Modeling and Measuring Association for Ordinal Data

A Thesis
Submitted to the Faculty of Graduate Studies and Research
In Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Statistics
University of Regina

by
Huihan Gong
Regina, Saskatchewan
July 2012

Copyright 2012: Huihan Gong
Huihan Gong, candidate for the degree of Master of Science in Statistics, has presented a thesis titled, *Modeling and Measuring Association for Ordinal Data*, in an oral examination held on July 16, 2012. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner: Dr. Lisa Fan, Department of Computer Science  
Supervisor: Dr. Dianliang Deng, Department of Mathematics & Statistics  
Committee Member: *Dr. Andrei Volodin, Department of Mathematics & Statistics  
Committee Member: Dr. Yang Y. Zhao, Department of Mathematics & Statistics  
Chair of Defense: Dr. Nader Mobed, Department of Physics  

*Not present at defense*
ABSTRACT

We propose the modified Pearson goodness-of-fit statistic and consider it in the cumulative contingency tables. For the non-cumulative tables, we provide the scoring systems and the variance minimization correction term to Pearson goodness-of-fit statistic. We also put forward the average of individual odds ratios as the common odds ratio and modify the Mantel-Haenszel estimator of common odds ratio in a $K \times K$ ordinal contingency table where the category numbers of the variables are the same. The $M^2$ statistic with scoring systems is also introduced. From the simulation results, we figure out that the chi-square statistics do not work except for those without the ordinal information. However, the $M^2$ statistic accurately detects the association between ordinal variables in a contingency table no matter if it is sparse or not. The scoring systems can be considered together with the $M^2$ statistic. The estimators of common odds ratio are appropriate as well. However, the estimators of common odds ratio are more appropriate for the lightly sparse tables because the test errors are relatively larger for the extremely sparse tables. The Mantel-Haenszel estimator of common odds ratio works better than the average of individual odds ratios in the extremely sparse tables.
ACKNOWLEDGEMENTS

Professor Dianliang Deng has been the ideal thesis supervisor. His sage advice and patient encouragement aided the writing of this thesis. I deeply appreciate the academic and financial support I have received from Professor Deng. I would like to acknowledge the contribution of Professor Yang Zhao and Professor Andrei Volodin, who have helped me to expand my horizons and broaden my knowledge.

This research is partially supported by FGSR (Faculty of Graduate Studies and Research), teaching assistantships and research assistantships provided by the Department of Mathematics and Statistics.
DEDICATION

This thesis is dedicated to my parents, who selflessly love and support me. It is also dedicated to my best friend Xiaoyan Lei, who assisted me with my studies.
# Contents

Abstract ......................................................... ii
Acknowledgements .............................................. iii
Dedication ......................................................... iv

1 Introduction and Motivation .................................. 1

2 Literature Review .............................................. 5
   2.1 Ordinal Responses ......................................... 5
   2.2 Models for Ordinal Responses .............................. 7
      2.2.1 General Cumulative Model ............................. 7
      2.2.2 Proportional Odds Model .............................. 9
      2.2.3 Proportional Hazards Model ......................... 11
      2.2.4 Extreme-maximal-value Distribution Model ........ 12
   2.3 Statistics for Measuring the Independence in the Contingency Tables 13
      2.3.1 Pearson Goodness-of-fit Statistic .................. 13
      2.3.2 Odds Ratio ........................................... 14
      2.3.3 $M^2$ Statistic ....................................... 19
2.4 Related Work ................................................. 21
  2.4.1 Multivariate Probit Model .............................. 22
  2.4.2 Semiparametrically Structured Ordinal Model ........ 23

3 Modification of Test Statistics in the Ordinal Contingency Tables 24
  3.1 Modified Pearson Goodness-of-fit Statistic .................. 25
    3.1.1 Modifications of Pearson Goodness-of-fit Statistic in the Cumulative Contingency Tables .......... 25
    3.1.2 Scoring Systems ........................................... 29
    3.1.3 Variance Minimization Correction Term .................... 34
  3.2 Modified Odds Ratio .......................................... 37
    3.2.1 Average of the Individual Odds Ratios .................... 37
    3.2.2 Mantel-Haenszel Odds Ratio in the $K \times K$ Ordinal Contingency Tables ......................... 38
    3.2.3 Add-on Value for the Extremely Sparse Tables .............. 39
  3.3 Modified $M^2$ Statistic ...................................... 41

4 Simulation .................................................. 42
  4.1 Data Generation ............................................. 42
  4.2 Data Processing .............................................. 44
    4.2.1 Blocks .................................................... 44
    4.2.2 The Compressed Contingency Tables ....................... 45
    4.2.3 Degree of Freedom ....................................... 45
  4.3 Hypotheses and the Criteria of Test Statistics .............. 46
    4.3.1 Hypotheses ............................................. 46
4.3.2 The Criteria of Test Statistics ........................................ 46
4.4 Results of Simulation .................................................... 46
4.5 Summary of Simulation .................................................. 52

5 Applications ................................................................... 53
  5.1 Data Sources and Features .............................................. 53
    5.1.1 Data Sources .......................................................... 53
    5.1.2 Features ............................................................... 54
  5.2 Test Results of the Medical Data ..................................... 60

6 Future Work and Conclusion ............................................. 66
  6.1 Future Work ............................................................. 66
    6.1.1 Analyzing the Longitudinal Data ............................... 66
    6.1.2 Non-uniform Scores ............................................... 67
    6.1.3 The Estimators of Common Odds Ratio in the $I \times J$ Ordinal Contingency Tables ......................... 68
  6.2 Conclusion .............................................................. 69

Bibliography ...................................................................... 71
List of Tables

2.1 A contingency table: example 1 ........................................ 16
3.1 A contingency table: example 2 ........................................ 28
3.2 The cumulative contingency table regarding Table 3.1 ............ 28
3.3 Van der Waerdon scores .................................................. 31
3.4 Joint scores (the combination of exponential scores and Van der Waer-
don scores) ................................................................. 33
3.5 Sparse contingency table: an example .................................. 40
4.1 A compressed table based on Table 3.4 ................................. 45
4.2 The simulation results of Chi-square and $M^2$ statistics in the non-
cumulative tables ............................................................ 47
4.3 Simulation results in the cumulative tables ............................ 50
4.4 Simulation results of the estimators of common odds ratio in the non-
cumulative tables ............................................................ 51
5.1 Deep pain sensations (X) vs. surface pain sensations (Y) at baseline
time ................................................................. 55
5.2 Deep pain sensations (X) vs. surface pain sensations (Y) at study visit time 1 ................................................................. 56
5.3 Vibration detection mode 1 (X) vs. vibration detection mode 2 (Y) at baseline time .......................................................... 57
5.4 Vibration detection mode 1 (X) vs. vibration detection mode 2 (Y) at study visit time 1 ....................................................... 58
5.5 Disease stage (X) vs. pain intensity (Y) at baseline time .......... 59
5.6 Test results of deep pain sensations vs. surface pain sensations . . . . 61
5.7 Test results of two modes of vibration detection .................. 62
5.8 Test results of disease stage vs. pain intensity ...................... 63
Chapter 1

Introduction and Motivation

Ordinal data usually arises in practice, especially in medical studies, assessment investigations, industrial quality assurances, etc. In the early days, especially before 1980, the analyzing methods for ordinal responses are centered on Likert’s Scale analysis. Chimka and Wolfe (2009) reviewed those methodologies of ordinal variable analysis before 1980. The regression models for ordinal data that proposed by McCullagh (1980) are a part of those remarkable ones in the last few decades. The proportional odds model and the proportional hazards model are widely used in practice because of the simplicity of the interpretation, see Fahrmeir and Tutz (1994) for details. A series of methodologies regarding the modeling of ordered categorical data have been put forward since then. According to Fahrmeir and Tutz (1994), cumulative models referring to the threshold approach, extended version of cumulative models, sequential models and two-step models can be considered for ordinal responses. Anderson (1984) described qualitative logistic regression model, stereotype ordered regression model and other types of models for a qualitative, categorical response variable. Beitler
and Landis (1985) proposed a two-way ANOVA model for multinomial data which is directly analogous to the mixed model for quantitative data. Varin and Czado (2010) composed the mixed autoregressive ordinal probit models with normally distributed random effects. Tutz (2003) put forward the generalized semiparametrically structured ordinal models for which the predictors are determined by global effects and category-specific effects.

As for the methodologies of detecting the association between ordinal responses, a great amount of the methods are extended based on the ordinal contingency tables. Some of the tables of real data sets are extremely sparse. Dong and Simonoff (1995) proposed geometric combination estimators for the \( d \)-dimensional sparse ordinal contingency tables. They showed these estimators are non-negative, and even for relatively small tables they are superior to non-negative kernel estimators. Koehler (1986) presented the goodness-of-fit test for log-linear models in sparse contingency tables. The normal approximation is more accurate in compare with the chi-squared approximation in sparse tables. Through the techniques of order-restricted inference, Kimeldorf, et al (1992) obtained the min and max scorings to analyze two-sample ordinal data. The choices of scores are distinguished if the range of the minimum and maximum values includes the critical value. The generalized Mantel-Haenszel estimators for \( K \times 2 \times J \) tables are provided by Greenland (1989) and were used for testing the “sudden infant death syndrome” data.

Here we introduce a real data set which is from a project of medical related studies. Except for the fundamental information of the patients such as age, gender, race, history of several types of disease and medication records. The data set also
involves a sort of ordinal variables which have 11 categories at most. However, since the patients drop out of the project over time, the number of observations is 33 at baseline time which is the most over the whole period of this project. We are interested in the associations between the ordinal variables, e.g. the deep pain sensations and surface pain sensations. Yet the contingency tables constructed by those ordinal variables are extremely sparse. For example, when the number of categories is 11 for both of the ordinal variables and a total 33 records are observed, the ordinal contingency table has 33 observations out of 121 grids. On average, each grid has 0.2727 observation in it, so that the table is extremely sparse.

In order to detect the association between ordinal variables in an extremely sparse contingency table, based on the cumulative models for ordinal responses, we provide the modified Pearson goodness-of-fit statistic, odds ratios and $M^2$ statistic. One of the modification methods is using cumulative probabilities and entries based on the cumulative models for ordinal responses. We also attempt the modifications regarding the cumulative tables, scoring systems and first-order correction term for Pearson goodness-of-fit statistic. As for the odds ratios, we propose the average of individual odds ratios, modify the Mantel-Haenszel estimator of common odds ratio and consider an add-on value when the contingency table is extremely sparse. Since the $M^2$ statistic involves the correlation coefficient where the category numbers are involved, the scoring systems can be considered as well.

In this thesis, we ascertain the following proper test statistics for the ordinal contingency tables which are extremely sparse. The modified Pearson goodness-of-fit statistics are not proper, but the estimators of common odds ratio and the $M^2$ statistic
in regards to the correlation coefficient are considered to be the appropriate statistics
to detect the association between ordinal variables. The estimators of common odds
ratio are better be used on the lightly sparse contingency tables. Several scoring
systems can be considered when using the $M^2$ statistic.

The remainder of this thesis is organized as follows. In Chapter 2, we give the
classical regression models for ordinal responses, the test statistics for detecting the
association between the ordinal variables, and the related work in the last few years.
The modification methods of test statistics for the ordinal contingency tables are
introduced in Chapter 3. In Chapter 4, we generate the simulated samples from
a uniform distribution with 1000 replications. Through the simulated results, we
filter the appropriate test statistics for the sparse ordinal contingency tables. The
application to the medical data is introduced in Chapter 5. The last chapter provides
the possible future work and conclusion.
Chapter 2

Literature Review

2.1 Ordinal Responses

Response variables that have more than two categories often are ordinal. That means the events described by the category numbers 1, ..., k can be considered as ordered, see Fahrmeir and Tutz (1994) for details. The values of a variable can be ranked, but the real distance between adjacent categories is unknown and non-numerically valued. Ordinal data can be absorbed from many statistical researches such as medical studies, assessment investigations, industrial quality assurances, radiology, psychology, biomedicine, biology and sociology.

According to Anderson’s study (1984), there are two major types of ordinal variables. The first type is “grouped continuous” which is related to a single, underlying continuous variable directly. In other words, it can be considered as the categorized version of a continuous variable. For example, McCullagh (1980) mentioned that the ordered categories of “Family income distribution in constant (1973) dollars for
Northeast U.S.” are 0-3000, 3001-5000 and so on. Also in the example of Fahrmeir and Tutz (1994), the “breathing test results” are ordered in three categories which are “normal”, “borderline” and “abnormal”. If this test result is measured by the volume of breath, and the three categories correspond to intervals of volume scale, then the variable “breathing test result” is a grouped continuous variable.

The other type of ordinal variables is “assessed ordered”. It is generated by an assessor who processes an indeterminate amount of information before providing his judgment of the grade of the ordered variable, see Anderson (1984) for details. For example, Varin and Czado (2010) used the headache diary data with categories “no headache”, “mild headache”, “moderate headache”, “painful headache”, “severe headache” and “intense headache”. According to Anderson (1984), in principle, there is a single, unobservable, continuous variable related to this ordered scale, but in practice, the descriptions are made due to some other information such as the durative of the headache and the activities or performance of the patients. One may use these related information to “assess” the category of certain observation.

To compare the two types of ordinal variables, since the assessment is based on many unknown underlying vectors, the grouped continuous approach is usually less attractive for assessed ordered categorical variables, and in principle it can usually be observed itself, see Anderson and Pemberton (1985) for details.
2.2 Models for Ordinal Responses

2.2.1 General Cumulative Model

The cumulative model is one of the most widely used models for ordinal responses based on the category boundaries or threshold approach. It is assumed that the observable variable $Y$ is merely a categorized version of a latent continuous variable $U$, see Fahrmeir and Tutz (1994) for details. For those “grouped continuous” ordinal variables, the latent variable $U$ can be considered as the unobservable underlying variable. While for those “assessed ordered” ordinal variables, $U$ is “other information” leading to the judgment on the underlying continuous scale. In both cases, one can use the latent variable $U$ to construct the cumulative model. The benefit of using a latent variable is that it can make interpretation simpler. However, others without reference to the underlying continuous variable may also apply for constructing the cumulative model.

**Definition 2.2.1.** Assume $x$ is a vector of the observed explanatory variables, $Y \in \{1, ..., k\}$ is the observable variable and $U$ is the unobservable latent variable. In the threshold approach, the response $Y$ and the latent variable $U$ are connected in such a way that the relationship is monotone.

\[(i)\] $Y = r \quad \Leftrightarrow \quad \lambda_{r-1} < U \leq \lambda_r, \quad r = 1, ..., k;$

\[(ii)\] $U = -x'\gamma + \epsilon.$

In the definition, $-\infty = \lambda_0 < \lambda_1 < ... < \lambda_k = \infty$. This is a necessary and sufficient condition and we will talk about the reason in the following part. $\gamma = (\gamma_1, ..., \gamma_p)'$ is
a vector of regression coefficients and $\epsilon$ denotes a random variable with cumulative distribution function $F$. So that latent variable $U$ is categorized into observable variable $Y$ determined by the thresholds $\lambda_1, ..., \lambda_{k-1}$.

From the assumptions above, one can easily give out the cumulative model for observed variable $Y$ as follows

$$P(Y \leq r \mid x) = F(\lambda_r + x'\gamma), \quad (2.2.1)$$

or in linearized form

$$F^{-1}(p_r) = \lambda_r + x'\gamma. \quad (2.2.2)$$

Since $Y$ is the categorized version of $U$, so the elements of $Y$ are independent from each other. The left side of equation (2.2.1) is the sum of the probabilities for each category $P(Y = 1 \mid x) + ... + P(Y = r \mid x)$. From this point of view, equation (2.2.1) is called cumulative model with cumulative distribution function $F$. On the other hand, because of the derivation based on the thresholds $\lambda_1, ..., \lambda_{k-1}$ of the latent variable $U$, equation (2.2.1) can also be considered as a threshold model. The linearized form of the cumulative model is shown in equation (2.2.2), $p_r$ denotes the cumulative response probability $P(Y \leq r)$.

From the previous part, because the cumulative model involves the latent variable, the condition $-\infty = \lambda_0 < \lambda_1 < ... < \lambda_k = \infty$ is necessary as well as sufficient. It ensures that the probabilities of the response categories are non-negative for all of the covariates $x$ and the regression coefficients $\gamma$. If this condition applies, by equation (2.2.2), the cumulative respond probabilities are ordered by $0 \leq p_1 \leq p_2 \leq ... \leq p_{k-1} \leq 1$. 

8
2.2.2 Proportional Odds Model

From the previous part, we know the general cumulative model has the form that the left hand side of the equation is a cumulative probability while the right side of the equation is a cumulative distribution function $F$. There are several distinct distribution functions $F$ leading to specific cumulative models. One of the commonly chosen distribution functions is the logistic distribution function

$$F(x) = \frac{1}{1 + \exp(-x)}.$$ 

The cumulative model corresponding to this distribution function then comes to

$$P(Y \leq r \mid x) = \frac{\exp(\lambda_r + x'\gamma)}{1 + \exp(\lambda_r + x'\gamma)}, \quad r = 1, ..., q, \quad q = k - 1. \tag{2.2.3}$$

The equivalent forms are given by

$$\log \left\{ \frac{P(Y \leq r \mid x)}{P(Y > r \mid x)} \right\} = \lambda_r + x'\gamma, \tag{2.2.4}$$

or

$$\frac{P(Y \leq r \mid x)}{P(Y > r \mid x)} = \exp(\lambda_r + x'\gamma). \tag{2.2.5}$$

The cumulative odds of the event $Y \leq r$ on the left side of equation (2.2.5) satisfies

$$\frac{P(Y \leq r \mid x)}{P(Y > r \mid x)} = \frac{P(Y \leq r \mid x)}{1 - P(Y \leq r \mid x)}.$$ 

Then the logarithms of the cumulative odds

$$\log \left\{ \frac{P(Y \leq r \mid x)}{P(Y > r \mid x)} \right\}, \quad r = 1, ..., q, \quad q = k - 1,$$
are determined by a linear form of the explanatory variables. Thus, equation (2.2.4) is an ordinal logistic regression, also called logistic cumulative model. Because of the special property of using odds, this model can be called a proportional odds model as well.

Consequently, if two populations are considered, and they have the explanatory variables $x$ and $z$ respectively, the ratio of the cumulative odds of the event $Y \leq r$ corresponding to $x$ and $z$ is

$$\frac{P(Y \leq r \mid x)/P(Y > r \mid x)}{P(Y \leq r \mid z)/P(Y > r \mid z)} = \exp\{(x - z)'\gamma\}.$$  

It is clear to see that the ratio is independent of category number $r$. That means the ratio of the cumulative odds for two populations is the same for all of the cumulative odds, see Fahrmeir and Tutz (1994) for details. It is only affected by the explanatory variables $x$ and $z$, and the vector of coefficients $\gamma$.

Another consideration is that, according to Armitage and Colton (2005), suppose matters from the same population such that $x_0 = 0$ is the baseline value of the covariates, it follows that $\exp(\lambda_r)$ is the baseline odds for the event $Y \leq r$. While $x_1$ is a non-baseline covariate value. From the following equation

$$\frac{P(Y \leq r \mid x_1)/P(Y > r \mid x_1)}{P(Y \leq r \mid x_0)/P(Y > r \mid x_0)} = \exp\{(x_1 - x_0)'\gamma\},$$

we can obtain the response odds at a non-baseline covariate value by multiplying the baseline odds and the factor $\exp\{(x_1 - x_0)'\gamma\}$. Note that, the baseline odds can be arbitrary.
2.2.3 Proportional Hazards Model

If the cumulative distribution function $F$ is chosen as the extreme-minimal-value distribution, that is

$$F(x) = 1 - \exp(-\exp(x)),$$

which is the distribution function of the logarithm of a random variable having the $F$ distribution, see Pettitt (1984) for detail. One can plug this function into equation (2.2.1) to get the model

$$P(Y \leq r \mid x) = 1 - \exp[-\exp(\lambda_r + x'\gamma)], \quad r = 1, ..., q, \quad q = k - 1,$$

(2.2.6)

or equivalently the complementary log-log function

$$\log\{-\log[1 - P(Y \leq r \mid x)]\} = \lambda_r + x'\gamma.$$  

(2.2.7)

In model (2.2.7), the right part of the equation is the linear regression of covariance $x$ which determines the cumulative hazard function $-\log[1 - P(Y \leq r \mid x)]$ on the left part. The first derivative of the cumulative hazard function is the proportional hazards to which the model is corresponding

$$\frac{d}{dr}\{-\log[1 - P(Y \leq r \mid x)]\} = \exp(\lambda_r + x'\gamma)\lambda'_r.$$  

(2.2.8)

Based on this point, the model is called grouped Cox model since it may be derived as a grouped version of the continuous Cox or proportional hazards model, which is a statistical technique for exploring the relationship between the survival of a patient and several explanatory variables, well known in survival analysis, see Fahrmeir and Tutz (1994) for details.
Comparing the densities of the latent response $U$ of proportional odds model with proportional hazards model, for small values of covariance $x$, the densities are rather similar. In other words, the models often yield to equivalent fits for small values of explanatory variables.

### 2.2.4 Extreme-maximal-value Distribution Model

When $F(x)$ is another type of extreme value distribution function, the extreme-maximal-value distribution function

$$F(x) = \exp(-\exp(-x)).$$

The cumulative model based on $F$ is changed to

$$P(Y \leq r \mid x) = \exp\{-\exp[-(\lambda_r + x'\gamma)]\}, \quad r = 1, ..., q, \quad q = k - 1, \quad (2.2.9)$$

or equivalently to log-log links

$$\log[-\log P(Y \leq r \mid x)] = -(\lambda_r + x'\gamma). \quad (2.2.10)$$

Model (2.2.6) is not equivalent to model (2.2.9), but it can be estimated by the latter one. To be more specific, suppose a variable $Y' = -Y + k + 1$ with inverse ordering of categories, then plug $Y'$ into model (2.2.9) to estimate. Next, multiply the obtained estimates by -1 to yield the estimates for model (2.2.6), also the order of the threshold parameters $\lambda_r$ has to be reversed, see Fahrmeir and Tutz (1994) for the detail.
2.3 Statistics for Measuring the Independence in the Contingency Tables

There is an implied condition for constructing models for ordinal response, that is, the responses have an association with the coefficients. Therefore, measuring the independence of the variables is an essential part of modeling. Looking back to the statistical analysis history, there are several representative statistics for measuring the independence of ordinal responses.

2.3.1 Pearson Goodness-of-fit Statistic

Pearson goodness-of-fit statistic also well know as Chi-square ($\chi^2$) statistic was introduced by Karl Pearson in 1900. It had a revolutionary impact on categorical data analysis, which had focused on describing associations, see Agresti (2002) for details. Following the previous notation, consider two ordinal responses with $I$ and $J$ categories respectively. A test of null hypothesis and alternative hypothesis

$$H_0 : \text{Two ordinal variables are independent;}$$

$$H_a : \text{Two ordinal variables are associated.}$$

is set up. The test statistics $X^2$ based on the hypotheses equals

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \sim \chi^2_{(I-1)(J-1)},$$

(2.3.1)

where $n_{ij}$ denotes the entry on the $i$th row and $j$th column of the contingency table that built by two categorical variables. Under the null hypothesis, the expected
values of \( n_{ij} \), called expected frequencies, are \( \mu_{ij} = E(n_{ij}) = Np_{i+}p_{+j} \) while the marginal proportions \( \{p_{i+}\} \) and \( \{p_{+j}\} \) are usually unknown. But we can use maximum likelihood to estimate their values. They are the sample marginal frequencies out of the sample size \( N \), \( \hat{p}_{i+} = n_{i+}/N \) and \( \hat{p}_{+j} = n_{+j}/N \). Correspondingly, the expected frequencies estimated from the sample are \( \{\hat{\mu}_{ij} = N\hat{p}_{i+}\hat{p}_{+j} = n_{i+}n_{+j}/N\} \).

The test statistic asymptotically follows chi-squared distribution under the null hypothesis. Under the assumption that two ordinal variables are independent, the degree of freedom is given by

\[
df = (IJ - 1) - (I - 1) - (J - 1) = (I - 1)(J - 1).
\] (2.3.2)

It is known that the contingency table is constructed by two ordinal variables with \( I \) and \( J \) categories separately, hence the marginal proportions of rows and columns \( \{\pi_{i+}\} \) and \( \{\pi_{+j}\} \) both sum to 1.

### 2.3.2 Odds Ratio

Odds ratio is another kind of fundamentally used statistic to measure the association between two binary variables. The ordinal variables can be processed to meet the binomial condition which is required when using odds ratio. In this part, we will go through the constitution of odds ratio, the odds ratios in \( I \times J \) ordinal contingency tables, a widespread estimator of the common odds ratio and the hypotheses as well as the criteria.

#### 1. The Constitution of Odds Ratio in General

The most simple odds ratio is for \( 2 \times 2 \) contingency tables. According to Agresti (2007), we give the definition of the general odds ratio as follows.
Definition 2.3.1. In a $2 \times 2$ ordinal contingency table, the probability of “success” is $\pi_1$ in row 1 and $\pi_2$ in row 2. Within row 1, the odds of success are defined to be

$$odds_1 = \frac{\pi_1}{1 - \pi_1};$$

within row 2, the odds of success are defined to be

$$odds_2 = \frac{\pi_2}{1 - \pi_2}.$$

The odds ratio is defined by

$$\theta = \frac{odds_1}{odds_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}.$$

2. Individual odds ratio

According to Pigeot (1990), in the situation of $K$ pairs of independent binomial random variables $X_{1r}$ and $X_{2r}$ with sample sizes $N_{1r}$ and $N_{2r}$ and the corresponding success probabilities $\pi_{1r}$ and $\pi_{2r}$, each individual odds ratio $\theta_{I(r)}$ is defined as

$$\theta_{I(r)} = \frac{\pi_{1r}/(1-\pi_{1r})}{\pi_{2r}/(1-\pi_{2r})}, \quad r = 1, ..., K.$$

3. Odds Ratios in the $I \times J$ Ordinal Contingency Tables

Next, we consider the odds ratios in an $I \times J$ contingency table where one ordinal variable contains $I$ categories while the other has $J$ categories. The following table is an example of the contingency tables.
Table 2.1: A contingency table: example 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Local odds ratio

Due to Agresti (2002), the local odds ratios use the cells in adjacent rows and adjacent columns. The expression of the local odds ratio is

$$\theta_{L(i,j)} = \frac{\pi_{i,j} / \pi_{i+1,j+1}}{\pi_{i+1,j} / \pi_{i+1,j+1}} = \frac{\pi_{i,j} \pi_{i+1,j+1}}{\pi_{i,j+1} \pi_{i+1,j}}, \quad i = 1, ..., I - 1, \quad j = 1, ..., J - 1. \quad (2.3.3)$$

In equation (2.3.5), $\pi_{i,j}$ is the probability that the entry of the cell on row $i$ and column $j$ divided by the population. $\pi_{i+1,j}$ and $\pi_{i,j+1}$ are the adjacent entries to $\pi_{i,j}$.

Particularly, in Table 2.1, when $r = 4$ and $s = 2$, the local odds ratio is equivalent to

$$\hat{\theta}_{L(4,2)} = \frac{2 \times 2}{1 \times 1} = 4.$$

Since one can use each of $\binom{I}{2} = I(I - 1)/2$ pairs of rows in combination with each of the $\binom{J}{2} = J(J - 1)/2$ pairs of columns in an $I \times J$ table, the number of different subsets of the local odds ratios is $(I - 1)(J - 1)$, see Agresti (2007) for details.
Odds ratio of rectangular pattern of cells

Another basic set according to Agresti (2002) is given by

\[
\theta_{R(i,j)} = \frac{\pi_{ij}}{\pi_{iJ}} = \frac{\pi_{iJ} \pi_{IJ}}{\pi_{iJ} \pi_{IJ}} = \frac{\pi_{ij}}{\pi_{IJ}}, \quad i = 1, \ldots, I - 1, \quad j = 1, \ldots, J - 1.
\] (2.3.4)

This odds ratio is determined by the cell on \(i\)th row and the \(j\)th column, and the cell on the last row and the last column. For instance, in Table 2.1, if \(i = 4\) and \(j = 2\), the odds ratio is equivalent to

\[
\hat{\theta}_{R(4,2)} = \frac{2 \times 2}{2 \times 1} = 2.
\]

For the odds ratios in equation (2.3.3) and (2.3.4), if given the marginal distributions \(\pi_{i+}\) and \(\pi_{+j}\) and the cell probabilities are non-negative, the \((I - 1)(J - 1)\) parameters can describe any association in an \(I \times J\) table. Independence is certified when all of \((I - 1)(J - 1)\) odds ratios equal to 1, see Agresti (2007) for details.

4. Mantel-Haenszel Odds Ratio

Here we introduce the general Mantel-Haenszel (MH) estimator of common odds ratio. The formulas especially for categorical data are what we are going to talk about in the next chapter.

**Definition 2.3.2.** Assume there are \(K\) strata and each one is a \(2 \times 2\) contingency table. The odds ratios are \(\theta_1, \theta_2, \ldots, \theta_K\) respectively. If there is an odds ratio such that

\[
\theta_C = \theta_1 = \theta_2 = \ldots = \theta_K,
\]

\(\theta_C\) is called a common odds ratio.
The Mantel-Haenszel estimator of the common odds ratio was put forward in 1959 and is still widely used nowadays. According to Higgins (2004), the MH estimator is obtained from the subsequent equations. To begin with, we define the quantities of numerator $A$ and denominator $B$ as

$$A = \sum_{r=1}^{K} \frac{N_{11r}N_{22r}}{N_r},$$

$$B = \sum_{r=1}^{K} \frac{N_{12r}N_{21r}}{N_r}.$$

$N_{11r}$ is the entry of the first row and the first column in the $r$th stratum; $N_r$ is the total of all the entries in the $r$th stratum.

The Mantel-Haenszel estimator of the common odds ratio is

$$\hat{\theta}_{MH} = \frac{A}{B}. \quad (2.3.5)$$

The logarithm of MH odds ratio $\log(\hat{\theta}_{MH})$ has an approximate normal distribution with mean $\log(\hat{\theta}_{MH})$ and variance which is given by

$$Var\left[\log(\hat{\theta}_{MH})\right] = \frac{\sum_{r=1}^{K} (N_{11r} + N_{22r})(N_{11r}N_{22r})/N_r^2}{2A^2} + \frac{\sum_{r=1}^{K} (N_{11r} + N_{22r})(N_{12r}N_{21r}) + (N_{12r} + N_{21r})(N_{11r}N_{22r})/N_r^2}{2AB} + \frac{\sum_{r=1}^{K} (N_{12r} + N_{21r})(N_{12r}N_{21r})/N_r^2}{2B^2}.$$

The normal approximation can be used to make a confidence interval for $\log(\hat{\theta}_{MH})$, see Higgins (2004) for details. For example, an approximate 95% confidence interval for $\log(\hat{\theta}_{MH})$ is

$$\left[\log(\hat{\theta}_{MH}) - 1.96 \sqrt{Var\left[\log(\hat{\theta}_{MH})\right]}, \log(\hat{\theta}_{MH}) + 1.96 \sqrt{Var\left[\log(\hat{\theta}_{MH})\right]}\right],$$
then it is exponentiated to obtain a confidence interval for $\theta_{MH}$, which is

$$\left[ \exp \left( \log(\hat{\theta}_{MH}) - 1.96 \sqrt{\text{Var} \left[ \log(\hat{\theta}_{MH}) \right]} \right), \exp \left( \log(\hat{\theta}_{MH}) + 1.96 \sqrt{\text{Var} \left[ \log(\hat{\theta}_{MH}) \right]} \right) \right].$$

5. The Hypotheses and the Criteria

The null and alternative hypotheses in terms of the odds ratio in general are

$$H_0 : \theta = 1;$$

$$H_a : \theta \neq 1.$$

The value of odds ratio is non-negative and ranges from 0 to infinity. Value 1.0 is a baseline for comparison. Due to the Definition 2.3.1, if the estimated odds ratio equals to 1.0, the odds that an observation falls in column 1 are the identical in row 1 and row 2. Otherwise, if the estimated odds ratio does not equal to 1.0, the odds where an observation is in column 1 are different in row 1 and row 2. This criteria applies for the odds ratios in $I \times J$ contingency tables as well.

For MH estimator of common odds ratio, the value of $\theta_{MH}$ falls between 0 and infinity and is fully efficient when the estimated value equal to 1. As it is indicated by Greenland (1989), the estimator $\theta_{MH}$ exceeds 90% efficiency relative to conditional maximum likelihood, and exceeds 95% when $0.2 \leq \theta_{MH} \leq 5.0$.

2.3.3 $M^2$ Statistic

Another statistic for testing the independence especially for ordinal data is considering the linear trend. A scoring system is usually used to detect the trend association. One of the benefits of using the scoring systems is that, the test statistic is sensitive to
positive or negative linear trend, as well as makes use of the correlation information in the data, see Agresti (2007) for details. The other benefit is that a scoring system specifies the ordering information and would be more powerful to analyze the ordinal variables.

For a contingency table comprising ordinal variables $X$ and $Y$, assign the scores for the rows $u_i$ and that for the columns $v_j$ satisfying the order

$$u_1 \leq u_2 \leq \cdots \leq u_I,$$

$$v_1 \leq v_2 \leq \cdots \leq v_J,$$

which is the same order as the categories. The correlation coefficient $r$ between variables $X$ and $Y$ equals to the sample covariance of them divided by the product of the sample standard deviations of $X$ and $Y$. That is,

$$r = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (u_i - \bar{u})(v_j - \bar{v})\hat{p}_{ij}}{\sqrt{[\sum_{i=1}^{I} (u_i - \bar{u})^2\hat{p}_{i+}][\sum_{j=1}^{J} (v_j - \bar{v})^2\hat{p}_{+j}]}},$$

(2.3.8)

In this equation, $\bar{u} = \sum_{i=1}^{I} u_i \hat{p}_{i+}$ denotes the sample mean of the row scores, $\bar{v} = \sum_{j=1}^{J} v_j \hat{p}_{+j}$ denotes the sample mean of the column scores, $\hat{p}_{i+}$ and $\hat{p}_{+j}$ are the marginal probabilities of rows and columns, and $\hat{p}_{ij} = n_{ij}/N$ is the sample probability on the $i$th row and the $j$th column. Therefore, the numerator is the covariance of variables $X$ and $Y$.

The value of correlation coefficient $r$ ranges from $-1$ to $+1$. If the variables are independent, $r$ equals $0$. On the contrary, the correlation coefficient $r$ that equals to $-1$ or $+1$ is the evidence that the variables have negative or positive association. The larger the correlation coefficient $r$ is in absolute value, the farther the data fall from independence in the linear dimension, see Agresti (2007) for details.
Against the hypotheses of a two-sided test

\[ H_0 : \text{Two ordinal variables are independent, or equivalently } r = 0; \]

\[ H_a : \text{Two ordinal variables are associated, or equivalently } r \neq 0. \]

The test statistic \( M^2 \) has components

\[ M^2 = r^2(N - 1) \sim \chi_1^2 \text{ for large sample,} \] \hspace{1cm} (2.3.9)

or equivalently in a square root form

\[ M = r\sqrt{N - 1} \sim N(0, 1) \text{ for large sample.} \] \hspace{1cm} (2.3.10)

\( M^2 \) is correlated to \( r^2 \) and the sample size \( N \). For a large sample, \( M^2 \) asymptotically follows chi-squared distribution with degree of freedom 1 under the assumption of independence. Whereas, when using the square root \( M \), \( M \) has an approximately standard normal distribution for large sample size.

Based on the theory from Agresti (2007), the \( M \) statistic applies for one-sided alternative hypothesis as well, i.e. \( H_a : r < 0. \)

### 2.4 Related Work

In this section, we will discuss several relatively new models for ordinal data. The existing modeling methods especially to deal with categorical responses are limited. So we also get an approach to those for longitudinal data. Because it is defined as the data with repeated observations of the same variables over a period of time where the variable types vary from binary, ordinal, nominal, etc. Thus in principle, if we omit the time vector, the model may apply for ordinal data. No matter how the form
changes, the subsequent models are all based on the general cumulative model that we introduced in equation (2.2.1).

### 2.4.1 Multivariate Probit Model

Varin and Czado (2010) proposed mixed ordered probit models with an autocorrelated component to capture subject-specific time series variability for the longitudinal data with binary and ordinal outcomes. At time point \( t \) for categorical data, if the distribution function of \( \epsilon \) that occurs in Definition 2.2.1 is normally distributed, say \( \epsilon \sim N(0, \sigma^2) \), such that the distribution function \( F \) equals to the cumulative probability function of a standard normal variable

\[
F(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt.
\]

Consequently, the cumulative model is resulted in a multivariate probit model with restrictions that the error \( \epsilon \) has unit variance, e.g. \( \sigma^2 = 1 \), and the first threshold \( \lambda_1 \) or, alternatively, the coefficient \( \gamma_1 \) is fixed to zero, see Agresti (2007) for details. Linking it to equation (2.2.1), the probit model can be represented by a cumulative probability that is given by

\[
P(Y \leq r \mid x) = F \left( \frac{\lambda_r + x^\prime \gamma}{\sqrt{\sigma^2 + 1}} \right),
\]

where the threshold \( \lambda = (\lambda_2, ..., \lambda_k) \), the coefficient \( \gamma \) and the variance component \( \sigma^2 \) are given.

Specially, when the subject \( Y \) experiences a certain level \( r \), as the variables follow ordinal scales, the probability can be written into the difference of two probabilities,
that is,
\[ P(Y = r \mid x) = P(U \in (\lambda_{r-1}, \lambda_r) \mid x) = F\left(\frac{\lambda_r + x'\gamma}{\sqrt{\sigma^2 + 1}}\right) - F\left(\frac{\lambda_{r-1} + x'\gamma}{\sqrt{\sigma^2 + 1}}\right). \]

### 2.4.2 Semiparametrically Structured Ordinal Model

Tutz (2003) proposed a semiparametrical model which is a multivariate generalized linear model that differs from the common ordinal models. As it is demonstrated in equation (2.2.1), the cumulative probability can be denoted by a generalized linear function \( F \) with the predicator \( \lambda_r + x'\gamma \) where the variable \( x \) does not depend on the category and both \( \lambda \) and \( \gamma \) are coefficients.

To construct a more structured model, assign the ordinal data be given by response \( Y \), explanatory variable \( x \) and additional covariate \( w = (w_1, ..., w_m)^T \). A semiparametrically structured ordinal model is obtained by the generalized linear function \( F \) with the predictor
\[ \eta = \lambda + x'\gamma + \sum_{j=1}^{m} \alpha_j(w_j), \]
where \( \alpha = (\alpha_1, ..., \alpha_m) \) are unspecified functions. Accordingly, the semiparametrically structured ordinal model has the following expression
\[ P(Y \leq r \mid x) = F(\eta) = F[\lambda + x'\gamma + \sum_{j=1}^{m} \alpha_j(w_j)]. \]

This model contains a partially generalized linear model \( F(\lambda + x'\gamma) \) and a term which retains the additive form yet remains unspecified \( F[\sum_{j=1}^{m} \alpha_j(w_j)] \). The method of estimating the parameters is based on a penalized expansion in basic function, which can be seen as an extension of penalized regression splines, see Tutz (2003) for details.
Since one of the aspects we are focusing on is testing the association between the ordinal variables where the test statistics are based on a two-way contingency table. Some unexpected problems may arise during calculations when it is an extremely sparse contingency table; see Dong and Simonoff (1995) for more information. A sparse contingency table is usually encountered in practice when the number of cells $I \times J$ is much larger compared to the total number of observations $N$, see Burman (1987) for details. The appearance of this type of table is that there are many cells that appear with entries 0 or 1. As a consequent, the test statistics should be reformed to avoid the error of calculations that may be caused by the sparse table.
3.1 Modified Pearson Goodness-of-fit Statistic

3.1.1 Modifications of Pearson Goodness-of-fit Statistic in the Cumulative Contingency Tables

We commence the modification from the definition of Pearson goodness-of-fit statistic.

1. Correction Term for Each Item of Pearson Chi-square Statistic

Definition 3.1.1. If $x$ is normally distributed with mean $\mu$ and standard deviation $\sigma$, then Pearson goodness-of-fit statistic $X^2$ which is defined by

$$X^2 = \left( \frac{x - \mu}{\sigma} \right)^2$$

follows chi-squared distribution with 1 degree of freedom under the assumption of independence. Note that, $X$ is the standardized statistic which follows the standard normal distribution $N(0,1)$.

The variance $\sigma^2$ is substituted by an approximately equivalent value which is the mean value $\hat{\mu}_{ij}$ in equation (2.3.1). However, if a more accurate form of the test statistic is required, one should use the real variance as it is in Definition 3.1.1. One can easily obtain the improved chi-square test statistic from the subsequent calculations.

Definition 3.1.2. Let $I, J$ be the number of rows and columns of a two-way contingency table with sample size $N$, the modified chi-square test statistic with the real
The variance is defined as

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{(1 - \hat{\pi}_{ij})\hat{\mu}_{ij}} \sim \chi^2_{(I-1)(J-1)}, \quad (3.1.1)$$

where $\pi_{ij} = p_i + p_j$, $\mu_{ij} = Np_i p_j$ and $\hat{\pi}_{ij}$ is the estimated value of $\pi_{ij}$.

We start from the definition of variance to calculate the denominator. That is,

$$\text{Var}(n_{ij}) = E(n_{ij}^2) - [E(n_{ij})]^2,$$

$$E(n_{ij}) = \mu_{ij} = Np_i p_j.$$

If we use indicator function, then the expected value of $n_{ij}^2$ is

$$E(n_{ij}^2) = E \left[ \left( \sum_{k=1}^{K} I\{X_k = i\}I\{Y_k = j\} \right) \left( \sum_{l=1}^{L} I\{X_l = i\}I\{Y_l = j\} \right) \right]$$

$$= E \left( \sum_{k=1}^{K} I\{X_k = i\}I\{Y_k = j\} \sum_{l \neq k} I\{X_l = i\}I\{Y_l = j\}I\{X_k = i\}I\{Y_k = j\} \right)$$

$$= Np_i p_j + N(N-1)p_i^2 p_j^2,$$

and the variance of $n_{ij}$ is

$$\text{Var}(n_{ij}) = \left[ Np_i p_j + N(N-1)p_i^2 p_j^2 \right] - N^2 p_i^2 p_j^2$$

$$= Np_i p_j - Np_i^2 p_j^2$$

$$= (1 - p_i p_j)Np_i p_j$$

$$= (1 - \pi_{ij})\mu_{ij}.$$
small for a sparse table because most of the entries are 0. So that $1 - p_{ij}$ term equals to 1 for most of the cells. That indeed is why we can use the mean value instead of the true variance in equation (2.3.1) to make the calculation simpler. In other words, in a non-cumulative table, $1 - p_{ij}$ term can be omitted.

2. Pearson Goodness-of-fit Statistic for the Cumulative Contingency Tables

When using the Pearson goodness-of-fit statistic to test the association in an $I \times J$ contingency table, the ordinal feature of the variables is not taken into consideration. To link with the cumulative models, we propose a cumulative contingency table which involves the information of ordinal data.

**Definition 3.1.3.** Given an $I \times J$ contingency table, one can construct a new table that the cells $m_{ij}$ are the cumulativeness of all the $n_{rs}$ on the $I \times J$ contingency table, such that

$$m_{ij} = \sum_{r=1}^{i} \sum_{s=1}^{j} n_{rs}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J.$$ 

The new table with cells $m_{ij}$ is called the cumulative contingency table.

**Remark 3.1.1.** In a cumulative contingency table, the entry $m_{IJ}$ equals the sample size of the corresponding non-cumulative contingency table.

For instance, assume that based on Table 3.1, the cumulative contingency table is given in Table 3.2. The cumulative entry $m_{22} = \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} = 1 + 2 + 2 + 0 = 5$. 

27
Table 3.1: A contingency table: example 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.2: The cumulative contingency table regarding Table 3.1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>10</td>
<td>12</td>
</tr>
</tbody>
</table>

For a cumulative contingency table, the Pearson chi-square statistic needs to be made the subsequential changes. To compare with equation (3.1.1), the non-cumulative entries \( n_{ij} \) are substituted by the cumulative entries \( m_{ij} \) and the probabilities are changed to the cumulative ones \( p'_{ij} = P(X \leq i, Y \leq j) \), such that

\[
X_C^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(m_{ij} - \hat{p}'_{ij})^2}{(1 - \hat{p}'_{ij}) \hat{p}'_{ij}} \sim \chi^2_{(I-1)(J-1)}, \tag{3.1.2}
\]

where \( m_{ij} = \sum_{k=1}^{i} \sum_{l=1}^{j} n_{ij} \), \( p'_{ij} = p'_{i+j}, \) \( p'_{ij} = N p'_{i+j} \), \( \hat{p}'_{ij} \) is the estimated value of \( p'_{ij} \) and the non-cumulative sample size \( N \) is remained. The cumulative marginal
probabilities are $p'_{i+} = P(X \leq i)$ and $p'_{+j} = P(Y \leq j)$. This test statistic is chi-squared distributed with $(I-1)(J-1)$ degree of freedom.

It is noteworthy that a cumulative table should not be used together with other methods that comprise ordinal information, for example, the scoring systems. Otherwise the ordinal information is considered iteratively.

3.1.2 Scoring Systems

The scoring system is another method to describe the property of ordinal. Ranks can be considered as scores that are used in place of the ordinal observations in a permutation test, see Higgins (2004) for details. Whereas there are several scoring systems which are commonly and effectively used other than ranking. For example, exponential scores, Van der Waerden scores, the combination of common scores, etc. In general, which one to choose depends on the distribution of the population from which the data are selected.
1. Exponential Scores

**Definition 3.1.4.** Let $I$ be the total number of categories, the expected values of the order statistics of the exponential distribution are given by

$$S_{e1} = \frac{1}{I},$$
$$S_{e2} = \frac{1}{I} + \frac{1}{I-1},$$
$$S_{e3} = \frac{1}{I} + \frac{1}{I-1} + \frac{1}{I-2},$$

... 

$$S_{eI} = \frac{1}{I} + \frac{1}{I-1} + \ldots + 1,$$

which are exponential scores.

**Remark 3.1.2.** Exponential scores sum to the total number of categories $I$. Or equivalently, if each of the above scores are subtracted by 1, one can obtain Savage scores which sum to 0.

2. Van der Waerden Scores

**Definition 3.1.5.** Let $\Phi^{-1}$ denotes the inverse of the cumulative distribution function of the standard normal distribution and $I$ is the total number of categories. The Van der Waerden scores are defined by

$$S_{VWi} = \Phi^{-1}\left(\frac{i}{I+1}\right), \quad i = 1, 2, \ldots, I.$$

**Remark 3.1.3.** Van der Waerden scores sum to 0.
As an example, when $I = 10$, then $S_{VW5} = \Phi^{-1}_{5/11} = Z_{0.4545} = -0.114$. Since the Van der Waerden (VW) scores are the inverse of the c.d.f. of the standard normal distribution, VW scores are considered as the approximation of the normal scores, see Higgins (2004) for details. In the following table, we can see the VW scores for the values of $I$ from 10 to 12.

Table 3.3: Van der Waerden scores

<table>
<thead>
<tr>
<th></th>
<th>N=10 Scores</th>
<th></th>
<th>N=11 Scores</th>
<th></th>
<th>N=12 Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.335</td>
<td>1</td>
<td>-1.383</td>
<td>1</td>
<td>-1.426</td>
</tr>
<tr>
<td>2</td>
<td>-0.908</td>
<td>2</td>
<td>-0.967</td>
<td>2</td>
<td>-1.020</td>
</tr>
<tr>
<td>3</td>
<td>-0.605</td>
<td>2</td>
<td>-0.674</td>
<td>3</td>
<td>-0.736</td>
</tr>
<tr>
<td>4</td>
<td>-0.349</td>
<td>4</td>
<td>-0.431</td>
<td>4</td>
<td>-0.502</td>
</tr>
<tr>
<td>5</td>
<td>-0.114</td>
<td>5</td>
<td>-0.210</td>
<td>5</td>
<td>-0.293</td>
</tr>
<tr>
<td>6</td>
<td>0.114</td>
<td>6</td>
<td>0.000</td>
<td>6</td>
<td>-0.097</td>
</tr>
<tr>
<td>7</td>
<td>0.349</td>
<td>7</td>
<td>0.210</td>
<td>7</td>
<td>0.097</td>
</tr>
<tr>
<td>8</td>
<td>0.605</td>
<td>8</td>
<td>0.431</td>
<td>8</td>
<td>0.293</td>
</tr>
<tr>
<td>9</td>
<td>0.908</td>
<td>9</td>
<td>0.674</td>
<td>9</td>
<td>0.502</td>
</tr>
<tr>
<td>10</td>
<td>1.335</td>
<td>10</td>
<td>0.967</td>
<td>10</td>
<td>0.736</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>1.383</td>
<td>11</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>1.426</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Joint Scores

Next we present one of the combination of the common scores, the “joint scores”.

**Definition 3.1.6.** Let $I$ be the total number of categories, the joint scores are defined by

\[
S_{j1} = \left( \frac{1}{I} \right) + \Phi^{-1}\left( \frac{1}{I+1} \right),
\]

\[
S_{j2} = \left( \frac{1}{I} + \frac{1}{I-1} \right) + \Phi^{-1}\left( \frac{2}{I+1} \right),
\]

...  

\[
S_{jI} = \left( \frac{1}{I} + \frac{1}{I-1} + ... + 1 \right) + \Phi^{-1}\left( \frac{I}{I+1} \right), \quad j = 1, 2, ..., I,
\]

which are the summations of exponential scores and VW scores for each category.

**Remark 3.1.4.** Because the terms of exponential scores sum to $I$ and the terms of VW scores sum to 0, thus the joint scores sum to $I$.

The joint scores for the values of $I$ from 10 to 12 are given in Table 3.4.
Table 3.4: Joint scores (the combination of exponential scores and Van der Waerdon scores)

<table>
<thead>
<tr>
<th></th>
<th>N=10 Scores</th>
<th></th>
<th>N=11 Scores</th>
<th></th>
<th>N=12 Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.235</td>
<td>1</td>
<td>1.292</td>
<td>1</td>
<td>1.343</td>
</tr>
<tr>
<td>2</td>
<td>-0.697</td>
<td>2</td>
<td>-0.776</td>
<td>2</td>
<td>-0.846</td>
</tr>
<tr>
<td>3</td>
<td>-0.269</td>
<td>3</td>
<td>-0.372</td>
<td>3</td>
<td>-0.462</td>
</tr>
<tr>
<td>4</td>
<td>0.130</td>
<td>4</td>
<td>-0.004</td>
<td>4</td>
<td>-0.117</td>
</tr>
<tr>
<td>5</td>
<td>0.532</td>
<td>5</td>
<td>0.360</td>
<td>5</td>
<td>0.217</td>
</tr>
<tr>
<td>6</td>
<td>0.960</td>
<td>6</td>
<td>0.737</td>
<td>6</td>
<td>0.556</td>
</tr>
<tr>
<td>7</td>
<td>1.445</td>
<td>7</td>
<td>1.147</td>
<td>7</td>
<td>0.917</td>
</tr>
<tr>
<td>8</td>
<td>2.034</td>
<td>8</td>
<td>1.618</td>
<td>8</td>
<td>1.313</td>
</tr>
<tr>
<td>9</td>
<td>2.937</td>
<td>9</td>
<td>2.194</td>
<td>9</td>
<td>1.772</td>
</tr>
<tr>
<td>10</td>
<td>4.264</td>
<td>10</td>
<td>2.987</td>
<td>10</td>
<td>2.339</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>4.403</td>
<td>11</td>
<td>3.123</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>4.529</td>
</tr>
</tbody>
</table>

4. Test Statistic with Scoring Systems

Definition 3.1.7. Take $1-p_{ij}$ terms as well as the scoring systems into consideration, a modified Pearson goodness-of-fit test statistic for the cumulative contingency tables
are defined by
\[
X^2_S = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(u_i v_j n_{ij} - u_i v_j \hat{\mu}_{ij})^2}{u_i v_j (1 - \hat{p}_{ij}) \hat{\mu}_{ij}}
\]
\[
= \sum_{i=1}^{I} \sum_{j=1}^{J} u_i v_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{(1 - \hat{p}_{ij}) \hat{\mu}_{ij}} \sim \chi^2_{(I-1)(J-1)},
\]
(3.1.3)
where \(u_i\) denote the scores of row categories and \(v_j\) denote the scores of column categories.

### 3.1.3 Variance Minimization Correction Term

Except for using the cumulative methods and the scoring systems, another consideration is appending a correction term. According to Farrington (1996), one can construct a first-order correction term to Pearson goodness-of-fit test statistic. One of the choices of the first-order correction term is variance minimization correction term. It is aiming to reduce the variance of the test statistic, so that the test statistic is more convergent to the true value, and the test is more powerful and persuasive.

**Definition 3.1.8.** Define the first-order correction term for a non-cumulative table as \(A\),
\[
A = \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{a}_{ij} (n_{ij} - \hat{\mu}_{ij}),
\]
so that for an \(I \times J\) ordinal contingency table, the adapted test statistic is
\[
X^2_A = A + \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{(1 - \hat{p}_{ij}) \hat{\mu}_{ij}} \sim \chi^2_{(I-1)(J-1)},
\]
(3.1.4)
where \(\hat{a}_{ij} = a_{ij} (\hat{\mu}_{ij})\) is a set of functions to be specified. Particularly, when the choice of \(a_{ij}\) is 0, \(X^2_A = X^2\) where \(X^2\) is the Pearson goodness-of-fit test statistic given in equation (2.3.1).
Theorem 3.1.1. If $a_{ij}$ in the first-order correction term $A$ has the following expression
\[ a_{ij} = -\frac{1 - 2p_{i+p+j}}{Np_{i+p+j}(1 - p_{i+p+j})}, \]
term $A$ minimizes the variance of the Pearson goodness-of-fit test statistic.

Proof. First we consider the situation for a non-cumulative contingency table. Now we proof when the variance of the test statistic is minimized, $a_{ij}$ has the expression that is given in Theorem 3.1.1. We know
\[ \text{Var}(n_{ij}) = Np_{ij}(1 - p_{ij}) = \mu_{ij}(1 - \frac{\mu_{ij}}{N}). \]
The first derivative of $\mu_{ij}$ equals
\[ \text{Var}'(n_{ij}) = \frac{d\text{Var}(n_{ij})}{d\mu_{ij}} = 1 - \frac{2\mu_{ij}}{N}. \]
In order to obtain the minimum value of variance, $a_{ij}$ need to satisfy
\[ a_{ij} = -\frac{\text{Var}'(n_{ij})}{\text{Var}(n_{ij})} = -\frac{1 - \frac{2\mu_{ij}}{N}}{\mu_{ij}(1 - \frac{\mu_{ij}}{N})} \]
\[ = -\frac{\text{Var}'(n_{ij})}{\text{Var}(n_{ij})} = -\frac{1 - \frac{2Np_{ij}}{N}}{Np_{ij}(1 - \frac{Np_{ij}}{N})} \]
\[ = \frac{2p_{ij} - 1}{Np_{ij}(1 - p_{ij})}, \]
see Farrington (1996) for details. Thus, the correction term $A$ for a non-cumulative table is
\[ A = \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{a}_{ij}(n_{ij} - \hat{\mu}_{ij}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{2\hat{p}_{ij} - 1}{N\hat{p}_{ij}(1 - \hat{p}_{ij})}(n_{ij} - N\hat{p}_{ij}), \]
where $p_{ij} = p_{i+p+j}$.
Following the similar steps, one can easily obtain the variance minimization correction term for the cumulative tables. That is,

\[ A' = \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{a}_{ij}' (m_{ij} - \hat{\mu}_{ij}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{2\hat{p}_{ij}' - 1}{N\hat{p}_{ij}'(1 - \hat{p}_{ij}')}(m_{ij} - N\hat{p}_{ij}') , \]

where \( p_{ij}' = p_{i+}p_{+j} \). Correspondingly, the chi-square test statistic with the correction term for cumulative table is given by

\[ X_{AC}^2 = A' + \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(m_{ij} - \hat{\mu}_{ij})^2}{(1 - \hat{p}_{ij}')\hat{\mu}_{ij}} \sim \chi^2_{(I-1)(J-1)}. \] (3.1.5)

The scoring systems and the variance minimization correction term can be considered at the same time for a non-cumulative table. There are several forms of the combination. The first form is that only the original chi-squared distribution term has scores, that is

\[ X_{A1}^2 = A + X_S^2 \]

\[ = \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{a}_{ij}(n_{ij} - \hat{\mu}_{ij}) + \sum_{i=1}^{I} \sum_{j=1}^{J} u_i v_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{(1 - \hat{p}_{ij}')\hat{\mu}_{ij}} \sim \chi^2_{(I-1)(J-1)}. \] (3.1.6)

The other form is that the chi-squared distribution term as well as the \( A \) correction term contain scores, that is

\[ X_{A2}^2 = A_S + X_S^2 \]

\[ = \sum_{i=1}^{I} \sum_{j=1}^{J} u_i v_j \hat{a}_{ij}(n_{ij} - \hat{\mu}_{ij}) + \sum_{i=1}^{I} \sum_{j=1}^{J} u_i v_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{(1 - \hat{p}_{ij}')\hat{\mu}_{ij}} \sim \chi^2_{(I-1)(J-1)}. \] (3.1.7)

As it is introduced in the previous subsection, exponential scores, VW scores and joint scores can be reckoned.
3.2 Modified Odds Ratio

To review the components of odds ratios that introduced in Section 2.3.2, we have given the general odds ratio in a 2 $\times$ 2 table, the individual odds ratios of $K$ pairs of independent binomial random variables, two distinct constructions of odds ratio in an $I \times J$ ordinal contingency table and MH estimator of the common odds ratio of $K$ 2 $\times$ 2 contingency tables. In this section, we compose a new methods to evaluate the common odds ratio in the $K \times K$ contingency tables where the category numbers of the variables are the same. After that, we introduce a method to deal with the extremely sparse tables.

3.2.1 Average of the Individual Odds Ratios

To begin with, in an $K \times K$ ordinal contingency table, the definition of individual odds ratio can be extended to that in the situation of $(K-1)^2$ pairs of independent binomial random variables. Let $\pi_{11r}$ denotes the joint probabilities where the observations from the first $r$ rows falls in the first $r$ columns. That is,

$$\pi_{11r} = P(X \leq r, Y \leq r)$$

where $r = 1, 2, ..., K - 1$. Next, let $N_{11r}$ denotes the total of entries where $X \leq r$ and $Y \leq r$. Similarly, $\pi_{12r}$ denotes the joint probabilities where $X \leq r$ and $r < Y \leq K$, $\pi_{21r}$ denotes the joint probabilities where $r < X \leq K$ and $Y \leq r$, $N_{22r}$ denotes the joint probabilities where $r < X \leq K$ and $r < Y \leq K$. Then the individual odds ratio in an $K \times K$ ordinal contingency table is defined by

$$\theta_{I(K)} = \frac{\pi_{11r}/\pi_{12r}}{\pi_{21r}/\pi_{22r}} = \frac{\pi_{11r}\pi_{22r}}{\pi_{12r}\pi_{21r}}. \quad (3.2.1)$$
The individual odds ratio of a sample equals
\[
\hat{\theta}_{I(K)} = \frac{p_{11r}p_{22r}}{p_{12r}p_{21r}} = \frac{N_{11r}N_{22r}}{N_{12r}N_{21r}}.
\] (3.2.2)

We consider the average of the individual odds ratios in one table to be an estimator of the common odds ratio of this table. The definitions of common odds ratio in the \(K \times K\) contingency tables are given here.

**Definition 3.2.1.** In a \(K \times K\) ordinal contingency table, if we only consider it when the category numbers of \(X\) and \(Y\) are equal, the estimator of common odds ratio is given by
\[
\hat{\theta}_{C(K)} = \frac{1}{K-1} \sum_{r=1}^{K-1} \frac{P(X \leq r, Y \leq r)/P(X \leq r, Y > r)}{P(X > r, Y \leq r)/P(X > r, Y > r)}
\]
\[
= \frac{1}{K-1} \sum_{r=1}^{K-1} \frac{N_{11r}/N_{12r}}{N_{21r}/N_{22r}}
\]
\[
= \frac{1}{K-1} \sum_{r=1}^{K-1} \frac{N_{11r}N_{22r}}{N_{12r}N_{21r}},
\] (3.2.3)

which is the average of the individual odds ratios of this table.

### 3.2.2 Mantel-Haenszel Odds Ratio in the \(K \times K\) Ordinal Contingency Tables

In the previous chapter, we have given the algorithm of Mantel-Haenszel estimator of common odds ratio in \(K\) strata of \(2 \times 2\) contingency tables. However, if we consider a \(K \times K\) contingency table to be \((K-1)(K-1)\) strata of \(2 \times 2\) tables, MH odds ratio may apply in this case.
Definition 3.2.2. In a $K \times K$ contingency table, if we only consider it when the number of the categories of $X$ and $Y$ are equivalent, the MH estimator of the common odds ratio of the table is

$$
\hat{\theta}_{MH(K)} = \frac{\sum_{r=1}^{K-1} N_{11r}N_{22r}/N_r}{\sum_{r=1}^{K-1} N_{12r}N_{21r}/N_r}. 
$$

(3.2.4)

The confidence interval for MH odds ratio is changed correspondingly.

3.2.3 Add-on Value for the Extremely Sparse Tables

The value of the estimator of common odds ratio is restricted from 0 to $\infty$. The value 1 of a common odds ratio indicates the independence of the ordinal variables. However, for a sparse table, there are still chances that the cumulative probabilities equal to 0 while the ordinal variables have an association. Thus we need to propose a modification method.

In a sparse contingency table, when the variable $r$ is small or approaching to the total number of category $K$, the cumulative probabilities usually equal to 0. For instance, in a $6 \times 6$ contingency table (Table 3.5), when $r = 2$, the cumulative probability $P(X \leq r, Y \leq r) = 0$. Consequently, the individual odds ratio equals 0. But it is not sufficient to indicate that two variables are not independent, because this result may be caused by the extremely sparse table. Similarly, when $r = 5$, the cumulative probability $P(X > r, Y > r) = 0$. So the individual odds ratio is infinity.
Table 3.5: Sparse contingency table: an example

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

To avoid the problems caused by sparse data, if there is any cumulative probability or equivalently $N$ term that equals zero, we slightly amend the individual odds ratio by adding a relatively small value of 0.5 on each $N$ term, see Agresti (2002) for details. Taking the average of individual odds ratios $\theta_{C(K)}$ as an example, it is changed to

$$\hat{\theta}_{C(K)} = \frac{1}{K - 1} \sum_{r=1}^{K-1} \frac{(N_{11r} + 0.5)(N_{22r} + 0.5)}{(N_{12r} + 0.5)(N_{21r} + 0.5)}. \quad (3.2.5)$$

Note that, we only do the modification for those if it is needed. That is, if there is no such cumulative probabilities resulting in the 0 or infinity value of the individual odds ratio, we do not use equation (3.2.5).

For MH estimator of common odds ratio $\theta_{MH(K)}$, we can add the relatively small value to avoid the zero probabilities as well. Such that,

$$\hat{\theta}_{MH(K)} = \frac{\sum_{r=1}^{K-1}(N_{11r} + 0.5)(N_{22r} + 0.5)/N_r}{\sum_{r=1}^{K-1}(N_{12r} + 0.5)(N_{21r} + 0.5)/N_r}. \quad (3.2.6)$$
The variance and the confidence interval change slightly corresponding to this change.

### 3.3 Modified $M^2$ Statistic

In the formula of $M^2$ statistic, the correlation coefficient $r$ between two variables contains the scores of rows and columns. Utilizing the scoring systems we mentioned in Section 3.1, the category numbers, exponential scores, VW scores and joint scores can be substituted into the function of correlation coefficient $r$ in equation (2.3.8). The equations of the correlation coefficient $r$ and $M_2$ statistic remain unchanged.
Chapter 4

Simulation

4.1 Data Generation

For the purpose of simulation, we generate the samples of ordinal random variables that meet the following conditions.

- Generating two ordinal variables and each variable consists of 11 categories;
- The variables are generated by uniform $(0, 1)$ distribution;
- Sample size is 33, 100, 200 respectively;
- Replications are 1000 times.

Here is the description of the conditions.

First, we generate two ordinal variables and each contains 11 categories, so that the contingency table constructed by these two variables is an $11 \times 11$ table. As we mentioned in the introduction, the reason for setting the total number of categories
to be 11 is that the number of variables of the medical data in the application is 11 at most. In order to verify if the test statistics can be used in the application for this medical data set, we use the same number of categories in the simulation.

Second, the simulated ordinal variables are randomly and independently generated. Let \((U_i, V_i)\) be the random sample that is generated from uniform \((0, 1)\) distribution, where \(i = 1, ..., N\), \(N\) is the sample size, and \(U_i\) is independent of \(V_i\). From each pair of \((U_i, V_i)\), we can get the category variables \((X_i, Y_i)\) where

\[
X_i = \begin{cases} 
1, & U_i \in [0, \frac{1}{11}] \\
2, & U_i \in (\frac{1}{11}, \frac{2}{11}] \\
& \vdots \\
11, & U_i \in (\frac{10}{11}, 1] 
\end{cases}
\]

\[
Y_i = \begin{cases} 
1, & V_i \in [0, \frac{1}{11}] \\
2, & V_i \in (\frac{1}{11}, \frac{2}{11}] \\
& \vdots \\
11, & V_i \in (\frac{10}{11}, 1] 
\end{cases}
\]

In this way, we obtain a sample that contains \(N\) observations. Next, we split the sample on a contingency table. For instance, if \(X_i = r\), \(Y_i = s\), we count once for the cell on row \(r\) and column \(s\).
Third, from the previous step, we obtain an $11 \times 11$ ordinal contingency table. In this table, we have a total 121 grids. Thus, when the sample size is 33, the table is extremely sparse. As it is mentioned in the introduction, in the medical data we have 33 observations at most. If the number of the observations is 100, the table is moderately sparse. When there are 200 observations, we consider the table lightly sparse.

Last but not least, the above steps for generating a data set are replicated 1000 times.

4.2 Data Processing

When the sample size is 33, the contingency table is extremely sparse. This property of a table may cause problems during the procedures of calculation. Therefore, we consider several possible methods to process the data for the purpose of making the table denser.

4.2.1 Blocks

Block analysis is useful in some particular situations. Unfortunately, it is not every contingency table that can be divided into blocks, since the blocks are in compliance with the condition that any two numbers from separate blocks are not in the same row or the same column. So block analysis does not work all the time.
4.2.2 The Compressed Contingency Tables

A compressed contingency table is obtained by canceling the rows and columns of the original table where all entries of them are zero. The category numbers in the compressed contingency table remain unchanged. It means, for example in Table 3.5, the entries of the first row and the second column are all zero. By canceling them we get a compressed table (Table 4.1), the compressed contingency table should have category numbers of rows and columns to be 2, 3, 4, 5, 6 and 1, 3, 4, 5, 6 respectively.

Table 4.1: A compressed table based on Table 3.4

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

4.2.3 Degree of Freedom

The cancelation results in the reduction of the degree of freedom. Since the zero rows and columns are canceled, the probabilities of them are not needed to be estimated. Then the total number of parameters to be estimated is reduced. To make it simple, the degree of freedom for chi-square statistic in Table 3.5 equals \((6 - 1)(6 - 1) = 25\). In comparison with that in the compressed table (Table 4.1), the degree of freedom
is reduced to \((5 - 1)(5 - 1) = 16\).

### 4.3 Hypotheses and the Criteria of Test Statistics

#### 4.3.1 Hypotheses

The hypotheses for simulation tests are

\[
H_0 : \text{Two ordinal variables are independent;}
\]

\[
H_a : \text{Two ordinal variables are associated.}
\]

#### 4.3.2 The Criteria of Test Statistics

Due to the truth that the simulated variables are randomly and independently generated from a uniform \((0, 1)\) distribution, whatever the test statistic is used, an appropriate test outcome should be the variables are not associated. That is, the null hypothesis is not rejected. Otherwise, the test statistic is considered improper or needs to be corrected.

### 4.4 Results of Simulation

In compliance with the above notations and rules of generating the simulated data, the simulation results are given below.
Table 4.2: The simulation results of Chi-square and $M^2$ statistics in the non-cumulative tables

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>33</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>$X^2$</td>
<td></td>
<td>102</td>
<td>86</td>
<td>99</td>
</tr>
<tr>
<td>$X^2$</td>
<td>with exp scores</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>$X^2$</td>
<td>with VW scores</td>
<td>254</td>
<td>180</td>
<td>180</td>
</tr>
<tr>
<td>$X^2$</td>
<td>with joint scores</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>$X^2 + A$</td>
<td></td>
<td>52</td>
<td>86</td>
<td>99</td>
</tr>
<tr>
<td>$X^2$</td>
<td>(exp)+A</td>
<td>1000</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>$X^2$</td>
<td>(VW)+A</td>
<td>254</td>
<td>181</td>
<td>180</td>
</tr>
<tr>
<td>$X^2$</td>
<td>(joint)+A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X^2$</td>
<td>(exp)+A(exp)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X^2$</td>
<td>(VW)+A(VW)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$X^2$</td>
<td>(joint)+A(joint)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M^2$</td>
<td>with category numbers</td>
<td>103</td>
<td>101</td>
<td>99</td>
</tr>
<tr>
<td>$M^2$</td>
<td>(exp)</td>
<td>100</td>
<td>103</td>
<td>97</td>
</tr>
<tr>
<td>$M^2$</td>
<td>(VW)</td>
<td>103</td>
<td>96</td>
<td>101</td>
</tr>
<tr>
<td>$M^2$</td>
<td>(joint)</td>
<td>98</td>
<td>100</td>
<td>97</td>
</tr>
</tbody>
</table>

Note: This table shows the number of rejecting $H_0$ in 1000 replications.
Table 4.2 shows the frequency of the null hypothesis being rejected out of 1000 replications under different significant level 0.10, 0.05 and 0.01. Based on the truth that two variables are selected independently, if the test statistics closely follow the chi-squared distribution, the times of rejecting $H_0$ are around 100, 50 and 10 out of 1000 replications under the significant levels 0.10, 0.05 and 0.01. When testing by the statistic $X^2$ without any modification, $H_0$ is rejected for about 100, 50 and 10 times respectively regardless the sample size which agrees with the truth. However, when testing ordinal variables, chi-square test statistic does not comprise the information of ordinal data. Strictly speaking, even though the simulation results are ideal, the test statistic is still not proper for ordinal variables.

When combining the scoring systems together with chi-square statistic, exponential scores and joint scores both result in 1000 rejections. The same results appear when the test statistics are $X^2 + A$ with exponential or joint scores. So neither exponential scores nor joint scores are appropriate to be used together with the chi-square statistic. The reason is that the difference of exponential scores for distinct categories is large. For a total of 11 categories, the score for the $1st$ category is 0.0909, while the score for the $11th$ category is 3.0199. The difference between joint scores is even expanded.

Next, we move to the chi-square statistics with VW scores. When testing by $X^2$, $X^2(VW) + A$ and $X^2(VW) + A(VW)$ statistics, under all tested sample sizes, $H_0$ are over-rejected. The rejection frequency is much more than 100, 50 and 10 respectively. It indicates that VW scores have great impact on the chi-square statistic, but not as much as the exponential scores and the joint scores.
It is noteworthy that the test statistic $X^2 + A$ without scores demonstrates a satisfactory result for a larger sample size. When the sample size is 100, $H_0$ is rejected 86, 42 and 9 times; while the sample size is 200, $H_0$ is rejected 99, 48 and 12 times under different significant levels. The test deviation is reasonable and acceptable. However, this test statistic does not take ordinal information into consideration.

The last part of Table 4.2 displays the test results of $M^2$ statistics. The $M^2$ statistic consists of scores in its expression while the scores convey the information of ordinal data. Meanwhile, the rejection frequencies are all around the critical value within a rational range. The results are not affected by the sample size, even for the extremely sparse table, the $M^2$ statistic performs well. The scoring systems are proper to be used together with the $M^2$ statistic. To compare with the other statistics, the $M^2$ statistic is more appropriate for the ordinal contingency tables.

On the other hand, when using the cumulative tables, the other methods that also comprise ordinal information are not utilized. The simulation results are given in Table 4.3.
Table 4.3: Simulation results in the cumulative tables

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.10</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>33</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>$X^2$</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$X^2 + A$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: This table shows the number of rejecting $H_0$ in 1000 replications.

It is evident to see that chi-square test statistic either with or without first-order correction term $A$ is not appropriate to test the ordinal data when using a cumulative table. The rejection frequencies is much different to the critical ones, no matter for small or large sample. The numbers of rejecting the null hypothesis are very close to 0, which demonstrates that the test statistics are smoothing in a cumulative contingency table.

The next table shows the simulation results of the estimators of common odds ratio. Odds ratio is defined regarding non-cumulative tables, thus the simulation only applies in the non-cumulative tables.
Table 4.4: Simulation results of the estimators of common odds ratio in the non-cumulative tables

<table>
<thead>
<tr>
<th>Test statistic</th>
<th>N=33</th>
<th>N=100</th>
<th>N=200</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\theta}_{C(11)}$</td>
<td>2.5617</td>
<td>1.2620</td>
<td>1.0919</td>
</tr>
<tr>
<td>$\hat{\theta}_{MH(11)}$</td>
<td>1.3440</td>
<td>1.0904</td>
<td>1.0290</td>
</tr>
</tbody>
</table>

Note: $\hat{\theta}_{C(11)}$ and $\hat{\theta}_{MN(11)}$ denote the average of 1000 replications of $\theta_{C(K)}$ and $\theta_{MH(K)}$ in $11 \times 11$ tables.

As it is noted under Table 4.4, $\theta_{C(K)}$ and $\theta_{MH(K)}$ are the average of individual odds ratios and MH estimator of common odds ratio introduced in Definition 3.2.1 and Definition 3.2.2 respectively.

According to the truth that two ordinal variables in the table are independent, both of the estimators of common odds ratio are around the critical value 1.0. But the values of the estimators are larger when the tables are extremely sparse. So the estimators of common odds ratio are better to be used for the lightly sparse tables.

In the meantime, it is apparent that no matter the table is sparse or not, MH odds ratio is closer to 1. When the sample size is 33, the ordinal contingency table is extremely sparse, and the difference between two odds ratios is $\theta_{C(11)} - \theta_{MH(11)} = 2.5617 - 1.3440 = 1.2177$. But for a large sample, $N = 200$, the difference is $\theta_{C(11)} - \theta_{MH(11)} = 1.0919 - 1.0290 = 0.0629$. Thus, when the ordinal contingency table is extremely sparse, MH estimator of common odds ratio demonstrates the as-
association between the ordinal variables better than the average of individual odds ratios. Whereas, when the table becomes lightly sparse, there is barely small difference between the average of individual odds ratios and MH estimator of common odds ratio.

4.5 Summary of Simulation

To sum up this chapter, according to the simulation results, the $M^2$ statistic works properly for either an extremely sparse table or a lightly sparse table. It detects the association between two ordinal variables in a contingency table. The category numbers, exponential scores, Van der Waerdon scores, and the joint scores can be used together with the $M^2$ test statistic in order to represent the ordinal information. Besides, the average of individual odds ratios and the Mantel-Haenszel estimator of common odds ratio are appropriate for sparse ordinal data as well, but some conditions may apply. For the extremely sparse contingency tables, the test error of the average of individual odds ratios is larger than that of MH estimator of odds ratio. However, for the lightly sparse tables, the difference of the test results between them is very little.
5.1 Data Sources and Features

5.1.1 Data Sources

As it is mentioned in the introduction, the medical data is from a project of medical related studies. It involves the patients’ pain scale data, history of certain disease treatments, chemotherapy and medication treatments on multiple time points, concomitant medication record, neurometer treatment, and individual’s demographics, etc.

The types of variables in the medical data are ordinal, nominal, string, interval, continuous and ratio. The associations of the next three pairs of ordinal variables are what we concern about. The first pair (data 1) is about neuropathic pain sensations, which are deep pain sensations and surface pain sensations after medical treatments. The second pair of variables (data 2) is about the modes of vibration detection to
exam the peripheral neuropathy. The third one (data 3) involves the disease stage and pain intensity after medical treatments.

Each data contains 5 parts that are collected at 5 different time points. The first time point is the baseline time without any treatment. The other time points are study visit times when the patients have had chemotherapy or medication treatments. Note that, for data 3, we only consider it at baseline time, because the disease stage as a part of the health information of the patients is only evaluated once at the beginning of this project.

5.1.2 Features

After sorting out the data to contingency tables (Table 5.1 - Table 5.5), several features of the data appear.
Table 5.1: Deep pain sensations (X) vs. surface pain sensations (Y) at baseline time

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td></td>
</tr>
</tbody>
</table>

Note: The empty grids are filled with “0”.
Table 5.2: Deep pain sensations (X) vs. surface pain sensations (Y) at study visit time 1

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>29</td>
</tr>
</tbody>
</table>

Note: The empty grids are filled with “0”. 
Table 5.3: Vibration detection mode 1 (X) vs. vibration detection mode 2 (Y) at baseline time

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>3</td>
<td>33</td>
</tr>
</tbody>
</table>

Note: The empty grids are filled with “0”.
Table 5.4: Vibration detection mode 1 (X) vs. vibration detection mode 2 (Y) at study visit time 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>1</td>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>6</td>
<td>29</td>
</tr>
</tbody>
</table>

Note: The empty grids are filled with “0”.
Table 5.5: Disease stage (X) vs. pain intensity (Y) at baseline time

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
</tbody>
</table>

Note: The empty grids are filled with “0”.

To begin with, all of the variables from the data are ordinal. For example, the pain scale data cannot be described in numerical ways yet the difference between adjacent categories does exist.

Secondly, at the baseline time, the first data is written in an $11 \times 11$ contingency table with 33 observations, the second data is in a $9 \times 9$ contingency table with 33 observations, and the third one is in a $4 \times 9$ contingency table with only 32 observations. Since the numbers of observations are very small in compare with the total grids on the table, the tables of the data are extremely sparse.

Another property is that the sample size on each time point is not the same for the reason that the patients drop out of the project over time. For example in data 1, there are 33 patients at baseline time while when it comes to study visit time 1, 2, 3 and 4, there are only 29, 26, 19 and 17 left. As we know, when the sample size is
extremely small, one has a bigger chance to have error during testing and modeling. In order to avoid this kind of error, we only test on the data at baseline time and study visit time 1.

Besides, the category numbers of the variables are from 0 to 10 in the original medical data. But using category number 0 can easily cause problems during calculations. So we change the category numbers to 1, 2, 3,..., 11.

5.2 Test Results of the Medical Data

As it is discussed in the previous chapter, the $M^2$ test statistic is efficient to test the association of ordinal variables while the estimators of common odds ratio can be used as well if some conditions are fulfilled.

After processing the data sets by the methods introduced in Section 4.2 and testing by the appropriate statistics, we obtain the test results in Table 5.6 to Table 5.8.
Table 5.6: Test results of deep pain sensations vs. surface pain sensations

<table>
<thead>
<tr>
<th>Test statistics</th>
<th>Baseline (N=33)</th>
<th>Study visit time 1 (N=28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^2$</td>
<td>17.3637 (R)</td>
<td>12.8443 (R)</td>
</tr>
<tr>
<td>$M^2$ (exp)</td>
<td>20.1286 (R)</td>
<td>7.1967 (R)</td>
</tr>
<tr>
<td>$M^2$ (VW)</td>
<td>16.1983 (R)</td>
<td>14.0339 (R)</td>
</tr>
<tr>
<td>$M^2$ (joint)</td>
<td>19.5947 (R)</td>
<td>8.1429 (R)</td>
</tr>
<tr>
<td>$\theta_{C(11)}$</td>
<td>72.6229</td>
<td>23.9346</td>
</tr>
<tr>
<td>$\theta_{MH(11)}$</td>
<td>22.9821</td>
<td>16.2838</td>
</tr>
</tbody>
</table>

Note: 1. “R” in brackets denotes rejecting null hypothesis; 2. The critical values are $\chi^2_{(1),0.10} = 2.7055$, $\chi^2_{(1),0.05} = 3.8415$ and $\chi^2_{(1),0.01} = 6.6349$. 

61
Table 5.7: Test results of two modes of vibration detection

<table>
<thead>
<tr>
<th>Test statistics</th>
<th>Baseline (N=33)</th>
<th>Study visit time 1 (N=29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^2$</td>
<td>29.0787 (R)</td>
<td>26.0913 (R)</td>
</tr>
<tr>
<td>$M^2$ (exp)</td>
<td>24.2081 (R)</td>
<td>25.0630 (R)</td>
</tr>
<tr>
<td>$M^2$ (VW)</td>
<td>28.8039 (R)</td>
<td>26.1709 (R)</td>
</tr>
<tr>
<td>$M^2$ (joint)</td>
<td>26.4928 (R)</td>
<td>25.5920 (R)</td>
</tr>
<tr>
<td>$\theta_{C(11)}$</td>
<td>257.7467</td>
<td>212.9733</td>
</tr>
<tr>
<td>$\theta_{MH(11)}$</td>
<td>74.3529</td>
<td>64.1250</td>
</tr>
</tbody>
</table>

Note: 1. “R” in brackets denotes rejecting null hypothesis; 2. The critical values are $\chi^2_{(1),0.10} = 2.7055$, $\chi^2_{(1),0.05} = 3.8415$ and $\chi^2_{(1),0.01} = 6.6349$.  

62
Table 5.8: Test results of disease stage vs. pain intensity

<table>
<thead>
<tr>
<th>Test statistics</th>
<th>Baseline (N=32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M^2$</td>
<td>1.8663 (FR)</td>
</tr>
<tr>
<td>$M^2$ (exp)</td>
<td>1.6641 (FR)</td>
</tr>
<tr>
<td>$M^2$ (VW)</td>
<td>0.6572 (FR)</td>
</tr>
<tr>
<td>$M^2$ (joint)</td>
<td>1.3323 (FR)</td>
</tr>
</tbody>
</table>

Note: 1. “FR” in brackets denotes that it is fail to reject null hypothesis; 2. The critical values are $\chi^2(1,0.10) = 2.7055$, $\chi^2(1,0.05) = 3.8415$ and $\chi^2(1,0.01) = 6.6349$.

From Table 5.6, we can conclude that the deep pain sensations and the surface pain sensations have certain types of associations at baseline time as well as study visit time 1. The values of $M^2$ statistics are positive and larger than the critical values under either 90% or 95% or 99% of confidence. It indicates that two variables have a positive correlation. The two estimators of common odds ratio have relatively large values in compare with the critical value 1, so they also support that two variables are associated. Since all the values of $M^2$ statistics and the common odds ratios are reduced at study visit time 1, the association of the variables is not as strong as it is at baseline time. If we look back in Table 5.1, we can figure out a trend of the distribution from top left to bottom right while most of the observations concentrate in grid $n_{11}$. But if we compare the distribution at baseline time to that at study visit time 1, it is easy to find that a little bit more data are distributed in the middle and
bottom part of Table 5.2. And it can be considered as a reason of the reducing. In practice, after the medical treatments, most of the patients have a strong feeling of discomfort and pain while different kinds of pain come at the same time. However, some types of the pain weaken over time and it is easier to tell which type of pain that still exists.

The test results of two modes of vibration detection are displayed in Table 5.7. We can draw a conclusion that two modes of vibration detection are correlated. Moreover, the association between them is extremely strong which we can conclude from the large value of the odds ratios. In Table 5.3, there is a clearly linear trend along the diagonal which assists the conclusion as well. Since the two modes of vibration detection are actually used to detect the same type of peripheral neuropathy, even though the modes are distinct, there should not be a huge difference between the outcomes of the detection.

We only list the test results of $M^2$ statistics at baseline time in Table 5.8. As it is mentioned when we introduce the data sources, the disease stage as a part of the health information of the patients is only evaluated once at baseline time. Besides, since the variables constitute a $4 \times 11$ table as it is shown in Table 5.5, yet we only consider the odds ratio in the situation that the ordinal contingency table has a $K \times K$ format, so that the values of the estimators of common odds ratio are not given in Table 5.8. All of the values of $M^2$ statistics are smaller than the critical values no matter under any level of confidence. The test results indicate that there is no apparent association between disease stage of a patient and the pain intensity. At the meantime, the distribution of the data in Table 5.5 agrees with the test results.
Looking back to Table 5.5, the data are centered on the left side of the table without any clear trend of the distribution. To sum up, the disease stage does not associate with pain intensity. No matter what stage of disease the patients are in, they feel a high level of pain after having the medical treatments.
Chapter 6

Future Work and Conclusion

6.1 Future Work

The possible future work involves analyzing the longitudinal data with ordinal outcomes; proposing other scoring systems for chi-square statistic, e.g. non-uniform scores; ascertaining the estimators of common odds ratio for the $I \times J$ ordinal contingency tables.

6.1.1 Analyzing the Longitudinal Data

Longitudinal data frequently appear in medical and biomedical studies as often as the ordinal data. Longitudinal data involves repeated observations of the same variables over a certain period of time. The repeated variables are often binary, nominal or ordinal. The methodologies we proposed in the previous chapters may also be used if different time is considered as an explanatory variable. For example, the general
cumulative model in equation (2.2.1) can be changed to

\[ P(Y_t \leq r_t \mid x_t) = F(\lambda_{rt} + x_t'\gamma_t). \]

Another example is that, the MH estimator of common odds ratio can be reputed for \( K \) strata of \( K \) time points.

Since the medical data we introduced in this thesis contains several time periods and the patients drop out of the project over time, one may propose the methods for analyzing the longitudinal data to analyze the medical data if the problem of missing data can be solved in the first place.

### 6.1.2 Non-uniform Scores

When testing with Pearson goodness-of-fit statistic, except the exponential scores, Van der Waerdon scores and the combination of them, other families of scores such as normal scores or non-uniform scores may apply for the sparse ordinal contingency tables. Explicitly, when there is a significant trend shown on the contingency table like the trend in the medical data, the categories that contain more observations may give more weight. For instance in Table 5.1, there are 20 out of 33 observations that fall in the first category of deep pain sensations (\( X \)) while they are under the first category of surface pain sensations (\( Y \)). Thus, we may give more weight to the first categories of variable \( X \) and \( Y \). As an example, the scores for 11 categories of \( X \) and \( Y \) could be 5, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1. As for different tables, the other combinations of the scores may also apply.
6.1.3 The Estimators of Common Odds Ratio in the $I \times J$ Ordinal Contingency Tables

In the chapters of simulation and application, we only detected the association by the average of individual odds ratios and MH estimator of common odds ratio in a $K \times K$ ordinal contingency table when the category numbers of $X$ and $Y$ are equivalent, for the purpose of the facilitation of calculations. However, in a more general case, such as in an $I \times J$ table, one may use the following equations to estimate the common odds ratio of the table.

In an $I \times J$ ordinal contingency table, the average of the individual odds ratios

$$\hat{\theta}_{C(I,J)} = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \frac{P(X \leq i, Y \leq j)/P(X \leq i, Y > j)}{P(X > i, Y \leq j)/P(X > i, Y > j)}$$

$$= \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \frac{N_{11(i,j)}/N_{12(i,j)}}{N_{21(i,j)}/N_{22(i,j)}}$$

$$= \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \frac{N_{11(i,j)}N_{22(i,j)}}{N_{12(i,j)}N_{21(i,j)}}$$

is an estimator of the common odds ratio of this $I \times J$ table. Particularly, when the number of categories of $X$ and $Y$ are equivalent, say it is a $K \times K$ table, the estimator of common odds ratio is

$$\hat{\theta}_{C(K,K)} = \frac{1}{(K-1)(K-1)} \sum_{i=1}^{K-1} \sum_{j=1}^{K-1} \frac{N_{11(i,j)}N_{22(i,j)}}{N_{12(i,j)}N_{21(i,j)}}$$

Moreover, since an $I \times J$ contingency table contains $(I-1)(J-1)$ strata of $2 \times 2$ tables, the MH estimator of common odds ratio of this $I \times J$ table is given by

$$\hat{\theta}_{MH(I,J)} = \frac{\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} N_{11(i,j)}N_{22(i,j)}/N_{(i,j)}}{\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} N_{12(i,j)}N_{21(i,j)}/N_{(i,j)}}$$
Specifically, when the number of categories of $X$ and $Y$ are the same, the MH estimator of common odds ratio is

$$\hat{\theta}_{MH(K,K)} = \frac{\sum_{i=1}^{K-1} \sum_{j=1}^{K-1} N_{11(i,j)} N_{22(i,j)}}{\sum_{i=1}^{K-1} \sum_{j=1}^{K-1} N_{12(i,j)} N_{21(i,j)}} / N(i,j).$$

The confidence interval for MH odds ratio is changed correspondingly.

### 6.2 Conclusion

The real data sets in various studies such as medical studies may be ordinal and the variables may consist of numerous categories. The contingency tables constructed by this type of ordinal variables are extremely sparse if the number of observations is relatively small.

For Pearson goodness-of-fit statistic, we provided the correction term for each item, the cumulative probabilities and entries, the scoring systems and the variance minimization correction term. When using the odds ratio to detect the association, we proposed the average of individual odds ratios as the common odds ratio and the Mantel-Haenszel estimator of common odds ratio for the $K \times K$ ordinal contingency tables when the category numbers of the variables are equivalent, as well as the add-on value for the extremely sparse tables. The $M^2$ statistic with the scoring systems is another option for analyzing the ordinal data.

From the results of simulation, we find that for the extremely sparse ordinal contingency tables, the $M^2$ statistic which involves the correlation coefficient as well as the estimators of common odds ratio can be used to detect the association between the ordinal variables. But in compare with that in the lightly sparse tables, there is a larger deviance when using the estimators of common odds ratio in the extremely
sparse tables. The MH estimators of common odds ratio is more appropriate than the average of individual odds ratios when the table is extremely sparse. Moreover, for an extremely sparse table, we can add a relatively small value on each term of the odds ratio to avoid the problems that may occur during calculations. On the other hand, the correlation coefficient in the $M^2$ statistic comprises scores. We can substitute the category numbers, exponential scores, Van der Waerden scores and the combination of exponential and VW scores in the $M^2$ statistic.
Bibliography


