INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

**THE ANALYSIS OF MULTIPLE RESPONSE DATA
IN MARKETING RESEARCH:
A RAO-SCOTT ADJUSTED CHI-SQUARED APPROACH**

By

Yves J. Decady

B. Sc. (Economics); M. A. (Public Administration)

**A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy**

**School of Business
Carleton University
Ottawa, Ontario**

September, 2000

**©Copyright
2000, Yves J. Decady**

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-57606-X

Canada

The undersigned recommend to the Faculty of Graduate Studies and Research

Acceptance of the thesis

**THE ANALYSIS OF MULTIPLE RESPONSE DATA
IN MARKETING RESEARCH:
A RAO-SCOTT ADJUSTED CHI-SQUARED APPROACH**

Submitted by Yves J. Decady, B. Sc. (Economics); M. A. (Public Administration)

in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Dr. Vinod Kumar, Director, School of Business

Dr. David Roland Thomas, Thesis Supervisor

Dr. U. N. Umesh, External Examiner

Carleton University

September, 2000

Abstract

In this study, Rao-Scott first- and second-order adjusted chi-squared tests for analysing one-way and two-way aggregate tables of multiple response data of the “pick any /r” type were developed. Extensions to other forms of multiple response data, such as “pick any”, “pick k/r”, “order r” and “order k/r” were also considered. These adjusted chi-squared tests were provided for five types of commonly used hypotheses, namely goodness-of-fit test, equality of response probabilities test, homogeneity test, independence test for cross-tabulations involving a single response variable and a multiple response variable, and independence test for tables involving two multiple response variables .

It appears that there are first-order tests for all cases that can be applied without access to the subject level data. For tables involving a single response variable and tables involving two multiple response variables, both a complete and a conjectured version of the first-order adjusted chi-squared tests were developed. In both cases, the conjectured tests worked as well as the complete versions. Monte Carlo simulation results suggest that these tests have good control of Type I error over a wide range of sample sizes and multiple counting indices.

The second-order tests provided particularly good control of Type I error over the full range of sample sizes and multiple counting settings. Wald statistics were also computed and compared to the Rao-Scott first- and second-order tests. A point of superiority of Rao-Scott methodology over the Wald procedure is that the Rao-Scott second-order test behaves properly across all sample sizes examined. The pseudo and modified chi-squared statistics recently proposed in the literature were found to be extremely conservative.

Acknowledgements

An endless search for knowledge, light and truth led me in June 1995 to the land of maple leaf where I found safe haven and a spiritual anchor for my hungry soul. It is with a heart full of gratitude that I pay tribute to this great country, Canada, its people and its institutions.

I wish to express my gratitude and appreciation to Dr. D. R. Thomas for his tireless guidance, direction, inspiration and unbounded patience in shaping this thesis from conception to completion. In addition to the invitation to glean from his field, he provided very generous financial support, which certainly eased the burdens associated with the dedication required for this study. Thanks for beaconing the way, throughout the development of this work, with high standard of excellence in statistical analyses.

To Dr. Judith Madill, I wish to extend a note of thanks for her kind support at critical times: entry into the program, thesis topic selection and generous financial contribution for a class project. I am particularly grateful for the intriguing questions raised at the time of my proposal defence and the accompanying challenges leading to uncovering possible extensions of the methods developed in this study.

Thanks to Dr. Georges Haines Jr. for attending the first presentation I made to Dr. D. R. Thomas before admission into the program and for the constant support provided since and at all phases of the program. Specially important were the suggestions and insights on simulation and random number generation testing. Thanks also for raising the questions of sparseness and aggregation, which led to developments found in Chapter 9 of this thesis.

Thanks to Dr. David Cray for accepting the invitation to be on my examining committee and for providing constant encouragement as well as insightful comments at proposal defence time. Thanks for the fundamental questions raised at proposal defence time on the desirable statistical properties of the statistics developed.

Thanks to Dr. Scott Bennett for accepting the invitation to be on my examining committee and for providing insightful comments at proposal defence time. Specially important was the suggestion to explore further the parallel between complex survey analysis and multiple response data analysis.

Thanks to Dr. U. N. Umesh for providing me with his two papers, his pioneering work in the field, on the analysis of multiple response data in marketing and for extracting, after a tedious amount of recoding, the unpublished data from his research, which he generously sent me. Thanks for accepting the invitation to be on my examining committee as external examiner.

As I witness the conclusion of this journey, I reflect upon the process and the new colleagues and friends made along the way, whose support and helpful hands were instrumental in bringing to pass this work. To Les Szamosi, Steve Murphy, Ajax Persaud and Michel Fuksa, my warm appreciation for your encouragement and your support -many thanks. I also give thanks to Greg Schmidt for saving me at sundry times from computational catastrophes.

To Greg Morrison, I express my appreciation for making the statistical section books on SAS accessible to me and for his timely advice, whenever it was required. His ongoing support has been reminiscent of our first encounter, when Dr. D. R. Thomas, on his first post-sabbatical days of work, introduced me to him and to Dr. Judith Madill while I was only visiting Ottawa.

For all the unspoken words of comfort and all the silent and discrete support from all professors and staff of the School of Business, I say thank you.

To my family, both nuclear and extended, I wish to express my gratitude for the emotional support, encouragement and understanding beyond measure that I have received.

Finally, I wholeheartedly express my gratitude to the Almighty for the light shed by and by on this project and for networking me with all these people, all of whom have had a profound influence on me and on this work.

Table of Contents

Abstract.....	i
Acknowledgements.....	ii
List of Tables.....	ix
Chapter 1.0	
Introduction	1
1.1 Definitions and Types of Multiple Response Data	1
1.2 The Importance of Multiple Response Data in Marketing	3
1.3 Tabular Formats for Aggregated Multiple Response Data	4
1.3.1 Multiple Response Data in Expanded Table Form	6
1.4 Identifying the Problem Areas	8
1.5 The General Research Objectives	10
1.6 Organization of the Dissertation	10
Chapter 2.0	
Methods Proposed to Date	12
2.1 Analysis of Disaggregated Multiple Response Data	12
2.1.1 The Family of Correspondence Analysis Methods	12
2.1.1.1 The Centroid Scaling Technique	12
2.1.1.2 Correspondence analysis	13
2.1.1.3 Dual scaling	14
2.1.2 Multidimensional scaling	16
2.1.3 Latent Class Analysis	18
2.1.4 Other Methods	19
2.2 Analysis of Aggregated Multiple Response Data	20
2.2.1 Weighted Least Squares and Wald Tests	20
2.2.2 Umesh's Indices of Predictive Association	22
2.2.3 The Pseudo and Modified Chi-Squared Statistics	23
2.2.4 Fractionally Adjusted Chi-Squared	24
2.2.5 Agresti-Liu's Approximate Chi-Squared Tests	24
2.3 Analysis of Multiple Response Data in Expanded Form	25
2.3.1 Loglinear Models	26
2.3.2 Generalized Loglinear Models	27
2.3.3 The Generalized Estimating Equations Approach	28
2.3.4 Resampling Methods	29
2.4 Summary and Conclusions	30

Chapter 3.0

Possible Approaches to the Analysis of Multiple Response Data	31
3.1 Assumptions on Which The Pearson Chi-Squared Statistic Is Based	31
3.2 Failure of the Pearson Statistic with Multiple Response Data	32
3.3 The Complex Nature of Multiple Response Data	32
3.4 Statistical Methods for Categorical Data from Complex Surveys	34
3.4.1 The Wald Weighted Least-Squares Approach	35
3.4.2 Rao-Scott Correction Methods	36
3.4.2.1 Rao-Scott First-Order Corrections	37
3.4.2.2 Rao-Scott Second-Order Corrections	40
3.5 Estimating Covariance Matrices for Use With Weighted Least Squares And Rao-Scott Techniques	42

Chapter 4.0

Testing Hypotheses in One-Way Multiple Response Tables: First-Order Rao-Scott Tests.	43
4.1 Simple Goodness of Fit	43
4.2 Equality of Response Probabilities	47

Chapter 5.0

Testing for Association in Two-Way Multiple Response Tables: First-Order Rao-Scott Tests.	53
5.1 Sample Design for Tests on Two-Way Tables of Single Response Counts	53
5.1.1 Independent Samples	53
5.1.2 A Single Random Sample (Single Response $\times$ Single Response)	56
5.1.3 Invariance to Distributional Assumptions	59
5.2 Two-Way Tables Involving Multiple Response Data	61
5.2.1 Single by Multiple Response Tables: Fixed Row Totals.	61
5.2.2 Single $\times$ Multiple Response, Random Row Totals	64

Chapter 6.0

Independence in Two-Way Tables When Both Variables Admit Multiple Response.	73
6.1 The Need for A Modified Chi-Squared Statistic	73
6.2 Derivation of the Modified Chi-Squared Statistic	73
6.3 Comparing the Correction Factor for the Multiple by Multiple Test with Correction Factors from the Other Tests	81

Chapter 7.0

Second-Order Rao-Scott and Weighted Least Squares Tests	82
7.1 A general Approach to the Second-Order Procedure	82
7.2 A General Approach to The Wald-WLS Methodology	84
7.2.1 Wald Goodness-of-Fit Test	86

7.2.2 The Wald Test of Equality of Item Response Probabilities	87
7.2.3 The Wald Test of Homogeneity	89
7.2.4 The Wald Test of Independence in the Single by Multiple Random Row Case	90
7.2.5 The Wald Test with Multiple by Multiple Tables	92
7.3 Application of the Second-Order Rao-Scott and WLS to Umesh's (1995) data set ..	92

Chapter 8

A Monte Carlo Simulation Study	97
8.1 Design of the Monte Carlo Study	97
8.2 The Statistics Examined	98
8.3 Simulation of the Goodness-of-Fit Test	100
8.3.1 Simulation Details	100
8.3.2 Simulation Results	101
8.4 Simulation of The Equality of Response Probabilities	104
8.4.1 Simulation Details	104
8.4.2 Simulation Results	104
8.5 Simulation of Homogeneity, i.e., Single (fixed) $\times$ Multiple Test	107
8.5.1 Simulation Details	107
8.5.2 Simulation Results	108
8.6 Simulation of the Independence, i.e., Single (random) $\times$ Multiple Test	110
8.6.1 Simulation Details	110
8.6.2 Simulation Results	111
8.7 Simulation of The Multiple $\times$ Multiple Test	114
8.7.1 Simulation Details	114
8.7.2 Simulation Results	116
8.8 Some Additional Simulation Results	120
8.9 Summary and Discussion of the Monte Carlo Results	120

Chapter 9

Extensions of the Proposed Methodology	123
9.1. Multiple Response Data of the "Pick k/r" Type	123
9.1.1 Goodness-of-fit, Equality of Response Probability and Homogeneity Tests	123
9.1.2 Single by Multiple and Multiple by Multiple Tests	124
9.2 Rao-Scott Corrections for "Order r" and "Order k/r Data"	125
9.2.1 "Order r"	126
9.2.2 "Order k/r"	129
9.3 Aggregation Across Multiple Response Variables	131
9.3.1 Summing Correlated Categories	131
9.3.2 Aggregation at the Incidence Matrix Level	133
9.4 Doubly Complex Multiple Response Data	134

Chapter 10

Summary and Conclusions	135
10.1 The Methods Developed	135
10.2 Technical Benefits	137
10.3 Application to Other Analytical Techniques	139
10.4 Managerial Relevance	141
10.5 Limitations	142

References	144
-------------------------	------------

Appendix 1

Distributions of Quadratic Forms	152
--	-----

Appendix 2

Correction Factor For Test of Equality of Item Probabilities in A One-way Array ...	157
---	-----

Appendix 3

Correction factor For A Homogeneity Test	161
--	-----

Appendix 4

Correction Factor for a Two-Way Multiple by Multiple Test	164
---	-----

Appendix 5

Correction Factor for A Single by Multiple Two-Way Table	170
--	-----

Appendix 6

Single by Multiple and Multiple by Multiple Tests via the Jackknife Procedure	172
---	-----

Appendix 7

Control of Type I Error Rates	173
-------------------------------------	-----

Appendix 8

Generating Correlated Binary Variates	177
---	-----

Appendix 9

SAS Codes Used in The Simulation	179
--	-----

List of Tables

Table 1.1	Country of Origin of Cars by Individual Respondent	4
Table 1.2	Effect of Country of Origin in Cars Selection	5
Table 1.3	Effect of Country of Origin and Gender in Car Selection	6
Table 1.4	Expanded Form of One-Way Aggregated Data	7
Table 5.1	Relationship Between Country of Origin and Gender in Choosing a Car (Single Response-Homogeneity)	54
Table 5.2	Probability Table Under Product-Multinomial Assumption	56
Table 5.3	Relationship Between Country of Origin and Gender in Choosing a Car (Single Response-Independence)	57
Table 5.4	Probability Table Under Single Sample Assumption	58
Table 5.5	Notation for A Single by Multiple Table: Umesh (1995) Example	61
Table 5.6	Effect of Country of Origin and Gender on Car Selection Random Margins (Multiple Response)	65
Table 6.1	Two-way Array of Multiple Response Data	74
Table 6.2	Probabilities Associated With Table 6.1	75
Table 6.3	Perceived Quality of Cars Versus Recommendation of Car for Purchase	80
Table 6.4	First-Order Rao-Scott Tests Based on Umesh's (1995) Data	81
Table 7.1	Second-Order Rao-Scott and Wald Tests Based on Umesh's (1995) Data	96
Table 8.1	Comparison of Actual Significance Levels for The Goodness-of-Fit Case Using $\tau' = [0.495 \ 0.728 \ 0.549]$, i.e. $\gamma = 1.772$ (Umesh values)	102

Table 8.2

Comparison of Actual Significance Levels for The Goodness-of-Fit Case

Using $\tau' = [0.20 \ 0.40 \ 0.60]$, i.e. $\gamma = 1.20$ 102**Table 8.3**

Comparison of Actual Significance Levels for The Goodness-of-Fit Case

Using $\tau' = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma = 1.80$ 103**Table 8.4**

Comparison of Actual Significance Levels for The Goodness-of-Fit Case

Using $\tau' = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma = 2.4$ 103**Table 8.5**

Comparison of Actual Significance Levels for The Equality

of Item Response Probabilities Case Using $\tau' = [0.59 \ 0.59 \ 0.59]$, i.e. $\gamma = 1.77$ 105**Table 8.6**

Comparison of Actual Significance Levels for The Equality

of Item Response Probabilities Case

Using $\tau' = [0.40 \ 0.40 \ 0.40]$, i.e. $\gamma = 1.20$ 105**Table 8.7**

Comparison of Actual Significance Levels for The Equality

of Item Response Probabilities Case Using

 $\tau' = [0.60 \ 0.60 \ 0.60]$, i.e. $\gamma = 1.80$ 106**Table 8.8**

Comparison of Actual Significance Levels for The Equality

of Item Response Probabilities Case

Using $\tau' = [0.80 \ 0.80 \ 0.80]$, i.e. $\gamma = 2.40$ 106**Table 8.9**

Comparison of Actual Significance Levels for

The Hypothesis of Homogeneity Case Using

 $\tau' = [0.495 \ 0.728 \ 0.549]$, i.e. $\gamma = 1.772$ and $p' = [0.54 \ 0.46]$ 108**Table 8.10**

Comparison of Actual Significance Levels for

The Hypothesis of Homogeneity Case Using

 $\tau' = [0.20 \ 0.40 \ 0.60]$, i.e. $\gamma = 1.2$ and $p' = [0.50 \ 0.50]$ 109**Table 8.11**

Comparison of Actual Significance Levels for

The Hypothesis of Homogeneity Case

Using $\tau' = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma = 1.80$ and $p' = [0.50 \ 0.50]$ 109**Table 8.12**

Comparison of Actual Significance Levels for

The Hypothesis of Homogeneity Case

Using $\tau' = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma = 2.40$ and $p' = [0.50 \ 0.50]$ 110

Table 8.13

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Single by Multiple Case
Using $\tau' = [0.5925 \ 0.6625 \ 0.5975]$, i.e. $\gamma = 1.8525$ and $p' = [0.54 \ 0.46]$ 112

Table 8.14

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Single by Multiple Case
Using $\tau' = [0.20 \ 0.40 \ 0.60]$, i.e. $\gamma = 1.2$ and $p' = [0.50 \ 0.50]$ 113

Table 8.15

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Single by Multiple Case
Using $\tau' = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma = 1.80$ and $p' = [0.50 \ 0.50]$ 113

Table 8.16

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Single by Multiple Case
Using $\tau' = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma = 2.40$ and $p' = [0.50 \ 0.50]$ 114

Table 8.17

Comparison of Actual Significance Levels for The Hypothesis of
Independence in The Multiple by Multiple Case Using $\tau'_i = [0.5925 \ 0.6625 \ 0.5975]$
and $\tau'_{.j} = [0.5425 \ 0.46 \ 0.2775]$, i.e. $\gamma_i = 1.8525$ and $\gamma_{.j} = 1.28$ 117

Table 8.18

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Multiple by Multiple Case Using
 $\tau'_i = [0.20 \ 0.40 \ 0.60]$ and $\tau'_{.j} = [0.20 \ 0.40 \ 0.60]$, i.e. $\gamma_i = 1.20$ and $\gamma_{.j} = 1.20$ 117

Table 8.19

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Multiple by Multiple Case Using
 $\tau'_i = [0.20 \ 0.40 \ 0.60]$ and $\tau'_{.j} = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma_i = 1.20$ and $\gamma_{.j} = 1.80$ 118

Table 8.20

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Multiple by Multiple Case Using
 $\tau'_i = [0.30 \ 0.60 \ 0.90]$ and $\tau'_{.j} = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma_i = 1.80$ and $\gamma_{.j} = 1.80$ 118

Table 8.21

Comparison of Actual Significance Levels for The Hypothesis of Independence in The
Multiple by Multiple Case Using
 $\tau'_i = [0.30 \ 0.60 \ 0.90]$ and $\tau'_{.j} = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma_i = 1.80$ and $\gamma_{.j} = 2.40$ 119

Table 8.22

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Multiple by Multiple Case Using
 $\tau'_i = [0.90 \ 0.80 \ 0.70]$ and $\tau'_{.j} = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma_i = 2.40$ and $\gamma_{.j} = 2.40$ 119

Table 9.1	
Typical Cross-Tabulation of "Order r" Data	126
Table 9.2	
Typical Cross-Tabulation of "Order k/r" Data	130
Table 10.1	
Rao-Scott First-Order Tests Developed for Multiple Response Data of The "Pick Any/r" Types	136

Chapter 1.0

Introduction

This thesis will focus on the analysis of multiple response data, a data format common in marketing research and other areas of business study. The primary objective will be to identify deficiencies and gaps in the array of statistical methods available for analysing multiple response data, and to remedy some of these deficiencies by developing new procedures that can be readily applied by marketing researchers. From the literature review, it will be seen that methods for analysing tables of aggregate multiple response data are particularly deficient, and the theoretical developments will therefore be focused on this area.

1.1 Definitions and Types of Multiple Response Data

Multiple response data in research arise from questions that require the respondents to make one or more choices from among a restricted or unrestricted set of alternatives. This type of data is referred to in the literature as “pick any” data for which Coombs (1964), in his seminal work on theory of data, provided a “typology of data collection methods”. In keeping with the terminology from data theory, the restricted and unrestricted types of multiple response data will be referred to in this study as “pick any/r” and “pick any”, respectively. “Pick k/r” is a variant of “pick any/r” whereby two constraints are imposed on the choice set by allowing respondents to select a predefined number of items (k) from a well defined universe (r). Collectively, all these types will be referred to as multiple response data (MRD).

As an example, a consumer making an automobile choice may select one or more cars from a large set of luxury automobiles, i.e., she may "pick any" from an undefined choice set. Alternatively, the consumer may select from a more limited choice set of say, seven luxury automobiles, e.g., {Acura, BMW, Cadillac, Corvette, Lexus, Mercedes, Porsche}, leading to "pick any/ r " data, where $r = 7$. As another example, a consumer may be asked to list the benefits she or he is seeking by consuming a product (pick any) or to choose a subset of k benefits from a predefined list (pick k out of r). As an example from secondary data analysis, a pick any data set can be constructed using purchase histories obtained from either diary or scanner panel data. Such a data set would contain ones for the brands chosen during a particular period and zeros for the brands not chosen during the same period of time. The data set for any purchase period can be presented as a matrix of rows representing different consumers and columns representing different brands.

With multiple response data of the "pick any" form, a non-chosen alternative may be either unacceptable, or acceptable but not considered, or acceptable and considered but not chosen (Levine, 1979). The identifying characteristic of these data is that the alternatives from which the preferred objects are selected may be assumed to vary from subject to subject. Also their choices may overlap completely, or partially, or not at all. Levine (1979) noted that such type of data carry a built-in ambiguity which makes them relatively difficult to analyse. These ambiguities are resolved to a large extent by the imposition of a constrained choice set, as in "pick any/ r " and "pick k/r " data, and the proposed research will focus primarily on data of the "pick any/ r ". Extensions to multiple response data of the "pick k/r ", "order r " and "order k/r " types will be considered briefly.

Nevertheless even with constrained choice sets, analytical and statistical problems remain, and the main objective of this research will be to identify and provide solutions to some of these statistical problems.

1.2 The Importance of Multiple Response Data in Marketing

One of the more typical kinds of data in marketing research for measuring consumer choice behaviour requires the consumers to select the brands used from a set of r brands, either using a constrained set, which corresponds to "pick any/ r " or using an unconstrained choice set, namely "pick any" (Green *et al.*, 1988; Hoffman and Franke, 1986).

In segmentation research, pick any data (e.g., benefits sought, or occasions of usage), have been used most often as the first stage in a multi-stage investigation, to identify the most important items. In the second stage, the retained meaningful sets of items are usually intervally scaled, which avoids the need to deal directly with the multiple response data in its original form. With intervally scaled data, classical multivariate analysis techniques can be used which automatically account for within subject correlations. Despite the fact that multiple response data are widely collected in marketing, anecdotal reports from the field and from the academic community agree that the analysis of this type of data is difficult, particularly when they are cross-tabulated, and that many researchers avoid the use of all kinds of multiple response data for this reason. Nishisato (1994, page 27) has even advised "avoiding as much as possible collecting multiple response data".

1.3 Tabular Formats for Aggregated Multiple Response Data

Multiple response data of the “pick any/r” type can be displayed in a variety of formats. In the first instance, a disaggregated multiple response table will be presented, followed by a table of aggregated data. In what follows, it will be assumed that each respondent makes at least one selection.

The basic format in which subjects (respondents) are placed in rows and responses (items) in columns is often called a response pattern table (Nishisato, 1980), an indicator matrix (Gifi, 1990) or an incidence matrix (Greenacre, 1984). This format, as illustrated below for pick any/3 data, represents the disaggregated form of multiple response data.

Table 1.1

Country of Origin of Cars by Individual Respondent

Respondent	Domestic	Japanese	European
1	0	1	1
...	...	...	...
n - 1	1	0	1
n	1	0	1

The disaggregated data of Table 1.1 can be aggregated over all respondents to yield a one-way table of counts, as shown in Table 1.2, which represents the number of

times (m_j , $j = 1, 2, 3$) each automobile origin is selected. Given that each respondent can select more than one automobile origin, $\sum m_j \geq n$. Whenever the inequality is strict, this will signal "multiple counting", i.e., at least one respondent will be counted in more than one cell.

Table 1.2
Effect of Country of Origin in Cars Selection

Country of Origin				
	Domestic	Japanese	European	Total
m_i	237	265	239	400

If the gender of all respondents in Table 1.1 is known, then aggregating over respondents by gender gives rise to a two-way cross-tabulation of a multiple response variable with a single response variable as in Umesh (1995). Table 1.3 offers an illustration, using the Umesh data. If the multiple response counts are denoted m_{ij} , $i = 1, 2$, $j = 1, 2, 3$ and the number of males and females are denoted n_1 and n_2 respectively, then it can be seen from Table 1.3 that the counts satisfy $\sum m_{1j} \geq n_1$ and $\sum m_{2j} \geq n_2$.

Table 1.3
Effect of Country of Origin and Gender
in Car Selection

Country of Origin				
Gender	Domestic	Japanese	European	Total
Male	146	131	138	216
Female	91	134	101	184
Total	237	265	239	400

1.3.1 Multiple Response Data in Expanded Table Form

There is one way of aggregating multiple response data that avoids the multiple counting problems associated with the data formats of Tables 1.2 and 1.3. Referred to as the expanded form of aggregated MRD, this format will be key to several of the analysis strategies to be reviewed in Section 2.3 of this thesis. Thus it will be briefly described here.

For convenience, a three-item multiple response variable will be considered, as illustrated in Table 1.1. The three individual item responses (0 or 1) can be represented by three binary variables with outcomes: yes = 1 (selected); no = 0 (not selected). Each combination of zeros and ones (the response patterns) in a single row of the incidence matrix (Table 1.1) can then be represented as a single entry in one of the cells of a $2 \times 2 \times 2$ cross-tabulation of the three binary variables. Aggregation over the rows of the incidence matrix (Table 1.1) then results in a $2 \times 2 \times 2$ table of aggregate counts, referred to as the

expanded form of Table 1.2. Such an expanded table, corresponding to the multiple response counts of Table 1.2, is illustrated in Table 1.4.

Table 1.4
Expanded Form of One-Way Aggregated Data

	Japanese = Yes		Japanese = No	
Domestic	European		European	
	Yes	No	Yes	No
Yes	106	47	32	52
No	50	62	51	0

Because only one cell of the $2 \times 2 \times 2$ representation of each row of the incidence matrix contains a 1 (representing a particular response pattern), the aggregate cells in the expanded table involve no multiple counting. As can be seen from Table 1.4, the eight cell counts sum to 400, the total sample size. This property of expanded tables will be very important in later application. Since it is assumed that the respondents provide at least one selection, the zero in the table constitutes a structural zero, and will not feature in the analysis. Expanded tables can be formed for multiple response variables having arbitrary number of items (denoted C), in which case they will contain $2^C - 1$ cells. The expanded table principle can also be applied to tables of the form of Table 1.3. For such a

table, each row would take an expanded form, resulting in a $2 \times (2^3 - 1)$ aggregation. For a table containing an arbitrary number R of rows (strata) and C items, the resulting expanded table would be an $R \times (2^C - 1)$ cross-tabulation.

1.4 Identifying the Problem Areas

Many techniques have been developed for analyzing multiple response data in the formats presented above. These focus primarily on the analysis of disaggregated data, i.e., on data sets in the form of incidence matrices (Table 1.1) that preserve the individual identities of respondents and items. Examples include correspondence analysis (Hoffman and Franke, 1986; Kaciak and Louviere, 1990), dual scaling (Nishisato, 1994), multidimensional scaling (Caroll and Green, 1997), latent class analysis (Bockenholt and Bockenholt, 1990) and the stochastic methods developed by DeSarbo and Cho (1989). These techniques answer questions relating to the relationships between and among items and respondents, and provide a representation of these relationships in a joint-low dimensional space. Their primary objective is description and they each satisfy this objective to a greater or lesser degree. However, some of the methods proposed for disaggregated data incorporate inferential measures in the form of variance estimates, and it will be shown in the literature review to follow that some inferential difficulties remain to be resolved.

From a reading of the literature, it is overwhelmingly evident that methods for analysing disaggregated multiple response data far outnumber the methods available for analysing multiple response data in the form of aggregated tables of counts. The latter remains a challenge for both experimental and survey-based researchers.

To test hypotheses relating to goodness-of-fit or association for aggregate data in the form shown in Tables 1.2 and 1.3, respectively, statistics of the familiar “chi-squared” type are a natural first choice. However, for aggregate multiple response data, the categories cease to be mutually exclusive, since some sampled subjects will be represented in more than two cells of a table, i.e., the cell counts may overlap partly or totally.

This is illustrated by the fact that the sum of the row totals of Table 1.3 is considerably greater than the total number of 400 respondents. As noted earlier, this characteristic of multiple response data is referred to as “multiple counting”. The net effect of multiple counting, which violates the assumption of exclusivity, is that the distribution of cell counts is no longer multinomial, and the classical “chi-squared” statistics no longer conform to their theoretical (large sample) chi-squared distribution. No doubt because of the difficulty noted above, cross-tabulations involving multiple response data have been to a considerable extent avoided by practitioners.

The literature review will describe several methods that have been proposed for testing hypotheses (and measuring association effects) on aggregate tables of multiple response data. However, it will be seen that these methods are, for the most part, complex and computationally intensive, and are often based on methodologies that are unfamiliar

to most practitioners. A general conclusion will be that methods for analysing aggregating multiple response data are relatively poorly developed when compared to the techniques available for disaggregated data.

1.5 The General Research Objectives

Thus the primary objective of this research will be to extend the array of methods available to practitioners for analysing and testing hypotheses on aggregate tables of multiple response data. A part of this objective will be to develop methods that are easy to apply and that are based on well known and familiar testing strategies. Both aspects of the research objective will be met by adapting a theoretical framework developed for the analysis of data drawn from complex surveys, i.e., surveys involving stratification and clustering.

1.6 Organization of the Dissertation

Chapter 2 will present a critical review of available methods for the analysis of both disaggregated and aggregated multiple response data. A detailed examination of the technical and statistical problems associated with multiple response data will be provided in Chapter 3. The complex nature of multiple response data will be explained and the parallel with complex survey analysis clarified. Some techniques developed recently for analysing complex survey data will be described, and their relevance to the analysis of cross-tabulated multiple response data will be discussed. Chapter 4 will describe the application of the simplest (first-order) Rao-Scott procedures to the problem of testing a

goodness-of-fit hypothesis on a one-way table of multiple response counts. This will be followed in Chapter 5 by the development of first-order Rao-Scott procedures for testing association in two-way tables involving a single- and a multiple-response variable. The more general case of association in two-way tables involving multiple response variables will be considered in Chapter 6.

Chapter 7 describes the application of the more accurate second-order Rao-Scott procedures and includes some specifics relating to Wald tests. A Monte Carlo simulation study designed to evaluate the finite sample properties of the proposed methods will be described in Chapter 8. This will include a comparison among variants of the proposed procedures, together with an evaluation of some of the alternative procedures that have been proposed.

Chapter 9 provides extensions of the Rao-Scott approach to multiple response data of the “pick k/r ”, “order k/r ” and “order r ” types of data. Cross-tabulations involving aggregation across multiple response variables are also considered in this chapter.

The benefits and the limitations of this study will be discussed in Chapter 10 and issues that require further research will be identified.

Chapter 2.0

Methods Proposed to Date

2.1 Analysis of Disaggregated Multiple Response Data

In this section, the techniques used to analyse multiple response data (MRD) in the disaggregated form of Table 1.1 will be presented. It will be shown that multiple response data displayed in this format are predominantly analysed by scaling techniques. For each technique, a brief description is given and some marketing applications are pinpointed.

2.1.1 The Family of Correspondence Analysis Methods

The title “correspondence analysis” is used to refer to a group of scaling techniques, such as correspondence analysis (Benzecri, 1973; Greenacre, 1984), dual scaling (Nishisato, 1980) and centroid scaling, (Levine, 1979). The rationale for this grouping is provided by various authors (DeSarbo and Cho, 1989; Tenenhaus and Young, 1985; Cox and Cox, 1994).

2.1.1.1 The Centroid Scaling Technique

This technique, due to Levine (1979), consists essentially in representing subjects as points which are, up to a scale factor, the centroids of points corresponding to the alternatives that have been chosen, with the alternatives located at points which are, up to a scale factor, the centroids of the locations of the subjects who have chosen them. It

provides a way of displaying subjects and all alternatives for all subjects as points in a comprehensive joint low-dimensional space. It assumes that, when asked to choose among alternatives, the subjects choose only a subset. Therefore the centroid scaling technique task consists of estimating or recovering the unknown joint space from the choices alone.

Green and DeSarbo (1981) applied Levine's technique in a survey asking the respondents to list all foreign sports cars that they would like to own, if money were not a major constraint (unrestricted choice set). Holbrook, Moore and Winer (1980), also applied Levine's method to study and represent competitive market positions.

2.1.1.2 Correspondence analysis

Correspondence analysis as advocated by the French school (Benczezi, 1973; Greenacre, 1984) is a descriptive statistical method that graphically represents the rows and columns of a matrix of categorical data in the same low-dimensional space. This data matrix can take form of an incidence matrix, or a contingency table or even a multi-way table. In correspondence analysis, numerical scores in corresponding units are assigned to the rows and columns of a data matrix so as to maximize their relationship. Simply expressed, it is a way of jointly scaling a matrix of non-negative data to represent objects described by rows or columns as points in a space of suitably low dimension.

Multiple correspondence analysis (MCA) is the multivariate extension of correspondence analysis (CA) to the joint analysis of multi-way tables. Binary choice, "pick any", or any other type of discrete and/or continuous data can be analysed by CA

(Kaciak and Louviere, 1990). Multiple response data must be “doubled” (Greenacre, 1984, p.171), and continuous data must be discretized or recoded into categories (Greenacre, p. 159). “Doubling” means that both the original (selected) and the reflected (not selected) forms of the data are submitted to the CA algorithm, a modification that establishes a symmetry between the poles of each bipolar variable so that the correspondence analysis is invariant to the choice of scale direction. It is worth noting that correspondence analysis (CA) and multiple correspondence analysis (MCA) can also be used to analyse contingency tables.

Hoffman and Franke (1986) applied CA to “pick any/r” data and used the doubling principle; Green *et al.*, (1988) also applied CA to data of the “pick k/r” type. CA and MCA are descriptive statistical tools and are not appropriate for hypothesis testing. However, it is possible to compute standard errors and replace the points in CA and MCA plots by 95 percent confidence ellipses (Gifi, 1990). However, it should be noted that these attempts to include statistical inferences in CA and MCA are only justified when the cell frequencies follow a multinomial distribution, which is not the case with multiple response data (MRD).

2.1.1.3 Dual scaling

Dual scaling involves quantification of rows and columns of a row-by-column categorical table so as to optimize a certain criterion in the least-squares sense, such as minimizing the within-row sum of squares and maximizing the between-row sum of squares (Nishisato, 1980). In the words of Nishisato (1980, p. 12), it is an extension of

Eckart and Young's (1936) decomposition through principal component analysis to categorical data.

As with CA, dual scaling can be applied to a subjects-by-questions matrix, or a contingency table, or even a multi-way table (see Nishisato, 1994, p.307-319). However, dual scaling quantifies a broader type of data than CA, such as paired comparisons, rank-order, ratings and sorting data (Nishisato, 1994). The duality of the scaling procedure is expressed by the symmetric relations between derived weight vectors for the columns and rows of the data matrix. Full characterization of this technique is provided in Nishisato (1980, 1994).

Hill's (1974) description of CA as an analogue of principal components analysis suggests that correspondence analysis and dual scaling are indistinguishable. In fact, in their basic form, both methods can be traced back to the same starting point, and in their simplest form they are indeed equivalent. However, Nishisato (1994) considers that both methods have passed the stage of basic formulation, and that separate developments and refinements have introduced sufficient differences for the two methods to be considered distinct. It should also be noted that there is a strong mathematic similarity between both these methods and Levine's centroid scaling technique.

Nishisato (1980,1994) considers multiple response data to be more complex to handle in dual scaling than is the standard data format of one response per question. Therefore, he recommends turning multiple choice questions into dichotomous questions. Alternatively, he also suggests that such data can be simplified by treating multiple responses as trend variables, e.g., giving the responses a score of $1/k$ (k being

the number of categories of the multiple response variable) , thus returning the correct respondent total weight of 1.

2.1.2 Multidimensional scaling

Multidimensional scaling (MDS) is a set of estimation procedures designed to obtain a multidimensional spatial representation of the structure in various types of data (Kruskal and Wish 1978; Carroll and Green, 1997). It is generally applied to data such as proximity measures (similarity or dissimilarity) or dominance measures (preferences, ratings). With some additional processing, it can be applied to discrete choice data, including data of the “pick any” type. A typical implementation of MDS is to construct a square matrix of proximity measures, where each matrix element represents the proximity of object (row) “i” to object (column) “j”. The MDS model then assumes the existence of a metric space containing the objects, in which the metric distance between the objects corresponds as closely as possible their measured proximities. The estimation process consists of choosing the dimension of the metric space and locating each object in that space.

Holbrook et al.(1982) used 0-1 brand choice data to construct multidimensional maps and compared them with maps constructed based on complete rank orderings data. Petit (1995) collected multiple response data of the “pick any/r” form and constructed proximity measures between firms and between products and used MDS to display the firms and the products graphically, i.e., in their respective metric space. Proximity measures between either the rows or the columns of an incidence matrix can be obtained

using some ad hoc measure of distance, such as the Jaccard distance (Romesberg, 1990).

Attempts to develop inferential measures for MDS have been based on the maximum likelihood (ML) principle (Ramsay, 1980; Takane, 1981; Zinnes and MacKay, 1983). To develop their maximum likelihood multidimensional scaling (MLMDS) approach these authors assume that the observed dissimilarity data are error-perturbed, and therefore the error component can be explicitly modelled using an hypothesized probability distribution, namely the normal or log-normal distribution. In marketing, both DeSarbo and Hoffman (1986) and DeSarbo and Cho (1989) have proposed an MLMDS approach based on the assumption that the underlying choice process of a particular person choosing a particular object is a Bernoulli process, a probability model that appropriately accounts for multiple response data. Along these lines, DeSarbo and Hoffman (1986, 1987), using the logit function, developed a stochastic MLMDS-based choice model for the estimation of joint space of subjects and objects from "pick any/r" and "pick any" data. DeSarbo and Cho (1989), using later the probit function, devised a new MLMDS-based choice model for the estimation of a joint space of individuals and stimuli from "pick any/r" binary choice data.

Elrod (1988), applying MDS to scanner/panel data, proposed a new stochastic brand choice model, "choice map", that simultaneously infers the distribution of consumer preferences and brand positions in a map. Maximum likelihood is used to estimate model parameters and to provide tests for the number of dimensions.

2.1.3 Latent Class Analysis

Latent class analysis (LCA) can be viewed as a data unmixing procedure (Dillon and Kumar, 1994; Green *et al.* 1976). With LCA, the population is assumed to be heterogeneous and consists of homogeneous subgroups of unknown size which are called components of the mixtures or simply components. The key tasks consist in finding the component parameters and the mixing proportions. When latent class analysis (LCA) is applied to dichotomous data, it is assumed that the k random (Bernoulli) variables are independent so that their joint pdf is the product of the marginal probabilities. Several computational procedures for finding maximum likelihood estimates of the parameters and proportions exist. Goodness of fit is tested by calculating a chi-square test based on actual versus fitted frequencies.

Ramaswamy et al. (1996) used a latent class analysis to perform a joint segmentation based on the collection of “pick any/r” data by having customers pick both their financial investment goals (benefits sought) from a list of 13 goals and the banking services (usage of services) they use from a list of 13 services. The segments, in this case, corresponded to the identified latent classes. Let us note in passing that each class can be considered to be aggregate multiple response counts from correlated binary variates so that the assumption of local independence may lead to considerable bias in the estimates of mixing proportions (Vacek, 1985).

Due to interests vested by researchers in the area of market segmentation, latent class multidimensional scaling (LCMDS) was developed to satisfy the need to perform simultaneously MDS and segmentation/cluster analysis. The solution to a LCMDS problem can be found using, for example, the expectation-maximization algorithm. The work by Bockenholt and Bockenholt (1990) typifies the use of LCMDS for analysing "pick any/r" data.

2.1.4 Other Methods

Cluster analysis refers to a group of multivariate techniques whose primary purpose is to group objects based on the characteristics they possess so that objects in the same cluster are more like one another than they are like objects in other clusters. Recently, Petit (1995) used "pick any/r" data to derive overlapping clusters in a study on market definition for underwater fiber cable systems.

"Pick any" data have also been used in market response models. Brand distance from ideal points derived from maps based on "pick any" data are modelled as a function of price, advertising, and lagged market share (Holbrook et al.1982). Winner and Moore (1989), a few years later, reported a study covering various applications of "pick any" data aimed at evaluating the effects of marketing-mix variables on brand positioning in which scanner panel data were used to create maps from "pick any" data. Also scanner panel data have been used to produce a series of longitudinal joint spaces for a number of time periods, bypassing the need to collect traditional multidimensional scaling (MDS) data (Moore and Winner, 1987).

2.2 Analysis of Aggregated Multiple Response Data

In this section, an effort is made to describe the extent to which multiple response data (MRD) have been or can be analysed at different levels of aggregation. Hypothesis testing appears to be a major focus of the techniques applied to multiple response data in aggregated table format, though some attention has also been focussed on measures of association (Umesh, 1995). Again, a brief description of each technique is provided, and marketing applications are indicated.

2.2.1 Weighted Least Squares and Wald Tests

The method of weighted least squares is a very general statistical technique that can be used to construct large sample tests for a very wide range of categorical data problems (Agresti, 1990). In its most general form, the WLS approach involves fitting a linear model to functions of observed proportions, and estimating the parameters of this model using the principle of weighted least squares. The weight matrix is taken to be the inverse of the covariance matrix of the functions of proportions. Corresponding tests on model parameters are formulated using Wald statistics, which have large sample chi-squared distributions. Details can be found, for example, in Agresti (1990). Loughin and Scherer (1998) reported simulation results for weighted least squares tests of association on tables of the form shown in Table 1.3. Results were not encouraging. However, simulation results to be discussed in this thesis suggest that WLS procedures can sometimes be of practical value.

A simple example of the application of WLS and Wald tests to multiple response data is provided by Wild and Seber (1993,1994) who proposed a method for testing the equality of pairs of item response probabilities in a one-way table of multiple response counts. They constructed a large sample normal test, equivalent to a chi-squared Wald test on one degree of freedom. To implement Wild and Seber's (1993) large sample test, which is equivalent to McNemar's test as these authors noted, either the full incidence matrix or the expanded form of the aggregate table is required (see Sections 1.3 and 4.2).

Of particular interest is the conservative version of their test that can be used given only the aggregated multiple response data, i.e., when neither the full data set nor the expanded multinomial table is available, as is often the case with published summary data. The problem with a conservative test, as the authors pointed out, is that if it is not significant, one cannot be sure whether the true value of the test statistic would be significant or not.

For a multiple response variable having C categories, Wild and Seber's technique could be applied separately to the $C(C-1)/2$ pairs of comparisons of proportions. For a large number of multiple-response items, this leads to a large number of pairwise tests, necessitating control of overall Type I error rate via the Bonferroni adjustment. Such an approach has not been applied in the marketing literature.

2.2.2 Umesh's Indices of Predictive Association

Cross-tabulations involving two multiple-response variables or a multiple-response variable and a single-response variable have been recently investigated by Umesh *et al.* (1992), and Umesh (1995). Umesh (1995) presented a method to estimate the association between nominal multiple response variables through the use of what he termed the multiple index of predictive association (MIPA). This index can be computed for: (1) the case in which the dependent variable is a single-response variable and the independent is a multiple-response variable; (2) the case in which both the dependent and independent variables are multiple-response variables; and (3) the case in which the dependent is a multiple-response variable and the independent is a single response variable.

Umesh *et al.*'s (1992) and Umesh's (1995) measures, lambda-L and lambda-M, are similar in principle to Goodman and Kruskal's (1979) lambda measures and provide a prediction of the most likely response of an individual to a nominally-scaled variable. As indicated by the author, Lambda-L, is useful in situation in which a researcher wants to use one variable to predict at least one of many choices or options of a consumer. Lambda-M is appropriate when a conservative decision has to be made in the context of a new-product introduction, for example. Unlike Goodman and Kruskal (1954), Umesh (1995) did not provide variance estimation for his measures and neither did he devise a testing procedure for them. As a result, they remain descriptive measures of predictive association.

2.2.3 The Pseudo and Modified Chi-Squared Statistics

Umesh (1995), in a search for a test of association between a multiple response variable and a single response suggested that it is possible to compute the analogue of a chi-squared test. However, he acknowledged that such a test should be termed a 'pseudo chi-square', because of the problem of multiple counting, which induces additional correlations among the response categories. Thus, he noted, the resulting statistic would not have a large sample chi-square distribution.

Loughin and Scherer (1998) re-derived Umesh's (1995) pseudo chi-square test when they proposed a modified chi-squared test for cross-tabulated multiple response data. They explored theoretically the asymptotic null distribution of their (and Umesh's) statistic and showed that it is equivalent to a linear combination of chi-squared random variables, each having one degree of freedom. Such distributions are difficult to tabulate. For the test of association on a cross-tabulation of a multiple response and a single response variable (e.g. the data of Table 1.3), Loughin and Scherer (1998) used a bootstrap resampling technique (see, for example Efron and Tibshirani, 1993) to estimate the sampling distribution of the pseudo chi-squared statistic, instead of basing a test on its asymptotic distribution. From their bootstrap distribution they obtained a p -value for their statistic. One advantage of this approach is that it uses a familiar test statistic; a major disadvantage, on the other hand, is the use of computer intensive bootstrapping.

2.2.4 Fractionally Adjusted Chi-Squared

Loughin and Scherer also reported a fractionally adjusted Pearson statistic for testing association in a single by multiple table in which each respondent's positive responses in the incidence matrix are replaced by the reciprocal of the total number of positive responses made by that respondent. With this procedure, each unit contributes a total frequency of one to the table, avoiding the multiple counting problem. This is reminiscent of Nishisato's (1980, 1994) recommendation for multiple response data (MRD) in dual scaling. Loughin and Scherer (1998) treated the fractionally adjusted Pearson statistic as a chi-squared random variable on the standard number of degrees of freedom, but in their simulation it turns out to be overly conservative. A theoretical explanation of this finding will be provided later.

2.2.5 Agresti-Liu's Approximate Chi-Squared Tests

Recently, Agresti and Liu (1999) have suggested the use of a "naive chi-squared" statistic for tests of association on single by multiple response tables (see Table 1.3). This naive chi-squared" is constructed by breaking the R (single) $\times$ C (multiple) table into C two-way tables of dimension $R \times 2$, the two-category variable representing selection or otherwise of the j^{th} item, $j = 1, \dots, C$, and then summing the resulting C statistics from the separate marginal tables, treating them as if they were independent. The degrees of freedom for the naive test is the sum of the individual df of the separate marginal tables, (i.e., $R - 1$), resulting in total degrees of freedom $C(R - 1)$. Using the same principle, Agresti and Liu (1999) also computed a naive likelihood-ratio statistic. It should be noted

that the assumption of independence of the marginal tables is untenable, particularly with multiple response data. Lewis and Burt (1949) showed that even for single response data, summing separate estimates of χ^2 to obtain a composite estimate is incorrect since these are not independent. Agresti and Liu (1999) also suggested an alternative based on the Bonferroni approach, to obtain a valid way of simultaneously using the marginal Pearson statistics to test multiple marginal independence. A drawback of this approach is that it is somewhat conservative.

The naive test has the nice property of being invariant to the switching of the “yes” and “no” labels for all items. It also gives test values that are comparable to those obtained from elaborated procedures based on maximum likelihood methods. Agresti and Liu (1999) showed that was the case with Loughin and Scherer’s (1998) data. It has been verified by the author that this property also holds for the Umesh (1995) data . It has also been verified that their test can easily be extended to association tests on two-way tables when both variables are of multiple response form.

2.3 Analysis of Multiple Response Data in Expanded Form

In this section, the techniques used to analyse multiple response data in expanded form will be presented. Once again, a brief description of each technique is given and some marketing applications are mentioned. Modelling is not the major preoccupation of the methods described in this section. Models are used only to construct relevant tests. It is also important to remember that methods based on expanded tables can never be applied to data in the aggregated form, illustrated in Tables 1.2 and 1.3.

2.3.1 Loglinear Models

A log-linear model specifies the logarithms of the expected cell frequencies of a cross-tabulation of two or more single-response categorical variables as linear functions of parameters representing characteristics of the categorical variables and their interactions with each other (Agresti, 1990). The extent to which the frequencies expected under the model approximate the frequencies actually observed is the key to assessing model fit (Agresti, 1990). With loglinear models, all variables are treated as response variables, and researchers are able to test a variety of hypotheses, such as complete statistical independence of variables, conditional independence of variables, marginal homogeneity, etc. Estimation and testing are based on solving maximum likelihood equations corresponding to the multinomial distribution.

Several authors, including Hoffman and Franke (1986), Umesh (1995) and Agresti and Liu (1998, 1999) have suggested the use of loglinear models for analysing particular arrays of expanded multiple response data. In particular, Umesh (1995) and later Agresti and Liu (1998) constructed a test of association for a two-way cross-tabulation of a multiple response variable and a single response variable (e.g., the setup shown in Table 1.3) by contrasting two specific loglinear models defined on the corresponding expanded (multinomial) table. Various marginal models having simpler joint distribution can be used to test association via log-linear modelling. For example, Agresti and Liu (1998) fit a quasi-symmetric loglinear model to the Loughin and Scherer (1998) data.

A difficulty with the loglinear model approach is that not all hypotheses relating to the multiple response items can be expressed in loglinear forms. In fact, there are many hypotheses of interest on aggregated multiple response tables that cannot be formulated at all using loglinear models on expanded tables. A simple example is the test of equality of response probabilities in a one-way array corresponding to a C-item multiple response variable.

2.3.2 Generalized Loglinear Models

To address hypotheses such as the equality of response probabilities, a more general model formulation such as the generalized loglinear model, discussed in detail by Lang and Agresti (1994), is required. Agresti and Liu (1999) recently used the generalized loglinear model to test for association in tables of the Umesh (1995) type, namely two-way cross-tabulations of an R-category single response variable by a C-category multiple response variable. They used the constrained maximization algorithm developed by Lang and Agresti (1994) to obtain maximum likelihood estimates of the parameters of their loglinear model, subject to the constraints implied by the independence hypothesis on the aggregate $R \times C$ table. Applying this technique, they constructed valid chi-squared and likelihood ratio tests on the expanded table corresponding to Loughin and Scherer's (1998) table of aggregate data (with $R = 5$, $C = 5$). However, Agresti and Lang's (1994) iterative methods are difficult to implement and are not readily available in statistical packages. Further algorithmic developments have recently been reported (Lang *et al.*, 1999) that make such approaches more feasible.

To address hypotheses such as the equality of response probabilities, Lang and Agresti's (1994) generalized loglinear model is the only way of obtaining maximum likelihood estimates and corresponding tests. As noted above, methods based on the generalized loglinear model cannot yet be considered a practical solution for most practitioners.

2.3.3 The Generalized Estimating Equations Approach

A quasi-likelihood approach for obtaining estimates in marginal models is that based on generalized estimating equations (GEE) (Liang and Zeger, 1986). With GEE, it is assumed that a transformation of the set of correlated response variable means is a linear function of the covariates and that the covariance matrix is a known function of the mean. An assumed working covariance matrix is used, which can be iteratively improved as necessary. It has been shown that the solutions to the GEE are asymptotically consistent and normally distributed, even if the intra-subject correlation structure is misspecified. One can use iteratively reweighted least squares to obtain a solution to the estimating equations. An application of GEE to the Loughin and Scherer (1998) data is provided by Agresti and Liu (1999), who note that in this special case it reduces to a standard WLS solution (see Section 2).

2.3.4 Resampling Methods

Resampling methods can, in principle, be used to analyse multiple response data in either disaggregate, expanded or aggregated form. Resampling methods consist of repeatedly drawing subsamples from the sample data in order to simulate the distribution of statistics of interests, and/or to obtain estimates of the variances of non-linear parameters. Resampling is a computer-intensive approach which has now become viable as a result of the computing speed currently available. Clearly, access to multiple response data in incidence matrix form is required.

Loughin and Scherer's (1998) approach is an example of resampling applied to the analysis of aggregate MRD. As reported earlier, they used a variant of resampling called bootstrapping to obtain the finite sampling distribution of their modified chi-squared statistic. Westfall and Young (1993) designed multiple testing procedures based on a variety of resampling methods, and offered various applications involving multiple response data.

In concept, resampling methods constitute a powerful approach for solving difficult estimation and testing problem. It appears that theoretical subtlety is replaced by programming and computer power. There are, however, hidden complications. Naive bootstrapping does not always work; also, with cross-classified multiple response data, the design of resampling methods is not straightforward, since there exists a variety of modelling approaches to testing the same hypothesis (Agresti and Liu, 1999), each with different impacts on test performance. Thus careful and complex design of resampling procedures is required.

2.4 Summary and Conclusions

This review of the methods proposed to date has been categorized in terms of the level of data aggregation involved. It could alternatively have been categorized as descriptive as opposed to inferential, since methods for analysing disaggregate data are primarily descriptive in nature, though attempts have been made to incorporate inference properties in some. For the researcher/practitioner who wants to analyse multiple response data in disaggregate form, a wide variety of methods are available. These methods are based on multivariate approaches and are somewhat complicated to apply and interpret.

One of the striking conclusions from the review of the methods for aggregate data is that, although many approaches have been suggested, the treatment of conceptually simple tests of hypothesis remains a constant challenge for both survey and experiment-based research. Although Umesh (1995) with his pseudo chi-squared test and lambda measures, Loughin and Scherer (1998) with their modified chi-squared test and resampling-based testing procedure, and Agresti and Liu (1999) with their model-based testing approach and naive chi-squared test, have made significant contributions to the analysis of aggregate data, significant problems remain. All these procedures suffer from one deficiency or another, and extension to other practical problems are difficult and require methodologies that are unfamiliar to most practitioners in the marketing field.

In this thesis therefore, theoretical efforts will be focused on developing easily applied tests for commonly encountered hypotheses on aggregate multiple response data in one- and two-way tables.

Chapter 3.0

Possible Approaches to the Analysis of Multiple Response Data

In this chapter, the statistical issues posed by multiple response data are examined and the parallel between multiple response data analysis and the analysis of complex survey is identified and discussed.

3.1 Assumptions on Which The Pearson Chi-Squared Statistic Is Based

The classical Pearson test of association for tabulated data has a large sample chi-squared distribution only when certain conditions are satisfied, namely:

- a) the data consist of one or more simple random samples of n independent observations (or subjects) drawn from one or more large populations;
- b) the observations (or subjects) can be classified into categories that are mutually exclusive and collectively exhaustive.

Under these conditions, the cell counts of the cross-tabulation follow a multinomial distribution.

3.2 Failure of the Pearson Statistic with Multiple Response Data

In a cross-classification involving multiple response variables, each subject can contribute one or (usually) more than one count to each cell. Thus the categories, or cells, will not be mutually exclusive, i.e., assumption (b) above is violated. Further, since the choices of one subject will tend to be more similar, one to another, than to responses from another subject, there will be a "within subject" correlation, which will induce a correlation between the cells that is different from the correlation implied by the classical multinomial assumption. Thus the multinomial distribution will not be the correct distribution for the observed multiple response counts aggregated over subjects.

A similar violation of the assumptions upon which a valid Pearson chi-squared statistic is predicated occurs when the data come from statistical survey designs involving clustering. Units within clusters are more similar to one another than to units in different clusters, inducing within cluster correlation, for which appropriate corrections to the Pearson statistic have to be made (Rao and Scott, 1981, 1984; Rao and Thomas, 1988).

3.3 The Complex Nature of Multiple Response Data

Complex data (see Thomas, 1993) arise in many fields of application. They are a natural outcome of complex survey designs, and are quite common in research. Agresti (1990) observed that most data sets collected in the social sciences have sampling designs more complex than simple random sampling and that such designs commonly employ stratification and/or clustering. It is also true, however, that social scientists generally ignore the impact of survey design on their statistical analysis.

Data may also become complex as a result of apparently innocuous designs. For example, repeated measurements of individuals over time as in longitudinal studies render a data set complex. Longitudinal data sets that arise often in social studies and in marketing studies based on panel data and scanner panel data are another type of complex data. Complex data arise also from different experimental setups. Thomas (1993) mentioned an example by Rosner (1982) involving measurement of both subject's eyes; similarly for measurements on both of each subject's ears (Coren and Hakstian, 1990).

With multiple response data, each subject can be thought of as a cluster, so that the counts in a cross-tabulation involving multiple response variables represent aggregated cluster counts, which cannot be described by a multinomial distribution. It can be seen, therefore, that the analysis of multiple response data should have much in common with the analysis of complex survey data, a topic that has received considerable study in the statistical literature over the last 15 years. The effect of multiple response data of the "pick k/r" type is to form clusters of equal size from a finite and well defined universe. By the same token, the effect of multiple response data of the "pick any/r" type is to form clusters of unequal size from a predefined or a known universe.

Whenever the data can be regarded as complex, it is now recognized in the survey literature that the standard techniques of analysis found in most statistical packages cannot be applied, since classical statistical inference techniques were developed under the simple random sampling assumption. The violation of the assumption of independence of basic units of analysis, which is a feature of clustering, implies that variances estimated by standard procedures are erroneous and statistical tests do not

maintain their nominal error rates (see Rao and Thomas, 1987; Thomas, 1993). Given the parallel between multiple response data and complex surveys, the techniques developed for the latter case will be examined in a search for a convenient way of analysing multiple response data sets.

3.4 Statistical Methods for Categorical Data from Complex Surveys

Rao and Thomas (1988) and Thomas (1993) identified three main families of procedures that have been applied in hypothesis testing on cross-classified categorical data drawn from complex sample surveys. The earliest family of tests to be developed was based on the weighted least squares approach, with tests provided by Wald statistics (Koch, Freeman and Freeman, 1975; Binder, 1983). The second family is represented by the Rao-Scott first and second-order corrections to the classical X^2 (Pearson) and G^2 (log-likelihood ratio tests) (see Rao and Scott, 1981, 1984, 1987). The third family is based on the principle of resampling, in particular, the methods developed by Fay (1979, 1982, 1985) for jackknifing the classical X^2 and G^2 tests. Jackknifing involves the creation of pseudo replicates of the test statistic through repetitive application of the “leave one out” principle. Since one of the goals of this research is to find methods that are familiar and easy to apply, the Fay approach (like Loughin and Scherer’s bootstrapping approach) will not be pursued further. One common thread among all these streams of research is that they all rely on some common method of variance and covariance estimation that accounts for the complexity of the data.

3.4.1 The Wald Weighted Least-Squares Approach

The weighted least squares approach is based on the direct estimation of the covariance matrix of the cell observations and its subsequent use as a weighting matrix in the estimation of the parameters of a non-linear regression model, and on the testing (via Wald statistics) of the model parameters and overall model fit. The difference between the complex survey applications of WLS and that described earlier in Section 2.2.1 is that the weighting matrix is not of multinomial form, and must be estimated according to the survey design. As example, consider a simple test of goodness of fit for k response proportions, $p_1, \dots, p_k$ written as a vector $\mathbf{p}$. In this simple case, there are no parameters to estimate, i.e., the only test involved is a goodness-of-fit hypothesis test, $H_0: \mathbf{p} = \mathbf{p}_0$, where $\mathbf{p}_0$ is a hypothesized vector (e.g., the equiprobable vector with elements given by $p_{0j} = 1/k$).

For the sake of the example, we will consider only single response data, in which case the k proportions will sum to one. The extension to multiple response data will be considered later. Denote the covariance matrix of the vector of the estimated proportions $\hat{\mathbf{p}}$, by $\mathbf{V}(\hat{\mathbf{p}})$, with estimate $\hat{\mathbf{V}}(\hat{\mathbf{p}})$. In the simple case of simple random sampling, the elements p_i will be estimated by $\hat{p}_i = n_i/n$, where n_i are the category counts of ultimate units and $n = \sum n_i$ is the total sample size. For complex surveys, the estimates $\hat{p}_i$ will be based on weighted-up cell counts. The Wald statistic for testing H_0 is given by the quadratic form

$$X_w = (\hat{\mathbf{p}} - \mathbf{p}_0)' [\hat{\mathbf{V}}(\hat{\mathbf{p}})]^{-1} (\hat{\mathbf{p}} - \mathbf{p}_0) \quad (3.1)$$

where $\hat{V}^-$ denotes any generalized inverse. $\hat{V}$ is singular because the elements of both $\hat{p}$ and p_0 sum to 1 in the classical case, i.e. $\mathbf{1}'(\hat{p} - p) = 0$. The statistic X_w^2 is invariant to the choice of generalized inverse (Rao, 1973, page 26). Rao and Scott avoided the use of generalized inverses by dropping the last element of $\hat{p}$ and p_0 , and by dropping the corresponding row and column of $\hat{V}(\hat{p})$. This statistic is distributed asymptotically as χ_{k-1}^2 , a central chi-square distribution with $k - 1$ degrees of freedom, under H_0 . The null hypothesis is rejected when $X_w^2 > \chi_{k-1}^2(\alpha)$, $\chi_{k-1}^2(\alpha)$ being the upper percent point of χ_{k-1}^2 at level α . In this treatment, the covariance matrix is quite general and can accommodate any complex design for which $\hat{V}(\hat{p})$ can be obtained. The method is not restricted to multinomially distributed counts. It should be noted that though Wald-WLS methods are very general in theory, they do have several drawbacks in practice. As showed by Thomas and Rao (1987) and Thomas, Roberts and Singh (1996), large number of clusters are required before Wald tests achieve their large sample theoretical test levels. For small and moderate numbers of clusters they can yield dangerously liberal tests.

3.4.2 Rao-Scott Correction Methods

Adjustment methods are based on the idea that if “standard” methods are applied to complex cross-classified data, the resulting test statistic can be adjusted or corrected by dividing it by a factor calculated from the data. This factor, usually referred to as a design effect (Kish, 1965), is selected to make the adjusted statistic approximate a chi-square distribution.

Various competing adjustments methods have been proposed in the statistical literature to take account of the within-cluster correlations. Some authors have used statistical models such as the beta-binomial (Crowder, 1978; Paul, 1982), and the Dirichlet-multinomial model (Brier, 1980; Koehler and Wilson, 1986) as a basis for deriving their test adjustment parameter. Others used the correlated-binomial model (Kupper and Haserman, 1978), the multiplicative-binomial model (Altham, 1978), and the correlated probit regression model (Ochi and Prentice, 1984).

A common limitation of the previous models is that they assume a specific model for the intracluster correlations (Rao and Scott, 1992). It may be difficult to decide a priori what distributions one's observations follow. Thus a big advantage of the Rao-Scott approach to test adjustment is that it makes no assumptions about distributions.

3.4.2.1 Rao-Scott First-Order Corrections

The Rao-Scott approach will also be introduced and illustrated using tests for single response data. Their approach assumes only central limit theorem for the estimated proportions of the table in question, either a one-way table for simple goodness-of-fit tests or a multi-way cross-classified table for the case of loglinear modelling. In other words, they assume

$$n^{1/2}(\hat{\boldsymbol{p}} - \boldsymbol{p}) \xrightarrow{D} N(\mathbf{0}, V(\hat{\boldsymbol{p}})) \quad (3.2)$$

where $\xrightarrow{D}$ denotes convergence in distribution as the sample size $n \rightarrow \infty$, and $\hat{\boldsymbol{p}}$ represents

a vector of cell proportions. Loosely speaking, this means that for large samples, the vector of estimated proportions, $\hat{\mathbf{p}}$, has an approximate multivariate normal distribution with mean $\mathbf{P}$ and covariance matrix $n^{-1/2}\mathbf{V}(\hat{\mathbf{p}})$. Here, $\mathbf{V}(\hat{\mathbf{p}})$ will depend on the design and can be estimated by the standard variance estimation methods of survey sampling (Wolter, 1985). The simple goodness of fit “chi-squared” test for simple response data will again be used to illustrate the application of the Rao-Scott approach. Consider a vector of proportions, $\hat{\mathbf{p}}$, whose elements consist of a set of multinomial-like proportions computed from the weighted up counts obtained from a complex sample survey. To test the null hypothesis that this vector of proportions equals some null vector, $\mathbf{p}_0$, against the alternative hypothesis that they are not equal, a test statistic that mimics the classical Pearson “chi-squared” test is

$$X^2 = n \sum_{j=1}^k (\hat{p}_j - p_{0j})^2 / p_{0j} = (\hat{\mathbf{p}} - \mathbf{p}_0) \mathbf{D}_0^{-1} (\hat{\mathbf{p}} - \mathbf{p}_0) \quad (3.3)$$

where $\hat{p}_j$ is the estimated proportion of sample units in category j , and $\mathbf{D}_0 = \text{diag}(\mathbf{p}_0)$ is a diagonal matrix with elements p_{0j} , $j = 1, \dots, k$.

Rao and Scott (1981, 1984) showed that X^2 is distributed asymptotically, under the null hypothesis, as a weighted sum $\delta_1 W_1 + \delta_2 W_2 + \dots + \delta_{k-1} W_{k-1}$ of $k-1$ independent chi-squared random variables, denoted W_i , $i = 1, \dots, k-1$. The weights δ_i are the non-zero eigenvalues of a “design effect” matrix given by $\mathbf{D}_0^{-1} \mathbf{V}$. In practice, $\mathbf{V}$ is replaced by $\hat{\mathbf{V}}$, the covariance matrix of $\hat{\mathbf{p}}$ where the estimation is done in a fashion consistent with the complex design (Wolter, 1985). The weights δ_i (and their estimates $\hat{\delta}_i$) are essentially

eigenvalues of a design effect matrix which measures the inflation (or deflation) of the true variances and covariances, $\mathbf{V}$, relative to those expected under the assumption of a multinomial distribution. When this approach is applied to multiple response data, the corresponding δ 's will capture the effects of the multiple responses.

Rao and Scott (1981) defined their design effect matrix in non-singular form as $[\mathbf{P}_0^{(T)}]^{-1} \mathbf{V}^T$ where the superscript T denotes trimmed covariance matrices, i.e. covariance matrices formed by trimming one row and one column, corresponding to the omitted element of the probability vector, $\mathbf{p}$. The matrix $\mathbf{P}$ denotes the multinomial variance matrix, given by $\mathbf{P}_0 = \text{diag}(\mathbf{P}) - \mathbf{P}\mathbf{P}'$. The eigenvalues of Rao and Scott's trimmed design effect matrix correspond exactly to the non-zero eigenvalues of $\mathbf{D}_0^{-1} \mathbf{V}$.

Rao and Scott (1981, 1984) proposed two easily applied methods based on the generalized design effects. The simplest, referred to as the first-order Rao-Scott corrections, consists of replacing the standard chi-square statistic (3.3) by

$$X_C^2 = X^2 / \hat{\delta}, \quad (3.4)$$

where $\hat{\delta}$ is the mean of the estimated design effect weights. This correction makes the first moment (the mean or expected value) of X_C^2 equal to that of a χ^2 distribution on $k-1$ degrees of freedom. It can be shown that X_C^2 performs well provided the δ_i 's are not too dissimilar in magnitude. In many cases, the first-order correction can be computed given only partial information on the covariance matrix of $\mathbf{p}$.

3.4.2.2 Rao-Scott Second-Order Corrections

A better approximation to the asymptotic distribution of any X^2 -type statistic can be obtained when the full estimated covariance matrix $\hat{V}$ of the cell probabilities is known. The Rao-Scott second-order correction uses a corrected chi-squared statistic that has the same mean and the same variance as the asymptotic distribution of X^2 (the first-order correction equates only the mean).

The second-order correction factor is also based on the generalized design effects. It is obtained by replacing the modified chi-squared statistic (3.4) by another of the form

$$X_S^2 = X^2 / [\hat{\delta}(1 + \hat{a}^2)] = X_C^2 / (1 + \hat{a}^2) \quad (3.5)$$

where X_S^2 denotes the second-order corrected statistic, and $\hat{a}$, which measures the degree of variation among the generalized design effects, has the form of a coefficient of variation. It is given by

$$\hat{a}^2 = \sum_{i=1}^v \delta_i^2 / [(v\hat{\delta}^2) - 1] \quad (3.6)$$

where v represents the degrees of freedom of the basic X^2 test in question. Under the null hypothesis, this second-order test statistic can be regarded as a chi-square random variable on $v^* = v/(1 + \hat{a}^2)$ degrees of freedom. As an historical note, the subscript S on the second-order corrected statistic refers to Satterthwaite (1946), who was the first to suggest such an approximation for distributions of weighted sums of chi-squared

variables.

The second-order Rao-Scott correction performs better than the first-order when the δ 's differ appreciably in magnitude. When the δ 's are equal, the first-order corrections are asymptotically exact (Rao and Scott, 1981). Based on a simulated study of clustered data, Thomas and Rao (1987) recommended using the second-order procedure whenever the full estimated covariance matrix $\hat{V}$ is available.

Though the Rao-Scott approach has been illustrated here for a simple goodness-of-fit test, its general form can be specified (Rao and Scott, 1984) for a loglinear model of the type

$$\ln(p) = X\beta, \quad (3.7)$$

and also for more general models (Roberts, Rao and Kumar, 1987). Thus the Rao-Scott approach offers a fully general method for the analysis of complex data.

This approach to the analysis of complex survey data immediately suggests a method that can be applied to the analysis of multiple response data. Given that each subject in a cross-classified multiple response variables can be regarded as a cluster, the Rao-Scott approach should be directly applicable.

It is also worth noting that an approach based on the Rao-Scott methodology will also be extendable to the case where the subjects providing the multiple responses are themselves sampled via a complex survey design. This extension is far more difficult to carry out using the recent Agresti (1999) approach, and fails completely in the case of Loughin and Scherer's (1998) bootstrapping approach. This extension will be a topic for further research.

3.5 Estimating Covariance Matrices for Use With Weighted Least Squares And Rao-Scott Techniques

Both the Wald and Rao-Scott approaches rely on a valid estimation of the covariance matrix under the specific design. In general, for complex surveys, estimation of variance is a major statistical topic in and of itself. Extensive presentation and discussion of variance estimation techniques are considered by Cochran (1963) and Wolter (1985). The latter presents various approaches to the estimation of variance for both linear and nonlinear estimators. However for multiple response data, it is relatively easy to use the data for individual subjects to obtain $\hat{V}(\hat{\beta})$, under the assumption that the subjects themselves are drawn from a simple random sampling process. As noted in Chapter 2, when multiple response data are aggregated in the form of an expanded table, the corresponding cell frequencies follow a multinomial distribution. Since the aggregated tables of counts in this thesis are marginal to an expanded table, the required covariance matrix for multiple response data can also be obtained as matrix transforms of the multinomial covariance of the expanded table.

To perform a test on two-way tables and to do loglinear modelling in general, variances of non-linear functions of the proportions are needed, e.g., the estimated expected values in a two-way table which are given by $\hat{p}_{i.}\hat{p}_{.j}$. The variances and covariances of such nonlinear estimates are obtained by the method of Taylor linearization (for detailed proofs, see Appendix 4). Alternatively, the ubiquitous jackknife approach (introduced by Tukey, 1958) can also be used for variance estimation, where the pseudo replicates are obtained via the leave one out method (see also Appendix 6).

Chapter 4.0

Testing Hypotheses in One-Way

Multiple Response Tables: First-Order Rao-Scott Tests.

In this chapter, Rao-Scott tests for one way tables will be developed, namely tests of goodness-of-fit and of the equality of response probabilities. Tests for two-way tables will be discussed in Chapter 5.

4.1 Simple Goodness of Fit

The simple goodness of fit “chi-squared” test will be considered first as it provides the simplest illustration of the Rao-Scott approach to multiple response data.

For a multiple response variable having C items, consider the counts m_j , $j = 1, \dots, C$, obtained by aggregating over all n respondents as in Table 1.2. The one-way array of C counts can be represented as a vector $\mathbf{m}$, i.e.,

$\mathbf{m}' = (m_1, \dots, m_C)$, where the prime denotes a vector transpose. A corresponding vector $\mathbf{t}' = (t_1, \dots, t_C)$ can be defined, where the element $t_j = m_j/n$ represents a sample estimate of the “item response” probability, τ_j , that a respondent will provide a positive response to multiple response item j . Here τ_j can be defined in terms of expected counts, i.e., $\tau_j = E(m_j)/n$, when n represents the total number of respondents sampled. It is assumed that n is fixed in advance. Under this setup, a goodness of fit null hypothesis can be represented as

$$H_0(GOF): \boldsymbol{\tau} = \boldsymbol{\tau}_0, \quad (4.1)$$

where τ is the $C \times 1$ vector of item response probabilities and τ_0 is a $C \times 1$ vector of hypothesized probabilities typically obtained from secondary data. Rao and Scott (1981) dealt with multinomial-like vectors of probabilities for which the individual probabilities summed to one. This constraint does not apply to the vector of item probabilities τ or to its estimates, $\hat{\tau}$, because of the multiple counting that occurs with multiple response variables. A conventional chi-squared statistic for testing $H_0(GOF)$ is given by

$$X^2(GOF) = n \sum_{j=1}^C \frac{(\hat{\tau}_j - \tau_{0j})^2}{\tau_{0j}} \quad (4.2)$$

which can be written in matrix form as

$$X^2(GOF) = (\hat{\tau} - \tau_0)' D_0^{-1} (\hat{\tau} - \tau_0), \quad (4.3)$$

where D_0 is a $C \times C$ diagonal matrix with diagonal elements τ_{0i} , $i = 1, \dots, C$. Under the assumption that $\hat{\tau}$ has an asymptotic, i.e., large sample, multivariate normal distribution with mean τ_0 , and covariance matrix V/n , it can be shown (following Rao and Scott, 1981) that X^2 is distributed asymptotically as a weighted sum of C independent χ_1^2 random variables, denoted W_i , $i = 1, \dots, C$, i.e., that $X^2 \sim \delta_1 W_1 + \delta_2 W_2 + \dots + \delta_C W_C$. The weights δ_i are eigenvalues of the matrix $D_0^{-1} V$. A proof of this result is provided in Appendix 1, Case 1. Estimates of the δ_i , denoted $\hat{\delta}_i$, are eigenvalues of $D_0^{-1} \hat{V}$, where $\hat{V}$ is an estimate of the asymptotic covariance matrix, V .

In this case, the first order Rao-Scott correction consists of dividing $X^2(GOF)$ by

$\hat{\delta}(GOF)$, the mean of the estimated weights $\hat{\delta}_1, \dots, \hat{\delta}_C$. Rao and Scott (1981) showed that, after the correction, their approximate first-order statistic has the same mean as a chi-squared random variable on the appropriate number of degrees of freedom, which is C in the present case. The required correction factor, $\hat{\delta}(GOF)$ can be expressed as

$$\hat{\delta}(GOF) = \text{trace}(\mathbf{D}_0^{-1} \hat{\mathbf{V}})/C. \quad (4.4)$$

Since the trace of a matrix is the sum of its diagonal elements, only the diagonal elements of $\hat{\mathbf{V}}$ need be estimated. These can be obtained directly from the multiple response counts themselves, and are given by $\mathbf{t}_j(1 - \mathbf{t}_j)$, since the multiple response counts can be regarded as correlated binomial variates. The diagonal elements of $\mathbf{D}_0^{-1} \hat{\mathbf{V}}$ become $\mathbf{t}_j(1 - \mathbf{t}_j)/\tau_j = 1 - \tau_j$, $j = 1, \dots, C$, so that from equation (4)

$$\hat{\delta}(GOF) = \sum_{j=1}^C (1 - \tau_j)/C = 1 - \sum_{j=1}^C \tau_j/C = 1 - \hat{\gamma}/C, \quad (4.5)$$

where the term $\hat{\gamma} = \sum_j \tau_j$ represents a multiple counting index. In terms of cell counts, the multiple counting index can be written as $\gamma = m_{\cdot}/n$, where $m_{\cdot}$ is the sum of the C multiple response counts in the one-way array. For a single response variable, γ will equal 1 while for a multiple response factor it can take on a maximum value of $\gamma = C$, corresponding to a positive response to all multiple response items. In practice, γ will be somewhere between these extremes. The first-order Rao-Scott corrected chi-squared test, denoted $X^2_C(GOF)$, then consists of referring the corrected statistic

$$X_C^2(GOF) = X^2(GOF) / \hat{\delta}(GOF) = X^2(GOF) / (1 - \hat{\gamma}/C), \quad (4.6)$$

to a chi-squared distribution on C degrees of freedom. The corresponding single response version of this test (Rao and Scott, 1981) has C - 1 degrees of freedom, due to the restriction that single response probabilities sum to one. One advantage of the Rao-Scott first-order test over competing procedures is that it requires only a knowledge of the aggregated multiple response counts (together with the sample size), a useful property when access to the full data set is not available.

Example 4.1. With the sample sizes of males and females in Table 1.3 considered fixed, vectors of estimated response probabilities can be obtained by dividing each row of Table 1.3 by the sample size, giving $\mathbf{t}'_M = (0.679, 0.606, 0.639)$, and $\mathbf{t}'_F = (0.495, 0.728, 0.549)$, for the males and females, respectively. For purposes of illustration only, we will treat the probabilities for males as known, fixed quantities, i.e., as if they were obtained from census data. We will then test the goodness of fit null hypothesis that the vector of probabilities for females is equal to the hypothesized (male) value, i.e., $H_0: \boldsymbol{\tau}_F = \boldsymbol{\tau}_0$, where the hypothesized vector is set equal to $\mathbf{t}_M$. With a sample size of 184, the value of the Pearson chi-squared test, $X^2(GOF)$, of equation (2) is calculated to be 16.2. Using estimated item probabilities $\mathbf{t}_F$, the multiple counting index can be calculated as $\gamma = (0.495 + 0.728 + 0.549) = 1.772$. Alternatively, using counts, it can be calculated as $(91 + 134 + 101)/400$, with the same result. The corrected chi-squared test then becomes $X_C^2(GOF) = 16.2 / (1 - 0.591) = 39.6$. This is compared

to the 5% critical value of a chi-squared random variable on three degrees of freedom, which is 7.815. The null hypothesis can therefore be rejected, i.e., it can be concluded that the response probabilities for females are significantly different from the hypothesized (male) probabilities. Note that in this case, the uncorrected test, ($X^2(GOF) = 16.2$) would also have lead to rejection. It is clear, however, that the uncorrected test is very conservative, yielding a test value that is approximately one-half of the correct value.

4.2 Equality of Response Probabilities

For a one-way array of C multiple response counts, consider the null hypothesis that all response probabilities are equal, or equivalently, that all expected counts are equal. Formally, this null hypothesis can be expressed as

$$H_0(E): \tau_1 = \dots = \tau_C = \tau, \quad (4.7)$$

where τ is the common, but unknown, scalar response frequency. Under $H_0(E)$, τ can be estimated by $\hat{\tau} = \sum_{j=1}^C \hat{\tau}_j / C$, where $\hat{\tau}_j = m_j / n$, and a conventional “chi-squared” statistic for testing H_0 can be constructed, namely

$$X^2(E) = n \sum_{j=1}^C \frac{(\hat{\tau}_j - \hat{\tau})^2}{\hat{\tau}}. \quad (4.8)$$

An alternative expression of the same test can be written in matrix form as

$$X^2(E) = (n/\mathbf{t})(\mathbf{t} - \mathbf{t}\mathbf{1})(\mathbf{t} - \mathbf{t}\mathbf{1})' \quad (4.9)$$

or as

$$X^2(E) = (n/\mathbf{t})\mathbf{t}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/C)\mathbf{t} \quad (4.10)$$

where $\mathbf{t} = (t_1, \dots, t_C)'$. Under the assumption that $\mathbf{t}$ has an asymptotic, i.e., large sample, multivariate normal distribution with mean $\mathbf{t}\mathbf{1}$ and covariance matrix $\mathbf{V}/n$, it is shown in Appendix 1 (case 2) that X^2 is distributed asymptotically as a weighted sum of $C - 1$ independent χ^2_1 random variables, denoted W_i , $i = 1, \dots, C$, i.e., that $X^2 \sim \delta_1 W_1 + \delta_2 W_2 + \dots + \delta_{C-1} W_{C-1}$. The weights δ_i are the non-zero eigenvalues of the matrix $(1/\tau)\mathbf{J}\mathbf{V}$, where $\mathbf{J} = \mathbf{I} - \mathbf{1}\mathbf{1}'/C$. Estimates of the δ_i , denoted $\hat{\delta}_i$, are eigenvalues of $(1/\tau)\mathbf{J}\hat{\mathbf{V}}$ where $\hat{\mathbf{V}}$ is an estimate of the asymptotic covariance matrix, $\mathbf{V}$.

In this case, the first order Rao-Scott adjustment consists of dividing $X^2(E)$ by $\hat{\hat{\delta}}(E)$ the mean of the weights $\hat{\delta}_1, \dots, \hat{\delta}_{C-1}$. After the adjustment, this approximate first-order statistic has the same mean as a chi-squared random variable on $C - 1$ degrees of freedom. The required adjustment factor, $\hat{\hat{\delta}}(E)$ can be expressed as

$$\hat{\hat{\delta}}(E) = \text{trace}[(1/\tau)\mathbf{J}\hat{\mathbf{V}}]/(C - 1) \quad (4.11)$$

In this case (see appendix 2), the first-order Rao-Scott correction factor becomes

$$\hat{\delta}(E) = (1 - \gamma/C)(1 - \bar{\rho}), \quad (4.12)$$

where $\gamma = m/n$, as before, and $\bar{\rho}$ is the average of the $C(C-1)/2$ correlations among the counts, m_j , $j = 1, \dots, C$. Each ij correlation can be routinely estimated using the incidence matrix entries for items i and j . Though easy to calculate, they do require access to the full data set. Alternatively, they could be calculated given the multinomial frequencies in the expanded table.

The first-order Rao-Scott corrected test then consists of comparing the corrected statistic $X_C^2(E) = X^2(E)/\hat{\delta}(E)$ to a χ^2 distribution on $C - 1$ degrees of freedom.

When the full data set is not available, it is possible to derive a conservative version of the equality of item probabilities test by noting that given a vector of item probabilities, τ_j , the correlations among the multiple responses are not free to vary from -1 to 1 (Emrich and Piedmonte, 1991; Lee, 1993; Lipsitz and Fitzmaurice, 1994); they are bounded below by

$$\max\{ -[(\tau_i \tau_j)/(1 - \tau_i)(1 - \tau_j)]^{1/2}, \{ -[(1 - \tau_i)(1 - \tau_j)/(\tau_i \tau_j)]^{1/2} \} \quad (4.13)$$

and above by

$$\min\{ [(\tau_i(1 - \tau_j)/(\tau_j)(1 - \tau_i)]^{1/2}, \{ [(\tau_j)(1 - \tau_i)/(\tau_i(1 - \tau_j)]^{1/2} \}. \quad (4.14)$$

Using the matrix of lower bounds correlations, an upper bound adjustment factor can be expressed as

$$\hat{\delta}(E_U) = (1 - \gamma/C)(1 - \hat{\rho}_L), \quad (4.15)$$

where $\hat{\rho}_L$ is the average of the $C(C-1)/2$ lower bounds correlations among the counts, m_i , $i = 1, \dots, C$. Likewise, using the matrix of upper bounds correlations, a lower bound adjustment factor can be expressed as

$$\hat{\delta}(E_L) = (1 - \gamma/C)(1 - \hat{\rho}_U), \quad (4.16)$$

where $\hat{\rho}_U$ is the average of the $C(C-1)/2$ upper bounds correlations among the counts, m_i , $i = 1, \dots, C$. If the test based on the upper bound adjustment factor leads to rejection of the null hypothesis, then a test based on the complete information would also lead to the rejection of the null hypothesis. Likewise, if the test based on the lower bound adjustment factor fails to reject the null hypothesis, the test based on the use of the complete data set would also fail to reject the null hypothesis. The advantage of these versions of the corrected test is that the lower and upper bounds can be obtained directly from the aggregate counts.

Letting $C = 2$, equation (4.8) becomes

$$X^2(E) = n \sum_{j=1}^2 \frac{(t_1 - t_2)^2}{t_1 + t_2} \quad (4.17)$$

and the conservative version of the Rao-Scott first-order correction factor turns out to be

$$\hat{\delta}(E_U) = (1 - \bar{t})/\bar{t} \quad (4.18)$$

where $\bar{t} = (t_1 + t_2)/2$. Therefore, the conservative version of the Rao-Scott first -order statistic can be written as

$$X^2(E_U) = n \sum_{j=1}^2 \frac{(t_1 - t_2)^2}{2(1 - \bar{t})} \quad (4.19)$$

which is equivalent to the square of the Wild and Seber (1993, 1994) conservative normal test for comparing two multiple response proportions from the same sample given by

$$Z_L = (t_1 - t_2) / \sqrt{\{\min(t_1 + t_2), [(1 - t_1) + (1 - t_2)]/n\}} \quad (4.20)$$

Example 4.2. Consider the vector of probabilities calculated from Table 1.3 for the female category, i.e $\boldsymbol{\tau}'_F = (0.495, 0.728, 0.549)$. For the purpose of illustration, we will test that the item response probabilities for females are all equal. With a sample size of 184, the value of the Pearson chi-squared test, $X^2(E)$, is calculated as 9.32. Using estimated probabilities, the multiple counting index can be calculated as $\gamma = (0.495 + 0.728 + 0.549) = 1.772$. The average of the $C(C-1)/2$ correlations among the counts, $\bar{\rho}$, is calculated to be -0.040342. The corrected chi-squared statistic then

becomes $\chi^2_C(E) = 9.32/(1 - 0.591)(1 + 0.0403) = 21.89$. On comparing $\chi^2_C(E) = 21.89$ to the 5% critical value of a chi-squared random variable on $C - 1$ degrees of freedom, which is 5.991, the null hypothesis can be rejected. In other words, it can be concluded that individual item response probabilities for females are significantly different, i.e., female respondents do exhibit distinct preferences for the automobile classes presented in the survey question.

The conservative version of the Rao-Scott first-order corrected chi-squared statistic was found to be $\chi^2_C(E_{ij}) = 9.32/(1 - 0.591)(1 + 0.69562) = 13.44$. This test also leads to the rejection of the null hypothesis. However, the problem with a conservative test is that if it is not significant, there is no way to know whether the true value of the test statistic is significant or not.

Chapter 5.0

Testing for Association in Two-Way Multiple Response Tables: First-Order Rao-Scott Tests.

In this chapter, Rao-Scott tests for two-way tables will be derived, namely homogeneity and independence tests. The different sampling setups leading to homogeneity and independence tests will first be discussed for single response variables, so that the sampling issues can be clearly separated from the other complexities introduced by multiple response. Based on the sample design, two types of distributions will be considered.

5.1 Sample Design for Tests on Two-Way Tables of Single Response Counts

5.1.1 Independent Samples

A researcher might draw a sample from two or more independent populations and estimate the proportions of each population that fall within each category of a single response variable. In this case, each of the two samples is an independent random sample from a distinct population. For notational convenience, the discussion in this section will feature two independent samples though all the ideas generalize to R independent samples from R distinct populations. If the data are displayed in a two-way table, with rows representing the individual independent samples, then the total number of subjects in each row will be fixed by the design, i.e., by the prior decision of how many subjects should be drawn from each population. For a single response variable, the relevant distribution is the product-binomial (for two categories) or product-multinomial

distribution (for three or more categories). The appropriate association hypothesis is that the category probabilities for each population are the same, i.e.,

$p_1 = p_2 = \dots = p_R = p$, where p_j is the $C \times 1$ vector of category probabilities in the j^{th} of the R independent populations, and p is an unknown common vector of probabilities. This is referred to as the homogeneity hypothesis.

As an example of the classical (single response) homogeneity case, consider a hypothetical research design in which separate samples of 200 male drivers and 200 female drivers were randomly and independently selected, and the country of origin of their favourite cars was registered. The following cross-tabulation would be typical of such a design.

Table 5.1
Relationship Between Country of Origin and Gender
in Choosing a Car (Single Response-Homogeneity)

Country of Origin				
Gender	Domestic	Japanese	European	Total
Male	74	35	91	200
Female	76	65	59	200
Total	150	100	150	400

Note that we can regard this table as a cross-tabulation of a single response variable having fixed (i.e., non-random) row marginal totals with a single response variable having non-fixed (i.e., random) column marginal totals. We will denote the

single-response cell frequencies in this single $\times$ single table as n_{ij} , $i = 1, 2$, $j = 1, \dots, C$.

The relevant question is whether the distributions of car origins in the two populations are the same or not. In this case, the association null hypothesis can be expressed as the homogeneity of a C category variable (car origin) across two populations, given by

$$H_0: \quad \mathbf{p}_1 = \mathbf{p}_2 = \mathbf{p} \quad (5.1)$$

where $\mathbf{p}_i = (p_{i1}, \dots, p_{iC})'$ denotes the population vector of row proportions p_{ij} in population i , $i = 1, 2$, and $\mathbf{p} = (p_1, \dots, p_C)$ denotes the unknown common vector of proportions under H_0 . The expected cell frequencies under the null hypothesis may be written as $n_{i.} p_j$, which must be estimated from the data. Here $n_{i.}$ are the fixed row totals given by $n_{i.} = \sum_j n_{ij}$. A consistent and unbiased estimator of the probability (or population proportion) p_{ij} in each cell is given by $\hat{p}_{ij} = n_{ij} / n_{i.}$. A Pearson test statistic for the homogeneity hypothesis in the case of $i = 2$ populations is given by

$$X^2(H_{SS}) = \sum_{i=1}^2 n_{i.} \sum_{j=1}^C \frac{(\hat{p}_{ij} - \hat{p}_j)^2}{\hat{p}_j} \quad (5.2)$$

where $\hat{p}_j = (n_{1.} \hat{p}_{1j} + n_{2.} \hat{p}_{2j}) / n_{..}$, $n_{..} = \sum_{i=1}^2 n_{i.}$ is the total number of subjects, and the subscript SS denotes the single response by single response classification. Alternatively $\hat{p}_j$ can be expressed in terms of marginal counts, in which case

$$X^2(H_{SS}) = \sum \sum (n_{ij} - \hat{E}(n_{ij}))^2 / \hat{E}(n_{ij}) \quad (5.3)$$

where $\hat{E}(n_{ij}) = n_{i.} n_{.j} / n_{..}$, where $n_{.j}$ denotes the (random) column marginal totals given

by $\sum_i n_{ij}$. The estimated probabilities corresponding to Table 5.1 under the assumption of two independent multinomial (or product-multinomial) sampling distributions are given in Table 5.2.

Table 5.2
Probability Table Under Product-Multinomial Assumption

Country of origin				
Gender	Domestic	Japanese	European	Total
Male	0.37	0.175	0.455	1
Female	0.38	0.325	0.295	1

Here it is evident that the row probabilities each sum to one.

5.1.2 A Single Random Sample (Single Response $\times$ Single Response)

A second sampling design might require the selection of one random sample of subjects of size n ($n = n_{..} = \sum_i \sum_j n_{ij}$), with observations cross-classified according to two criteria. Here we again consider the row and column variables to be single response. This gives rise to a classical contingency table in which only the grand total of subjects ($n = n_{..}$) is fixed by design. This setup differs from the independent samples in that the row marginal totals as well as the column marginal totals are random. The joint distribution of the two variables now becomes relevant, since the rows of the tables are no longer independent. The appropriate hypothesis is now that of independence (as opposed

to homogeneity), and the relevant distribution is the multinomial distribution for the counts $n_{11}, \dots, n_{1C}, n_{21}, \dots, n_{2C}$. Note that since we have assumed that both row and column variables are single response then the classic Pearson test of independence (or association) is the relevant test.

As an example, consider a hypothetical research design in which 400 drivers were randomly selected and the country of origin of the cars, as well as the gender of the driver was registered. The following cross-tabulation would be typical of such a design.

Table 5.3
Relationship Between Country of Origin and Gender
in Choosing a Car (Single Response-Independence)

Country of Origin				
Gender	Domestic	Japanese	European	Total
Male	90	35	91	216
Female	60	65	59	184
Total	150	100	150	400

Now the relevant question becomes whether the two variables are associated or not. For single response variables, this hypothesis can be formulated in the following way:

$$H_0 : p_{ij} = p_{i.} p_{.j} \quad i = 1, 2; j = 1, \dots, C., \quad (5.4)$$

The p_{ij} are unknown cell probabilities (or population proportions) and $p_{i.}$ and $p_{.j}$ are the corresponding row and column marginal proportions. The unknown cell proportions are consistently and unbiasedly estimated by the sample proportion $\hat{p}_{ij} = n_{ij}/n$, where $n = n_{..}$ is the fixed size of the sample. Corresponding estimators of the marginal proportions are denoted $\hat{p}_{i.}$ and $\hat{p}_{.j}$, where $\hat{p}_{i.} = \sum_j \hat{p}_{ij}$ and $\hat{p}_{.j} = \sum_i \hat{p}_{ij}$. Computing the estimated cell proportions under the single sample assumption gives rise to the following table, where all probabilities sum to one in this single response $\times$ single response case.

Table 5.4

Probability Table Under Single Sample Assumption

Country of Origin				
Gender	Domestic	Japanese	European	Row Marginal
Male	0.225	0.09	0.225	0.54
Female	0.15	0.16	0.15	0.46
Column Marginal	0.375	0.25	0.375	1

The Pearson chi-squared statistic for the test of independence can be written as

$$X^2(I_{SS}) = n \sum_{i=1}^r \sum_{j=1}^c \frac{(\hat{p}_{ij} - \hat{p}_{i.} \hat{p}_{.j})^2}{\hat{p}_{i.} \hat{p}_{.j}} \quad (5.5)$$

Equation (5.5) can be alternatively formulated as

$$X^2(I_{SS}) = \sum \sum (n_{ij} - \hat{E}(n_{ij}))^2 / \hat{E}(n_{ij}) \quad (5.6)$$

where $\hat{E}(n_{ij}) = n_{i.} n_{.j} / n$.

Thus it can be seen that for *single response variables*, the homogeneity test statistic (5.3) and the independence (or association) statistic (5.6) have equivalent algebraic forms.

5.1.3 Invariance to Distributional Assumptions

The classical theory for large samples of single response variables leads to an interesting result, namely:

The distribution of $X^2(SS)$ given by equation (5.2) or (5.3) under the assumption of independent samples is identical to the distribution of $X^2(SS)$, given by equations (5.5) or (5.6) under the assumption of a single sample.

In other words, for the classical case of large samples of single response variables, it does not matter whether the row totals $n_{1.}$ and $n_{2.}$ are fixed by design or not.

In general, this invariance property applies to two-way and multi-way tables under simple random sampling for hypotheses that can be represented via loglinear models, which include homogeneity and independence (Agresti, 1996). For large samples, independent sampling leads to a product-multinomial distribution for the cell counts of the full table. Single sampling leads to a full multinomial distribution. In both cases, the sampling distributions of the relevant test statistics (homogeneity and independence) are the same.

Another way of representing the invariance is to note that if we first assume a multinomial distribution for the whole table, and then work out the conditional distribution, conditional on fixed row totals, we end up with the distributional formula for a product multinomial distribution. In other words, the setup of Section 5.1.2 conditional on fixed row totals also leads to results of Section 5.1.1.

The relevance of this discussion to the multiple response case is that:

The general equivalence of sampling models and test procedures for single response variables cannot be assumed to extend to aggregate cross-tabulations of a single variable and a multiple response variable. Aggregate multiple response tables do not have multinomial distributions on which the invariance property depends.

In other words, for the cross-tabulations of a *single* by a *multiple* response variable, different analyses are required depending on whether the row totals,

$n_{i.}$, $i = 1, \dots, R$, are fixed or random.

5.2 Two-Way Tables Involving Multiple Response Data

5.2.1 Single by Multiple Response Tables: Fixed Row Totals.

In this section, an association test will be developed for two-way tables featuring a single response variable and a multiple response column variable where the number of subjects per row is fixed. First, let us review the notation that will be used for a *single* by *multiple* table with fixed row totals.

Table 5.5

Notation for A Single by Multiple Table: Umesh (1995) Example

Gender	Domestic	Japanese	European	Subjects
Male	m_{11}	m_{12}	m_{13}	n_1
Female	m_{21}	m_{22}	m_{23}	n_2
Total	$m_{.1}$	$m_{.2}$	$m_{.3}$	$n = n_1 + n_2$

In Table 5.5 n_i denotes the fixed number of subjects sampled from population i falling in row i . We denote multiple response counts by m_{ij} , in contrast to the n_{ij} notation used for single response tables. Further, in this case we will use the single notation n_i for the fixed number of subjects per row instead of $n_{i.}$ and the sum $\sum_i n_i$ will be denoted n , rather than $n_{..}$. Note that $n_i \neq \sum_j m_{ij} = m_{i.}$, i.e., there is multiple counting. The cell count m_{ij} denotes the number of subjects (out of n_i) in row i who respond positively to item j , and $m_{.j} = \sum_i m_{ij}$ = total number of subjects responding to item j (there is no double counting in the column totals). The item response probabilities for multiple response data will be

denoted τ_{ij} , $i = 1, \dots, R$; $j = 1, \dots, C$, where $\tau_{ij} = E(m_{ij})/n_i$, and R is the number of independent samples (i.e. rows). For each of the two populations, the vectors of response probabilities will be denoted $\tau_i = (\tau_{i1}, \dots, \tau_{iC})'$, $i = 1, \dots, R$.

As for the case of a single by single response table, the association test in this case is equivalent to a test of homogeneity of item response probabilities. The case of two independent samples will be considered first, based on the Umesh (1995) data of Table 1.3. For these data, the homogeneity hypothesis can be formally specified as

$$H_0(H_{SM}): \quad \tau_1 = \tau_2 = \tau, \quad (5.7)$$

where τ is the *unknown* common value of the response probabilities. Thus, under the null hypothesis, the counts in both rows of Table 1.3 are samples from distributions having common probability vector τ , so that an estimate, $\hat{\tau}$, of the common response probabilities is given by $(m_1 + m_2)/(n_1 + n_2)$, where m_1 and m_2 represent vectors of multiple response counts in the first and second rows of Table 1.3, respectively. The count vectors can be expressed in the form $m_i = n_i \tau_i$, $i = 1, 2$, so that the estimate of τ can be written as

$$\hat{\tau} = \theta_1 \hat{\tau}_1 + \theta_2 \hat{\tau}_2, \quad (5.8)$$

where $\theta_i = n_i/(n_1 + n_2) = n_i/n$, $i = 1, 2$, and $\theta_1 + \theta_2 = 1$. A conventional chi-squared

statistic for testing $H_0(H_{SM})$, which is identical to Umesh's (1995) pseudo chi-squared test, is then given by

$$X^2(H_{SM}) = n \sum_{i=1}^2 \theta_i \sum_{j=1}^C \frac{(\mathbf{t}_{ij} - \mathbf{t}_j)^2}{\mathbf{t}_j}, \quad (5.9)$$

where $\mathbf{t}_{ij}$ and $\mathbf{t}_j$ represent the j^{th} elements of the estimated probability vectors $\mathbf{t}_i$, $i=1, 2$, and $\mathbf{t}$, respectively. As in the goodness of fit case, a large sample Rao-Scott analysis can be used to develop an approximate first-order adjusted statistic whose large sample distribution has the same mean as a chi-squared random variable on C degrees of freedom. The sample estimate, $\hat{\delta}(H_{SM})$ of the theoretical adjustment factor, $\bar{\delta}(H_{SM})$ is again of the form $1 - \gamma/C$, where the multiple counting index γ is given by

$$\gamma = \sum_{j=1}^C (\theta_1 \mathbf{t}_{1j} + \theta_2 \mathbf{t}_{2j}) = m_{..}/n. \quad (510)$$

Here $m_{..} = \sum_i \sum_j m_{ij}$ is the total of the multiple response counts in the two rows of the table, and n is the total number of subjects. A detailed proof of this result is given in Appendix 3. The adjusted test then consists of referring $X^2(H_{SM})$ to the $100\alpha\%$ critical point of a chi-squared distribution on C degrees of freedom.

This adjusted test can be extended (Appendix 3) to compare more than two row vectors of probabilities. For the case of R rows, the correction factor is again of the form $\hat{\delta}(H_{SM}) = 1 - \gamma/C$, with γ now given by

$$\gamma = \sum_{j=1}^C (\theta_1 t_{1j} + \dots + \theta_R t_{Rj}) = m_{..}/n \quad (5.11)$$

Here $\theta_i = n_i / n$, $i = 1, \dots, R$, n_i is the number of subjects in the i^{th} of the R rows,

$n = \sum_{i=1}^R n_i$, and $m_{..} = \sum_{i=1}^R \sum_{j=1}^C m_{ij}$, the total of the multiple response counts in the $R \times C$ table.

Example 5.1 For the Umesh data of Table 1.3 the multiple counting index is calculated to be $\gamma = (237+265+239)/400 = 1.8525$. Umesh 's (1995) pseudo chi-squared statistic, which is algebraically identical to $X^2(H_{SM})$, has the value 9.06, so that the adjusted first-order Rao-Scott statistic becomes $X_C^2(H_{SM}) = 9.06/(1-.618) = 23.68$. When referred to the 5% critical point of a chi-squared on 3 degrees of freedom (7.815), this is clearly significant, i.e., automobile choice is associated with gender. The pseudo chi-squared test itself would also have lead to rejection in this case, though it is clearly very conservative. It is interesting to compare the results of the approximate Rao-Scott test with the loglinear model test also reported by Umesh (1995). In that case, he obtained a likelihood ratio chi-squared value of 21.9, again on 3 degrees of freedom, quite close to the value of the first-order adjusted Rao-Scott statistic.

5.2.2 Single $\times$ Multiple Response, Random Row Totals

Agresti and Liu (1999) analysed the Loughin and Scherer (1998) data under the assumption that the row totals were fixed. However, Loughin and Scherer (1998)

assumed that the number of subjects in each row was not fixed by design. They assumed that only the total number of subjects was fixed, i.e., their analysis dealt with a single random response variable cross-classified with a random multiple response variable. Thus, in this section, the first order Rao-Scott tests described in 5.2.1 will be extended to the case of random row variable totals, which is the more typical case encountered in practice.

When only the overall sample size is fixed by design, the analysis must be based on the individual cell probabilities τ_{ij} defined as $\tau_{ij} = E(m_{ij})/n$, with n being the fixed number of subjects sampled. The τ_{ij} represent the joint probability that a randomly selected subject will fall in row i of the table and will respond positively to item j of the multiple response. These probabilities are estimated by $t_{ij} = m_{ij}/n$ as illustrated in Table 5.6.

Table 5.6
Effect of Country of Origin and Gender on Car Selection
Random Margins (Multiple Response)

Gender	Domestic	Japanese	European	Row marginal
Male	0.365	0.327	0.345	0.54
Female	0.227	0.335	0.252	0.46
Row marginal	0.592	0.662	0.597	

We will represent this table of estimated cell probabilities using the $RC \times 1$ vector τ , given by

$$\mathbf{t} = (t_{11}, \dots, t_{1c}, t_{21}, \dots, t_{2c}) \quad (5.12)$$

with elements $t_{ij} = m_{ij}/n$. The (random) row marginal probabilities are now given by

$\tau_i = E(n_i)/n$, where n_i represent the number of individual subjects in row i , which is not equal to $\sum m_{i.}$. The column marginal probabilities are given by $\tau_{.j} = E(m_{.j})/n$.

Corresponding estimates are $\hat{t}_i = n_i/n$ and $\hat{t}_{.j} = m_{.j}/n$. Here the subscript on the marginal τ 's are chosen to match their definition in terms of expected counts.

Umesh (1995) and Loughin and Scherer (1998) showed that for tables of this type, the form of the classical test of association given by equation (5.6) (with n_{ij} replaced by m_{ij} , $n_{i.}$ and $n_{.j}$ replaced by $m_{i.}$ and $m_{.j}$) does not apply when the row variable is of multiple response type. Rather than using the "naive" classical expected values $\hat{E}(m_{ij}) = m_{i.}m_{.j}/m_{..}$, the appropriate form for the expected values is given by $\hat{E}(m_{ij}) = n_i m_{.j}/n$. To see why the classical expected values are incorrect, we must consider the form of the hypothesis of independence when one (column) variable is of multiple response type and the other (row) variable is of single response type. In this case, the appropriate independence hypothesis is

$$H_0(I_{SM}): \tau_{ij} = \tau_i \tau_{.j} \quad (5.13)$$

where τ_i is the marginal probability of the selected individual falling in the i -th row and $\tau_{.j}$ is the probability of the selected individual responding positively to category j of the column variable. As noted above, the probability τ_i is estimated as $\hat{\tau}_i = n_i / n$ and $\tau_{.j}$ is estimated by $\hat{\tau}_{.j} = m_{.j} / n$. In general, $n = \sum_i n_i$ where the n_i are random but n is fixed by design. Since the expected value of cell count m_{ij} is given by $E(m_{ij}) = n\tau_{ij}$, it can be expressed as

$$E(m_{ij}) = n\tau_{ij} \quad (5.14)$$

$$= n\tau_i\tau_{.j}, \text{ under } H_0(I_{SM}), \quad (5.15)$$

which can be estimated as

$$\hat{E}(m_{ij}) = n(n_i/n)(m_{.j}/n) \quad (5.16)$$

$$= n_i m_{.j} / n \neq m_i m_{.j} / m_{..} \quad (5.17)$$

Thus the "naive" expected values are inappropriate, i.e., they are inconsistent estimates of the true expected values and must be replaced by the expected values given by equation (5.17). The appropriate form of a modified chi-squared statistic for testing $H_0(I_{SM})$ then becomes :

$$X_{MOD}^2(I_{SM}) = \sum \sum (m_{ij} - \hat{E}(m_{ij}))^2 / \hat{E}(m_{ij}) \quad (5.18)$$

or alternatively

$$X_{MOD}^2(I_{SM}) = n \sum [\tau_{ij} - (\tau_i \tau_{.j})]^2 / \tau_i \tau_{.j} \quad (5.19)$$

This statistic is precisely the modified (pseudo) X^2 independently developed by Umesh (1995) and by Loughin and Scherer (1998). The association/homogeneity statistic given in equation (5.9) of section 5.2.1 is in fact equivalent to the independence statistic of equation (5.19), though the interpretation of the τ_{ij} are different. For statistic (5.9) the τ_{ij} are based on individual rows (which have fixed number of subjects). In the present context of the random row totals, they would have to be interpreted as conditional probabilities, i.e., the probability that a subject will respond positively to item j , given that the subject is in row i . In other words, the true difference between the homogeneity approach of section 5.2.1 and the independence approach of this section is not the numerical value of the statistic, but the sampling distribution of the statistic, which we must assume to be different in the two cases.

It must be stressed that although statistic (5.18) or (5.19) is based on the correct estimated expected values (unlike the simple generalization of the classical formula for two-way tests of association), it will still not have the “classical” χ^2 asymptotic distribution because of the additional correlation among the individual τ_{ij} ’s (or m_{ij} ’s) induced by the multiple counts. Thus a detailed asymptotic analysis of the distribution of $X^2(I_{SM})$ must be carried out. The asymptotic analysis is more complicated than for the case of fixed n_i , for two reasons:

- 1) The estimated expected value terms in equation (5.19) are the product of two random variables, $t_{.i}$ and $t_{.j}$, and,
- 2) The covariance matrix of the vector

$$\mathbf{t} = (t_{11}, \dots, t_{1c}, t_{22}, \dots, t_{2c}) \quad (5.20)$$

does not have the block diagonal structure which makes the Rao-Scott analysis of the independent case relatively straight forward (see Appendix 3). However it is possible to perform a large sample analysis to yield a Rao-Scott adjustment factor. Unlike the homogeneity case, this factor is not directly estimable from the aggregate table itself, but requires access to the full data set, or equivalently to the aggregate data in expanded table format.

As in the homogeneity test case, a large sample Rao-Scott analysis can be used to develop an approximate first-order adjusted statistic whose large sample distribution has the same mean as a chi-squared random variable on $(R-1)C$ degrees of freedom.

In this case, the relevant Rao-Scott correction factor, $\bar{\delta}(I_{SM})$, can be written in the form

$$\bar{\delta}(I_{SM}) = \left\{ 1 - \frac{E(\gamma)}{C} \right\} + 2/C \left\{ (R-1)E(\gamma) - \sum_{i=1}^R \frac{\text{cov}(m_{i.}/n, n_i/n)}{\tau_i} \right\} \quad (5.21)$$

where $\gamma = m_{.j}/n$ is the usual multiple counting factor.

Thus the correction factor consists of a first term whose estimate, $(1 - \gamma/C)$, is identical to that obtained for the case of fixed n_j , together with a second term that accounts for the randomness in the n_j . The derivation of the correction factor

$\bar{\delta}(I_{SM})$ of equation (5.21) is given in Appendix 5, where it is presented as a special case of the multiple by multiple table to be discussed in chapter 6 (and Appendix 4).

It is interesting to consider the second term in equation (5.21) in more detail. If we make the assumption that $m_{.j}/n$ can be approximately written as $m_{.j}/n \doteq E(\gamma)n_j/n$, then

$$\text{cov}(m_{.j}/n, n_j/n) = E(\gamma)\text{var}(n_j/n) = E(\gamma)\tau_j(1 - \tau_j) \quad (5.22)$$

in which case

$$2 \left\{ (R - 1)E(\gamma) - \sum_{j=1}^R \frac{\text{cov}(m_{.j}/n, n_j/n)}{\tau_j} \right\} = (R - 1)E(\gamma) - \frac{E(\gamma)}{n} (R - \sum \tau_j) \quad (5.23)$$

$$= (R - 1)E(\gamma) - E(\gamma)(R - 1) = 0. \quad (5.24)$$

Thus, under this plausible approximation (true for single response variables), the results for random row totals is the same as for fixed row totals, resulting in a test that is identical in form to the homogeneity test.

The first-order Rao-Scott adjusted chi-squared test consists of referring the adjusted statistic

$$X_C^2(I_{SM}) = X_{MOD}^2(I_{SM}) / \hat{\delta}(I_{SM}) \quad (5.25)$$

to a chi-squared distribution on $(R-1)C$ degrees of freedom. Here $\hat{\delta}(I_{SM})$ is the estimate of $\delta(I_{SM})$ given in equation (5.21), obtained by replacing population quantities by their sample estimates, e.g., $E(m_{.j})/n$ by $m_{.j}/n$. The test based on the conjectured factor derived via equation (5.22), i.e., the test having a form identical to the homogeneity test, will be denoted $X_Y^2(I_{SM})$. Its performance will be explored empirically later.

The covariance term in equation (5.21) can be estimated easily if the individual responses of the sampled individuals are available. For each sampled subject, $k = 1, \dots, n$, construct an $R \times C$ multiple response table. For a particular row, i say, let $m_{i.}^{(k)}$ represent the total number of item responses in that row, which will be an integer between 0 and C . Similarly let $n_i^{(k)}$ represent the number of subjects in that row, which will be either 0 or 1. Then the i^{th} covariance term in (5.21) can be estimated by

$$estcov(m_{i.}/n, n_i/n) = [n/(n-1)]^2 \sum_{k=1}^n (m_{i.}^{(k)} - n_{i.}/n)(n_i^{(k)} - n_i/n) \quad (5.26)$$

Thus the required estimate of the correction factor for the test of independence in a single by multiple table, with random row totals, is given by

$$\hat{\delta}(I_{SM}) = \left\{ 1 - \frac{\hat{\gamma}}{C} \right\} + 2/C \left\{ (R-1)\hat{\gamma} - \sum_{i=1}^R \frac{estcov(m_{i.}/n, n_i/n)}{\hat{t}_i} \right\} \quad (5.27)$$

Example 5.2. For the Umesh data of Table 1.3 the estimated value of $\hat{\delta}(I_{SM})$ of equation (5.27) becomes 0.376, while the corresponding value for the case of fixed row totals (see section 5.2.1) was 0.3825. Thus, for this case, the second factor in equation (5.27) has the value -0.006, i.e. the absolute value of the percentage difference between the fixed and random row total correction factor is only 1.6%. Thus the approximation given in equation (5.24) appears to be very accurate in this case. Umesh's (1995) pseudo chi-squared statistic, which is algebraically identical to $X_{MOD}^2(I_{SM})$, has the value 9.06, so that the adjusted first-order Rao-Scott statistic becomes $X_{MOD}^2(I_{SM}) = 9.06/0.37577 = 24.11$, i.e., very close to the value of 23.7 for the corrected homogeneity test. When referred to the 5% critical point of a chi-squared on 3 degrees of freedom (7.815), this is significant as before, confirming that automobile choice is associated with gender. As noted earlier, the pseudo chi-squared test itself would also have lead to rejection in this case, though it is very conservative.

Chapter 6.0

Independence in Two-Way Tables When Both Variables

Admit Multiple Response.

In this chapter, a modified chi-squared statistic for tables involving two multiple response variables is proposed and the first-order Rao-Scott correction is applied to it.

6.1 The Need for A Modified Chi-Squared Statistic

The form of the modified chi-squared statistic that is the basis of both the test of homogeneity and the test of independence for *single* by *multiple* response variables cannot be applied to *multiple* by *multiple* tables. The difficulty is that the expected values given in equation (5.15) are specifically defined with a single response row variable in mind (for either fixed or random n_i). To deal with the test of independence when the row variable also allows multiple responses, the expected values must be redefined. It then follows that the distribution of the test statistic are different from either of the previous cases, so that a third version of the Rao-Scott first-order test is required.

6.2 Derivation of the Modified Chi-Squared Statistic

We consider a sample of independent subjects who respond to two multiple response variables having R and C categories, respectively. The responses are arrayed in an $R \times C$ aggregate table of multiple response counts.

Let m_{ij} be the observed joint frequency for the i^{th} and the j^{th} multiple response categories. Such a cross-classification is illustrated in Table 6.1. The total number of subjects represented in the cross-classification is n .

Table 6.1
Two-way Array of Multiple Response Data

	Column 1	Column 2	...	Column C	One-Way Marginal
Row 1	m_{11}	m_{12}	...	m_{1C}	$m_{1.}$
Row 2	m_{21}	m_{22}	...	m_{2C}	$m_{2.}$
...	...	...	...	...	...
Row R	m_{R1}	m_{R2}	...	m_{RC}	$m_{R.}$
One-Way Marginal	$m_{.1}$	$m_{.2}$	...	$m_{.C}$	n

Let $m_{i.}$, $i = 1, \dots, R$ represent the counts, obtained by aggregating over all n respondents for the multiple response variable having R items, irrespective of their choice of items j . Similarly, let $m_{.j}$, $j = 1, \dots, C$ represent the counts, obtained by aggregating over all n respondents for the multiple response variable having C items, irrespective of their choice of items i . These represent, in other words, the observed frequencies for the two multiple response variables, each one considered separately, i.e., they represent one-way marginals.

Note that $m_{.j} \neq m_{j.} = \sum_i m_{ij}$ and $m_{i.} \neq m_{.i} = \sum_j m_{ij}$. For a single response variable, $m_{ij} = n_{ij}$, the item response probabilities associated with this table can be represented in Table 6.2.

Table 6.2
Probabilities Associated With Table 6.1

	Column 1	Column 2	...	Column c	One-Way Marginal
Row 1	τ_{11}	τ_{12}	...	τ_{1c}	$\tau_{1.}$
Row 2	τ_{21}	τ_{22}	...	τ_{2c}	$\tau_{2.}$
...	...	...	...	...	...
Row r	τ_{r1}	τ_{r2}	...	τ_{rc}	$\tau_{r.}$
One-Way Marginal	$\tau_{.1}$	$\tau_{.2}$	...	$\tau_{.c}$	

The probabilities τ_{ij} represent the probability that a given randomly selected subject will respond positively to item i of the row variable and item j of the column variable. The elements $\tau_{i.} = E(m_{i.})/n$ represent the probability that a respondent will provide a positive response to the row variable, irrespective of his response to the column variable. Similarly, $\tau_{.j} = E(m_{.j})/n$ represent the probability that a respondent will provide a positive response to the column variable, irrespective of his response to the row variable. Estimates of these row and column one-way marginal probabilities are given by

$t_{i.} = m_{i.} / n$ and $t_{.j} = m_{.j} / n$, respectively. The null hypothesis to be tested here is that of the independence of row and column variables. With the above notation, the hypothesis of row $\times$ column independence can be written as

$$H_0(I_{MM}): \quad \tau_{ij} = \tau_{i.} \tau_{.j}, \quad i = 1, \dots, R, \quad j = 1, \dots, C. \quad (6.1)$$

The expected cell frequencies are given by $E(m_{ij}) = E(m_{i.})E(m_{.j})/n$, which can be estimated by $\hat{E}(m_{ij}) = m_{i.} m_{.j} / n$. Thus the modified chi-squared statistic for testing independence in a *multiple by multiple* table is given by

$$X^2(I_{MM}) = \sum_i \sum_j \frac{[m_{ij} - (m_{i.} m_{.j} / n)]^2}{m_{i.} m_{.j} / n} \quad (6.2)$$

or equivalently,

$$X^2(I_{MM}) = n \sum_i \sum_j \frac{[t_{ij} - (t_{i.} t_{.j})]^2}{t_{i.} t_{.j}} \quad (6.3)$$

In comparison, the naive unmodified chi-square statistic would be given by

$$X^2(I) = \sum_i \sum_j \frac{[m_{ij} - (m_{i.} m_{.j}) / n]^2}{(m_{i.} m_{.j}) / n} \quad (6.4)$$

which involves incorrect (i.e. inconsistent) expected values since the marginal totals $m_{i.}$ and $m_{.j}$ involve multiple counting of subjects.

Again, though the equivalent statistics (6.2) and (6.3) involve the correct, i.e. consistent estimated, expected values, they do not have the familiar asymptotic χ^2 distribution. Further analysis is therefore required to determine appropriate first-order Rao-Scott corrections. The analysis made here proceeds from first principles, as the results given by Rao and Scott (1984) assume some of the properties of single response variables. As an example they assume that $\tau_{i.}$ and $\tau_{.j}$ sum to one, respectively, which is not true in the multiple response case.

The expression $\chi^2(I_{MM})$, given in equation (6.3) can be written in matrix form as

$$\chi^2(I_{MM}) = n(\mathbf{t} - \mathbf{t}_{i.} \otimes \mathbf{t}_{.j})' [\mathbf{D}(\mathbf{t}_{i.} \otimes \mathbf{t}_{.j})]^{-1} (\mathbf{t} - \mathbf{t}_{i.} \otimes \mathbf{t}_{.j}) \quad (6.5)$$

where $\mathbf{t}$ is now a vector of dimension $RC \times 1$ containing the item response probabilities arranged row-wise, i.e., $\mathbf{t} = (\tau_{11}, \tau_{12}, \dots, \tau_{1C}, \dots, \tau_{R1}, \tau_{R2}, \dots, \tau_{RC})'$, and $\mathbf{t}_{i.} = (\tau_{i1}, \tau_{i2}, \dots, \tau_{iC})'$ and $\mathbf{t}_{.j} = (\tau_{1j}, \tau_{2j}, \dots, \tau_{Cj})'$. The term $\mathbf{t} - \mathbf{t}_{i.} \otimes \mathbf{t}_{.j}$ represents the residual differences between the observed proportions and their expected values under H_0 , and $\mathbf{D}[(\mathbf{t}_{i.} \otimes \mathbf{t}_{.j})]^{-1}$ is a diagonal matrix with elements corresponding to the expected values, with $\otimes$ denoting the Kronecker product. The analysis will proceed under the assumption that

$n^{1/2}(\mathbf{t} - \boldsymbol{\tau}) \xrightarrow{D} N_{RC}(\mathbf{0}, \mathbf{V})$, i.e., the expected probabilities have a large sample normal distribution with mean $\boldsymbol{\tau} = E(\mathbf{t})$ and covariance matrix $n^{-1}\mathbf{V}$. Since $\mathbf{t}_{.j}$ and $\mathbf{t}_{.i}$ do not satisfy the constraint that their elements sum to one, the matrix $\mathbf{V}$ is non-singular. As shown in Appendix 1 (Case 1) $X^2(I_{MM})$ is asymptotically distributed as a weighted sum of RC independent χ^2_1 random variates, $\delta_1 W_1 + \delta_2 W_2 + \dots + \delta_{RC} W_{RC}$. The weights δ_i are eigenvalues of the estimated design effect matrix

$$\hat{\mathbf{D}}_{MM} = [\mathbf{D}^{-1}(\mathbf{t}_{.i}) \otimes \mathbf{D}^{-1}(\mathbf{t}_{.j})] \hat{\mathbf{V}}_h \quad (6.6)$$

where $\hat{\mathbf{V}}_h$, the estimated variance of the residuals $\mathbf{t} - \mathbf{t}_{.i} \otimes \mathbf{t}_{.j}$ can be obtained via Taylor series methods or via Jackknife methods as described Appendix 6. The first order Rao-Scott correction factor is the mean of the estimated weights, given by

$$(RC)\bar{\delta}(I_{MM}) = \frac{\sum_i \sum_j v(\mathbf{t}_{ij})/\mathbf{t}_{.i} \mathbf{t}_{.j} + (\sum_i \mathbf{t}_{.i}) \sum_j v(\mathbf{t}_{.j})/\mathbf{t}_{.j} + (\sum_j \mathbf{t}_{.j}) \sum_i v(\mathbf{t}_{.i})/\mathbf{t}_{.i}}{2 \sum_j c(\mathbf{t}_{.j}, \mathbf{t}_{.j})/\mathbf{t}_{.j} + 2 \sum_i c(\mathbf{t}_{.i}, \mathbf{t}_{.i})/\mathbf{t}_{.i} + 2c(\mathbf{t}_{..}, \mathbf{t}_{..})} \quad (\text{see}$$

Appendix 4, equation 37).

The first-order Rao-Scott adjusted chi-squared test consists of referring the adjusted statistic

$$X^2_C(I_{MM}) = X^2(I_{MM}) / \bar{\delta}(I_{MM}) \quad (6.7)$$

to a chi-squared distribution on RC degrees of freedom, where $\hat{\delta}(I_{MM})$ is an estimate of $\bar{\delta}(I_{MM})$. An expression for $\hat{\delta}(I_{MM})$ is derived in Appendix 4. This can be evaluated in a straightforward way, and removes the necessity for evaluating individual eigenvalues of $\hat{D}_{MM}$ (equation 6.6). Note, however, that to estimate $\hat{\delta}(I_{MM})$ one needs either the full data set in incidence matrix form, or the expanded table of dimension

$$(2^R - 1) \times (2^C - 1).$$

An approximation similar to (5.22) suggests a simplified form, which is offered here as a conjecture

$$X_Y^2(I_{MM}) = (1 - \tau_{\cdot}/R)(1 - \tau_{\cdot}/C). \quad (6.8)$$

This simplified form will be empirically explored later.

Example 6.1. Table 6.3 illustrates a *multiple by multiple* table using another example provided by Umesh (1995). In this example, Umesh cross-classified the multiple response variable “perceived quality of cars attributable to origin” (Domestic, Japanese and European) with the multiple response variable “cars parents would recommend for purchase” (Domestic, Japanese and European).

Table 6.3**Perceived Quality of Cars Versus Recommendation of Car for Purchase**

	Domestic	Japanese	Europea n	m_{·j}	m_j
Domestic	165	97	61	323	237
Japanese	141	148	65	354	265
European	128	111	85	324	239
m_{i·}	434	356	211	m _{..} = 1001	m _· = 741
m_i	217	184	111	m _· = 512	n = 400

The total ratio of counts to total number of independent subjects in this example is $1001/400 = 2.5025$, less than the maximum (RC), but large enough to signify serious multiple counting. Clearly, the Pearson chi-squared test is not valid in this case. The modified Pearson statistic yields (equation 6.3) $X^2(I_{MM}) = 23.9$. The first-order correction factor is calculated using Equation (12) of Appendix 4 and is found to be 0.2123, which gives a Rao-Scott adjusted test value of $X_C^2(I_{MM}) = 112.58$, i.e., the unmodified test statistic has been inflated by a factor close to five. This dramatically illustrates the catastrophic conservativeness of the unmodified statistic. With a tabulated critical value of $\chi_{0.05}^2$ of 31.410, clearly, the hypothesis of no association between perceived quality of cars and cars parents would recommend for purchase is rejected. In this case, it can be seen that use of the uncorrected statistic ($X^2(I_{MM}) = 23.9$) would not have led to rejection.

6.3 Comparing the Correction Factor for the Multiple by Multiple Test with Correction Factors from the Other Tests

In the examples examined in Chapter 4 and Chapter 5, the use of first-order Rao-Scott corrected statistic leads to the rejection of the null hypothesis. For these particular examples, the use of the uncorrected (modified) chi-square statistic also leads to rejection. However, it can be seen from the examples presented that multiple counting results in extremely conservative test results. In practice, therefore, one can expect to see many cases where hypotheses based on uncorrected tests are accepted when in fact appropriate use of the corrected tests would lead to rejection, the above multiple by multiple case being one such example. The following table presents a summary of the test results obtained to this point using Umesh's (1995) data. The summary clearly illustrates the extreme conservativeness of the uncorrected tests.

Table 6.4
First-Order Rao-Scott Tests Based on Umesh's (1995) Data

Test Types	$X^2(1)$	X^2_c	Correction Factor
Goodness-of-Fit	16.2	39.56	0.409
Equality of Item Probabilities	9.32	21.89	0.425
Fixed Single by Multiple	9.06	23.68	0.382
Random Single by Multiple	9.06	24.11	0.376
Random Multiple by Multiple	23.9	112.58	0.212

$X^2(1)$ in Table 6.4 refers to a test based on the correct (consistent) expected values.

Chapter 7.0

Second-Order Rao-Scott and Weighted Least Squares Tests

In this chapter, second-order Rao-Scott and weighted least squares (WLS) tests will be developed. These tests will be illustrated using the examples provided in Chapters 4, 5 and 6.

7.1 A general Approach to the Second-Order Procedure

As presented in Chapters 4, 5 and 6, Rao and Scott (1981,1984) showed that for specific hypotheses, X^2 is distributed asymptotically under the null hypothesis as a weighted sum of independent χ^2_1 random variables. The weights δ_i are eigenvalues of a design effect matrix which measures the inflation (or deflation) of the true variances and covariances, $\hat{V}$, relative to those expected under the assumption of a multinomial distribution. In the cases developed in the three previous chapters, the δ 's captured the effects of the multiple response. The general form of the first-order corrections (X^2 divided by the mean of the δ 's) was also specified. In some cases, e.g., for goodness-of-fit, homogeneity and (approximately) for single by multiple independence tests, full information on $\hat{V}$ was not required.

A better approximation to the asymptotic distribution of any X^2 -type statistic, however, can be obtained when the full estimated covariance matrix $\hat{V}$ of the cell probabilities is known. The Rao-Scott second-order correction uses a chi-squared approximation that has the same mean and the same variance as the asymptotic distribution of X^2 . The first-order correction equates only the mean.

The second-order correction factor is also based on the generalized design effects.

It is obtained by replacing the modified chi-squared statistic by

$$X_s = X^2 / \hat{\delta}(1 + \hat{a}^2) = X_c^2 / (1 + \hat{a}^2) \quad (7.1)$$

where X_c^2 denotes the first-order corrected statistic, and $\hat{a}$, which measures the degree of variation among the generalized design effects, has the form of a coefficient of variation.

It is given by

$$\hat{a}^2 = \frac{\sum_{i=1}^v \hat{\delta}_i^2}{v \hat{\delta}^2} - 1 \quad (7.2)$$

where v represents the degrees of freedom of the basic X^2 test in question. Under the null hypothesis, this second-order test statistic can be regarded as a chi-square random variable on $v^* = v/(1 + \hat{a}^2)$ degrees of freedom, where v is the degrees of freedom associated with the uncorrected X^2 test. For goodness-of-fit, equality of item response probability, homogeneity, single by multiple independence and multiple by multiple independence tests v is given by R , $R - 1$, $(R - 1)C$, $(R - 1)C$ and RC respectively.

Equation (7.2) for $\hat{a}^2$ can be evaluated immediately once the individual $\hat{\delta}_i$ have been obtained from an eigenvalue decomposition of the relevant design effect matrix. It is worth noting that Rao-Scott (1981) showed that the eigenvalue decomposition can be avoided and that equation (7.2) can instead be expressed in terms of the trace of $\hat{V}^2$. Though this does involve fewer computational operations, the speed of modern

computers renders this a trivial issue. Thus in this thesis, all second-order corrections have been obtained directly from the computed eigenvalues δ_j .

The second-order Rao-Scott correction performs better than the first-order when the δ 's differ in magnitude. When the δ_j 's are equal, the first-order corrections are asymptotically exact (Rao and Scott, 1981). The second-order adjustment provides a better approximation to a χ^2 distribution. It should be stressed that computation of the second-order correction requires access to the full data set; the aggregate table itself is not sufficient.

7.2 A General Approach to The Wald-WLS Methodology

The Wald-weighted least squares approach is based on the direct estimation of the covariance matrix of the cell observations and its subsequent use as a weighting matrix in the estimation of the model parameters and in the testing of the parameters via Wald (1943) statistics. Wald-WLS methods as discussed in Chapter 3 will be applied here to the examples of Chapters 4, 5 and 6 based on Umesh's (1995) data set. The test considered in this thesis can all be specified directly in terms of the item response probabilities (τ_j 's or $\tau_{j\cdot}$'s for one-way and two-way tables, respectively). Thus the estimation of model parameters can be side-stepped and the Wald tests can be directly specified. All tests of interest can be specified in general as

$$g(\tau) = 0 \quad (7.3)$$

where g is a possibly non-linear vector function of the vector τ of item response probabilities. When a direct (and consistent) estimate of τ , namely $\hat{\tau}$, is available (as in the cases discussed in this thesis), a general form for a Wald statistic is given by

$$g'(\hat{\tau})[\hat{V}(g(\hat{\tau}))^{-1}g(\hat{\tau})] \quad (7.4)$$

where the term in square brackets denotes the estimate of the variance of $g(\hat{\tau})$, denoted $V[g(\hat{\tau})]$.

Using standard large sample methods (Agresti, 1990), $V[g(\hat{\tau})]$ can be expressed (asymptotically) as

$$V[g(\hat{\tau})] = HV(\hat{\tau})H' \quad (7.5)$$

where H is the matrix of partial derivatives of $g(\tau)$ with respect to τ and $V(\hat{\tau})$ is the covariance matrix of $\hat{\tau}$. Thus the estimated covariance matrix can be expressed as

$$\hat{V}[g(\hat{\tau})] = \hat{H}V(\hat{\tau})\hat{H}' \quad (7.6)$$

where $\hat{H}$ is the value of H evaluated at $\tau = \hat{\tau}$.

The Wald statistic (7.4) then has a large sample chi-squared distribution, the degrees of freedom being given by the rank (dimension) of the non-singular covariance matrix $\hat{V}[g(\hat{\tau})]$. In some cases, this covariance matrix will be singular, in which case the

inverse in equation (7.4) can be replaced by a generalized inverse (Scott, Rao and Thomas, 1990). As shown by these authors, there are alternatives to the use of generalized inverses and in this thesis, Wald statistics based on singular covariance matrices will be reduced to the non-singular case by trimming specific elements of $\mathbf{g}(\mathbf{t})$ and corresponding rows and columns of $\hat{\mathbf{V}}[\mathbf{g}(\mathbf{t})]$. Details will be provided for the specific tests.

7.2.1 Wald Goodness-of-Fit Test

The Wald goodness-of-fit test statistic does not involve any singular matrices and can be expressed as

$$X_{\mathcal{W}}^2(GOF) = (\mathbf{t} - \mathbf{\tau}_0)' [\hat{\mathbf{V}}(\mathbf{t})]^{-1} (\mathbf{t} - \mathbf{\tau}_0), \quad (7.7)$$

where $\mathbf{t}$ is a $C \times 1$ vector of estimated item response probabilities (given by $\mathbf{m}/n$ and $\hat{\mathbf{V}}$ is the $C \times C$ covariance matrix of $\mathbf{t}$. This variance matrix can be easily computed using the individual level data set and is given by

$$\mathbf{V}(\mathbf{t}) = [n/(n - 1)] \sum_{k=1}^n (\mathbf{m}^{(k)} - \mathbf{t})(\mathbf{m}^{(k)} - \mathbf{t})' \quad (7.8)$$

where $\mathbf{m}^{(k)}$ is the vector of counts (ones and zeros) for the k^{th} subject. Under the null hypothesis, the Wald statistic is distributed as a chi-squared random variable on C degrees of freedom in the case of the goodness-of-fit test.

7.2.2 The Wald Test of Equality of Item Response Probabilities

The null hypothesis for the test of equality of item response probabilities is given by (see equation 4.7)

$$H_0: \tau_1 = \tau_2 = \dots, \tau_C = \tau \quad (7.9)$$

where τ is the unknown common response probability. In vector form, this can be written as

$$H_0: \mathbf{t} - \tau \mathbf{1}_C = \mathbf{t} - 1/C(\sum_j \tau_j) \mathbf{1}_C = \mathbf{g}(\tau) \quad (7.10)$$

Thus $\mathbf{g}(\mathbf{t})$, needed for the Wald Statistic, is given by

$$\mathbf{g}(\mathbf{t}) = \mathbf{t} - C^{-1} \mathbf{1}_C (\sum_{j=1}^C t_j) = (\mathbf{I}_C - C^{-1} \mathbf{1}_C \mathbf{1}_C') \mathbf{t} \quad (7.11)$$

where $\mathbf{I}_C$ is an identity matrix of dimension $C \times C$ and $\mathbf{1}_C$ is a $C \times 1$ matrix of ones.

The estimated variance of $\mathbf{g}(\mathbf{t})$, a linear transformation of $\mathbf{t}$ in this case, is given by

$$V[\mathbf{g}(\mathbf{t})] = \mathbf{J} \hat{V}(\mathbf{t}) \mathbf{J}' \quad (7.12)$$

where $J_c = I_c - C^{-1} \mathbf{1}_c \mathbf{1}_c'$ and $V(\mathbf{t})$ is given by equation (7.8). Since $J_c \mathbf{1}_c = \mathbf{0}$, so also is $\hat{V}[\mathbf{g}(\mathbf{t})] \mathbf{1}_c$. The singularity can be removed by dropping, or "trimming", the last element of $\mathbf{g}(\mathbf{t})$, which yields

$$\mathbf{g}^{(n)}(\mathbf{t}) = J_c^{(n)} \mathbf{t} = (I_c^{(n)} - C^{-1} \mathbf{1}_c^{(n)} \mathbf{1}_c^{(n)'}) \mathbf{t} \quad (7.13)$$

where $I_c^{(n)}$ is obtained by removing the last row of the identity and $\mathbf{1}_c^{(n)}$ is obtained by removing the last element of $\mathbf{1}_c$. The corresponding trimmed covariance matrix becomes

$$V^{(n)}[\mathbf{g}(\mathbf{t})] = J_c^{(n)} \hat{V}(\mathbf{t}) J_c^{(n)'} \quad (7.14)$$

The Wald statistic, on $C-1$ degrees of freedom, can then be expressed

$$X_{\#}^2(E) = \mathbf{t}' J_c^{(n)} [J_c^{(n)} \hat{V}(\mathbf{t}) J_c^{(n)'}]^{-1} J_c^{(n)} \mathbf{t} \quad (7.15)$$

No information is lost by removing the last element of $\mathbf{g}(\mathbf{t})$, namely the element

$\tau_c - \tau$, since this term is redundant, i.e., it can be determined from the $C-1$ terms, since the linear restriction $\sum_{j=1}^C (\tau_j - \tau) = 0$ must be satisfied.

7.2.3 The Wald Test of Homogeneity

The hypothesis of homogeneity is given (see equation 5.7) as

$$\tau_1 = \tau_2 = \dots \tau_R = \tau \quad (7.16)$$

where the common vector τ is unknown. By analogy with the test of equality of response probabilities described in the previous section, this can alternatively be expressed as

$$g(T) = \{[I_R - R^{-1} \mathbf{1}_R \mathbf{1}_R'] \otimes I_C\} T \quad (7.17)$$

where I_R is the identity matrix of rank R , $\mathbf{1}_R$ is a $R \times 1$ vector of ones, and $T = (\tau_1', \tau_2', \dots, \tau_R')'$. The vector $g(T)$ is of dimension $RC \times 1$, with its i^{th} block representing the term $\tau_i - \tau$. The last block, $\tau_R - \tau$, is redundant because of the linear restriction on the estimates, namely $\sum_{i=1}^R (\tau_i - \tau) = 0$. The last block can thus be removed with no loss of information, and a Wald test based on the trimmed vector $g^{(n)}(T)$ results in a non-singular Wald test given by

$$X_W^2(H) = \hat{T}' K^{(n)} [K^{(n)} \hat{V}^{(n)}(\hat{T}) K^{(n)}]^{-1} K^{(n)} \hat{T} \quad (7.18)$$

where $K^{(n)} = \{[I_R^{(n)} - R^{-1} \mathbf{1}_R^{(n)} \mathbf{1}_R^{(n)}'] \otimes I_C, I_R^{(n)}\}$ and $\mathbf{1}_R^{(n)}$ are obtained from I_R and $\mathbf{1}_R$ by removal of the last row and element, respectively, and $\hat{V}^{(n)}(\hat{T})$ is the trimmed version of the estimated covariance matrix of T . Since the latter is a block diagonal matrix having R

blocks of dimension $C \times C$, the trimmed version has dimension $(R - 1)C \times (R - 1)C$, containing $(R - 1)$ blocks $\hat{\nu}(\hat{\mathbf{t}}_i)$, each given by equation (7.8) with n replaced by n_i , the number of elements in the i^{th} row of the two-way table.

7.2.4 The Wald Test of Independence in the Single by Multiple Random Row Case

The hypothesis of independence for a single by multiple table was given in section 5.2.2 as

$$H_0: \tau_{ij} = \tau_{i.} \tau_{.j} \quad (7.19)$$

where $\tau_{i.}$ and $\tau_{.j}$ denote marginal probabilities. In equation (5.20), the $RC \times 1$ vector of table probabilities was denoted τ with estimate $\hat{\mathbf{t}}$. Thus the function $g(\tau)$, defining the null hypothesis, and on which the Wald test is based, can be expressed as

$$g(\tau) = \tau - \tau_{i.} \otimes \tau_{.j} = 0 \quad (7.20)$$

where $\tau_{i.}$ and $\tau_{.j}$ are vectors of dimension $R \times 1$ and $C \times 1$, with elements $\tau_{i.}$ and $\tau_{.j}$, respectively.

To define the Wald test, the variance of $g(\hat{\mathbf{t}})$ is required. Unlike the previous tests considered, $g(\hat{\mathbf{t}})$ is a non-linear function of the cell probabilities, by virtue of the product term $\tau_{i.} \otimes \tau_{.j}$. Thus the linearized form shown in equation (7.6), which uses the matrices of

partial derivatives, H , is required. The appropriate formula is given in Appendix 5. The variance matrix $V(\mathbf{t})$ is easily obtained, by analogy with equation (7.8), i.e.,

$$V(\mathbf{t}) = [n/(n-1)] \sum_{k=1}^n (\mathbf{m}^k - \mathbf{t})(\mathbf{m}^k - \mathbf{t})' \quad (7.21)$$

where $\mathbf{m}^k$ now denotes the $RC \times 1$ vector of table counts (ones and zeros) for the k^{th} subject, $k = 1, \dots, n$.

Because the row marginal probabilities in the single by multiple case sum to one, the vector $\mathbf{g}(\hat{\mathbf{T}})$ and the covariance matrix $\hat{V}(\hat{\mathbf{T}})$ must be trimmed. Thus $\mathbf{g}^{(n)}(\hat{\mathbf{T}})$ is obtained from $\mathbf{g}(\hat{\mathbf{T}})$ by dropping its last C element, and $\hat{V}^{(n)}[\mathbf{g}(\hat{\mathbf{T}})]$ is obtained from $\hat{V}[\mathbf{g}(\hat{\mathbf{T}})]$ by dropping its last C rows and columns. The resulting Wald statistic is given by

$$X_{\mathcal{W}}^2(I_{SM}) = [\mathbf{g}^{(n)}(\mathbf{t})]' [\hat{V}^{(n)}[\mathbf{g}(\mathbf{t})]]^{-1} [\mathbf{g}^{(n)}(\mathbf{t})] \quad (7.22)$$

Under the null hypothesis, $X_{\mathcal{W}}^2(I_{SM})$ will be distributed in large samples as a chi-squared random variable on $(R-1)C$ degrees of freedom.

Note also the variance matrix $V(\mathbf{t})$ can be estimated via a matrix transformation of the multinomial covariance matrix of the expanded table (of dimension $(R \cdot 2^C - 1)(R \cdot 2^C - 1)$). Similarly, the linearization process to estimate $V[\mathbf{g}(\mathbf{t})]$ can be avoided by a direct use of the jackknife approach as developed in Appendix 6, where the pseudo replicates are obtained via the leave one out method.

7.2.5 The Wald Test with Multiple by Multiple Tables

The Wald test of independence for a multiple by multiple table is very similar to that described for single by multiple table in the previous section. For the multiple by multiple case, there is no need for any trimming of the $g(\mathbf{t})$ vector or the corresponding covariance matrix, as neither set of marginal probabilities sums to one. Thus the Wald test becomes

$$X_W^2(I_{MM}) = [g^{(n)}(\mathbf{t})]' [\hat{V}^{(n)}[g(\mathbf{t})]]^{-1} [g^{(n)}(\mathbf{t})], \quad (7.23)$$

where in this case, $g(\mathbf{t}) = \mathbf{t} - \mathbf{t}_{i.} \otimes \mathbf{t}_{.j}$. Again, the variance matrix can be obtained by linearization (Appendix 4) or via the jackknife approach (Appendix 6). Under the null hypothesis, the Wald statistic is distributed as a chi-squared random variable on RC degrees of freedom in the case of the multiple by multiple test.

7.3 Application of the Second-Order Rao-Scott and WLS to Umesh's (1995) data set

Example 7.1: Goodness-of-Fit Test. Using $\hat{V}$ computed from the multiple response in disaggregated form and $\hat{a}^2$ given by equation (7.2), the second-order Rao-Scott corrections, $(1 + \hat{a}^2)$, is equal to 1.0837. Since the first-order test value obtained in Section 4.1 was 39.6, the adjusted Rao-Scott second-order test is now given by

$X_S^2(GOF) = 39.6/1.08267 = 36.54$. This value is compared to the 5% chi-squared distribution on $v = 2.768$ degrees of freedom, which corresponds to a critical value of

7.409. Clearly, the null hypothesis is rejected. The Wald statistic is in this case equal to 37.70 and is compared to the 5% critical value of a chi-squared random variable on three degrees of freedom, which is 7.815. The null hypothesis is again rejected by the Wald test, which gives results quite similar to the second-order Rao-Scott test. These two procedures give results similar to the first-order Rao-Scott correction, which suggests that for these data, “the design effects” do not differ greatly. Thus these tests reinforce the conclusion that the response probabilities for females are significantly different from the hypothesized (male) probabilities.

Example 7.2: Equality of Item Response Probability Test. With eigenvalues obtained from the design effect matrix defined in Section 4.2, and $\hat{a}^2$ given by equation (7.2), the second-order Rao-Scott correction, $(1 + \hat{a}^2)$, is found to be equal to 1.00228, which results in an adjusted chi-squared test value of $X_S^2(E) = 21.89/1.00228 = 21.84$. This test value is compared to a critical value of 5.983 on $v = 1.995$ degrees of freedom. The null hypothesis of equality of item response probabilities is rejected. The Wald statistic (equation 7.15) is equal to 22.49 and is compared to the 5% critical value of a chi-squared random variable on two degrees of freedom, which is 5.991. Both second-order Rao-Scott and Wald tests give results quite close to the first-order Rao-Scott result, and all of them lead to the rejection of the null hypothesis. This reinforces earlier conclusion that the individual response probabilities for females are significantly different.

Example 7.3: Homogeneity Test. For the test of homogeneity, the covariance matrix is a block diagonal formed by the individual covariance matrices corresponding to the population represented by each row of the table, i.e., in the Umesh (1995) case by the separate covariance matrices for men and women. With eigenvalues computed from the design effect matrix given in Appendix 3 and using equation (7.2) to compute $\hat{a}^2$, the second-order correction, $(1 + \hat{a}^2)$, is found to be equal to 1.0093., which gives an adjusted second-order Rao-Scott statistic of $X_S^2(H) = 23.68/1.00882 = 23.47$. This value is compared to the 5% critical point of 7.767 of a chi-squared distribution on $v = 2.972$ degrees of freedom. The Wald statistic (equation 7.18) has the value of 24.63 and is referred to the critical point of a chi-squared on 3 degrees of freedom (7.815). Again, both second-order Rao-Scott and Wald statistic are significant, as was the first-order Rao-Scott test. This reinforces the earlier conclusion that automobile choice is associated with gender. The pseudo chi-squared test itself would also have lead to rejection in this case, though it is clearly very conservative.

Example 7.4: Single by Multiple Test. The design effect matrix for the single by multiple test of independence can be computed directly from the formula given in Appendix 5 for $\hat{D}_{SM}$. A total of C of these eigenvalues will be identically zero and can be ignored. Equation (7.2), based on the non-zero eigenvalues, can now be used to recalculate $\hat{a}^2$. The latter evaluates at 0.01131 and the second-order Rao-Scott correction factor, $(1 + \hat{a}^2)$, is found to be 1.01131. This leads to an adjusted second-order Rao-Scott statistic of $X_S^2(I_{SM}) = 24.11/1.01131 = 23.84$. This value is compared to the 5% critical

point of a chi-squared distribution on $v = 2.97$ degrees of freedom, which is 7.767. The Wald statistic has the value of 23.62, to be referred to a chi-squared on 3 degrees of freedom (7.815). As in the earlier examples, both second-order Rao-Scott and the Wald procedures are similar in outcome to the first-order Rao-Scott, indicating that for these data, the first-order test is sufficient.

Example 7.5: Multiple by Multiple Test. Using the design effect matrix given in Appendix 4, equation (2), and $\hat{a}^2$ from equation (7.2) the second-order Rao-Scott correction factor, $(1 + \hat{a}^2)$, is found to be 1.2868, which leads to an adjusted second-order Rao-Scott statistic of $X_S^2(I_{MM}) = 112.52/1.2868 = 87.44$. This value is compared to the 5% critical point of a chi-squared distribution on $v = 6.994$ degrees of freedom, which corresponds to value of 14.058. The Wald statistic has the value of 104.92, to be referred to a chi-squared on 9 degrees of freedom, i.e., a critical value of 31.41. Though both tests reject the null (that there is no association between perceived quality and parental recommendation), these test results are not quite as close to the first-order results as in previous examples. Even in the multiple by multiple case, however, the differences among the two Rao-Scott procedures and the Wald procedure are not major, suggesting that all these methods provide a viable way to analyse cross tabulated multiple response data. The closeness of the second-order correction factor to one also suggests little variation in the design effects, for each of the examples studied, though the variation is clearly greatest for this multiple by multiple example.

Table 7.1 presents a summary of first- and second-order Rao-Scott tests and Wald test results for one- and two-way tables. A detailed comparison of all these methods, based on simulation, will be described in the next chapter.

Table 7.1
Second-Order Rao-Scott and Wald Tests
Based on Umesh's (1995) Data

Test Types	X^2_c (df)	X^2_s (df)	Wald (df)
Goodness-of-Fit	39.56 (3)	36.54 (2.768)	37.79 (3)
Equality of Item Probabilities	21.89 (2)	21.84 (1.995)	22.49 (2)
Fixed Single by Multiple	23.68 (3)	23.47 (2.972)	24.62 (3)
Random Single by Multiple	24.11 (3)	23.84 (2.972)	23.62 (3)
Random Multiple by Multiple	112.58 (9)	87.48 (6.994)	104.92 (9)

Chapter 8

A Monte Carlo Simulation Study

In this chapter the behaviour of the methods proposed in Chapters 4 to 7 will be examined for realistic sample sizes, a necessary check since they all rely on asymptotic, i.e., large sample, results for their inferential justification.

8.1 Design of the Monte Carlo Study

The Monte Carlo simulation is designed to evaluate the actual Type I error rates for small, moderate and large sample sizes (e.g., for samples of size 50, 100, 300, 600 and 1000), for the statistics included in the study, namely the unmodified and modified chi-squared statistics, the Rao-Scott first- and second-order corrections, and the corresponding Wald procedures. The standard 5% normal test level will be assumed.

By evaluating the actual Type I error control of the various statistics, for each test setup, it will be possible to see how well an approximate test (derived using large sample theory) works in practice, for small and moderate sample sizes. The data for each test setup (goodness-of-fit, equality, homogeneity, etc.) will be generated from a model that satisfies the relevant null hypothesis. A test that generates 7% rejection (at the nominal 5% level) will be regarded as adequate. One that generates over 10% rejection will be regarded as excessively liberal.

For all one-way and two-way tests, the correlation structure was chosen to match the correlation structure found in Umesh's (1995) published data. Each run of the Monte

Carlo experiment, for each test, involved 1000 independent computer generated samples. For each of the 1000 realizations, the test statistic in question was computed and compared to the appropriate critical value and a “reject-accept” decision recorded. Estimates of the Type I error control of each statistic were then be computed as the proportion of the number of rejections per 1000 realizations. Separate experiments were performed for samples of varying size, from small ($n=50$) to large ($n=1000$). The Monte Carlo study focussed only on Type I error, i.e., data were in all cases generated from a probability distribution that conformed to the null hypothesis, as noted before. Studies of test power have been deferred for future research.

Emrich and Piedmonte (1991)’s method for generating correlated binary variates was used to generate the data. Their algorithm is described in Appendix 8.

8.2 The Statistics Examined

The study focussed only on χ^2 , modified χ^2 (i.e., an χ^2 test based on consistently estimated expected values), first- and second-order Rao-Scott corrected tests, and Wald tests. Loughin and Scherer’s bootstrap approach and Agresti’s generalized loglinear model techniques were not investigated due to their complexity and computational intensiveness. These comparisons remain for future study. The specific tests compared were:

- 1) goodness-of-fit tests using the uncorrected, Rao-Scott first- and second-order adjusted chi-squared statistics, respectively denoted by $X^2(GOF)$, $X_c^2(GOF)$.

 $X_s^2(GOF)$ and $X_w^2(GOF)$ as they are defined in Sections 4.1, 7.1 and 7.2.1.
- 2) equality of probabilities test for the uncorrected, Rao-Scott first- and second-order adjusted chi-squared statistics, respectively denoted by $X^2(E)$, $X_c^2(E)$ and $X_s^2(E)$ and $X_w^2(E)$ as they are defined in Sections 4.2, 7.1 and 7.2.2.
- 3) homogeneity test using the unmodified, modified, Rao-Scott first- and second-order adjusted chi-squared statistics, respectively denoted by $X_U^2(H_{SM})$, $X_{MOD}^2(H_{SM})$, $X_C^2(H_{SM})$ and $X_S^2(H_{SM})$ and $X_W^2(I_{HM})$ as they are defined in Sections 5.2.1, 7.1 and 7.2.3.
- 4) test of independence (single by multiple) using the unmodified, modified, Rao-Scott first- and second-order adjusted chi-squared statistics, respectively denoted by $X_U^2(I_{SM})$, $X_{MOD}^2(I_{SM})$, $X_C^2(I_{SM})$, $X_S^2(I_{SM})$ and $X_W^2(I_{SM})$ as they are defined in Sections 5.2.2, 7.1 and 7.2.4 respectively.
- 5) test of independence (multiple by multiple) using the unmodified, modified, Rao-Scott first- and second-order adjusted chi-squared statistics, respectively denoted by $X_U^2(I_{MM})$, $X_{MOD}^2(I_{MM})$, $X_C^2(I_{MM})$, $X_S^2(I_{MM})$ and $X_W^2(I_{MM})$ as they are defined in Sections 6.2, 7.1 and 7.2.5 respectively.

8.3 Simulation of the Goodness-of-Fit Test

8.3.1 Simulation Details

In order to create the data set needed for a goodness-of-fit test, $X^2(\text{GOF})$, the correlation structure of the binary variates coupled with the vector of multiple response probabilities were used. The data are first simulated using the Umesh data set with a vector of multiple response probabilities given by $\tau' = [0.495 \ 0.728 \ 0.549]$, i.e., a multiple response index of $\gamma = 1.772$. The following matrix of correlations was used:

$$Corr_{female} = \begin{Bmatrix} 1 & -0.10438 & 0.02291 \\ -0.10438 & 1 & -0.18547 \\ 0.02291 & -0.18547 & 1 \end{Bmatrix} \quad (8.1)$$

One-way aggregated multiple response tables were created by summing responses over all subjects. At each run, the observed vector of probabilities were compared to the hypothesized probabilities. The behaviour of the procedures were also studied with small, moderate and high multiple counting indices. These were set at 1.2, 1.8 and 2.4. In these cases, the actual expected aggregate counts must necessarily differ from the Umesh's (1995) values.

8.3.2 Simulation Results

All results given represent the proportion of actual rejections of a true hypothesis at a 5 percent nominal level, based on 1000 independent Monte Carlo trials. [Note that the 95% confidence interval for an estimated rejection rate of approximately 5% is $\pm 1.4\%$. Thus an estimate of 0.05 yields a 95% confidence interval of (0.036, 0.064)].

Tables 8.1 to 8.4 give the actual significance level (SL) for goodness-of-fit tests and all examined procedures. Both first- and second-order Rao-Scott corrected tests maintain their nominal significance level, i.e. the proportion of rejections is very close to 0.05. There seems to exist no need to compute the second-order test, since it virtually has the same performance as the first-order. The Wald goodness-of-fit test also offers excellent control of the test levels, although it is slightly higher for samples of size 50. The uncorrected test ($\chi^2(GOF)$) is in every situation very conservative and should never be recommended. Rao-Scott first- and second-order tests and the Wald test have values that are relatively close and emerge as competing tests for testing goodness-of-fit hypotheses involving vectors of multiple item response probabilities.

Table 8.1**Comparison of Actual Significance Levels for The Goodness-of-Fit Case****Using $\tau' = [0.495 \ 0.728 \ 0.549]$, i.e. $\gamma = 1.772$ (Umesh values)**

Sample Size	$X^2(GOF)$	$X^2_c(GOF)$	$X^2_s(GOF)$	$X^2_w(GOF)$
50	0	0.057	0.055	0.07
100	0	0.049	0.045	0.063
300	0	0.056	0.053	0.065
600	0.001	0.05	0.048	0.058
1000	0	0.054	0.052	0.057

Table 8.2**Comparison of Actual Significance Levels for The Goodness-of-Fit Case****Using $\tau' = [0.20 \ 0.40 \ 0.60]$, i.e. $\gamma = 1.20$**

Sample Size	$X^2(GOF)$	$X^2_c(GOF)$	$X^2_s(GOF)$	$X^2_w(GOF)$
50	0.004	0.053	0.045	0.09
100	0.006	0.068	0.055	0.076
300	0.005	0.054	0.048	0.052
600	0.004	0.049	0.047	0.049
1000	0.006	0.054	0.05	0.049

Table 8.3**Comparison of Actual Significance Levels for The Goodness-of-Fit Case****Using $\tau' = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma = 1.80$**

Sample Size	$X^2(GOF)$	$X_C^2(GOF)$	$X_S^2(GOF)$	$X_W^2(GOF)$
50	0.005	0.065	0.051	0.107
100	0.001	0.067	0.045	0.09
300	0.001	0.058	0.046	0.059
600	0	0.066	0.052	0.057
1000	0	0.056	0.047	0.056

Table 8.4**Comparison of Actual Significance Levels for The Goodness-of-Fit Case****Using $\tau' = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma = 2.4$**

Sample Size	$X^2(GOF)$	$X_C^2(GOF)$	$X_S^2(GOF)$	$X_W^2(GOF)$
50	0	0.062	0.053	0.1
100	0	0.055	0.046	0.084
300	0	0.055	0.05	0.063
600	0	0.058	0.047	0.053
1000	0	0.054	0.043	0.049

8.4 Simulation of The Equality of Response Probabilities

8.4.1 Simulation Details

To test for equality of response probabilities, $X^2(E)$, one-way aggregated multiple response tables were generated using a vector of multiple response probabilities given by $\tau' = [0.59 \ 0.59 \ 0.59]$, i.e., a multiple response index of $\gamma = 1.77$ corresponding to the Umesh data set. The correlation matrix defined in equation (8.1) was used. One-way aggregated multiple response tables were created by summing responses over all subjects. Again, the observed vectors of probabilities were compared to the hypothesized probabilities which were a vector of identical values. The average number of positive responses per subject was fixed at 1.2, 1.8 and 2.4..

8.4.2 Simulation Results

Tables 8.5 to 8.8 compare the actual significance levels of $X^2(E)$, $X_C^2(E)$, $X_S^2(E)$ and the Wald statistic $X_W^2(E)$ for equality of item response probabilities tests. Both Rao-Scott first- and second-order tests perform well and are virtually identical, the second-order correction factor being almost 1. The significance level is quite stable. The Wald statistic is quite insensitive to the sample size. Once again, both Rao-Scott first- and second-order corrected tests and the Wald statistic are competitive tests when equality of item response probabilities test is considered, though the first-order test $X_C^2(E)$ is easiest to apply. The uncorrected test is overly conservative. Tables 8.5, 8.6, 8.7 and 8.8 also provide simulation results for small, moderate and large multiple counting indices.

Table 8.5
Comparison of Actual Significance Levels for The Equality
of Item Response Probabilities Case Using
 $\tau' = [0.59 \ 0.59 \ 0.59]$, i.e. $\gamma = 1.77$

Sample Size	$X^2(E)$	$X_C^2(E)$	$X_S^2(E)$	$X_W^2(E)$
50	0.002	0.049	0.043	0.061
100	0.002	0.046	0.045	0.062
300	0	0.055	0.055	0.056
600	0	0.054	0.054	0.057
1000	0.001	0.046	0.046	0.049

Table 8.6
Comparison of Actual Significance Levels for The Equality
of Item Response Probabilities Case Using
 $\tau' = [0.40 \ 0.40 \ 0.40]$, i.e. $\gamma = 1.20$

Sample Size	$X^2(E)$	$X_C^2(E)$	$X_S^2(E)$	$X_W^2(E)$
50	0.033	0.055	0.054	0.077
100	0.024	0.043	0.042	0.062
300	0.022	0.046	0.046	0.05
600	0.032	0.053	0.053	0.056
1000	0.028	0.052	0.052	0.057

Table 8.7
Comparison of Actual Significance Levels for The Equality
of Item Response Probabilities Case Using
 $\tau' = [0.60 \ 0.60 \ 0.60]$, i.e. $\gamma = 1.80$

Sample Size	$X^2(E)$	$X_C^2(E)$	$X_S^2(E)$	$X_W^2(E)$
50	0.001	0.058	0.056	0.079
100	0.002	0.053	0.052	0.063
300	0.001	0.054	0.054	0.056
600	0	0.055	0.055	0.06
1000	0.003	0.047	0.047	0.049

Table 8.8
Comparison of Actual Significance Levels for The Equality
of Item Response Probabilities Case Using
 $\tau' = [0.80 \ 0.80 \ 0.80]$, i.e. $\gamma = 2.40$

Sample Size	$X^2(E)$	$X_C^2(E)$	$X_S^2(E)$	$X_W^2(E)$
50	0	0.051	0.047	0.13
100	0	0.052	0.047	0.11
300	0	0.054	0.054	0.07
600	0	0.064	0.064	0.066
1000	0	0.05	0.05	0.052

8.5 Simulation of Homogeneity, i.e., Single (fixed) $\times$ Multiple Test

8.5.1 Simulation Details

For tables involving a single response variable with fixed margins and a multiple response variable, the data creation process was the following. For each population, the individual marginal probabilities and the correlations among the different binary variates of the multiple response variable were computed. Since under the null hypothesis, the vector of probabilities are equal across populations, only one of these vectors with its corresponding correlations matrix was needed to generate the multiple response data. From Umesh's (1995) data set, the vector of multiple response probabilities for the female population was used. It is given by $\tau' = [0.495 \ 0.728 \ 0.549]$, i.e., a multiple response index of $\gamma = 1.772$.

A different seed number was used to generate the data for each population. Then, the data so created were merged in order to form the appropriate tables. The multiple counting indices were fixed at 1.2, 1.8 and 2.4, in order to simulate small, moderate and high response density, respectively. Computation was also made for a five by five table just for the purpose of checking bootstrap results by Loughin and Scherer (1998) (see Appendix 5). The size of each sample was held fixed. Once again, the correlation structure among the binary variates of the multiple response variable for each population was preserved and the model assumption of homogeneity respected.

8.5.2 Simulation Results

Comparison of the Type I error rates of $X_U^2(H_{SM})$, $X_{MOD}^2(H_{SM})$, $X_C^2(H_{SM})$, $X_S^2(H_{SM})$ and the Wald statistic for homogeneity of response probabilities test using Umesh's (1995) data set are given in Tables 8.9, 8.10, 8.11 and 8.12. Once again, both Rao-Scott first- and second-order tests are quite stable and perform well and maintain the nominal significance level value. The Wald statistic also performs well, except for samples of size 50 where it becomes slightly liberal. The Wald test is still a competing procedure for Rao-Scott corrected tests, though its relative complexity and lack of familiarity to practitioners will be a disadvantage. Both the unmodified and the modified (Loughin and Scherer, 1998) or the pseudo chi-squared test (Umesh, 1995), should never be used. They are overly conservative.

Table 8.9
Comparison of Actual Significance Levels for
The Hypothesis of Homogeneity Case Using
 $\tau' = [0.495 \ 0.728 \ 0.549]$, i.e. $\gamma = 1.772$ and $p' = [0.54 \ 0.46]$

Sample Size	$X_U^2(H_{SM})$	$X_{MOD}^2(H_{SM})$	$X_C^2(H_{SM})$	$X_S^2(H_{SM})$	$X_W^2(H_{SM})$
50	0	0	0.051	0.041	0.098
100	0	0.002	0.055	0.051	0.072
300	0	0	0.056	0.053	0.066
600	0	0	0.045	0.044	0.046
1000	0	0	0.051	0.051	0.059

Table 8.10
Comparison of Actual Significance Levels for
The Hypothesis of Homogeneity Case Using
 $\tau' = [0.20 \ 0.40 \ 0.60]$, i.e. $\gamma = 1.2$ and $p' = [0.50 \ 0.50]$

Sample Size	$X_U^2(H_{SM})$	$X_{MOD}^2(H_{SM})$	$X_C^2(H_{SM})$	$X_S^2(H_{SM})$	$X_S^2(H_{SM})$
50	0.001	0.001	0.058	0.05	0.072
100	0	0.001	0.057	0.048	0.056
300	0.001	0.005	0.056	0.047	0.052
600	0	0.006	0.059	0.056	0.068
1000	0	0.007	0.048	0.045	0.053

Table 8.11
Comparison of Actual Significance Levels for
The Hypothesis of Homogeneity Case Using
 $\tau' = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma = 1.80$ and $p' = [0.50 \ 0.50]$

Sample Size	$X_U^2(H_{SM})$	$X_{MOD}^2(H_{SM})$	$X_C^2(H_{SM})$	$X_S^2(H_{SM})$	$X_S^2(H_{SM})$
50	0.002	0.002	0.058	0.044	0.07
100	0	0	0.069	0.052	0.069
300	0.001	0.001	0.066	0.054	0.062
600	0.001	0.001	0.063	0.05	0.049
1000	0.001	0.002	0.055	0.049	0.05

Table 8.12
Comparison of Actual Significance Levels for
The Hypothesis of Homogeneity Case Using
 $\tau' = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma = 2.40$ and $p' = [0.50 \ 0.50]$

Sample Size	$X_U^2(H_{SM})$	$X_{MOD}^2(H_{SM})$	$X_C^2(H_{SM})$	$X_S^2(H_{SM})$	$X_S^2(H_{SM})$
50	0	0	0.063	0.044	0.058
100	0	0	0.057	0.048	0.063
300	0	0	0.062	0.056	0.058
600	0	0	0.06	0.051	0.052
1000	0	0	0.056	0.049	0.052

8.6 Simulation of the Independence, i.e., Single (random) $\times$ Multiple Test

8.6.1 Simulation Details

In order to perform a test of independence, $X^2(I_{SM})$, on a cross-tabulation involving a single response variable with random margins, and a multiple response variable, the individual marginal probabilities and the correlations among the different binary variates for the multiple response variable were used to generate the multiple response data for each simulated subject. The vector of multiple response item probabilities corresponding to the row marginal of Table 1.3 was used, i.e. $\tau' = [0.5925 \ 0.6625 \ 0.5975]$, and $\gamma = 1.8525$. For the single response variable, the data were generated from the binomial distribution using the vector of single response proportion computed

from the same table, i.e. $p' = [0.54 \ 0.46]$. The correlation matrix corresponding to the vector of “perceived quality “ probabilities was also computed from the Umesh (1995) data set and was found to be

$$Corr_{PerQual} = \begin{Bmatrix} 1 & -0.04317 & -0.03743 \\ -0.04317 & 1 & -0.02520 \\ -0.03743 & -0.02520 & 1 \end{Bmatrix} \quad (8.2)$$

Individual units were then associated with multiple response vectors to build up the required two-way tables. The average number of responses per subject was fixed at 1.2, 1.8 and 2.4 in order to simulate small, moderate and high multiple counting indices, respectively. A different seed number was used for each data generation process, to ensure the row/column independence specified by $H_0(I_{SM})$. Once again, the correlation structure among the binary variates of the multiple response variable were preserved and the model assumption of independence respected.

8.6.2 Simulation Results

Comparison of the Type I error rates of $X_U^2(I_{SM})$, $X_{MOD}^2(I_{SM})$, $X_C^2(I_{SM})$, $X_S^2(I_{SM})$ and the Wald statistic for single by multiple test using Umesh’s (1995) data set are given in Tables 8.13, 8.14, 8.15 and 8.16. Both the unmodified and the modified statistic are extremely conservative, and should never be used. Both the first- and second-order Rao-Scott tests provide good control of test significance. $X_Y^2(I_{SM})$, the approximate statistic

derived by applying the formula for single by multiple with fixed row margins to the case where the margins are random was also studied. This also performs very well and provides test significance control almost identical to that provided by the legitimate Rao-Scott first-order correction. The agreement between these two procedures becomes increasingly marked as the sample size increases, suggesting that these tests are asymptotically equivalent. In other words, it can be conjectured that equation (5.22) is asymptotically correct, as opposed to being just an approximation. The second-order Rao-Scott and the Wald procedure behave as in the previous cases.

Table 8.13

**Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Single by Multiple Case
Using $\tau' = [0.5925 \ 0.6625 \ 0.5975]$, i.e $\gamma = 1.8525$ and $\rho' = [0.54 \ 0.46]$**

Sample Size	$X_U^2(I_{SM})$	$X_{MOD}^2(I_{SM})$	$X_\gamma^2(I_{SM})$	$X_C^2(I_{SM})$	$X_S^2(I_{SM})$	$X_W^2(I_{SM})$
50	0	0.001	0.049	0.059	0.052	0.082
100	0	0.001	0.063	0.064	0.061	0.073
300	0	0	0.057	0.059	0.059	0.062
600	0	0	0.062	0.062	0.062	0.059
1000	0	0	0.048	0.048	0.048	0.05

Table 8.14
Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Single by Multiple Case
Using $\tau' = [0.20 \ 0.40 \ 0.60]$, i.e. $\gamma = 1.2$ and $p' = [0.50 \ 0.50]$

Sample Size	$X_U^2(I_{SM})$	$X_{MOD}^2(I_{SM})$	$X_Y^2(I_{SM})$	$X_C^2(I_{SM})$	$X_S^2(I_{SM})$	$X_W^2(I_{SM})$
50	0	0.006	0.065	0.07	0.067	0.09
100	0	0.002	0.055	0.057	0.053	0.072
300	0	0.001	0.046	0.047	0.045	0.05
600	0	0.002	0.059	0.06	0.053	0.056
1000	0	0.007	0.051	0.051	0.047	0.055

Table 8.15
Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Single by Multiple Case
Using $\tau' = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma = 1.80$ and $p' = [0.50 \ 0.50]$

Sample Size	$X_U^2(I_{SM})$	$X_{MOD}^2(I_{SM})$	$X_Y^2(I_{SM})$	$X_C^2(I_{SM})$	$X_S^2(I_{SM})$	$X_W^2(I_{SM})$
50	0	0	0.06	0.066	0.048	0.072
100	0	0	0.067	0.07	0.049	0.062
300	0.001	0.002	0.067	0.068	0.054	0.06
600	0.003	0.004	0.051	0.052	0.051	0.055
1000	0.002	0.004	0.057	0.058	0.05	0.055

Table 8.16
Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Single by Multiple Case
Using $\tau' = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma = 2.40$ and $p' = [0.50 \ 0.50]$

Sample Size	$X_U^2(I_{SM})$	$X_{MOD}^2(I_{SM})$	$X_\gamma^2(I_{SM})$	$X_C^2(I_{SM})$	$X_S^2(I_{SM})$	$X_W^2(I_{SM})$
50	0	0	0.054	0.061	0.049	0.083
100	0	0	0.059	0.067	0.065	0.077
300	0	0	0.065	0.065	0.055	0.054
600	0	0	0.056	0.058	0.051	0.056
1000	0	0	0.055	0.055	0.046	0.049

8.7 Simulation of The Multiple $\times$ Multiple Test

8.7.1 Simulation Details

In order to perform a test of independence on a cross-classification involving two multiple response variables, $X^2(I_{MM})$, the individual marginal probabilities and the correlation among the different binary variates for each multiple response variable were computed separately and generated independently using different seed numbers. The vectors of one-way marginal response probabilities for the multiple response were $\tau'_i = [0.5925 \ 0.6625 \ 0.5975]$, i.e. $\gamma = 0.617$ and $\tau'_j = [0.5425 \ 0.46 \ 0.2775]$, i.e. $\gamma = 0.426$. The correlations among the multiple response categories were computed from the Umesh (1995) data. These were, for the multiple response variable “perceive quality”

$$Corr_{PerQual} = \begin{Bmatrix} 1 & -0.04317 & -0.03743 \\ -0.04317 & 1 & -0.02520 \\ -0.03743 & -0.02520 & 1 \end{Bmatrix} \quad (8.3)$$

and for the multiple response variable “parental recommendation”

$$Corr_{ParRec} = \begin{Bmatrix} 1 & -0.38079 & -0.24899 \\ -0.38079 & 1 & -0.12390 \\ -0.24899 & -0.12390 & 1 \end{Bmatrix} \quad (8.4)$$

Each multiple response variable was generated using its own marginal probabilities and corresponding matrices of correlations. A different seed number was used each time.

Since under the null hypothesis $\mathbf{t} = \mathbf{t}_i \otimes \mathbf{t}_j$, for each Monte Carlo trial the appropriate multiple by multiple response table was formed by taking the Kronecker product of the marginal probabilities in order to obtained the observed values. One again, the correlations among the individual binary variates for each multiple response variable were preserved and the hypothesized independence of rows and columns respected. The average number of responses per subject was fixed at 1.2, 1.8 and 2.4 in order to simulate small moderate and high multiple counting indices.

8.7.2 Simulation Results

Comparison of actual significance levels in the independence test case involving two multiple response variables are provided from Table 8.17 to Table 8.22 for all procedures examined, including the conjectured test X^2_{MM} . The first-order Rao-Scott corrected test shows a slight inflation of the error rate with samples of size 50. So also does the conjectured form described in equation (6.8). However, both are well within the acceptable range. Rao-Scott second-order test provides good control of test levels for all examined sample sizes.

As in the case of single by multiple independence test, X^2_{MM} turns out to be a good approximation to the Rao-Scott first-order corrected test and should be recommended because of its ease of use. The Wald statistic exhibits poor control of the test significance levels for sample sizes 50 to 100 and performs well for sample of sizes 300, 600 and 1000. Both the unmodified and the modified chi-squared tests turn out to be extremely conservative as in all previous cases, and should never be used. The second order corrected test should be preferred to the other procedures in multiple by multiple analyses, whenever individual data are available and whenever its slightly greater complexity is not an issue.

Table 8.17

Comparison of Actual Significance Levels for The Hypothesis of Independence in
The Multiple by Multiple Case Using $\tau'_{i.} = [0.5925 \ 0.6625 \ 0.5975]$
and $\tau'_{.j} = [0.5425 \ 0.46 \ 0.2775]$, i.e. $\gamma_{i.} = 1.8525$ and $\gamma_{.j} = 1.28$.

Sample Size	$X^2_{U(MM)}$	$X^2_{MOD(MM)}$	$X^2_Y(MM)$	$X^2_C(MM)$	$X^2_S(MM)$	$X^2_W(MM)$
50	0	0	0.076	0.088	0.053	0.164
100	0	0	0.055	0.059	0.040	0.11
300	0	0	0.059	0.06	0.043	0.065
600	0	0	0.062	0.064	0.047	0.051
1000	0	0	0.063	0.065	0.051	0.063

Table 8.18

Comparison of Actual Significance Levels for
The Hypothesis of Independence in The Multiple by Multiple Case Using
 $\tau'_{i.} = [0.20 \ 0.40 \ 0.60]$ and $\tau'_{.j} = [0.20 \ 0.40 \ 0.60]$, i.e. $\gamma_{i.} = 1.20$ and $\gamma_{.j} = 1.20$

Sample Size	$X^2_{U(MM)}$	$X^2_{MOD(MM)}$	$X^2_Y(MM)$	$X^2_C(MM)$	$X^2_S(MM)$	$X^2_W(MM)$
50	0	0	0.061	0.067	0.042	0.173
100	0	0	0.071	0.076	0.054	0.109
300	0	0	0.066	0.065	0.052	0.079
600	0	0	0.052	0.055	0.046	0.057
1000	0	0	0.051	0.05	0.047	0.048

Table 8.19**Comparison of Actual Significance Levels for****The Hypothesis of Independence in The Multiple by Multiple Case**Using $\tau'_{i.} = [0.20 \ 0.40 \ 0.60]$ and $\tau'_{.j} = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma_{i.} = 1.20$ and $\gamma_{.j} = 1.80$

Sample Size	$X^2_{U(MM)}$	$X^2_{MOD(MM)}$	$X^2_{\gamma(MM)}$	$X^2_{C(MM)}$	$X^2_{S(MM)}$	$X^2_{W(MM)}$
50	0	0	0.07	0.085	0.054	0.15
100	0	0	0.079	0.086	0.054	0.124
300	0	0	0.067	0.071	0.046	0.065
600	0	0	0.089	0.085	0.058	0.063
1000	0	0	0.065	0.066	0.046	0.058

Table 8.20**Comparison of Actual Significance Levels for****The Hypothesis of Independence in The Multiple by Multiple Using** $\tau'_{i.} = [0.30 \ 0.60 \ 0.90]$ and $\tau'_{.j} = [0.30 \ 0.60 \ 0.90]$, i.e. $\gamma_{i.} = 1.80$ and $\gamma_{.j} = 1.80$

Sample Size	$X^2_{U(MM)}$	$X^2_{MOD(MM)}$	$X^2_{\gamma(MM)}$	$X^2_{C(MM)}$	$X^2_{S(MM)}$	$X^2_{W(MM)}$
50	0	0	0.081	0.098	0.042	0.199
100	0	0	0.099	0.103	0.047	0.128
300	0	0	0.089	0.092	0.054	0.067
600	0	0	0.09	0.09	0.054	0.071
1000	0	0	0.079	0.078	0.055	0.062

Table 8.21**Comparison of Actual Significance Levels for****The Hypothesis of Independence in The Multiple by Multiple Case**

Using $\tau'_{i.} = [0.30 \ 0.60 \ 0.90]$ and $\tau'_{.j} = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma_{i.} = 1.80$ and $\gamma_{.j} = 2.40$

Sample Size	$X_U^2(I_{MM})$	$X_{MOD}^2(I_{MM})$	$X_Y^2(I_{MM})$	$X_C^2(I_{MM})$	$X_S^2(I_{MM})$	$X_W^2(I_{MM})$
50	0	0	0.093	0.1	0.051	0.207
100	0	0	0.076	0.084	0.042	0.12
300	0	0	0.069	0.073	0.05	0.126
600	0	0	0.074	0.076	0.047	0.066
1000	0	0	0.077	0.076	0.044	0.055

Table 8.22**Comparison of Actual Significance Levels for****The Hypothesis of Independence in The Multiple by Multiple Case**

Using $\tau'_{i.} = [0.90 \ 0.80 \ 0.70]$ and $\tau'_{.j} = [0.90 \ 0.80 \ 0.70]$, i.e. $\gamma_{i.} = 2.40$ and $\gamma_{.j} = 2.40$

Sample Size	$X_U^2(I_{MM})$	$X_{MOD}^2(I_{MM})$	$X_Y^2(I_{MM})$	$X_C^2(I_{MM})$	$X_S^2(I_{MM})$	$X_W^2(I_{MM})$
50	0	0	0.081	0.085	0.053	0.204
100	0	0	0.09	0.094	0.056	0.17
300	0	0	0.063	0.065	0.044	0.126
600	0	0	0.058	0.06	0.04	0.069
1000	0	0	0.062	0.063	0.053	0.073

8.8 Some Additional Simulation Results

Simulations involving cross-tabulations of multiple response variables having each four categories were also performed. Results are reported in Appendix 7 from Table 1 to Table 5 . In some cases, the first-order Rao-Scott test is slightly liberal. Whenever this happens, the second order Rao-Scott test offers good control of the test significance level.

The behaviour of the statistics in five $\times$ five single by multiple tables was also studied. Results are reported in Table 6 of Appendix 7. In general, the statistics examined behave in a manner comparable to results reported for cross-tabulations involving multiple response variables with three categories. However, for samples of size 50, the Rao-Scott first-order statistic, its approximate version and the second-order Rao-Scott statistic are conservative, while the Wald statistic is extremely liberal. This may be due to sparseness in the aggregate table, which has, in this case, 25 cells.

The Agresti and Liu's (1999) naive chi-squared statistic discussed in Section 2.2.5 was also investigated in single by multiple tables. Results are reported in Table 7 of Appendix 7. For all sample sizes studied and all multiple counting indices considered, it provides good control of type I error rate.

8.9 Summary and Discussion of the Monte Carlo Results

Data simulated from correlated binary variates were used to examine the Type I error performance of chi-squared test statistics for different types of hypotheses involving multiple response data. A study of test statistics from two different methods, the Rao-

Scott adjustment approach and the weighted least squares approach via Wald statistics suggest that the better methods overall are provided by the Rao-Scott adjustment approach. Both first- and second-order test perform well. The second-order test provides particularly good control of Type I error over the full range of sample sizes and multiple counting settings. In practice, users can do well with only the first-order tests, in most cases.

A special strength of the Rao-Scott first-order test resides in the fact that it can be easily computed from the aggregate tables. In more complex cases, such as in both of the independence test cases, easily computed approximations are available. These approximate statistics, namely $X^2_\gamma(I_{SM})$ and $X^2_\gamma(I_{MM})$, use corrections factors that mimic the results obtained from the homogeneity test, i.e, they use correction factors of the type $1 - \gamma/C$. In general, these tests exhibit very similar control of Type I error to the first-order corrected tests. The similarity is so marked that it invites the conjecture that the full theoretical forms of the correction factor for the single by multiple (equation 5.21) and the multiple by multiple (equation 6.7) are asymptotically equal to the simplified form given in equations (5.22) and (6.8), respectively. These two conjectures remain an issue of further research.

The only situation in which the Rao-Scott first-order tests, namely $X^2_\gamma(I_{SM})$ and $X^2_\gamma(I_{MM})$, exhibit difficulty in controlling Type I error is the multiple by multiple case, for 100 or fewer numbers of subjects. In this case, empirical Type I error reach about 10 %, outside of the 95% confidence interval of (3.6%,6.4%).

The Rao-Scott second-order test is the most stable of all compared procedures, as noted. It always achieves the desired significance level within acceptable bounds.

However it requires knowledge of the individual data level, which is almost never available with published papers in the marketing literature. Hence, another virtue of the first-order test is its applicability to published aggregate tables of multiple response data. Whenever the full data set is available, it is advisable to use the second-order Rao-Scott test. Based on a simulated study of complex survey data, Thomas and Rao (1987) also recommended using the second-order procedure whenever the full estimated matrix $\hat{V}$ is available.

It is worth noting in passing that an emerging competitor for the Rao-Scott first-order correction is Agresti and Liu's (1999) naive chi-squared test discussed, in Section 2.2.5. In addition to giving a test value comparable to certain maximum likelihood-based procedures, and comparable to the like Rao-Scott first-order test results, early simulation results show that the naive statistic exhibits a behaviour similar to the Rao-Scott first-order statistic in terms of control of Type I error rates (see Appendix 7). This will be explored elsewhere.

The Wald procedure requires large sample sizes before attaining good Type error properties. It also requires knowledge of the individual data level. This feature of Wald statistics has also been noted in study of test behaviour under cluster sampling (Thomas and Rao, 1987; Thomas *et al.*, 1996). A point of superiority of Rao-Scott methodology over the Wald procedure is that their second-order test behaves properly across all sample sizes examined.

Chapter 9

Extensions of the Proposed Methodology

In this chapter extensions of the Rao-Scott approach to multiple response data of the “pick k/r” type will be outlined. For this type of data, it will be shown that the tests developed in Chapters 4, 5, 6 and 7 can be directly used. Some specifics of “pick k/r” data will be clarified. It will also be shown that Rao-Scott theory extends to cross-tabulations involving “order r” and “order k/r” types of data. Consideration will also be given to displays showing aggregation across multiple response variables.

9.1. Multiple Response Data of the “Pick k/r” Type

9.1.1 Goodness-of-fit, Equality of Response Probability and Homogeneity Tests

With “pick k/r” data, each multiple response category has response probability τ_j as before. For a one-way array of multiple response counts $\sum_1^k \tau_j = k$. Therefore, by design, the correction factor is known at the outset. It is again of the form of $1 - \gamma/C$ and given in the simple goodness-of-fit-test by

$$\hat{\delta}(GOF) = 1 - k/C. \quad (9.1)$$

In the equality of response probability case, the first-order Rao-Scott corrections factor is again given by $\hat{\delta} = (1 - \tau)(1 - \rho)$, which can alternatively be expressed as

$$\hat{\delta}(E) = (1 - k/C)(1 - \hat{\rho}). \quad (9.2)$$

Here, as in Section 4.2, $\hat{\rho}$ is the estimate of the average correlation among the multiple response counts.

For each row of a cross-tabulation of a single by a multiple response variable $\sum_1^k t_{ij} = \gamma$. In the homogeneity case, since γ is known at the outset by design, it follows that the first-order Rao-Scott corrections is given, as in the goodness-of-fit case, by

$$\hat{\delta}(H_{SM}) = 1 - k/C. \quad (9.3)$$

9.1.2 Single by Multiple and Multiple by Multiple Tests

With multiple response data of the “pick k/r” type, the expanded table has many structural zeros. For a multiple response with R categories, the number of non-empty cells is given by the binomial factor C_R^k . Non-empty cells for the multiple response variable with C categories are similarly defined. The expanded table is truncated and is not of the form $(2^R - 1) \times (2^C - 1)$. It is rather of the form $C_R^k \times C_C^k$. The formulae for multiple response variables of the “pick any/r” form still apply. However, the corresponding matrices **A**, **π** , **F₁** and **F_J** defined in Appendix 4 need to be reviewed due to the structural zeros. Again, the first-order Rao-Scott can be derived directly from the information available from the aggregate table using equations (48) and (49) of Appendix 4. Alternatively, the variance matrix can also be obtained through the jackknife method.

The single by multiple correction factor can also be obtained as a specialization of the multiple by multiple correction factor theoretical results in a fashion similar to the proofs provided in Appendices 4 and 5.

It follows that the first-order Rao-Scott corrections factor can be derived as per Appendix 5, equation (5) . The variance matrix can also be obtained through the jackknife procedure.

It is worth noting that the first-order correction factor can be easily used since the correction factors are known a priori by design. The approximate statistics defined via equations (5.22) and (6.8) can also be used.

9.2 Rao-Scott Corrections for “Order r ” and “Order k/r Data”

“Order k/r data” (order k out of r) emerge in marketing research whenever respondents are asked to rank a subset k out of r listed products. “Order k/r ” is a special case of “order r ” data, whereby all respondents are asked to assign a rank to all items on a given list. Strictly speaking, “order k/r ” and “order r ” types of data and multiple response data of the pick k/r and pick any/ r form belong to the same family from a data theory perspective. In our treatment of “order k/r ” and “order r ”, we will only note with a one or a zero whether a given rank is assigned to an item or not. No attempts to treat the ranks as ordinal data will be made.

A detailed discussion of the “order r ” case will be given first; the “order k/r ” will then be obtained as a modification.

9.2.1 "Order r "

The following table is typical of tabulations involving "order r " type of data.

Table 9.1

Typical Cross-Tabulation of "Order r " Data

Products	Rank 1	...	Rank r	Total
Item 1	n_{11}	...	n_{1r}	n
...	...	...	...	...
Item r	n_{r1}	...	n_{rr}	n
Total	n	n	n	nr

For the analysis of "order r " type of data in cross-tabulated format, the usual chi-squared test cannot be used. Since each respondent examines all r products and uses all r ranks there is a violation of the independence assumption despite the fact that the categories are mutually exclusive. The cell counts do not involve multiple response. However the marginal totals do, and in fact, are multiplied by a multiple counting factor r , which is known a priori.

If we were to test the hypothesis that all items have the same probability of being ranked first, no correction would be necessary. A chi-squared test for an individual hypothesis of this nature will not require adjustment; each subject selects only one item.

However, if we want to simultaneously test the null hypothesis that all items have the same probability of receiving a given rank, irrespective of the rank, standard chi-squared tests will not suffice, because of the correlation across ranks. Thus Rao-Scott corrections are necessary in this case.

One can compute a naive chi-squared statistic, X_j^2 , for each rank using

$X_j^2 = C \sum_i (n_{ij} - n/r)^2/n$ and the composite naive chi-squared statistic can be obtained

by summing the r chi-squared statistics, i.e., $X^2 = \sum X_j^2$. Anderson (1959) used this

approach and noted that the naive chi-squared statistic does not approach χ^2 with $(r-1)^2$

degrees of freedom (the classical result for an $R \times C$ table with $R = C = r$) and he proposed

an adjustment to the naive test. This turns out to be the same as the Rao-Scott first-order correction factor.

Let us note that the naive chi-squared can also be written in matrix form

$$[n - (n/r)\mathbf{1}_r]'\mathbf{D}[n - (n/r)\mathbf{1}_r], \quad (9.4)$$

where $\mathbf{1}_r$ is a vector of ones of dimension $r \times 1$, and where $\mathbf{D}$ is a diagonal matrix of dimension $r^2 \times r^2$, having diagonal elements equal to r/n .

To implement the Rao-Scott approach, what is needed is the design effects matrix of the form of $\mathbf{D}\mathbf{V}$, where $\mathbf{V}$ denotes the covariance matrix of the vector $\mathbf{n}$, of cell counts,

i.e., $\mathbf{n} = (n_{11}, n_{12}, \dots, n_{1r}, \dots, n_{r1}, \dots, n_{rr})'$.

Since individual n_{ij} can be regarded as binomial, (but not independent) counts, the diagonal structure of V (sufficient to define Rao-Scott first-order corrections) is given by

$$V = n \begin{pmatrix} 1/r(1 - 1/r) & . & . & . \\ . & 1/r(1 - 1/r) & . & . \\ . & . & . & . \\ . & . & . & 1/r(1 - 1/r) \end{pmatrix} \quad (9.5)$$

Because the row and column cell counts are all equal to n , the number of independent subjects, the rank of V must equal

$r^2 - (r - 1) - (r - 1) - 1 = (r - 1)^2$. Then the average, $\hat{\delta}$, of the eigenvalues of the design effects matrix DV is given by

$$(r - 1)^2 \hat{\delta} = \text{trace}(D_0 V) = r^2(r/n)n(1/r)(1 - 1/r) = r(r - 1) \quad (9.6)$$

Thus the correction factor for the naive X^2 turns out to be $r/(r - 1)$, a factor known a priori, i.e., by design. A first-order test thus consists of referring $[(r - 1)/r]X^2$ to a χ^2 on $(r - 1)^2$.

Anderson (1959) developed an adjustment to the chi-squared procedure to perform a test of varietal differences in a consumer preference study involving three

varieties of snap beans. His adjusted test value turns out to be the same as the test value obtained when Rao-Scott theory is applied to his data set. In fact, Anderson (1959) showed that in this case, his test is asymptotically correct. This means that for the “order r ” problem, the first-order Rao-Scott procedure is asymptotically valid. This is so because when the full correlation structure of V is evaluated (Anderson, 1959), it turns out that all non-zero eigenvalues of DV are equal, i.e., the well known constant design effect case (Rao and Scott, 1981).

When ties among ranks are permitted, two basic chi-squared assumptions are violated, namely: the exclusivity and the independence assumptions. The former requires an upward adjustment of the test, the latter a downward adjustment of the test. The cumulative effect of these needed adjustments is not studied in this thesis, though the foundation and orientation of further research on the subject will be laid.

9.2.2 “Order k/r ”

In the case of “order k/r ” data, a similar approach can be followed. In that case, a typical table is shown below. Table 9.2 differs from Table 9.1 in that the row totals no longer sum to n , the total number of subjects. The following table is typical of “order k/r ” data.

Table 9.2**Typical Cross-Tabulation of "Order k/r" Data**

Products	Rank 1	...	Rank k	Total
Product 1	n_{11}	...	n_{1k}	$n_{1.}$
Product 2				
...	...	...		
...	...	...	...	...
Product r	n_{r1}	...	n_{rk}	$n_{r.}$
Total	n	n	n	kn

In fact it is easy to see that there is no constraint on the row marginal tables, i.e.

$\sum_{j=1}^k n_{ij} = n_{i.}$. Again, the uncorrected chi-squared statistic is computed as in the case of "order r" type of data. Also, V is the same, (at least as far as diagonal elements go).

Because the column cell counts all equal to n , the number of independent subjects, the rank of V must equal $(r - 1)k$. Thus the correction factor is given by

$$k(r - 1)\hat{\delta}_0 = \text{trace}(\mathbf{D}_0^{-1} \hat{V}), \quad (9.7)$$

which reduces to

$$k(r - 1)\hat{\delta}_0 = k(r - 1) \quad (9.8)$$

with r being the number of listed items (product) and k the maximum assigned rank. In other words, the first-order Rao-Scott “correction” in this case is equal to one, i.e., no correction. This is somewhat misleading, in that the reference test statistic is a $\chi^2_{k(r-1)}$ random variable, i.e., the degrees of freedom have been corrected from the “naive” value of $(k-1)(r-1)$ to $k(r-1)$. Once this has been done, first-order Rao-Scott theory simply states that no other scaling (correction) of the statistic is necessary.

9.3 Aggregation Across Multiple Response Variables

Aggregation across multiple response variables occurs in various practical situations. Aggregation can also be performed in order to transform “pick any” data into “pick any/n”. When the sample size is small or the aggregate table is highly sparse, aggregation of multiple response categories can be very useful. In what follows, two methods of aggregation will be considered and appropriate examples will be provided. The first example deals with multiple response data of the pick any/n form; the second involves multiple response data of the “pick any” type.

9.3.1 Summing Correlated Categories

First, consider a survey in which the respondents were asked to check all the magazines they read on a list of n magazines. The researcher then aggregates the data by summing the relevant original counts by magazine category, i.e., sports, fashion and automobile magazines. Since correlated counts are being aggregated to provide a single count for a category, the individual category counts will no longer be binomial. Hence, the

simplification used in deriving the $1 - \gamma/C$ correction factor does not apply. However, when data in the incidence form are available, the variances of the aggregate multiple response variable can be easily computed using known formulae. As an example, consider that the aggregated sports magazine category (S) consists of three items (x, y and z), then the variance of the aggregated category can be calculated as

$$V(S) = V(x) + V(y) + V(z) + 2cov_{xy} + 2cov_{xz} + 2cov_{yz}. \quad (9.10)$$

Alternatively, because subjects are independent, formula (7.8) can be used directly to find the variances of the aggregated counts and their covariances. Because the aggregated counts are not correlated binomial, the methods of Sections 4 through 7 cannot be applied directly. However, when the variances and covariances of the aggregated counts are obtained via equation (9.10), the previously derived correction formulas can be modified to accept the estimated variances and covariances. For the single by multiple and the multiple by multiple tests of independence, the formulas given in equations (10) and (48) of Appendix (4) are already specified in terms of general variances and covariances. These can thus be immediately adapted.

9.3.2 Aggregation at the Incidence Matrix Level

Aggregation can also take place when the data are yet in the incidence matrix form. What is really needed here is simply a rule for aggregating multiple response variable categories. Let us illustrate with a multiple response variable having seven categories, each with a cell multiple count equal to 30. Further, let us assume that the first three categories can be aggregated using some meaningful criterion and that the remaining four categories are aggregated in two new other categories, each with two of the four remaining original categories. The newly formed categories do not have 90, 60 and 60 as cell counts, since this aggregation process involves summing only the number of times the original categories forming an aggregate multiple response variable category are checked at least once, irrespective of the choices of the other aggregate categories. For this type of aggregation, the correction factors are computed based on the principles described previously and Rao-Scott theory applies in the same fashion.

Another version of the previous example of aggregation of multiple response data occurs when researchers convert multiple response data of the “pick any” type into multiple response data of the “pick any/n” type. Such transformation may be suggested by sparseness in the aggregate table itself. As an example, consider a survey in which respondents are asked to list all the magazines they would buy if money were not a limiting factor. At the analysis stage, the researchers might again want to aggregate the data so obtained into magazine category, i.e., sports, fashion and automobile magazines, i.e, transforming a “pick any” data type into a “pick any/n” type.

9.4 Doubly Complex Multiple Response Data

Multiple response data turn out to be doubly complex when they are collected in the context of surveys that involve stratification and possibly several stages of selection (i.e., multi-stage sampling). To their natural difficulty associated with multiple response data is added the complexity of the survey design itself.

Rao-Scott theory was primarily conceived for the analysis of cross-tabulations from complex surveys. It is therefore quite natural to apply it in context of true complex surveys where multiple response data are collected. It is noteworthy observing that both variance inflation (for the complex survey) and deflation factors (for the multiple responses) are at work in such doubly complex situations. The cumulative effect of the needed adjustments is not studied in this thesis. It is worth noting, however, that the correction factors (37) and (5) for multiple by multiple and single by multiple tests of independence given in Appendices 4 and 5 can be directly applied, provided only that appropriate complex survey techniques of variance and covariance estimation are employed (Wolter, 1985).

Chapter 10

Summary and Conclusions

In this chapter the research methods developed in the preceding chapters will be integrated. The benefits and limitations of this study will be presented and tasks for future research identified.

10.1 The Methods Developed

In this study, Rao-Scott first- and second-order adjusted chi-squared tests for analysing one-way and two-way tables of multiple response data of the “pick any /r” type were developed. These tests were provided for five types of commonly used hypotheses, namely goodness-of-fit test, equality of response probabilities test, homogeneity test, independence test for cross-tabulations involving a single response variable and a multiple response variable, and independence test for tables involving two multiple response variables. Extensions to other forms of multiple response data, such as “pick any”, “pick k/r”, “order r” and “order k/r” multiple response data were also considered.

The most striking conclusion drawn from this study is that there are first-order tests for all five testing situations considered that can be applied without access to the subject level data. For equality in one-way arrays, this requires use of the conservative test based on the lower limits of the correlations among estimated probabilities. For the single by multiple and multiple by multiple tests of independence, avoiding use of subject level data implies use of the conjectured versions of the first-order corrections. For both cases, however, the conjectured tests work as well as the complete versions.

The second order, being of the general form $(1 + a^2)$, the following table only provides a synoptic view of the first-order Rao-Scott approach.

Table 10.1
Rao-Scott First-Order Tests Developed for Multiple Response Data of
The “Pick Any/r” Types

Test Types	Aggregated Tables	Disaggregated Tables
$X^2(GOF)$	$1 - \hat{\gamma}/C$	$1 - \hat{\gamma}/C$
$X^2(E)$	$(1 - \hat{\gamma}/C)(1 - \hat{\rho}_r)$	$(1 - \hat{\gamma}/C)(1 - \hat{\rho})$
$X^2(H)$	$1 - \hat{\gamma}/C$	$1 - \hat{\gamma}/C$
$X^2(I_{SM})$	$1 - \hat{\gamma}/C$	$(R - 1)C\bar{\delta} = \sum_i \sum_j v(t_{ij})/t_i t_j - \sum_j v(t_{.j})/t_{.j} +$ $(\sum_j t_{.j}) \sum_i v(t_{i.})/t_{i.} - 2 \sum_i (t_{i.}, t_{.i})/t_{i.}$
$X^2(I_{MM})$	$(1 - \tau_{.}/R)(1 - \tau_{.}/C)$	$(RC)\bar{\delta}(I_{MM}) = \sum_i \sum_j v(t_{ij})/t_i t_j + (\sum_i t_{i.}) \sum_j v(t_{.j})/t_{.j}$ $+ (\sum_j t_{.j}) \sum_i v(t_{i.})/t_{i.} - 2 \sum_j c(t_{.j}, t_{.j})/t_{.j}$ $- 2 \sum_i c(t_{i.}, t_{i.})/t_{i.} + 2c(t_{.}, t_{.})$

This study has also revealed that the second-order corrected test should be preferred to the other procedures in multiple by multiple analyses, whenever individual data are available and whenever its slightly greater complexity is not an issue.

It appears that for all five testing situations considered, the Wald procedures can be used, whenever large sample sizes and access to the full data set are available.

Given the varied ways in which multiple response data can be used in the task of theory-building across all domains of the management discipline, all areas of business research will benefit from the diffusion of these procedures specially designed for analysing aggregated multiple response data. In what follows, some specific technical and marketing research benefits are outlined.

10.2 Technical Benefits

The research reported in this thesis addresses the commonly encountered challenges in the analysis and interpretation of aggregated multiple response data, and offers a set of theoretically grounded and practically relevant solutions that can be easily applied by market research analysts and readily comprehended by marketing managers.

The review of the marketing literature summarized in Chapter 2 has revealed that the dominant use of multiple response data is for the construction of perceptual maps, a purely descriptive statistical application. With the methods developed in this study, inferential analysis based on one- and two-way array of aggregate multiple response data will now be possible.

For the single by multiple test of independence, Umesh (1995) did not provide a sampling distribution to which his pseudo chi-squared statistic could be compared. Loughin and Scherer (1998), on the other hand, used a very computationally intensive bootstrap process to simulate a sampling distribution. As a result of the current research, it will now be possible, through the first-order Rao-Scott approach, to compare the adjusted version of the pseudo chi-squared statistic to critical values obtained from published tables of the chi-squared distribution.

The first-order Rao-Scott approach also provides easy to apply methods for other testing situations, including goodness-of-fit and equality of response probabilities in one-way tables, as well as independence for multiple by multiple tables. The benefit is tests that need access only to the aggregate data and one-way marginal tables.

The usefulness of this development to marketing researchers should not be underestimated. Anecdotal evidence suggest that new methods of analysis take twenty years or more from development to routine adoption by analysts and applied researchers. The techniques developed in this thesis are all based on tests of the familiar chi-squared form, and get referred to standard chi-squared tables. The only modification required is division of a simple correction factor based on aggregate table counts and one-way marginals. There is no need for additional theory and no need for the practitioner to become familiar with the statistical technology. Thus one can hope that adoption of the new procedures will be rapid. Simulation results suggest that the proposed methods are similar in accuracy to the more complex and computationally involved techniques proposed by Loughin and Scherer (1998) and Agresti and Liu (1999).

10.3 Application to Other Analytical Techniques

As an example where the new tests may be used to extend existing technologies, Green and Carmone's (1977) *segment congruence analysis* might be considered. This is a method of selecting among competing segmentation bases, and is based on an a posteriori collection of multiple response data. Once the segments are obtained, respondents are classified on the basis of their simultaneous membership into segments from all bases. In its essence, this recoding process, critical to answering the question of the best segmentation basis, is equivalent to asking a respondent to check all the segments s/he thinks s/he belongs to on a list of n segments, i.e., multiple response data. With each segmentation basis yielding a two-segment solution, a multiway table is easily constructed.

Green and Carmone (1977) used loglinear models on the corresponding expanded table in order to answer the question of the best segmentation basis. However, a researcher might need to decide whether the membership probabilities for all segments are equal or, in other words, may want to decide whether to consider the market as undifferentiated or segmented. Such test can only be performed on the aggregated table. The proposed research provides an appropriate testing procedure not readily available through other means.

Another technique that might benefit is *chi-square automatic interaction detector* (CHAID), since chi-squared tests for cross-tabulations of multiple response variables are now readily available.

Grover and Dillon (1988) noted that for more than three decades marketing researchers have been using brand switching data derived from aggregate panel data for understanding market structure. At the methodology level, loglinear models (Novak, 1993; Iacobucci and Henderson, 1997) and weighted least squares (Novak and Stangor, 1987) have been used on force switching matrices whereby individuals are classified by their first and second choices among r brands. With the methods developed in this thesis, it will now be possible to test relevant hypotheses on aggregate brand switching data matrices. It will also be possible to test relevant association hypotheses on aggregate brand switching data matrices admitting multiple product choices. Therefore, it is expected that variety seeking, brand loyalty and market share, as areas of interest to marketing managers and researchers alike, will benefit from the diffusion and adoption of the methods developed in this thesis.

Finally, it should be noted that the new techniques provide an alternative to the scaling of multiple response data that is often resorted to. In marketing research, multiple response data (e.g., benefits sought, or occasions of usage), have been used most often as the first stage in a multi-stage investigation, to identify the most important items. In the second stage, the retained meaningful sets of items are usually intervally scaled, which avoids the need to deal directly with the multiple response data in its original form. With intervally scaled data, classical multivariate analysis techniques can be used which automatically account for within subject correlations. The products-by-uses analysis and the benefit segmentation study typify such approaches.

In the products-by-uses analysis, samples of customers are given a target product or brand and asked to suggest as many uses for that product as possible. They are then asked for other products or brands suggested by such uses and then additional uses for the expanded product list. Then another sample of respondents is asked to judge whether they would consider using each product for each situation. Researchers typically avoid the need to deal directly with multiple response data by asking respondents to judge the appropriateness of each of m products for each of n usage scenarios on a k -point “likelihood of use” scale. These data are often transformed to average suitability measures within a product-by-uses matrix and then factor-analysed. This methodology is referred to as products-by-uses analysis (Day et al., 1979; Bourgeois et al., 1987).

The research reported in this study offers a more direct approach for dealing with multiple response data and provides ways to avoid the need to transform them into another scale.

10.4 Managerial Relevance

The managerial usefulness of a set of valid test procedures for analysing cross-classified multiple response data can be illustrated in terms of a wide variety of marketing strategies.

As an example featuring market penetration strategy, managers interested in achieving widespread market coverage, channel acceptance, volume sales and profits could, given store level data in a market area, explore the association between the stores in which a group of products is distributed, and shoppers’ use of these stores.

As for the development of markets for mature products, managers could, given a matrix of products by usage, test the association between nature of usage (new versus old usages) and multiple product choice. for example. Managers could, in a context of experimental design, test the existence of an association between benefits sought in a product in different market areas, while controlling for some relevant marketing stimuli.

Finally, managers interested in investigating customer satisfaction could cross-classify benefits expected with benefits provided, which are two natural multiple response variables, and test the hypothesis of independence of the two variables. The rejection of the null hypothesis could be followed by estimation of Umesh 's (1995) measures of association. A strong association in this case would indicate the existence of customer satisfaction, given that customer satisfaction can be defined as the absence of deviation between what the product is expected to deliver and the actual experience of customers with the product.

Similar practical uses for the proposed methods can be found in decisions context involving all the components of the marketing mix (e.g., product, pricing, distribution and promotion).

10.5 Limitations

- (1) The proposed methods rely on large sample results which necessarily implies that they are approximate in practice. In itself, this is not unusual since the classical χ^2 procedure itself is only asymptotically correct. The proposed procedures differ in that they (first-order and second-order) are approximate even for large samples. However, the simulation results suggest that they offer good approximations, and may be more accurate than some asymptotically correct procedures such as those based on weighted least squares.
- (2) The Monte Carlo simulation study did not include the Loughin and Scherer (1998) and Agresti and Liu (1999) methods because of their complexity. To implement a simulation study that includes them would require separate investigations of efficient numerical computational strategies. Again, this represents an issue for further research.
- (3) The methods considered in this study are only available for one- and two-way tables. Extension of these methods to multi-way tables remains a task for future research.

- (4) Exploration of power, i.e., rejection probabilities for data conforming to the alternative hypothesis, was not undertaken. This remains a task for future research.
- (5) The most practically useful version of the single by multiple and multiple by multiple independence tests have only been empirically validated. Their theoretical forms rely on conjectures inspired by some approximate relations. These conjectures need to be theoretically explored and validated.
- (6) The work reported on “pick k/r ”, “order r ”, and order k/r ” data types is preliminary. Further work remains to be done, specifically to define other appropriate hypothesis of interests to practitioners, and to develop corresponding tests.

References

- Agresti, Alan (1990), *Categorical Data Analysis*, John Wiley & Sons, Inc.
- Agresti, Alan (1996), *An Introduction to Categorical Data Analysis*, New York: Wiley
- Agresti, Alan and Liu I-Ming (1998), "Modelling Responses to a Categorical Variable Allowing Arbitrarily Many Category Choices," July; <http://www.stat.ufl.edu/~aa> [Accessed December 1998].
- Agresti, Alan and Liu I-Ming (1999), "Modelling a Categorical Variable Allowing Arbitrarily Many Category Choices," *Biometrics*, 55, 936-943.
- Altham, P. M. E. (1978), "Two Generalizations of the Binomial Distribution," *Applied Statistics* 27, 162-167.
- Anderson, R. L. (1959), "Use of Contingency Table in the Analysis of Consumer Preference Studies," *Biometrics*, 15, 582-590.
- Benzecri, J. P. et al. (1973), *L'Analyse des Donnees: II, L'Analyse des Correspondances*, Paris: Dunod.
- Binder, D. A. (1983), "On the Variances of Asymptotically Normal Estimators from Complex Surveys," *International Statistical Review*, 51, 279-292.
- Bockenholt, U. And Bockenholt, I. (1990), "Modelling Individual Differences in Unfolding Preference Data: A Restricted Latent Class Approach," *Applied Psychological Measurement*, 14 (3), p. 257-269.
- Bourgeois, Jacques C., Haines, Georges Jr., and Sommers, Montrose S. (1987), "Product Market Structure: Problems, Definitions and Issues," in Michael J. Houston (Ed.), *Review of Marketing*, American Marketing Association.
- Brier, S. S. (1980), "Analysis of Contingency Tables Under Cluster Sampling," *Journal of the American Statistical Association*, 80:148-157.
- Carroll, J. Douglas and Green, Paul E., (1987), "Psychometrics Methods in Marketing Research: Part II, Multidimensional Scaling," *Journal of Marketing Research*, vol. 34 193-204.
- Cochran, William G. (1963), *Sampling Techniques*, second edition, John Wiley & Sons, Inc.

Coombs, Clyde. H. (1964), *A Theory of Data*, John Wiley & Sons, Inc.

Coren, S. and Hakstian, A. R. (1990), "Methodological Implications for Interaural Correlation: Count Head not Ears," *Perceptions and Psychographics*, 48, 291-294.

Cox, Trevor F., and Cox, Michael A. A. Cox (1994), "Multidimensional Scaling", in Cox, D. R. And Snell, E. J., *Analysis of Binary Data*, London: Chapman and Hall.

Crowder, M. J. (1978), "Beta-binomial ANOVA for Proportions," *Applied Statistics* 27, 34-37.

Day, George S., Shoker Allan D., Srivastava, Rajendra K. (1979), "Customer-Oriented Approaches to Identifying Product-Markets," *Journal of Marketing*, vol. 43 (Fall). 8-19.

DeSarbo, W. S. and Hoffman, D. L. (1986), "A New Unfolding Threshold Model for the Spatial Representation of the Binary Choice Data," *Applied Psychological Measurement*, 10, 247-64.

DeSarbo, W. S. and Hoffman, D. L. (1987), "Constructing MDS Joint Spaces from Binary Choice Data: a Multidimensional Unfolding Threshold Model for Marketing Research," *Journal of marketing Research*, 24 40-54.

DeSarbo, W. S. and Cho J., (1989), " A Stochastic Multidimensional Scaling Vector Threshold Model for the Spatial Representation of 'Pick Any/N' Data," *Psychometrika*, 54, 105-30.

DeSarbo, W., Monroe, A. K. and Monroe, L. A. (1993), "Non-Spatial Tree Models for the Assessment of Competitive Market Structure: an Integrated Review of The Marketing and Psychometric Literature," In Eliashberg and Lilien (eds). *Handbooks in Operations Research and Management Science: Marketing*, Amsterdam, Netherlands: Elsevier Science.

Dillon, W. R. and Kumar, A. (1994), "Latent Structure and Other Mixture models in Marketing: An Integrative Survey and Overview," in Richard Bagozzi (eds), *Advanced in Marketing Research*.

Eckart, C. and Young, G (1936), "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, 1, 211-218.

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Elrod, Terry (1988), "Choice Map: Inferring Product Attributes and Heterogeneous Consumer Preferences from Choice Data," *Marketing science*, 7, 21-40.

Emrich, Lawrence J. and Piedmonte, Marion R. (1991), "A Method for Generating High-Dimensional Multivariate Binary Variates," *The American Statistician*, November, Vol. 45, No. 4.

Fay, R. E. (1979), "On Adjusting the Pearson Chi-Square Statistic for Cluster Sampling," in *Proceedings of the Social Statistics Section, American Statistical Association*, Washington, DC: American Statistical Association, Pp. 402-405.

Fay, R. E. (1982), "Contingency Tables Analysis for Complex Sample Designs: CPLX," in *Proceedings of the Survey Research Methods Section, American Statistical Association*, Washington, DC: American Statistical Association, Pp 44-52.

Fay, R. E. (1985), "A Jackknifed Chi-Squared Test for Complex Samples," *Journal of The American Statistical Association*, 80: 148-157.

Gifi, A. (1990), *Nonlinear Multivariate Analysis*, New York: Wiley.

Goodman, Leo A. and Kruskal, William H. (1979), *Measures of Association for Cross-Tabulations*, New York: Springer-Verlag.

Green, Paul E., Carmone, Frank J. and Wachspress, David P. (1976), "Consumer Segmentation via Latent Class Analysis," *Journal of Consumer Research*, Vol. 3, December, pp. 170-174.

Green, Paul E. and Carmone, Frank J. (1977) "Segment Congruence Analysis: A method for Analysing Association Among Alternative Bases for Market Segmentation," *Journal of Consumer Research*, 3 (March) 217-222.

Green, Paul E. and DeSarbo, Wayne S. (1981), "Two models for Representing Unrestricted Choice Data," in Kent B. Monroe and Ann Arbor (Ed) *Advances in Consumer Research*, Association for Consumer Research, Vol. 8, pp. 309-312.

Green, Paul E., Schaffer, Catherine M. and Patterson, Karen M., (1988), "A Reduced-Space Approach to the Clustering of Categorical Data in Market Segmentation," *Journal of the Market Research Society*, volume 30, number 3.

Greenacre, Michael J., (1984), *Theory and Application of Correspondence Analysis*, London: Academic Press Inc.

Grover, R. and Dillon W. R., (1988) "Understanding Market Characteristics from Aggregate Brand Switching Data by the Method of Spectral Decomposition", *International Journal of Research in Marketing*, 5 p. 77-89.

Hill, M. O., (1974), "Correspondence Analysis: A Neglected Multivariate Method," *Journal of Applied Statistics*, Vol. 23, No. 3, p. 340.

Hoffman, Donna L. and Franke, George R. (1986), "Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research," *Journal of Marketing Research*, Vol. 23 (August), 213-27.

Holbrook, Morris B. Moore, William L. and Winer, Russell (1982), "Constructing Joint Spaces from Pick-Any Data: A New Tool for Consumer Analysis," *Journal of Consumer Research*, Vol. 9, June.

Iacobucci, D. and Henderson, G., (1997), "Log-Linear Models for Consumer Brand Switching Behaviour: What a Manager Can Learn from studying Standardized Residuals", in *Advance in Consumer Research*, Association for Consumer Research, vol. 24, p. 375-380.

Kaciak, E. and Louviere, J., ((1990), "Multiple Correspondence Analysis of Multiple Choice Experiment Data," *Journal of Marketing Research*, Vol. 27 (November), 455-65.

Kish, Leslie (1965), *Survey Sampling*, John Wiley & Sons, Inc.

Koch, G.G., Freeman, D. H. and Freeman, J. L. (1975). "Strategies in the Multivariate Analysis of Data from Complex Surveys," *International Statistical Review*, 43, 59-78.

Koehler, K. J and Wilson, J. R (1986). "Chi-Square Tests for Comparing Vectors of Proportions for Several Cluster Samples," *Communications in Statistics*, 15, 1235-1249.

Kruskal, J. B. and Wish, M., (1978), *Multidimensional Scaling*, Beverly Hills: Sage Publications.

Kupper, L. and Haseman, J. K. (1978), "The Use of a Correlated Binomial Model for the Analysis of Certain Toxicological Experiments," *Biometrics* 34, 69-76.

Lang, J. B. and Agresti, A. (1994), "Simultaneously Modelling Joint and Marginal Distributions of Multivariate Categorical Responses," *Journal of the American Statistical Association*, 89, 625-632.

- Lang, J. B., McDonald, J. W. and Smith P. W. F., (1999), "Association-Marginal Modelling of Multivariate Categorical Responses: A Maximum Likelihood Approach," *Journal of the American Statistical Association*, Dec. Vol. 94, No. 448.
- Lee, A. J. (1993), "Generating Random Binary Deviates Having Fixed Marginal Distributions and Specified Degrees of Association," *The American Statistician*, August, Vol. 47, No. 3.
- Levine, H. J. (1979), "Joint-Space Analysis of Pick-Any Data: Analysis of Choices From an Unconstrained Set of Alternatives," *Psychometrika*, March, vol. 44, No. 1.
- Lewis, Don and Burke, C. J. (1949) "The Use and Misuse of The Chi-Square Test," *Psychological Bulletin*, Vol.46, No.6.
- Liang, Kung-Yee and Zeger, Scott L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, pp. 13-22.
- Lipsitz, S. R., and Fitzmaurice, G., (1994), "Sample Size for Repeated Measures Studies with Binary Responses," *Statistics in Medicine*, Vol. 13, 1233-1239.
- Loughin, Thomas M. And and Scherer, Peter N. (1998), "Testing Association in Contingency Tables with Multiple Column Responses," *Biometrics*, 54, 630-637.
- Moore, L. William, and Winer, S. Russell (1987), " A Panel-Data Based Method For Merging Joint Space Market Response Function Estimation," *Marketing Science*, vol. 6, No. 1, Winter, 25-42.
- Nishisato, S. (1980), *Analysis of Categorical Data: Dual Scaling and its applications*, University of Toronto Press.
- Nishisato, S.(1994), *Elements of Dual scaling*, Lawrence Elbaum.
- Novak, Thomas P., and Stangor, Charles (1987), "Testing Competitive Market Structures: An Application of Weighted Least Squares Methodology to Brand Switching Data," *Marketing Science*, vol.6, No. 1 Winter, p. 82-97.
- Novak, Thomas P. (1993), "Log-Linear Trees: Models of Market Structure in Brand Switching Data," *Journal of Marketing Research*, 30, 267-287.
- Ochi, Y. and Prentice, R. L. (1984), "Likelihood Inference in a Correlated Probit Regression Model," *Biometrika* 71, 531-543.

Paul, S. R., (1982), "Analysis of Proportions of Affected Foetuses in Teratological Experiments," *Biometrics* 38, 361-370

Petit, Jean-Luc E. (1995), *Market Definition for Underwater Fiber Optic Cable Systems*, M. M. S Thesis, School of Business, Carleton University.

Ramaswamy, V., Chatterjee, R. and Cohen S. H. (1996), "Joint Segmentation on Distinct Interdependence Bases with Categorical Data," *Journal of marketing Research*, 33 (August 337-50).

Ramsay, J. O. (1980), "Some Small Sample Results for Maximum Likelihood Estimation in Multidimensional Scaling," *Psychometrika* 45, pp. 139-144.

Ramsay, J. O (1982) "Some Statistical Approaches to Multidimensional Scaling Data," *Journal of the Royal Statistical Society, A* 145, 285-312.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, 2nd ed., Wiley, New York.

Rao, J. N. K and Scott, A. J. (1981), "The Analysis of Categorical Data from Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables," *Journal of The American Statistical Association* 76:221-30.

Rao, J. N. K and Scott, A. J. (1984), "On Chi-Squared Tests for Multi-Way Tables with Cell Proportions Estimated from Survey Data," *Annals of Statistics* 12:46-60.

Rao, J. N. K and Scott, A. J. (1987). "On simple Adjustments to Chi-Square Tests with Sample Survey Data," *Annals of Statistics*, 15, 385-397.

Rao, J. K. N. and Thomas, D. R. (1988). "The Analysis of Cross-Classified Categorical Data from Complex Sample Surveys," *Sociological Methodology*, 18, 213-269.

Rao, J. N. K. and Scott, A. J. (1992), "A Simple Method for the Analysis of Clustered Data," *Biometrics*, 48, 557-585.

Roberts, G., Rao, J. N. K. and Kumar, S. (1987), "Logistic Regression Analysis of Sample Survey Data," *Biometrika* 74:1-12.

Romesberg, H. Charles (1990), *Cluster Analysis for Researchers*, Malabar, Florida: Krieger.

Rosner, B. (1982), "Statistical Methods in Ophthalmology: An Adjustment for The Intraclass Correlation Between Eyes," *Biometrics*, 38, 105-114.

Rosner, B., (1984), "Multivariate Methods in Ophthalmology With Applications to Other Paired-data Situations," *Biometrics*, 40, 1025-1035.

SAS Institute, Inc. (1990), *SAS/IML Software: Usage and Reference, Version 6, First Edition*, Cary, NC: SAS Institute, Inc.

SAS Institute, Inc. (1990), *SAS Guide to Macro Processing, Version 6, Second Edition*, Cary, NC: SAS Institute, Inc.

Satterwhaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics* 2, 110-114.

Scott, A. J., Rao, J. N. K., and Thomas, D. R. (1990), "Weighted Least-Squares and Quasilikelihood Estimation for Categorical Data Under Singular Models," *Linear Algebra and its Applications*, 127:427-447.

Takane, Y. (1981), "Multidimensional Successive Categories Scaling: a Maximum Likelihood Method," *Psychometrika*, 46, 9-28.

Tenenhaus, M., & Young, F. (1985), "An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis, and Other Methods of Quantifying Categorical, Multivariate Data," *Psychometrika*, 50, 91-119.

Thomas, D. R. and Rao, J. N. K. (1987), " Small-Sample Comparisons of Level and Power for Simple Goodness-of-fit Statistics Under Cluster Sampling," *Journal of the American statistical Association*, 82: 630-636.

Thomas, D. R. (1993), "Inference using Complex Data from Surveys and Experiments," *Canadian Psychology*, 34, no.4 (October).

Thomas, D. R., Singh, A. C. and Roberts, G. R. (1996), "Tests of Independence on Two-Way Tables Under Cluster Sampling: An Evaluation," *International Statistical Review* 64: 295-311

Tukey, J. W. (1958), "Bias and Confidence in Not Quite Large Samples," *Annals of Mathematical Statistics* 29: 614.

Umesh, U. N. (1995), "Predicting Nominal Variable Relationships with Multiple Response," *Journal of Forecasting*, Vol. 14, 585-596.

Umesh, U. N., Tan, M. and Stem, D. (1992), "Analysis of Multiple Response in Marketing Research: Estimating the Degree of Association," *Marketing Letters*, 3, 107-114.

- Vacek, P.M. (1985), "The Effect of Conditional Dependence on The Evaluation of Diagnostic Tests," *Biometrics*, Vol. 41, 959-968.
- Wald, A. (1943), "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society* 54 :426-482.
- Westfall, P. H. and Young, S. S. (1993), *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*, New-York: John Wiley & Sons, Inc.
- Wild, C. J., and Seber, G. A. F. (1993), "Comparing Two Proportions from The Same Survey," *The American Statistician*, 47, No. 4 178-181.
- Wild, C. J., and Seber, G. A. F. (1994), "Comparing Two Proportions from The Same Survey," *The American Statistician*, (Correction) 48, No. 3 p.269.
- Winer, S. Russell and Moore, L. William (1989), "Evaluating the Effects of Marketing-Mix Variables on Brand Positioning," *Journal of Advertising Research*, Vol. 29, No. 1.
- Wolter, K. M. (1985), *Introduction to Variance Estimation*, New York: Springer.
- Zeger, S. L. and Liang, K. (1986), "Longitudinal Data Analysis for Discrete and Continuous Outcomes," *Biometrics* 42, 121-130.
- Zinnes, J. L. and MacKay, D. B. (1983), "Probabilistic Multidimensional Scaling: Complete and Incomplete data," *Psychometrika*, 48,27-48.

Appendix 1

Distributions of Quadratic Forms

Consider $x'Ax$, where $x \sim N(0, V)$ and A is a $p \times p$ symmetric matrix. We have $A = Q\Lambda Q'$ where the columns of Q are the orthonormal eigenvectors of A (which is symmetric) and Λ is a diagonal matrix, with diagonal elements, λ_i , given by the eigenvalues of A . Diagonal elements of Λ are arranged in decreasing order, and columns of Q are arranged accordingly. If A is singular, then an appropriate number of the bottom diagonal elements of Λ will be zero.

We will use the following results: $A^{1/2} = Q\Lambda^{1/2}Q'$ where $\Lambda^{1/2}$ is a diagonal matrix with elements given by $\lambda_i^{1/2}$. Then $A = A^{1/2}A^{1/2}$, with $A^{1/2}$ symmetric.

Case 1 A , non-singular of rank p , V non-singular, of dimension $p \times p$.

Let $y = A^{1/2}x$. Then $x'Ax = y'y$ and $y \sim N(0, A^{1/2}VA^{1/2})$. Consider the decomposition of $A^{1/2}VA^{1/2}$. We have

$$A^{1/2}VA^{1/2} = P\Delta P' = P\Delta^{1/2}\Delta^{1/2}P \quad (1)$$

where P is the matrix of orthonormal eigenvectors of $A^{1/2}VA^{1/2}$ and $\Delta = \text{diag}(\delta_i)$ is the diagonal matrix of eigenvalues of $A^{1/2}VA^{1/2}$. Let

$$Z = \Delta^{-1/2}P'y; \quad (2)$$

then

$$Z \sim N(0, \Delta^{-1/2}P'(P\Delta P')P\Delta^{-1/2}), \quad (3)$$

which reduces to

$$\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}) \quad (4)$$

since $\mathbf{P}'\mathbf{P} = \mathbf{I}$ Now $\mathbf{y}$ can be written as $\mathbf{y} = \mathbf{P}\Delta^{1/2}\mathbf{Z}$ (from (2), so that

$$\mathbf{y}'\mathbf{y} = \mathbf{Z}'\Delta^{1/2}\mathbf{P}'\mathbf{P}\Delta^{1/2}\mathbf{Z} = \mathbf{Z}'\Delta\mathbf{Z}. \quad (5)$$

Therefore, it follows that the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{y}$ is distributed as

$\mathbf{Z}'\Delta\mathbf{Z} = \sum_i^p \delta_i w_i^2$ where w_i^2 are independent single degree of freedom chi-squared random variables. Also, the δ_i are the p positive eigenvalues of $\mathbf{A}^{-1/2}\mathbf{V}\mathbf{A}^{-1/2}$, identical to the eigenvalues of $\mathbf{A}\mathbf{V}$.

Case 2 A Singular (of rank $q < p$), V Non-Singular

We can still define $\mathbf{y} = \mathbf{A}^{-1/2}\mathbf{x}$ as above, though $\mathbf{A}^{-1/2}$ is non-singular; in the non-singular case, at least one of the trailing diagonal elements of $\mathbf{A}$ will be zero. Thus,

$$\mathbf{A}^{-1/2}\mathbf{V}\mathbf{A}^{-1/2} = \mathbf{P}\Delta\mathbf{P}' = (\mathbf{P}_1, \mathbf{P}_2) \begin{pmatrix} \Delta_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{P}_1' \\ \mathbf{P}_2' \end{pmatrix} = \mathbf{P}_1\Delta_1\mathbf{P}_1' \quad (6)$$

where $\mathbf{P}_1$ is the set of eigenvalues corresponding to the non-zero eigenvalues of $\mathbf{A}^{-1/2}\mathbf{V}\mathbf{A}^{-1/2}$, $\mathbf{P}_2$ is the set of eigenvalues corresponding to the zero eigenvalues of $\mathbf{A}^{-1/2}\mathbf{V}\mathbf{A}^{-1/2}$, and Δ_1 is a diagonal matrix having non-zero diagonal elements (the non-zero eigenvalues). Let $\mathbf{Z} = \Delta_1^{-1/2}\mathbf{P}_1'\mathbf{y}$; then

$$\mathbf{Z} \sim N(\mathbf{0}, \Delta_1^{-1/2}\mathbf{P}_1'(\mathbf{P}_1\Delta_1\mathbf{P}_1')\mathbf{P}_1\Delta_1^{-1/2}). \quad (7)$$

But

$$(\mathbf{P}_1', \mathbf{P}_2')(\mathbf{P}_1, \mathbf{P}_2) = \begin{pmatrix} \mathbf{P}_1' \\ \mathbf{P}_2' \end{pmatrix} (\mathbf{P}_1, \mathbf{P}_2) = \begin{pmatrix} \mathbf{P}_1' \mathbf{P}_1 & \mathbf{P}_1' \mathbf{P}_2 \\ \mathbf{P}_2' \mathbf{P}_1 & \mathbf{P}_2' \mathbf{P}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad (8)$$

because $\mathbf{P}$ is an orthonormal matrix. Thus $\mathbf{P}_1' \mathbf{P}_1 = \mathbf{I}_q$ where $q < p$ is the rank of $\mathbf{A}$ and equation (7) reduces to $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_q)$. Now consider $\mathbf{x}' \mathbf{A} \mathbf{x} = \mathbf{y}' \mathbf{y}$. We have

$$\mathbf{P} \mathbf{P}' = \mathbf{I} = (\mathbf{P}_1, \mathbf{P}_2) \begin{pmatrix} \mathbf{P}_1' \\ \mathbf{P}_2' \end{pmatrix} = \mathbf{P}_1 \mathbf{P}_1' + \mathbf{P}_2 \mathbf{P}_2'. \quad (9)$$

Therefore,

$$\mathbf{y}' \mathbf{y} = \mathbf{y}' \mathbf{P}_1 \mathbf{P}_1' \mathbf{y} + \mathbf{y}' \mathbf{P}_2 \mathbf{P}_2' \mathbf{y} \quad (10)$$

Also,

$$\mathbf{z} = \Delta_1^{-1/2} \mathbf{P}_1' \mathbf{y} \Rightarrow \Delta_1^{1/2} \mathbf{z} = \mathbf{P}_1' \mathbf{y} \quad (11)$$

so that

$$\mathbf{y}' \mathbf{y} = \mathbf{z}' \Delta_1^{1/2} \Delta_1^{1/2} \mathbf{z} + \mathbf{y}' (\mathbf{P}_2 \mathbf{P}_2') \mathbf{y} \quad (12)$$

$$= \mathbf{z}' \Delta_1 \mathbf{z} + \mathbf{y}' (\mathbf{P}_2 \mathbf{P}_2') \mathbf{y}. \quad (13)$$

Since

$$\mathbf{y} \sim N(\mathbf{0}, \mathbf{A}^{-1/2} \mathbf{V} \mathbf{A}^{-1/2}) \quad (14)$$

or equivalently

$$\mathbf{y} \sim N(\mathbf{0}, \mathbf{P}_1 \Delta_1 \mathbf{P}_1') \quad (15)$$

it follows that $\mathbf{y} \in \mathcal{C}(\mathbf{P}_1 \Delta_1 \mathbf{P}_1')$, i.e., $\mathbf{y}$ belongs to the column space of $\mathbf{P}_1 \Delta_1 \mathbf{P}_1'$. Thus

$$y = P_1 \Delta P_1' a \quad (16)$$

for some vector a .

Consider $P_2' y$; P_2 is orthogonal to the space of P_1 by definition. Therefore

$$P_2' y = 0. \quad (17)$$

Finally

$$y' y = z' \Delta_1 z. \quad (18)$$

It has already been shown that

$$z \sim N(0, I_q). \quad (19)$$

Hence $x' A x = y' y$ is distributed as $\sum_i^q \delta_i w_i^2$, where w_i^2 are independent single degree of freedom chi-squared random variables. The δ_i are eigenvalues of $A^{1/2} V A^{1/2}$, where $A^{1/2} = Q \Lambda^{1/2} Q'$ and $\Lambda^{1/2}$ is singular. But eigenvalues of $A^{1/2} V A^{1/2}$ are equal to eigenvalues of $A V$ (see proof below), even when A is singular.

Proof:

Let δ_i be eigenvalues of $A^{1/2} V A^{1/2}$. Let φ_i be the eigenvalues of $A V$. Then

$$\text{trace}(A^{1/2} V A^{1/2}) = \sum \delta_i = \text{trace}(A^{1/2} A^{1/2} V) = \sum \varphi_i \quad (20)$$

$$\text{trace}(A^{1/2} V A^{1/2} A^{1/2} V A^{1/2}) = \sum \delta_i^2 = \text{trace}(A V A V) = \sum \varphi_i^2 \quad (21)$$

$$\text{trace}(A^{1/2} V A^{1/2} A^{1/2} V A^{1/2} A^{1/2} V A^{1/2}) = \sum \delta_i^3 = \text{trace}(A V A V A V) = \sum \varphi_i^3 \quad (22)$$

etc.

In general,

$$\text{trace}(\mathbf{A}^{1/2} \mathbf{V} \mathbf{A}^{1/2})^m = \sum \delta_i^m = \text{trace}(\mathbf{A} \mathbf{V})^m = \sum \varphi_i^m, \quad (23)$$

for any integer m . Thus, $\delta_i = \varphi_i$

Case 3 $\mathbf{A}$ non-singular, $\mathbf{V}$ singular, of rank $q < p$.

The proof is the same as for case 2. where all we assume is that $\mathbf{A}^{1/2} \mathbf{V} \mathbf{A}^{1/2}$ is singular, of rank q . Also, eigenvalues of $\mathbf{A}^{1/2} \mathbf{V} \mathbf{A}^{1/2}$ are equal to eigenvalues of

$$\mathbf{A}^{-1/2} \mathbf{A}^{1/2} \mathbf{V} \mathbf{A}^{1/2} \mathbf{A}^{1/2} = \mathbf{V} \mathbf{A} \quad (24)$$

which are equal to the eigenvalues of

$$\mathbf{A} \mathbf{V} \mathbf{A} \mathbf{A}^{-1} = \mathbf{A} \mathbf{V} \quad (25)$$

since $\mathbf{A}$ is non-singular.

Appendix 2

Correction Factor For Test of Equality of Item Probabilities in A One-way Array

A conventional “chi-squared” statistic to test the equality of item probabilities was provided as

$$X^2(E) = n \sum_{j=1}^C \frac{(\mathbf{t}_j - \mathbf{t})^2}{\mathbf{t}} \quad (1)$$

where $\mathbf{t} = \sum_{j=1}^C \mathbf{t}_j / C$. An alternative expression of the same test can be written in matrix form as

$$X^2(E) = (\mathbf{t} - \mathbf{t}\mathbf{1})' \mathbf{D}_{(\mathbf{t}\mathbf{1})}^{-1} (\mathbf{t} - \mathbf{t}\mathbf{1}), \quad (2)$$

where $\mathbf{D}(\mathbf{t}\mathbf{1})$ is a $C \times C$ diagonal matrix with identical elements $\mathbf{t}$. Consider that $\mathbf{t}_j - \mathbf{t} = \mathbf{t}_j - \mathbf{1}'\mathbf{t}/C$, where the vector $\mathbf{t} = (\mathbf{t}_1, \dots, \mathbf{t}_C)$. Then (2) can be written as

$$X^2(E) = [(\mathbf{I} - \mathbf{1}\mathbf{1}'/C)\mathbf{t}]' \mathbf{D}(\mathbf{t}\mathbf{1}) [(\mathbf{I} - \mathbf{1}\mathbf{1}'/C)\mathbf{t}] \quad (3)$$

or equivalently

$$X^2(E) = \mathbf{t}' \mathbf{J} \mathbf{D}(\mathbf{t}\mathbf{1}) \mathbf{J} \mathbf{t} \quad (4)$$

$$= (\mathbf{t}' \mathbf{J} \mathbf{J} \mathbf{t}) / \mathbf{t} = \mathbf{t}' \mathbf{J} \mathbf{t} / \mathbf{t}, \quad (5)$$

where $\mathbf{J} = (\mathbf{I} - \mathbf{1}\mathbf{1}'/C)$ and $\mathbf{J}^2 = \mathbf{J}$.

The variance matrix of $\mathbf{t}$ under the null hypothesis is given by

$$\mathbf{V} = \tau(1 - \tau) \begin{pmatrix} 1 & . & . & \rho_{IC} \\ \rho_{JI} & 1 & . & \rho_{IJ} \\ . & . & 1 & . \\ \rho_{CI} & . & . & 1 \end{pmatrix} \quad (6)$$

Then, assuming the $\mathbf{t}$ has a large sample normal distribution with variance matrix $\mathbf{V}/n$, and using results from Appendix 1, Case 2, the generalized design effects matrix becomes $\mathbf{JV}/\tau$. Therefore the Rao-Scott first-order correction factor can be stated as

$$(C - 1)\bar{\delta} = \text{trace}(\mathbf{JV}/\tau) \quad (7)$$

$$(C - 1)\bar{\delta} = \text{trace}(\mathbf{JV})/\tau = (1/\tau)(1 - \tau) \text{trace} \left\{ (\mathbf{I} - \mathbf{11}'/C) \begin{pmatrix} 1 & . & \rho_{IJ} \\ . & . & . \\ \rho_{IJ} & . & 1 \end{pmatrix} \right\}$$

..

(8)

$$= (1 - \tau) \text{trace} \left\{ \begin{pmatrix} 1 & . & \rho_{IJ} \\ . & . & . \\ \rho_{IJ} & . & 1 \end{pmatrix} - \mathbf{11}'/C \begin{pmatrix} 1 & . & \rho_{IJ} \\ . & . & . \\ \rho_{IJ} & . & 1 \end{pmatrix} \right\} \quad (9)$$

$$= C(1 - \tau) - (1 - \tau)/C \operatorname{trace} \left\{ \mathbf{I} \mathbf{I} \begin{pmatrix} 1 & \cdot & \rho_{ij} \\ \cdot & \cdot & \cdot \\ \rho_{ij} & \cdot & 1 \end{pmatrix} \right\} \quad (10)$$

$$= C(1 - \tau) - (1 - \tau)/C \quad C[(C - 1)\bar{\rho} + 1] \quad (11)$$

$$= (1 - \tau)[(C - (C - 1)\bar{\rho}) - 1] \quad (12)$$

Thus we have

$$(C - 1)\bar{\delta} = (1 - \tau)(C - 1)(1 - \bar{\rho}) \quad (13)$$

which yields a correction given by

$$\bar{\delta}(E) = (1 - \tau)(1 - \bar{\rho}) \quad (14)$$

Again, Rao-Scott theory shows that the distribution of this statistic is a weighted sum of one-degree of freedom chi-squares (see Appendix 1, Case 2), not the familiar χ^2_{C-1} . In this case, the estimated first-order Rao-Scott adjustment factor can equivalently be expressed

$$\hat{\delta}(E) = (1 - \gamma/C)(1 - \hat{\rho}), \quad (15)$$

where $\gamma = m/n$, as before, and $\hat{\rho}$ is the average of the $C(C-1)/2$ correlations among the estimated probabilities, $\hat{p}_j$, $j = 1, \dots, C$.

Appendix 3

Correction factor For A Homogeneity Test

For the general case of R population (rows), the statistic is given in equation (5.9) for two population homogeneity test can be generalized as

$$X^2(H_{SM}) = n \sum_{i=1}^R \theta_i \sum_{j=1}^C \frac{(\hat{t}_{ij} - \hat{t}_j)^2}{\hat{t}_j}, \quad (1)$$

where $\theta_i = n_i/n, i = 1, \dots, R$ and $n = \sum n_i$. In order to use the Rao-Scott methodology $X^2(H_{SM})$ must be written in matrix form, namely

$$X^2(H_{SM}) = (\hat{T} - T)' [\{D(\theta) - \theta\theta'\} \otimes D(\tau)^{-1}] (\hat{T} - T), \quad (2)$$

where $(\hat{T} - T)' = ((\hat{t}_1 - \tau)' , \dots, (\hat{t}_R - \tau)'), \theta$ is an $R \times 1$ vector with elements

$\theta_i, i = 1, \dots, R, D(\theta)$ and $D(\tau)$ are diagonal matrices with elements corresponding to the vectors θ and τ , respectively, and $\otimes$ denotes the Kronecker product. It will be assumed that the vectors of estimated probabilities, $\hat{t}_i$ have large sample normal

distribution with means τ_i (equal to τ under H_0) and variance matrices V_i/n_i . Then, from the results given in Appendix 1, case 2, it can be seen that $X^2(H_{SM})$ will be distributed as a linear combination of $(C - 1)R$ one degree-of-freedom chi-squared variables, where the weights δ_i of the linear combination are the non-zero eigenvalues of the matrix

$D_H = [\{D(\theta) - \theta\theta'\} \otimes D(\tau)^{-1}] V$, where V is a block diagonal matrix of dimension

$RC \times RC$, with blocks given by $\theta_i' V_i$. The first order Rao-Scott correction factor is the

mean of the estimated weights, given by $\bar{\delta}(H_{SM}) = [(R-1)C]^{-1} \text{trace}(\hat{D}_H)$, where $\hat{D}_H$ is

the estimated design effect matrix obtained by replacing τ and V in the expression for $D(H)$ by their estimates $\hat{\tau}$ and $\hat{V}$, respectively. Under the null hypothesis, the diagonals of V_i are given by $\tau_j(1 - \tau_j)$ where τ_j is the j th element of the vector τ . The diagonal elements of $D(\tau)$ is τ_j 's.

We note that

$$\bar{\delta}(H_{SM}) = [(R-1)C]^{-1} \text{trace}(D_H) \quad (4)$$

can also be written as

$$\bar{\delta}(H_{SM}) = [(R-1)C]^{-1} \sum_{i=1}^R (1 - \theta_i) \text{trace}(D_i) \quad (5)$$

where $D_i = D^{-1}(\tau)V_i$. When we evaluate under the null hypothesis, the expression for

$\bar{\delta}(H_{SM})$ becomes

$$\bar{\delta}(H_{SM}) = \sum_{i=1}^R \sum_{j=1}^C (1 - \theta_i)(1 - \tau_j)/(R - 1)C, \quad (5)$$

$$= \sum_{i=1}^R (1 - \theta_i)(C - \sum_j \tau_j)/(R - 1)C \quad (6)$$

$$= (C - \tau_j)(R - 1)/(R - 1)C = 1 - \tau_j/C. \quad (7)$$

This can be estimated as

$$\hat{\delta}(H_{SM}) = 1 - m_{\gamma\gamma}/(n_{\gamma}C) = 1 - \hat{\gamma}/C. \quad (8)$$

Appendix 4

Correction Factor for a Two-Way Multiple by Multiple Test

Let us recall that the modified chi-squared statistic can be written as

$$X_{MOD}^2(I_{MM}) = n(\mathbf{t} - \mathbf{t}_{i.} \otimes \mathbf{t}_{.j})' \mathbf{D}^{-1} (\mathbf{t}_{i.} \otimes \mathbf{t}_{.j}) (\mathbf{t} - \mathbf{t}_{i.} \otimes \mathbf{t}_{.j}) \quad (1)$$

and that the design effects matrix can also be written as

$$\hat{\mathbf{D}}_{MM} = [\mathbf{D}^{-1} (\mathbf{t}_{i.}) \otimes \mathbf{D}^{-1} (\mathbf{t}_{.j})] \hat{\mathbf{V}}_h. \quad (2)$$

where $\hat{\mathbf{V}}_h$ is the variance of $\mathbf{t} - \mathbf{t}_{i.} \otimes \mathbf{t}_{.j}$, which will be obtained by linearization. Let us

note that

$$\mathbf{t} - \mathbf{t}_{i.} \otimes \mathbf{t}_{.j} = \mathbf{t} - [\mathbf{t}_{i.} + (\mathbf{t}_{i.} - \mathbf{t}_{i.})] \times [\mathbf{t}_{.j} + (\mathbf{t}_{.j} - \mathbf{t}_{.j})] \quad (3)$$

$$= \mathbf{t} - \mathbf{t}_{i.} \mathbf{t}_{.j} - \mathbf{t}_{i.} (\mathbf{t}_{.j} - \mathbf{t}_{.j}) - (\mathbf{t}_{i.} - \mathbf{t}_{i.}) \mathbf{t}_{.j} + \dots, \quad (4)$$

where $\dots$ denotes a remainder term of lower order asymptotically. Thus the random term contributing to the variance is given by

$$\mathbf{w} = \mathbf{t} - \mathbf{t}_{i.} \mathbf{t}_{.j} - \mathbf{t}_{i.} \mathbf{t}_{.j}. \quad (5)$$

In other words the terms $n^{1/2}(\mathbf{t} - \mathbf{t}_{i.} \mathbf{t}_{.j})$ and $n^{1/2}(\mathbf{w} - \mathbf{w})$ are asymptotically equivalent.

i.e. they both have mean zero and (scaled) variance matrix $\hat{\mathbf{V}}_h$.

First, let us note that it has been shown that the aggregate table can be expressed in expanded form. Similarly, one can move from the expanded to the corresponding aggregate table. The latter approach will be taken to derive the required variance

expression. It will then be shown that expressions based on the expanded table can be converted to expressions defined only on the aggregate quantities.

Let the vector π denote cell probabilities defined on the expanded table of dimension $R^* \times C^*$ where $R^* = 2^R - 1$ and $C^* = 2^C - 1$. Thus π has dimension $R^*C^* \times 1$. Further, let p denote the vector of estimates of π , defined in terms of the observed cell proportions. Thus $\pi = E(p)$, where $E(\cdot)$ denotes expectation. Next, let A be a matrix that defines additive functions of the π_k 's, (and p_k 's) that equate to the τ_{ij} 's (and t_{ij} 's) of the aggregate table. Thus $\tau = A\pi$ and $t = Ap$ where A has dimension $R^*C^* \times R^*C^*$. Similarly denote by F_I the appropriate $(R \times R^*C^*)$ matrix needed in order to obtain $\tau_{i\cdot}$ from the expanded table, i.e. $\tau_{i\cdot} = F_I\pi$ and $t_{i\cdot} = F_Ip$. F_I is a matrix of zeros and ones of dimension $R \times R^*C^*$ and can be expressed as $(I_R \otimes 1_C')A$. Similarly, F_J is defined as $(1_R' \otimes I_C)A$ in order to obtain $\tau_{\cdot j}$, i.e. $\tau_{\cdot j} = F_J\pi$ and $t_{\cdot j} = F_Jp$.

To get an expression for V_h , we find the variance of $n^{1/2}(\hat{w} - w)$, where $\hat{w}$ can be expressed in terms of expanded table properties as

$$\hat{w} = Ap - F_I\pi(F_Jp) - (F_Ip)F_J\pi \quad (6)$$

This can be written as

$$\hat{w} = Ap - [(F_I\pi) \otimes F_J]p - [F_I \otimes (F_J\pi)]p, \quad (7)$$

from which it follows that we want the covariance matrix of

$$n^{1/2}[A - (F_I\pi) \otimes F_J - F_I \otimes (F_J\pi)]p. \quad (8)$$

Thus the linearized estimate of the covariance matrix of the residual cell probabilities is given by

$$[A - (F, \pi) \otimes F, - F, \otimes (F, \pi)] V [A' - (\pi' F') \otimes F', F' \otimes (\pi' F')] \quad (9)$$

where V is defined via the large sample assumption, $n^{1/2}(\rho - \pi) \xrightarrow{D} N(0, V)$.

Equation (9) contains nine terms, i.e.,

$$\begin{aligned} V_h = & A V A' + (F, \pi \otimes F) V (\pi' F' \otimes F') + (F, \otimes F, \pi) V (F' \otimes \pi' F') \\ & - A V (\pi' F' \otimes F') - A V (F' \otimes \pi' F') - (F, \pi \otimes F) V A' \\ & + (F, \pi \otimes F) V (F' \otimes \pi' F') - (F, \otimes F, \pi) V A' + (F, \otimes F, \pi) V (\pi' F' \otimes F') \end{aligned} \quad (10)$$

The generalized design effects matrix can be expressed as

$$D_{MM} = [D^{-1}(F, \pi) \otimes D^{-1}(F, \pi)] V_h, \quad (11)$$

with V_h as described in equation (10). A closed form expression of the average design effect (i.e., the first-order Rao-Scott correction factor) can be found in terms of the trace of D_{MM} , i.e.,

$$\bar{\delta}(I_{MM}) = (RC)^{-1} \text{trace}(D_{MM}) \quad (12)$$

Note that D_{MM} of equation (11) can be expressed in terms of τ 's, namely

$$D_{MM} = D^{-1}(\tau_i) \otimes D^{-1}(\tau_j) V_h \quad (13)$$

The trace of D_{MM} has nine terms which will be considered separately below.

$$1) \text{ First term} = \text{trace}[D^{-1}(\tau_i) \otimes D^{-1}(\tau_j) V(\tau)] \quad (14)$$

$$= \sum_i \sum_j v(\tau_{ij}) / \tau_i \tau_j \quad (15)$$

since $V(\tau) = A'VA$. Here $v(\tau_{ij})$ is the ij^{th} diagonal elements of $V(\tau)$.

$$2) \text{ Second term} = \text{trace}\{[D^{-1}(F_i, \pi) \otimes D^{-1}(F_j, \pi)] [(F_i, \pi \otimes F_j) V(\pi' F_i' \otimes F_j')]\} \quad (16)$$

$$= \text{trace}\{[D^{-1}(F_i, \pi) F_i, \pi] \otimes [D^{-1}(F_j, \pi) F_j, \pi]\} V(\pi' F_i' \otimes F_j') \quad (17)$$

$$= \text{trace}\{V[\pi' F_i' \otimes F_j'] D^{-1}(F_j, \pi) F_j, \pi]\} \quad (18)$$

$$= (\sum_i \tau_i) \text{trace}[V F_j' D(\tau_j)^{-1} F_j] \quad (19)$$

$$= (\sum_i \tau_i) \sum_j v(\tau_j) / \tau_j \quad (20)$$

where $v(\tau_j)$ is the j^{th} element of the scaled covariance matrix of τ_j .

$$3) \text{ Third term} = \text{trace}\{[D^{-1}(F_i, \pi) \otimes D^{-1}(F_j, \pi)] [(F_i \otimes F_j, \pi) V(F_i' \otimes \pi' F_j')]\} \quad (21)$$

$$= \text{trace}\{[D^{-1}(F_j, \pi) F_j, \pi] \otimes [D^{-1}(F_i, \pi) F_i, \pi]\} V(\pi' F_j' \otimes F_i') \quad (22)$$

$$= \text{trace}\{V[(\pi' F_j' \otimes F_i') D^{-1}(F_i, \pi) F_i, \pi]\} \quad (23)$$

$$= (\sum_j \tau_j) \text{trace}[V F_i' D(\tau_i)^{-1} F_i] \quad (24)$$

$$= (\sum_j \tau_j) \sum_i v(\tau_i) / \tau_i \quad (25)$$

$$4) \text{ Fourth term} = - \text{trace}\{[D^{-1}(F_i, \pi) \otimes D^{-1}(F_j, \pi)] [A V(\pi' F_i' \otimes F_j')]\} \quad (26)$$

$$= - \text{trace}\{AV[\pi'F_i'D^{-1}(F_i\pi)]\otimes[F_j'D^{-1}(F_j\pi)]\} \quad (27)$$

$$= - \text{trace}\{AV[1_i'\otimes F_j'D^{-1}(F_j\pi)]\} \quad (28)$$

$$= -\text{trace}\{A_{ij}VF_j/\tau_j\} = -c(\mathbf{t}_{ij}, \mathbf{t}_{ij})/\tau_j = - \sum_j c(\mathbf{t}_{.j}, \mathbf{t}_{.j})/\tau_j \quad (29)$$

where A_{ij} is the $(ij)^{\text{th}}$ row of A , F_j is the j^{th} row of F , and $c(\mathbf{t}_{ij}, \mathbf{t}_{ij})$ denotes the scaled covariance matrix of τ_j and τ_i .

$$5)\text{fifth term} = - \text{trace}\{[D^{-1}(F_i\pi)\otimes D^{-1}(F_j\pi)][AV(F_i'\otimes \pi'F_j')]\} \quad (30)$$

$$= - \text{trace}\{AV[F_i'D^{-1}(F_i\pi)\otimes \pi'F_j'D^{-1}(F_j\pi)]\} \quad (31)$$

$$= - \sum_j c(\mathbf{t}_{i.}, \mathbf{t}_{i.})/\tau_i \quad (32)$$

by analogy to the fourth term.

6)sixth term: identical to the fourth term, since the trace of a matrix is equal to the trace of its transpose.

$$7)\text{seventh term} = \text{trace}\{[D^{-1}(F_i\pi)\otimes D^{-1}(F_j\pi)][(F_i\pi\otimes F_j)V(F_i'\otimes \pi'F_j')]\} \quad (33)$$

$$= \text{trace}\{[1_i'\otimes D^{-1}(F_j\pi)F_j]V(F_i'\otimes \pi'F_j')\} \quad (34)$$

$$= \text{trace}\{V(F_i'1_i\otimes 1_j'F_j)\} \quad (35)$$

Since the outer product of two vectors is the same as their Kronecker product, this becomes

$$= \text{trace}\{1_j'F_jVF_i'1_i\} = c(\sum_j \mathbf{t}_j, \sum_i \mathbf{t}_i), \quad (36)$$

where c denotes the scaled covariance operator as before.

8) eight term: Identical to the fifth term (trace of transpose).

9) ninth term: Identical to the seventh term (trace of transpose)

Finally,

$$\begin{aligned} (RC)\bar{\delta}(I_{MM}) = & \sum_i \sum_j v(\mathbf{t}_{ij})/\mathbf{t}_{i.}\mathbf{t}_{.j} + (\sum_i \mathbf{t}_{i.})\sum_j v(\mathbf{t}_{.j})/\mathbf{t}_{.j} + (\sum_i \mathbf{t}_{.j})\sum_i v(\mathbf{t}_{i.})/\mathbf{t}_{i.} \\ & - 2\sum_j c(\mathbf{t}_{.j},\mathbf{t}_{.j})/\mathbf{t}_{.j} - 2\sum_i c(\mathbf{t}_{i.},\mathbf{t}_{i.})/\mathbf{t}_{i.} + 2c(\mathbf{t}_{..},\mathbf{t}_{..}) \end{aligned} \quad (37)$$

The estimate of the correction factor, $\hat{\delta}(I_{MM})$, is obtained by substituting estimated probabilities, variances and covariances into equation (37).

Appendix 5

Correction Factor for A Single by Multiple Two-Way Table

The single by multiple case is just a specialization of the multiple by multiple.

With single by multiple tables $t_{.j} = t_{.j}$ and $\sum t_{.j} = 1$. Again, as previously defined, the

design effect matrix is $\hat{D}_{SM} = D^{-1}(t_{.j}) \otimes D^{-1}(t_{.j}) \hat{V}_h$. Because $\sum t_{.j} = 1$ in this case, the

rank of V_h is $(R-1)C$. Thus

$$(R-1)C \bar{\delta}(I_{SM}) = \text{trace} \hat{D}_{SM}. \quad (1)$$

The right hand side of (1) can be obtained from expression (12) of Appendix 4 by

specializing to the case $\sum t_{.j} = 1$. Note also that $t_{.j} = t_{.j}$ for the single by multiple

case. We will now develop the expression $\bar{\delta}(I_{SM})$, or the right hand side of equation (37).

Appendix 4.

1) First term: No change, except that $t_{.j} = t_{.j}$ in this case.

$$2) \text{ Second term} = \sum_j v(\tau_{.j})/\tau_{.j} = \sum_j v(\tau_{.j})/\tau_{.j}$$

(2)

since $\sum t_{.j} = 1$

3)Third term: No change

$$= \sum_j \tau_j \sum_i v(\tau_i)/\tau_i = \sum_j \tau_j \sum_i v(\tau_i)/\tau_i \quad (3)$$

4) Fourth term = $-2 \sum_j v(\tau_j)/\tau_j$

(4)

5)Fifth term: No change

$$= \sum_i c(\tau_i, \tau_i)/\tau_i \quad (5)$$

6)Sixth term = 0 since $\sum_i \tau_i = 1$

As a result, the expression $\bar{\delta}(I_{SM})$ becomes

$$(R - 1)C\bar{\delta} = \sum_i \sum_j v(\tau_j)/\tau_i \tau_j - \sum_j v(\tau_j)/\tau_j + (\sum_j \tau_j) \sum_i v(\tau_i)/\tau_i - 2 \sum_i (\tau_i, \tau_i)/\tau_i \quad (6)$$

Appendix 6

Single by Multiple and Multiple by Multiple Tests via the Jackknife Procedure

Leaving one observation out, we compute the residuals $r_{ij}^k = t_{ij}^k - t_{i.}^k \otimes t_{.j}^k$ using the remaining $n - 1$ observations, where the superscript k refers to a pseudo-replicate. Repeating this procedure n times, omitting a different observation each time, we obtain a matrix of residuals of dimension of $(n \times RC)$ from which V_{jk} is estimated. The jackknife estimate is denoted $\hat{V}_{jk}$

In the multiple by multiple case, this variance matrix is

$$V_{jk} = n^{-2} \sum_{k=1}^n \{t_{ij} - t_{i.} \otimes t_{.j} - \hat{E}(t_{ij} - t_{i.} \otimes t_{.j})\}^{(k)'} \{t_{ij} - t_{i.} \otimes t_{.j} - \hat{E}(t_{ij} - t_{i.} \otimes t_{.j})\}^{(k)} / (n - 1) \quad (1)$$

Again, the design effect matrix is given by

$$\hat{D}_{MM} = D^{-1}(t_i) \otimes D^{-1}(t_j) \hat{V}_{jk} \quad (2)$$

This approach can also be used for both the multiple by multiple and the single by multiple independence cases.

Appendix 7

Control of Type I Error Rates

Table 1

**Percentage of 1000 Independent Monte Carlo Trials in Which the Null Hypothesis of Independence is Rejected When the Null Hypothesis is True for A Multiple by Multiple With $\tau'_{i.}=[0.20\ 0.30\ 0.50\ 0.60]$ and $\tau'_{.j}=[0.20\ 0.30\ 0.50\ 0.60]$,
i.e. $\gamma_{i.} = 1.60$ and $\gamma_{.j} = 1.60$**

Sample Size	$X_U^2(I_{MM})$	$X_{MOD}^2(I_{MM})$	$X_Y^2(I_{MM})$	$X_C^2(I_{MM})$	$X_S^2(I_{MM})$	Wald
50	0	0	0.06	0.081	0.034	0.434
100	0	0	0.064	0.069	0.037	0.164
300	0	0	0.074	0.077	0.054	0.092
600	0	0	0.066	0.066	0.052	0.052
1000	0	0	0.068	0.066	0.059	0.052

Table 2

Percentage of 1000 Independent Monte Carlo Trials in Which the Null Hypothesis of Independence is Rejected When the Null Hypothesis is True for A Multiple by Multiple With $\tau'_{i.}=[0.20\ 0.30\ 0.50\ 0.60]$ and $\tau'_{.j}=[0.30\ 0.50\ 0.70\ 0.90]$, i.e. $\gamma_{i.} = 1.60$ and $\gamma_{.j} = 2.40$

Sample Size	$X_U^2(I_{MM})$	$X_{MOD}^2(I_{MM})$	$X_Y^2(I_{MM})$	$X_C^2(I_{MM})$	$X_S^2(I_{MM})$	Wald
50	0	0	0.104	0.113	0.048	0.355
100	0	0	0.075	0.083	0.044	0.172
300	0	0	0.089	0.093	0.057	0.086
600	0	0	0.078	0.079	0.05	0.064
1000	0	0	0.078	0.077	0.05	0.049

Table 3

Percentage of 1000 Independent Monte Carlo Trials in Which the Null Hypothesis of Independence is Rejected When the Null Hypothesis is True for A Multiple by Multiple With $\tau'_{i.}=[0.20\ 0.30\ 0.50\ 0.60]$ and $\tau'_{.j}=[0.70\ 0.80\ 0.80\ 0.90]$, i.e. $\gamma_{i.} = 1.60$ and $\gamma_{.j} = 3.20$

Sample Size	$X_U^2(I_{MM})$	$X_{MOD}^2(I_{MM})$	$X_Y^2(I_{MM})$	$X_C^2(I_{MM})$	$X_S^2(I_{MM})$	Wald
50	0	0	0.067	0.071	0.034	0.373
100	0	0	0.069	0.065	0.04	0.186
300	0	0	0.058	0.064	0.041	0.09
600	0	0	0.074	0.075	0.063	0.082
1000	0	0	0.064	0.066	0.04	0.06

Table 4

Percentage of 1000 Independent Monte Carlo Trials in Which the Null Hypothesis of Independence is Rejected When the Null Hypothesis is True for A Multiple by Multiple With $\tau'_{i.}=[0.70 \ 0.80 \ 0.80 \ 0.90]$ and $\tau'_{.j}=[0.70 \ 0.80 \ 0.80 \ 0.90]$, i.e. $\gamma_{i.} = 3.20$ and $\gamma_{.j} = 3.20$

Sample Size	$X_U^2(I_{MM})$	$X_{MOD}^2(I_{MM})$	$X_Y^2(I_{MM})$	$X_C^2(I_{MM})$	$X_S^2(I_{MM})$	Wald
50	0	0	0.069	0.082	0.043	0.368
100	0	0	0.066	0.067	0.039	0.193
300	0	0	0.078	0.084	0.045	0.112
600	0	0	0.052	0.055	0.038	0.068
1000	0	0	0.074	0.071	0.053	0.073

Table 5

Percentage of 1000 Independent Monte Carlo Trials in Which the Null Hypothesis of Independence is Rejected When the Null Hypothesis is True for A Multiple by Multiple With $\tau'_{i.}=[0.30 \ 0.50 \ 0.70 \ 0.60]$ and $\tau'_{.j}=[0.30 \ 0.50 \ 0.70 \ 0.90]$, i.e. $\gamma_{i.} = 2.40$ and $\gamma_{.j} = 2.40$

Sample Size	$X_U^2(I_{MM})$	$X_{MOD}^2(I_{MM})$	$X_Y^2(I_{MM})$	$X_C^2(I_{MM})$	$X_S^2(I_{MM})$	Wald
50	0	0	0.098	0.118	0.055	0.419
100	0	0	0.095	0.101	0.056	0.166
300	0	0	0.095	0.099	0.05	0.108
600	0	0	0.108	0.106	0.053	0.081
1000	0	0	0.087	0.089	0.049	0.072

Table 6

Percentage of 1000 Independent Monte Carlo Trials in Which the Null Hypothesis of Independence is Rejected When the Null Hypothesis is True for A Single by Multiple With $\tau' = [0.1679 \ 0.3435 \ 0.3625 \ 0.50 \ 0.3549]$ and $\rho' = [0.33358 \ 0.0611 \ 0.1183 \ 0.4313 \ 0.053]$, i.e. $\gamma_1 = 1.7288$

Sample Size	$X_U^2(I_{SM})$	$X_{MOD}^2(I_{SM})$	$X_Y^2(I_{SM})$	$X_C^2(I_{SM})$	$X_S^2(I_{SM})$	Wald
50	0	0	0.012	0.009	0.001	0.52
100	0	0	0.039	0.051	0.019	0.203
300	0	0.004	0.056	0.059	0.041	0.119
600	0	0.002	0.061	0.064	0.044	0.084
1000	0	0.001	0.053	0.052	0.045	0.069

Table 7

Percentage of 1000 Independent Monte Carlo Trials in Which the Null Hypothesis of Independence is Rejected When the Null Hypothesis is True for A Single by Multiple Naive Test (Agresti and Liu 1999)

Sample Size	$\gamma = 1.77$	$\gamma = 1.20$	$\gamma = 1.80$	$\gamma = 2.4$
50	0.047	0.07	0.045	0.042
100	0.063	0.061	0.056	0.057
300	0.058	0.069	0.062	0.052
600	0.055	0.063	0.061	0.059
1000	0.042	0.06	0.067	0.053

Appendix 8

Generating Correlated Binary Variates

Emrich and Piedmonte (1991) developed an algorithm for generating correlated binary variates regardless of the orientation of the correlations among the variates. Their algorithm was used and will be presented in what follows.

8.1 Description of the Algorithm

Let us generate J binary outcomes, $Y_1, \dots, Y_J$ such that $E(Y_j) = p_j$ and $\text{corr}(Y_j, Y_k) = \delta_{jk}$, where $\{p_j, j = 1, \dots, J\}$ and $\{\delta_{jk}, j = 1, \dots, J-1; k = 2, \dots, J\}$ are given and where $J \geq 2$. δ_{jk} is not free to vary from -1 to 1, given τ_j and τ_k . It is bounded below by

$$\max\{ -[\tau_j \tau_k / (1 - \tau_j)(1 - \tau_k)]^{1/2}, -[(1 - \tau_j)(1 - \tau_k) / \tau_j \tau_k]^{1/2} \} \quad (1)$$

and above by

$$\min\{ [\tau_j(1 - \tau_k) / \tau_k(1 - \tau_j)]^{1/2}, [\tau_k(1 - \tau_j) / \tau_j(1 - \tau_k)]^{1/2} \} \quad (2)$$

These restrictions guarantee that the joint mass function for the outcomes be nonnegative for all outcomes.

The first step of the algorithm consists in solving the equations

$$\Phi[z(\tau_j), z(\tau_k), \rho_{jk}] = \delta_{jk} [\tau_j(1 - \tau_j) \tau_k(1 - \tau_k)]^{1/2} + \tau_j \tau_k \quad (3)$$

for ρ_{jk} , ($j = 1, \dots, j-1$; $k = 2, \dots, k$), where $z(\tau)$ is the p^{th} quantile of the standard normal distribution. The equations are solved for any δ_{jk} satisfying the set of restrictions.

The second step in the algorithm is to generate a J-dimensional multivariate normal random variable,

$$Z = Z_1, \dots, Z_j \quad (4)$$

with the correlation matrix $\Sigma = \rho_{jk}$.

Finally, for $j = 1, \dots, j$, set $Y_j = 1$ if $Z_j \leq z(\tau)$ and set $Y_j = 0$ otherwise.

Appendix 9

SAS Codes Used in The Simulation

9.1 Goodness-of-Fit Test

```

%macro create; %let runs=1000;

Proc iml worksize=900 symsize=900;

sum1={0};sum2={0};sum3={0};sum5={0};

c={0};cc={0};ccc={0};p={0};pp={0};ppp={0};sum6=0;

/* multinomial, p(l)=.1, l=1,2,...7 */

%do l=1 %to &runs;

    n=300;    y&l=rantbl(j(n,1,0),
0.436902,0.253356,0.108633,0.112692,0.048388,0.028233,0.011796);

    b&l=(y&l=j(n,1,1)) ;
    c&l=(y&l=j(n,1,2));
    d&l=(y&l=j(n,1,3));
    e&l=(y&l=j(n,1,4));
    f&l=(y&l=j(n,1,5));
    g&l=(y&l=j(n,1,6));
    h&l=(y&l=j(n,1,7));

    *:bin&l=(b&l||c&l||d&l||e&l||f&l||g&l||h&l)*{1 0 1 0 1 0 1,0 1 1 0 0 1 1,0 0 0 1 1 1 1}';

```

```

mu&i=j(1,nrow(bin&i),1)*(b&i||c&i||d&i||e&i||f&i||g&i||h&i)*{1 0 1 0 1 0 1,0 1 1 0 0 1
1,0 0 0 1 1 1 1}' ;free bb&i cc&i dd&i ee&i ff&i
gg&i HH&i; free b&i c&i d&i e&i f&i g&i h&i;
tau&i=(mu&i)/nrow(bin&i);print tau&i;
tau0={0.436902 0.253356 0.108633 0.112692 0.048388 0.028233 0.011796}*{1 0 1 0 1
0 1,0 1 1 0 0 1 1,0 0 0 1 1 1 1}' ;
XX&i=bin&i- j(nrow(bin&i),1,1)#(J(nrow(bin&i),{1},{1})*tau&i);
COV&i=(XX&i`*XX&i)/(nrow(bin&i)); free XX&i;
WLS&i=nrow(bin&i)#((tau&i-Tau0)*inv(cov&i)*(tau&i-tau0)`);
X2&i=nrow(bin&i)#((tau&i-tau0)*inv(diag(tau0))*(tau&i-tau0)`);
des&i=1-(sum(tau&i)/ncol(bin&i));Rao&i=x2&i/des&i;
dess&i=(inv(diag(tau&i))*cov&i);free cov&i;
c2&i=trace(dess&i*dess&i)/(3*(des&i**2));
Rao2&i=Rao&i/c2&i;
tot&i=(x2&i||rao&i||Rao2&i||wls&i); sum5=sum5+tot&i;
vec&i=(des&i||((trace(dess&i)/3)||c2&i));print vec&i;
com&i=(tot&i)>=j(1,2,7.815
)||((cinv(0.95,ncol(bin&i)/c2&i))||7.815);sum6=sum6+com&i;
free tau&i;
FREE x2&i des&i rao&i wls&i TOT&i C&i CC&i CCC&i P&i PP&i PPP&i COV&i
XX&i;
%end; typec=sum6/&runs; print typec;

```

```

mean=sum5/&runs;print mean;FREE SUM1 SUM2 SUM3 sum5 TYPEC TYPECC
TYPECCC;

%mend create;

%create;

```

9.2 Equality of Item Response Probabilities

```

%macro create; %let runs=1000;

Proc iml worksize=900 symsize=900; sum={0};summ={0}; sun={0};

    /* multinomial, p(i)=.1, i=1,2,...,7 */

%do i=1 %to &runs;

    n=50;

    y&i=rantbl(j(n,1,0),0.009196,
                0.008746,
                0.081475,
                0.009046,
                0.081194,
                0.080014,
                0.730329);

    b&i= (y&i=j(n,1,1)) ;bb&i=sum(b&i);

    c&i= (y&i=j(n,1,2));cc&i=sum(c&i);

    d&i= (y&i=j(n,1,3));dd&i=sum(d&i);

```

```

e&i=(y&i=j(n,1,4));ee&i=sum(e&i);

f&i=(y&i=j(n,1,5));ff&i=sum(f&i);

g&i=(y&i=j(n,1,6));gg&i=sum(g&i);

h&i=(y&i=j(n,1,7));hh&i=sum(h&i);

free y&i;

*;bin&i=(b&i||c&i||d&i||e&i||f&i||g&i||h&i)*{1 0 1 0 1 0 1,0 1 1 0 0 1 1,0 0 0 1 1 1 1}'`;

free b&i c&i d&i e&i f&i g&i h&i;

mu&i=(bb&i||cc&i||dd&i||ee&i||ff&i||gg&i||hh&i);free bb&i cc&i dd&i ee&i ff&i gg&i
HH&i;

m1&i={1 0 1 0 1 0 1}*mu&i`;

m2&i={0 1 1 0 0 1 1}*mu&i`;

m3&i={0 0 0 1 1 1 1}*mu&i`;

tau&i=(m1&i||m2&i||m3&i)/nrow(bin&i); free m1&i m2&i m3&i;

tau0&i=(j(1,3,1)*tau&i`)/ncol(bin&i); tau0&i=j(1,3,tau0&i);print tau&i;

/*tau0&i={0.1287711 0.1301556 0.1458431 0.1307697 0.148065 0.1484893 0.167906
}*{1 0 1 0 1 0 1,0 1 1 0 0 1 1,0 0 0 1 1 1 1}'`; */

XX&i=bin&i- j(nrow(bin&i),1,1)#(J(nrow(bin&i),{1},{1})*tau&i);

COV&i=(XX&i`*XX&i)/(nrow(bin&i));

WLS&i=n*((tau&i-Tau0&i)*inv(cov&i)*(tau&i-tau0&i)`);

X2&i=n#((tau&i-tau0&i)*inv(diag(tau0&i))*(tau&i-tau0&i)`);

co&i=corr(bin&i);print co&i;

n=nrow(bin&i);

```



```

SUM&i=bin&i[+.);
XPX&i=bin&i`*bin&i-SUM&i`*SUM&i/N;
S&i=DIAG(1/SQRT(VECDIAG(XPX&i)));
CORR&i=S&i*XPX&i*S&i; print corr&i;

ncov&i=j(3,1,1)*(tau&i#(1-tau&i))#co&i;
des&i=(1-(sum(tau&i)/ncol(bin&i)))*(1-(((sum(co&i) - trace
(co&i))/6)));ARao&i=x2&i/des&i;
j=i(3)-j(3,1,1)*j(1,3,1)/3;
dess&i=inv(diag(tau&i))*(j*cov&i*j);tr&i=trace(dess&i)/2;free cov&i;
Rao&i=X2&i/tr&i;
c2&i=trace(dess&i*dess&i)/(2*(tr&i)**2);
Rao2&i=Rao&i/c2&i; vec&i=(des&i||tr&i||c2&i);print vec&i;summ=summ+vec&i;
tot&i=(x2&i||Arao&i||Rao&i||Rao2&i||wls&i);print tot&i;sum=sum+tot&i;
com&i=(tot&i)>=(j(1,3,cinv(0.95,2))||cinv(0.95,2/c2&i)||cinv(0.95,2));sun=sun+com&i;
FREE x2&i des&i arao&i dess&i rao&i wls&i TOT&i C&i CC&i CCC&i P&i PP&i
PPP&i COV&i XX&i;

%end; typec=summ/&runs; typecc=sun/&runs; print typec;print typecc;

FREE SUM1 SUM2 SUM3 sum5 TYPEC TYPECC TYPECCC;

%mend create;

%create;

```

:

9.3 Homogeneity Test

```
%macro create; %let runs=1000;

Proc iml worksize=900 symsize=900; sum1={0};sum2={0}; sum3={0};sum5={0};

c={0};cc={0};ccc={0};p={0};pp={0};ppp={0};uu={0}; sum6=0;

/* multinomial, p(i)=.1, i=1,2,...7 */

%do i=1 %to &runs;

seed=8091017; fseed=1017809;n=500;

y&i=rantbl(j(n,1,seed+&i), 0.0815217, 0.2336957, 0.1141304, 0.1358696,
0.076087, 0.1576087, 0.201087);

do&i=ranbin((seed+&i),j(n,1,1),0.4945);

ja&i=ranbin((seed+&i),j(n,1,1),0.7282);

eu&i=ranbin((seed+&i),j(n,1,1), 0.5489);

bin&i=(do&i||ja&i||eu&i); free do&i ja&i eu&i;

b&i= (y&i=j(n,1,1)) ;bb&i=sum(b&i); free b&i;

c&i= (y&i=j(n,1,2));cc&i=sum(c&i); free c&i;

d&i= (y&i=j(n,1,3));dd&i=sum(d&i); free d&i;

e&i= (y&i=j(n,1,4));ee&i=sum(e&i);free e&i;

f&i= (y&i=j(n,1,5));ff&i=sum(f&i); free f&i;

g&i= (y&i=j(n,1,6));gg&i=sum(g&i); free g&i;

h&i= (y&i=j(n,1,7));hh&i=sum(h&i);free h&i;
```

*,

mu&i=(bb&i||cc&i||dd&i||ee&i||ff&i||gg&i||hh&i);free bb&i cc&i dd&i ee&i ff&i gg&i

HH&i;

m1&i={1 0 0 1 1 0 1}*mu&i`;

m2&i={0 1 0 1 0 1 1}*mu&i`;

m3&i={0 0 1 0 1 1 1}*mu&i`;

tau&i=(m1&i||m2&i||m3&i)/nrow(bin&i);

XX&i=bin&i-j(nrow(bin&i),1,1)#(J(nrow(bin&i),{1},{1})*tau&i);

COV&i=(XX&i`*XX&i)/(nrow(bin&i)-1); free XX&i;

*,

fy&i=rantbl(j(n,1,fseed+&i), 0.0815217, 0.2336957, 0.1141304, 0.1358696, 0.076087,

0.1576087, 0.201087);

fdo&i=ranbin((fseed+&i),j(n,1,1),0.4945);

fja&i=ranbin((fseed+&i),j(n,1,1),0.7282);

feu&i=ranbin((fseed+&i),j(n,1,1), 0.5489);

fbin&i=(fdo&i||fja&i||feu&i); free fdo&i fja&i feu&i;

fb&i= (fy&i=j(n,1,1)) ;fbb&i=sum(fb&i); free fb&i;

fc&i= (fy&i=j(n,1,2));fcc&i=sum(fc&i); free fc&i;

fd&i= (fy&i=j(n,1,3));fdd&i=sum(fd&i); free fd&i;

fe&i= (fy&i=j(n,1,4));fee&i=sum(fe&i);free fe&i;

ff&i= (fy&i=j(n,1,5));fff&i=sum(ff&i); free ff&i;

fg&i= (fy&i=j(n,1,6));fgg&i=sum(fg&i); free fg&i;

```
fh&i= (fy&i=j(n,1,7));fhh&i=sum(fh&i);free fh&i;
```

```
*;
```

```
fmu&i=(fbb&i||fcc&i||fdd&i||fee&i||fff&i||fgg&i||fhh&i);free fbb&i fcc&i fdd&i fee&i
```

```
fff&i fgg&i fHH&i;
```

```
fm1&i={1 0 0 1 1 0 1}*fmu&i`;
```

```
fm2&i={0 1 0 1 0 1 1}*fmu&i`;
```

```
fm3&i={0 0 1 0 1 1 1}*fmu&i`;
```

```
ftau&i=(fm1&i||fm2&i||fm3&i)/nrow(fbin&i);
```

```
;
```

```
*;
```

```
XY&i=bin&i- j(nrow(fbin&i),1,1)#(J(nrow(fbin&i),{1},{1})*ftau&i);
```

```
CO&i=(Xy&i`*XY&i)/(nrow(fbin&i)-1); free Xy&i;
```

```
nn=nrow(bin&i)+nrow(fbin&i);
```

```
obs&i=(m1&i||m2&i||m3&i||fm1&i||fm2&i||fm3&i); /*observed values*/
```

```
to&i= (m1&i||m2&i||m3&i)+(fm1&i||fm2&i||fm3&i);
```

```
exp&i= ((nrow(bin&i)||nrow(fbin&i))@to&i)/(nn); /*correct expected values*/
```

```
tou&i= (sum(m1&i||m2&i||m3&i))/(sum(fm1&i||fm2&i||fm3&i));
```

```
free m1&i m2&i m3&i fm1&i fm2&i fm3&i ;
```

```
exu&i=(tou&i`@to&i)/(sum(to&i));/*incorrect expected values*/
```

```
xun&i= ((obs&i-exu&i)##2)/exu&i;XuM&i=sum(xun&i);/*unmodified chi-square*/
```

```
X&i=((obs&i-exp&i)##2)/exp&i;XM&i=sum(x&i);free x&i; /*modified chi-square*/
```

```
yy&i=1 - (sum(to&i)/(nn*ncol(bin&i)));ARao&i=xM&i/yy&i; /*first-order corrections*/
```

```

dess&i=(inv(diag(tau&i+ftau&i))*(cov&i+co&i));

tr&i=trace(dess&i)/3;desq&i=(dess&i*dess&i); asq&i=trace(desq&i)/(3*(tr&i**2));

wls&i=nn*(tau&i-ftau&i)*inv(cov&i+co&i)*(tau&i-ftau&i)';

rao&i=xm&i/tr&i;

Rao2&i=Rao&i/asq&i;

tot&i=(xum&i||Xm&i||ARao&i||Rao&i||Rao2&i||wls&i);sum5=sum5+tot&i;creaTE
TOT&I FROM TOT&I;APPEND FROM Tot&i;

com&i=(tot&i)>={7.815 7.815 7.815 7.815 7.815 7.815};sum6=sum6+com&i;

free obs&i to&i exp&i tou&i exu&i;

*;

%end;typec=sum6/&runs;

print typec;

FREE SUM1 SUM2 SUM3 sum5 TYPEC TYPECC TYPECCC;

%mend create;

%create;

%macro names(name, number);

%do n=1 %to &number;

&name&n

%end;

%mend names;

data cc; set %names(tot,1000);proc print data=cc;run;

proc means data=cc mean std alpha=0.05 lclm uclm; var _numeric_;

```

```
run;
```

9.4 Single by Multiple Test

```
%macro create; %let runs=1000;

Proc iml worksize=900 symsize=900; sum1={0}; sum2={0};sum3={0};

    /* multinomial, p(i)=.1, i=1,2,...7 */

    %do i=1 %to &runs;n=50; tii&i=ranbin(0,j(n,1,1),0.50);/*random row totals*/

y&i=rantbl(j(n,1,0),0.436902,0.253356,0.108633,0.112692,0.048388,0.028233,0.011796

);

b&i= (y&i=j(n,1,1)) ;

c&i= (y&i=j(n,1,2));

d&i= (y&i=j(n,1,3));

e&i= (y&i=j(n,1,4));

f&i= (y&i=j(n,1,5));

g&i= (y&i=j(n,1,6));

h&i= (y&i=j(n,1,7));

*; free y&i;

Kl&i=(b&i||c&i||d&i||e&i||f&i||g&i||h&i); free b&i c&i d&i e&i f&i g&i h&i;

Kl&i=kl&i*{1 0 1 0 1 0 1,0 1 1 0 0 1 1,0 0 0 1 1 1 1}' ;

*;
```

```

tyb&i=(tii&i=j(n,1,1)); tya&i=(tii&i=j(n,1,0));co&i=(tya&i||tyb&i);free tii&i;

mat&i=(co&i`*kll&i);pi&i=(mat&i[1,1:7]||mat&i[2,1:7])`/n; free kll&i;

ffi={1 1 1 0 0 0,0 0 0 1 1 1};

ffj={1 1 0 0 0 0,0 0 1 1 0 0,0 0 0 0 1 1};

tius&i= hdir(co&i,kll&i)*ffi`;

NOBb&i= NROW(co&i);

MEaA&i= co&i[:,]; * BASIC DESCRIPTIVE STATISTICS;

co&i=co&i- J(NOBb&i,{1},{1})*MEaA&i;    * COVARIANCE MATRIX; free

meaa&i;

COVt&i=(co&i`*co&i)/(NOBb&i); free nobb&i tii&i meaa&i;

new&i=(tius&i||co&i); free tius&i co&i;

NOBS= NROW(new&i);

MEA&i= new&i[:,]; * BASIC DESCRIPTIVE STATISTICS;

new&i=new&i- J(NOBS,{1},{1})*MEA&i;    * COVARIANCE MATRIX;

kov&i=(new&i`*new&i)/(NOBS);

covv&i=diag(pi&i)-(pi&i*pi&i`); free mea&i new&i;

obs&i=mat&i*{1 0 1 0 1 0 1,0 1 1 0 0 1 1,0 0 0 1 1 1 1}`;/*observed values*/

free mat&i;

ob&i=obs&i[1,1:3] ||obs&i[2,1:3];

to&i= obs&i[1,1:3] +obs&i[2,1:3];

ep&i= ((sum(tya&i)||sum(tyb&i))@to&i)/(n); ti&i= (sum(tya&i)||sum(tyb&i))/n;free

tya&i tyb&i kll&i; print ti&i;

```

```

tou&i= sum(obs&i[1,1:3])/sum(obs&i[2,1:3]); free obs&i;
obbs&i=(tou&i`@to&i)/sum(tou&i);free tou&i;
xun&i= ((ob&i-obbs&i)##2)/obbs&i;XuM&i=sum(xun&i);/*unmodified chi-square*/
free xun&i;
X&i=((ob&i-ep&i)##2)/ep&i;XM&i=sum(x&i);free x&i; /*modified chi-square*/free
x&i;
free obbs&i;
yy&i=1 - (sum(to&i)/(n*3));yRao&i=xM&i/yy&i; /*first-order corrections using the
homogeneity design effects*/

tj&i=(j(1,n,1)*kl&i)/n; print tj&i;
NO&i= NROW(kl&i);
ME&i= kl&i[:,]; * BASIC DESCRIPTIVE STATISTICS;
zy&i=kl&i- J(NO&i,{1},{1})*ME&i; * COVARIANCE MATRIX; free me&i;
COVz&i=(zy&i`*zy&i)/(NO&i); free no&i zy&i me&i;
ts={1 0 1 0 1 0 1,0 1 1 0 0 1 1,0 0 0 1 1 1 1};
eij=block(ts,ts);
fi=i(2)@j(1,7,1); fj=j(1,2,1)@{1 0 1 0 1 0 1,0 1 1 0 0 1 1,0 0 0 1 1 1 1};
Di&i=diag(ti&i@tj&i); *ctij&i=eij*covv&i*eij`;
des&i=(trace(inv(diag(ti&i@tj&i))*(eij*covv&i*eij`))- trace(inv(diag(tj&i))*covz&i)
+sum(tj&i)* trace(inv(diag(ti&i))*covt&i)
-2*trace(inv(diag(ti&i))*kov&i[1:2,3:4]))/3;

```



```

/*Vres&i=eij*covv&i*eij`+((fi*pi&i)@fj)*covv&i*((fi*pi)@fj)`
+(fi@(fj*pi&i))*covv&i*(fi@(fj*pi&i))`
-eij*covv*((fi*pi)@fj)`-eij*covv*(fi@(fj*pi))`-((fi*pi)@fj)*covv*eij`
+((fi*pi&i)@fj)*covv&i*(fi@(fj*pi&i))`-(fi@(fj*pi&i))*covv&i*eij`
+(fi@(fj*pi&i))*covv&i*((fi*pi&i)@fj)`;

print vres&i; */

vres&i=(eij-(fi*pi&i)@fj- fi@(fj*pi&i))*covv&i*(eij-(fi*pi&i)@fj- fi@(fj*pi&i))`;

deff&i=inv(di&i)*vres&i;

DESQ&i=TRACE(deff&i*deff&i); sec&i=yy&i*(sum(tj&i)/sum(ti&i));

xse&i=xm&i/sec&i;

DESS&i=TRACE(deff&i)/3; ASQ&i=(DESQ&i)/(3*(DESS&i)**2);free desq&i deff&i;

wald&i=n*((ob&i/n)-(ep&i/n))[1,1:3]*inv(vres&i[1:3,1:3])*((ob&i/n)-(ep&i/n))[1,1:3]`;

/*waldd&i=n*((ob&i/n)-(ep&i/n))[1,4:6]*inv(vres&i[4:6,4:6])*((ob&i/n)-(ep&i/n))[1,4:6
]`;*/free obs&i ep&i covv&i;

free ob&i ep&i;

Rao1&i=xm&i/dess&i;rao2&i=rao1&i/asq&i; vec&i=(yy&i||dess&i||asq&i);

sum2=sum2+vec&i;free vec&i;

all&i=(xum&i||xm&i||yrao&i||Rao1&i||rao2&i||wald&i);

co&i=(all&i)>=(j(1,4,7.815)||((cinv(0.95,3/asq&i))||7.815));sum1=sum1+co&i;free co&i

xum&i xm&i rao&i raop&i; ;

free yy&i des&i dess&i asq&i;

%end;typec=sum1/&runs;typecc=sum2/&runs; gamma=sum3/&runs;

```

```
print typec;print typecc;
```

```
FREE SUM1 ;
```

```
%mend create;
```

```
%create;
```

9.5 Multiple by Multiple Test

```
%macro create; %let runs=1000;
```

```
Proc iml worksize=900 symsize=900; reset fw=9; sum1={0}; sum2={0};sum3={0};
```

```
/* multinomial, p(i)=.1, i=1,2,...7 */
```

```
%do i=1 %to &runs;n=1000; /*random row totals*/
```

```
*seed=(1:n)#&i; /*seeda=j(1,n,(floor(1000000000*(sqrt(time())))));*/
```

```
vfy&i=rantbl(j(n,1,0),0.436902,0.253356,0.108633,0.112692,0.048388,0.028233,0.0117
```

```
96);
```

```
vfb&i= (vfy&i=j(n,1,1)) ;
```

```
vfc&i= (vfy&i=j(n,1,2));
```

```
vfd&i= (vfy&i=j(n,1,3));
```

```
vfe&i= (vfy&i=j(n,1,4));
```

```
vvff&i= (vfy&i=j(n,1,5));
```

```
vfg&i= (vfy&i=j(n,1,6));
```

```
vfh&i= (vfy&i=j(n,1,7));
```

```
free vfy&i;
```

```

*:fkk&i=(vfb&i||vfc&i||vfd&i||vfe&i||vvff&i||vfg&i||vfh&i);
free vfb&i vfc&i vfd&i vfe&i vvff&i vfg&i vfh&i ;

fk&i=fkk&i*{0 0 0 1 1 1 1,0 1 1 0 0 1 1,1 0 1 0 1 0 1}`;

y&i=rantbl(j(n,1,0),0.013949,0.053511,0.127448,0.024619,0.056448,0.217173,0.506803
);

b&i= (y&i=j(n,1,1)) ;

c&i= (y&i=j(n,1,2));

d&i= (y&i=j(n,1,3));

e&i= (y&i=j(n,1,4));

f&i= (y&i=j(n,1,5));

g&i= (y&i=j(n,1,6));

h&i= (y&i=j(n,1,7));

*: free y&i; kf&i= (b&i||c&i||d&i||e&i||f&i||g&i||h&i);

free b&i c&i d&i e&i f&i g&i h&i;

k&i= kf&i*{0 0 0 1 1 1 1,0 1 1 0 0 1 1,1 0 1 0 1 0 1}`;

fin&i= fkk&i`*kf&i;free fkk&i kf&i;

tius&i=hdir(k&i,fk&i)*(j(1,3,1)@i(3))`;

tjus&i=hdir(k&i,fk&i)*(i(3)@j(1,3,1))`;

NOB&i= NROW(tjus&i);new&i=(tjus&i||k&i);newm&i=new&i[:,];

new&i=new&i-j(nrow(new&i), 1,1)*newm&i;

cone&i=(new&i`*new&i)/(nrow(new&i));

nob&i=nrow(tius&i);anew&i=(tius&i||fk&i);free tjus&i tius&i newm&i` new&i;

```

```
anwm&i=new&i[:,];new&i=new&i-j(nrow(new&i), 1,1)*anwm&i;
```

```
acon&i=(new&i`*new&i)/(nrow(new&i));free fk&i k&i new&i anwm&i new&i;
```

```
fmk&i={0    0    1,
```

```
    0    1    0,
```

```
    0    1    1,
```

```
    1    0    0,
```

```
    1    0    1,
```

```
    1    1    0,
```

```
    1    1    1
```

```
}`*fin&i*j(7,1,1);
```

```
km&i=j(1,7,1)*fin&i*{
```

```
    0    0    1,
```

```
    0    1    0,
```

```
    0    1    1,
```

```
    1    0    0,
```

```
    1    0    1,
```

```
    1    1    0,
```

```
    1    1    1});
```

```
ob&i= {0    0    1,
```

```
    0    1    0,
```

```
    0    1    1,
```

```

1      0      0,
1      0      1,
1      1      0,
1      1      1}'*fin&i*{0      0      1,
                                0      1      0,
                                0      1      1,
                                1      0      0,
                                1      0      1,
                                1      1      0,
                                1      1      1};

```

```

pi&i=(fin&i[1,1:7]||fin&i[2,1:7]||fin&i[3,1:7]||fin&i[4,1:7]||fin&i[5,1:7]||fin&i[6,1:7]||fin
&i[7,1:7])'/n;

```

```

free fin&i;

```

```

exp&i=(km&i@fmk&i)/n;free km&i fmk&i;

```

```

obs&i=ob&i[1,1:3]||ob&i[2,1:3]||ob&i[3,1:3];

```

```

ep&i=exp&i[1,1:3]||exp&i[2,1:3]||exp&i[3,1:3];

```

```

*;to&i=j (1,3,1)*ob&i;

```

```

tou&i= ob&i*j(3,1,1);free ob&i;

```

```

obbs&i=(tou&i`@to&i)/sum(tou&i);free to&i tou&i;

```

```

xun&i=((obs&i-obbs&i)##2)/obbs&i;XuM&i=sum(xun&i);free xun&i; /*unmodified

```

```

chi-square*/ free xun&i;

```

```

X&i=((obs&i-ep&i)##2)/ep&i;XM&i=sum(x&i);free x&i; /*modified chi-square*/free

```

```

x&i;

eij&i={0 0 0 1 1 1 1,0 1 1 0 0 1 1,1 0 1 0 1 0 1}@{0 0 0 1 1 1 1,0 1 1 0 0 1 1,1 0 1 0 1 0
1};

fi&i={0 0 0 1 1 1 1,0 1 1 0 0 1 1,1 0 1 0 1 0 1}@j(1,7,1); fj&i=j(1,7,1)@{0 0 0 1 1 1 1,0
1 1 0 0 1 1,1 0 1 0 1 0 1};

covv&i=diag(pi&i)-(pi&i*pi&i);

*ctij&i=eij&i*covv&i*eij&i`/*set of proofs ap&i=fi&i*pi&i;print
ap&i;app&i=fj&i*pi&i;print app&i;appp&i=(j(1,n,1)*fk&i)/n;print appp&i;
apppp&i=(j(1,n,1)*k&i)/n;print apppp&i; */

covz&i=fj&i*covv&i*fj&i`; coz&i=fi&i*covv&i*fi&i`;

*cov&i=eij&i*covv&i*eij&i`;

ti&i=fi&i*pi&i;tj&i=fj&i*pi&i;free pi&i;

cof&i=sum(fi&i*covv&i*fj&i`);

des&i=(trace(inv(diag(ti&i@tj&i))*diag(eij&i*covv&i*eij&i`))+sum(ti&i)*trace(inv(dia
g(tj&i))*(covz&i))+
sum(tj&i)*trace(inv(diag(ti&i))*(coz&i))
-2*trace(inv(diag(ti&i))*(acon&i[1:3,4:6]))-
2*trace(inv(diag(tj&i))*(cone&i[1:3,4:6]))+2*cof&i)/9;free coz&i covz&i cof&i ctij&i
acon&i cone&i;

/*design effects computed using the formula for multiple by multiple*/

tott&i=diag(ti&i@tj&i);

/*Vres&i=eij&i*covv&i*eij&i`+((fi&i*pi&i)@fj&i)*covv&i*((fi&i*pi&i)@fj&i)`

```

```

+(f&i@(fj&i*pi&i))*covv&i*(f&i@(fj&i*pi&i))`
-eij&i*covv&i*((f&i*pi&i)@fj&i)`-eij&i*covv&i*(f&i@(fj&i*pi&i))`
-((f&i*pi&i)@fj&i)*covv&i*eij&i`
+((f&i*pi&i)@fj&i)*covv&i*(f&i@(fj&i*pi&i))`-(f&i@(fj&i*pi&i))*covv&i*eij&i`
+(f&i@(fj&i*pi&i))*covv&i*((f&i*pi&i)@fj&i)`;*/
vres&i=(eij&i-(ti&i)@fj&i-f&i@(tj&i))*covv&i*(eij&i-(ti&i)@fj&i-f&i@(tj&i))`;
free eij&i f&i fj&i covv&i;
gam&i=(1-sum(tj&i)/3)*(1-sum(ti&i)/3);
Xgam&i=xm&i/gam&i; sec&i=gam&i*(sum(tj&i)/sum(ti&i));free ti&i tj&i;
deff&i=(inv(tott&i)*vres&i); free tott&i;desq&i=trace(deff&i*deff&i);
DESS&i=trace(DEFF&i)/9;free deff&i; ASQ&i=(DESQ&i)/(9*(DESS&i**2));free
desq&i;rao1&i=xm&i/dess&i;rao2&i=rao1&i/asq&i;
wald&i=n*(obs&i/n-ep&i/n)*inv(vres&i)*(obs&i/n-ep&i/n)`; free obs&i
ep&i;vec&i=(gam&i||des&i||dess&i||asq&i);sum2=sum2+vec&i;
free to&i exp&i tou&i; all&i=(xum&i||xm&i||xgam&i||Rao1&i||Rao2&i||Wald&i);free
xum&i xm&i rao1&i rao2&i;
free vec&i gam&i des&i dess&i; sum3=sum3+all&i;print all&i;
co&i=(all&i)>=(j(1,4,16.919)||cinv(0.95,9/asq&i)||16.919);sum1=sum1+co&i;free all&i
asq&i co&i seed fseed;
%end;typec=sum1/&runs; typecc=sum2/&runs;typeccc=sum3/&runs;
print typec; print typecc;print typeccc;
FREE sum1 ;

```

```
%mend create;
```

```
%create;
```

9.6 Generating Correlated Binary Data

```
%macro create; %let runs=1;
```

```
Proc iml worksize=900 symsize=900;
```

```
%do i=1 %to &runs;
```

```
seed=0;
```

```
n=1000;*Number of observations;
```

```
dim =3;*number of categories;
```

```
marg={0.30,0.60,0.90};*Marginal probabilities;
```

```
COREST={1 -0.04317 -0.03743,-0.04317 1 -0.02520,-0.03743 -0.02520 1};
```

```
*Correlation matrix;
```

```
*Subroutine of bivariate standard normal CDF;
```

```
Start binorcdf(pp1,pp2,rhoti);
```

```
za=probit(pp1);
```

```
zb=probit(pp2);
```

```
numint=1000;
```

```
interval=(zb+5)/numint;
```

```
z2vec=cusum({-5}||J(1,numint-1,interval))||zb;
```

```
fvec=probnorm((za-rhoti*z2vec)/(1-rhoti##2))#
```

```
((1/sqrt(2*3.1415926))*exp(-(z2vec##2)/2));
```



```

coevec={1}||repeat({4 2},1,numint/2-1)||{4 1};

intf=(1/3)*interval*coevec*fvec`;

Return(intf);

finish;

*****;

Start comor(pp,cor,n);

comor=J(n,n,0);

do i=1 to n-1;do j=i+1 to n;

pp1=pp[i];pp2=pp[j];

delta=cor[i,j];

qq1=1-pp1;qq2=1-pp2;

check=min(sqrt(pp1*qq2/(pp2*qq1)),sqrt(pp2*qq1/(pp1*qq2)));

if delda > check

then print "ERROR, NEED TO FIX DELTA" ;

goal=sqrt(pp1*(1-pp1)*pp2*(1-pp2))*delta+pp1*pp2;

low=0;high=1;

error=1;iternum=1;

:
do while (error > 0.0001 & iternum < 50);

mid=(low+high)/2;

cdfmid=binorcdf(pp1,pp2,mid);

if cdfmid > goal then high=mid;

else if cdfmid < goal then low=mid;

```

```

rhot=mid;

iternum=iternum+1;

error=abs(cdfmid -goal);

end;

cornor[i,j]=rhot;

end;end;

cornor=cornor+cornor`+I(n);

If any(eigval(cornor) <0) then print "cornor is not npd";

Return(cornor);

finish;

*****;

*FIND THE COR FOR GENERATING Y1. .... YL IN EMRICH AND PIEDMONTE
(1991);

cornor1=cornor(marg,correst,dim);

yobs&i=( (half(cornor1)`*normal(J(dim,n,seed))) <= (probit(marg)*J(1,n,1)) );

print yobs&i ;*/ kl&l=yobs&i`;free obs&i;

*****

```