	.077†
	b	.004*	.017*	.101†	.103†	.002*	.016*	.095†	.098†
	c	.001*	.009*	.054	.055	.002*	.010*	.057	.058
.75	c′	.000*	.005*	.037	.038	.001*	.003*	.034	.035
	d	.008*	.033	.145†	.148†	.007*	.031	.144†	.147†
	d′	.012*	.066	.213†	.218†	.020*	.064	.228†	.233†
	a	.016*	.046	.168†	.170†	.018*	.045	.165†	.168†
	b	.028	.062	.175†	.178†	.019*	.059	.176†	.179†
	c	.013*	.041	.123†	.125†	.013*	.034	.133†	.135†
.40	c′	.011*	.033	.102†	.104†	.005*	.027	.091†	.093†
	d	.042	.099†	.235†	.237†	.036	.085†	.235†	.237†
	d′	.071	.140†	.308†	.311†	.066	.134†	.308†	.312†

Note: See the note from Table 15.

normal counterparts. In general however, Type I error rates were only slightly affected by departures from the multivariate normality assumption.

The results associated with N = 45 for the pooled procedures have not been tabled here, but are contained in Appendix B in Tables B1 and B2. The FWRs are comparable to those obtained for N = 30. In general, however, the error rates associated with this larger sample size tend to be slightly less deviant, so that there was less discrepancy between empirical and nominal values, particularly for the normal data. For the nonnormal data this is not always the case. As a consequence, error rates are still liberal under departures from the multisample sphericity assumption; the maximum value attained was .34, when K = 8.

Error rates for the five procedures which employ a test statistic based on a nonpooled estimate of error variance are contained in Tables 17 and 18. Again, only those results associated with the complete null hypothesis for N = 30 are reported here.

Table 17 reveals that regardless of whether the data were normally or nonnormally distributed, none of the procedures which employ a nonpooled test statistic were liberal when the RM factor had four levels. When the multivariate normality assumption was satisfied, the empirical values for the M procedure were consistently larger than those obtained for the H/MSB procedures. However, all three procedures occasionally resulted in conservative values when the data was highly nonspherical (ϵ = .40). The R error rates exceeded those for the S procedure across all conditions, but nevertheless were largely conservative, particularly for

Table 17

Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis; K = 4; N = 30)

		Normal				χ^2			
		S	R	H/MSB	M	S	R	H/MSB	M
ϵ									
	a	.013*	.023*	.038	.043	.008*	.019*	.029	.032
	b	.013*	.025	.039	.045	.010*	.022*	.032	.037
	c	.011*	.025	.040	.046	.008*	.019*	.029	.033
1.00	c′	.011*	.023*	.037	.043	.007*	.018*	.029	.034
	d	.019*	.037	.047	.057	.017*	.030	.041	.049
	d′	.030	.046	.055	.063	.024*	.040	.047	.059
	a	.011*	.024*	.038	.042	.007*	.016*	.026	.033
	b	.016*	.030	.045	.053	.009*	.019*	.027	.033
	c	.012*	.024*	.034	.041	.009*	.021*	.031	.035
.75	c′	.011*	.023*	.036	.042	.008*	.018*	.029	.034
	d	.019*	.035	.046	.052	.014*	.024*	.034	.042
	d′	.031	.042	.052	.063	.020*	.034	.047	.055
	a	.007*	.013*	.022*	.025	.005*	.009*	.013*	.015*
	b	.006*	.013*	.019*	.023*	.007*	.015*	.020*	.023*
	c	.007*	.015*	.023*	.025	.006*	.013*	.018*	.021*
.40	c′	.006*	.014*	.020*	.022*	.007*	.012*	.019*	.021*
	d	.014*	.022*	.028	.031	.014*	.022*	.032	.036
	d′	.018*	.026	.032	.038	.020*	.031	.036	.043

Note: See the note from Table 15.

nonspherical data. The values for the S procedure rarely exceeded the lower bound of Bradley's (1978) criterion.

The results for the nonnormal data in Table 17 reveal generally lower FWRs as compared with those for normal data. In particular, the S values were conservative for all 18 of the conditions investigated. The R procedure only exceeded the lower bound of Bradley's (1978) criterion for negative pairings of group sizes and covariance matrices. Finally, the H/MSB and M methods were largely conservative under extreme degrees of nonsphericity, with minimum values of .013 and .015 respectively.

The $K = 8$ results associated with the smaller sample size, for both normal and nonnormal data, are contained in Table 18. With respect to the error rates obtained when the multivariate normality assumption was satisfied, the S method was very conservative for all of the conditions investigated, with a mean FWR of .003. The R procedure only surpassed Bradley's (1978) lower bound when the most disparate group sizes were negatively paired with covariance matrices (condition d′), across all values of ϵ. For the most part, the H/MSB and M procedures provided good control of the FWR when the RM factor had eight levels. However, the latter tended to be liberal for this same d′ condition, attaining a maximum value of .106. The empirical FWRs for the stepwise Bonferroni procedures which used a nonpooled test statistic were only slightly greater than the upper bound of Bradley's criterion when $\epsilon = 1.0$ for this negative pairing condition (i.e., $\hat{\alpha} = .076$).

Table 18

Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis; K = 8; N = 30)

		Normal				χ^2			
		S	R	H/MSB	M	S	R	H/MSB	M
ϵ									
	a	.002*	.008*	.037	.050	.000*	.003*	.020*	.029
	b	.004*	.012*	.043	.055	.001*	.005*	.027	.037
	c	.002*	.010*	.036	.050	.001*	.005*	.024*	.038
1.00	c′	.002*	.008*	.035	.049	.001*	.003*	.021*	.031
	d	.003*	.015*	.047	.072	.002*	.010*	.042	.057
	d′	.014*	.038	.076†	.106†	.007*	.020*	.050	.073
	a	.001*	.008*	.035	.048	.001*	.003*	.022*	.029
	b	.003*	.015*	.046	.061	.001*	.005*	.029	.040
	c	.001*	.010*	.033	.047	.000*	.005*	.029	.040
.75	c′	.001*	.009*	.036	.048	.001*	.005*	.021*	.030
	d	.005*	.018*	.047	.067	.004*	.010*	.040	.054
	d′	.023*	.033	.069	.097†	.007*	.021*	.054	.078†
	a	.001*	.006*	.029	.036	.001*	.004*	.021*	.027
	b	.001*	.008*	.032	.043	.002*	.008*	.029	.039
	c	.001*	.006*	.028	.037	.002*	.006*	.023*	.029
.40	c′	.001*	.007*	.027	.035	.001*	.004*	.017*	.025
	d	.004*	.011*	.039	.053	.004*	.014*	.040	.051
	d′	.011*	.027	.055	.075	.011*	.031	.061	.081†

Note: See the note from Table 15.

As seen from Table 18, the FWRs were generally lower for nonnormal data than for normal data when K = 8, consistent with the findings for K = 4. The M procedure was slightly liberal for negative pairings of the more disparate group sizes and unequal covariance matrices when $\epsilon < 1.0$, attaining a maximum value of .081, but otherwise provided good FWR control. The H/MSB nonpooled procedures were slightly conservative for all values of ϵ when the design was balanced and covariance matrices were equal, and when group sizes and covariance matrices were positively paired, with a minimum value of .017.

The N = 45 results for the nonpooled tetrad contrast procedures have not been tabled here, but can be found in Appendix B in Tables B3 and B4. As reported for the pooled procedures, error rates tended to be less deviant from the .05 significance level as compared with the values associated with N = 30, and as a result, no liberal values are found in these two tables.

Table 19 contains the Type I error rates, averaged across both sample size conditions, as well as the normality and ϵ conditions, for both K = 4 and K = 8. As revealed in this table, the S procedure was conservative for both balanced and unbalanced conditions. The R procedure only exceeded the lower bound of Bradley's (1978) criterion once, when group sizes and covariance matrices were negatively paired and group sizes reflected the larger degree of imbalance. On the other hand, the average FWRs for the H, MSB, and M procedures were neither conservative nor liberal, but the error rates associated with the M procedure were always greater than those associated with the former two procedures. The maximum value attained was

Table 19

Average Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis)

	K = 4				K = 8			
	S	R	H/MSB	M	S	R	H/MSB	M
a	.008*	.016*	.028	.031	.001*	.004*	.028	.036
b	.010*	.020*	.031	.035	.001*	.007*	.035	.043
c	.008*	.018*	.030	.033	.001*	.006*	.030	.039
c′	.009*	.018*	.029	.033	.001*	.005*	.029	.037
d	.013*	.024*	.037	.042	.002*	.010*	.040	.053
d′	.019*	.032	.042	.048	.007*	.021*	.053	.071
μ	.011*	.021*	.033	.037	.002*	.009*	.036	.046

Note: See the note from Table 15; The μ values represent empirical rates that have been averaged across all conditions.

.071, and was associated with the d′ condition. Finally, the difference between the average values for the H/MSB and M procedures was greater when K = 8 (i.e., mean difference = .01) than when K = 4 (i.e., mean difference = .004).

The data obtained for the tetrad contrast procedures under a partial null hypothesis have not been reported here, but are available in Appendix C. Separate results are reported for the H and MSB procedures in these tables. Trends in findings were similar to those reported for the data tabled in this chapter, and in Appendix B. The empirical FWRs for the procedures employing a pooled test statistic were largely conservative when K = 4 for both values of total sample size. However, error rates for the H, MSB, and M procedures did exceed the upper bound of Bradley's (1978) liberal criterion when very unequal group sizes and covariance matrices were negatively paired (i.e., condition d′), even when ϵ = 1.0. When the number of RM factor levels was increased to eight, the error rates for the H, MSB, and M procedures were typically liberal for both the d and d′ conditions. The maximum value obtained was .229.

For the nonpooled procedures when K = 4, error rates were consistently lower than the .05 level of significance for a partial null hypothesis for both values of total sample size. As expected, and consistent with previous findings for partial null hypotheses (Keselman, 1993, 1994; Seaman et al., 1991), the FWRs were generally either less than the lower bound of Bradley's (1978) liberal criterion, or approached it in value. However, in contrast with the findings for the complete null hypothesis, the values for both Bonferroni procedures were marginally larger than the M values across several conditions, for both normal and nonnormal data, when K = 4. This pattern was not evident when K = 8, for which the latter always produced larger FWRs than the H and MSB procedures.

Power Rates

Since the procedures employing a pooled test statistic could not provide FWR control under violations of the multisample sphericity and normality assumptions, only the procedures which used a nonpooled test statistic are considered with respect to power. In comparing the power rates for the five procedures, Einot and Gabriel's (1975) criterion of denoting power differences (PDs) less than .10 as negligible, and differences greater than .20 as substantial will be used. The power values contained in Tables 20 and 21 have been averaged across population effect sizes, values of ϵ, and population shapes, since the individual empirical power rates obtained for various combinations of these conditions followed the same trends that will be highlighted in the ensuing discussion. Also contained in these tables are the mean power values (i.e., μ values), which have been averaged across all investigated conditions.

Table 20 reveals that in terms of average per-comparison power rates, there were negligible differences between the five nonpooled procedures when K = 4. As expected, the S procedure was the least powerful, but only slightly less powerful than the R procedure. By comparing the μ values for these two procedures it is apparent that the PD was approximately .04 for both values of N. Overall, the M procedure was marginally less powerful than either stepwise Bonferroni procedure. The

Table 20

Empirical Average Per-Comparison Power Rates for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Averaged Across Effect Sizes, Sphericity Values, and Population Shapes)

	N = 30					N = 45				
	S	R	H	MSB	M	S	R	H	MSB	M
					K = 4					
a	.65	.69	.73	.74	.72	.76	.79	.83	.84	.82
b	.60	.64	.69	.70	.68	.72	.76	.80	.81	.79
c	.64	.68	.72	.73	.72	.75	.79	.82	.83	.82
c′	.66	.70	.74	.75	.73	.77	.80	.84	.84	.83
d	.53	.58	.62	.63	.62	.68	.72	.76	.76	.75
d′	.42	.47	.49	.50	.51	.59	.64	.67	.68	.67
μ	.58	.63	.66	.68	.66	.71	.75	.79	.80	.78
					K = 8					
a	.46	.55	.65	.66	.66	.62	.69	.78	.78	.78
b	.40	.50	.60	.60	.61	.56	.65	.74	.74	.74
c	.45	.54	.65	.65	.65	.61	.68	.77	.77	.77
c′	.47	.57	.67	.67	.67	.63	.70	.78	.79	.78
d	.36	.42	.48	.51	.55	.50	.59	.68	.68	.69
d′	.22	.30	.34	.34	.41	.39	.48	.57	.58	.60
μ	.39	.48	.56	.57	.59	.55	.63	.72	.72	.73

Note: See the notes from Tables 15 and 19.

exception was for N = 30 when the most disparate group sizes were negatively paired with group covariance matrices. The MSB procedure was more powerful than the H method, although only by a very small degree; the difference in the μ values was approximately .01 for both values of N.

Further examination of the K = 4 results in Table 20 reveal that the nature of the pairing of group sizes and covariances had a substantial impact on power. Empirical power rates were greatest when the most disparate group sizes were positively paired with covariance matrices (i.e., condition c′), and smallest for negative pairings of very unequal sample sizes and covariance matrices (i.e., condition d′). To illustrate, for Shaffer's procedure, the PD between these two conditions is .24 for N = 30 and .16 for N = 45.

The K = 8 data in Table 20 show greater discrepancies between the tetrad contrast procedures in terms of average per-comparison power rates. When N = 30, the M procedure was most powerful, as reflected in the μ values. The PDs were .20 and .11 for this procedure and the S and R procedures, respectively. However, based on the average power rates, the differences between the M and stepwise Bonferroni procedures were small. For the MSB method the difference was .02, while for the H procedure it was .03. The N = 45 power values reveal even smaller differences between the M and Bonferroni procedures, although the superiority of the former over both the S and R methods was similar.

Table 21 contains the all-comparison power rates for the five nonpooled procedures. The μ values for K = 4 reveal that the PD between the most powerful

Table 21

Empirical All-Comparison Power Rates for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Averaged Across Effect Sizes, Sphericity Values, and Population Shapes)

	N = 30					N = 45				
	S	R	H	MSB	M	S	R	H	MSB	M
					K = 4					
a	.36	.42	.49	.50	.47	.54	.58	.64	.64	.62
b	.31	.36	.44	.45	.41	.50	.54	.60	.61	.58
c	.36	.41	.49	.50	.46	.54	.58	.63	.64	.61
c′	.38	.44	.51	.52	.49	.55	.59	.65	.65	.63
d	.23	.28	.36	.36	.33	.42	.48	.55	.56	.52
d′	.13	.17	.21	.22	.20	.31	.37	.44	.45	.42
μ	.30	.35	.42	.42	.39	.48	.52	.58	.59	.56
					K = 8					
a	.07	.19	.28	.28	.28	.26	.39	.52	.52	.50
b	.03	.12	.22	.22	.21	.20	.32	.44	.44	.43
c	.06	.18	.28	.28	.27	.25	.38	.50	.50	.48
c′	.08	.21	.30	.30	.30	.27	.41	.53	.53	.51
d	.04	.05	.15	.10	.15	.11	.24	.35	.35	.34
d′	.01	.01	.01	.01	.03	.03	.11	.19	.19	.20
μ	.05	.13	.21	.20	.21	.19	.31	.42	.42	.41

Note: See the notes from Tables 15 and 19.

procedure, MSB, and the least powerful procedure, S, was only slightly greater than .10 for both values of total sample size. As well, the data again reveal substantial differences in power rates for positive (i.e., the c′ condition) and negative (i.e., the d′ condition) pairings of group sizes and covariance matrices for all procedures. For the MSB procedure, a PD of .30 exists between these two conditions for N = 30, and this difference was .20 for N = 45.

The K = 8 data reveal that the H and MSB procedures were essentially equally powerful for both N = 30 and N = 45. Overall, the differences between these two procedures and S were close to .20 for N = 30, and slightly greater than this criterion for N = 45. However, the R method was only slightly less powerful than either H or MSB according to Einot and Gabriel's (1975) criteria. For N = 45, rates for both the H and MSB procedures only slightly exceeded the M procedure rates for all but the d′ condition, where the latter was slightly more powerful (PD = .01).

Synopsis of Simulation Results

As anticipated, those procedures which employed a pooled estimate of error variance could not control the FWR to α under departures from multisample sphericity, particularly when the design was unbalanced. As well, the Scheffe (1953) procedure which relied on a nonpooled test statistic was predictably conservative. In keeping with theoretical expectations, the Studentized maximum root procedure yielded higher rates of error than the Scheffe procedure, but the majority of conditions investigated did not produce values surpassing the lower bound of

Bradley's (1978) liberal criterion (i.e., $\hat{\alpha}$ of .025). For the most part, the procedures which utilized a nonpooled test statistic and either a Hochberg (1988) step-up Bonferroni, Shaffer (1986) modified sequentially rejective Bonferroni, or Studentized maximum modulus critical value provided good control of the FWR under violations of multisample sphericity in unbalanced designs, even when the data were sampled from a nonnormal population. However, all three of these procedures became quite conservative under the combined effects of nonnormality, large departures from sphericity, and small values of K.

Power differences among the five procedures which used a nonpooled test statistic were not large when the number of levels of the repeated measures factor was small and according to Einot and Gabriel's (1975) criteria, most would be considered negligible. However differences between the least and most powerful procedures became more pronounced when K was increased in value. Based on the recommendations of Einot and Gabriel, Scheffe's procedure could be declared substantially less powerful than all but the Studentized maximum root procedure under most of the conditions investigated. Among the most powerful procedures, Hochberg (1988), Shaffer (1986), and the Studentized maximum modulus, there was no uniform superiority in terms of either average per-comparison or all-comparison rates. The Studentized maximum root procedure was never substantially less powerful than these procedures, but always had marginally less power for detecting nonnull tetrad contrasts.

CHAPTER 7

APPLICATIONS, SUMMARY, AND CONCLUSIONS

Applications of Results

As Seaman et al. (1991) have noted, the adoption of any statistical procedure by applied researchers is largely dependent on its practicality. Based on the results of the simulation study, it would appear that the best procedure to use when conducting tetrad contrasts would be one which incorporates a nonpooled test statistic in combination with either a Hochberg (1988) step-up Bonferroni, Shaffer (1986) modified sequentially rejective Bonferroni, or Studentized maximum modulus CV. The purpose of the following discussion is to illustrate the application of these procedures for probing interactions using data for a mixed design. The procedures will be considered for a mixed design containing a single independent groups factor and a single RM factor. To place these results in a context meaningful to C&T researchers, the example will be based on a real study by Lawson and Lorentzen (1990).

In Lawson and Lorentzen's (1990) research, a number of different sports bras were evaluated for perceived comfort and support by women who had been classified into groups on the basis of bra cup size. Although the actual data for this experiment are not available, Appendix D contains a set of hypothetical data for the perceived support dependent variable. As the authors noted, support scores were based on the sum of ten 5-point Likert scales; total scores for each subject could range from ten to 50. with higher scores indicating a more favourable evaluation of support. In keeping with the results reported by Lawson and Lorentzen, the design is unbalanced, and the group sizes selected for illustrative purposes, according to bra cup size, are: $n_A = 9$, $n_B = 13$, $n_C = 8$, and $n_D = 5$. Further, suppose that this hypothetical data set is based on the results obtained for three different bras (i.e., $K = 3$). Although Lawson and Lorentzen did not make any statements regarding random assignment of the bras to each study participant, for purposes of this example it is assumed that randomization was employed as a means to reduce possible carry-over effects in the evaluation of bra support.

Lix and Keselman (in press) developed a statistical program written in the SAS/IML (SAS Institute Inc., 1989a) programming language which uses the GLM approach described in Chapter 4 to develop one or more hypotheses and test them using Johansen's (1980) approximate df solution. To conduct the set of all possible tetrad contrasts for the example data set in Appendix D using a nonpooled statistic, the researcher must specify a contrast vector for both the grouping factor and the RM factor and each of these must form a pairwise contrast on the levels of the appropriate factor. In using Lix and Keselman's statistical program, it is assumed that the data are entered in a particular order, so that the scores associated with the subjects in Group A are followed by the scores for the subjects in Group B, and so on. Appendix D contains the SAS/IML program written by Lix and Keselman and the additional programming statements required to conduct all possible tetrad contrasts.

Table 22 contains the means and variance-covariance matrices for the repeated measurements for the four groups of subjects. Figure 1 contains a plot of the mean support scores for each bra condition and cup size group. A variety of symbols are used to denote the mean values associated with a particular group and lines connect these symbols. Such plots of the data may be conducted using any graphics or statistical software package. The nonparallel lines in this graph are useful in illustrating the existence of an interaction between cup size and type of bra and provide a visual representation of the nature of the data. This graph illustrates that individuals in the B and C groups gave very similar evaluations of support to all of the bra conditions, while individuals in groups A and D responded in a much different manner.

Table 23 contains the tetrad contrast t statistics, df, and p values produced using the program developed by Lix and Keselman (in press). It is important to note that the program computes F statistics (i.e., Equation 4.5), but these are easily converted to t statistics via the relationship $t = \sqrt{F}$.

Table 24 provides the significance criteria for each of the Hochberg (1988), Shaffer (1986), and Studentized maximum modulus methods for the set of 18 tetrad contrasts. In this particular example all three methods produce the same significant results, however, this may not always be the case.

With Hochberg's (1988) step-up Bonferroni method, after ranking the p values corresponding to the nonpooled t statistics, one begins by comparing the largest p value, which corresponds to the t statistic involving the second and third levels of the grouping factor and the first and second levels of the RM factor (i.e., $p_{(B-C;\ 1-2)}$), to

Table 22

Means ($\bar{Y}$) and Variance-Covariance Matrices ($\hat{\Sigma}$) for Hypothetical Data Set

	BR1	BR2	BR3
Group A (n_A = 9)			
$\bar{Y}_A$	20.556	35.778	45.667
$\hat{\Sigma}_A$	9.778	6.014	5.333
		15.194	4.542
			9.750
Group B (n_B = 13)			
$\bar{Y}_B$	14.385	22.769	36.308
$\hat{\Sigma}_B$	12.090	3.346	3.122
		10.859	-1.756
			10.731
Group C (n_C = 8)			
$\bar{Y}_C$	13.875	22.50	33.875
$\hat{\Sigma}_C$	4.125	-2.357	2.268
		10.000	2.500
			13.554
Group D (n_D = 5)			
$\bar{Y}_D$	25.200	44.600	27.600
$\hat{\Sigma}_D$	13.700	4.600	-3.900
		14.800	2.300
			2.300

Note: BR1 - BR3 = Bra conditions 1 through 3.

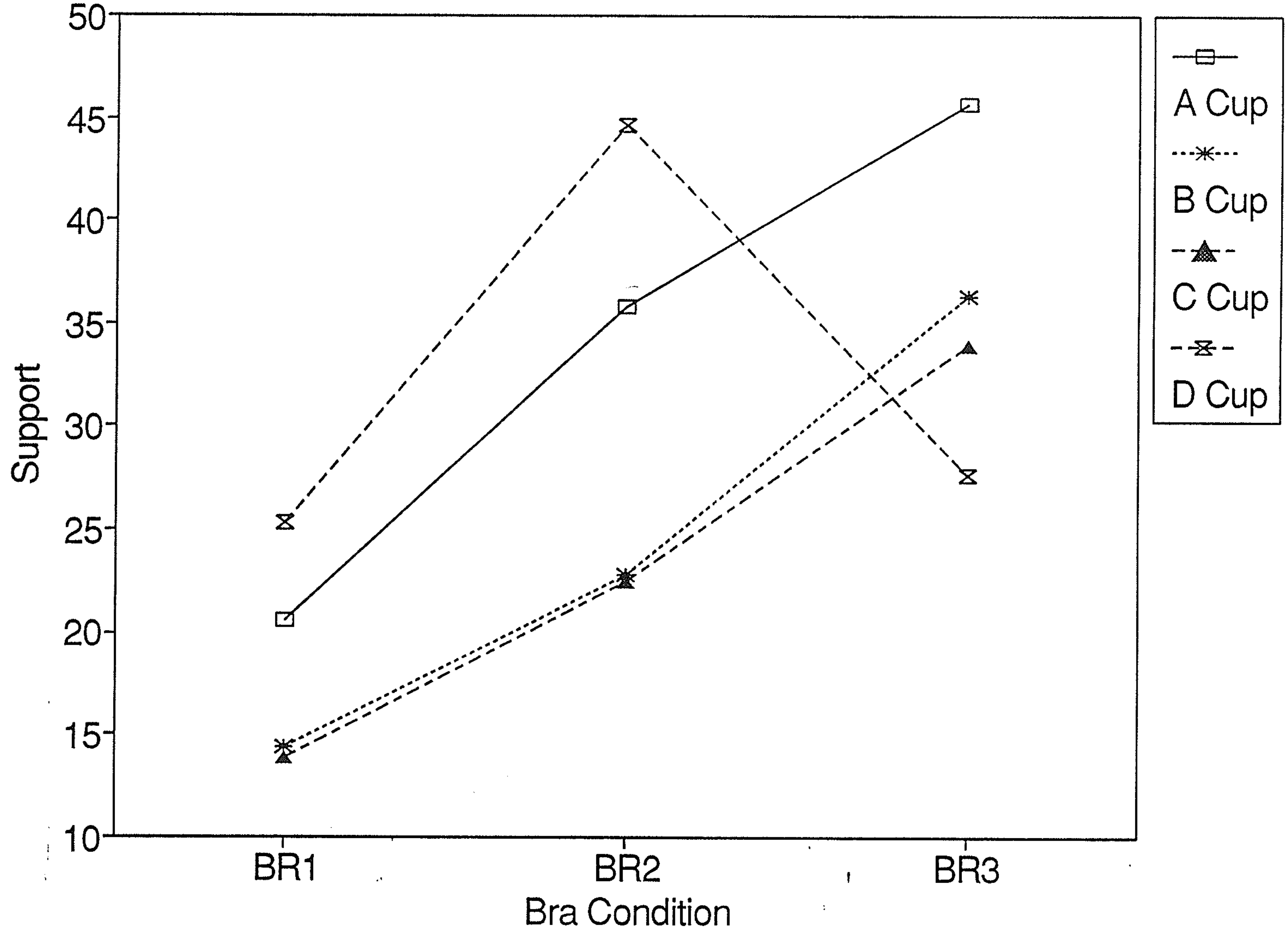

Figure 1. Graphical Representation of Mean Support Scores by Cup Size and Bra Condition

Table 23

Tetrad Contrast Results for Hypothetical Data Set

Group Factor Levels in Tetrad Contrast		RM Factor Levels in Tetrad Contrast: BR1-BR2	BR1-BR3	BR2-BR3
A-B	t	4.170	2.121	1.900
	ν_s	18.590	19.894	19.516
	p	.0005	.0467	.0724
A-C	t	3.387	3.153	3.795
	ν_s	13.692	13.623	14.403
	p	.0045	.0073	.4738
A-D	t	1.815	9.476	13.024
	ν_s	7.047	5.704	9.313
	p	.1121	.0001	< .0001
B-C	t	.126	1.126	1.049
	ν_s	14.090	16.341	16.740
	p	.9010	.2765	.3089
B-D	t	4.873	7.947	14.508
	ν_s	6.774	6.280	10.481
	p	.0020	.0002	< .0001
C-D	t	4.322	6.955	12.926
	ν_s	8.551	6.776	9.964
	p	.0022	.0003	< .0001

Note: t = tetrad contrast t statistic; ν_s = error df; p = p value.

Table 24

Significance Criteria for Tetrad Contrasts

Group Factor Levels in Tetrad Contrast		RM Factor Levels in Tetrad Contrast: BR1-BR2	BR1-BR3	BR2-BR3
A-B	H	.0042*	.0071	.0083
	MSB	.0042*	.0071	.0083
	M	3.388*	3.362	3.369
A-C	H	.0056	.0062	.0250
	MSB	.0056	.0062	.0250
	M	3.542	3.544	3.513
A-D	H	.0100	.0033*	.0029*
	MSB	.0100	.0042*	.0042*
	M	4.151	4.481*	3.836*
B-C	H	.0500	.0125	.0167
	MSB	.0500	.0125	.0167
	M	3.526	3.449	3.437
B-D	H	.0045*	.0036*	.0028*
	MSB	.0050*	.0042*	.0028*
	M	4.206*	4.318*	3.730*
C-D	H	.0050*	.0038*	.0030*
	MSB	.0050*	.0042*	.0042*
	M	3.920*	4.206*	3.773*

Note: H = Hochberg's (1988) step-up Bonferroni; MSB = Shaffer's (1986) modified sequentially rejective Bonferroni; M = Studentized maximum modulus; * = significant tetrad contrast.

α = .05. This statistic is not significant, and accordingly, the corresponding hypothesis, H_o: $(\mu_{B1} - \mu_{B2}) - (\mu_{C1} - \mu_{C2}) = 0$, is retained. Testing proceeds by comparing the next-largest p value, $p_{(A\text{-}C;\ 2;3)}$ to .05/2 = .025, which is also nonsignificant. The corresponding hypothesis can not be rejected, and neither can the seven subsequent ones (in rank order) according to the criteria given in Table 24. However, $p_{(A\text{-}C;\ 1\text{-}2)}$ = .0045, which is compared to the criterion .05/9 = .0056, produces a significant result. The corresponding hypothesis, H_o: $(\mu_{A3} - \mu_{A2}) - (\mu_{C1} - \mu_{C2}) = 0$, is rejected, as are all of the remaining hypotheses which have not been tested.

According to Shaffer (1986, Table 3, p. 829), the set of tetrad contrast hypotheses that could possibly be true in a 4 x 3 design is given by A_{18} = {0 - 10, 12, 18}. This implies that for testing the *m*th hypothesis, where m = 1 ,..., 18, the significance level α/C^*_m is computed by finding the maximum value in the set A_{18} which is less than C - m + 1. Applying this rule, if one hypothesis was rejected, then at most 12 remaining hypotheses could be true (i.e., C^*_2 = 12) and the significance level for testing $p_{(2)}$ is .05/12. If, for example, seven hypotheses were rejected then, at most, 10 of the remaining hypotheses could be true and C^*_8 = 10.

With Shaffer's (1986) method, testing is also conducted on the ranked p values, but begins with the smallest p value, $p_{(B\text{-}D;\ 2\text{-}3)}$, which is evaluated with the criterion .05/18 = .0028. Since this result is significant, the corresponding hypothesis, H_o: $(\mu_{B2} - \mu_{B3}) - (\mu_{D2} - \mu_{D3}) = 0$, is rejected. Testing proceeds by comparing the next-smallest p value, $p_{(A\text{-}D;\ 2\text{-}3)}$ to .05/12 = .0042, which also produces a significant result. The next five p values (in ascending order of magnitude) are also evaluated using this criterion, and all are declared significant; the two subsequent p values, $p_{(B\text{-}D;\ 1\text{-}2)}$ and $p_{(C\text{-}D;\ 1\text{-}2)}$, which are both evaluated using the criterion .05/10 = .005, are also significant. However, the next p value, which corresponds to the hypothesis, H_o: $(\mu_{A1} - \mu_{A3}) - (\mu_{C1} - \mu_{C3}) = 0$, is compared to .05/9 = .0062. Since this value is greater than the corresponding criterion, the corresponding hypothesis is retained, as are all of the remaining hypotheses.

In order to apply the Studentized maximum modulus method to the data, a table of CVs must be consulted. Such tables are available in a number of different sources, such as Maxwell and Delaney's (1990) text, but the most comprehensive set is given by Hochberg and Tamhane (1987). Since the error df for the tetrad contrast t statistics presented in Table 23 are fractional, in order to obtain an exact CV, linear interpolation in $1/\nu_s$ is necessary, and can be accomplished using the formula

$$M_{\nu_s} = M_{\nu_{s-}} - \left[\frac{1/\nu_{s-} - 1/\nu_s}{1/\nu_{s-} - 1/\nu_{s+}}\right](M_{\nu_{s-}} - M_{\nu_{s+}}) \ , \qquad (7.1)$$

where ν_{s-} is the integer portion of ν_s, $\nu_{s+} = \nu_{s-} + 1$, and M_x is the CV for the selected df, obtained from the Studentized maximum modulus distribution.

In interpreting these findings, first consider the information conveyed by a single tetrad contrast. For example, the signficant result associated with groups A and B and the first and second bra conditions (i.e., the upper left-hand result in Table 24), indicates that individuals with A and B cup sizes differ significantly in their evaluation of the support provided by the first and second bras. The most interesting results given in Table 24 are that individuals with B and C cup sizes differ significantly in their evaluations of the differences in support offered by all of the bras; this holds true for individual wearing C and D cup sizes as well.

These results are more meaningful than those that could be obtained by conducted simple effect tests. Suppose, for example, that the researcher elected to compare support scores for the three bras within a particular cup size group. Such an analysis would give no indication of the differences in perceived support offered by different bras. Similarly, if support scores for a particular bra were compared across the four groups of study participants, this would give no indication of how women of different cup sizes differed in their evaluations of the investigated bras.

The findings of the analysis involving tetrad contrasts provides important information that could be used by manufacturers of the sports bras considered in this investigation. In order to differentiate their own product from competing products in the marketplace, manufacturers need to attend to the bra attributes which contribute to support, and provide different support features for women with different cup sizes. This appears to be particularly important for women of larger cup sizes (i.e., C and D), since women who wore A and B cup sizes tended to more similar in their perceptions of support across the different treatment conditions.

Summary of Results

The purpose of this study was twofold: (a) to examine applications of RM methodologies in the C&T literature, particularly with respect to mixed designs and

(b) to investigate various MCPs that are appropriate for probing interactions in mixed designs. This research was undertaken to add to the extremely limited body of C&T literature which has focussed on inadequacies in statistical analyses encountered in the discipline. This study also contributes to the body of knowledge on valid analyses for RM data, much of which is found in the psychological and statistical literature.

With respect to the first part of this project, a content analysis of the C&T literature revealed a number of important characteristics of the use of RM methodologies in this discipline. Research designs which incorporate correlated data are used most often by researchers who are interested in understanding how individuals perceive and evaluate clothing or textile products, but are used infrequently in studies that have a retailing or marketing focus. In recognition of the fact that many factors may simultaneously impinge on individual responses to clothing or textile products, C&T researchers typically use RM designs which contain more than a single factor, but for unknown reasons, do not take full advantage of the factorial data structure to test for the presence of variable interactions. When factorial designs are used, C&T researchers are more often interested in understanding how separate groups of individuals, who are classified on the basis of such variables as sex and age, differ in their responses. Furthermore, when interactions are tested in factorial designs, it is typically the case that researchers do not adopt procedures which will aid in the interpretation of the nature and source of the significant result. Finally, in conducting tests of correlated effects, C&T researchers usually rely on conventional methods of analysis, which are generally considered to be inappropriate due to the stringent assumptions that the data must satisfy in order for statistical validity concerns to be overcome in hypothesis testing.

Given this background, alternative, robust methods for testing hypotheses on omnibus interaction effects in mixed designs and for probing interactions were discussed. With respect to the latter issue, the use of interaction, or tetrad, contrasts was recommended, as such contrasts involve individual interaction components and are not confounded by the presence of marginal effects, as is the case with tests of simple effects.

The results of a simulation study provided empirical evidence of the extent of the bias that may result from adopting a tetrad contrast procedure which relies on a test statistic that incorporates the conventional pooled estimate of error variance and hence is dependent on the stringent assumption of multisample sphericity. None of the investigated procedures could control the familywise rate of Type I error when the multisample sphericity assumption was not satisfied, regardless of whether the design was balanced or unbalanced. However, the tetrad contrast procedures which used a nonpooled test statistic that does not assume multisample sphericity, rarely resulted in inflated Type I error rates, even under the most extreme departures from this assumption and when the data were nonnormal in form. When this test statistic was used in conjunction with either of two stepwise Bonferroni methods or a method that incorporated Studentized maximum modulus critical values, the resulting test procedures provided the greatest power to detect a nonnull effect.

Conclusions

Investigations which incorporate multiple measurements on units of analysis are popular in a wide variety of disciplines and can be useful in the study of research problems which have a C&T orientation. A problem exists, however, in the lack of adoption of procedures that will produce results which are both valid and meaningful. More specifically, in situations where mixed designs are employed, C&T researchers may not recognize that traditional methods of analyzing correlated effects are unlikely to be appropriate choices given the stringent derivational assumptions on which they rest. Furthermore, traditional strategies for examining variable interactions in such designs may not provide the information needed to adequately describe the nature and source of these interactions. The data-analytic problems identified in this study are not unique to the C&T discipline, as is evidenced by research conducted in other fields of scientific inquiry. However, they are of sufficient importance to be reemphasized in an attempt to improve methodological practice.

Although the application of omnibus procedures to test for the presence of an interaction may be a popular approach, the nonpooled tetrad contrast procedures considered in this paper should be regarded as viable and appealing methods for probing interactions. An omnibus procedure can only be regarded as a preliminary test which provides no information concerning the specific factor level combinations which contribute to the presence of a significant interaction. Through a content analysis of the C&T literature it was observed that researchers often choose to conduct tests of simple main effects following a significant overall result; from a theoretical viewpoint, the correct strategy is to conduct mean comparisons which reduce the overall interaction effect into its component parts, and, as a consequence, allow for identification of the specific source(s) of the interaction.

Often a chief concern in conducting statistical tests is the availability of a statistical software that will allow the researcher to conduct analyses of interest. The procedures for probing interaction effects considered in this paper are easily applied to data obtained from mixed designs using a recently developed SAS/IML (SAS Institute Inc., 1989a) program (Lix & Keselman, in press). The use of this program for conducting tetrad contrasts was demonstrated using an example data set.

Recommendations for Future Research

The current research project provides the basis for future research in a number of areas, including statistical knowledge of applied researchers, testing variable interactions, and approximate degrees of freedom test procedures. The following discussion considers each of these areas in turn.

One of the most important, but not unexpected, findings of the current review of the C&T RM literature was that researchers continue to cling to conventional methods of analyzing correlated effects, even though a series of studies have shown such procedures to be inappropriate in the majority of data-analytic situations. A study of the knowledge base and statistical decision-making strategies of researchers who are likely to make use of RM designs might help to pinpoint reasons for failing to consider alternative procedures and any barriers to the use of such procedures. For example, researchers may be aware of the derivational assumptions underlying the statistical tests which they use, but may not perceive violations of these assumptions as being relevant in their data analysis. However, as Moran (1986) has noted, researchers should be encouraged to take a proactive role in removing the potential for errors to occur in reported results.

The simulation study which investigated the use of tetrad contrasts for probing interaction effects was limited to a situation in which only a single dependent variable was studied. Since many research problems are concerned with the simultaneous investigation of more than one dependent variable, it would be beneficial to consider the operating characteristics of the recommended procedures in multivariate mixed designs (e.g., Robey & Barcikowski, 1986). Such a study could provide relevant information to C&T researchers, as Damhorst (1990) has suggested that multivariate designs should routinely be adopted because they will provide answers to the kinds of complex research questions which C&T researchers should be addressing to promote theory development in the field.

Furthermore, while the tetrad contrast procedures considered in this project were only marginally affected by the degree and form of nonnormality investigated, it would be worthwhile to investigate procedures that may be robust to the effects of nonnormality. Wilcox (1993) suggests that more extreme degrees of skewness than that considered in this study are likely to be encountered in social science data. He considered the application of Yuen's (1974) method for trimming aberrant scores from the sample data (see also Yuen & Dixon, 1973) prior to computing an omnibus test statistic in the simple RM design, and recommends this approach for a variety of forms of nonnormal data. However, Yuen's approach was designed for use with symmetric distributions in which nonnormality arises because of outliers and full consideration how such an approach might operate for extreme degrees of skewness has not yet been undertaken.

REFERENCES

Algina, J., & Oshima, T. C. (1994). Type I error rates for Huynh's general approximation and improved general approximation tests. British Journal of Mathematical and Statistical Psychology, 47, 151-165.

Barcikowski, R. S., & Robey, R. R. (1984). Decisions in single group repeated measures analysis: Statistical tests and three computer packages. American Statistician, 38(2), 148-150.

Barcikowski, R., & Robey, R. R. (1990, April). Interpreting repeated measures data through "special" mean contrasts. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Ball, D., McKenry, P. C., & Price-Bonham, S. (1983). Use of repeated-measures designs in family research. Journal of Marriage and the Family, 45, 885-896.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. Proceedings of the Royal Society, A901, 268-282.

Bartlett, M. S. (1939). A note on tests of significance in multivariate analysis. Proceedings of the Cambridge Philosophical Society, 35, 180-185.

Belli, G. M. (1988, April). Type I error rates of MANOVA of repeated measures under group heterogeneity in unbalanced designs. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Boik, R. J. (1975). Interactions in the analysis of variance: A procedure for interpretation and a Monte Carlo comparison of univariate and multivariate methods for repeated measures designs. Dissertation Abstracts International, 36, 2908B. (University Microfilms No. 75-27,837).

Boik, R. J. (1991). Scheffe's mixed model for multivariate repeated measures: A relative efficiency evaluation. Communications in Statistics, Theory and Methods, 20, 1233-1255.

Boik, R. J. (1993). The analysis of two-factor interactions in fixed effects linear models. Journal of Educational Statistics, 18, 1-40.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and correlation between errors in the two-way classification. Annals of Mathematical Statistics, 25, 484-498.

Bradley, J. V. (1978). Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144-152.

Bradu, D., & Gabriel, K. R. (1974). Simultaneous statistical inference on interactions in two-way analysis of variance. Journal of the American Statistical Association, 69, 428-436.

Brigham, F. R. (1974). A critical note on the application of analysis of variance techniques in Ergonomics research. Ergonomics, 17, 259-265.

Carmer, S. G., & Swanson, M. R. (1973). An evaluation of ten multiple comparison procedures by Monte Carlo methods. Journal of the American Statistical Association, 68, 66-74.

Cicchetti, D. V. (1972). Extension of multiple-range tests to interaction tables in the analysis of variance: A rapid approximate solution. Psychological Bulletin, 77, 405-408.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cole, J. W. L., & Grizzle, J. E. (1966). Applications of multivariate analysis to repeated measurement experiments. Biometrics, 22, 810-828.

Collier, R. O., Baker, F. B., Mandeville, G. K., & Hayes, T. F. (1967). Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. Psychometrika, 32, 339-353.

Cornell, J. E., Young, D. M., Seaman, S. L., & Kirk, R. E. (1992). Power comparisons of eight tests for sphericity in repeated measures designs. Journal of Educational Statistics, 17, 233-249.

Damhorst, M. L. (1990). In search of a common thread: Classification of information communicated through dress. Clothing and Textiles Research Journal, 8(2), 1-12.

Davidson, M. L. (1972). Univariate versus multivariate tests in repeated measures experiments. Psychological Bulletin, 77, 446-452.

Duncan, D. B. (1955). Multiple range and multiple F tests. Biometrics, 11, 1-42.

Dunn, O. J. (1961). Multiple comparisons among means. Journal of the American Statistical Association, 56, 52-64.

Dunnett, C. W., & Tamhane, A. C. (1992). A step-up multiple test procedure. Journal of the American Statistical Association, 87, 162-170.

Einot, I., & Gabriel, K. R. (1975). A study of the powers of several methods of multiple comparisons. Journal of the American Statistical Association, 70, 574-583.

Ekstrom, D., Quade, D., & Golden, R. N. (1990). Statistical analysis of repeated measures in psychiatric research. Archives of General Psychiatry, 47, 770-772.

Fisher, R. A. (1935). The design of experiments. Edinburgh: Oliver and Boyd.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32, 675-701.

Gabriel, K. R., Putter, J., & Wax, Y. (1973). Simultaneous confidence intervals for product-type interaction contrasts. Journal of the Royal Statistical Society, Series B, 35, 234-244.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate designs. Annals of Mathematical Statistics, 29, 885-891.

Glass, G. V, & Hopkins, K. D. (1984). Statistical methods in education and psychology (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Green, K. B. (1984). Home economics research: A macro view. Home Economics Research Journal, 12, 514-522.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. Psychometrika, 24, 95-112.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. Psychological Bulletin, 82, 1-2.

Harman, H. H. (1976). Modern factor analysis (3rd ed.). Chicago: Unversity of Chicago Press.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75, 800-802.

Hochberg, Y., & Tamhane, A. C. (1987). Multiple comparison procedures. New York: Wiley.

Holm. S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65-70.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika, 75, 383-386.

Hotelling, H. (1931). The generalization of Student's ratio. Annals of Mathematical Statistics, 2, 360-378.

Hotelling, H. (1951). A generalized t test and measure of multivariate dispersion. In J. Neyman (Ed.), Proceedings of the second Berkeley symposium on mathematical statistics and probability, Vol. 2 (pp. 23-41). Berkeley, CA: University of California Press.

Hutton, S. S. (1984). State of the art: Clothing as a form of human behavior. Home Economics Research Journal, 12, 340-353.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. Psychometrika, 32, 161-175.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurements designs have exact F distributions. Journal of the American Statistical Association, 65, 1582-1589.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. Journal of Educational Statistics, 1, 69-82.

Huynh, H., & Feldt, L. S. (1980). Performance of traditional F tests in repeated measures designs under covariance heterogeneity. Communications in Statistics: Theory and Methods, A9, 61-74.

International Mathematical and Statistical Library. (1987). IMSL library (9th ed.). Houston, TX: Author.

Jaccard, J., Becker, M. A., & Wood, G. (1984). Pairwise multiple comparison procedures. Psychological Bulletin, 96, 589-596.

James, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. Biometrika, 38, 19-43.

James, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. Biometrika, 41, 19-43.

Johansen, S. (1980). The Welch-James approximation to the distribution of the residual sum of squares in a weighted linear regression. Biometrika, 67, 85-92.

Johnson, D. E. (1976). Some new multiple comparison procedures for the two-way AOV model with interaction. Biometrics, 32, 929-934.

Kaiser, S. B., & Damhorst, M. L. (1991). Conceptualizing critical linkages in textiles and clothing subject matter. In S. B. Kaiser & M. L. Damhorst (Eds.) ITAA special publication #4: Critical linkages in textiles and clothing subject matter: Theory, method, and practice (pp. 1-10). Monument, CO: International Textile and Apparel Association.

Kaufman, N. J., Dudley-Marling, C., & Serlin, S. L. (1986). An examination of statistical interactions in the special education literature. Journal of Special Education, 20, 31-42.

Keselman, H. J. (1982). Multiple comparisons for repeated measures means. Multivariate Behavioral Research, 17, 87-92.

Keselman, H. J. (1993). Stepwise multiple comparisons of repeated measures means under violations of multisample sphericity. In F. M. Hoppe (Ed.), Multiple comparisons, selection, and aplications in biometry (pp. 167-186). New York: Marcel Dekker.

Keselman, H. J. (1994). Stepwise and simultaneous multiple comparison procedures of repeated measures means. Journal of Educational Statistics, 19, 127-162.

Keselman, H. J., Carriere, K. C., & Lix, L. M. (1993). Testing repeated measures hypotheses when covariance matrices are heterogeneous. Journal of Educational Statistics, 18, 305-309.

Keselman, H. J., & Keselman, J. C. (1988). Repeated measures multiple comparison procedures: Effect of violating multisample sphericity in unbalanced designs. Journal of Educational Statistics, 13, 215-226.

Keselman, H. J., & Keselman, J. C. (1993). Analysis of repeated measurements (pp. 105-145). In L. K. Edwards (Ed.), Applied analysis of variance in behavioral science (pp. 105-145). New York: Marcel Dekker.

Keselman, H. J., Keselman, J. C., & Lix, L. M. (in press). The analysis of repeated measurements: Univariate tests, mutivariate tests, or both? British Journal of Mathematical and Statistical Psychology.

Keselman, J. C., & Keselman, H. J. (1990). Analysing unbalanced repeated measures designs. British Journal of Mathematical and Statistical Psychology, 43, 265-282.

Keselman, J. C., Lix, L. M., & Keselman, H. J. (1994). The analysis of repeated measurements: A quantitative research synthesis. Manuscript submitted for publication.

Keuls, M. (1952). The use of the 'Studentized range' in conjunction with an analysis of variance. Euphytica, 1, 112-122.

Kirk, R. E. (1995). Experimental design: Procedures for the behavioral sciences (3rd ed.). Pacific Grove, CA: Brooks/Cole.

LaTour, S. A., & Miniard, P. W. (1983). The misuse of repeated measures analysis in marketing research. Journal of Marketing Research, 20, 45-57.

Lavori, P. (1990). ANOVA, MANOVA, My black hen: Comments on repeated measures. Archives of General Psychiatry, 47, 775-778.

Lawley, D. N. (1938). A generalization of Fisher's z test. Biometrika, 30, 180-187, 467-469.

Lawson, L., & Lorentzen, D. (1990). Selected sports bras: Comparisons of comfort and support. Clothing and Textiles Research Journal, 8(4), 55-60.

Lecoutre, B. (1991). A correction for the e-tilde approximate test in repeated measures designs with two or more independent groups. Journal of Educational Statistics, 16, 371-372.

Lindsey, J. K. (1993). Models for repeated measurements. New York: Oxford University Press.

Lix, L. M. (1995). Pairwise comparisons of means in clothing and textiles research. Clothing and Textiles Research Journal, 13(2), 102-110.

Lix, L. M., & Keselman, H. J. (in press). Approximate degrees of freedom tests: A unified perspective on testing for mean equality. Psychological Bulletin.

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1995). Analysis of single-group repeated measures designs: A quantitative review. Manuscript submitted for publication.

Looney, S. W., & Stanley, W. B. (1989). Exploratory repeated measures analysis for two or more groups: Review and update. American Statistician, 43(4), 220-225.

Lovie, A. D. (1981). On the early history of ANOVA in the analysis of repeated measure designs in psychology. British Journal of Mathematical and Statistical Psychology, 34, 1-5.

Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. American Educational Research Journal, 7, 397-421.

Marascuilo, L. A., & Levin, J. R. (1976). The simultaneous investigation of interaction and nested hypotheses in two-factor analysis of variance designs. American Educational Research Journal, 13, 61-65.

Marascuilo, L. A., & McSweeney, M. (1977). Nonparametric and distribution-free methods for the social sciences. Monterey, CA: Brooks/Cole.

Maxwell, S. E. (1980). Pairwise multiple comparisons in repeated measures designs. Journal of Educational Statistics, 5, 269-287.

Maxwell, S. E., & Arvey, R. D. (1982). Small sample profile analysis with many variables. Psychological Bulletin, 92, 778-785.

Maxwell, S. E., & Delaney, H. D. (1990). Designing experiments and analyzing data: A model comparison perspective. Belmont, CA: Wadsworth.

Mendoza, J. L., Toothaker, L. E., & Crain, B. R. (1976). Necessary and sufficient conditions for F ratios in the LxJxK factorial design with two repeated factors. Journal of the American Statistical Association, 71, 992-993.

Micceri, T. (1989). The unicorn, the normal distribution, and other improbable creatures. Psychological Bulletin, 105, 156-166.

Moran, J. D., III. (1986). Let the t-test rest in peace: A note on the control of error rates. Home Economics Research Journal, 14, 390-383.

Nagasawa, R. H., Kaiser, S. B., & Hutton, S. S. (1989). Theoretical development in clothing and textiles: Are we stuck in concrete? Clothing and Textiles Research Journal, 7(2), 23-31.

Newman, D. (1939). The distribution of the range in samples from a normal population expressed in terms of an independent estimate of standard deviation. Biometrika, 31, 20-30.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. Psychological Bulletin, 97, 316-333.

Olejnik, S., & Huberty, C. J. (1993, April). Preliminary statistical tests. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Oliver, B. A., & Mahoney, M. Y. (1991). The Clothing and Textiles Research Journal: An empirical examination. Clothing and Textiles Research Journal, 9(3), 22-27.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. Annals of Mathematical Statistics, 26, 117-121.

Personal Bibliographic Software, Inc. (1992). Pro-Cite 2.0 for the IBM (IBM version 2.0) [Computer software]. Ann Arbor, MI: Author.

Quintana, S. M., & Maxwell, S. E. (1985, April). A better-than-average estimate of e. Paper presented at the annual meeting of the American Educational Research Association.

Quintana, S. M., & Maxwell, S. E. (1994). A monte carlo comparison of seven e-adjusted procedures in repeated measures designs with small samples. Journal of Educational Statistics, 19, 57-72.

Ramsey, P. H. (1978). Power differences between pairwise multiple comparisons. Journal of the American Statistical Association, 73, 479-485.

Robey, R. R., & Barcikowski, R. S. (1986, April). Type I error in the single group repeated measures design with multiple measures per occasion. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Rogan, J. C., Keselman, H. J., & Mendoza, J. L. (1979). Analysis of repeated measurements. British Journal of Mathematical and Statistical Psychology, 32, 269-286.

Rom, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. Biometrika, 77, 663-665.

Rosnow, R. L., & Rosenthal, R. (1989a). Definition and interpretation of interaction effects. Psychological Bulletin, 105, 143-146.

Rosnow, R. L., & Rosenthal, R. (1989b). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284.

Rosnow, R. L., & Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at only the interaction (unless all main effects are zero). Psychological Bulletin, 110, 574-576.

Rouanet, H., & Lepine, D. (1970). Comparison between treatments in a repeated-measurement design: ANOVA and multivariate methods. British Journal of Mathematical and Statistical Psychology, 23, 147-163.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. Annals of Mathematical Statistics, 24, 220-238.

Sanik, M. M. (1983). Repeated measure design: A time use application. Home Economics Research Journal, 12, 122-126.

SAS Institute Inc. (1989a). SAS/IML software: Usage and reference, Version 6 (1st ed.). Cary, NC: Author.

SAS Institute Inc. (1989b). SAS/STAT user's guide, Version 6 (4th ed., Vol. 2). Cary, NC: Author.

Satterthwaite, F. E. (1941). Synthesis of variance. Psychometrika, 6, 309-316.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. Biometrics, 2, 110-114.

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and type II error properties of the t test to departures from population normality. Psychological Bulletin, 111, 352-360.

Scheffe, H. (1953). A method for judging all contrasts in analysis of variance. Biometrika, 40, 87-104.

Schlick, P. J. (1991). Classification systems: Alternative research strategies for historic costume. In S. B. Kaiser & M. L. Damhorst (Eds.) ITAA special publication #4: Critical linkages in textiles and clothing subject matter: Theory, method, and practice (pp. 150-159). Monument, CO: International Textile and Apparel Association.

Schumm, W. R. (1982). Integrating theory, measurement and data analysis in family studies survey research. Journal of Marriage and the Family, 44, 983-998.

Schumm, W. R., Bugaighis, M. A., & Jurich, A. P. (1985). Using repeated measures designs in program evaluation of family therapy. Journal of Marital and Family Therapy, 11, 87-95.

Schumm, W. R., Barnes, H. L., Bollman, S. R., Jurich, A. P., Milliken, G. A. (1985). Approach to the statistical analysis of family data. Home Economics Research Journal, 14, 112-122.

Seaman, M. A., Levin, J. R., & Serlin, R. C. (1991). New developments in pairwise multiple comparisons: Some powerful and practicable procedures. Psychological Bulletin, 110, 577-586.

Searle, S. (1987). Linear models for unbalanced data. New York: Wiley.

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. Journal of the American Statistical Association, 81, 826-831.

Snedecor, G. W. (1937). Statistical methods (1st ed.). Ames, IA: Collegiate Press Inc.

Tang, K. L., & Algina, J. (1993). Performance of four multivariate tests under variance-covariance heteroscedasticity. Multivariate Behavioral Research, 28, 391-405.

Timm, N. H. (1994, April). Analysis of interactions: Another look. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Tukey, J. W. (1953). The problem of multiple comparisons. Unpublished manuscript, Princeton University.

Turnbull, S. G., & Lix, L. M. (1991). Looking back: A content analysis of clothing and textiles research. Unpublished manuscript, University of Manitoba, Winnipeg, MB.

Vercruyssen, M., & Hendrick, H. W. (1989). Behavioral research and analysis: An introduction to statistics within the context of experimental designs (2nd ed.). Lawrence, KS: Ergosyst Associates.

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. Biometrika, 34, 28-35.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. Biometrika, 38, 330-336.

Wilcox, R. R. (1993). Analysing repeated measures or randomized block designs using trimmed means. British Journal of Mathematical and Statistical Psychology, 46, 63-76.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. Biometrika, 24, 471-494.

Winer, B. J. (1962). Statistical principles in experimental design. New York: McGraw-Hill.

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. Biometrika, 61, 165-170.

Yuen, K. K., & Dixon, W. J. (1973). The approximate behavior and performance of the two-sample trimmed t. Biometrika, 60, 369-374.

Appendix A

Statistical References Cited in Clothing and Textiles Repeated Measures Articles

STATISTICAL REFERENCES

Anderson, V. L., & McLean, R. A. (1974). Design of experiments: A realistic approach. New York: Marcel Dekker.

Bradley, J. V. (1968). Distribution-free statistical tests. Englewood Cliffs, NJ: Prentice-Hall.

Cicchetti, D. V. (1972). Extension of multiple-range tests to interaction tables in the analysis of variance. Psychological Bulletin, 77, 405-408.

Conover, W. J. (1980). Practical nonparametric statistics (2nd ed.). New York: Wiley.

Fontenelle, G., Phillips, A., & Lane, D. (1985). Generalizing across stimuli as well as subjects: A neglected aspect of external validity. Journal of Applied Psychology, 101, 151-152.

Games, P. A. (1981). The ANOVR series: Analyses for repeated measure designs with optional robust procedures when assumptions are violated. American Statistician, 35(2), 109-110.

Games, P. A., Gray, G. S., & Herron, W. L. (1980). ANOVR: Analysis of variance with repeated measures. Behavior Research Methods & Instrumentation, 12, 467.

Guilford, J. P., & Fruchter, B. (1978). Fundamental statistics in psychology and education (6th ed.). New York: McGraw-Hill.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measure designs have exact F-distributions. Journal of the American Statistical Association, 65, 1582-1589.

Kendall, M. G., & Stuart, A. (1977). The advanced theory of statistics (4th ed.). New York: Macmillan.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

Keppel, G. (1982). Design and analysis: A researcher's handbook (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kirk, R. E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole.

Marascuilo, L. A., & McSweeney, M. (1977). Nonparametric and distribution-free methods for the social sciences. Monterey, CA: Brooks/Cole.

Miller, I., & Freund, J. E. (1977). Probability and statistics for engineers (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Ostle, B., & Mensing, R. W. (1975). Statistics in research: Basic concepts and techniques for research workers (3rd ed.). Ames, IA: Iowa State University Press.

Scheffe, H. (1959). The analysis of variance. New York: Wiley.

Seigel, S. (1956). Non-parametric statistics for the behavioral sciences. New York: McGraw-Hill.

Spiegel, M. R. (1961). Schaum's outline of theory and problems of statistics. New York: McGraw-Hill.

Norusis, M. J. (1988). SPSS/PC+ advanced statistics, 2.0. Chicago: SPSS.

Steel, R. G. D., & Torrie, J. H. (1980). Principles and procedures of statistics: A biometrical approach. New York: McGraw- Hill.

Vercruyssen, M., & Hendrick, H. W. (1989). Behavioral research and analysis: An introduction to statistics within the context of experimental designs (2nd ed.). Lawrence, KS: Ergosyst Associates.

Winer, B. J. (1962). Statistical principles in experimental design. New York: McGraw-Hill.

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Zar, J. H. (1984). Biostatistical analysis. Englewood Cliffs, NJ: Prentice-Hall.

Appendix B

Simulation Results for the Complete Null Hypothesis When N = 45

Table B1

Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Complete Null Hypothesis; K = 4; N = 45)

		Normal				χ^2			
		S	R	H/MSB	M	S	R	H/MSB	M
ϵ									
	a	.005*	.015*	.034	.035	.008*	.021*	.045	.046
	b	.014*	.029	.056	.058	.018*	.032	.054	.057
	c	.007*	.015*	.035	.036	.006*	.017*	.034	.034
1.00	c′	.005*	.008*	.018*	.019*	.004*	.009*	.024*	.024*
	d	.029	.052	.097†	.099†	.027	.053	.094	.098†
	d′	.050	.086†	.144†	.147†	.047	.079†	.133†	.136†
	a	.021*	.040	.073	.075	.019*	.037	.069	.072
	b	.031	.051	.082†	.084†	.030	.052	.080†	.081†
	c	.019*	.032	.056	.057	.016*	.028	.051	.052
.75	c′	.014*	.024*	.043	.044	.015*	.026	.041	.042
	d	.053	.085†	.122†	.125†	.047	.075	.118†	.121†
	d′	.088†	.120†	.170†	.172†	.071	.108†	.163†	.167†
	a	.064	.090†	.139†	.142†	.063	.096†	.139†	.142†
	b	.074	.010*	.141†	.143†	.072	.101†	.146†	.150†
	c	.053	.076†	.110†	.113†	.055	.081†	.119†	.122†
.40	c′	.037	.055	.082†	.084†	.040	.056	.085†	.087†
	d	.099†	.136†	.185†	.187†	.103†	.147†	.193†	.196†
	d′	.132†	.168†	.217†	.220†	.139†	.184†	.238†	.241†

NOTE: S = Scheffe (1953); R = Studentized maximum root; H/MSB = Hochberg (1988) step-up Bonferroni/Shaffer (1986) modified sequentially rejective Bonferroni; M = Studentized maximum modulus; a = pairings of equal covariance matrices and equal group sizes; b = pairings of unequal covariance matrices and equal group sizes; c/c′ = positive pairings of covariance matrices and group sizes [c: n_j = 12, 15, 18; c′: n_j = 9, 15, 21]; d/d′ = negative pairings of covariance matrices and group sizes [d: n_j = 18, 15, 12; d′: n_j = 21, 15, 9]; * = empirical value < .025; † = empirical value > .075.

Table B2

Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Complete Null Hypothesis; K = 8; N = 45)

		Normal				χ^2			
		S	R	H/MSB	M	S	R	H/MSB	M
ϵ									
	a	.000*	.002*	.035	.037	.000*	.003*	.047	.048
	b	.001*	.007*	.065	.066	.000*	.008*	.071	.072
	c	.000*	.002*	.034	.035	.000*	.003*	.040	.041
1.00	c′	.000*	.001*	.023*	.024*	.000*	.002*	.022*	.022*
	d	.002*	.015*	.120†	.122†	.003*	.019*	.118†	.120†
	d′	.008*	.040	.198†	.200†	.010*	.041	.198†	.202†
	a	.002*	.009*	.074	.076†	.002*	.008*	.065	.067
	b	.003*	.020*	.096†	.098†	.003*	.017*	.094†	.095†
	c	.002*	.013*	.060	.061	.003*	.012*	.063	.064
.75	c′	.001*	.005*	.033	.034	.001*	.007*	.042	.042
	d	.010*	.035	.148†	.150†	.009*	.032	.148†	.151†
	d′	.021*	.065	.228†	.231†	.020*	.061	.223†	.225†
	a	.016*	.047	.156†	.158†	.010*	.043	.167†	.169†
	b	.023*	.063	.184†	.185†	.023*	.064	.189†	.191†
	c	.016*	.043	.130†	.131†	.014*	.045	.131†	.133†
.40	c′	.008*	.028	.092†	.094†	.008*	.025	.093†	.094†
	d	.042	.093†	.242†	.244†	.034	.090†	.233†	.236†
	d′	.070	.147†	.317†	.319†	.066	.141†	.312†	.314†

Note: See the note from Table B1.

Table B3

Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis; K = 4; N = 45)

		Normal				χ^2			
		S	R	H/MSB	M	S	R	H/MSB	M
ϵ									
	a	.007*	.018*	.037	.041	.009*	.017*	.035	.037
	b	.012*	.023*	.040	.044	.009*	.019*	.035	.038
	c	.012*	.023*	.040	.044	.008*	.018*	.036	.039
1.00	c′	.009*	.022*	.036	.040	.008*	.017*	.030	.034
	d	.013*	.025	.044	.049	.013*	.025	.041	.046
	d′	.015*	.028	.043	.048	.016*	.029	.042	.047
	a	.009*	.018*	.036	.039	.005*	.013*	.027	.029
	b	.011*	.024*	.039	.043	.009*	.020*	.035	.038
	c	.007*	.019*	.033	.035	.006*	.016*	.031	.034
.75	c′	.011*	.020*	.037	.042	.009*	.018*	.033	.036
	d	.012*	.024*	.041	.046	.010*	.020*	.036	.040
	d′	.021*	.036	.050	.057	.013*	.027	.039	.044
	a	.005*	.011*	.018*	.019*	.004*	.009*	.019*	.020*
	b	.007*	.014*	.024*	.026	.007*	.012*	.020*	.021*
	c	.006*	.013*	.021*	.022*	.007*	.013*	.024*	.025
.40	c′	.008*	.015*	.025	.026	.006*	.010*	.018*	.020*
	d	.005*	.011*	.020*	.023*	.011*	.018*	.030	.032
	d′	.010*	.016*	.023*	.027	.014*	.026	.035	.038

Note: See the note from Table B1.

Table B4

Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Complete Null Hypothesis; K = 8; N = 45)

		Normal				χ^2			
		S	R	H/MSB	M	S	R	H/MSB	M
ϵ									
	a	.001*	.004*	.034	.040	.000*	.002*	.024*	.029
	b	.000*	.005*	.043	.049	.000*	.004*	.032	.038
	c	.001*	.006*	.039	.047	.001*	.004*	.029	.036
1.00	c′	.000*	.006*	.044	.052	.000*	.003*	.026	.033
	d	.003*	.011*	.043	.057	.000*	.004*	.034	.042
	d′	.004*	.015*	.058	.073	.001*	.010*	.045	.057
	a	.001*	.006*	.038	.044	.000*	.003*	.022*	.027
	b	.002*	.008*	.045	.051	.000*	.005*	.033	.039
	c	.001*	.006*	.040	.047	.000*	.004*	.028	.035
.75	c′	.000*	.006*	.038	.043	.001*	.002*	.028	.033
	d	.002*	.010*	.044	.054	.000*	.004*	.036	.043
	d′	.005*	.015*	.049	.064	.003*	.012*	.041	.052
	a	.000*	.005*	.029	.033	.000*	.002*	.020*	.023*
	b	.001*	.005*	.029	.034	.001*	.005*	.032	.037
	c	.001*	.003*	.024*	.029	.000*	.005*	.026	.031
.40	c′	.001*	.006*	.031	.035	.000*	.002*	.023*	.026
	d	.001*	.007*	.034	.041	.002*	.008*	.035	.045
	d′	.003*	.013*	.041	.050	.006*	.015*	.043	.052

Note: See the note from Table B1.

Appendix C

Simulation Results for a Partial Null Hypothesis

Table C1

Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Partial Null Hypothesis; K = 4; N = 30)

		Normal					χ^2				
		S	R	H	MSB	M	S	R	H	MSB	M
ϵ											
	a	.003*	.009*	.026	.031	.023*	.003*	.010*	.029	.036	.026
	b	.007*	.013*	.029	.033	.025	.005*	.012*	.030	.035	.024*
	c	.003*	.006*	.017*	.018*	.014*	.004*	.007*	.018*	.020*	.015*
1.00	c′	.001*	.004*	.010*	.012*	.010*	.001*	.003*	.008*	.010*	.007*
	d	.011*	.022*	.053	.059	.046	.013*	.025*	.050	.057	.044
	d′	.024*	.041	.082†	.091†	.073	.026	.048	.092†	.101†	.082†
	a	.011*	.017*	.038	.042	.033	.006*	.015*	.036	.040	.031
	b	.006*	.012*	.029	.032	.025*	.006*	.010*	.028	.031	.023*
	c	.002*	.007*	.018*	.021*	.015*	.003*	.006*	.016*	.019*	.014*
.75	c′	.001*	.003*	.009*	.011*	.008*	.001*	.004*	.008*	.009*	.007*
	d	.010*	.021*	.049	.052	.043	.011*	.024*	.052	.058	.045
	d′	.023*	.044	.081†	.088†	.073	.021*	.035	.076†	.083†	.067
	a	.026	.013*	.071	.076†	.064	.027	.041	.072	.077†	.065
	b	.012*	.020*	.041	.043	.035	.010*	.020*	.043	.044	.038
	c	.006*	.013*	.030	.032	.026	.005*	.009*	.022*	.024*	.019*
.40	c′	.003*	.007*	.015*	.017*	.014*	.006*	.008*	.016*	.018*	.014*
	d	.019*	.032	.059	.061	.053	.021*	.038	.063	.066	.056
	d′	.034	.052	.086†	.091†	.079†	.034	.051	.087†	.092†	.081†

NOTE: S = Scheffe (1953); R = Studentized maximum root; H = Hochberg (1988) step-up Bonferroni; MSB = Shaffer (1986) modified sequentially rejective Bonferroni; M = Studentized maximum modulus; a = pairings of equal covariance matrices and equal group sizes; b = pairings of unequal covariance matrices and equal group sizes; c/c′ = positive pairings of covariance matrices and group sizes [c: n_j = 8, 10, 12; c′: n_j = 6, 10, 14]; d/d′ = negative pairings of covariance matrices and group sizes [d: n_j = 12, 10, 8; d′: n_j = 14, 10, 6]; * = empirical value < .025; † = empirical value > .075.

Table C2

Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Partial Null Hypothesis; K = 4; N = 45)

		Normal					χ^2				
		S	R	H	MSB	M	S	R	H	MSB	M
ϵ											
	a	.005*	.011*	.032	.034	.026	.004*	.010*	.029	.034	.023*
	b	.007*	.013*	.031	.037	.024*	.004*	.010*	.031	.035	.024*
	c	.003*	.005*	.017*	.020*	.012*	.003*	.007*	.021*	.023*	.017*
1.00	c′	.002*	.005*	.011*	.012*	.009*	.002*	.005*	.014*	.015*	.011*
	d	.011*	.020*	.051	.056	.041	.011*	.020*	.049	.052	.039
	d′	.020*	.041	.085†	.091†	.070	.023*	.040	.085†	.091†	.072
	a	.012*	.020*	.047	.052	.039	.007*	.018*	.045	.049	.034
	b	.005*	.012*	.034	.035	.027	.004*	.010*	.032	.036	.025
	c	.001*	.006*	.018*	.019*	.014*	.002*	.004*	.016*	.017*	.013*
.75	c′	.000*	.003*	.010*	.012*	.007*	.001*	.003*	.011*	.013*	.008*
	d	.010*	.023*	.054	.059	.044	.009*	.018*	.048	.052	.041
	d′	.017*	.034	.080†	.086†	.068	.019*	.036	.072	.077†	.060
	a	.028	.039	.072	.075	.062	.020*	.032	.063	.065	.055
	b	.012*	.020*	.043	.046	.033	.011*	.020*	.043	.046	.037
	c	.007*	.009*	.026	.028	.022*	.006*	.010*	.025	.027	.021*
.40	c′	.003*	.006*	.012*	.013*	.009*	.003*	.007*	.015*	.016*	.013*
	d	.015*	.030	.061	.064	.052	.016*	.026	.058	.062	.048
	d′	.030	.046	.088†	.090†	.075	.033	.051	.089†	.077†	.060

Note: See the note from Table C1; For c condition, n_j = 12, 15, 18; For c′ condition, n_j = 9, 15, 21; For d condition, n_j = 18, 15, 12; For d′ condition, n_j = 21, 15, 9.

Table C3

Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Partial Null Hypothesis; K = 8; N = 30)

		Normal					χ^2				
		S	R	H	MSB	M	S	R	H	MSB	M
ϵ											
	a	.000*	.002*	.026	.027	.024*	.000*	.002*	.027	.028	.025
	b	.001*	.004*	.036	.038	.034	.001*	.005*	.040	.042	.037
	c	.000*	.002*	.020*	.021*	.019*	.000*	.002*	.022*	.024*	.020*
1.00	c′	.000*	.000*	.011*	.012*	.011*	.000*	.001*	.012*	.014*	.012*
	d	.001*	.009*	.075	.077†	.070	.002*	.008*	.073	.077†	.069
	d′	.004*	.022*	.140†	.144†	.131†	.005*	.025	.133†	.138†	.125†
	a	.001*	.006*	.049	.050	.046	.001*	.006*	.049	.050	.047
	b	.001*	.007*	.046	.046	.042	.001*	.007*	.049	.050	.045
	c	.000*	.001*	.026	.027	.023*	.000*	.003*	.028	.030	.026
.75	c′	.000*	.002*	.014*	.015*	.013*	.000*	.002*	.018*	.020*	.017*
	d	.003*	.014*	.081†	.084†	.078†	.004*	.017*	.090†	.094†	.086†
	d′	.009*	.030	.134†	.139†	.128†	.009*	.030	.138†	.142†	.134†
	a	.008*	.030	.113†	.117†	.110†	.006*	.024*	.111†	.114†	.108†
	b	.008*	.027	.103†	.105†	.099†	.006*	.022*	.103†	.106†	.100†
	c	.006*	.017*	.075	.078†	.072	.003*	.014*	.072	.074	.070
.40	c′	.001*	.010*	.045	.046	.043	.003*	.008*	.049	.051	.048
	d	.016*	.048	.157†	.162†	.151†	.014*	.043	.152†	.155†	.147†
	d′	.030	.074	.215†	.220†	.210†	.024*	.073	.212†	.216†	.206†

Note: See the note from Table C1.

Table C4

Empirical FWRs for Tetrad Contrast Procedures Employing a Pooled Test Statistic (Partial Null Hypothesis; K = 8; N = 45)

		Normal					χ^2				
		S	R	H	MSB	M	S	R	H	MSB	M
ϵ											
	a	.000*	.001*	.030	.030	.027	.000*	.001*	.033	.034	.029
	b	.000*	.003*	.044	.044	.038	.001*	.003*	.038	.039	.033
	c	.000*	.001*	.026	.026	.023*	.000*	.001*	.025	.025	.022*
1.00	c′	.000*	.008*	.010*	.016*	.009*	.000*	.000*	.013*	.014*	.011*
	d	.001*	.008*	.079†	.080†	.071	.001*	.009*	.082†	.083†	.071
	d′	.004*	.017*	.137†	.138†	.123†	.005*	.019*	.136†	.138†	.126†
	a	.001*	.005*	.049	.050	.043	.002*	.006*	.052	.053	.048
	b	.001*	.006*	.049	.050	.043	.001*	.006*	.047	.048	.043
	c	.001*	.003*	.029	.030	.026	.001*	.005*	.032	.033	.029
.75	c′	.000*	.002*	.018*	.018*	.016*	.000*	.002*	.019*	.019*	.017*
	d	.003*	.011*	.074	.075	.066	.003*	.014*	.086†	.089†	.078†
	d′	.006*	.023*	.135†	.136†	.124†	.007*	.030	.140†	.141†	.129†
	a	.008*	.028	.122†	.123†	.116†	.007*	.027	.129†	.131†	.117†
	b	.009*	.032	.109†	.110†	.102†	.006*	.026	.105†	.106†	.097†
	c	.003*	.013*	.072	.073	.066	.003*	.011*	.069	.070	.065
.40	c′	.002*	.009*	.044	.045	.040	.002*	.009*	.053	.053	.048
	d	.015*	.047	.164†	.166†	.149†	.012*	.039	.150†	.152†	.138†
	d′	.029	.076†	.226†	.229†	.211†	.025	.066	.200†	.202†	.192†

Note: See the notes from Tables C1 and C2.

Table C5

Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Partial Null Hypothesis; K = 4; N = 30)

		Normal					χ^2				
		S	R	H	MSB	M	S	R	H	MSB	M
ϵ											
	a	.008*	.013*	.025*	.030	.023*	.003*	.010*	.020*	.024*	.019*
	b	.012*	.018*	.028	.033	.028	.006*	.015*	.026	.029	.026
	c	.010*	.018*	.029	.034	.029	.005*	.009*	.019*	.023*	.018*
1.00	c′	.008*	.015*	.026	.031	.025	.006*	.012*	.021*	.024*	.021*
	d	.010*	.020*	.029	.034	.031	.008*	.016*	.026	.030	.028
	d′	.013*	.022*	.029	.032	.034	.012*	.022*	.032	.035	.034
	a	.007*	.014*	.025	.031	.023*	.004*	.010*	.017*	.022*	.017*
	b	.009*	.017*	.028	.034	.028	.006*	.013*	.023*	.027	.023*
	c	.008*	.013*	.028	.033	.026	.004*	.011*	.021*	.026	.022*
.75	c′	.008*	.016*	.029	.032	.027	.005*	.010*	.019*	.023*	.019*
	d	.010*	.021*	.031	.038	.034	.006*	.014*	.024*	.031	.025
	d′	.016*	.025	.033	.038	.036	.007*	.015*	.023*	.029	.027
	a	.007*	.013*	.022*	.026	.019*	.003*	.006*	.013*	.015*	.011*
	b	.007*	.011*	.024*	.028	.021*	.005*	.010*	.018*	.022*	.019*
	c	.006*	.011*	.025*	.028	.021*	.004*	.009*	.016*	.018*	.015*
.40	c′	.004*	.010*	.021*	.025*	.018*	.004*	.009*	.017*	.018*	.014*
	d	.008*	.014*	.024*	.028	.024*	.008*	.015*	.026	.029	.025
	d′	.010*	.015*	.025*	.028	.025*	.012*	.020*	.027	.029	.027

Note: See the note from Table C1.

Table C6

Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Partial Null Hypothesis; K = 4; N = 45)

		Normal					χ^2				
		S	R	H	MSB	M	S	R	H	MSB	M
ϵ											
	a	.006*	.012*	.030	.033	.024*	.003*	.007*	.022*	.024*	.018*
	b	.008*	.016*	.032	.037	.029	.005*	.013*	.028	.033	.025*
	c	.006*	.013*	.031	.035	.025	.005*	.012*	.028	.031	.023*
1.00	c′	.005*	.011*	.032	.037	.026	.004*	.009*	.024*	.029	.020*
	d	.007*	.015*	.032	.037	.030	.006*	.012*	.026	.031	.024*
	d′	.009*	.015*	.028	.034	.028	.008*	.015*	.028	.032	.028
	a	.005*	.011*	.030	.036	.026	.005*	.011*	.025*	.027	.020*
	b	.007*	.013*	.031	.035	.025	.006*	.014*	.028	.031	.024*
	c	.006*	.013*	.033	.036	.026	.005*	.008*	.022*	.027	.017*
.75	c′	.005*	.011*	.031	.034	.023*	.003*	.008*	.024*	.026	.018*
	d	.008*	.020*	.035	.039	.032	.007*	.014*	.025	.030	.023*
	d′	.006*	.012*	.028	.031	.026	.004*	.012*	.023*	.028	.021*
	a	.005*	.011*	.026	.028	.021*	.002*	.005*	.016*	.017*	.013*
	b	.006*	.012*	.026	.029	.022*	.005*	.011*	.022*	.024*	.019*
.40	c	.003*	.008*	.022*	.025*	.018*	.005*	.010*	.022*	.025*	.019*
	c′	.005*	.009*	.025	.027	.021*	.003*	.007*	.021*	.023*	.016*
	d	.004*	.008*	.023*	.027	.018*	.005*	.011*	.020*	.023*	.018*
	d′	.006*	.012*	.025	.028	.020*	.008*	.016*	.027	.032	.026

Note: See the notes from Tables C1 and C2.

Table C7

Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Partial Null Hypothesis; K = 8; N = 30)

		Normal					χ^2				
		S	R	H	MSB	M	S	R	H	MSB	M
ϵ											
	a	.001*	.006*	.026	.029	.032	.001*	.002*	.015*	.017*	.019*
	b	.002*	.007*	.028	.032	.040	.001*	.005*	.018*	.020*	.025
	c	.001*	.005*	.026	.030	.035	.001*	.004*	.014*	.016*	.021*
1.00	c′	.002*	.007*	.028	.029	.036	.001*	.004*	.016*	.018*	.020*
	d	.002*	.009*	.036	.040	.049	.001*	.006*	.028	.032	.037
	d′	.006*	.017*	.045	.049	.064	.003*	.008*	.028	.030	.039
	a	.002*	.006*	.025	.027	.031	.000*	.003*	.017*	.019*	.024*
	b	.001*	.006*	.029	.032	.039	.001*	.004*	.020*	.022*	.026
	c	.002*	.006*	.027	.029	.037	.001*	.003*	.021*	.023*	.027
.75	c′	.001*	.005*	.025	.026	.034	.001*	.003*	.017*	.020*	.025
	d	.002*	.010*	.035	.037	.047	.002*	.008*	.027	.029	.036
	d′	.009*	.021*	.048	.051	.063	.004*	.013*	.031	.033	.042
	a	.000*	.002*	.020*	.020*	.025	.001*	.004*	.015*	.016*	.018*
	b	.002*	.006*	.025	.027	.031	.001*	.004*	.019*	.020*	.025
	c	.001*	.005*	.021*	.023*	.027	.000*	.004*	.018*	.019*	.023*
.40	c′	.001*	.005*	.019*	.021*	.025	.001*	.003*	.013*	.014*	.019*
	d	.002*	.007*	.030	.032	.038	.001*	.005*	.024*	.025	.030
	d′	.006*	.014*	.034	.036	.047	.004*	.013*	.034	.037	.047

Note: See the note from Table C1.

Table C8

Empirical FWRs for Tetrad Contrast Procedures Employing a Nonpooled Test Statistic (Partial Null Hypothesis; K = 8; N = 45)

		Normal					χ^2				
		S	R	H	MSB	M	S	R	H	MSB	M
ϵ											
	a	.001*	.004*	.027	.029	.029	.000*	.001*	.016*	.017*	.017*
	b	.001*	.005*	.033	.035	.035	.000*	.002*	.020*	.021*	.022*
	c	.001*	.004*	.028	.030	.029	.001*	.003*	.019*	.019*	.019*
1.00	c′	.000*	.002*	.025	.027	.027	.000*	.002*	.018*	.019*	.020*
	d	.001*	.007*	.035	.037	.039	.000*	.006*	.030	.032	.032
	d′	.000*	.008*	.034	.037	.041	.001*	.007*	.032	.034	.038
	a	.000*	.003*	.030	.031	.032	.000*	.003*	.020*	.021*	.020*
	b	.000*	.004*	.030	.030	.031	.000*	.003*	.024*	.026	.026
	c	.000*	.002*	.029	.030	.031	.000*	.001*	.021*	.022*	.023*
.75	c′	.000*	.004*	.028	.029	.029	.000*	.001*	.020*	.021*	.021*
	d	.000*	.005*	.028	.031	.033	.001*	.004*	.027	.029	.030
	d′	.002*	.006*	.029	.032	.037	.000*	.005*	.024*	.028	.030
	a	.001*	.003*	.028	.028	.026	.000*	.001*	.017*	.017*	.016*
	b	.000*	.003*	.025	.026	.026	.001*	.003*	.025	.025	.024*
	c	.000*	.002*	.025	.025	.022*	.000*	.002*	.020*	.026	.019*
.40	c′	.001*	.003*	.022*	.023*	.021*	.000*	.002*	.019*	.019*	.018*
	d	.001*	.004*	.024*	.025	.025	.001*	.003*	.022*	.023*	.023*
	d′	.003*	.008*	.031	.032	.036	.002*	.008*	.028	.030	.032

Note: See the notes from Tables C1 and C2.

Appendix D

Example Data Set and Computer Programming Statements for Tetrad Contrasts Example

Hypothetical Data Set

	BR1	BR2	BR3
Group A (n_A = 9)			
	19	31	42
	26	40	47
	18	39	41
	21	42	50
	18	33	43
	19	35	45
	23	36	49
	24	35	47
	17	31	47
Group B (n_B = 13)			
	14	20	36
	11	23	31
	15	21	38
	10	19	42
	16	30	40
	21	26	37
	10	20	37
	12	23	32
	13	21	34
	13	26	35
	20	20	40
	17	21	37
	15	26	33
Group C (n_C = 8)			
	17	19	33
	16	21	37
	15	25	31
	13	20	35
	11	25	30
	14	22	37
	13	28	39
	12	20	29
Group D (n_D = 5)			
	21	46	30
	27	50	28
	30	45	26
	22	40	27
	26	42	27

Note: BR1 - BR3 = Support scores for bra conditions 1 through 3

SAS/IML Program from Lix and Keselman (in press)[1]

```
***INVOKE THE IML PROCEDURE & DEFINE THE MODULE WJGLM***;
PROC IML;
RESET NONAME;
START WJGLM;
**********PERFORM DIAGNOSTICS AND DEFINE MATRICES**********;
IF NROW(U)=0 THEN U=I(NCOL(Y));
IF NROW(C) > NCOL(C) THEN PRINT
  'ERROR: NUMBER OF ROWS OF C EXCEEDS NUMBER OF COLUMNS';
IF NCOL(U) > NROW(U) THEN PRINT
  'ERROR: NUMBER OF COLUMNS OF U EXCEEDS NUMBER OF ROWS';
DO I=1 TO NCOL(NX);
  X1=J(NX[I],1,I);
  IF I=1 THEN X=X1;
    ELSE X=X//X1;
  END;
X=DESIGN(X);
NTOT=NROW(Y);
WOBS=NCOL(Y);
BOBS=NCOL(X);
WOBS1=WOBS-1;
***********FORM SIGMA MATRIX AND VECTOR OF MEANS**********;
BHAT=INV(X`*X)*X`*Y;
MUHAT=SHAPE(BHAT,WOBS#BOBS);
SIGMA=J(WOBS#BOBS,WOBS#BOBS,0);
DF=NX-1;
DO I=1 TO BOBS;
  SIGB=(Y#X[,I]-X[,I]*BHAT[I,])`*(Y#X[,I]-X[,I]*BHAT[I,])/DF[I];
  F=I#WOBS-WOBS1;
  L=I#WOBS;
  SIGMA[F:L,F:L]=SIGB/NX[I];
END;
**********CALCULATE TEST STATISTIC, DF, AND P-VALUE**********;
R=C@U`;
T=(R*MUHAT)`*INV(R*SIGMA*R`)*(R*MUHAT);
A=0;
IMAT=I(WOBS);
```

[1]From "Approximate Degrees of Freedom Tests: A Unified Perspective on Testing for Mean Equality" by L. M. Lix and H. J. Keselman, in press, Psychological Bulletin.

```
DO I=1 TO BOBS;
  QMAT=J(BOBS#WOBS,BOBS#WOBS,0);
  F=I#WOBS-WOBS1;
  L=I#WOBS;
  QMAT[F:L,F:L]=IMAT;
  PROD=(SIGMA*R`)*INV(R*SIGMA*R`)*R*QMAT;
  A=A+(TRACE(PROD*PROD)+TRACE(PROD)**2)/DF[I];
END;
A=A/2;
DF1=NROW(R);
DF2=DF1#(DF1+2)/(3#A);
CVAL=DF1+2#A-6#A/(DF1+2);
RESULTS=J(4,1,0);
RESULTS[1]=T/CVAL;
RESULTS[2]=DF1;
RESULTS[3]=DF2;
RESULTS[4]=1 - PROBF(RESULTS[1],DF1,DF2);
*************************PRINT RESULTS**************************;
PRINT 'WELCH-JAMES APPROXIMATE DF SOLUTION';
PRINT 'CONTRAST MATRIX:';
PRINT R[FORMAT=4.1],;
MUHAT=MUHAT`;
PRINT 'MEAN VECTOR:';
PRINT MUHAT[FORMAT=10.4],;
PRINT 'SIGMA MATRIX:';
PRINT SIGMA[FORMAT=10.4],;
RESLAB={"TEST STATISTIC"  "NUMERATOR DF"  "DENOMINATOR DF"
"P-VALUE"};
PRINT 'SIGNIFICANCE TEST RESULTS:';
PRINT RESULTS[ROWNAME=RESLAB FORMAT=10.4]/;
************************END OF MODULE************************;
FINISH;
```

At this point, the SAS/IML code needed to run the program for a particular research design is input.

SAS/IML Programming Statements for Tetrad Contrasts

```
Y = {19 31 42, 26 40 47, 18 39 41, 21 42 50, 18 33 43, 19 35 45, 23 36 49,
24 35 47, 17 31 47, 14 20 36, 11 23 31, 15 21 38, 10 19 42, 16 30 40,
21 26 37, 10 20 37, 12 23 32, 13 21 34, 13 26 35, 20 20 40, 17 21 37,
15 26 33, 17 19 33, 16 21 37, 15 25 31, 13 20 35, 11 25 30, 14 22 37,
13 28 39, 12 20 29, 21 46 30, 27 50 28, 30 45 26, 22 40 27, 26 42 27};
NX = {9 13 8 5};
C = {1 -1 0 0};
U = {1, -1, 0};
PRINT 'A VS B ON BR1 & BR2';
RUN WJGLM;
U = {1, 0, -1};
PRINT 'A VS B ON BR1 & BR3';
RUN WJGLM;
U = {0, 1, -1};
PRINT 'A VS B ON BR2 & BR3';
RUN WJGLM;
C = {1 0 -1 0};
U = {1, -1, 0};
PRINT 'A VS C ON BR1 & BR2';
RUN WJGLM;
U = {1, 0, -1};
PRINT 'A VS C ON BR1 & BR3';
RUN WJGLM;
U = {0, 1, -1};
PRINT 'A VS C ON BR2 & BR3';
RUN WJGLM;
C = {1 0 0 -1};
U = {1, -1, 0};
PRINT 'A VS D ON BR1 & BR2';
RUN WJGLM;
U = {1, 0, -1};
PRINT 'A VS D ON BR1 & BR3';
RUN WJGLM;
U = {0, 1, -1};
PRINT 'A VS D ON BR2 & BR3';
RUN WJGLM;
C = {0 1 0 -1};
U = {1, -1, 0};
PRINT 'B VS D ON BR1 & BR2';
RUN WJGLM;
U = {1, 0, -1};
PRINT B VS D ON BR1 & BR3';
```

```
RUN WJGLM;
U = {0, 1, -1};
PRINT B VS D ON BR2 & BR3';
RUN WJGLM;
C = {0 0 1 -1};
U = {1, -1, 0};
PRINT 'C VS D ON BR1 & BR2';
RUN WJGLM;
U = {1, 0, -1};
PRINT 'C VS D ON BR1 & BR3';
RUN WJGLM;
U = {0, 1, -1};
PRINT 'C VS D ON BR2 & BR3';
RUN WJGLM;
```

THE UNIVERSITY OF MANITOBA

DEPARTMENT OF PSYCHOLOGY

Winnipeg, Manitoba
Canada R3T 2N2

March 22, 1995

Karen Thomas
American Psychological Association

Dear Ms. Thomas:

This letter is in regards to the copyright for the manuscript "Approximate degrees of freedom tests: A unified perspective on testing for mean equality" which is to be published in *Psychological Bulletin*. This paper was authored by myself and Dr. H. J. Keselman of the Psychology Department, University of Manitoba. I wish to be granted permission to include the computer program, which is found in Appendix B of the article, in my doctoral dissertation entitled "Probing interactions in repeated measures designs: Applications in clothing and textiles research". Please note that I was the individual who originally developed this computer program, and that I plan to reproduce the entire program in my dissertation.

Sincerely,

Lisa M. Lix
Ph. D. Candidate