

Marginal Loglinear Models for Three Multiple-Response Categorical Variables

by

Thilini Surendra

B.Sc., University of Sri Jayewardenepura, 2012

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

© **Thilini Surendra 2016**
SIMON FRASER UNIVERSITY
Fall 2016

All rights reserved.

However, in accordance with the *Copyright Act of Canada*, this work may be reproduced without authorization under the conditions for “Fair Dealing.” Therefore, limited reproduction of this work for the purposes of private study, research, education, satire, parody, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: Thilini Surendra

Degree: Master of Science (Statistics)

Title: *Marginal Loglinear Models for Three
Multiple-Response Categorical Variables*

Examining Committee: **Chair:** Tim Swartz
Professor

Thomas Loughin
Senior Supervisor
Professor

Rachel Altman
Supervisor
Associate Professor

Joan Hu
Examiner
Professor

Date Defended: 9 December 2016

Abstract

A lot of survey questions include a phrase like, “Choose all that apply”, which lets the respondents choose any number of options from predefined lists of items. Responses to these questions result in multiple-response categorical variables (MRCVs). This thesis focuses on analyzing and modeling three MRCVs. There are 232 possible models representing different combinations of associations. Parameters are estimated using generalized estimating equations generated by a pseudo-likelihood and variances of the estimators are corrected using sandwich methods. Due to the large number of possible models, model comparisons based on nested models would be inappropriate. As an alternative, model averaging is proposed as a model comparison tool as well as to account for model selection uncertainty. Further the calculations required for computing the variance of the estimators can exceed 32-bit machine capacity even for a moderately large number of items. This issue is addressed by reducing dimensions of the matrices.

Keywords: Multiple-response categorical variables; loglinear models; pseudo-likelihood; model averaging

Acknowledgements

I would first like to thank my senior supervisor Prof. Tom Loughin for his continuous support and guidance throughout the past two years. Thanks for bearing with me and patiently showing me the correct path to proceed the research. Also, I should thank Prof. Tim Swartz for accepting me as a graduate student to the department. And I'm grateful to all the professors who lectured me during the past two years. I very much appreciate the financial support and research suggestions provided by Procter and Gamble.

Being an expecting mom is quite challenging at this stage of academic life. But I was managed to balance both the academic and personal life thanks to my beloved husband Menuka. His immense support and caring gave me a peaceful mind to concentrate on the studies. Also I should thank our little baby girl 'Tiara' who's on the way, for not giving me a hard time throughout the pregnancy. I'm grateful to my beloved parents and in-laws who always loved me unconditionally and cared for me even though we are oceans apart.

I should thank my lab colleagues; Will, Andrew and Sharla for their helpful comments on my research. Their useful feedback shaped my research in a better way. Last but not least, thanks to all my friends for their support and caring.

Table of Contents

Approval	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
1 Introduction	1
2 Review of the literature	4
2.1 Testing for association between an SRCV and an MRCV	4
2.2 Testing for association between 2 MRCVs	5
2.2.1 Representation of MRCVs	6
2.2.2 Specification of SPMI	7
2.2.3 Modified test statistic and its sampling distribution	9
2.3 Modeling an SRCV and an MRCV	10
2.4 Modeling two MRCVs	11
2.4.1 Identification of models	12
2.4.2 Inference for generalized loglinear models	13
2.4.3 Model comparisons	15
3 Models for three MRCVs	16
3.1 Model building process for three MRCVs	16
3.1.1 Odds ratios within a subcube	16
3.1.2 Odds ratios across the subcubes	18
3.1.3 Model building for three MRCVs	19
3.2 Inference on the models	22
3.3 Model comparisons	23

3.3.1	Takeuchi Information Criterion	23
3.3.2	Model Averaging	25
4	Computation of variance in the presence of large number of items	28
4.1	Limitations of the current method	28
4.2	Solution to the problem	28
4.3	Example	31
5	Analysis Example	33
5.1	Description of the dataset	33
5.2	Fitting the models	36
5.3	Model averaging	36
6	Conclusions and discussion	39
	Bibliography	41
	Appendix A An R Code to compare run times between two methods	42
	Appendix B An R Code to compute model averaged ORs and 95% confidence intervals	46

List of Tables

Table 2.1	Veterinary information sources and education groups. Source: Loughin and Scherer (1998).	4
Table 2.2	Waste storage methods and Veterinary information sources. Source: Bilder and Loughin (2004).	5
Table 2.3	Item response table.	6
Table 2.4	Joint table for Kansas farmer data.	8
Table 3.1	Notation for conditional ORs for each face within a (W_i, Y_j, Z_k) subcube.	17
Table 3.2	All possible models for three MRCVs.	19
Table 3.3	Models for each cell of (W_i, Y_j, Z_k) th subcube.	20
Table 4.1	Joint table for a hypothetical situation.	30
Table 4.2	Cross-classification of row and column positions of non-zero joint counts.	30
Table 4.3	Run times (seconds) for computing variance using two different approaches. An “x” means that the program terminated with an error code “Cannot allocate vector of size xxxGB”.	31
Table 5.1	Table of observed conditional ORs along with 95% confidence intervals (in parentheses). Highlighted cells denote ORs that do not include 1.	34
Table 5.2	Table of observed conditional ORs for each subcube. Highlighted cells denote significant ORs.	35
Table 5.3	Table of TIC and evidence weights when 3-way association is assumed within a subcube, and this association is allowed to vary across the subcubes according to the pattern given in the first column.	37
Table 5.4	Table of model-averaged conditional ORs along with 95% confidence intervals (in parentheses). Highlighted cells denote ORs that do not include 1.	37

List of Figures

Figure 3.1	Item response table for three MRCVs. (Source: http://3.bp.blogspot.com/-o_nuM70hxIk/UEkpqCunRII/AAAAAAAAAPc/mrd5CfbQDxI/s1600/234.jpg)	17
Figure 5.1	Cumulative weights of top 16 models.	38

Chapter 1

Introduction

Categorical data occur in response to many kinds of data collection schemes. For example, surveys often ask respondents to pick one response to a question out of a list of possible items. Such variables are called single-response categorical variables (SRCVs).

A lot of survey questions include a phrase like, “Choose all that apply” (CATA), which lets the respondents choose any number of options from predefined lists. Typically, they may choose any combination of options or sometimes may not select anything at all. Responses to CATA questions result in categorical variables that are known as multiple-response categorical variables (MRCVs) (Bilder and Loughin, 2004) because they have multiple response options for each subject.

An example that involves MRCVs is described in Bilder and Loughin (2007). A survey was conducted among Kansas farmers. They were asked the following questions about their swine waste management practices:

1. Choose all swine waste storage methods used
 - (a) lagoon
 - (b) pit
 - (c) natural drainage
 - (d) holding tank
2. Choose all contaminants tested for
 - (a) nitrogen
 - (b) phosphorus

- (c) salt

Research questions of interest given the above two MRCVs are (1) is waste storage independent of what contaminants it's tested for and (2) if they are dependent, what is the association structure?

The regular methods of analyzing categorical variables such as generalized linear models (GLMs) assume independence of responses. However, since respondents may provide multiple responses to a CATA question, these responses may be correlated. One should not simply ignore the within-subject dependence and analyze the responses as if they were independent (Bilder and Loughin, 2007). Furthermore, the within-subject association can take any arbitrary form, including mixed positive and negative, which makes certain models for clustered data and repeated measurements such as generalized linear mixed models (GLMMs) and generalized estimating equations (GEEs) questionable (Bilder and Loughin, 2014).

Numerous authors have considered the analysis and modeling of two MRCVs (see Bilder and Loughin, 2007 for a summary). This thesis focuses on analyzing associations among three MRCVs. For example, the survey of the Kansas farmers contains another MRCV which gives sources of veterinary information.

3. Choose all sources of veterinary information
 - (a) professional consultant
 - (b) veterinarian
 - (c) state or local extension service
 - (d) magazines
 - (e) feed companies and representatives

It is conceivable that the source(s) that farmers use for information about swine waste storage and testing might impact their practices, leading to possible complex interactions among the three variables. Generally, these questions can take the form,

1. Are the three MRCVs independent of each other?
2. If not, is there an association between any two MRCVs?
 - (a) If so, does the association between the two MRCVs change across the levels of the third MRCV?

The organization of the thesis is as follows. Chapter 2 reviews the literature regarding analysis of two MRCVs. Chapter 3 generalizes the analysis of two MRCVs to three MRCVs.

Chapter 4 discusses computational considerations that arise during the analysis of MRCVs in the presence of large numbers of options. Chapter 5 applies the models to a real world dataset, and Chapter 6 presents conclusions and discussions.

Chapter 2

Review of the literature

2.1 Testing for association between an SRCV and an MRCV

Loughin and Scherer (1998) first introduce testing for associations between an SRCV and an MRCV. They report on a survey that asked a sample of 262 Kansas farmers, “What are your primary sources of veterinary information? Choose all that apply”. The outcome categories were A) professional consultant, B) veterinarian, C) state or local extension service, D) magazines and E) feed companies and representatives. Further, the farmers were asked about their highest level of education. Table 2.1 provides response counts when veterinary information source is cross-classified with the education.

Table 2.1: Veterinary information sources and education groups. Source: Loughin and Scherer (1998).

Education	Information Source					Total
	A	B	C	D	E	
High School	19	38	29	47	40	173
Vocational	2	6	8	8	4	28
2 yr college	1	13	10	17	14	55
4 yr college	19	29	40	53	29	170
Others	3	4	8	6	6	27
Total	44	90	95	131	93	453

Notice that the 262 farmers provide a total of 453 responses in this table. Since an individual may contribute to multiple column items, the counts of Table 2.1 may be correlated.

Loughin and Scherer (1998) argue that the usual Pearson chi-square test statistic for the null hypothesis of independence between the row and column variables has a distribution

that is no longer a chi-square distribution. They present a modification to the Pearson test statistic that compares the observed cell counts in the contingency table with their proper expected values under the null hypothesis. They show that its null asymptotic distribution of this modified statistic is that of a linear combination of chi-square random variables, each having 1 degree of freedom (df). Extremely large sample sizes may be needed to make the asymptotic distribution a reasonable approximation to the finite-sample distribution of the modified test statistic. Therefore they propose using a bootstrap as an alternative to find the sampling distribution of the modified test statistic.

2.2 Testing for association between 2 MRCVs

Bilder and Loughin (2004) introduce testing for independence between two MRCVs. They use additional data from the same survey reported in Loughin and Scherer (1998), where the same Kansas farmers were asked about their swine waste storage practices. The farmers were allowed to select as many responses as applied from a list of options. Table 2.2 summarizes the response counts for each category combination of the two MRCVs.

Table 2.2: Waste storage methods and Veterinary information sources. Source: Bilder and Loughin (2004).

Waste storage methods	Veterinary information sources				
	Professional consultant	Veterinarian	State service	Magazines	Feed companies
Lagoon	34	54	50	63	41
Pit	17	33	34	43	37
Natural drainage	6	23	30	49	34
Holding tank	1	4	4	6	2

A total of 279 farmers provided responses to both of these questions, and since they are allowed to select any number of options from the lists, they may contribute to more than one cell in the contingency table. As a result, the total number of responses adds up to 565 in Table 2.2, which again exceeds the sample size. Correlated counts again invalidate the asymptotic distribution of a Pearson statistic applied to the counts in this table.

Agresti and Liu (1999) also consider the problem of testing for independence between two MRCVs. They refer to the categories for each MRCV as *items*, and note that representing MRCVs using a regular contingency table presents only the counts of those who responded positively to each item. It provides no information on the respondents who responded

negatively to a particular item. Thus, some important information is missing in this representation. The next section describes a better way to represent MRCVs.

2.2.1 Representation of MRCVs

Bilder and Loughin (2004) treat each item of the MRCVs as a yes/no binary response. Let W and Y denote the MRCVs for a cross-tabulation's row and column variables, respectively. Corresponding to the data in Table 2.2, W represents waste storage methods and Y represents sources of veterinary information. In this example, there are four items under W and five items under Y . In general, suppose W has I items and Y has J items. Assuming that n subjects are sampled at random, for $i = 1, 2, \dots, I$, let

$$W_i = \begin{cases} 1, & \text{if responded 'yes' (positive response) for item } i \\ 0, & \text{if responded 'no' (negative response) for item } i \end{cases}$$

Let Y_j be similarly defined for $j = 1, 2, \dots, J$.

Notice that a separate 2×2 contingency table can be prepared for each pair of items (W_i, Y_j) . The full set of contingency tables created for all IJ pairs of items is called an *item-response table* (IRT) (Bilder and Loughin, 2007). An example is given in Table 2.3. It summarizes all possible responses to item pairs, without regard to responses to other items. Hence, the counts within the IRT are marginal counts taken across all of the other items. The IRT is capable of overcoming the difficulties caused by the regular $I \times J$ contingency table, as it summarizes both the respondents and non-respondents to a given item pair. Notice that the regular contingency table corresponds to just the counts in the (1,1) cell of each subtable. Also each subtable adds up to the sample size n .

Table 2.3: Item response table.

		Y ₁		Y ₂		Y ₃		Y ₄		Y ₅	
		0	1	0	1	0	1	0	1	0	1
W ₁	0	126	10	100	36	91	45	68	68	84	52
	1	109	34	89	54	93	50	80	63	102	41
W ₂	0	172	27	142	57	138	61	111	88	143	56
	1	63	17	47	33	46	34	37	43	43	37
W ₃	0	156	38	127	67	129	65	112	82	135	59
	1	79	6	62	23	55	30	36	49	51	34
W ₄	0	223	43	180	86	175	91	141	125	175	91
	1	12	1	9	4	9	4	7	6	11	2

Instead of summarizing marginal counts between each pair of items separately, one may be interested in summarizing counts among all possible combinations of row and column items. This is known as the *joint table* (Bilder and Loughin, 2004) and is similar to the expanded and complete tables described in Loughin and Scherer (1998) and Agresti and Liu (1999), respectively. Table 2.4 gives the joint table of Kansas farmer data. The joint cell counts are used to find the counts of cell (a, b) for $a = 0, 1$ and $b = 0, 1$ in the (W_i, Y_j) subtable of Table 2.3 for $i = 1, \dots, I$, and $j = 1, \dots, J$. The joint table has 2^I rows and 2^J columns. The joint table is sparse when the sample size is much smaller than $2^{(I+J)}$, or when the counts are concentrated within a few combinations of certain row- or column-item responses. Both of these situations hold for the Kansas farmer data.

2.2.2 Specification of SPMI

The main objective of Bilder and Loughin (2004) is to test for independence between two MRCVs. In particular, a convenient representation of the association between the two MRCVs can be obtained by considering whether items of W are associated with items of Y , without regard to which combinations of W – or Y – items are chosen. Agresti and Liu (1999) call this a test for *simultaneous pairwise marginal independence (SPMI)*, because the null hypothesis is that each of the subtables in the IRT is created under independence between the respective row and column items. This test can be performed to help determine whether each source of veterinary information is simultaneously independent of each swine waste storage method. If SPMI is rejected, individual 2×2 marginal tables should be further examined to determine where associations occur.

Let the cell counts in the joint table be denoted by n_{kl} for the k th possible combination of (W_1, \dots, W_I) and l th possible combination of (Y_1, \dots, Y_J) , where $k = 1, \dots, 2^I$ and $l = 1, \dots, 2^J$. The corresponding cell probability is denoted by τ_{kl} and the joint cell counts are assumed to follow a multinomial distribution with probabilities τ_{kl} , since exactly one combination of each item is chosen by each individual. From the probability model on the joint table, corresponding probabilities and expected counts for the item-response table can be constructed.

Specifically, let $m_{ab(ij)}$ be the observed counts of the item response table; i.e., they represent the number of joint occurrences of $(W_i = a, Y_j = b)$, where $a = 0, 1$, $b = 0, 1$, $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, J$. Let corresponding expected counts under the probability model be $\mu_{ab(ij)}$ and let the corresponding probability $P(W_i = a, Y_j = b)$ be given by $\pi_{ab(ij)}$.

Then,

$$E(m_{ab(ij)}) = n \sum_{\{k,l:W_i=a&Y_j=b\}} \tau_{kl}$$

$$\mu_{ab(ij)} = n\pi_{ab(ij)}$$

In the case of SRCVs, independence between two variables holds when their joint probability distribution can be obtained from products of the row- and column-total probabilities of each variable. Agresti and Liu (1999) extend this definition for two MRCVs through SPMI, which is the simultaneous independence between each pair of items.

Therefore, the hypotheses for a test of SPMI are,

$$H_0 : \pi_{ab(ij)} = \pi_{a.(ij)}\pi_{.b(ij)} \quad \text{for } a = 0, 1, \quad b = 0, 1, \quad i = 1, \dots, I \quad \text{and} \quad j = 1, \dots, J,$$

$$H_1 : \text{at least one equality does not hold,}$$

where $\pi_{a.(ij)}$ and $\pi_{.b(ij)}$ are the corresponding row- and column-total probabilities for $W_i = a$ and $Y_j = b$, respectively.

2.2.3 Modified test statistic and its sampling distribution

Agresti and Liu (1999) suggest using a modified test statistic that sums up all the Pearson statistics of the subtables in the IRT to test for the independence between an SRCV and an MRCV. It expands the modified test statistic by Loughin and Scherer (1998) to consider the ‘no’ outcomes as well as the ‘yes’ outcomes in the marginal tables. Therefore the statistic is invariant to the switching of the ‘yes’ and ‘no’ labels for all the items. Following their suggestion, Bilder and Loughin (2004) develop a test for SPMI between two MRCVs. Let the Pearson statistic for testing independence between W_i and Y_j be $X_{s,i,j}^2$. The modified statistic for testing SPMI is simply the sum of all such statistics across all subtables in the item-response table,

$$X_s^2 = \sum_{i=1}^I \sum_{j=1}^J X_{s,i,j}^2$$

If the IJ Pearson statistics are naively assumed as independent, X_s^2 is asymptotically chi-square distributed with IJ df. But in most cases the Pearson statistics are not independent of each other, because the marginal counts in different subtables are based on the same joint-table counts. Therefore Bilder and Loughin (2004) investigate several ways to approximate the asymptotic distribution of X_s^2 . They show that it has an asymptotic distribution which

is a linear combination of independent chi-square random variables with $df = 1$. Rao and Scott (1981) propose approximations to distributions of this type. The first-order correction adjusts X_s^2 so that it has the same mean as a χ_{IJ}^2 random variable. Thomas and Decady (2004) and Bilder and Loughin (2004) show that the adjustment factor is 1. Therefore, under the first order correction, X_s^2 is compared to sum of χ_1^2 random variables which results in the same testing method as if they were independent. Bilder and Loughin (2001) show that the first-order adjusted statistic does not hold the correct size when MRCVs are involved. Therefore, second-order corrections are used to better approximate the distribution.

A second-order correction further adjusts the modified test statistic so that it has the same mean and variance as a χ_ν^2 random variable, for some value ν . Bilder and Loughin (2004) show that the second-order adjusted statistic, $X_{RS2}^2 = IJX_s^2 / \sum_{p=1}^{IJ} \lambda_p^2$, behaves approximately as a χ^2 random variable with $\nu = I^2 J^2 / \sum_{p=1}^{IJ} \lambda_p^2$ df, where λ_p , $p = 1, \dots, IJ$, are the coefficients of the linear combination.

Bilder and Loughin (2004) show that the sampling distribution of X_s^2 can instead be approximated by nonparametric bootstrap procedure. Bootstrap procedures are generally used to estimate a test statistic's sampling distribution when its distribution is mathematically hard to derive or when assumptions behind large-sample approximations are violated. In the present context, the data are resampled by randomly selecting a row response combination (w_1, \dots, w_I) and combining it with an independently chosen column response combination (y_1, \dots, y_J) . This process is repeated n times to form a resample, and a large number of resamples, B , are taken. The test statistic $X_{s,b}^{2*}$ is calculated for each resample and the p-value is computed as the proportion of test statistics greater than or equal to the original test statistic; i.e., $\text{p-value} = (\#\text{of } X_{s,b}^{2*} \geq X_s^2) / B$.

Bilder and Loughin (2004) also suggest using a Bonferroni adjustment as an alternative to the bootstrap approach. For each subtable calculate the p-value p_{ij} using the usual χ_1^2 approximation for $X_{s,i,j}^2$ and reject the null hypothesis of SPMI if any $p_{ij} < \alpha / IJ$. The Bonferroni-adjusted p-value is $\tilde{p} = IJ \min_{ij}(p_{ij})$. SPMI is rejected when $\tilde{p} < \alpha$.

Bilder and Loughin (2004) further show that out of the above mentioned methods, the bootstrap method generally holds the correct size and has adequate power to detect various alternative hypotheses. The Bonferroni and second-order adjusted X_s^2 are conservative sometimes.

2.3 Modeling an SRCV and an MRCV

Agresti and Liu (1999) discuss the modeling of a multiple-response categorical variable. Using the data from Loughin and Scherer (1998), they treat each category of the MRCV

as a yes/no binary response. They then develop a simultaneous logit model for each of the binary components of the multivariate response. Let variable Y_1 indicate whether a farmer said ‘yes’ to source A , variable Y_2 indicate whether a farmer said ‘yes’ to source B , and so forth. Also let X be the random variable representing the highest education level. The simplest model is one that assumes simultaneous independence between X and Y_1 , X and Y_2 , and so forth in five separate two-way marginal tables. This condition is referred to as *multiple marginal independence (MMI)*. A model for the separate tables’ probabilities under MMI is:

$$\log \left(\frac{\pi_{j|i}}{1 - \pi_{j|i}} \right) = \beta_j, \quad i = 1, \dots, I, j = 1, \dots, J, \quad (2.1)$$

where, $\pi_{j|i}$ denotes the probability of responding ‘yes’ for item j given the $X = i$. According to model 2.1 the probability of responding ‘yes’ for item j is the same for all levels of X . i.e., each item is independent of the education level but the probability of a positive response may vary among the different binary items.

A more general model for these probabilities is

$$\log \left(\frac{\pi_{j|i}}{1 - \pi_{j|i}} \right) = \beta_{ij}, \quad i = 1, \dots, I, j = 1, \dots, J, \quad (2.2)$$

where the probability of responding ‘yes’ for item j is different for levels of X and across items. Since no assumptions are made on the association structure, the number of parameters equals the number of probabilities being modeled. In other words, this is a saturated model. Both of these models represent marginal constraints on a multinomial model fitted to the joint table of counts. The joint table in this context is the $I \times 2^J$ cross tabulation of responses to a row category and a *combination* of column items. Agresti and Liu (1999) use maximum likelihood techniques to find parameter estimates under constraints 2.1 or 2.2. They test for MMI with large samples using likelihood ratio test and the Pearson test statistic to compare the models 2.1 and 2.2. These statistics have large sample chi-square distributions with $df = (I - 1)J$. The large-sample approximations for the sampling distributions of these test statistics may not be very good, because the joint table is likely to be very sparse.

2.4 Modeling two MRCVs

Generally loglinear models are used to model associations in regular contingency tables. Many examples are given in Bilder and Loughin (2014). For an ordinary SRCV, let W

represent the row variable with I levels and Y represent the column variable with J levels. The loglinear model assuming independence between W and Y can be written as,

$$\log(\mu_{ij}) = \beta_0 + \beta_i^W + \beta_j^Y, \quad i = 1, \dots, I, j = 1, \dots, J, \quad (2.3)$$

where μ_{ij} is the mean count in cell (i, j) , β_0 is the log mean count of cell $(1, 1)$, β_i^W and β_j^Y operate on the row and column margins respectively, and $\beta_1^W = \beta_1^Y = 0$. Notice that,

$$\beta_i^W = \log(\mu_{i1}) - \log(\mu_{11}) = \dots = \log(\mu_{iJ}) - \log(\mu_{1J}) = \log(\mu_{i.}) - \log(\mu_{1.}).$$

Similarly,

$$\beta_j^Y = \log(\mu_{1j}) - \log(\mu_{11}) = \dots = \log(\mu_{IJ}) - \log(\mu_{I1}) = \log(\mu_{.j}) - \log(\mu_{.1}).$$

The loglinear model that allows association between W and Y can be written as,

$$\log(\mu_{ij}) = \beta_0 + \beta_i^W + \beta_j^Y + \beta_{ij}^{WY}, \quad i = 1, \dots, I, j = 1, \dots, J, \quad (2.4)$$

where β_{ij}^{WY} is the interaction term which allows the difference of log mean counts between cells in two rows to change across columns and vice versa, and $\beta_1^W = \beta_1^Y = \beta_{1j}^{WY} = \beta_{i1}^{WY} = 0$.

2.4.1 Identification of models

Bilder and Loughin (2007) generalize loglinear models to test for SPMI and to describe any patterns of associations when SPMI does not hold.

Now let W and Y be MRCVs. First consider the loglinear model that assumes independence between items W_i and Y_j . Let $\mu_{ab(ij)}$ be the expected count for row a and column b of (i, j) th subtable of Table 2.3. The SPMI model assumes independence in all subtables,

$$\log(\mu_{ab(ij)}) = \beta_{0(ij)} + \beta_{a(ij)}^W + \beta_{b(ij)}^Y, \quad a = 0, 1, \quad b = 0, 1, \quad i = 1, \dots, I, \quad j = 1, \dots, J \quad (2.5)$$

This is the same as equation 2.3 with extra subscripts (i, j) added to identify which subtable is being modeled. Odds ratios of all the subtables are 1 for the SPMI model given in equation 2.5.

Bilder and Loughin (2007) investigate models that allow odds ratios not to be equal to 1 and to change across row and column items in different ways. Some of the possible models are,

$$\text{Homogeneous association : } \log(\mu_{ab(ij)}) = \beta_{0(ij)} + \beta_{a(ij)}^W + \beta_{b(ij)}^Y + \lambda_{ab}$$

$$\text{W-main effects : } \log(\mu_{ab(ij)}) = \beta_{0(ij)} + \beta_{a(ij)}^W + \beta_{b(ij)}^Y + \lambda_{ab} + \lambda_{ab(i)}^W$$

$$\text{Y-main effects : } \log(\mu_{ab(ij)}) = \beta_{0(ij)} + \beta_{a(ij)}^W + \beta_{b(ij)}^Y + \lambda_{ab} + \lambda_{ab(j)}^Y$$

$$\text{W- and Y-main effects : } \log(\mu_{ab(ij)}) = \beta_{0(ij)} + \beta_{a(ij)}^W + \beta_{b(ij)}^Y + \lambda_{ab} + \lambda_{ab(i)}^W + \lambda_{ab(j)}^Y$$

$$\text{Saturated model : } \log(\mu_{ab(ij)}) = \beta_{0(ij)} + \beta_{a(ij)}^W + \beta_{b(ij)}^Y + \lambda_{ab} + \lambda_{ab(i)}^W + \lambda_{ab(j)}^Y + \lambda_{ab(ij)}^{WY}$$

The homogeneous association model assumes that the odds ratios are the same in every subtable and uses a single parameter, λ_{ab} , to represent the common log odds ratio. The W-main effects model assumes that the odds ratios change across the row items, but are the same across the column items. Similarly, the Y-main effects model assumes that the odds ratios change across the column items, but are the same across the row items. These models have parameters for log odds ratios that may change only with i or j , respectively. The W- and Y-main effects model allows odds ratios to vary across both the row items and column items, but the differences of log odds ratios between subtables in two rows are constant across the columns and vice versa. In all the models described above the odds ratios are allowed to vary in a structured way subject to certain restrictions. But in the saturated model, the odds ratios are allowed to adapt freely to the data without any restriction.

2.4.2 Inference for generalized loglinear models

In the context of two MRCVs, each subtable in the IRT is a 2×2 contingency table and can be modeled by Poisson distribution. Since each subtable is merely a different marginal arrangement of the same joint-table counts, the subtables are not independent of one another. Therefore, specifying a full likelihood across all the subtables involve specifying associations across items, which results in a very complex model with many parameters. Instead, Bilder and Loughin (2007) propose to use a pseudo-likelihood function which assumes independence across the subtables. Therefore, the pseudo-likelihood is simply the product of each of the IJ Poisson likelihood functions from each subtable.

The parameter estimators can be obtained by maximizing the pseudo-likelihood function and solving a set of GEEs. They are called pseudo maximum likelihood estimators (pseudo-MLEs). This approach is quite similar to the maximum likelihood estimation except for the

fact that it is not based on a full likelihood. The estimates are consistent and asymptotically normally distributed similar to MLEs.

However, the variances of the pseudo-MLEs depend on the underlying association structure across the subtables that we misspecify as independence in the model. Therefore, the variance estimates produced by treating the pseudo-likelihood as a regular likelihood may be smaller than they should be, as they ignore the correlations among the counts in IJ subtables. Liang and Zeger (1986) propose a method to correct the variances, resulting in variance estimators that are called “sandwich” estimators as they are mathematically written as a product of three matrices, where the same matrix is used at the each end. Bilder and Loughin (2007) provide details, which are summarized below.

The relationship between item-response table counts and joint table counts is,

$$\mathbf{m} = B\mathbf{n},$$

where \mathbf{m} is a $4IJ \times 1$ vector of observed counts of the item-response table ($m_{ab(ij)}$), \mathbf{n} is a $2^{(I+J)} \times 1$ vector of joint table counts (n_{kl}) and B is a $4IJ \times 2^{(I+J)}$ matrix that contains only 0’s and 1’s. Specifically, B can be written as

$$B = \begin{bmatrix} G \otimes H \\ G \otimes (J_{J \times 2^J} - H) \\ (J_{I \times 2^I} - G) \otimes H \\ (J_{I \times 2^I} - G) \otimes (J_{J \times 2^J} - H) \end{bmatrix}, \quad (2.6)$$

where G is a $I \times 2^I$ matrix containing all possible $(W_1, \dots, W_I)'$ vectors of 0’s and 1’s, H is similarly defined for all possible $(Y_1, \dots, Y_J)'$ vectors, and $J_{r \times c}$ is a $r \times c$ matrix of 1’s.

The asymptotic variance of \mathbf{m} can be estimated by,

$$\hat{V} = \widehat{V(\mathbf{m})} = nB [\text{diag}(\hat{\boldsymbol{\tau}}) - \hat{\boldsymbol{\tau}}\hat{\boldsymbol{\tau}}'] B', \quad (2.7)$$

where $\hat{\boldsymbol{\tau}}$ is a vector of $2^{(I+J)} \times 1$ containing estimated joint probabilities ($\hat{\tau}_{kl}$).

For any given model, let $\boldsymbol{\beta}$ be a vector containing all of the model’s parameters as described in Section 2.4.1, let X be the corresponding design matrix, and let $\hat{\boldsymbol{\mu}}$ be the vector of model-predicted counts. Bilder and Loughin (2007) show that covariance matrix for $\hat{\boldsymbol{\beta}}$ can be estimated by $\hat{\Sigma} = (X' \text{diag}(\hat{\boldsymbol{\mu}}) X)^{-1} X' \hat{V} X (X' \text{diag}(\hat{\boldsymbol{\mu}}) X)^{-1}$.

2.4.3 Model comparisons

Since no true likelihood function is used, usual information criteria such as AIC and BIC cannot be used for model comparisons. Instead, Bilder and Loughin (2007) use hypothesis testing of nested models to perform model comparisons.

One approach uses Rao-Scott-adjusted Pearson statistics and modified asymptotic chi-square distributions. The Pearson statistic calculated to compare two models, when one is nested within the other, is

$$X_M^2 = \sum_{a,b,i,j} \frac{\left(\hat{\mu}_{ab(ij)}^{(A)} - \hat{\mu}_{ab(ij)}^{(0)}\right)^2}{\hat{\mu}_{ab(ij)}^{(0)}}, \quad a = 0, 1, \quad b = 0, 1, \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

where $\hat{\mu}_{ab(ij)}^{(0)}$ and $\hat{\mu}_{ab(ij)}^{(A)}$ are model predicted counts under the null and alternative hypothesis, respectively. As described in Section 2.2.3, first-order Rao-Scott uses a chi-square distribution with IJ df as an approximation to the true asymptotic distribution of the Pearson statistic. Second-order Rao-Scott adjustment attempts to correct both the test statistic and the asymptotic distribution so that the result has the same mean and variance as the correct asymptotic distribution. Similar calculations can be applied to statistics created using a likelihood-ratio formulation.

The first-order correction can lead to liberal test. Further, Bilder and Loughin (2007) discover that the second-order correction for model comparison can be conservative sometimes. An alternative approach suggested by Bilder and Loughin (2007) to compare models is to use the bootstrap to estimate the distribution of the test statistic. They use semi-parametric resampling as in Gange (1995) to generate correlated binary data with features similar to the original data. The models specified under both the null and alternative hypotheses are fitted to the resampled data and model comparison statistic $X_{M,b}^{2*}$ computed for each resample. A p-value is computed as $(\# \text{ of } X_{M,b}^{2*} \geq X_M^2) / B$.

Following the model comparisons, when an adequate model is found, model-estimated odds ratios and their confidence intervals are used to investigate the associations between MRCVs. Bilder and Loughin (2007) also show how to use standardized residuals to identify huge deviations from the specified models.

Bilder and Loughin (2007) describe briefly how to model three MRCVs. They present a few special models, but do not consider the many more complicated models one can apply when modeling three MRCVs. Developing these models and understanding their features is the main focus of this thesis and is discussed in detail in the next chapter.

Chapter 3

Models for three MRCVs

3.1 Model building process for three MRCVs

3.1.1 Odds ratios within a subcube

When there are three MRCVs, models can be built similar to the ones described in Section 2.4.1. Let Z_k be the indicator variable for the third MRCV for $k = 1, \dots, K$ similarly defined as W_i and Y_j . The joint table of counts is now a three-way cross-classification of counts for all possible combinations of responses to the items in W, Y , and Z . The IRT in this case consists of three-way cross-classifications of (W_i, Y_j, Z_k) item combinations, each of which results in a $2 \times 2 \times 2$ contingency table. Geometrically each subtable can be depicted as a $2 \times 2 \times 2$ subcube residing within each cell of a larger $I \times J \times K$ cube as shown in Figure 3.1. In this case, there are IJK subcubes and all these subcubes together form the IRT.

In the case of two MRCVs, the models are built by fully parameterizing the main effects in each subtable and allowing the odds ratios (OR) to vary in a structured way as explained in Section 2.4.1. In the case of three MRCVs the main effects are once again saturated with parameters in each subcube. However, since there are six 2×2 faces in each subcube, six different odds ratios can be considered for each subcube. These odds ratios are conditional, as they are the odds ratios between two items evaluated at a fixed level of the third item. Table 3.1 shows the notation used throughout the thesis for conditional OR of each face within a (W_i, Y_j, Z_k) subcube.

There are several ways that these six conditional odds ratios could vary within a subcube.

Figure 3.1: Item response table for three MRCVs. (Source: http://3.bp.blogspot.com/-o_nuM70hxIk/UEkpcCunRII/AAAAAAAAAPc/mrd5CfbQDxI/s1600/234.jpg)

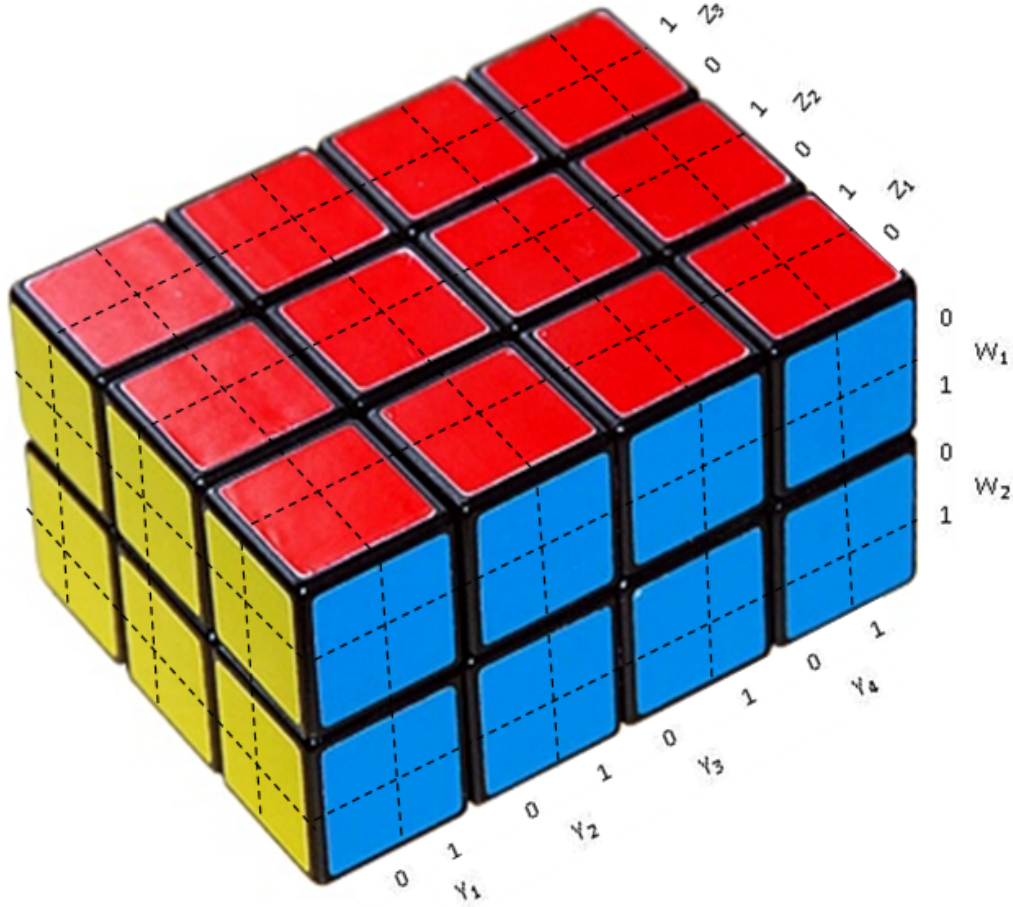


Table 3.1: Notation for conditional ORs for each face within a (W_i, Y_j, Z_k) subcube.

Type of conditional OR	Conditional OR at third item= 0	Conditional OR at third item= 1
Conditional OR between W_i and Y_j	$OR_{ij(k,0)}$	$OR_{ij(k,1)}$
Conditional OR between W_i and Z_k	$OR_{i(j,0)k}$	$OR_{i(j,1)k}$
Conditional OR between Y_j and Z_k	$OR_{(i,0)jk}$	$OR_{(i,1)jk}$

- Mutual independence: All the three items W_i , Y_j and Z_k are independent of each other, denoted by (W_i, Y_j, Z_k) . All the conditional ORs are 1; i.e., $OR_{ij(k,0)} = OR_{ij(k,1)} = OR_{i(j,0)k} = OR_{i(j,1)k} = OR_{(i,0)jk} = OR_{(i,1)jk} = 1$.
- Joint independence: Two items are jointly independent of the third but are associated with each other. This is denoted by $(W_i Y_j, Z_k)$, $(W_i Z_k, Y_j)$ or $(Y_j Z_k, W_i)$. The

conditional ORs between the two associated items are not equal to 1 but are the same across the levels of the third item; i.e., $OR_{ij(k,0)} = OR_{ij(k,1)}$, $OR_{i(j,0)k} = OR_{i(j,1)k}$ or $OR_{(i,0)jk} = OR_{(i,1)jk}$, respectively.

- Conditional independence: Two items are independent given the third, denoted by $(W_i Y_j, W_i Z_k)$, $(W_i Z_k, Y_j Z_k)$ or $(W_i Y_j, Y_j Z_k)$. Consider $(W_i Y_j, W_i Z_k)$ for an example. This model specifies that Y_j and Z_k are independent, given W_i . Furthermore, The conditional ORs between W_i , Y_j and W_i , Z_k are all not equal to 1, but are the same for the two levels of the missing variable; i.e., $OR_{ij(k,0)} = OR_{ij(k,1)}$ and $OR_{i(j,0)k} = OR_{i(j,1)k}$.
- Homogeneous associations: Denoted by $(W_i Y_j, W_i Z_k, Y_j Z_k)$, the conditional ORs between any two items are not equal to 1, but they are the same across the levels of the third item. i.e, $OR_{ij(k,0)} = OR_{ij(k,1)}$, $OR_{i(j,0)k} = OR_{i(j,1)k}$ and $OR_{(i,0)jk} = OR_{(i,1)jk}$.
- Saturated (heterogeneous associations) : Denoted by $(W_i Y_j Z_k)$, the conditional ORs between any two items change across the levels of the third item.

3.1.2 Odds ratios across the subcubes

Once associations within a subcube are identified, models can be extended to describe how these associations may vary across the items that are represented by the different subcubes. A given association within a subcube can change or be constant across its variables in other subcubes. However a given association within a subcube must be constant across the level of any variables not involved in the association. For example, in the joint independence model $(W_i Y_j, Z_k)$, the $W_i Y_j$ association can change across W items $i = 1, \dots, I$, and/or across Y items, $j = 1, \dots, J$. However, this association cannot change across different Z items, $k = 1, \dots, K$. This is because, when an association does not involve the third variable, it can be measured on the marginal sums across the levels of that variable. In the three-way IRT, these marginal totals do not change across levels of the variable for fixed levels of the other two variables. For example, the marginal counts for $W_i = a$, $Y_j = b$ across the two levels of Z_k are the same for all k . This reduces the number of possible models that can be fitted. Table 3.2 summarizes all possible models that can be fitted to three MRCVs.

According to Table 3.2, When there are no associations within a subcube (mutual independence), there is nothing to change across the subcubes, so this association structure is labeled as being constant (C). Hence there is only one possible way to model it. When there is a single 2-way association within a subcube (e.g., row 2 of Table 3.2), the association may remain constant (C), change across one variable only (W or Y), change across two variables simultaneously with the changes being constant across the levels of the other variable

Table 3.2: All possible models for three MRCVs.

Associations within a subcube	Associations across subcubes	No. of models
1. independent	C	1
2. $(W_i Y_j, Z_k)$	C, W, Y, W+Y, WY	5
3. $(W_i Z_k, Y_j)$	C, W, Z, W+Z, WZ	5
4. $(Y_j Z_k, W_i)$	C, Y, Z, Y+Z, YZ	5
5. $(W_i Y_j, W_i Z_k)$	(C,C), (C,W),.....,(WY, WZ)	25
6. $(W_i Y_j, Y_j Z_k)$	(C,C), (C,Y),.....,(WY, YZ)	25
7. $(Y_j Z_k, W_i Z_k)$	(C,C), (C,W),.....,(YZ, WZ)	25
8. $(W_i Y_j, W_i Z_k, Y_j Z_k)$	(C,C,C),(C,C,Y),.....,(WY, WZ, YZ)	125
9. $(W_i Y_j Z_k)$	C,W,Y,Z,W+Y, W+Z, Y+Z, W+Y+Z, WY, WZ, YZ, (WY, YZ), (WY,WZ), (YZ, WZ), (WY, WZ, YZ), (WYZ)	16

(W+Y) or change across two variables without any restrictions (WY). Hence there are 5 possible models through which a 2-way association can manifest itself across the subcubes in the IRT. The same explanation applies for rows 3 and 4 of Table 3.2. When there are two 2-way associations within a subcube (rows 5, 6 and 7 of the Table 3.2), each association can vary in 5 different ways. Considering all possible combinations the both associations can form, there are 25 possible ways the associations can change across the subcubes. Similarly, when there are three 2-way associations within a subcube (row 8 of the Table 3.2), each association can vary in 5 different ways result in all possible combinations can form 125 models. When there is a 3-way association (row 9 of the Table 3.2), it can vary in 16 possible ways. Therefore, altogether there are 232 ways the three MRCVs can be modeled.

3.1.3 Model building for three MRCVs

Once the associations within a subcube and patterns of associations across the subcubes are identified as given in Table 3.2, one can write the models for the associations similar to the way it is done for two MRCVs in Section 2.4.1. For a given association within a subcube, only the highest possible order of association across the subcubes is discussed and modeled below. The models are not shown and explained for the lower order associations across the subcubes, since they are the subsets of the models of highest order associations, and one can figure them out easily by omitting certain unnecessary parameters.

In the case of mutual independence, there is no association within and across the subcubes. The model can be written as,

$$\log(\mu_{abc(ijk)}) = \beta_{0(ijk)} + \beta_{a(ijk)}^W + \beta_{b(ijk)}^Y + \beta_{c(ijk)}^Z,$$

where $\mu_{abc(ijk)}$ is the expected count for the (a, b, c) cell of (W_i, Y_j, Z_k) th subcube of IRT for $a = 0, 1$, $b = 0, 1$, $c = 0, 1$, $i = 1, \dots, I$, $j = 1, \dots, J$, and $k = 1, \dots, K$. Models for each cell of (W_i, Y_j, Z_k) th subcube can be written as follows.

Table 3.3: Models for each cell of (W_i, Y_j, Z_k) th subcube.

a	b	c	Model
0	0	0	$\log(\mu_{000(ijk)}) = \beta_{0(ijk)}$
0	1	0	$\log(\mu_{010(ijk)}) = \beta_{0(ijk)} + \beta_{1(ijk)}^Y$
1	0	0	$\log(\mu_{100(ijk)}) = \beta_{0(ijk)} + \beta_{1(ijk)}^W$
1	1	0	$\log(\mu_{110(ijk)}) = \beta_{0(ijk)} + \beta_{1(ijk)}^W + \beta_{1(ijk)}^Y$
0	0	1	$\log(\mu_{001(ijk)}) = \beta_{0(ijk)} + \beta_{1(ijk)}^Z$
0	1	1	$\log(\mu_{011(ijk)}) = \beta_{0(ijk)} + \beta_{1(ijk)}^Y + \beta_{1(ijk)}^Z$
1	0	1	$\log(\mu_{101(ijk)}) = \beta_{0(ijk)} + \beta_{1(ijk)}^W + \beta_{1(ijk)}^Z$
1	1	1	$\log(\mu_{111(ijk)}) = \beta_{0(ijk)} + \beta_{1(ijk)}^W + \beta_{1(ijk)}^Y + \beta_{1(ijk)}^Z$

Based on Table 3.3, the parameters are estimated as

$$\begin{aligned}\hat{\beta}_{0(ijk)} &= \log(\hat{\mu}_{000(ijk)}), \\ \hat{\beta}_{1(ijk)}^W &= \log(\hat{\mu}_{1..(ijk)}) - \log(\hat{\mu}_{0..(ijk)}), \\ \hat{\beta}_{1(ijk)}^Y &= \log(\hat{\mu}_{.1.(ijk)}) - \log(\hat{\mu}_{.0.(ijk)}) \text{ and} \\ \hat{\beta}_{1(ijk)}^Z &= \log(\hat{\mu}_{..1(ijk)}) - \log(\hat{\mu}_{..0(ijk)}).\end{aligned}$$

$\hat{\beta}_{1(ijk)}^W$, $\hat{\beta}_{1(ijk)}^Y$ and $\hat{\beta}_{1(ijk)}^Z$ estimate the main effects of W_i , Y_j and Z_k respectively within (W_i, Y_j, Z_k) th subcube.

When there is a 2-way association within a subcube (e.g., row 2 of Table 3.2) and it varies across both the variables without any restrictions (WY), the model can be written as,

$$\log(\mu_{abc(ijk)}) = \beta_{0(ijk)} + \beta_{a(ijk)}^W + \beta_{b(ijk)}^Y + \beta_{c(ijk)}^Z + \lambda_{ab} + \lambda_{ab(i)}^W + \lambda_{ab(j)}^Y + \lambda_{ab(ij)}^{WY}.$$

Models can be written for each subcube similar to given in Table 3.3; the additional parameters are estimated as

$$\begin{aligned}
\hat{\lambda}_{11} &= \log \widehat{OR}_{11(1,0)}, \\
\hat{\lambda}_{11(i)}^W &= \log \widehat{OR}_{i1(1,0)} - \log \widehat{OR}_{11(1,0)}, \\
\hat{\lambda}_{11(j)}^Y &= \log \widehat{OR}_{1j(1,0)} - \log \widehat{OR}_{11(1,0)} \text{ and} \\
\hat{\lambda}_{11(ij)}^{WY} &= \log \widehat{OR}_{ij(1,0)} + \log \widehat{OR}_{11(1,0)} - \log \widehat{OR}_{i1(1,0)} - \log \widehat{OR}_{1j(1,0)}.
\end{aligned}$$

When there are two 2-way associations within a subcube (e.g., row 7 of Table 3.2) and they may vary freely across subcubes (YZ, WZ). The model is,

$$\begin{aligned}
\log(\mu_{abc(ijk)}) &= \beta_{0(ijk)} + \beta_a^W(ijk) + \beta_b^Y(ijk) + \beta_c^Z(ijk) + \delta_{ac} + \delta_{ac(i)}^W + \delta_{ac(k)}^Z + \delta_{ac(ik)}^{WZ} \\
&+ \gamma_{bc} + \gamma_{bc(j)}^Y + \gamma_{bc(k)}^Z + \gamma_{bc(jk)}^{YZ}.
\end{aligned}$$

The association parameters are estimated similar to the previous case as

$$\begin{aligned}
\hat{\delta}_{11} &= \log \widehat{OR}_{1(1,0)1}, \\
\hat{\delta}_{11(i)}^W &= \log \widehat{OR}_{i(1,0)1} - \log \widehat{OR}_{1(1,0)1}, \\
\hat{\delta}_{11(k)}^Z &= \log \widehat{OR}_{1(1,0)k} - \log \widehat{OR}_{1(1,0)1}, \\
\hat{\delta}_{11(ik)}^{WZ} &= \log \widehat{OR}_{i(1,0)k} + \log \widehat{OR}_{1(1,0)1} - \log \widehat{OR}_{i(1,0)1} - \log \widehat{OR}_{1(1,0)k}, \\
\hat{\gamma}_{11} &= \log \widehat{OR}_{(1,0)11}, \\
\hat{\gamma}_{11(j)}^Y &= \log \widehat{OR}_{(1,0)j1} - \log \widehat{OR}_{(1,0)11}, \\
\hat{\gamma}_{11(k)}^Z &= \log \widehat{OR}_{(1,0)1k} - \log \widehat{OR}_{(1,0)11} \text{ and} \\
\hat{\gamma}_{11(jk)}^{YZ} &= \log \widehat{OR}_{(1,0)jk} + \log \widehat{OR}_{(1,0)11} - \log \widehat{OR}_{(1,0)j1} - \log \widehat{OR}_{(1,0)1k}.
\end{aligned}$$

When there is a 3-way association (row 9 of the Table 3.2) and (WYZ) across the subcubes which is the saturated model is given as,

$$\begin{aligned}
\log(\mu_{abc(ijk)}) &= \beta_{0(ijk)} + \beta_a^W(ijk) + \beta_b^Y(ijk) + \beta_c^Z(ijk) \\
&+ \lambda_{ab} + \lambda_{ab(i)}^W + \lambda_{ab(j)}^Y + \lambda_{ab(k)}^Z + \lambda_{ab(ij)}^{WY} + \lambda_{ab(ik)}^{WZ} + \lambda_{ab(jk)}^{YZ} + \lambda_{ab(ijk)}^{WYZ} \\
&+ \delta_{ac} + \delta_{ac(i)}^W + \delta_{ac(j)}^Y + \delta_{ac(k)}^Z + \delta_{ac(ij)}^{WY} + \delta_{ac(ik)}^{WZ} + \delta_{ac(jk)}^{YZ} + \delta_{ac(ijk)}^{WYZ} \\
&+ \gamma_{bc} + \gamma_{bc(i)}^W + \gamma_{bc(j)}^Y + \gamma_{bc(k)}^Z + \gamma_{bc(ij)}^{WY} + \gamma_{bc(ik)}^{WZ} + \gamma_{bc(jk)}^{YZ} + \gamma_{bc(ijk)}^{WYZ} \\
&+ \eta_{abc} + \eta_{abc(i)}^W + \eta_{abc(j)}^Y + \eta_{abc(k)}^Z + \eta_{abc(ij)}^{WY} + \eta_{abc(ik)}^{WZ} + \eta_{abc(jk)}^{YZ} + \eta_{abc(ijk)}^{WYZ}.
\end{aligned}$$

The 2-way association parameters $\lambda_{11(ik)}^{WZ}$, $\lambda_{11(jk)}^{YZ}$, $\delta_{11(ij)}^{WY}$, $\delta_{11(jk)}^{YZ}$, $\gamma_{11(ij)}^{WY}$ and $\gamma_{11(ik)}^{WZ}$ are estimated as the same way explained above. The additional association parameters are estimated as,

$$\hat{\lambda}_{111(ijk)}^{WYZ} = \left(\log \widehat{OR}_{ij(k,0)} + \log \widehat{OR}_{11(k,0)} - \log \widehat{OR}_{i1(k,0)} - \log \widehat{OR}_{1j(k,0)} \right) - \left(\log \widehat{OR}_{ij(1,0)} + \log \widehat{OR}_{11(1,0)} - \log \widehat{OR}_{i1(1,0)} - \log \widehat{OR}_{1j(1,0)} \right)$$

$\hat{\delta}_{ac(ijk)}^{WYZ}$ and $\hat{\gamma}_{bc(ijk)}^{WYZ}$ can be obtained similarly.

$\hat{\eta}_{111} = \log \widehat{OR}_{11(1,1)} - \log \widehat{OR}_{11(1,0)}$ and it's the same for conditional association between any two variables. Let D_{ijk} be the difference between the two conditional log ORs in (W_i, Y_j, Z_k) th subcube; i.e., $D_{ijk} = \log OR_{ij(k,1)} - \log OR_{ij(k,0)}$. Thus the additional parameters can be estimated in terms of \hat{D}_{ijk} as

$$\begin{aligned} \hat{\eta}_{111} &= \hat{D}_{111}, \\ \hat{\eta}_{111(i)}^W &= \hat{D}_{i11} - \hat{D}_{111}, \\ \hat{\eta}_{111(j)}^Y &= \hat{D}_{1j1} - \hat{D}_{111}, \\ \hat{\eta}_{111(k)}^Z &= \hat{D}_{11k} - \hat{D}_{111}, \\ \hat{\eta}_{111(ij)}^{WY} &= \hat{D}_{111} + \hat{D}_{ij1} - \hat{D}_{1j1} - \hat{D}_{i11}, \\ \hat{\eta}_{111(ik)}^{WZ} &= \hat{D}_{111} + \hat{D}_{i1k} - \hat{D}_{11k} - \hat{D}_{i11}, \\ \hat{\eta}_{111(jk)}^{YZ} &= \hat{D}_{111} + \hat{D}_{1jk} - \hat{D}_{1j1} - \hat{D}_{11k} \text{ and} \\ \hat{\eta}_{111(ijk)}^{WYZ} &= \left(\hat{D}_{ijk} + \hat{D}_{11k} - \hat{D}_{i1k} - \hat{D}_{1jk} \right) - \left(\hat{D}_{111} + \hat{D}_{ij1} - \hat{D}_{i11} - \hat{D}_{1j1} \right). \end{aligned}$$

3.2 Inference on the models

Similar to the inferences developed for the loglinear models involving two MRCVs, inferences for three MRCVs can be made. Parameters are estimated using GEEs generated by a pseudo-likelihood and variances of the estimates are corrected using sandwich methods.

In the case of three MRCVs, the relationship between item-response table counts and joint table counts is $\mathbf{m} = A\mathbf{n}$, where \mathbf{m} is a $8IJK \times 1$ vector of observed counts of the IRT ($m_{abc(ijk)}$), \mathbf{n} is a $2^{(I+J+K)} \times 1$ vector of joint table counts (n_{klm}) and A is a $8IJK \times 2^{(I+J+K)}$ matrix that contains only 0's and 1's. Analogous to the two-MRCV case, A can be written as

$$A = \begin{bmatrix} G \otimes H \otimes L \\ G \otimes H \otimes (J_{K \times 2^K} - L) \\ G \otimes (J_{J \times 2^J} - H) \otimes L \\ (J_{I \times 2^I} - G) \otimes H \otimes L \\ (J_{I \times 2^I} - G) \otimes (J_{J \times 2^J} - H) \otimes L \\ (J_{I \times 2^I} - G) \otimes H \otimes (J_{K \times 2^K} - L) \\ G \otimes (J_{J \times 2^J} - H) \otimes (J_{K \times 2^K} - L) \\ (J_{I \times 2^I} - G) \otimes (J_{J \times 2^J} - H) \otimes (J_{K \times 2^K} - L) \end{bmatrix},$$

where L is similarly defined as G and H for all possible $(Z_1, \dots, Z_K)'$ vectors. The asymptotic variance of \mathbf{m} can be estimated by,

$$\hat{W} = \widehat{V(\mathbf{m})} = nA [\text{diag}(\hat{\boldsymbol{\tau}}) - \hat{\boldsymbol{\tau}}\hat{\boldsymbol{\tau}}'] A', \quad (3.1)$$

where $\hat{\boldsymbol{\tau}}$ is a vector of length $2^{(I+J+K)}$ containing estimated joint probabilities $(\hat{\tau}_{klm})$. The covariance matrix for $\hat{\boldsymbol{\beta}}$ can be estimated by $\hat{\Sigma} = (X' \text{diag}(\hat{\boldsymbol{\mu}}) X)^{-1} X' \hat{W} X (X' \text{diag}(\hat{\boldsymbol{\mu}}) X)^{-1}$.

3.3 Model comparisons

3.3.1 Takeuchi Information Criterion

As mentioned in Section 2.4.3, usual information criteria such as AIC cannot be used for model comparisons, since estimation is not based on a full likelihood. Takeuchi (1976) proposes a version of AIC, now called the Takeuchi Information Criterion (TIC), that can be used with likelihoods where the model is misspecified.

Suppose a random sample $\mathbf{y} = y_1, \dots, y_n$ has an unknown density $f(y)$. Let the density for a proposed model be $g(y) = \prod g(y_i)$, where g depends on a vector of unknown model parameters, $\boldsymbol{\theta}$. Suppose that a log-likelihood function is formed from this model as $\log L(\boldsymbol{\theta}) = \sum_{i=1}^n \log g(y_i, \boldsymbol{\theta})$. Then TIC generalizes AIC as,

$$TIC = -2 \log L(\hat{\boldsymbol{\theta}}) + 2 \text{tr}(\hat{Q}^{-1} \hat{\Omega}),$$

where

$$\hat{Q} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta \partial \theta^t} \log g(y_i, \hat{\theta}) \quad \text{and}$$

$$\hat{\Omega} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial}{\partial \theta} \log g(y_i, \hat{\theta}) \frac{\partial}{\partial \theta} \log g(y_i, \hat{\theta})^t \right).$$

Notice that when $g(y) = f(y)$, then $\hat{Q} = \hat{\Omega}$ and the second term in TIC is twice the number of parameters in the model, exactly as in AIC. However, TIC can be computed even when the model is misspecified. Hence, it can be used for our marginal models that are fitted using pseudo-likelihood. In general, the marginal loglinear model involving two MRCVs can be written as $m_{ab(ij)} \sim \text{Poisson}(\mu_{ab(ij)})$, with

$$\log(\mu_{ab(ij)}) = X_{ab(ij)}^t \boldsymbol{\beta},$$

where X is $4IJ \times p$ design matrix and p is the number of parameters in the model.

Therefore, within this context, \hat{Q} can be derived as below:

$$\begin{aligned} \log g(m_{ab(ij)}, \boldsymbol{\beta}) &= -\mu_{ab(ij)} + m_{ab(ij)} \log \mu_{ab(ij)} - \log m_{ab(ij)}! \\ &= -\exp(X_{ab(ij)}^t \boldsymbol{\beta}) + m_{ab(ij)} X_{ab(ij)}^t \boldsymbol{\beta} - \log m_{ab(ij)}! \\ \frac{\partial}{\partial \boldsymbol{\beta}} \log g(m_{ab(ij)}, \boldsymbol{\beta}) &= -\exp(X_{ab(ij)}^t \boldsymbol{\beta}) X_{ab(ij)} + m_{ab(ij)} X_{ab(ij)} \\ &= (m_{ab(ij)} - \exp(X_{ab(ij)}^t \boldsymbol{\beta})) X_{ab(ij)} \\ \frac{\partial}{\partial \boldsymbol{\beta}} \log g(m_{ab(ij)}, \boldsymbol{\beta})^t &= (m_{ab(ij)} - \exp(X_{ab(ij)}^t \boldsymbol{\beta})) X_{ab(ij)}^t \\ \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \log g(m_{ab(ij)}, \boldsymbol{\beta}) &= -X_{ab(ij)} \exp(X_{ab(ij)}^t \boldsymbol{\beta}) X_{ab(ij)}^t \end{aligned}$$

Therefore,

$$\hat{Q} = \frac{1}{4IJ} \sum_{a,b,i,j} (\mathbf{X}_{ab(ij)} \hat{\mu}_{ab(ij)} \mathbf{X}_{ab(ij)}^t)$$

$$\hat{Q} = \frac{1}{4IJ} X^t \text{diag}(\hat{\boldsymbol{\mu}}) X,$$

where $diag(\hat{\boldsymbol{\mu}})$ is a $4IJ \times 4IJ$ diagonal matrix whose diagonal is given by $\hat{\mu}_{ab(ij)}$ s. Derivation of $\hat{\Omega}$ is as follows.

$$\begin{aligned}\hat{\Omega} &= \frac{1}{4IJ} \sum_{a,b,i,j} \left(m_{ab(ij)} - \exp \left(X_{ab(ij)}^t \boldsymbol{\beta} \right) \right) X_{ab(ij)} \left(m_{ab(ij)} - \exp \left(X_{ab(ij)}^t \boldsymbol{\beta} \right) \right) X_{ab(ij)}^t \\ &= \frac{1}{4IJ} \sum_{a,b,i,j} \left[\left(m_{ab(ij)} - \hat{\mu}_{ab(ij)} \right)^2 X_{ab(ij)} X_{ab(ij)}^t \right] \\ &= \frac{1}{4IJ} X^t diag \left[(\mathbf{m} - \hat{\boldsymbol{\mu}})^2 \right] X,\end{aligned}$$

where $diag(\mathbf{m} - \hat{\boldsymbol{\mu}})$ is a diagonal matrix of $4IJ \times 4IJ$ whose diagonal is given by $(m_{ab(ij)} - \hat{\mu}_{ab(ij)})$ s.

Similarly, TIC for three MRCVs can be obtained. Let $m_{abc(ijk)} \sim Poisson(\mu_{abc(ijk)})$, with

$$\log(\mu_{abc(ijk)}) = X_{abc(ijk)}^t \boldsymbol{\beta},$$

Then \hat{Q} and $\hat{\Omega}$ can be extended for three MRCVs as,

$$\begin{aligned}\hat{Q} &= \frac{1}{8IJK} X^t diag(\hat{\boldsymbol{\mu}}) X \\ \hat{\Omega} &= \frac{1}{8IJK} X^t diag \left[(\mathbf{m} - \hat{\boldsymbol{\mu}})^2 \right] X,\end{aligned}$$

where $diag(\hat{\boldsymbol{\mu}})$ and $diag(\mathbf{m} - \hat{\boldsymbol{\mu}})$ are $8IJK \times 8IJK$ diagonal matrices whose diagonals are $\hat{\mu}_{abc(ijk)}$ s and $(m_{abc(ijk)} - \hat{\mu}_{abc(ijk)})$ s respectively.

3.3.2 Model Averaging

As described in Section 2.4.3, hypothesis testing of nested models is used to perform comparisons among models for two MRCVs. In the case of three MRCVs, the number of models possible is huge and the presence of different sets of association parameters means that many interesting models are not nested. Therefore, performing model comparisons using nested models would be inadequate, because competing models with different association parameters cannot be compared. Using information criteria (IC) for model comparisons removes this restriction.

When an IC is used for model selection, the model with the smallest IC is typically selected as the best model and further inferences are based on that model. But there could be many

models with ICs very close to the smallest value, indicating that there is some uncertainty regarding which model is truly the best. Basing estimation and inferences on a single model in this case ignores potentially useful information contained in the competing models, and also ignores the uncertainty associated with deciding which model is “correct”. Model averaging is a technique that can be used to account for model-selection uncertainty in further inferences by considering all the possible models.

Bayesian model averaging (BMA) (Hoeting et al., 1999) computes Bayesian information criterion (BIC) values for each model and uses these to estimate the posterior probability that each model is correct, given the data. Burnham and Anderson (2002) extend the model averaging procedure to other ICs. The quantities computed for each model are no longer posterior probabilities, since the connection to the Bayesian paradigm is lost with the change of IC. They are instead called “evidence weights” and are calculated for each model as below.

Let M be the total number of models fitted, TIC_m and TIC_0 be the TIC for model m and smallest TIC of all the models respectively. Define $\Delta_m = TIC_m - TIC_0 \geq 0$.

Then the evidence weight for model m is defined as,

$$w_m = \frac{e^{-\frac{\Delta_m}{2}}}{\sum_{a=1}^M e^{-\frac{\Delta_a}{2}}}, \quad m = 1, \dots, M$$

This scales the weights so that they resemble probabilities, in that they lie between 0 and 1 and the weights sum to 1 across all models. A model with a high evidence weight is better supported by the data than one with a relatively smaller weight. Evidence weights are also useful for identifying important parameters and estimating them while accounting for uncertainty regarding which model is best.

Let θ be any parameter estimated or fixed to a constant value in the models. Denote the parameter estimate from model m as $\hat{\theta}_m$ and the corresponding variance estimate from that model as $\widehat{Var}(\hat{\theta}_m)$. The model-averaged estimate is given by

$$\hat{\theta}_{MA} = \sum_{m=1}^M \hat{w}_m \hat{\theta}_m,$$

and the variance is estimated by,

$$\widehat{Var}(\hat{\theta}_{MA}) = \sum_{m=1}^M \hat{w}_m \left[(\hat{\theta}_m - \hat{\theta}_{MA})^2 + \widehat{Var}(\hat{\theta}_m) \right].$$

In the case of MRCVs, we are mostly interested in predicting odds ratios for each subtable. Let $\hat{\phi}_{s,m}$ be the estimated OR for sth subtable ($s = 1, \dots, 6IJK$) from the m th model ($m = 1, \dots, M$). The model-averaged estimate of sth $\log(\text{OR})$ is

$$\log(\hat{\phi}_{s,MA}) = \sum_{m=1}^M \hat{w}_m \log(\hat{\phi}_{s,m}),$$

and the variance of the model-averaged estimate of sth $\log(\text{OR})$ is

$$\widehat{Var}(\log(\hat{\phi}_{s,MA})) = \sum_{m=1}^M \hat{w}_m \left[(\log(\hat{\phi}_{s,m}) - \log(\hat{\phi}_{s,MA}))^2 + \widehat{Var}(\log(\hat{\phi}_{s,m})) \right].$$

Thereby a 95% C.I for model-averaged ORs can be obtained.

Chapter 4

Computation of variance in the presence of large number of items

4.1 Limitations of the current method

As explained in Section 2.4.2, marginal loglinear models can be developed to model two MRCVs and the variances of the parameter estimates can be corrected using sandwich methods. This computation requires an estimate, \hat{V} , of $4IJ \times 4IJ$ asymptotic variance matrix of the observed marginal counts from the IRT. Let $T = \text{diag}(\hat{\tau}) - \hat{\tau}\hat{\tau}'$ be a matrix of dimension $2^{(I+J)} \times 2^{(I+J)}$ which denotes the covariance matrix of joint probabilities. Then,

$$\hat{V} = \widehat{V(\mathbf{m})} = nBTB'. \quad (4.1)$$

Notice that when I and/or J are moderately large, T and B can be huge. In one example we worked on, $I = 41$ and $J = 19$, so that T had 2^{120} entries. Merely enumerating the matrix indexes surpasses current machine memory capacity, which ultimately leads to runtime errors. The next section develops a new approach to find the variance of observed counts in the IRT by reducing the dimensions of the matrices.

4.2 Solution to the problem

Elements of T are given by,

$$T_{ij} = \begin{cases} \hat{\tau}_i(1 - \hat{\tau}_i); & i = j \\ -\hat{\tau}_i\hat{\tau}_j; & i \neq j \end{cases}.$$

Since the joint table is typically sparse, most of the elements of $\hat{\tau}$ are zeroes, which leads the corresponding rows and columns of T to become zero too. In order to reduce the size of T , we use the positions of the non-zero counts to identify which rows and columns of T will be zero, and develop a new matrix \tilde{T} by considering only the non-zero elements of $\hat{\tau}$ vector.

Let q be the number of non-zero counts in the joint table—presumably $q \ll 2^{(I+J)}$ —and let $\tilde{\tau}$ be the $q \times 1$ vector of non-zero joint probabilities. Then, \tilde{T} has dimension $q \times q$ and can be obtained as,

$$\tilde{T} = \text{diag}(\tilde{\tau}) - \tilde{\tau}\tilde{\tau}'. \quad (4.2)$$

The matrix B given in equation 2.6 can be written as,

$$B = \begin{bmatrix} g_{11}H & g_{12}H & \cdot & \cdot & g_{12^I}H \\ g_{21}H & g_{22}H & \cdot & \cdot & g_{22^I}H \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ g_{I1}^c H^c & g_{I2}^c H^c & \cdot & \cdot & g_{I2^I}^c H^c \end{bmatrix},$$

where g_{rc} is the (r, c) element of G , $r = 1, \dots, I$ and $c = 1, \dots, 2^I$, and $H^c = (J_{J \times 2^J} - H)$. The value of g_{rc} is either 0 or 1. Therefore, each element given in B is either a matrix of 0's or H (or H^c depending on the row of B). In order to reduce the dimension of B , only columns of B that correspond to non-zero rows of T are retained. Let \tilde{B} be the reduced version of B . Then \tilde{B} can be obtained by considering only certain columns of G and H . The following algorithm explains the procedure to obtain \tilde{B} .

1. First identify row and column positions of non-zero counts in the joint table. Let \mathbf{R} be an $a \times 1$ vector of row positions defined as $\mathbf{R} = (R_1, R_2, \dots, R_a)$ where a is the number of non-zero row positions ($a \leq 2^I$). Also, let $\mathbf{h}_b; b = 1, \dots, a$ be variable-length vectors containing all of the column indices of non-zero counts within the b th row.
2. \mathbf{R} is used to decide which columns of G (and G^c , where $G^c = (J_{I \times 2^I} - G)$) should be used and \mathbf{h}_b is used to decide which columns of H (and H^c) should be used with each column of G .

3. Then \tilde{B} has dimension $4IJ \times q$ can be written in terms of certain columns of G and H as,

$$\tilde{B} = \begin{bmatrix} G[, R_1] \otimes H[, \mathbf{h}_1] & G[, R_2] \otimes H[, \mathbf{h}_2] & \cdots & G[, R_a] \otimes H[, \mathbf{h}_a] \\ G[, R_1] \otimes H^c[, \mathbf{h}_1] & G[, R_2] \otimes H^c[, \mathbf{h}_2] & \cdots & G[, R_a] \otimes H^c[, \mathbf{h}_a] \\ G^c[, R_1] \otimes H[, \mathbf{h}_1] & G^c[, R_2] \otimes H[, \mathbf{h}_2] & \cdots & G^c[, R_a] \otimes H[, \mathbf{h}_a] \\ G^c[, R_1] \otimes H^c[, \mathbf{h}_1] & G^c[, R_2] \otimes H^c[, \mathbf{h}_2] & \cdots & G^c[, R_a] \otimes H^c[, \mathbf{h}_a] \end{bmatrix},$$

where the elements within the square brackets indicate which columns of each matrix are used.

Let's consider the following simple example that illustrates the computation of \tilde{B} . Consider two MRCVs with two items each. Table 4.1 gives the joint table of hypothetical counts.

Table 4.1: Joint table for a hypothetical situation.

	Y_1	0	0	1	1
	Y_2	0	1	0	1
0	0	2	1	0	0
0	1	0	3	0	2
1	0	0	0	0	0
1	1	2	0	0	0
W_1	W_2				

According to Table 4.1, there are 10 observations in total and only 5 non-zero counts. Row positions and column positions of non-zero joint counts can be summarized as given in Table 4.2.

Table 4.2: Cross-classification of row and column positions of non-zero joint counts.

row positions	column positions
1	1 , 2
2	2 , 4
4	1

For an example, there are two non-zero counts in the first row of Table 4.1 and they are at column indices 1 and 2 respectively.

Then $\mathbf{R} = (1, 2, 4)$, and

$$\begin{aligned}
R_1 &= 1, & \mathbf{h}_1 &= (1, 2) \\
R_2 &= 2, & \mathbf{h}_2 &= (2, 4) \\
R_3 &= 4, & \mathbf{h}_3 &= (1).
\end{aligned}$$

Then, \tilde{B} can be obtained as,

$$\tilde{B} = \begin{bmatrix} G[, 1] \otimes H[, (1, 2)] & G[, 2] \otimes H[, (2, 4)] & G[, 4] \otimes H[, (1)] \\ G[, 1] \otimes H^c[, (1, 2)] & G[, 2] \otimes H^c[, (2, 4)] & G[, 4] \otimes H^c[, (1)] \\ G^c[, 1] \otimes H[, (1, 2)] & G^c[, 2] \otimes H[, (2, 4)] & G^c[, 4] \otimes H[, (1)] \\ G^c[, 1] \otimes H^c[, (1, 2)] & G^c[, 2] \otimes H^c[, (2, 4)] & G^c[, 4] \otimes H^c[, (1)] \end{bmatrix}.$$

4.3 Example

By using the above approach, dimensions of T and B can be substantially reduced. This method can be used to compute the variances of the parameter estimates in the presence of any number of items as long as the number of combinations of W and Y that have non-zero counts is manageable.

A real dataset of two MRCVs with 41 and 19 items is analyzed using the standard variance computation method—i.e., $V = nBTB'$ —and using the dimension reduction approach; i.e., $V = n\tilde{B}\tilde{T}\tilde{B}'$. Table 4.3 compares the run times of the two approaches when the number of items is increased gradually.

Table 4.3: Run times (seconds) for computing variance using two different approaches. An “x” means that the program terminated with an error code “Cannot allocate vector of size xxxGB”.

I	J	standard method	dim.reduction method
5	6	2.4	5.1
8	7	x	6.0
10	12	x	6.5
23	18	x	10.5
41	19	x	20.4

According to Table 4.3, the standard method is able to produce results only when the dataset has a very limited number of items. In the other situations it is unable to produce results as it reaches the total allocated memory. However, the new approach generates

results even for the total of 60 items. Appendix A provides the R code for the comparisons between the two methods.

Chapter 5

Analysis Example

5.1 Description of the dataset

The dataset analyzed in this thesis is from the survey of farmers reported in Loughin and Scherer (1998). A total of 279 farmers responded to the following three MRCVs.

1. Which of the following do you test your swine waste for? Binary responses (1=Positive, 0-Negative) are provided for each category.
 - (a) W_1 : Nitrogen
 - (b) W_2 : Phosphorus
 - (c) W_3 : Salt
2. What swine waste disposal methods do you use? Binary responses (1=Positive, 0-Negative) are provided for each category.
 - (a) Y_1 : Lagoon
 - (b) Y_2 : Pit
 - (c) Y_3 : Natural drainage
 - (d) Y_4 : Holding tank
3. What are your primary sources of veterinary information? Binary responses (1=Positive, 0-Negative) are provided for each category.
 - (a) Z_1 : Professional consultant
 - (b) Z_2 : Veterinarian

- (c) Z_3 : State or local extension service
- (d) Z_4 : Magazines
- (e) Z_5 : Feed companies and representatives

The full data set represents a $3 \times 4 \times 5$ structure, and is therefore somewhat cumbersome to explore completely. For the sake of simplicity and ease of demonstration only a $2 \times 2 \times 2$ subset of these data is analyzed. The subset consists of items W_1, W_2, Y_1, Y_2, Z_1 and Z_2 . Hence $I = J = K = 2$ and the IRT is a $2 \times 2 \times 2$ cube, each consisting of a $2 \times 2 \times 2$ subcube of marginal counts. Each cell in the main cube represents one combination of chemical, storage method, and information source, and each subcube represents counts of all combinations of positive and negative responses to those three items. Table 5.1 shows the observed conditional ORs along with 95% confidence intervals between each pair of items for a given level of the third variable. Each of these is computed from one 2×2 table (half of the subcube) within its respective main-cube cell.

Table 5.1: Table of observed conditional ORs along with 95% confidence intervals (in parentheses). Highlighted cells denote ORs that do not include 1.

WY OR	$Z_1 = 0$	$Z_1 = 1$	$Z_2 = 0$	$Z_2 = 1$
$W_1 Y_1$	2.5 (1.1,5.5)	0.9 (0.1,5.1)	3.1 (1.2,8.0)	1.1 (0.4,3.5)
$W_1 Y_2$	1.8 (0.8,3.9)	1.8 (0.4,8.3)	1.6 (0.6,4.1)	2.0 (0.7,5.8)
$W_2 Y_1$	2.9 (1.2,7.4)	1.9 (0.2,18.2)	4.5 (1.4,14.5)	1.4 (0.4,5.0)
$W_2 Y_2$	1.9 (0.8,4.6)	1.2 (0.2,6.3)	1.6 (0.6,4.5)	1.9 (0.6,6.4)

WZ OR	$Y_1 = 0$	$Y_1 = 1$	$Y_2 = 0$	$Y_2 = 1$
$W_1 Z_1$	2.6 (0.5,13.9)	0.9 (0.3,2.5)	1.3 (0.4,4.2)	1.3 (0.4,4.7)
$W_1 Z_2$	2.7 (0.8,8.5)	1.0 (0.4,2.3)	1.3 (0.5,3.2)	1.6 (0.5,4.7)
$W_2 Z_1$	1.9 (0.2,17.1)	1.2 (0.5,3.5)	2.0 (0.6,6.5)	1.3 (0.3,5.4)
$W_2 Z_2$	3.0 (0.7,12.7)	0.9 (0.4,2.4)	1.3 (0.5,3.6)	1.5 (0.4,5.2)

YZ OR	$W_1 = 0$	$W_1 = 1$	$W_2 = 0$	$W_2 = 1$
$Y_1 Z_1$	4.6 (2.0,10.5)	1.6 (0.3,9.1)	4.0 (1.8,8.9)	2.6 (0.3,26.1)
$Y_1 Z_2$	1.9 (1.1,3.3)	0.7 (0.2,2.6)	1.8 (1.1,3.2)	0.6 (0.1,2.9)
$Y_2 Z_1$	1.7 (0.8,3.6)	1.7 (0.4,7.9)	1.8 (0.9,3.7)	1.2 (0.2,6.5)
$Y_2 Z_2$	1.6 (0.9,3.0)	2.0 (0.6,7.3)	1.7 (0.9,3.0)	2.0 (0.5,8.9)

Perusing this table reveals some patterns that anticipate what the more formal analysis will reveal. In particular, all the WZ conditional ORs have confidence intervals that contain 1. This suggests that there is no clear evidence of an association between chemical testing and information source, regardless of the types of waste storage method used. The WY odds ratio is significantly greater than 1 in certain cases that form a clear pattern. The farmers

test for chemicals more often when they use a lagoon than when they do not, but only if they do not get veterinary information from a professional consultant or a veterinarian. However, when they do use either one of these sources of information, there is no apparent association between waste storage method and chemical testing. Further, YZ odds ratios are significantly greater than 1 in certain situations. The farmers use lagoon more often when they use either information source than when they don't, but only if they don't test for any chemicals. But this association does not seem to hold when the farmers test for either one of the chemicals. Also it can be noticed that none of the information sources and none of the chemicals tested for are involved in associations with 'pit' as a waste storage method.

By looking at the patterns of observed conditional ORs one can guess the type of models that would be well fitted to the dataset. In order to identify the models, within-subcube associations and across-subcube associations should be identified. Table 5.2 below summarizes six conditional ORs of each subcube, with a value of 1 assumed if it was not excluded by the respective confidence interval.

Table 5.2: Table of observed conditional ORs for each subcube. Highlighted cells denote significant ORs.

Subcube	$OR_{ij(k,0)}$	$OR_{ij(k,1)}$	$OR_{i(j,0)k}$	$OR_{i(j,1)k}$	$OR_{(i,0)jk}$	$OR_{(i,1)jk}$
(W_1, Y_1, Z_1)	2.5	1	1	1	4.6	1
(W_2, Y_1, Z_1)	2.9	1	1	1	4.0	1
(W_1, Y_1, Z_2)	3.1	1	1	1	1.9	1
(W_2, Y_1, Z_2)	4.5	1	1	1	1.8	1
(W_1, Y_2, Z_1)	1	1	1	1	1	1
(W_1, Y_2, Z_2)	1	1	1	1	1	1
(W_2, Y_2, Z_1)	1	1	1	1	1	1
(W_2, Y_2, Z_2)	1	1	1	1	1	1

In the first four subcubes— (W_1, Y_1, Z_1) , (W_2, Y_1, Z_1) , (W_1, Y_1, Z_2) and (W_2, Y_1, Z_2) —the conditional OR between the two items change across the levels of the third item except for the association $W_i Z_k$ given Y_j . It is equal to 1 and the same across the levels of Y_j . The conditional ORs in these four subcubes can be summarized as,

$$\begin{aligned}
 OR_{ij(k,0)} &\neq OR_{ij(k,1)}; & i = 1, 2 \quad j = 1 \quad \text{and } k = 1, 2 \\
 OR_{(i,0)jk} &\neq OR_{(i,1)jk}; & i = 1, 2 \quad j = 1 \quad \text{and } k = 1, 2 \\
 OR_{i(j,0)k} &= OR_{i(j,1)k} = 1; & i = 1, 2 \quad j = 1 \quad \text{and } k = 1, 2
 \end{aligned}$$

Therefore, within-subcube associations in these four subcubes can be considered heterogeneous (saturated) and is denoted by $(W_i Y_j Z_k)$ for $i = 1, 2, j = 1$ and $k = 1, 2$. According

to Table 5.2, all the conditional ORs of the last four subcubes are equal to 1. Hence all the three items are independent of each other (mutual independence) and can be denoted by (W_i, Y_j, Z_k) for $i = 1, 2, j = 2$ and $k = 1, 2$. However considering all eight subcubes together, a saturated model would be necessary to provide the required flexibility for 1 and non-1 ORs.

According to the results of Table 5.2, it's very clear that the heterogeneous associations within the subcubes change across Y items. They may change across W and Z items as well. We can expect the models with different combinations of these three items change across the subcubes given $(W_i Y_j Z_k)$ within the subcube and they will be well fitted to the given dataset.

5.2 Fitting the models

As described in Section 3.1.3, all 232 possible models are fitted to the dataset. For all the models fitted, TIC and evidence weights are calculated. Models featuring complete independence, one 2-way association, or two 2-way associations within a subcube all have relatively high TIC which lead to essentially zero evidence weights regardless of whether their ORs are allowed to vary across the models. However, when three 2-way associations within a subcube are assumed, some models have relatively smaller TICs than the previous scenario and produce slightly larger weights (around 0.001). When 3-way association is assumed within a subcube, all the models result in relatively smaller TIC and result in higher weights than the previous scenarios as given in Table 5.3. The smallest TIC (304.5) and the highest evidence weight (0.146) are produced by the saturated model (WYZ across the models).

Figure 5.1 shows the cumulative evidence weights produced by the models given in Table 5.3. The first 11 models cover up to 0.98 cumulative weight and the rest of the models do not seem to add much contribution. Subsequently, our model-averaging calculations are based only on these 11 models.

5.3 Model averaging

Model averaging is carried out as explained in Section 3.2.2. Model-averaged ORs and their 95% confidence intervals are given in Table 5.4. The R code is given in Appendix B. The pattern of associations is the same as was seen in the conditional odds ratios in Table 5.1.

Table 5.3: Table of TIC and evidence weights when 3-way association is assumed within a subcube, and this association is allowed to vary across the subcubes according to the pattern given in the first column.

Across-subcube model	TIC	Evidence weights
WYZ	304.5	0.146
WY, WZ, YZ	305.2	0.106
WY, YZ	305.2	0.106
YZ	305.3	0.099
WZ, YZ	305.3	0.098
WY, WZ	305.9	0.073
WY	305.9	0.073
Y + Z	306.0	0.070
W + Y + Z	306.0	0.070
Y	306.0	0.069
W + Y	306.0	0.069
C	312.9	0.002
Z	313.1	0.002
W	313.1	0.002
W + Z	313.3	0.002
WZ	313.5	0.002

Therefore the same conclusions can be made from the model-averaging process.

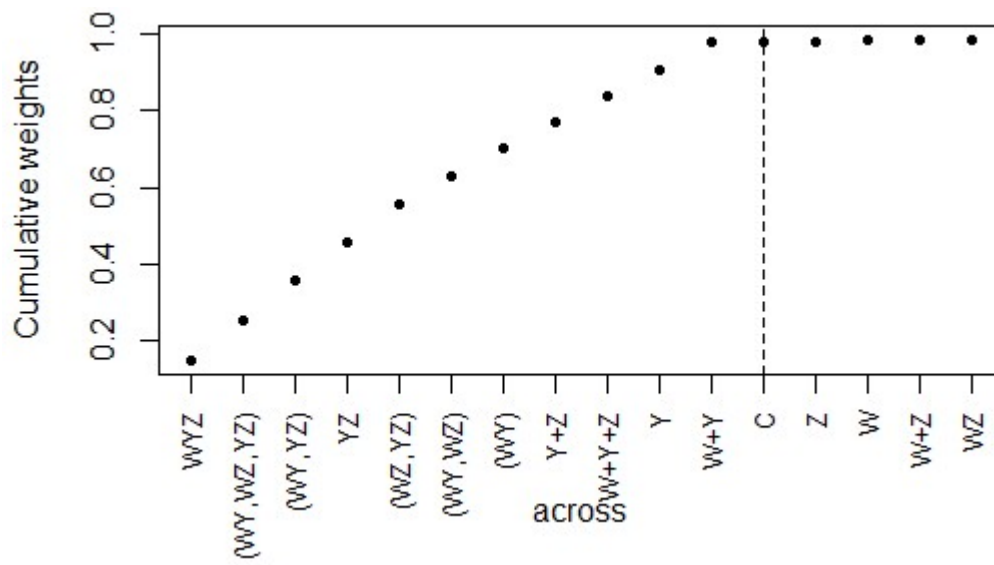
Table 5.4: Table of model-averaged conditional ORs along with 95% confidence intervals (in parentheses). Highlighted cells denote ORs that do not include 1.

WY OR	$Z_1 = 0$	$Z_1 = 1$	$Z_2 = 0$	$Z_2 = 1$
W_1Y_1	2.5 (1.1,5.3)	0.9 (0.2,4.7)	3.2 (1.3,7.9)	1.1 (0.4,3.3)
W_1Y_2	1.8 (0.8,3.9)	1.7 (0.4,6.8)	1.6 (0.7,4.0)	1.9 (0.7,5.5)
W_2Y_1	3.1 (1.2,7.8)	1.4 (0.2,8.4)	4.2 (1.4,12.3)	1.5 (0.4,5.1)
W_2Y_2	1.8 (0.7,4.3)	1.5 (0.4,6.3)	1.7 (0.6,4.4)	1.8 (0.6,5.6)

WZ OR	$Y_1 = 0$	$Y_1 = 1$	$Y_2 = 0$	$Y_2 = 1$
W_1Z_1	2.4 (0.5,11.3)	0.9 (0.4,2.4)	1.4 (0.5,3.9)	1.3 (0.4,4.3)
W_1Z_2	2.7 (0.9,8.3)	1.0 (0.4,2.2)	1.3 (0.5,3.1)	1.5 (0.5,4.5)
W_2Z_1	2.6 (0.4,15.3)	1.2 (0.4,3.1)	1.8 (0.6,5.4)	1.5 (0.4,5.6)
W_2Z_2	2.7 (0.7,10.6)	1.0 (0.4,2.4)	1.3 (0.5,3.4)	1.4 (0.4,4.6)

YZ OR	$W_1 = 0$	$W_1 = 1$	$W_2 = 0$	$W_2 = 1$
Y_1Z_1	4.5 (2.0,9.2)	1.7 (0.4,8.4)	4.2 (1.9,9.2)	1.9 (0.3,11.8)
Y_1Z_2	1.9 (1.1,3.1)	0.7 (0.2,2.4)	1.8 (1.1,3.1)	0.7 (0.1,2.8)
Y_2Z_1	1.7 (0.8,3.5)	1.6 (0.4,6.4)	1.7 (0.8,3.5)	1.4 (0.3,6.3)
Y_2Z_2	1.7 (0.9,3.0)	2.0 (0.6,6.6)	1.7 (1.0,3.0)	1.8 (0.5,6.9)

Figure 5.1: Cumulative weights of top 16 models.



Chapter 6

Conclusions and discussion

The thesis focuses on analyzing and modeling three MRCVs. It extends the approach of Bilder and Loughin (2007), to allow for three MRCVs and explores the potential complications faced when modeling three MRCVs. There are 232 possible models representing different combinations of associations. Parameters are estimated using GEEs generated by a pseudo-likelihood and variances of the estimates are corrected using sandwich methods. Due to the large number of possible models, model comparisons based on hypothesis testing of nested models would be computationally intensive and inefficient. As an alternative, model averaging is proposed as a model comparison tool which can be also used to account for model selection uncertainty.

Further it is noticed that the calculations required for computing the variance of the estimates can exceed 32-bit machine capacity even for a moderately large number of items. This issue is addressed in Chapter 4 by identifying and eliminating rows and columns of zeroes from the sparse matrix. The approach is demonstrated only for two MRCVs but can be extended for three or more MRCVs. The new approach does have certain limitations. It works efficiently only when the joint table is largely sparse; i.e., when the number of non-zero joint counts (q) is much less than the total number of cells in the joint table (2^{I+J}). It is not helpful if q is close to 2^{I+J} . However, this would happen only if either $I + J$ is small or if the sample size is immense, and if all or most combinations of items are at least somewhat likely to occur together.

Chapter 2 reviews the literature related to modeling two MRCVs and presents 6 possible ways of modeling associations. However, Chapter 3 shows that when another MRCV is added, there are 232 possible ways of representing different combinations of associations. Just by adding one more MRCV, the number of possible models increases drastically. One might be interested in identifying the models for 4 or even more MRCVs. This is certainly conceptually possible, but the higher the dimensions the harder it would be to visualize the

patterns in which the odds ratios might change. Therefore the difficulty of constructing and fitting all possible models to higher dimensions would be one of the major challenges in the context of modeling any number of MRCVs.

All the methods and models built so far assume that the simple random sample of units are drawn from the population. But in practice, the surveys may have a complex design and use strategies such as stratification and clustering. When a complex sampling design is used, the observations are not independent of each other, which invalidates the methods and models developed. Therefore, when CATA questions result from a complex survey sample, appropriate inference techniques should be developed.

The models built so far involve only categorical variables. But one might be interested in incorporating continuous covariates into these models and in studying how the associations between MRCVs change across continuous covariates. This would be a great direction for future research.

Bibliography

- Agresti, A. and Liu, I.-M. (1999). Modeling a categorical variable allowing arbitrarily many category choices. *Biometrics*, 55:936–943.
- Bilder, C. and Loughin, T. M. (2001). On the first-order rao-scott correction of the umesh-loughin-scherer statistic. *Biometrics*, 57:1253–1255.
- Bilder, C. and Loughin, T. M. (2004). Testing for marginal independence between two categorical variables with multiple responses. *Biometrics*, 60:241–248.
- Bilder, C. and Loughin, T. M. (2007). Modeling association between two or more categorical variables that allow for multiple category choices. *Communications in Statistics: Theory and Methods*, 36:433–451.
- Bilder, C. and Loughin, T. M. (2014). *Analysis of Categorical Data With R*. CRC Press.
- Burnham, K. and Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2nd edition.
- Gange, S. J. (1995). Generating multivariate categorical variates using the iterative proportional fitting algorithm. *American Statistical Association*, 49:134–138.
- Hoeting, J. et al. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14:382–417.
- Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.
- Loughin, T. M. and Scherer, P. N. (1998). Testing for association in contingency tables with multiple column responses. *Biometrics*, 54:630–637.
- Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, 76:221–230.
- Takeuchi, K. (1976). Distributions of information statistics and criteria for adequacy of models. *Mathematical Science*, 153:12–18.
- Thomas, D. R. and Decady, Y. J. (2004). Testing for association using multiple response survey data: approximate procedures based on the rao-scott approach. *Int. J. Testing*, 4:43–59.

Appendix A

An R Code to compare run times between two methods

The following R code was used to compare run times (seconds) for computing variance using the standard method and the dimension reduction method.

```
rm(list=ls())

##Set the seed
set.seed(558562316)

##Load the packages
library(MRCV)
library(compositions)
library(plyr)

##Read the dataset
data=read.csv('C:\\Users\\menuk\\Google Drive\\thesis
\\analysis\\CATA.csv')
##Change data and I, J accordingly
I =41
J =19
n= nrow(data)

##Compute run time for standard method
start.time <- Sys.time()
W.counts <- as.data.frame(table(data[, 1:I]))
cols <- c(1:I)
W.counts <- W.counts[do.call("order", as.data.frame(
W.counts[,cols]))], ]
Y.counts <- as.data.frame(table(data[, (I+1):(I+J)]))
cols <- c(1:J)
Y.counts <- Y.counts[do.call("order", as.data.frame(
```

```

Y.counts[, cols])), ]
n.counts <- as.data.frame(table(data))
cols <- c(1:(ncol(data)))
n.counts <- n.counts[do.call("order", as.data.frame(
n.counts[, cols])), ]
G <- t(data.matrix(W.counts[, 1:I]) - 1)
H <- t(data.matrix(Y.counts[, 1:J]) - 1)

tau <- n.counts[, ncol(n.counts)]/n
Jr <- matrix(data = 1, nrow = I, ncol = 2^I)
Jc <- matrix(data = 1, nrow = J, ncol = 2^J)
B.matrix <- rbind(kronecker(G, H), kronecker(G,
(Jc - H)), kronecker((Jr - G), H),
kronecker((Jr - G), (Jc - H)))

V <- n * B.matrix %%% tcrossprod((diag(tau) -
tcrossprod(tau)), B.matrix)

end.time <- Sys.time()
time.taken <- end.time - start.time
cat("time□taken", time.taken)

##Run time using mew method

start.time <- Sys.time()

##Function to convert decimal to binary vector
##given the number
convert_to_binary <- function(number, noBits)
{
  number <- number-1
  i <- 0
  string <- numeric(noBits)
  while(number > 0)
  {
    string[noBits - i] <- number %% 2
    number <- number %/% 2
    i <- i + 1
  }
  return(string)
}

##Function to convert binary to decimal
convert_to_decimal <- function(binary_vec)
{
  binary_char <- paste(binary_vec, collapse = "")

```

```

    decimal_val <- unbinary(binary_char)+1

    return(decimal_val)
}

##joint table for non_zero counts
joint_nonzero=count(data, vars = colnames(data))

##n vectorof non zero counts
n.nonzero.vec <- joint_nonzero$freq
Q <- length(n.nonzero.vec)

##tau vectorof non zero probabilities
tau.nonzero.vec <- n.nonzero.vec/sum(n.nonzero.vec)

##T for non_zero probabilities
T.nonzero.mat <- diag(tau.nonzero.vec)-
tcrossprod(tau.nonzero.vec)

##Get non zero positions of joint and
#convert to decimal
rows.nonzero <- apply(joint_nonzero[,1:I],1,
convert_to_decimal)
cols.nonzero <- apply(joint_nonzero[, (I+1):(I+J)],
,1,convert_to_decimal)

g_cols <- unique(rows.nonzero)
B_rows <- 1:4
G_vec <- c(1,1,2,2)
H_vec <- c(1,2,1,2)

##Function to compute B.tilda: performs
#kronecker at each iteration
calculate_B <- function(g_cols,B_rows)
{
    h_cols <- cols.nonzero[which(rows.nonzero==g_cols)]
    H_sub <- sapply(h_cols,function(x) convert_to_binary(x,J))
    G_sub <- as.matrix(convert_to_binary(g_cols,I),ncol=1)

    Jr <- matrix(data = 1,nrow= I, ncol = ncol(G_sub))
    Jc <- matrix(data = 1,nrow= J, ncol = ncol(H_sub))

    G_list <- list(G_sub,Jr-G_sub)
    H_list <- list(H_sub,Jc-H_sub)

    B<- kronecker(G_list[[G_vec[B_rows]]],H_list[[
H_vec[B_rows]]])

```

```

    return(B)
}

B_mat <- sapply(g_cols,function(x) apply(sapply(B_rows,
calculate_B,g_cols=x,simplify = "array"),2,c),
simplify = "array")
B_mat2<-do.call("cbind", B_mat)

##Compute the variance
Var <- n * B_mat2 %*% tcrossprod(T.nonzero.mat, B_mat2)
end.time <- Sys.time()
time.taken <- end.time - start.time
cat("time□taken", time.taken)

```

Appendix B

An R Code to compute model averaged ORs and 95% confidence intervals

The following R code was written to compute model averaged ORs and 95% confidence intervals.

```
rm(list=ls())

##set the seed
set.seed(558562316)

##Load the packages
library(stringr)
library(MRCV)
library(psych)

##Create the dataset
data=farmer3
new_data <- data[,c(1,2,4,5,8,9)]
I=2;J=2;K=2
data<- new_data

##Across-subcube associations
wy_ac <- c("C","W","Y","W+Y","W:Y")
wz_ac <- c("C","W","Z","W+Z","W:Z")
yz_ac <- c("C","Y","Z","Y+Z","Y:Z")

wyz_ac <-c("C","W", "Y","Z","W+Y","W+Z", "Y+Z","W+Y+Z","W:Y",
           "W:Z","Y:Z","W:Y,Y:Z", "W:Y,W:Z","W:Z,Y:Z",
           "W:Y,W:Z,Y:Z", "W:Y:Z")
```

```

wy.wz <- as.vector(outer(wy_ac,wz_ac,paste,sep=","))
wy.yz <- as.vector(outer(wy_ac,yz_ac,paste,sep=","))
yz.wz <- as.vector(outer(yz_ac,wz_ac,paste,sep=","))

wy.wz.yz <- as.vector(outer(wy.wz,yz_ac,paste,sep=","))

first.row <- data.frame(within="C",across="C")
model_table <- expand.grid(c("wi:yj"),wy_ac)
model_table <- rbind(model_table, expand.grid(c("wi:zk"),
wz_ac))
model_table <- rbind(model_table, expand.grid(c("yj:zk"),
yz_ac))
model_table <- rbind(model_table, expand.grid(c("wi:yj,
wi:zk"),wy.wz))
model_table <- rbind(model_table, expand.grid(
c("wi:yj,yj:zk"),wy.yz))
model_table <- rbind(model_table, expand.grid(c("yj:zk,
wi:zk"),yz.wz))
model_table<-rbind(model_table,expand.grid(c("wi:yj,
wi:zk,yj:zk"),wy.wz.yz))
model_table <- rbind(model_table, expand.grid(c("wi:yj:zk"),
wyz_ac))

names(model_table) <- names(first.row)
model_table <- rbind(first.row,model_table)

##Model for complete independence
const = "count~-1+W:Y:Z+wi%in%W:Y:Z+yj%in%W:Y:Z+zk%in%W:Y:Z"
model.list <- list()

###Function to create 1-2 way interactions
within_across <- function(wi,ac, three.way)
{
  terms.ac <- nchar(ac)
  wi.vec <- c(wi)

  if(ac == "C")
  {
    wi.vec <- wi.vec
  }else if(terms.ac == 1){
    within.eff <- paste(wi, ac, sep="%in%")
    wi.vec <- c(wi.vec,within.eff)

  }else if(grepl('+',ac,fixed=TRUE)){

    split.terms <- unlist(strsplit(ac,'+', fixed = TRUE))
    within.eff <- as.vector(outer(wi, split.terms,paste,

```

```

        sep= "%in%"))
    wi.vec <- c(wi.vec,within.eff)

}else if(grepl(':',ac,fixed=TRUE) & three.way == FALSE)
{
    split.terms <- unlist(strsplit(ac,':', fixed = TRUE))
    interact <- paste(wi, ac, sep="%in%")
    within.eff <- as.vector(outer(wi, split.terms,paste,
        sep= "%in%"))
    wi.vec <- c(wi.vec,within.eff, interact)

}else if(grepl(':',ac,fixed=TRUE)&
    grepl(',',ac,fixed=TRUE) & three.way == TRUE){
    split.int.terms <- unlist(strsplit(ac,',',
        fixed = TRUE))
    interact <- paste(wi, split.int.terms, sep="%in%")
    split.terms<-unique(unlist(strsplit(split.int.terms,
        ':',fixed = TRUE)))
    within.eff <- as.vector(outer(wi, split.terms,paste,
        sep= "%in%"))
    wi.vec <- c(wi.vec,within.eff, interact)

}else if(grepl(':',ac,fixed=TRUE)&grepl(',',ac,
fixed=TRUE)== FALSE &three.way == TRUE){
    split.terms<-unlist(strsplit(ac,':',fixed = TRUE))
    interact.three <- paste(wi, ac, sep="%in%")

    combinations.two <- t(combn(split.terms,2))
    interact.two <- apply(combinations.two, 1, paste,
        collapse=":")
    within.eff.two <- as.vector(outer(wi,interact.two,
    paste,sep= "%in%"))
    within.eff <- as.vector(outer(wi, split.terms,
    paste,sep= "%in%"))
    wi.vec <- c(wi.vec,within.eff, within.eff.two,
        interact.three)

}

return(wi.vec)

}

##Function to create 2-2 way and 2-3 way interactions
within_across.two.three.int <- function(wi,ac)
{

```



```

within.vec <- unlist(str_split(wi,","))
across.vec <- unlist(str_split(ac,","))
w.ac.matrix <- rbind(within.vec, across.vec)
row.names(w.ac.matrix) <- NULL
w.ac.matrix<- rbind(w.ac.matrix,rep(FALSE,
ncol(w.ac.matrix)))

int.matrix <- apply(w.ac.matrix, 2,
function(x)do.call(within_across, as.list(x)))

return(int.matrix)
}

##Function to calculate TIC
compute_TIC<- function(model)
{
model <- as.formula(model)
n <- nrow(data)

nvars <- 2 + is.numeric(K)
model.data.unsorted <- MRCV:::data.format(data = data,
I = I,J = J,K = K, nvars = nvars,
add.constant = add.constant)

if (nvars == 2) {
model.data <- model.data.unsorted[order(
-model.data.unsorted$wi,
-model.data.unsorted$yj), ]
}
if (nvars == 3) {
model.data <- model.data.unsorted[order(
-model.data.unsorted$wi,
-model.data.unsorted$yj,
-model.data.unsorted$zk), ]
}

for (i in 1:I) {
parm <- paste("W", i, sep = "")
if (length(agrep(parm, model, max.distance = 0))>0){
model.data <- data.frame(model.data, as.numeric(
(model.data[,1] == names(data)[i])))
colnames(model.data)[ncol(model.data)] <- parm
}
}
for (j in 1:J) {
parm <- paste("Y", j, sep = "")

```

```

    if (length(agrep(parm, model, max.distance = 0))>0) {
      model.data <- data.frame(model.data, as.numeric(
        (model.data[,2] == names(data)[(I + j)])))
      colnames(model.data)[ncol(model.data)] <- parm
    }
  }
  if (nvars == 3) {
    for (k in 1:K) {
      parm <- paste("Z", k, sep = "")
      if (length(agrep(parm, model, max.distance = 0)) >
        0) {
        model.data <- data.frame(model.data, as.numeric(
          (model.data[,3] == names(data)[(I + J + k)])))
        colnames(model.data)[ncol(model.data)] <- parm
      }
    }
  }
}

mod.fit0 <- MRCV:::genloglin.fit(data = model.data,
model = model, nvars = nvars)
X <- model.matrix(mod.fit0)
p <- ncol(X)

mod.fit <- genloglin(data = data, I=I, J = J, K = K,
model = model, boot = FALSE, B=1)
model.est <- mod.fit
mod0 <- mod.fit$mod.fit
mu.hat = as.vector(mod0$fitted.values)
m=mod0$data$count
one.vec = rep(1,8*I*J*K)

Q=(t(X)%*%diag(mu.hat)%*%X)/(8*I*J*K)
omega=(t(X)%*%diag((m-mu.hat)^2)%*%X)/(8*I*J*K)
loglike=2*(t(mu.hat)%*%one.vec-t(m)%*%log(mu.hat)+
          t(lfactorial(m))%*%one.vec)
TIC=loglike+2*tr(solve(Q)%*%omega)

return(list(TIC,p,model.est))
}

##Function to compute weights
compute_weights<- function(TIC.vec)
{
  TIC0 <- min(TIC.vec)
  diff.TIC <- TIC.vec-TIC0

```

```

exp.diff.TIC <- exp(-diff.TIC/2)
one.vec.TIC <- rep(1,length(TIC.vec))
evidence.weight <- exp.diff.TIC/(t(exp.diff.TIC)
%*%one.vec.TIC)

return(evidence.weight)
}

##Function to get the top models
get_top_weight_models<- function(weight)
{

weight.data <- data.frame(model.index=c(1:length(weight))
,weights=weight)
ordered.weight <- weight.data[order(weight.data$weights
,decreasing =TRUE),]
cum_weights <- cumsum(ordered.weight$weights)
weights.plot<- plot(cum_weights[1:20], pch=20,
ylab="cummulative_weights",xlab = "models",
main="Cummulative_weights_of_top_20_models")

return(list(ordered.weight, weights.plot))
}

##Function to compute MA ORs
compute_OR_MA <- function(models, weight)
{

pairs <- c("WY", "YZ", "WZ")
all.OR<- unlist(lapply(pairs,function(x)
as.vector(rownames(predict(models[[length(models)]],
pair=x)$OR.model.asymp))))

OR_table <- data.frame(matrix(0,nrow=length(models) ,
ncol=length(all.OR)))

colnames(OR_table) <- all.OR

var.ln.OR <- data.frame(matrix(0,nrow=length(models),
ncol=length(all.OR)))
colnames(var.ln.OR) <- all.OR

##for each model get predicted OR
for(a in 1:length(models))
{

```

```

model_OR <- unlist(lapply(pairs,function(x)
  predict(models[[a]],pair=x)$OR.model.asymp[,1]))
col.num<- which(names(model_OR) %in% all.OR)
OR_table[a,col.num]<- as.vector(model_OR)
##var(ln ORj,m)
upper <- as.vector(unlist(lapply(pairs,function(x)
  predict(models[[a]], pair=x)$OR.model.asymp[,3])))
lower <- as.vector(unlist(lapply(pairs,function(x)
  predict(models[[a]], pair=x)$OR.model.asymp[,2])))

var.ln.OR[a,col.num]<-((log(upper)-log(lower))/
(1.96*2))^2

}
##get ln_OR
ln.OR.table <- log(OR_table)
lower.ln.OR <- data.frame(matrix(0,nrow=length(models),
                                ncol=length(all.OR)))
colnames(lower.ln.OR) <- all.OR
upper.ln.OR <- lower.ln.OR

for(b in 1:ncol(ln.OR.table))
{
  lower.ln.OR[,b] <- ln.OR.table[,b]
  -1.96*sqrt(var.ln.OR[,b])
  upper.ln.OR[,b] <- ln.OR.table[,b]
  +1.96*sqrt(var.ln.OR[,b])
}

##compute Model averaged ln(OR)
MA.ln.OR=data.frame(MA.ln.OR=apply(ln.OR.table,2,
                                function(x) t(x)%%weight))
MA.OR=as.vector(apply(OR_table,2,
                      function(x) t(x)%%weight))
obs.OR <- as.vector(unlist(lapply(pairs,function(x)
  predict(models[[a]], pair=x)$OR.obs[,1])))

num.col<- c(1:ncol(ln.OR.table))
MA.ln.OR <- data.frame(MA.ln.OR,var.MA.ln.OR=
  unlist(lapply(num.col,function(x) t(weight) %*%
  ((ln.OR.table[,x]-MA.ln.OR[x,1])^2 + var.ln.OR[,x]))))
lower.MA.OR <-exp(MA.ln.OR[,1]-1.96*sqrt(MA.ln.OR[,2]))
upper.MA.OR <- exp(MA.ln.OR[,1]+1.96*sqrt(MA.ln.OR[,2]))
MA.ln.OR <- data.frame(MA.ln.OR, MA.OR=MA.OR,

```

```

obs.OR = obs.OR,lower.MA.OR=lower.MA.OR,
upper.MA.OR=upper.MA.OR)

lower.ln.MA.OR <-MA.ln.OR[,1]-1.96*sqrt(MA.ln.OR[,2])
upper.ln.MA.OR <-MA.ln.OR[,1]+1.96*sqrt(MA.ln.OR[,2])

MA.ln.OR.CI <- data.frame(MA.ln.OR,lower=
lower.ln.MA.OR,upper=upper.ln.MA.OR)

return(MA.ln.OR)
}

##Identify the interaction type in 'within'
for(i in 1:nrow(model_table))
{
wi <- as.character(model_table$within[i])
ac <- as.character(model_table$across[i])

if(wi=="C")
{
model_table[i,3]="independent"
vec <- c()

}else if(str_count(wi,",")==0 &
str_count(wi,":")== 1){
model_table[i,3]="1-2_away"
vec <- within_across(wi,ac,FALSE)

}else if(str_count(wi,",")==0 &
str_count(wi,":")== 2){
model_table[i,3]="1-3_away"
vec <- within_across(wi,ac,TRUE)
vec <- c("wi:yj","wi:yj%in%W","wi:yj%in%Y","wi:yj%in%Z"
,"wi:yj%in%W:Y","wi:yj%in%W:Z", "wi:yj%in%Y:Z",
"wi:yj%in%W:Y:Z","wi:zk", "wi:zk%in%W",
"wi:zk%in%Y","wi:zk%in%Z","wi:zk%in%W:Z",
"wi:zk%in%Y:Z","wi:zk%in%W:Y",
"wi:zk%in%W:Y:Z","yj:zk", "yj:zk%in%W",
"yj:zk%in%Y", "yj:zk%in%Z", "yj:zk%in%Y:Z",
"yj:zk%in%W:Z","yj:zk%in%W:Y","yj:zk%in%W:Y:Z",vec)

}else if(str_count(wi,",") ==1){
model_table[i,3]="2-2_away"
vec <- within_across.two.three.int(wi, ac)

```

```

}else if(str_count(wi,",") ==2){
  model_table[i,3]="3-2-way"
  vec <- within_across.two.three.int(wi, ac)
}

model_table[i,4] <- ifelse(length(vec)==0,const,
paste(const,paste(unlist(vec), collapse = "+")
, sep = "+"))
}

##compute TIC and P
list.TIC=lapply(model_table[,4],compute_TIC)
row.TIC= c(1:nrow(model_table))
TIC.p.list=lapply(row.TIC,function(x)unlist(
list.TIC[[x]][1:2]))

TIC.p=matrix(unlist(TIC.p.list),ncol=2,byrow=TRUE)
model_table = cbind(model_table, TIC.p)

##Get the names and the estimates
estimates_names <- lapply(row.TIC, function(x)
names(summary(list.TIC[[x]][3][[1]]))
$coefficients[,1]))
estimates_est <- lapply(row.TIC, function(x)
as.vector(summary(list.TIC[[x]][3][[1]]))
$coefficients[,1]))
estimates_SE <- lapply(row.TIC, function(x)
as.vector(summary(list.TIC[[x]][3][[1]]))
$coefficients[,2]))

##calaculate evidence weights

model_table[,7] <- compute_weights(model_table[,5])

sorted=sort(model_table[,7],decreasing = TRUE)[1:11]

higher_weights=model_table[model_table$V7 %in%
sorted,c(1,2,6,7)]

```

```

saturated_terms <- estimates_names[[length(estimates_names)]]

### compute MA_OR
plot.weight <- get_top_weight_models(model_table[,7])[[2]]
##Change the threshold accordingly
weight.thresh <- 11
abline(v=weight.thresh,lty=2)

top_weights<- get_top_weight_models(model_table[,7])[[1]]
$weights[1:weight.thresh]
top_models_index <- get_top_weight_models(model_table[,7])[[1]]
$model.index[1:weight.thresh]

top_models <- lapply(top_models_index, function(x)
                    list.TIC[[x]][3][[1]])

##Scale the weights
scale_weights <- top_weights/sum(top_weights)
MA.OR <- compute_OR_MA(top_models,scale_weights)

colnames(model_table)<- c("within","across", "type","model",
"TIc","Para","evidence_weights")

```