STANDARD AND NON-STANDARD LOG-LINEAR MODELS

FOR 2×2 CONTINGENCY TABLES

By

G M Toufiqul Hoque

Bachelor of Science - Applied Statistics
University of Dhaka
2015

Master of Science - Mathematics
Lamar University
2021

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science - Mathematical Sciences

Department of Mathematical Sciences
College of Sciences
The Graduate College

University of Nevada, Las Vegas
May 2024

UNLV | GRADUATE COLLEGE

Thesis Approval

The Graduate College
The University of Nevada, Las Vegas

April 5, 2024

This thesis prepared by

G M Toufiqul Hoque

entitled

Standard and Non-Standard Log-Linear Models For 2 × 2 Contingency Tables

is approved in partial fulfillment of the requirements for the degree of

Master of Science - Mathematical Sciences
Department of Mathematical Sciences

Petros Hadjicostas, Ph.D.
Examination Committee Chair

Hokwon Cho, Ph.D.
Examination Committee Member

Dieudonne Phanord, Ph.D.
Examination Committee Member

Ashok Singh, Ph.D.
Graduate College Faculty Representative

Alyssa Crittenden, Ph.D.
Vice Provost for Graduate Education & Dean of the Graduate College

ABSTRACT

Standard and non-standard log-linear models
for 2×2 contingency tables
by
G M Toufiqul Hoque

Examination Committee Chair: Dr. Petros Hadjicostas

Log-linear models can be used to model the joint relationship of two or more categorical variables in a multiway contingency table. In a log-linear model, the logarithm of the expected joint counts (or the logarithm of the joint probabilities) in a contingency table can be written as a linear model.

Most log-linear models used in practice are standard. Standard log-linear models include the traditional parameter terms we see in ANOVA models: an overall effect, main effects, and various kinds of interaction terms.

Standard log-linear models are divided into hierarchical and non-hierarchical. Hierarchical models satisfy the hierarchy principle: if a higher-order term is included in the log-linear model, then so are all the lower-order terms.

Most of the standard log-linear models used in practice are hierarchical models. Non-hierarchical standard models appear rarely in the literature because they are difficult to interpret.

In this thesis, we examine standard and non-standard log-linear models for 2×2 contingency tables. To show their application, we use a thromboembolism data set that first appeared in Vessey and Doll (1968) and was later analyzed by Worcester (1971) using a multiplicative model, which can be equivalently written as a non-standard log-linear model.

Although the above data were collected in a one-to-two matching design, Worcester (1971) analyzed them using multinomial sampling where only the total was fixed.

In this thesis, however, we also examine a product multinomial sampling design for these data, which is a more correct probability model for a matched design.

We use the free statistical software **R** to estimate the above log-linear models. We compare the estimated log-linear models using the Pearson chi-square test, the G-square test, and the AIC, and we discuss the results.

ACKNOWLEDGEMENTS

I would like to express my thanks to Dr. Petros Hadjicostas for his major role in my thesis work. Dr. Petros Hadjicostas provided me with every bit of direction, guidance, assistance, and expertise that I required over the last year.

I am very glad and fortunate to have Dr. Hokwon Cho, Dr. Dieudonné Phanord, and Dr. Ashok Singh on my committee.

I am grateful to the faculty, staff, and colleagues in the Department of Mathematical Sciences at UNLV for their support. I want to convey a special gratitude to Dr. Nahid Hasan for his support and time.

Finally, I would like to pay special regards to my parents, wife, and child.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Contingency tables cross-classify two or more nominal or ordinal categorical variables. Observed contingency tables usually contain frequencies (counts) or percentages, which can be conditional or unconditional.

Log-linear models can be used to model the joint relationship of two or more categorical variables in a multiway contingency table. In a log-linear model, the logarithm of the expected joint counts (or the logarithm of the joint probabilities) in a contingency table can be written as a linear model using ANOVA-type notation. See, for example, Agresti [1] and Bishop *et al.* [2].

Most log-linear models used in practice are *standard*. Standard log-linear models include the traditional parameter terms we see in ANOVA models: an overall effect, main effects, and various kinds of interaction terms. Because the number of such parameters is usually larger than the number of expected counts, various kinds of restrictions on these parameters are usually imposed.

Standard log-linear models are divided into *hierarchical* and *non-hierarchical*. Hierarchical models satisfy the *hierarchy principle*: if a higher-order term is included in the log-linear model, then so are all the lower-order terms. (Many times, but not always, the hierarchy principle is followed in traditional ANOVA models.)

Most of the standard log-linear models used in practice are hierarchical models. Non-hierarchical standard models appear rarely in the literature because they are difficult to interpret.

Non-standard log-linear models are also rare in practice, but they sometimes appear in relation to *synergy*. For example, in his recent Ph.D. dissertation, Hasan [10] used non-standard log-linear models to study synergy of two categorical variables in affecting a third one. We do not examine synergy in this thesis. (For other examples of the use of non-standard log-linear models, see von Eye and Mun [13] and von Eye *et al.* [14].)

In this thesis, we examine standard and non-standard log-linear models for 2×2 contingency tables. To show their application, we use a thromboembolism data set that first appeared in Vessey and Doll [11] and was later analyzed by Worcester [15] using a multiplicative model, which can be equivalently written as a non-standard log-linear model.

The aforementioned thromboembolism data set was collected from a hospital index when in May 1967 the Medical Research Council of the Statistical Research Unit of the University College Medical School London looked into the risks of thromboembolic disease in women taking oral contraceptives. The researchers examined women who were admitted during the period 1964–1966 with the age range between 16–40 and having been diagnosed with some kind of thromboembolism.

Although the above data were collected in a one-to-two matching design (one woman with thromboembolism was matched to two women with no thromboembolism), Worcester [15] analyzed them using multinomial sampling where only the total (in this case, 116 women) was fixed.

In this thesis, however, we also examine a product multinomial sampling design for these data, which is a more correct probability model for a matched design (where the number of women with thromboembolism and the number of women with no thromboembolism are fixed).

We review the theory for the standard hierarchical log-linear models for the above data (that are cross-classified in a 2×2 contingency table). We also review the theory for the standard non-hierarchical log-linear models that include the overall effect (i.e., the intercept). Finally, we review the theory of the log-linear version of Worcester's [15] multiplicative model mentioned above. This model was also examined by Hasan [10], and it is non-standard.

We use the free statistical software **R** to estimate the parameters of the above log-linear models (both standard and non-standard). We compare the estimated log-linear models and discuss the results.

In more detail, the organization of the thesis is as follows. In Chapter 2, we describe various thromboembolism data sets that appeared in Vessey and Doll [11], which were collected in 1967. Only one of the data sets is analyzed in this thesis (see Table 2.1). Some of the other data sets in Chapter 2 were analyzed by Hasan [10] in his Ph.D. dissertation, Funo [5, 6], and Worcester [15].

In Section 3.1, we define the multinomial and product multinomial sampling schemes for a 2×2 contingency table. In Section 3.2, we give Worcester's multiplicative parametrization of a 2×2 contingency table that she used to analyze the aforementioned thromboembolism

data set in Table 2.1.

In Sections 3.3 and 3.4, we discuss various standard and non-standard log-linear models to analyze the data in Table 2.1 of Chapter 2. (The non-standard log-linear model we consider here is the logarithmic version of Worcester's multiplicative parametrization of a 2×2 contingency table that we examine in Section 3.2.) In particular, in Section 3.3 we use multinomial sampling for a 2×2 table (and thus we work with unconditional expected counts), while in Section 3.4 we use product multinomial sampling for a 2×2 table (and thus we work with conditional expected counts).

In Section 4.1, we describe the matrix formulation of log-linear models and explain how to calculate the MLE of the expected counts using multinomial sampling and product multinomial sampling. In Section 4.2, we state the asymptotic distributions of the MLE of the expected counts in the cases of multinomial sampling and product multinomial sampling.

In Section 4.3, we discuss the asymptotic distribution of the estimate of the log odds ratio under multinomial sampling and product multinomial sampling. In Section 4.4, we review the test statistics for the Pearson chi-square test and the likelihood ratio test. In Section 4.5, we discuss the Akaike Information Criterion (AIC) under the two sampling schemes.

In Section 4.6, we estimate the expected counts and find the estimated asymptotic standard errors (ASE) of the logarithms of their estimates. Furthermore, we calculate the goodness-of-fit statistics for the unconditional and conditional log-linear models along with their AICs and discuss the results.

One thing we observed with the analysis in this thesis is that the 'best' log-linear model is sometimes a standard non-hierarchical model. Such models, however, are difficult to interpret. For more details, see Section 4.7.

CHAPTER 2

DESCRIPTION OF THE DATA

In this chapter, we describe various thromboembolism data sets in Vessey and Doll [11], which were collected in Britain in 1967. We shall analyze some of these data in this thesis using standard and non-standard log-linear models.

Even though, in this thesis, we only analyze a 2×2 table cross-classifying two categorical variables (see Table 2.1), we describe many related datasets in Vessey and Doll [11]. Some of these are actually a refinement of the data we examine and they include more categorical variables.

Some of these data sets were analyzed by authors that discussed synergy and antagonism of two categorical variables on a response using non-standard log-linear models; see Bishop *et al.* [2], Funo [5, 6], Hasan [10], von Eye and Mun [13], von Eye, Schuster, and Rogers [14], and Worcester [15].

In May 1967, the Medical Research Council of the Statistical Research Unit of the University College Medical School London looked into the risks of thromboembolic disease in women taking oral contraceptives. The data set was collected from the hospital index.

The researchers examined women who were admitted during the period 1964–1966 with the age range between 16–40 and having been diagnosed with phlebitis, thrombophlebitis, thrombosis, or embolism. All women satisfying these criteria were selected except if they were single or widowed, had any other reason for developing the disease, were pregnant, suffered superficial thrombophlebitis, or were not interviewed because they had died.

Moreover, two control patients were selected for each patient. The control patients were matched with the affected patients based on their age, date of admission, parity, and absence of the traits used to remove patients from the affected group.

The information about these patients appears in Tables 2.1, 2.2, and 2.3 of this chapter. These were copied from Tables IV, VI, and VIII in Vessey and Doll [11, pp. 201–202], respectively. The three tables compare different factors about these patients.

Table 2.1 shows the number of patients in the treatment and control groups who had been taking oral contraceptives and the number of patients in the treatment and control groups who did not take oral contraceptives. This is a 2×2 contingency table and both factors are nominal.

Table 2.2 is an extension of Table 2.1. It now includes another nominal factor: whether or not a patient had thromboembolism in the past. This is a $2 \times 2 \times 2$ contingency table and all three factors are nominal.

Table 2.3 is another extension of Table 2.1. The smoking habit of a patient has now been added. The number of cigarettes smoked on a given day has been grouped into 3 categories. This is a $2 \times 2 \times 3$ contingency table and the smoking habit is an ordinal factor.

In this thesis, as mentioned above, we analyze only the data in Table 2.1 using hierarchical and non-hierarchical log-linear models and compare the results. We also use non-standard log-linear models, but the interpretation of the latter models is very problematic.

Table 2.1: Affected and control patients classified by the use of oral contraceptives during the month before the onset of a disease episode

	Oral contraceptive		
Diagnostic group	Used	Did not use	Total
Thromboembolism	26	32	58
Control	10	106	116
Total	36	138	174

Table 2.2: History of previous venous thrombosis or pulmonary embolism in affected and control patients

Diagnostic group	Oral contraceptive	Previous thromboembolism		Total
		Present	Absent	
Thromboembolism	Used	9	17	26
	Did not use	12	20	32
	Total	21	37	58
Control	Used	2	8	10
	Did not used	3	103	106
	Total	5	111	116

Table 2.3: Prevalence of cigarettes smoked at the onset of an episode for cases and controls

Diagnostic group	Oral contraceptive	No of cigar. smoked			Total
		0	1–14	15+	
Thromboembolism	Used	8	4	14	26
	Did not use	11	14	7	32
	Total	19	18	21	58
Control	Used	3	5	2	10
	Did not used	49	35	22	106
	Total	52	40	24	116

CHAPTER 3

THEORETICAL DISCUSSION OF VARIOUS LOG-LINEAR MODELS FOR 2×2 CONTINGENCY TABLES

In this chapter, we give a theoretical discussion of various standard and non-standard log-linear models which can be used to analyze the data in Table 2.1 of Chapter 2. The table cross-classifies the categorical variables "use of oral contraceptive" and "diagnostic group".

3.1 Introduction

Denote by i the level of the diagnostic group ($i = 1$ for cases and $i = 2$ for controls) and by j the level of the contraceptive group ($j = 1$ for use and $j = 2$ for not use). The frequencies f_{ij} for Table 2.1, in symbolic notation, appear in Table 3.1. We write

$$\boldsymbol{f} = (f_{11}, f_{12}, f_{21}, f_{22})^\top.$$

Table 3.1: Cell frequencies in symbolic notation

	Oral contraceptive		
Diagnostic group	Used ($j = 1$)	Did not use ($j = 2$)	Total
Thromboembolism ($i = 1$)	f_{11}	f_{12}	$f_{1\cdot} = N_1$
Control ($i = 2$)	f_{21}	f_{22}	$f_{2\cdot} = N_2$

There are two ways to analyze the data in Table 2.1: using multinomial sampling and product multinomial sampling. Because of the way the data were collected (see Chapter

2), the most appropriate method of analysis is product multinomial sampling (even though Bishop *et al.* [2] and Worcester [15] used multinomial sampling).

(a) Multinomial sampling

For $i, j \in \{1,2\}$, let π_{ij} be the probability that a patient belongs to the cell (i, j). See Table 3.2. We denote by $\pi_{i\cdot}$ the marginal probability that a patient belongs to diagnostic group i, where $i = 1$ for the thromboembolism patients and $i = 2$ for the control group. Let also $\pi_{\cdot j}$ be the marginal probability that a patient took ($j = 1$) or did not take an oral contraceptive ($j = 2$). We have $\sum_{i,j} \pi_{ij} = 1$. Also

$$\pi_{i\cdot} = \pi_{i1} + \pi_{i2} \quad \text{and} \quad \pi_{\cdot j} = \pi_{1j} + \pi_{2j}.$$

Assume $\sum_{i,j} f_{ij} = N$ and N is fixed. Let m_{ij} be the expected number of people in cell (i, j) under multinomial sampling. In such a case,

$$m_{ij} = N\pi_{ij} \quad \text{for } i, j \in \{1, 2\}. \tag{3.1}$$

Table 3.2: Cell probabilities for multinomial sampling

	Oral contraceptive		
Diagnostic group	Used ($j = 1$)	Did not use ($j = 2$)	Total
Thromboembolism ($i = 1$)	π_{11}	π_{12}	$\pi_{1\cdot}$
Control ($i = 2$)	π_{21}	π_{22}	$\pi_{2\cdot}$
Total	$\pi_{\cdot 1}$	$\pi_{\cdot 2}$	1

Table 3.3: Expected counts for multinomial sampling

	Oral contraceptive		
Diagnostic group	Used ($j = 1$)	Did not use ($j = 2$)	Total
Thromboembolism ($i = 1$)	m_{11}	m_{12}	$m_{1\cdot}$
Control ($i = 2$)	m_{21}	m_{22}	$m_{2\cdot}$
Total	$m_{\cdot 1}$	$m_{\cdot 2}$	N

Assume the probabilities $\pi_{ij} = \pi_{ij}(\boldsymbol{\theta})$ depend on an unknown parameter vector $\boldsymbol{\theta}$. The expected counts $m_{ij} = m_{ij}(\boldsymbol{\theta})$ also depend on $\boldsymbol{\theta}$. Then the likelihood function is

$$L(\boldsymbol{\theta};\boldsymbol{f}) = N!\prod_{i=1}^{2}\prod_{j=1}^{2}\frac{\pi_{ij}(\boldsymbol{\theta})^{f_{ij}}}{f_{ij}!} \propto \prod_{i=1}^{2}\prod_{j=1}^{2}\pi_{ij}(\boldsymbol{\theta})^{f_{ij}}.$$

From Eq. (3.1), we get

$$L(\boldsymbol{\theta};\boldsymbol{f}) = N!\prod_{i=1}^{2}\prod_{j=1}^{2}\frac{\left(\frac{m_{ij}(\boldsymbol{\theta})}{N}\right)^{f_{ij}}}{f_{ij}!} \propto \prod_{i=1}^{2}\prod_{j=1}^{2}m_{ij}(\boldsymbol{\theta})^{f_{ij}}.$$

The log-likelihood is

$$\ell(\boldsymbol{\theta};\boldsymbol{f}) = C_1 + \sum_{i,j} f_{ij}\log\pi_{ij}(\boldsymbol{\theta}) = C_2 + \sum_{i,j} f_{ij}\log m_{ij}(\boldsymbol{\theta}),$$

where C_1 and C_2 are constants. By maximizing the above log-likelihood with respect to $\boldsymbol{\theta}$, subject to the constraint

$$\sum_{i,j} m_{ij}(\boldsymbol{\theta}) = N,$$

we may find the MLE's of the π_{ij}'s or m_{ij}'s corresponding to multinomial sampling. These estimates are given in Chapter 4.

(b) Product multinomial sampling

Let $\pi_{j|i=1}$ be the conditional probability that an affected patient used ($j = 1$) or did not use ($j = 2$) oral contraceptives given that the patient had thromboembolism. Also, let $\pi_{j|i=2}$ be the conditional probability that the patient used ($j = 1$) or did not use ($j = 2$) oral contraceptives given that the patient was in the control group. See Table 3.4.

Table 3.4: Cell probabilities for product multinomial sampling

	Oral contraceptive		
Diagnostic group	Used ($j = 1$)	Did not use ($j = 2$)	Total
Thromboembolism ($i = 1$)	$\pi_{1\|1}$	$\pi_{2\|1}$	1
Control ($i = 2$)	$\pi_{1\|2}$	$\pi_{2\|2}$	1

Table 3.5: Expected counts for product multinomial sampling

Diagnostic group	Oral contraceptive: Used ($j=1$)	Oral contraceptive: Did not use ($j=2$)	Total
Thromboembolism ($i=1$)	$m_{1\|1}$	$m_{2\|1}$	N_1
Control ($i=2$)	$m_{1\|2}$	$m_{2\|2}$	N_2

Let N_1 and N_2 be the marginal totals for the patients belonging to diagnostic groups $i=1$ and $i=2$, respectively.

Let $m_{j|i=1} = N_1\pi_{j|i=1}$ be the conditional expected number of patients that used ($j=1$) or did not use ($j=2$) oral contraceptives given that they had thromboembolism. Also, let $m_{i|j=2} = N_2\pi_{j|i=2}$ be the conditional expected number of patients that used ($j=1$) or did not use ($j=2$) oral contraceptives given that they were in the control group.

Assume the probabilities $\pi_{j|i} = \pi_{(j|i)}(\boldsymbol{\theta})$ depend on an unknown parameter vector $\boldsymbol{\theta}$ for $i,j \in \{1,2\}$. The conditional expected counts $m_{j|i} = m_{j|i}(\boldsymbol{\theta})$ also depend on $\boldsymbol{\theta}$. Then the likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta};\boldsymbol{f}) &= N_1!\prod_{j=1}^{2}\frac{\pi_{(j|i=1)}(\boldsymbol{\theta})^{f_{1j}}}{f_{1j}!}\cdot N_2!\prod_{j=1}^{2}\frac{\pi_{(j|i=2)}(\boldsymbol{\theta})^{f_{2j}}}{f_{2j}!} \\ &\propto \prod_{j=1}^{2}\pi_{(j|i=1)}(\boldsymbol{\theta})^{f_{1j}}\cdot\prod_{j=1}^{2}\pi_{(j|i=2)}(\boldsymbol{\theta})^{f_{2j}}. \end{aligned}$$

From the equation $m_{j|i} = N_i\pi_{j|i}$, we get

$$\begin{aligned} L(\boldsymbol{\theta};\boldsymbol{f}) &= N_1!\prod_{j=1}^{2}\frac{\left(\frac{m_{(j|i=1)}(\boldsymbol{\theta})}{N_1}\right)^{f_{1j}}}{f_{1j}!}\cdot N_2!\prod_{j=1}^{2}\frac{\left(\frac{m_{(j|i=2)}(\boldsymbol{\theta})}{N_2}\right)^{f_{2j}}}{f_{2j}!} \\ &\propto \prod_{j=1}^{2}m_{(j|i=1)}(\boldsymbol{\theta})^{f_{1j}}\cdot\prod_{j=1}^{2}m_{(j|i=2)}(\boldsymbol{\theta})^{f_{2j}}. \end{aligned}$$

The log-likelihood function is

$$\begin{aligned} \ell(\boldsymbol{\theta};\boldsymbol{f}) &= C_3 + \sum_j f_{1j}\log\pi_{j|i=1}(\boldsymbol{\theta}) + \sum_j f_{2j}\log\pi_{j|i=2}(\boldsymbol{\theta}) \\ &= C_4 + \sum_j f_{1j}\log m_{j|i=1}(\boldsymbol{\theta}) + \sum_j f_{2j}\log m_{j|i=2}(\boldsymbol{\theta}), \end{aligned}$$

where C_3 and C_4 are constants. By maximizing the above log-likelihood with respect to $\boldsymbol{\theta}$, subject to the constraints

$$\sum_{i,j} m_{j|i=1}(\boldsymbol{\theta}) = N_1 \quad \text{and} \quad \sum_{i,j} m_{j|i=2}(\boldsymbol{\theta}) = N_2,$$

we may find the MLE's of the $\pi_{j|i}$'s or $m_{j|i}$'s corresponding to product multinomial sampling. These estimates are given in Chapter 4.

3.2 Worcester's parametrization of a 2×2 contingency table

Before we describe some non-standard log-linear models, we give Worcester's [15] multiplicative parametrization of a 2×2 contingency table. These non-standard log-linear models, which can be obtained from Worcester's [15] multiplicative models, are described in Section 3.3.2 of this thesis and they are variations of models that appear in Bishop *et al.* [2] and Hasan [10].

(a) Multinomial sampling

Worcester [15] considered the parametrization shown in Table 3.6 with

$$\boldsymbol{\theta} = (X_A, X_B, X_{AB})^\top,$$

where the parameters X_A, X_B, and X_{AB} are nonnegative. After some algebra, we find that

$$X_A = \frac{\pi_{12}}{\pi_{22}} = \frac{m_{12}}{m_{22}}, \quad X_B = \frac{\pi_{21}}{\pi_{22}} = \frac{m_{21}}{m_{22}},$$

$$\text{and} \quad X_{AB} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{m_{11}m_{22}}{m_{12}m_{21}}. \tag{3.2}$$

The parameter X_{AB} is the well-known *odds ratio.*

The null hypothesis

$$H_0 : X_{AB} = 1$$

corresponds to the independence of factors A and B because

$$X_{AB} = 1 \Leftrightarrow \pi_{12}\pi_{21} = \pi_{11}\pi_{22} \Leftrightarrow \pi_{11} = \pi_{1\cdot}\pi_{\cdot 1}.$$

The last equivalence follows from the fact that both equations are equivalent to the equation $\pi_{11} = \pi_{11}^2 + \pi_{11}\pi_{12} + \pi_{11}\pi_{21} + \pi_{12}\pi_{21}$. We can also prove that

$$X_{AB} = 1 \Leftrightarrow (\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j} \quad \text{for all} \quad i, j). \tag{3.3}$$

Table 3.6: Worcester's parametrization for multinomial sampling

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$m_{11} = NX_AX_BX_{AB}/D$	$m_{12} = NX_A/D$	
Control	$m_{21} = NX_B/D$	$m_{22} = N/D$	
	$D = X_AX_BX_{AB} + X_A + X_B + 1$		N

(b) Product multinomial sampling

Even though Worcester [15] did not consider product multinomial sampling, the parametrization in Table 3.6 implies the conditional parametrization in Table 3.7. We have

$$X_B = \frac{\pi_{1|2}}{\pi_{2|2}} = \frac{m_{1|2}}{m_{2|2}} \quad \text{and}$$

$$X_{AB} = \frac{\pi_{1|1}\pi_{2|2}}{\pi_{2|1}\pi_{1|2}} = \frac{m_{1|1}m_{2|2}}{m_{1|2}m_{2|1}}. \tag{3.4}$$

(We condition on the categories of factor A.) The parameter X_A cannot be expressed in terms of the conditional probabilities $\pi_{j|i}$, so it does not appear here. The parameter X_{AB} is again the odds ratio.

The null hypothesis

$$H_0 : X_{AB} = 1$$

can be stated equivalently as follows:

$$X_{AB} = 1 \Leftrightarrow \frac{\pi_{1|1}}{\pi_{2|1}} = \frac{\pi_{1|2}}{\pi_{2|2}} \Leftrightarrow \frac{\pi_{1|1}}{\pi_{2|1} + \pi_{1|1}} = \frac{\pi_{1|2}}{\pi_{2|2} + \pi_{1|2}} \Leftrightarrow \pi_{1|1} = \pi_{1|2}. \tag{3.5}$$

This means that $X_{AB} = 1$ if and only if the conditional probability of smoking given thromboembolism is equal to the conditional probability of smoking given no thromboembolism.

Table 3.7: Worcester's parametrization for product multinomial sampling

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$m_{1\|1} = \frac{N_1X_BX_{AB}}{X_BX_{AB}+1}$	$m_{2\|1} = \frac{N_1}{X_BX_{AB}+1}$	N_1
Control	$m_{1\|2} = \frac{N_2X_B}{X_B+1}$	$m_{2\|2} = \frac{N_2}{X_B+1}$	N_2

3.3 Log-linear models for a 2×2 contingency table for multinomial sampling

We use two different kinds of log-linear models for the data in Tables 2.1 and 3.1: (a) Standard log-linear models and (b) non-standard log-linear models. In this section, we use multinomial sampling for the 2×2 table, and thus we work with the unconditional expected counts, m_{ij}.

3.3.1 Using standard log-linear models for multinomial sampling

The general standard log-linear model for Table 3.3 (for multinomial sampling) is

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \quad (i, j \in \{1, 2\}). \tag{3.6}$$

The total number of parameters here is $1 + 2 + 2 + 4 = 9$. Usually, we impose one of the following three sets of restrictions in order to reduce the number of parameters to $2 \cdot 2 = 4$, which is the number of cells in the table.

i. **Zero-sum constraints:**

$\sum_i \lambda_i^A = 0$, $\sum_j \lambda_j^B = 0$, $\sum_i \lambda_{ij}^{AB} = 0$ for each j, and $\sum_j \lambda_{ij}^{AB} = 0$ for each i.

In this case,

$$\begin{aligned}
\log m_{11} &= \lambda + \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB}, \\
\log m_{12} &= \lambda + \lambda_1^A - \lambda_1^B - \lambda_{11}^{AB}, \\
\log m_{21} &= \lambda - \lambda_1^A + \lambda_1^B - \lambda_{11}^{AB}, \\
\log m_{22} &= \lambda - \lambda_1^A - \lambda_1^B + \lambda_{11}^{AB}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\lambda &= \frac{1}{4}\log m_{11} + \frac{1}{4}\log m_{12} + \frac{1}{4}\log m_{21} + \frac{1}{4}\log m_{22}, \\
\lambda_1^A &= \frac{1}{4}\log m_{11} + \frac{1}{4}\log m_{12} - \frac{1}{4}\log m_{21} - \frac{1}{4}\log m_{22}, \\
\lambda_1^B &= \frac{1}{4}\log m_{11} - \frac{1}{4}\log m_{12} + \frac{1}{4}\log m_{21} - \frac{1}{4}\log m_{22}, \\
\lambda_{11}^{AB} &= \frac{1}{4}\log m_{11} - \frac{1}{4}\log m_{12} - \frac{1}{4}\log m_{21} + \frac{1}{4}\log m_{22}.
\end{aligned}$$

In addition, $\lambda_1^A = -\lambda_2^A$, $\lambda_1^B = -\lambda_2^B$, and $\lambda_{11}^{AB} = -\lambda_{12}^{AB} = -\lambda_{21}^{AB} = \lambda_{22}^{AB}$.

ii. **Zero constraints for the first categories (R constraints):**

$\lambda_1^A = 0$, $\lambda_1^B = 0$, $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$.

In this case,

$$\begin{aligned}
\log m_{11} &= \lambda, \\
\log m_{12} &= \lambda + \lambda_2^B, \\
\log m_{21} &= \lambda + \lambda_2^A, \\
\log m_{22} &= \lambda + \lambda_2^A + \lambda_2^B + \lambda_{22}^{AB}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\lambda &= \log m_{11}, \\
\lambda_2^B &= \log m_{12} - \log m_{11}, \\
\lambda_2^A &= \log m_{21} - \log m_{11}, \\
\lambda_{22}^{AB} &= \log m_{11} - \log m_{12} - \log m_{21} + \log m_{22}.
\end{aligned}$$

iii. **Zero constraints for the last categories (SAS constraints):**

$\lambda_2^A = 0$, $\lambda_2^B = 0$, $\lambda_{i2}^{AB} = 0$ for each i, and $\lambda_{21}^{AB} = 0$.

In this case,

$$\begin{aligned}
\log m_{11} &= \lambda + \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB}, \\
\log m_{12} &= \lambda + \lambda_1^A, \\
\log m_{21} &= \lambda + \lambda_1^B, \\
\log m_{22} &= \lambda.
\end{aligned}$$

It follows that

$$\begin{aligned}
\lambda &= \log m_{22}, \\
\lambda_1^A &= \log m_{12} - \log m_{22}, \\
\lambda_1^B &= \log m_{21} - \log m_{22}, \\
\lambda_{11}^{AB} &= \log m_{11} - \log m_{12} - \log m_{21} + \log m_{22}.
\end{aligned}$$

Standard log-linear models are divided into (1) hierarchical log-linear models (Section 3.3.1.1) and (2) non-hierarchical log-linear models (Section 3.3.1.2). As mentioned before, a hierarchical model is one where, if a higher order term is included in the log-linear model, then so are all the lower order terms.

3.3.1.1 Using standard hierarchical log-linear models (for multinomial sampling)

It is well-known (e.g., see Agresti [1]) that no matter which restrictions on the λ parameters we use, for each pair (i, j), the estimates of the parameter m_{ij} in a standard hierarchical model are the same.

We estimate the following hierarchical submodels of the saturated model (3.6):

i. $\log m_{ij} = \lambda$

ii. $\log m_{ij} = \lambda + \lambda_i^A$

iii. $\log m_{ij} = \lambda + \lambda_j^B$

iv. $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$

Log-linear model iv above is equivalent to the independence model (3.3) in Section 3.2. We prove this claim in Appendix A.1.

The above unconditional standard hierarchical log-linear models are estimated in Tables 4.3–4.7 in Section 4.6.

3.3.1.2 Using standard non-hierarchical log-linear models (for multinomial sampling)

Standard non-hierarchical log-linear submodels of the saturated model (3.6) are those that are missing at least one of the terms λ, λ_i^A, λ_j^B, and λ_{ij}^{AB}. In this thesis, we always include the intercept λ in a standard log-linear model corresponding to multinomial sampling.

In standard non-hierarchical models, we get different results when we impose different kinds of zero constraints (and that is one of the reasons these are not used in practice).

Because of the above problems and because of time constraints, we only study the following standard non-hierarchical model under three different sets of parameter restrictions:

i. $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ with $\sum_i \lambda_{ij}^{AB} = 0$ for each j and $\sum_j \lambda_{ij}^{AB} = 0$ for each i. Thus we have

$$\begin{aligned}
\log m_{11} &= \lambda + \lambda_{11}^{AB}, \\
\log m_{12} &= \lambda - \lambda_{11}^{AB}, \\
\log m_{21} &= \lambda - \lambda_{11}^{AB}, \\
\log m_{22} &= \lambda + \lambda_{11}^{AB}.
\end{aligned}$$

ii. $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ with $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$. Thus we have

$$\begin{aligned}
\log m_{11} &= \lambda, \\
\log m_{12} &= \lambda, \\
\log m_{21} &= \lambda, \\
\log m_{22} &= \lambda + \lambda_{22}^{AB}.
\end{aligned}$$

iii. $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ with $\lambda_{i2}^{AB} = 0$ for each i, and $\lambda_{21}^{AB} = 0$. Thus we have

$$\begin{aligned}
\log m_{11} &= \lambda + \lambda_{11}^{AB}, \\
\log m_{12} &= \lambda, \\
\log m_{21} &= \lambda, \\
\log m_{22} &= \lambda.
\end{aligned}$$

The above three unconditional standard non-hierarchical log-linear models are estimated in Tables 4.8–4.10 in Section 4.6.

3.3.2 Using non-standard log-linear models for multinomial sampling

In this section, for multinomial sampling, we give the log-linear version of the Worcester multiplicative model given in Table 3.6. Let

$$w_A = \log X_A, \quad w_B = \log X_B, \quad w_{AB} = \log X_{AB}, \quad w = \log\left(\frac{N}{D}\right),$$

$$\delta_1 = \delta_{11} = 1, \quad \text{and} \quad \delta_2 = \delta_{12} = \delta_{21} = \delta_{22} = 0.$$

The parameters in Table 3.6 can be then written in the following way:

$$\begin{aligned}
\log m_{11} &= \log \frac{N}{D} + \log X_A + \log X_B + \log X_{AB} \\
&= w + \delta_1 w_A + \delta_1 w_B + \delta_{11} w_{AB}, \\
\log m_{12} &= \log \frac{N}{D} + \log X_A \\
&= w + \delta_1 w_A + \delta_2 w_B + \delta_{12} w_{AB}, \\
\log m_{21} &= \log \frac{N}{D} + \log X_B \\
&= w + \delta_2 w_A + \delta_1 w_B + \delta_{21} w_{AB}, \\
\log m_{22} &= \log \frac{N}{D} \\
&= w + \delta_2 w_A + \delta_2 w_B + \delta_{22} w_{AB}.
\end{aligned}$$

For multinomial sampling, we may thus write Worcester's multiplicative model as a non-standard log-linear model:

$$\log m_{ij} = w + \delta_i w_A + \delta_j w_B + \delta_{ij} w_{AB} \quad \text{for } i, j \in \{1, 2\}. \tag{3.7}$$

The above formulation of Worcester's model in a non-standard log-linear form is similar to the one in Bishop *at al.* [2, pp. 111–114] for $2 \times 2 \times 2$ tables. See also Hasan [10].

In Sections 3.2 and 3.3.1.1 of this thesis, we have seen that the independence model is equivalent to $X_{AB} = 1$; i.e, $w_{AB} = 0$. As in Appendix A.1, we may prove that

i. $\log m_{ij} = \lambda \iff X_A = X_B = X_{AB} = 1 \iff w_A = w_B = w_{AB} = 0$

ii. $\log m_{ij} = \lambda + \lambda_i^A \iff X_B = X_{AB} = 1 \iff w_B = w_{AB} = 0$

iii. $\log m_{ij} = \lambda + \lambda_j^B \iff X_A = X_{AB} = 1 \iff w_A = w_{AB} = 0$

iv. $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B \iff X_{AB} = 1 \iff w_{AB} = 0$

Using Eq. (3.7), we may thus reformulate models i-iv above as follows:

i. $\log m_{ij} = w$

ii. $\log m_{ij} = w + \delta_i w_A$

iii. $\log m_{ij} = w + \delta_j w_B$

iv. $\log m_{ij} = w + \delta_i w_A + \delta_j w_B$

We estimate submodels i-iv and the saturated models (3.6)/(3.7) in Tables 4.3–4.7 in Section 4.6.1. In addition, asymptotic variance-covariance matrices of $\hat{\boldsymbol{w}}$ under multinomial sampling are given in Appendix B.1.

3.4 Log-linear models for a 2×2 contingency table for product multinomial sampling

3.4.1 Using standard log-linear models for product multinomial sampling

The general standard log-linear model for Table 3.5 (for product multinomial sampling) is

$$\log m_{j|i} = \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \quad (i, j \in \{1, 2\}). \tag{3.8}$$

The total number of parameters here is $2 + 2 + 4 = 8$. Usually, we impose one of the following three sets of restrictions in order to reduce the number of parameters to $2 \cdot 2 = 4$, which is the number of cells in the table.

i. **Zero-sum constraints:**

$\sum_j \lambda_j^B = 0$, $\sum_i \lambda_{ij}^{AB} = 0$ for each j, and $\sum_j \lambda_{ij}^{AB} = 0$ for each i.

In this case,

$$\begin{aligned}
\log m_{1|1} &= \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB}, \\
\log m_{1|2} &= \lambda_2^A + \lambda_1^B - \lambda_{11}^{AB}, \\
\log m_{2|1} &= \lambda_1^A - \lambda_1^B - \lambda_{11}^{AB}, \\
\log m_{2|2} &= \lambda_2^A - \lambda_1^B + \lambda_{11}^{AB}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\lambda_1^A &= \frac{1}{2}\log m_{1|1} + \frac{1}{2}\log m_{2|1}, \\
\lambda_2^A &= \frac{1}{2}\log m_{1|2} + \frac{1}{2}\log m_{2|2}, \\
\lambda_1^B &= \frac{1}{4}\log m_{1|1} + \frac{1}{4}\log m_{1|2} - \frac{1}{4}\log m_{2|1} - \frac{1}{4}\log m_{2|2}, \\
\lambda_{11}^{AB} &= \frac{1}{4}\log m_{1|1} - \frac{1}{4}\log m_{1|2} - \frac{1}{4}\log m_{2|1} + \frac{1}{4}\log m_{2|2}.
\end{aligned}$$

In addition, $\lambda_1^B = -\lambda_2^B$, and $\lambda_{11}^{AB} = -\lambda_{12}^{AB} = -\lambda_{21}^{AB} = \lambda_{22}^{AB}$.

ii. **Zero constraints for the first categories (R constraints):**

$\lambda_1^B = 0$, $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$.

In this case,

$$\begin{aligned}
\log m_{1|1} &= \lambda_1^A, \\
\log m_{1|2} &= \lambda_2^A, \\
\log m_{2|1} &= \lambda_1^A + \lambda_2^B, \\
\log m_{2|2} &= \lambda_2^A + \lambda_2^B + \lambda_{22}^{AB}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\lambda_1^A &= \log m_{1|1}, \\
\lambda_2^A &= \log m_{1|2}, \\
\lambda_2^B &= \log m_{2|1} - \log m_{1|1}, \\
\lambda_{22}^{AB} &= \log m_{1|1} - \log m_{1|2} - \log m_{2|1} + \log m_{2|2}.
\end{aligned}$$

iii. **Zero constraints for the last categories (SAS constraints):**

$\lambda_2^B = 0$, $\lambda_{i2}^{AB} = 0$ for each i, and $\lambda_{21}^{AB} = 0$.

In this case,

$$\begin{aligned}
\log m_{1|1} &= \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB}, \\
\log m_{1|2} &= \lambda_2^A + \lambda_1^B, \\
\log m_{2|1} &= \lambda_1^A, \\
\log m_{2|2} &= \lambda_2^A.
\end{aligned}$$

It follows that

$$\begin{aligned}
\lambda_1^A &= \log m_{2|1}, \\
\lambda_2^A &= \log m_{2|2}, \\
\lambda_1^B &= \log m_{1|2} - \log m_{2|2}, \\
\lambda_{11}^{AB} &= \log m_{1|1} - \log m_{1|2} - \log m_{2|1} + \log m_{22}.
\end{aligned}$$

3.4.1.1 Using standard hierarchical log-linear models (for product multinomial sampling)

We shall see in Section 4.6 that, no matter which restrictions on the λ parameters we use, for each pair (i, j), the estimates of the parameter $m_{j|i}$ in a standard hierarchical model are the same.

We estimate the following hierarchical submodels of the model (3.8):

i. $\log m_{j|i} = \lambda_i^A$

ii. $\log m_{j|i} = \lambda_i^A + \lambda_j^B$

Log-linear model ii above is equivalent to the independence model (3.5) in Section 3.2. We prove this claim in Appendix A.2.

The above conditional standard hierarchical log-linear models are estimated in Tables 4.11–4.13 in Section 4.6.

3.4.1.2 Using standard non-hierarchical log-linear models (for product multinomial sampling)

Standard non-hierarchical log-linear submodels of the saturated model (3.8) are those that are missing at least one of the terms λ_i^A, λ_j^B, and λ_{ij}^{AB}. In this thesis, we always include the intercept λ_i^A in a standard log-linear model corresponding to product multinomial sampling.

In standard non-hierarchical models, we get different results when we impose different kinds of zero constraints (and that is one of the reasons these are not used in practice).

Because of the above problems and because of time constraints, we only study the following standard non-hierarchical model under three different sets of parameter restrictions:

i. $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ with $\sum_i \lambda_{ij}^{AB} = 0$ for each j and $\sum_j \lambda_{ij}^{AB} = 0$ for each i. Thus we have

$$\begin{aligned}
\log m_{1|1} &= \lambda_1^A + \lambda_{11}^{AB}, \\
\log m_{2|1} &= \lambda_1^A - \lambda_{11}^{AB}, \\
\log m_{1|2} &= \lambda_2^A - \lambda_{11}^{AB}, \\
\log m_{2|2} &= \lambda_2^A + \lambda_{11}^{AB}.
\end{aligned}$$

ii. $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ with $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$. Thus we have

$$\begin{aligned}
\log m_{1|1} &= \lambda_1^A, \\
\log m_{2|1} &= \lambda_1^A, \\
\log m_{1|2} &= \lambda_2^A, \\
\log m_{2|2} &= \lambda_2^A + \lambda_{22}^{AB}.
\end{aligned}$$

iii. $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ with $\lambda_{i2}^{AB} = 0$ for each i, and $\lambda_{21}^{AB} = 0$. Thus we have

$$\begin{aligned}
\log m_{1|1} &= \lambda_1^A + \lambda_{11}^{AB}, \\
\log m_{2|1} &= \lambda_1^A, \\
\log m_{1|2} &= \lambda_2^A, \\
\log m_{2|2} &= \lambda_2^A.
\end{aligned}$$

The above three conditional standard non-hierarchical log-linear models are estimated in Tables 4.14–4.16 in Section 4.6.

3.4.2 Using non-standard log-linear models for product multinomial sampling

In this section, for product multinomial sampling, we give the log-linear version of the Worcester multiplicative model given in Table 3.7. Let

$$\begin{aligned}
&v_1 = \log\left(\frac{N_1}{D_1}\right), \quad v_2 = \log\left(\frac{N_2}{D_2}\right), \quad v_B = \log X_B, \quad v_{AB} = \log X_{AB}, \\
&\delta_1 = \delta_{1|1} = 1, \quad \text{and} \quad \delta_2 = \delta_{2|1} = \delta_{1|2} = \delta_{2|2} = 0,
\end{aligned}$$

where $D_1 = X_B X_{AB} + 1$ and $D_2 = X_B + 1$. In other words, δ_i and $\delta_{j|i}$ are equal to 1 if and only if all of their indices are equal to 1; otherwise, they are equal to 0.

The parameters in Table 3.7 can be then written in the following way:

$$\begin{aligned}
\log m_{1|1} &= \log \frac{N_1}{D_1} + \log X_B + \log X_{AB} \\
&= v_1 + \delta_1 v_B + \delta_{1|1} v_{AB}, \\
\log m_{2|1} &= \log \frac{N_1}{D_1} \\
&= v_1 + \delta_2 v_B + \delta_{2|1} v_{AB}, \\
\log m_{1|2} &= \log \frac{N_2}{D_2} + \log X_B \\
&= v_2 + \delta_1 v_B + \delta_{1|2} v_{AB}, \\
\log m_{2|2} &= \log \frac{N_2}{D_2} \\
&= v_2 + \delta_2 v_B + \delta_{2|2} v_{AB}.
\end{aligned}$$

For product multinomial sampling, we may thus write Worcester's multiplicative model as a non-standard log-linear model:

$$\log m_{j|i} = v_i + \delta_j v_B + \delta_{j|i} v_{AB} \quad \text{for } i, j \in \{1, 2\}. \tag{3.9}$$

In Section 3.2 of this thesis, we have seen that the independence model is equivalent to $X_{AB} = 1$; i.e, $v_{AB} = 0$. As in Appendix A.2, we may prove that

i. $\log m_{j|i} = \lambda_i^A \iff X_B = X_{AB} = 1 \iff v_B = v_{AB} = 0$

ii. $\log m_{j|i} = \lambda_i^A + \lambda_j^B \iff X_{AB} = 1 \iff v_{AB} = 0$

Using Eq. (3.9), we may thus reformulate models i-ii above as follows:

i. $\log m_{j|i} = v_i$

ii. $\log m_{j|i} = v_i + \delta_j v_B$

We estimate submodels i-ii and the saturated models (3.8)/(3.9) in Tables 4.11–4.13 in Section 4.6.2.

CHAPTER 4

ANALYSIS OF THE EFFECT OF ORAL CONTRACEPTIVE ON THROMBOEMBOLISM USING LOG-LINEAR MODELS

In this chapter, we fit the various log-linear models developed in Chapter 3 of this thesis to the data in Table 2.1, which cross-classifies the categorical variables "having thromboembolism" and "using oral contraceptives". We give estimates and standard errors for the expected counts in the aformentioned 2×2 contingency table under each model. We do that both for multinomial sampling and product multinomial sampling.

4.1 Matrix formulation of a log-linear model

A general log-linear model for a 2×2 contingency table can be written in the form

$$\log \boldsymbol{m} = \boldsymbol{X}\boldsymbol{\theta}, \tag{4.1}$$

where $\boldsymbol{X}$ is a $4 \times q$ model matrix, $\boldsymbol{\theta}$ is a $q \times 1$ vector of population parameters, and

$$\boldsymbol{m} = \begin{cases} (m_{11}, m_{12}, m_{21}, m_{22})^\top, & \text{in the unconditional case;} \\ (m_{1|1}, m_{2|1}, m_{1|2}, m_{2|2})^\top, & \text{in the conditional case.} \end{cases}$$

See Tables 3.3 and 3.5 in Section 3.1.

In this thesis, we assume that $\boldsymbol{X}$ is a full column rank matrix and that $\boldsymbol{\theta}$ has no redundant parameters. As a result, the $q \times q$ square matrix $\boldsymbol{X}^\top \boldsymbol{X}$ is non-singular (i.e., invertible).

For example, for the unconditional log-linear model $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$ with the contraints $\lambda_2^A = \lambda_2^B = 0$, we have

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\theta} = \begin{pmatrix} \lambda \\ \lambda_1^A \\ \lambda_1^B \end{pmatrix}.$$

As another example, for the conditional log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_j^B$ with the contraints $\lambda_1^A + \lambda_2^A = 0 = \lambda_1^B + \lambda_2^B$, we have

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\theta} = \begin{pmatrix} \lambda_1^A \\ \lambda_1^B \end{pmatrix}.$$

The calculation of the MLE of $\boldsymbol{\theta}$ using multinomial sampling and product multinomial sampling was discussed in Section 3.1.

If $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$, then $\hat{\boldsymbol{m}} = \boldsymbol{m}(\hat{\boldsymbol{\theta}})$ is the MLE of $\boldsymbol{m}$ by the invariance property of the MLE. If Ω_0 is the linear space [1] defined by the model $\log \boldsymbol{m} = \boldsymbol{X}\boldsymbol{\theta}$, then (in both cases) the MLE $\hat{\boldsymbol{m}}$ of $\boldsymbol{m}$ satisfies the matrix equation

$$\boldsymbol{P}_{\Omega_0} \hat{\boldsymbol{m}} = \boldsymbol{P}_{\Omega_0} \boldsymbol{f}, \tag{4.2}$$

where $\boldsymbol{P}_{\Omega_0}$ is the orthogonal projection on the linear space Ω_0 . Here

$$\boldsymbol{f} = (f_{11}, f_{12}, f_{21}, f_{22})^\top.$$

If $\boldsymbol{X}$ is of full column rank, we have that $\boldsymbol{P}_{\Omega_0} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top$. This is the well-known hat matrix $\boldsymbol{H}$ in regression. For example, in the model $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$, we have

$$\boldsymbol{P}_{\Omega_0} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & \frac{3}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \end{pmatrix},$$

[1] Strictly speaking, $\Omega_0 = \{\boldsymbol{\mu} \in \mathbb{R}^4 | \exists \boldsymbol{\theta} \in \mathbb{R}^q : \boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\theta}\}$, where q is the number of columns of $\boldsymbol{X}$. We assume that $\boldsymbol{\mu} = \log \boldsymbol{m} \in \Omega_0$.

while in the model $\log m_{j|i} = \lambda_i^A + \lambda_j^B$, we have

$$\boldsymbol{P}_{\Omega_0} = \begin{pmatrix} \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Because the MLE $\hat{\boldsymbol{m}}$ depends on the $q \times 1$ vector $\boldsymbol{\theta}$, the system (4.2) has q unknowns and 4 equations. It can be solved using the Newton-Raphson method. We use the statistical software **R** to solve the system for each one of the log-linear models in Chapter 3.

If $\boldsymbol{X}$ is of full column rank and we pre-multiply both sides of (4.2) by $\boldsymbol{X}^\top$, we get

$$\boldsymbol{X}^\top \hat{\boldsymbol{m}} = \boldsymbol{X}^\top \boldsymbol{f}. \tag{4.3}$$

See Christensen [4, p. 400].

Equations (4.2) and (4.3) are valid both for multinomial sampling and product multinomial sampling.

Remark 4.1.1. If the design matrix $\mathbf{X}$ is of full column rank, it follows from Eq.(4.1) that

$$\boldsymbol{\theta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\log \boldsymbol{m}).$$

Thus, the estimated $\boldsymbol{\theta}$ equals

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\log \hat{\boldsymbol{m}}). \tag{4.4}$$

This follows from the invariance property of the MLE.

4.2 Asymptotic variance-covariance matrices of the vector MLE

In this section, we state the asymptotic distributions of the MLE $\hat{\boldsymbol{m}}$ in the cases of multinomial sampling and product multinomial sampling.

4.2.1 Asymptotic results for multinomial sampling

In multinomial sampling, we have $\boldsymbol{m} = (m_{11}, m_{12}, m_{21}, m_{22})^\top$. According to Christensen [4, Section 12.3], in the case, we have

$$\log \hat{\boldsymbol{m}} - \log \boldsymbol{m} \sim \text{ MVN}(\mathbf{0}, (\boldsymbol{A}(\boldsymbol{m}) - \boldsymbol{A}_{\boldsymbol{z}}(\boldsymbol{m}))\boldsymbol{D}^{-1}(\boldsymbol{m})) \tag{4.5}$$

approximately for large N, where $N = \sum_{i,j} f_{ij}$. Here $\mathbf{0}$ is an $IJ \times 1 = 4 \times 1$ zero vector, $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{m})$ is an $IJ \times IJ = 4 \times 4$ diagonal matrix with the vector $\boldsymbol{m}$ in its main diagonal,

$$\boldsymbol{A}(\boldsymbol{m}) = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{D}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{D}, \quad \text{and} \tag{4.6}$$

$$\boldsymbol{A}_{\boldsymbol{z}}(\boldsymbol{m}) = \boldsymbol{J}(\boldsymbol{J}^\top \boldsymbol{D}\boldsymbol{J})^{-1}\boldsymbol{J}^\top \boldsymbol{D} = \boldsymbol{J}\boldsymbol{J}^\top \boldsymbol{D}/\boldsymbol{N}. \tag{4.7}$$

In the last equation, we denote by $\boldsymbol{J}$ the $IJ \times 1 = 4 \times 1$ vector of 1's.

If $\hat{m}_{ij}$ is the k^{th} element of the vector $\hat{\boldsymbol{m}}$, then the square root of the k^{th} diagonal element of the asymptotic variance-covariance matrix $(\boldsymbol{A}(\boldsymbol{m}) - \boldsymbol{A}_{\boldsymbol{z}}(\boldsymbol{m}))\boldsymbol{D}^{-1}(\boldsymbol{m})$ gives the theoretical asymptotic standard error of $\log \hat{m}_{ij}$. We get the estimated asymptotic standard error of $\hat{m}_{ij}$ by replacing $\boldsymbol{m}$ with $\hat{\boldsymbol{m}}$. We calculate these for our data in Table 2.1 and the different models (under multinomial sampling) in Section 4.3.

4.2.2 Asymptotic results for product multinomial sampling

In product multinomial sampling, we have

$$\boldsymbol{m}^{(1)\top} = (m_{1|1}, m_{2|1}) \quad \text{and} \quad \boldsymbol{m}^{(2)\top} = (m_{1|2}, m_{2|2})^\top.$$

Also, $\boldsymbol{m} = (\boldsymbol{m}^{(1)\top}, \boldsymbol{m}^{(2)\top})^\top$. According to Christensen [4, Section 10.3], in this case, we have

$$\log \hat{\boldsymbol{m}} - \log \boldsymbol{m} \sim \text{ MVN}(\mathbf{0}, (\boldsymbol{A}(\boldsymbol{m}) - \boldsymbol{A}_{\boldsymbol{z}}(\boldsymbol{m}))\boldsymbol{D}^{-1}(\boldsymbol{m})). \tag{4.8}$$

approximately for large N_1 and N_2, where $N_1 = f_{1|1} + f_{2|1}$ and $N_2 = f_{1|2} + f_{2|2}$

Here $\mathbf{0}$ is an $IJ \times 1 = 4 \times 1$ zero vector, $\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{m})$ is an $IJ \times IJ = 4 \times 4$ diagonal matrix with the vector $\boldsymbol{m} = (\boldsymbol{m}^{(1)\top}, \boldsymbol{m}^{(2)\top})^\top$ in its main diagonal,

$$\boldsymbol{A}(\boldsymbol{m}) = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{D}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{D}, \quad \text{and} \tag{4.9}$$

$$\boldsymbol{A}_{\boldsymbol{z}}(\boldsymbol{m}) = \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{D}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top \boldsymbol{D}. \tag{4.10}$$

In addition,

$$\boldsymbol{Z} = \begin{pmatrix} \mathbf{J} & 0 \\ 0 & \mathbf{J} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

If $\hat{m}_{j|i}$ is the k^{th} element of the vector $\hat{\boldsymbol{m}}$, then the square root of the k^{th} diagonal element of the asymptotic variance-covariance matrix $(\boldsymbol{A}(\boldsymbol{m}) - \boldsymbol{A}_{\boldsymbol{z}}(\boldsymbol{m}))\boldsymbol{D}^{-1}(\boldsymbol{m})$ gives the

theoretical asymptotic standard error of $\log \hat{m}_{ij}$. We get the estimated asymptotic standard error of $\log \hat{m}_{ij}$ by replacing $\boldsymbol{m}$ with $\hat{\boldsymbol{m}}$. We calculate these for our data in Table 2.1 and the different models (under product multinomial sampling) in Section 4.3.

Remark 4.2.1. Because of Eqs. (4.4), (4.5), and (4.8), it follows that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \sim \text{ MVN}(\mathbf{0}, (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X} \text{V}(\log \hat{\boldsymbol{m}}) \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}). \tag{4.11}$$

This is true for both sampling schemes. For multinomial sampling Eq. (4.11) is approximately true for large N, while for product multinomial sampling Eq. (4.11) is approximately true for large N_1 and N_2.

4.3 The asymptotic distribution of the estimate of the log odds ratio

The *odds ratio* for a 2×2 table was defined by Eqs. (3.2) and (3.4) in Section 3.2. In particular, for multinomial sampling (see Table 3.3) the odds ratio is

$$X_{AB} = \frac{m_{11} m_{22}}{m_{12} m_{21}},$$

while for product multinomial sampling (see Table 3.5) the odds ratio is

$$X_{AB} = \frac{m_{1|1} m_{2|2}}{m_{1|2} m_{2|1}}.$$

Thus, the *log odds ratio* is

$$\log X_{AB} = \boldsymbol{c}^\top \log \boldsymbol{m},$$

where

$$\boldsymbol{c} = (1, -1, -1, 1)^\top \quad \text{and}$$

$$\boldsymbol{m} = \begin{cases} (m_{11}, m_{12}, m_{21}, m_{22})^\top, & \text{in the unconditional case;} \\ (m_{1|1}, m_{2|1}, m_{1|2}, m_{2|2})^\top, & \text{in the conditional case.} \end{cases}$$

4.3.1 The asymptotic distribution of $\log \hat{X}_{AB}$ under multinomial sampling

In the case of multinomial sampling, a point estimate of the log odds ratio, $\log X_{AB}$, is

$$\log \hat{X}_{AB} = \log \hat{m}_{11} - \log \hat{m}_{12} - \log \hat{m}_{21} + \log \hat{m}_{22} = \boldsymbol{c}^\top \log \hat{\boldsymbol{m}}.$$

The estimate has an asymptotic normal distribution with mean $\boldsymbol{c}^\top \log \boldsymbol{m}$ and asymptotic variance is given by

$$\mathrm{V}(\log \hat{X}_{AB}) = \boldsymbol{c}^\top (\boldsymbol{A} - \boldsymbol{A}_{\boldsymbol{z}}) \boldsymbol{D}^{-1} \boldsymbol{c} = \boldsymbol{c}^\top \boldsymbol{A} \boldsymbol{D}^{-1} \boldsymbol{c}, \tag{4.12}$$

where $\boldsymbol{D}$ is an $IJ \times IJ = 4 \times 4$ diagonal matrix with the vector $\boldsymbol{m}$ in its main diagonal, $\boldsymbol{A}$ is defined by Eq. (4.6), and $\boldsymbol{A}_{\boldsymbol{z}}$ is defined by Eq. (4.7). As usual, we get the estimated asymptotic variance of $\log \hat{X}_{AB}$ by replacing $\boldsymbol{m}$ with the MLE $\hat{\boldsymbol{m}}$.

For most of the unconditional log-linear models in Chapter 3, $\log \hat{X}_{AB} = 1$, and the asymptotic variance is zero. For the saturated model in Eq. (3.6), however, the estimated asymptotic variance is

$$\hat{\mathrm{V}}(\log \hat{X}_{AB}) = \frac{1}{f_{11}} + \frac{1}{f_{12}} + \frac{1}{f_{21}} + \frac{1}{f_{22}}.$$

See Christensen [4, pp. 40–41].

4.3.2 The asymptotic distribution of $\log \hat{X}_{AB}$ under product multinomial sampling

In the case of product multinomial sampling, a point estimate of the log odds ratio, $\log X_{AB}$, is

$$\log \hat{X}_{AB} = \log \hat{m}_{1|1} - \log \hat{m}_{2|1} - \log \hat{m}_{1|2} + \log \hat{m}_{2|2} = \boldsymbol{c}^\top \log \hat{\boldsymbol{m}}.$$

The estimate has an asymptotic normal distribution with mean $\boldsymbol{c}^\top \log \boldsymbol{m}$ and asymptotic variance is given by

$$\mathrm{V}(\log \hat{X}_{AB}) = \boldsymbol{c}^\top (\boldsymbol{A} - \boldsymbol{A}_{\boldsymbol{z}}) \boldsymbol{D}^{-1} \boldsymbol{c} = \boldsymbol{c}^\top \boldsymbol{A} \boldsymbol{D}^{-1} \boldsymbol{c}, \tag{4.13}$$

where $\boldsymbol{D}$ is an $IJ \times IJ = 4 \times 4$ diagonal matrix with the vector $\boldsymbol{m}$ in its main diagonal, $\boldsymbol{A}$ is defined by Eq. (4.9), and $\boldsymbol{A}_{\boldsymbol{z}}$ is defined by Eq. (4.10). As usual, we get the estimated asymptotic variance of $\log \hat{X}_{AB}$ by replacing $\boldsymbol{m}$ with the MLE $\hat{\boldsymbol{m}}$.

For most of the conditional log-linear models in Chapter 3, $\log \hat{X}_{AB} = 1$, and the asymptotic variance is zero. For the saturated model in Eq. (3.8), however, the estimated asymptotic variance is

$$\hat{\mathrm{V}}(\log \hat{X}_{AB}) = \frac{1}{f_{11}} + \frac{1}{f_{12}} + \frac{1}{f_{21}} + \frac{1}{f_{22}}.$$

Remark 4.3.1. Even though the asymptotic variance-covariance matrices of $\log \hat{\boldsymbol{m}}$ for multinomial sampling and product multinomial sampling are different, Eqs. (4.12) and (4.13) show that the asymptotic variance of $\log \hat{X}_{AB}$ are the same for both sampling schemes. This is also true for multiway contingency tables.

4.4 Goodness-of-fits tests for log-linear models

Two of the most common goodness-of-fit statistics for log-linear models are the Pearson chi-square test statistic and the likelihood ratio test statistic.

4.4.1 The Pearson chi-square test statistic

Assuming multinomial sampling, consider Tables 3.1 and 3.3 in Section 3.1. The *Pearson chi-square test statistic* is defined by

$$X^2 = \sum_{i,j} \frac{(f_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}. \tag{4.14}$$

For product multinomial sampling, we consider Tables 3.1 and 3.5 and the Pearson chi-square test statistic is given by

$$X^2 = \sum_{i,j} \frac{(f_{ij} - \hat{m}_{j|i})^2}{\hat{m}_{j|i}}. \tag{4.15}$$

For each type of sampling, under any of the log-linear models in Chapter 3, the X^2 statistic has an asymptotic chi-square distribution with degrees of freedom equal to

$$df = IJ - (\text{number of free parameters}),$$

where I is the number of rows and J is the number of columns of the contingency table.

Tables 4.1 and 4.2 give the number of free parameters and the residual degrees of freedom for the hierarchical log-linear models in Chapter 3.

Table 4.1: Degrees of freedom for multinomial sampling

Model	Number of free parameters	Degrees of freedom
$\log m_{ij} = \lambda$	1	$IJ - 1 = 3$
$\log m_{ij} = \lambda + \lambda_i^A$	$I = 2$	$IJ - I = 2$
$\log m_{ij} = \lambda + \lambda_j^B$	$J = 2$	$IJ - J = 2$
$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$	$I + J - 1 = 3$	$(I-1)(J-1) = 1$
$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$	$IJ = 4$	0

Table 4.2: Degrees of freedom for product multinomial sampling

Model	Number of free parameters	Degrees of freedom
$\log m_{j\|i} = \lambda_i^A$	$I = 2$	$IJ - I = 2$
$\log m_{j\|i} = \lambda_i^A + \lambda_j^B$	$I + J - 1 = 3$	$(I-1)(J-1) = 1$
$\log m_{j\|i} = \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$	$IJ = 4$	0

4.4.2 The likelihood ratio test statistic

Assuming multinomial sampling, consider Tables 3.1 and 3.3 in Section 3.1. The *likelihood ratio test statistic* is defined by

$$G^2 = -2 \sum_{i,j} f_{ij} \log \frac{\hat{m}_{ij}}{f_{ij}}. \tag{4.16}$$

For product multinomial sampling, we consider Tables 3.1 and 3.5 and the likelihood ratio test statistic is given by

$$G^2 = -2 \sum_{i,j} f_{ij} \log \frac{\hat{m}_{j|i}}{f_{ij}}. \tag{4.17}$$

For each type of sampling, under any one of the log-linear models in Chapter 3, the G^2 statistic has an asymptotic chi-square distribution with degrees of freedom equal to

$$df = IJ - (\text{number of free parameters}),$$

where I is the number of rows and J is the number of columns in the contingency table.

For a given log-linear model, the degrees of freedom for X^2 are the same as the degrees of freedom for G^2.

4.5 The Akaike Information Criterion

The different likelihoods of the two sampling schemes were discussed in Section 3.1. For the multinomial sampling scheme, the log-likelihood is

$$\ell(\boldsymbol{\theta}; \boldsymbol{f}) = \log N! - N \log N - \sum_{i,j} \log f_{ij}! + \sum_{i,j} f_{ij} \log m_{ij}(\boldsymbol{\theta}).$$

For the product multinomial sampling scheme, the log-likelihood is

$$\ell(\boldsymbol{\theta}; \boldsymbol{f}) = \log N_1! - N_1 \log N_1 - \sum_j \log f_{1j}! + \sum_j f_{1j} \log m_{j|i=1}(\boldsymbol{\theta})$$
$$+ \log N_2! - N_2 \log N_2 - \sum_j \log f_{2j}! + \sum_j f_{2j} \log m_{j|i=2}(\boldsymbol{\theta}).$$

For each sampling scheme, the *Akaike Information Criterion* of a log-linear model is defined by

$$\text{AIC} = 2k\text{–}2\ell(\hat{\boldsymbol{\theta}}; \boldsymbol{f}),$$

where k is the number of free parameters and $\hat{\boldsymbol{\theta}}$ is the estimate of parameter vector $\boldsymbol{\theta}$. When comparing log models using the AIC we prefer the model with the lowest AIC value.

4.6 Estimation of the expected counts

In this section, we use the statistical software **R** and the formulas in Sections 4.1 and 4.2 to estimate the expected counts, m_{ij} or $m_{j|i}$, and find the estimated asymptotic standard errors (ASE) of the logarithms of their estimates. We also estimate the log odds ratio using the theory in Section 4.3.

Asymptotic variance-covariance matrices for the estimates of the $\log \boldsymbol{m}$ under both sampling schemes are given in Appendix B.2.

4.6.1 Estimation of the parameters is the case of multinomial sampling

Consider the five standard hierarchical log-linear models of Section 3.3.1 under multinomial sampling. Tables 4.3–4.7 contain the estimates of the expected counts and their natural logarithms. The tables also include the estimates of the asymptotic standard errors (ASE) of the estimates of the logarithms of the counts.

Below each table, we give the estimate of the log odds ratio and the estimate of the asymptotic standard error of the estimate. It is non-zero only for the saturated model (see Table 4.7).

It turns out that the five non-standard log-linear models in Section 3.3.2 (the submodels of the log-linear version of Worcester's multiplicative model from Table 3.6) are equivalent to the previous five standard hierarchical log-linear models. For contingency tables that cross-classify three or more categorical variables, this is not necessarily true (see Hasan [10]).

Table 4.3: Estimation of the expected counts for the log-linear models $\log m_{ij} = \lambda$ and $\log m_{ij} = w$ (multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{11} = 43.5$ $(\log \hat{m}_{11} = 3.7728)$ (ASE = 0)	$\hat{m}_{12} = 43.5$ $(\log \hat{m}_{12} = 3.7728)$ (ASE = 0)	87
Control	$\hat{m}_{21} = 43.5$ $(\log \hat{m}_{21} = 3.7728)$ (ASE = 0)	$\hat{m}_{22} = 43.5$ $(\log \hat{m}_{22} = 3.7728)$ (ASE = 0)	87
Total	87	87	$N = 174$

$(\log \hat{X}_{AB} = 0, \text{ASE} = 0)$

Table 4.4: Estimation of the expected counts for the log-linear models $\log m_{ij} = \lambda + \lambda_i^A$ and $\log m_{ij} = w + \delta_i w_A$ (multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{11} = 29$ $(\log \hat{m}_{11} = 3.3673)$ (ASE = 0.1071)	$\hat{m}_{12} = 29$ $(\log \hat{m}_{12} = 3.3673)$ (ASE = 0.1071)	58
Control	$\hat{m}_{21} = 58$ $(\log \hat{m}_{21} = 4.0604)$ (ASE = 0.0536)	$\hat{m}_{22} = 58$ $(\log \hat{m}_{22} = 4.0604)$ (ASE = 0.0536)	116
Total	87	87	$N = 174$

$(\log \hat{X}_{AB} = 0, \text{ASE} = 0)$

Table 4.5: Estimation of the expected counts for the log-linear models $\log m_{ij} = \lambda + \lambda_j^B$ and $\log m_{ij} = w + \delta_j w_B$ (multinomial sampling)

Diagnostic group (A)	Oral contraceptive (B)		Total
	Used	Did not use	
Thromboembolism	$\hat{m}_{11} = 18$ $(\log \hat{m}_{11} = 2.8904)$ (ASE = 0.1484)	$\hat{m}_{12} = 69$ $(\log \hat{m}_{12} = 4.2341)$ (ASE = 0.0387)	87
Control	$\hat{m}_{21} = 18$ $(\log \hat{m}_{21} = 2.8904)$ (ASE = 0.1484)	$\hat{m}_{22} = 69$ $(\log \hat{m}_{22} = 4.2341)$ (ASE = 0.0387)	87
Total	36	138	$N = 174$

$(\log \hat{X}_{AB} = 0, \text{ASE} = 0)$

Table 4.6: Estimation of the expected counts for the log-linear models $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$ and $\log m_{ij} = w + \delta_i w_A + \delta_j w_B$ (multinomial sampling)

Diagnostic group (A)	Oral contraceptive (B)		Total
	Used	Did not use	
Thromboembolism	$\hat{m}_{11} = 12$ $(\log \hat{m}_{11} = 2.4849)$ (ASE = 0.1830)	$\hat{m}_{12} = 46$ $(\log \hat{m}_{12} = 3.8286)$ (ASE = 0.1140)	58
Control	$\hat{m}_{21} = 24$ $(\log \hat{m}_{21} = 3.1781)$ (ASE = 0.1578)	$\hat{m}_{22} = 92$ $(\log \hat{m}_{22} = 4.5218)$ (ASE = 0.0661)	116
Total	36	138	$N = 174$

$(\log \hat{X}_{AB} = 0, \text{ASE} = 0)$

Table 4.7: Estimation of the parameters for the saturated log-linear models $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$ and $\log m_{ij} = w + \delta_i w_A + \delta_j w_B + \delta_{ij} w_{AB}$ (multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{11} = 26$ $(\log \hat{m}_{11} = 3.2581)$ (ASE = 0.1809)	$\hat{m}_{12} = 32$ $(\log \hat{m}_{12} = 3.4657)$ (ASE = 0.1597)	58
Control	$\hat{m}_{21} = 10$ $(\log \hat{m}_{21} = 2.3026)$ (ASE = 0.3070)	$\hat{m}_{22} = 106$ $(\log \hat{m}_{22} = 4.6634)$ (ASE = 0.0607)	116
Total	36	138	$N = 174$

$(\log \hat{X}_{AB} = 2.1532, \text{ASE} = 0.4233)$

Consider the three standard non-hierarchical log-linear models of Section 3.3.1.2 under multinomial sampling. (As mentioned before, due to time constraints, we do not consider all the standard non-hierarchical log-linear models in this thesis.) Tables 4.8–4.10 contain the estimates of the expected counts and their natural logarithms. The tables also include the estimates of the asymptotic standard errors (ASE) of the estimates of the logarithms of the counts.

Below each table, we give the estimate of the log odds ratio and the estimate of the asymptotic standard error of the estimate. We comment about the value of the estimated log odds ratio only for the best models that we choose in Section 4.7.

Table 4.8: Estimation of the expected counts for the non-hierarchical log-linear model $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ with $\sum_i \lambda_{ij}^{AB} = 0$ for each j and $\sum_j \lambda_{ij}^{AB} = 0$ for each i (multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{11} = 66$ $(\log \hat{m}_{11} = 4.1897)$ (ASE = 0.0428)	$\hat{m}_{12} = 21$ $(\log \hat{m}_{12} = 3.0445)$ (ASE = 0.1344)	87
Control	$\hat{m}_{21} = 21$ $(\log \hat{m}_{21} = 3.0445)$ (ASE = 0.1344)	$\hat{m}_{22} = 66$ $(\log \hat{m}_{22} = 4.1897)$ (ASE = 0.0428)	87
Total	87	87	$N = 174$

$(\log \hat{X}_{AB} = 2.2904, \text{ASE} = 0.3542)$

Table 4.9: Estimation of the expected counts for the non-hierarchical log-linear model $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ with $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$ (multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{11} = 22.67$ $(\log \hat{m}_{11} = 3.1210)$ (ASE = 0.0946)	$\hat{m}_{12} = 22.67$ $(\log \hat{m}_{12} = 3.1210)$ (ASE = 0.0946)	45.33
Control	$\hat{m}_{21} = 22.67$ $(\log \hat{m}_{21} = 3.1210)$ (ASE = 0.0946)	$\hat{m}_{22} = 106$ $(\log \hat{m}_{22} = 4.6634)$ (ASE = 0.0607)	128.67
Total	45.33	128.67	$N = 174$

$(\log \hat{X}_{AB} = 1.5424, \text{ASE} = 0.1552)$

Table 4.10: Estimation of the expected counts for the non-hierarchical log-linear model $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ with $\lambda_{i2}^{AB} = 0$ for each i, and $\lambda_{21}^{AB} = 0$ (multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{11} = 26$ $(\log \hat{m}_{11} = 3.2581)$ (ASE = 0.1809)	$\hat{m}_{12} = 49.33$ $(\log \hat{m}_{12} = 3.8985)$ (ASE = 0.0318)	75.33
Control	$\hat{m}_{21} = 49.33$ $(\log \hat{m}_{21} = 3.8985)$ (ASE = 0.0318)	$\hat{m}_{22} = 49.33$ $(\log \hat{m}_{22} = 3.8985)$ (ASE = 0.0318)	98.67
Total	75.33	98.67	$N = 174$

$(\log \hat{X}_{AB} = -.6404, \text{ASE} = 0.2126)$

4.6.2 Estimation of the parameters is the case of product multinomial sampling

Consider the three standard hierarchical log-linear models of Section 3.4.1 under product multinomial sampling. Tables 4.11–4.13 contain the estimates of the expected counts and their natural logarithms. The tables also include the estimates of the asymptotic standard errors (ASE) of the estimates of the logarithms of the counts.

It turns out that the three non-standard log-linear models in Section 3.4.2 (the submodels of the log-linear version of Worcester's multiplicative model from Table 3.7) are equivalent to the previous three standard hierarchical log-linear models.

Table 4.11: Estimation of the expected counts for the log-linear models $\log m_{j|i} = \lambda_i^A$ and $\log m_{j|i} = v_i$ (product multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{1\|1} = 29$ $(\log \hat{m}_{1\|1} = 3.3673)$ (ASE = 0)	$\hat{m}_{2\|1} = 29$ $(\log \hat{m}_{2\|1} = 3.3673)$ (ASE = 0)	$N_1 = 58$
Control	$\hat{m}_{1\|2} = 58$ $(\log \hat{m}_{1\|2} = 4.0604)$ (ASE = 0)	$\hat{m}_{2\|2} = 58$ $(\log \hat{m}_{2\|2} = 4.0604)$ (ASE = 0)	$N_2 = 116$
Total	87	87	174

$(\log \hat{X}_{AB} = 0, \text{ASE} = 0)$

Table 4.12: Estimation of the expected counts for the log-linear models $\log m_{j|i} = \lambda_i^A + \lambda_j^B$ and $\log m_{j|i} = v_i + \delta_j v_B$ (product multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{1\|1} = 12$ $(\log \hat{m}_{1\|1} = 2.4849)$ (ASE = 0.1484)	$\hat{m}_{2\|1} = 46$ $(\log \hat{m}_{2\|1} = 3.8286)$ (ASE = 0.0387)	$N_1 = 58$
Control	$\hat{m}_{1\|2} = 24$ $(\log \hat{m}_{1\|2} = 3.1781)$ (ASE = 0.1484)	$\hat{m}_{2\|2} = 92$ $(\log \hat{m}_{2\|2} = 4.5218)$ (ASE = 0.0387)	$N_2 = 116$
Total	36	138	174

$(\log \hat{X}_{AB} = 0, \text{ASE} = 0)$

Table 4.13: Estimation of the parameters for the saturated log-linear models $\log m_{j|i} = \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$ and $\log m_{j|i} = v_i + \delta_j v_B + \delta_{j|i} v_{AB}$ (product multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{1\|1} = 26$ $(\log \hat{m}_{1\|1} = 3.2581)$ (ASE = 0.1457)	$\hat{m}_{2\|1} = 32$ $(\log \hat{m}_{2\|1} = 3.4657)$ (ASE = 0.1184)	$N_1 = 58$
Control	$\hat{m}_{1\|2} = 10$ $(\log \hat{m}_{1\|2} = 2.3026)$ (ASE = 0.3023)	$\hat{m}_{2\|2} = 106$ $(\log \hat{m}_{2\|2} = 4.6634)$ (ASE = 0.0285)	$N_2 = 116$
Total	36	138	174

$(\log \hat{X}_{AB} = 2.1532, \text{ASE} = 0.4233)$

Consider the three standard non-hierarchical log-linear models of Section 3.4.1.2 under product multinomial sampling. (As mentioned before, due to time constraints, we do not consider all the standard non-hierarchical log-linear models in this thesis.) Tables 4.14–4.16 contain the estimates of the expected counts and their natural logarithms. The tables also include the estimates of the asymptotic standard errors (ASE) of the estimates of the logarithms of the counts.

Below each table, we give the estimate of the log odds ratio and the estimate of the asymptotic standard error of the estimate. We comment about the value of the estimated log odds ratio only for the best models that we choose in Section 4.7.

Table 4.14: Estimation of the expected counts for the non-hierarchical log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ with $\sum_i \lambda_{ij}^{AB} = 0$ for each j and $\sum_j \lambda_{ij}^{AB} = 0$ for each i (product multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{1\|1} = 44$ $(\log \hat{m}_{1\|1} = 3.7842)$ (ASE = 0.0428)	$\hat{m}_{2\|1} = 14$ $(\log \hat{m}_{2\|1} = 2.6391)$ (ASE = 0.1344)	58
Control	$\hat{m}_{1\|2} = 28$ $(\log \hat{m}_{1\|2} = 3.3322)$ (ASE = 0.1344)	$\hat{m}_{2\|2} = 88$ $(\log \hat{m}_{2\|2} = 4.4773)$ (ASE = 0.0428)	116
Total	72	102	$N = 174$

$(\log \hat{X}_{AB} = 2.2902, \text{ASE} = 0.3545)$

Table 4.15: Estimation of the expected counts for the non-hierarchical log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ with $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$ (product multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{1\|1} = 29$ $(\log \hat{m}_{1\|1} = 3.3672)$ (ASE = 0.0000)	$\hat{m}_{2\|1} = 29$ $(\log \hat{m}_{2\|1} = 3.3672)$ (ASE = 0.0000)	58
Control	$\hat{m}_{1\|2} = 10$ $(\log \hat{m}_{1\|2} = 2.3025)$ (ASE = 0.3022)	$\hat{m}_{2\|2} = 106$ $(\log \hat{m}_{2\|2} = 4.6634)$ (ASE = 0.0285)	116
Total	39	135	$N = 174$

$(\log \hat{X}_{AB} = 2.3608, \text{ASE} = 0.3307)$

Table 4.16: Estimation of the expected counts for the non-hierarchical log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ with $\lambda_{i2}^{AB} = 0$ for each i, and $\lambda_{21}^{AB} = 0$ (product multinomial sampling)

	Oral contraceptive (B)		
Diagnostic group (A)	Used	Did not use	Total
Thromboembolism	$\hat{m}_{1\|1} = 26$	$\hat{m}_{2\|1} = 32$	58
	$(\log \hat{m}_{1\|1} = 3.2580)$	$(\log \hat{m}_{2\|1} = 3.4657)$	
	(ASE = 0.1456)	(ASE = 0.1184)	
Control	$\hat{m}_{1\|2} = 58$	$\hat{m}_{2\|2} = 58$	116
	$(\log \hat{m}_{1\|2} = 4.0604)$	$(\log \hat{m}_{2\|2} = 4.0604)$	
	(ASE = 0.0000)	(ASE = 0.0000)	
Total	84	90	$N = 174$

$(\log \hat{X}_{AB} = -0.2077, \text{ASE} = 0.2638)$

4.7 Comparison of the different log-linear models

Table 4.17 contains the results of the Pearson X^2 test and the likelihood ratio G^2 test for all the models we examined in Section 4.6 under multinomial sampling. In each case, we reject the model in favor of the saturated model $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$.

Table 4.18 gives the AICs of all the models under multinomial sampling. Clearly, the saturated model has the smallest AIC and it is thus the best model for fitting the data. [2] (The AIC from **R** corresponds to Poisson sampling and it differs from the actual AIC under multinomial sampling by a constant.)

From Table 4.7, we see that, for the saturated model, $\log \hat{X}_{AB} = 2.1532$ and $\text{ASE}(\log \hat{\text{X}}_{\text{AB}}) = 0.1791$. Thus, an asymptotic 95% confidence interval for the true log odds ratio, $\log X_{AB}$, is

$$2.1532 \pm 1.96 \times 0.1791 = (1.8021, 2.5042).$$

Since the asymptotic 95% confidence interval does not contain 0, the $\log X_{AB}$ is positive with 95% confidence, and thromboembolism and oral contraceptives are positively associated.

[2] There is a standard non-hierarchical model with a smaller AIC, but we do not consider it in this thesis.

Table 4.17: Goodness-of-fit statistics for the unconditional models [a b c]

Model	Df	Pearson X^2 test		G^2 test	
		X^2 statistic	p-value	G^2 statistic	p-value
$\log m_{ij} = \lambda$	3	125.68	0	113.01	0
$\log m_{ij} = \lambda + \lambda_i^A$	2	80.07	0	93.3	0
$\log m_{ij} = \lambda + \lambda_j^B$	2	46.79	0	49.21	0
$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$	1	30.89	0	29.5	0
$\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (R_1)	2	60.01	0	64.12	0
$\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (R_2)	2	11.41	0	12.84	0
$\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (R_3)	2	102.54	0	102.54	0
$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$	0	0	NA	0	NA

[a] R_1: $\sum_i \lambda_{ij}^{AB} = 0$ for each j and $\sum_j \lambda_{ij}^{AB} = 0$ for each i.
[b] R_2: $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$.
[c] R_3: $\lambda_{i2}^{AB} = 0$ for each i, and λ_{21}^{AB}=0.

Table 4.18: AIC for the unconditional models [a b c]

Model	AIC (from **R**)	AIC (actual)
$\log m_{ij} = \lambda$	136.0793	129.0814
$\log m_{ij} = \lambda + \lambda_i^A$	118.3710	111.3731
$\log m_{ij} = \lambda + \lambda_j^B$	74.2800	67.2821
$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$	56.5717	49.5738
$\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (R_1)	89.1914	82.1935
$\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (R_2)	37.90926	30.8914
$\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (R_3)	127.5935	120.6156
$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$	29.0712	22.0733

[a] R_1: $\sum_i \lambda_{ij}^{AB} = 0$ for each j and $\sum_j \lambda_{ij}^{AB} = 0$ for each i.
[b] R_2: $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$.
[c] R_3: $\lambda_{i2}^{AB} = 0$ for each i, and λ_{21}^{AB}=0.

We get similar conclusions by looking at the goodness-of-fit statistics and the AIC for the conditional models (the ones under the product multinomial sampling). See Tables 4.19 – 4.20. The saturated model $\log m_{j|i} = \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$ is the best model. (The estimate of $\log X_{AB}$ and the ASE of the estimate of $\log X_{AB}$ in Table 4.13 are the same as those in Table 4.7.)

Table 4.19: Goodness-of-fit statistics for conditional models [a b c]

Model	Df	Pearson X^2 test		G^2 test	
		X^2 statistic	p-value	G^2 statistic	p-value
$\log m_{j\|i} = \lambda_i^A$	2	125.68	0	113.01	0
$\log m_{j\|i} = \lambda_i^A + \lambda_j^B$	1	30.89	0	29.5	0
$\log m_{j\|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (R_1)	1	45.76	0	44.41	0
$\log m_{j\|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (R_2)	1	0.62	0.43	0.62	0.43
$\log m_{j\|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (R_3)	1	79.45	0	92.68	0
$\log m_{j\|i} = \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$	0	0	NA	0	NA

[a] R_1: $\sum_i \lambda_{ij}^{AB} = 0$ for each j and $\sum_j \lambda_{ij}^{AB} = 0$ for each i.
[b] R_2: $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$.
[c] R_3: $\lambda_{i2}^{AB} = 0$ for each i, and λ_{21}^{AB}=0.

Table 4.20: AIC for the conditional models [a b c]

Model	AIC (from **R**)	AIC (actual)
$\log m_{j\|i} = \lambda_i^A$	118.3710	105.8769
$\log m_{j\|i} = \lambda_i^A + \lambda_j^B$	56.5717	44.0776
$\log m_{j\|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (R_1)	71.4831	58.9890
$\log m_{j\|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (R_2)	27.6930	15.1989
$\log m_{j\|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (R_3)	119.7492	107.2550
$\log m_{j\|i} = \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$	29.0712	16.5771

[a] R_1: $\sum_i \lambda_{ij}^{AB} = 0$ for each j and $\sum_j \lambda_{ij}^{AB} = 0$ for each i.
[b] R_2: $\lambda_{1j}^{AB} = 0$ for each j, and $\lambda_{21}^{AB} = 0$.
[c] R_3: $\lambda_{i2}^{AB} = 0$ for each i, and λ_{21}^{AB}=0.

There is one exception to what we said above. From the three standard non-hierarchical models (under product multinomial sampling) that we considered in this thesis, the one with the equation $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ and restriction (R_2) has an AIC smaller than the AIC of the saturated model. In addition, using both the Pearson X^2 test and the G^2 likelihood ratio test, we fail to reject the null model in favor of the alternative model.

Choosing the above standard non-hierarchical log-linear (under product multinomial sampling) as the best model might be controversial because such models are difficult to interpret.

APPENDIX A

APPENDIX TO CHAPTER 3

A.1 Appendix to Section 3.3

In this appendix, we prove that the standard hierarchical log-linear model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B \tag{A.1}$$

is equivalent to the independence model

$$\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}, \quad \text{for all} \quad i, j, \tag{A.2}$$

under the assumption of multinomial sampling. See Table 3.2 in Section 3.1 and the theory of Section 3.3.

(a) (A.1) $\Rightarrow$ (A.2). Assume Eq. (A.1) holds. Since $m_{ij} = N\pi_{ij}$, we get

$$\begin{aligned} N\pi_{ij} &= e^{\lambda} e^{\lambda_i^A} e^{\lambda_j^B} \\ \Rightarrow \pi_{ij} &= \frac{1}{N} e^{\lambda} e^{\lambda_i^A} e^{\lambda_j^B} \\ \Rightarrow \pi_{ij} &= g(i)h(j), \end{aligned}$$

where $g(i) = \frac{1}{N} e^{\lambda} e^{\lambda_i^A}$ and $h(j) = e^{\lambda_j^B}$. By Lemma 4.2.7 in Casella and Berger [3], we conclude that factors A and B are independent, and thus Eq. (A.2) holds.
(b) (A.2) $\Rightarrow$ (A.1). Assume Eq. (A.2) holds. Since $m_{ij} = N\pi_{ij}$, we have

$$m_{ij} = N\pi_{i\cdot}\pi_{\cdot j} \Rightarrow \log m_{ij} = \log N + \log \pi_{i\cdot} + \log \pi_{\cdot j}$$

We may not write $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$ with $\lambda = \log N$, $\lambda_i^A = \log \pi_{i\cdot}$ and $\lambda_j^B = \log \pi_{\cdot j}$ because we need one of the three sets of restrictions of Section 3.3.1 to hold.

i. To get the zero-sum constraints, we let

$$\begin{aligned}\lambda &= \log N(\sqrt{\pi_{1\cdot}\pi_{2\cdot}\pi_{\cdot 1}\pi_{\cdot 2}}),\\ \lambda_i^A &= \log \frac{\pi_{i\cdot}}{\sqrt{\pi_{1\cdot}\pi_{2\cdot}}}, \quad \textit{and}\\ \lambda_j^B &= \log \frac{\pi_{\cdot j}}{\sqrt{\pi_{\cdot 1}\pi_{\cdot 2}}}.\end{aligned}$$

Then Eq. (A.1) holds with $\lambda_1^A + \lambda_2^A = 0$ and $\lambda_1^B + \lambda_2^B = 0$.

ii. To get zero constraints for the first categories (**R** constraints), we let

$$\begin{aligned}\lambda_1^A &= \lambda_1^B = 0,\\ \lambda &= \log(N\pi_{1\cdot}\pi_{\cdot 1}),\\ \lambda_2^A &= \log \frac{\pi_{2\cdot}}{\pi_{1\cdot}},\\ \lambda_2^B &= \log \frac{\pi_{\cdot 2}}{\pi_{\cdot 1}}.\end{aligned}$$

Then Eq. (A.1) holds with $\lambda_1^A = \lambda_1^B = 0$.

iii. To get zero constraints for the last categories (**SAS** constraints), we let

$$\begin{aligned}\lambda_2^A &= \lambda_2^B = 0,\\ \lambda &= \log(N\pi_{2\cdot}\pi_{\cdot 2}),\\ \lambda_1^A &= \log \frac{\pi_{1\cdot}}{\pi_{2\cdot}},\\ \lambda_1^B &= \log \frac{\pi_{\cdot 1}}{\pi_{\cdot 2}}.\end{aligned}$$

Then Eq. (A.1) holds with $\lambda_2^A = \lambda_2^B = 0$.

A.2 Appendix to Section 3.4

In this appendix, we prove that the standard hierarchical log-linear model

$$\log m_{j|i} = \lambda_i^A + \lambda_j^B, \tag{A.3}$$

which is suitable for product multinomial sampling, is equivalent to the independence model

$$\pi_{1|1} = \pi_{1|2}. \tag{A.4}$$

See Table 3.4 in Section 3.1 and the theory of Section 3.4 .

(a) (A.3) $\Rightarrow$ (A.4). Assume Eq. (A.3) holds. Since $m_{j|i} = N_i \pi_{j|i}$, we get that

$$\log m_{j|i} = \lambda_i^A + \lambda_j^B \Rightarrow \pi_{j|i} = \frac{1}{N_i} e^{\lambda_i^A} e^{\lambda_j^B}.$$

Now,

$$\frac{\pi_{1|1}\pi_{2|2}}{\pi_{2|1}\pi_{1|2}} = \frac{\frac{1}{N_1 N_2} e^{\lambda_1^A} e^{\lambda_1^B} e^{\lambda_2^A} e^{\lambda_2^B}}{\frac{1}{N_1 N_2} e^{\lambda_1^A} e^{\lambda_2^B} e^{\lambda_2^A} e^{\lambda_1^B}} = 1,$$

and so,

$$\frac{\pi_{1|1}}{\pi_{2|1}} = \frac{\pi_{1|2}}{\pi_{2|2}} \Rightarrow \frac{\pi_{1|1}}{\pi_{2|1} + \pi_{1|1}} = \frac{\pi_{1|2}}{\pi_{2|2} + \pi_{1|2}} \Rightarrow \pi_{1|1} = \pi_{1|2}.$$

Thus Eq. (A.4) holds.

(b) (A.4) $\Rightarrow$ (A.3). Assume Eq. (A.4) holds.

i. To get the zero-sum constraints, we let

$$\begin{aligned} \lambda_1^A &= \log(N_1 \pi_{1|2}^{1/2} \pi_{2|1}^{1/2}), \\ \lambda_2^A &= \log(N_2 \pi_{1|2}^{1/2} \pi_{2|2}^{1/2}), \quad \text{and} \\ \lambda_1^B &= -\lambda_2^B = \frac{1}{2} \log \frac{\pi_{1|2}}{\pi_{2|2}}. \end{aligned}$$

Since $m_{j|i} = N_i \pi_{j|i}$ and $\pi_{1|1} = \pi_{1|2}$, we may prove that $\log m_{j|i} = \lambda_i^A + \lambda_j^B$ and $\sum_j \lambda_j^B = 0$.

ii. To get zero constraints for the first categories (**R** constraints), we let

$$\begin{aligned} \lambda_1^A &= \log(N_1 \pi_{1|1}), \\ \lambda_2^A &= \log(N_2 \pi_{1|2}), \quad \text{and} \\ \lambda_2^B &= \log \frac{\pi_{2|1}}{\pi_{1|1}}. \end{aligned}$$

Since $m_{j|i} = N_i \pi_{j|i}$ and $\pi_{1|1} = \pi_{1|2}$, we may prove that $\log m_{j|i} = \lambda_i^A + \lambda_j^B$ and $\lambda_1^B = 0$.

iii. To get zero constraints for the last categories (**SAS** constraints), we let

$$\begin{aligned}\lambda_1^A &= \log(N_1\pi_{2|1}),\\ \lambda_2^A &= \log(N_2\pi_{2|2}), \quad \text{and}\\ \lambda_1^B &= \log\frac{\pi_{1|1}}{\pi_{2|1}}.\end{aligned}$$

Since $m_{j|i} = N_i\pi_{j|i}$ and $\pi_{1|1} = \pi_{1|2}$, we may prove that $\log m_{j|i} = \lambda_i^A + \lambda_j^B$ and $\lambda_2^B = 0$.

APPENDIX B

APPENDIX TO CHAPTER 4

B.1 Variance-covariance matrices for the estimates of the w-parameters in Section 3.3.2

To derive the results below, we use the theory and formulas of Chapter 4.

B.1.1 Variance-covariance matrices of the estimates of w for multinomial sampling

We list the variance-covariance matrices of $\hat{\boldsymbol{w}}$ under multinomial sampling for the unconditional log-linear models.

Model $\log m_{ij} = w$:

$$\hat{\boldsymbol{w}} = (\hat{w}) = \begin{pmatrix} 3.7728 \end{pmatrix} \quad \text{and} \quad \hat{V}(\hat{\boldsymbol{w}}) = \begin{pmatrix} 0 \end{pmatrix}$$

Model $\log m_{ij} = w + \delta_i w_A$:

$$\hat{\boldsymbol{w}} = \begin{pmatrix} \hat{w} \\ \hat{w}_A \end{pmatrix} = \begin{pmatrix} 4.0604 \\ -0.6931 \end{pmatrix} \quad \text{and} \quad \hat{V}(\hat{\boldsymbol{w}}) = \begin{pmatrix} 0.0029 & -0.0086 \\ -0.0086 & 0.0259 \end{pmatrix}$$

Model $\log m_{ij} = w + \delta_j w_B$:

$$\hat{\boldsymbol{w}} = \begin{pmatrix} \hat{w} \\ \hat{w}_B \end{pmatrix} = \begin{pmatrix} 4.2341 \\ -1.3438 \end{pmatrix} \quad \text{and} \quad \hat{V}(\hat{\boldsymbol{w}}) = \begin{pmatrix} 0.0015 & -0.0072 \\ -0.0072 & 0.0350 \end{pmatrix}$$

Model $\log m_{ij} = w + \delta_i w_A + \delta_j w_B$:

$$\hat{\boldsymbol{w}} = \begin{pmatrix} \hat{w} \\ \hat{w}_A \\ \hat{w}_B \end{pmatrix} = \begin{pmatrix} 4.5218 \\ -.6932 \\ -1.3437 \end{pmatrix} \quad \text{and} \quad \hat{V}(\hat{\boldsymbol{w}}) = \begin{pmatrix} 0.0044 & -0.0086 & -.0072 \\ -0.0086 & 0.0259 & -0.0000 \\ -0.0072 & 0.0000 & .0350 \end{pmatrix}$$

Model $\log m_{ij} = w + \delta_i w_A + \delta_j w_B + \delta_{ij} w_{AB}$:

$$\hat{\boldsymbol{w}} = \begin{pmatrix} \hat{w} \\ \hat{w}_A \\ \hat{w}_B \\ \hat{w}_{AB} \end{pmatrix} = \begin{pmatrix} 4.6634 \\ -1.1977 \\ -2.3608 \\ 2.1532 \end{pmatrix} \text{ and } \hat{V}(\hat{\boldsymbol{w}}) = \begin{pmatrix} 0.0037 & -0.0094 & -0.0094 & 0.0094 \\ -0.0094 & 0.0407 & 0.0094 & -0.0407 \\ -0.0094 & 0.0094 & 0.1094 & -0.1094 \\ 0.0094 & -0.0407 & -0.1094 & 0.1791 \end{pmatrix}$$

B.2 Appendix to Section 4.6

B.2.1 Variance-covariance matrices of the estimates of $\log m_{ij}$ for multinomial sampling

Here $\hat{\boldsymbol{m}} = (m_{11}, m_{12}, m_{21}, m_{22})^\top$. We list the variance-covariance matrices of $\hat{\boldsymbol{m}}$ under multinomial sampling for the unconditional log-linear models in Section 3.3.

Log-linear model $\log m_{ij} = \lambda$ (multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Log-linear model $\log m_{ij} = \lambda + \lambda_i^A$ (multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0115 & 0.0115 & -0.0058 & -0.0058 \\ 0.0115 & 0.0115 & -0.0058 & -0.0058 \\ -0.0058 & -0.0058 & 0.0029 & 0.0029 \\ -0.0058 & -0.0058 & 0.0029 & 0.0029 \end{pmatrix}$$

Log-linear model $\log m_{ij} = \lambda + \lambda_j^B$ (multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0220 & -0.0058 & 0.0220 & -0.0058 \\ -0.0058 & 0.0015 & -0.0058 & 0.0015 \\ 0.0220 & -0.0058 & 0.0220 & -0.0058 \\ -0.0058 & 0.0015 & -0.0058 & 0.0015 \end{pmatrix}$$

Log-linear model $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B$ (multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0335 & 0.0058 & 0.0163 & -0.0115 \\ 0.0058 & 0.0130 & -0.0115 & -0.0043 \\ 0.0163 & -0.0115 & 0.0249 & -0.0029 \\ -0.0115 & -0.0043 & -0.0029 & 0.0044 \end{pmatrix}$$

Log-linear model $\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$ (multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0327 & -0.0058 & -0.0058 & -0.0058 \\ -0.0058 & 0.0255 & -0.0058 & -0.0058 \\ -0.0058 & -0.0058 & 0.0943 & -0.0058 \\ -0.0058 & -0.0058 & -0.0058 & 0.0037 \end{pmatrix}$$

Log-linear model $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (zero sum constraint, multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0018 & -0.0057 & -0.0057 & 0.0018 \\ -0.0057 & 0.0181 & 0.0181 & -0.0057 \\ -0.0057 & 0.0181 & 0.0181 & -0.0057 \\ 0.0018 & -0.0057 & -0.0057 & 0.0018 \end{pmatrix}$$

Log-linear model $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (first category zero constraint, multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0090 & 0.0090 & 0.0090 & -0.0057 \\ 0.0090 & 0.0090 & 0.0090 & -0.0057 \\ 0.0090 & 0.0090 & 0.0090 & -0.0057 \\ -0.0057 & -0.0057 & -0.0057 & 0.0036 \end{pmatrix}$$

Log-linear model $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ (last category zero constraint, multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0327 & -0.0057 & -0.0057 & -0.0057 \\ -0.0057 & 0.0010 & 0.0010 & 0.0010 \\ -0.0057 & 0.0010 & 0.0010 & 0.0010 \\ -0.0057 & 0.0010 & 0.0010 & 0.0010 \end{pmatrix}$$

B.2.2 Variance-covariance matrices of the estimates of $\log m_{ij}$ for product multinomial sampling

Here $\hat{\boldsymbol{m}} = (m_{1|1}, m_{2|1}, m_{1|2}, m_{2|2})^\top$. We list the variance-covariance matrices of $\hat{\boldsymbol{m}}$ under product multinomial sampling for the conditional log-linear models in Section 3.4.

Log-linear model $\log m_{j|i} = \lambda_i^A$ (product multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_j^B$ (product multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0220 & -0.0058 & 0.0220 & -0.0058 \\ -0.0058 & 0.0015 & -0.0058 & 0.0015 \\ 0.0220 & -0.0058 & 0.0220 & -0.0058 \\ -0.0058 & 0.0015 & -0.0058 & 0.0015 \end{pmatrix}$$

Log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$ (product multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0212 & -0.0172 & 0.0000 & 0.0000 \\ -0.0172 & 0.0140 & 0.0000 & 0.0000 \\ 0.0000 & -0.0000 & 0.0914 & -0.0086 \\ 0.0000 & -0.0000 & -0.0086 & 0.0008 \end{pmatrix}$$

Log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (zero sum constraint, product multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0018 & -0.0057 & -0.0057 & 0.0018 \\ -0.0057 & 0.0180 & 0.0180 & -0.0057 \\ -0.0057 & 0.0180 & 0.0181 & -0.0057 \\ 0.0018 & -0.0057 & -0.0057 & 0.0018 \end{pmatrix}$$

Log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (first category zero constraint, product multinomial sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0 & 0 & 0.0000 & 0.0000 \\ 0 & 0 & 0.0000 & 0.0000 \\ 0 & 0 & 0.0914 & -0.0086 \\ 0 & 0 & -0.0086 & 0.0008 \end{pmatrix}$$

Log-linear model $\log m_{j|i} = \lambda_i^A + \lambda_{ij}^{AB}$ (last category zero constraint, product multinomial

sampling):

$$\hat{V}(\log \hat{\boldsymbol{m}}) = \begin{pmatrix} 0.0212 & -0.0172 & 0 & 0 \\ -0.0172 & 0.0140 & 0 & 0 \\ 0.0000 & 0.0000 & 0 & 0 \\ 0.0000 & 0.0000 & 0 & 0 \end{pmatrix}$$

B.3 R programs for some of the calculations in Section 4.6

In this part of the appendix, we give some **R** programs to estimate the parameters of the log-linear models in Section 3.3. We use three kinds of log-linear models. The results are summarized in Section 4.6. For brevity, we only include **R** programs for multinomial sampling.

(a) Standard hierarchical log-linear models (with zero constraints for the first categories)

```
## lambda
freq <- c(26,32,10,106)
m<-data.frame(diagonostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagonostic<-relevel(diagonostic, ref = "1"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "1"))
m
##   diagnostic contraceptive
##   1          1             1
##   2          1             2
##   3          2             1
##   4          2             2
X <- model.matrix( ~ diagonostic*contraceptive,
data = m, contrasts.arg = list(diagonostic = "contr.sum",
contraceptive="contr.sum"))
X<-X[,-c(2:4)]
X
## 1 2 3 4
## 1 1 1 1
```

```
fit_a <- glm(freq~X, family = poisson)
summary(fit_a)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
##      1       2       3       4
## -2.870  -1.830  -6.132   7.989

## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.77276    0.07581   49.77   <2e-16 ***
## X                 NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 113.01  on 3  degrees of freedom
## Residual deviance: 113.01  on 3  degrees of freedom
## AIC: 136.08

## Number of Fisher Scoring iterations: 5
fit_a$fitted.values
##    1    2    3    4
## 43.5 43.5 43.5 43.5
-----------------------------------------------------
## lambda+lambda(A)
freq <- c(26,32,10,106)
m<-data.frame(diagonostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagonostic<-relevel(diagonostic, ref = "1"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "1"))
m
```

```
##   diagonostic contraceptive
## 1           1             1
## 2           1             2
## 3           2             1
## 4           2             2
X <- model.matrix( ~ diagonostic*contraceptive,
data = m, contrasts.arg = list(diagonostic = "contr.sum",
contraceptive="contr.sum"))
X<-X[,-1]
X<-X[,-c(2,3)]
X
##  1  2  3  4
##  1  1 -1 -1
fit_a <- glm(freq~X, family = poisson)
##summary(fit_a)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
##       1        2        3        4
## -0.5671   0.5479  -7.8002   5.6423
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.71387    0.08041   46.19  < 2e-16 ***
## X           -0.34657    0.08041   -4.31 1.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 113.01  on 3  degrees of freedom
## Residual deviance:  93.30  on 2  degrees of freedom
```

```
## AIC: 118.37
##
## Number of Fisher Scoring iterations: 5
fit_a$fitted.values
##  1  2  3  4
## 29 29 58 58
##-----------------------------------------------------
## lambda+lambda(B)
freq <- c(26,32,10,106)
m<-data.frame(diagonostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagonostic<-relevel(diagonostic, ref = "1"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "1"))
m
##   diagonostic contraceptive
## 1           1             1
## 2           1             2
## 3           2             1
## 4           2             2
X <- model.matrix( ~ diagonostic*contraceptive,
data = m, contrasts.arg = list(diagonostic = "contr.sum",
contraceptive="contr.sum"))
X<-X[,-c(1,2,4)]
X
##  1  2  3  4
##  1 -1  1 -1
fit_a <- glm(freq~X, family = poisson)
summary(fit_a)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
##      1       2       3       4
##  1.767  -4.982  -2.060   4.125
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.56224    0.09357   38.07  < 2e-16 ***
## X           -0.67187    0.09357   -7.18 6.96e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 113.008  on 3  degrees of freedom
## Residual deviance:  49.209  on 2  degrees of freedom
## AIC: 74.28
##
## Number of Fisher Scoring iterations: 4
fit_a$fitted.values
##  1  2  3  4
## 18 69 18 69
##------------------------------------------------------------
## lambda+lambda(A)+lambda(B)
freq <- c(26,32,10,106)
m<-data.frame(diagonostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagonostic<-relevel(diagonostic, ref = "1"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "1"))
m
##   diagonostic contraceptive
## 1           1             1
## 2           1             2
## 3           2             1
## 4           2             2
X <- model.matrix( ~ diagonostic*contraceptive,
data = m, contrasts.arg = list(diagonostic = "contr.sum",
contraceptive="contr.sum"))
X<-X[,-c(1,4)]
```

```
X
##   diagonostic1 contraceptive1
## 1            1              1
## 2            1             -1
## 3           -1              1
## 4           -1             -1
fit_b <- glm(freq~X, family = poisson)
summary(fit_b)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
##      1       2       3       4
##  3.494  -2.185  -3.239   1.425
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      3.50335    0.09734   35.99  < 2e-16 ***
## Xdiagonostic1   -0.34657    0.08041   -4.31 1.63e-05 ***
## Xcontraceptive1 -0.67187    0.09357   -7.18 6.97e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 113.01  on 3  degrees of freedom
## Residual deviance:  29.50  on 1  degrees of freedom
## AIC: 56.572
##
## Number of Fisher Scoring iterations: 5
fit_b$fitted.values
##  1  2  3  4
## 12 46 24 92
```

```
##-------------------------------------------------------
## lambda+lambda(A)+lambda(B)+lambda(AB)
freq <- c(26,32,10,106)
m<-data.frame(diagonostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagonostic<-relevel(diagonostic, ref = "1"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "1"))
m
##   diagonostic contraceptive
## 1           1             1
## 2           1             2
## 3           2             1
## 4           2             2
X <- model.matrix( ~ diagonostic*contraceptive,
data = m, contrasts.arg = list(diagonostic = "contr.sum",
contraceptive="contr.sum"))
X<-X[,-1]
X
##   diagonostic1 contraceptive1 diagonostic1:contraceptive1
## 1            1              1                           1
## 2            1             -1                          -1
## 3           -1              1                          -1
## 4           -1             -1                           1
fit_b <- glm(freq~X, family = poisson)
summary(fit_b)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
## [1]  0  0  0  0
##
## Coefficients:
##                             Estimate Std. Err. z val. Pr(>|z|)
## (Intercept)                  3.42246  0.10581  32.344  < 2e-16 ***
```

```
## Xdiagonostic1                -0.06055  0.10581  -0.572    0.567
## Xcontraceptive1              -0.64212  0.10581  -6.068 1.29e-09 ***
## Xdiagonostic1:contraceptive1  0.53830  0.10581   5.087 3.63e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance:  1.1301e+02  on 3  degrees of freedom
## Residual deviance: -7.5495e-15  on 0  degrees of freedom
## AIC: 29.071
##
## Number of Fisher Scoring iterations: 3
fit_b$fitted.values
##   1   2   3   4
##  26  32  10 106
##-------------------------------------------------------------
```

(b) Standard non-hierarchical log-linear models

For brevity, we only show some of the calculations for the standard non-hierarchical log-linear model $\log m_{ij} = \lambda + \lambda_{ij}^{AB}$ with the restrictions $\lambda_{1j}^{AB} = 0$ for each j and $\lambda_{21}^{AB} = 0$.

```
##lambda+lambda(AB)
freq <- c(26,32,10,106)
m<-data.frame(diagonostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagonostic<-relevel(diagnostic, ref = "1"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "1"))
m
## diagnostic contraceptive
##  1           1            1
##  2           1            2
##  3           2            1
##  4           2            2
X <- model.matrix( ~ diagonostic*contraceptive,
+                    data = m)
```

```
x_ab <- c(0,0,0,1)
X <- cbind(X,x_ab)
X<-X[,-c(1:4)]
X
## 1 2 3 4
## 0 0 0 1
fit_ab <- glm(freq~X, family = poisson)
summary(fit_ab)

## Call:
## glm(formula = freq ~ X, family = poisson)

## Deviance Residuals:
##      1        2        3        4
## 0.6839   1.8448  -2.9945   0.0000

## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.1209     0.1213  25.736   <2e-16 ***
## X             1.5425     0.1554   9.928   <2e-16 ***
## ---
## Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

## (Dispersion parameter for poisson family taken to be 1)

##    Null deviance: 113.008  on 3  degrees of freedom
## Residual deviance:  12.838  on 2  degrees of freedom
## AIC: 37.909

## Number of Fisher Scoring iterations: 4

fit_ab$fitted.values
##        1         2         3         4
```

```
##  22.66667  22.66667  22.66667 106.00000
##-----------------------------------------------------
```

(c) Non-standard log-linear models (log-linear versions of Worcester's models)

```
## w
freq <- c(26,32,10,106)
m<-data.frame(diagnostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagnostic<-relevel(diagnostic, ref = "2"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "2"))
m
##   diagnostic contraceptive
## 1          1             1
## 2          1             2
## 3          2             1
## 4          2             2
X<-model.matrix( ~ diagnostic*contraceptive, data = m)
X<-X[,-c(2:4)]
X
## 1 2 3 4
## 1 1 1 1
fit_b <- glm(freq~X, family = poisson)
summary(fit_b)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
##      1       2       3       4
## -2.870  -1.830  -6.132   7.989
##
## Coefficients: (1 not defined because of singularities)
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.77276    0.07581   49.77   <2e-16 ***
## X                 NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 113.01  on 3  degrees of freedom
## Residual deviance: 113.01  on 3  degrees of freedom
## AIC: 136.08
##
## Number of Fisher Scoring iterations: 5
fit_b$fitted.values
##    1    2    3    4
## 43.5 43.5 43.5 43.5
##-------------------------------------------------------------
## w+w(A)
freq <- c(26,32,10,106)
m<-data.frame(diagnostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagnostic<-relevel(diagnostic, ref = "2"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "2"))
m
##   diagnostic contraceptive
## 1          1             1
## 2          1             2
## 3          2             1
## 4          2             2
X<-model.matrix( ~ diagnostic*contraceptive, data = m)
X<-X[,-1]
X<-X[,-c(2,3)]
X
## 1 2 3 4
## 1 1 0 0
```

```
fit_a <- glm(freq~X, family = poisson)
summary(fit_a)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
##       1        2        3        4
## -0.5671   0.5479  -7.8002   5.6423
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.06044    0.09285   43.73  < 2e-16 ***
## X           -0.69315    0.16082   -4.31 1.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 113.01  on 3  degrees of freedom
## Residual deviance:  93.30  on 2  degrees of freedom
## AIC: 118.37
##
## Number of Fisher Scoring iterations: 5
fit_a$fitted.values
##  1  2  3  4
## 29 29 58 58
##------------------------------------------------------------
## w+w(B)
freq <- c(26,32,10,106)
m<-data.frame(diagnostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagnostic<-relevel(diagnostic, ref = "2"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "2"))
m
```

```
##   diagnostic contraceptive
## 1          1             1
## 2          1             2
## 3          2             1
## 4          2             2
X<-model.matrix( ~ diagnostic*contraceptive, data = m)
X<-X[,-c(1,2,4)]
X
## 1 2 3 4
## 1 0 1 0
fit_b <- glm(freq~X, family = poisson)
summary(fit_b)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
##      1       2       3       4
##  1.767  -4.982  -2.060   4.125
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.23411    0.08512   49.74  < 2e-16 ***
## X           -1.34373    0.18715   -7.18 6.96e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 113.008  on 3  degrees of freedom
## Residual deviance:  49.209  on 2  degrees of freedom
## AIC: 74.28
##
## Number of Fisher Scoring iterations: 4
```

```
fit_b$fitted.values
##  1  2  3  4
## 18 69 18 69
##_______________________________________________________
## w+w(A)+w(B)
freq <- c(26,32,10,106)
m<-data.frame(diagnostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagnostic<-relevel(diagnostic, ref = "2"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "2"))
m
##   diagnostic contraceptive
## 1          1             1
## 2          1             2
## 3          2             1
## 4          2             2
X<-model.matrix( ~ diagnostic*contraceptive, data = m)
X<-X[,-c(1,4)]
X
##   diagnostic1 contraceptive1
## 1           1              1
## 2           1              0
## 3           0              1
## 4           0              0
fit_b <- glm(freq~X, family = poisson)
summary(fit_b)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
##      1       2       3       4
##  3.494  -2.185  -3.239   1.425
##
## Coefficients:
```

```
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       4.5218     0.1006   44.95  < 2e-16 ***
## Xdiagnostic1     -0.6931     0.1608   -4.31 1.63e-05 ***
## Xcontraceptive1  -1.3437     0.1871   -7.18 6.97e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 113.01  on 3  degrees of freedom
## Residual deviance:  29.50  on 1  degrees of freedom
## AIC: 56.572
##
## Number of Fisher Scoring iterations: 5
fit_b$fitted.values
##  1  2  3  4
## 12 46 24 92
##_____________________________________________________
## w+w(A)+w(B)+w(AB)
freq <- c(26,32,10,106)
m<-data.frame(diagnostic=gl(2,2,4), contraceptive=gl(2,1,4))
m<-within(m, diagnostic<-relevel(diagnostic, ref = "2"))
m<-within(m, contraceptive<-relevel(contraceptive, ref = "2"))
m
##   diagnostic contraceptive
## 1          1             1
## 2          1             2
## 3          2             1
## 4          2             2
X<-model.matrix( ~ diagnostic*contraceptive, data = m)
X<-X[,-1]
X
##   diagnostic1 contraceptive1 diagnostic1:contraceptive1
## 1           1              1                          1
```

```
## 2           1              0                          0
## 3           0              1                          0
## 4           0              0                          0
fit_b <- glm(freq~X, family = poison)
##
##
summary(fit_b)
##
## Call:
## glm(formula = freq ~ X, family = poisson)
##
## Deviance Residuals:
## [1]  0  0  0  0
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  4.66344    0.09713  48.013  < 2e-16 ***
## Xdiagnostic1                -1.19770    0.20170  -5.938 2.89e-09 ***
## Xcontraceptive1             -2.36085    0.33081  -7.137 9.56e-13 ***
## Xdiagnostic1:contraceptive1  2.15321    0.42326   5.087 3.63e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance:  1.1301e+02  on 3  degrees of freedom
## Residual deviance: -9.3259e-15  on 0  degrees of freedom
## AIC: 29.071
##
## Number of Fisher Scoring iterations: 3
fit_b$fitted.values
##   1   2   3   4
##  26  32  10 106
##_____________________________________________________
```

BIBLIOGRAPHY

[1] Agresti, A. (2013), *Categorical Data Analysis (3rd ed.)*, Hoboken, New Jersey: John Wiley & Sons.

[2] Bishop, Y. M., Fienberg, S. E., and Holland, P.W. (2007), *Discrete Multivariate Analysis: Theory and Applications*, New York, New York: Springer.

[3] Casella, G., and Berger, R. L. (2001) *Statistical Inference (2nd ed.)*, Duxbury, USA.

[4] Christensen, R. (1997), *Log-Linear Models and Logistic Regression (2nd ed.)*, New York, NY, USA.

[5] Funo, E. (2002), Worcester's log-linear model: theory and practice, *Nature-People-Society: Science and the Humanities*, **33(7)**, 25–47

[6] Funo, E. (2004a), Worcester's log-linear model for three–dimensional contingency table, *Quarterly Journal of Economics (Society of Economics, Kanto Gakuin University)*, **221(10)**, 1–13.

[7] Funo, E. (2004b), Worcester's log-linear model for $2 \times 2 \times 3$ contingency table, *Nature-People-Society: Science and the Humanities*, **37(7)**, 41–50.

[8] Funo, E. (2006), Worcester's log-linear model for four or more dimensions, *Quarterly Journal of Economics (Society of Economics, Kanto Gakuin University)*, **227(4)**, 120–135.

[9] Funo, E. (2007), Note on the analysis of two dimensional contingency tables, *Kanto Gakuin University Institute of Economics and Management Annual Report)*, **29(3)**, 156–162.

[10] Hasan, M. N. (2023), *Synergy and Antagonism in Log-Linear Models*, Ph.D. dissertation, University of Nevada, Las Vegas.

[11] Vessey, M. P., and Doll, R. (1968), Investigation of relation between use of oral contraceptives and thromboembolic disease, *British Medical Journal*, **2(5599)**, 199–205.

[12] Vessey, M. P., and Doll, R. (1969), Investigation of relation between use of oral contraceptives and thromboembolic disease. A further report. *British Medical Journal*, **2(5658)**, 651–657.

[13] von Eye, A., and Mun, E. Y. (2013), *Log-linear Modeling: Concepts, Interpretation, and Application*, Hoboken, New Jersey: John Wiley & Sons.

[14] von Eye, A., Schuster C., and Rogers W. M. (1998), Modelling synergy using manifest categorical variables, *International Journal of Behavioral Development*, **22(3), 537-557**.

[15] Worcester, J. (1971), The relative odds in the 2^3 contingency table, *American Journal of Epidemiology*, **93(3)**, 145–149.

CURRICULUM VITAE

Graduate College

University of Nevada, Las Vegas

G M Toufiqul Hoque

Email: toufiqulhaq@gmail.com

Degrees:

Bachelor of Science in Applied Statistics, 2015

Master of Science in Mathematics, 2021

Thesis Title: Standard and Non-Standard Log-Linear Models for 2×2 Contingency Tables

Thesis Examination Committee:

Chairperson: Petros Hadjicostas, Ph.D.

Committee Member: Hokwon Cho, Ph.D.

Committee Member: Dieudonné Phanord, Ph.D.

Graduate Faculty Representative: Ashok Singh, Ph.D.