Forum Statistika dan Komputasi : *Indonesian Journal of Statistics*
ISSN : 0853-8115
Vol. 17 No.2, Oktober 2012, p: 9-14
available online at:
journal.ipb.ac.id/index.php/statistika

THE LOG LINEAR MODELS FOR TWO DIMENSIONAL CONTINGENCY TABLES UNDER THE MULTINOMIAL SAMPLING DESIGN

Robertus Dole Guntur[1]
[1] Mathematics Department, Faculty of Science and Engineering, Nusa Cendana University
email : robertusguntur@yahoo.com

Abstract

The main purpose of this paper is to investigate the model that could be used for modeling the association patterns between two variables in the contigency table by using the multinomial sampling scheme. The result shows that for two dimensional contingency tables and by using the multinomial sampling design, we could use the log linear independence model and the log linear saturated model in order to investigated the association patterns between two variables. In the log linear independence model, the logit for the binary response does not depend on the level of the explanatory variables. Meanwhile, the direct relationship between the odds ratio and the association paratemers exists in the log linear saturated model. The association parameters could be interpreted in terms of odds ratio.

Key words : The log linear model, Independence model, Saturated model, multinomial sampling scheme.

INTRODUCTION

Modeling data in the contigency table becomes wellknown nowdays. For instance, in medical research, we may be interested in studying the association among categorical variables such as smoking habits, age, gender, and breathing test. Similarly, in educational research, we may want to investigate the relationship among variables such as grade, sex, type of school, and accomplishment in mathematics. The researchers try to used a lot of methods in order to explore the association patterns among variables. One of the approachs which is commonly used to investigate the model of association among variables is log linear model. This model could be used to explore the association patterns among a set of categorical response variables. Moreover, Tunaru (2001) and Preisser & Koch (1997) explained that the log linear models are very appropriate for contingency table in which each classification could be treated as a response variable.

Agresti (2007) shown that the log linear model is the special cases of generalized linear models which assume that each cell counts in the contigency tables derived from Poisson distribution. The total number of counts and the marginal counts are random variables. However, in particular situations the researchers could fix the total number of counts in the contingency table. For example, in medical research, Tiensuwan et al (2005) had used 3203 cancer patients record in order to investigate the associated factors that contribute for cancer cases in Thailand. In this case the inference statistics for investigating the associated factors related to cancer cases based on the multinomial sampling scheme.

When we have two categorical variables, we could use the two dimensional contingency table to express data of these variables. The sampling distribution of the data in the contingency table plays an important role in explaining the association pattern between two variables. Powers and Xie (2000) shown that each sampling distribution of the data would produce the specific joint distribution function of the two variables response in every cell of the contingency table. Moreover, when we conduct a research, we have to understand well the potential model that would be investigated in order to describe the association patterns between variables. Therefore, the main purpose of this paper is to investigate the model that could be used for modeling the association patterns between two variables in the contigency table under the assumption that the multinomial sampling was applied for data in the contingency table.

This paper consists of three main parts. Firstly, we discuss the multinomial sampling design as one of the sampling designs that could be applied in two dimensional contingency tables. Secondly, we investigate the log linear independence model for two dimensional contingency tables. The independence model of the log linear model in the two ways contingency table could be used for modeling of the association between two categorical variables. In this part, we interpret the parameter of the independence model by treating one variable as a function of the other variables. This method is very useful when one of the variables is binary response variable. Thirdly, we explore the log linear saturated model. The interpretation of parameters in the log linear saturated model could be done by using the odds ratio between two variables.

MULTINOMIAL SAMPLING DESIGN

The probability model for each cell count in the contingency table depends on the sampling dessign applied to categorical data. In the particular situations, the total number of the observations is fixed by the researcher. For instance, the investigation in medical research was conducted by Tiensuwan *et al.* (2009) in order to explore the association between cancer cases and age. They took sample based on 3203 cancer patients record in the National Cancer Institute Thailand. So in this investigation, the total sample size is fixed, however, the total number of the participants who are suffering cancer in each age group is not fixed. For these situations, a single multinomial distribution could be applied. When we apply the single multinomial sampling distribution for the two ways the contingency table, the possibility outcomes which are *IJ* cell counts have the single multinomial distribution. It is defined by

$$p(n_{ij}) = n! \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{\pi_{ij}^{n_{ij}}}{n_{ij}!} \qquad (1)$$

where $(\pi_{ij} > 0,\ i = 1,2,\dots,I; j = 1,2,\dots,J;\ n_{ij} \geq 0; \geq\ \sum_i \sum_j n_{ij} = n)$.

Suppose we have data derived from the single multinomial sample size *n*. When this data is represented in the *I X J* contingency table, then each cell is the representative of the joint probabilities function π_{ij} of the multinomial sampling distributions for two categorical responses. The association between these responses is not dependent statistically if the joint probabilities function equals the product of marginal probabilities of each response, which is

$$\pi_{ij} = \pi_{i+}\pi_{+j} \qquad (2)$$

$for\ i = 1,2,\dots,I\ and\ j = 1,2,\dots,J.$

The expected value for single multinomial sample could be described by $\mu_{ij} = n\pi_{ij}$. When the independent association between categorical data in row and column exits, the expected value for each cell count could be determined by

$$\mu_{ij} = n\pi_{i+}\pi_{+j}\ ,\ for\ \forall\ i\ and\ j \qquad (3)$$

THE LOG LINEAR INDEPENDENCE MODEL FOR TWO DIMENSIONAL CONTINGENCY TABLES

In order to investigate the association between two variables, we have to specify the model which is explaining the association patterns between these variables. Agresti[2] defined a log linear model is a linear model for the natural logarithms of the expected cell counts in a contingency table. Taking the log of both sides of equation (3) yields

$$log\ \mu_{ij} = log\ n\ +\ log\ \pi_{i+} +\ log\ \pi_{+j} \qquad (4)$$

As we can see, model in equation (4) has the additive form which is adding an *i^th^* row effect and a *j^th^* column effect. Suppose we have two variables, variable *A* was used to describe the row variable and variable *B* was used to describe the column variable. Then, equation (4) is equivalent to

$$log\ \mu_{ij} = \lambda\ +\ \lambda_i^A\ \ +\lambda_j^B \qquad (5)$$

where

$$\lambda = \log\ n + \left[\sum_{h=1}^{I}\ log\ \pi_{h+}\right]/I\ + \left[\sum_{h=1}^{J}\ log\ \pi_{+h}\right]/J$$

$$\lambda_i^A\ =\ \ log\ \pi_{i+} - \left[\sum_{h=1}^{I}\ log\ \pi_{h+}\right]/I$$

$$\lambda_j^B\ =\ \ log\ \pi_{+j} - \left[\sum_{h=1}^{J}\ log\ \pi_{+h}\right]/J$$

The parameters λ_i^A and λ_j^B satisfy the condition that $\sum_{i=1}^{I} \lambda_i^A\ = \sum_{j=1}^{I} \lambda_j^B\ = 0.$

The Model in equation (5) is called the log linear independence model for two dimensional contingency tables. The parameter λ_i describes the effect of classification in row *i* and the parameter λ_j describes the effect of classification in column *j*. For instance, table 1 below is a 4 *X 3* table of μ_{ij} that satisfies the log linear model of independence for two ways contingency table.

Table 1. The expectation of the observation in a *4 x3* contingency table

	Variable B		
Variable A	μ_{11}	μ_{12}	μ_{13}
	μ_{21}	μ_{22}	μ_{23}
	μ_{31}	μ_{32}	μ_{33}
	μ_{41}	μ_{42}	μ_{43}

All μ_{ij} for *i* =1, 2, 3, 4 and *j* =1, 2, 3 in Table 1 satisfy the log linear independence model in equation (5), which specification of the parameter model could be defined by :

$$\lambda = \log\ n + \left[\sum_{h=1}^{4}\ log\ \pi_{h+}\right]/4 + \left[\sum_{h=1}^{3}\ log\ \pi_{+h}\right]/3$$

$$\lambda_1^A\ =\ \ log\ \pi_{1+} - \left[\sum_{h=1}^{4}\ log\ \pi_{h+}\right]/4$$

$$= \frac{(\ 3\ log\,\pi_{1+} -\ \ log\,\pi_{2+} -\ \ log\,\pi_{3+} -\ \ log\,\pi_{4+})}{4} \tag{6}$$

$$\lambda_2^A = \quad log\,\pi_{2+} - \left[\sum_{h=1}^{4} log\,\pi_{h+}\right]/4$$

$$= \frac{(\ 3\ log\,\pi_{2+} -\ \ log\,\pi_{1+} -\ \ log\,\pi_{3+} -\ \ log\,\pi_{4+})}{4} \tag{7}$$

$$\lambda_3^A = \quad log\,\pi_{3+} - \left[\sum_{h=1}^{4} log\,\pi_{h+}\right]/4$$

$$= \frac{(\ 3\ log\,\pi_{3+} -\ \ log\,\pi_{1+} -\ \ log\,\pi_{1+} -\ \ log\,\pi_{4+})}{4} \tag{8}$$

$$\lambda_4^A = \quad log\,\pi_{4+} - \left[\sum_{h=1}^{4} log\,\pi_{h+}\right]/4$$

$$= \frac{(\ 3\ log\,\pi_{4+} -\ \ log\,\pi_{1+} -\ \ log\,\pi_{3+} -\ \ log\,\pi_{1+})}{4} \tag{9}$$

Then, we could use the value in equation (6), (7), (8), and (9). So we found that

$$\sum_{i=1}^{4} \lambda_i^A = \lambda_1^A + \lambda_2^A + \lambda_3^A + \lambda_4^A = 0$$

Next, for parameter $\lambda_j^B \ \ for\ j = 1, 2,\ and\ 3$

$$\lambda_1^B = \quad log\,\pi_{+1} - \left[\sum_{h=1}^{3} log\,\pi_{+h}\right]/3$$

$$= \frac{(\ 2\ \ log\,\pi_{+1} -\ \ \ log\,\pi_{+2} -\ \ \ log\,\pi_{+3})}{3} \tag{10}$$

$$\lambda_2^B = \quad log\,\pi_{+2} - \left[\textstyle\sum_{h=1}^{3} log\,\pi_{+h}\right]/3$$

$$= \frac{(\ 2\ \ log\,\pi_{+2} -\ \ \ log\,\pi_{+1} -\ \ \ log\,\pi_{+3})}{3} \tag{11}$$

$$\lambda_3^B = \quad log\,\pi_{+3} - \left[\textstyle\sum_{h=1}^{3} log\,\pi_{+h}\right]/3$$

$$= \frac{(\ 2\ \ log\,\pi_{+3} -\ \ \ log\,\pi_{+2} -\ \ \ log\,\pi_{+1})}{3} \tag{12}$$

Then, from equation (10), (11), and (12) we found that

$$\sum_{j=1}^{3} \lambda_j^B = \lambda_1^B + \lambda_2^B + \lambda_3^B = 0$$

Therefore, λ_i^A and λ_j^B satisfy the condition $\sum_{i=1}^{4} \lambda_i^A = \sum_{j=1}^{3} \lambda_j^B = 0$ as required in the log linear independent model in equation (5)

The Hypothesis test could be conducted to investigate that there is association between variables A and variables B. Let consider an *I X J* contingency table with the multinomial sampling model for each cell count. The null hypothesis for test of the independent association between two variables is the model in equation (5) holds, and the alternative hypothesis is the model in equation (5) does not hold. Under the null hypothesis is true, the maximum likelihood estimation of π_{ij} for each cell in the contingency table could be attained. Guntur (2012) shown that the maximum likelihood estimator for the mean of multinomial distribution could be defined by $\hat{\mu}_{ij} = n_{i.}n_{.j}/n..$ Therefore, we could test this independence hypothesis by using the X^2 and G^2 test which is the Pearson statistics and the likelihood ratio statistics respectively. For those tests, we could use the fitted value of the model $(\hat{\mu}_{ij})$ which is obtained by getting the solution of the set of likelihood equation. Agresti (1990) shown that for two ways contingency table, the likelihood ratio statistics is defined by formula

$$G^2 = 2\textstyle\sum_{i=1}^{I} \quad \sum_{j=1}^{J} n_{ij}\, log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right) \tag{13}$$

and the Pearson statistics is determined by

$$X^2 = \textstyle\sum_{i=1}^{I} \quad \sum_{j=1}^{J} \left(\frac{\left(n_{ij} - \hat{\mu}_{ij}\right)^2}{\hat{\mu}_{ij}}\right) \tag{14}$$

The distribution for these tests is χ^2 with degree of freedom *(I-1)(J-1)*

INTERPRETATION PARAMETER OF THE LOG LINEAR INDEPENDENCE MODEL

The interpretation of parameters in the log linear model is not straightforward, because we treat variables in rows and columns as response variables. However, in the case of the product multinomial model with a binary response, a simple interpretation is possible. This approach is very suitable particularly when we have two levels of categorical response. For instance, we have the log linear model of independence for *I X 2* tables. The logit of binary response for every row *i* =1, 2, ..., I becomes

$$\log\left(\frac{\pi_{1|i}}{\pi_{2|i}}\right) = log\left(\frac{\mu_{i1}}{\mu_{i2}}\right)$$

$$= \left(\lambda + \lambda_i^A \quad + \lambda_1^B\right) - \left(\lambda + \lambda_i^A \quad + \lambda_2^B\right)$$

$$= \lambda_1^B - \lambda_2^B \tag{15}$$

It is clear from this result that the logit of the model does not depend on the row. In other words, the logit for the binary response does not depend on the level of the explanatory variables. For zero sum constraint on parameters, equation (15) equals $2\lambda_1^B$. Therefore, the odds of response in level one factor *B* than level two factors *B* equal $\exp(2\lambda_1^B)$.

For example, we disscus data from Agresti (2002) in the study of the association between smoking status and breathing test results, by the age less than 40. In this study, 1350 people who have never smoked and current smokers participated. The number who showed normal breathing test is 577 of the 611 people who never smoked, and 682 of the 739 people who is current smokers. The null hypothesis for this test is Ho : The independence model between breathing test result and smoking status. The result of fitting the log linear independence model conducted using R software is summarized in Table 2. From this table, we can see that the model fits well using as criteria the Likelihood ratio test and the Pearson Chi-Square because the p-value for those tests is greater than alpha either 0.01 or 0.05. So, based on this data there is no association between breathing test result and smoking status. For the constraints used, $\lambda_1^B = 0$ and $\lambda_2^B = -2.627$. Therefore, the estimated odds of normal breathing test result for people who were age less than 40 was exp(-2.627)= 0.072 for each smoking status.

Table 2 The result of fitting independence log linear model to cross-classification of breathing test result by smoking status.

Criteria for assessing goodness of fit model			
Criterion	Df	Value	Sig.
Likelihood Ratio	1	2.489	0.115
Pearson Chi-Square	1	2.456	0.117
Parameter Estimates			
Parameter		Estimates	Standart Error
Constant		6.536	0.038
Smoking	Never Smoked	-0.190	0.055
Smoking	Current Smoked	0	0
Breathing Test Result	Normal	-2.627	0.109
Breathing Test Result	Not normal	0	0

THE LOG LINEAR SATURATED MODEL FOR TWO DIMENSIONAL CONTINGENCY TABLES

The saturated model is the most complex model for the two ways contingency table, because in this model the number of parameters equals the number of cells in the table. The interaction between two variables is included in the saturated model. By using this model we could get a result that is the fitted values are exactly equal to the observed value. Therefore, the saturated model fits the data perfectly. Agresti (1984) and Christensen (1990) explained that the log linear saturated model for two ways table is defined by

$$log\,\mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \qquad (16)$$

where

$\lambda_i^A = \eta_{i.} - \eta_{..}$
$\lambda_j^B = \eta_{.j} - \eta_{..}$
$\lambda_{ij}^{AB} = \eta_{ij} - \eta_{i.} - \eta_{.j} + \eta_{..}$

$$\lambda = \eta_{..} = \sum_{i=1}^{I} \sum_{j=1}^{J} \eta_{ij} /IJ$$

In this case, η_{ij} is defined by $\eta_{ij} = \log \mu_{ij}$. The parameters λ_{ij}^{AB} satisfy the condition that

$$\sum_{i=1}^{I} \lambda_{ij}^{AB} = \sum_{j=1}^{J} \lambda_{ij}^{AB} = 0 \qquad (17)$$

Moreover, the parameters λ_i^A and λ_j^B satisfy the condition that

$$\sum_{i=1}^{I} \lambda_i^A = \sum_{j=1}^{J} \lambda_j^B = 0 \qquad (18)$$

For instance, table 3 below is a 4 *X 3* table of μ_{ij} that satisfies the log linear model of saturated model for two ways contingency table.

Table 3. The expectation of the observation in a 4 x 3 contingency table

	Variable B		
Variable A	μ_{11}	μ_{12}	μ_{13}
	μ_{21}	μ_{22}	μ_{23}
	μ_{31}	μ_{32}	μ_{33}
	μ_{41}	μ_{42}	μ_{43}

All μ_{ij} for *i* =1, 2, 3, 4 and *j* =1, 2, 3 in table 3 satisfy the log linear saturated model as be stated in equation (16), which specifications of the parameter model are

$$\lambda_{1j}^{AB} = \eta_{1j} - \eta_{1.} - \eta_{.j} + \eta_{..} \qquad (19)$$

$$\lambda_{2j}^{AB} = \eta_{2j} - \eta_{2.} - \eta_{.j} + \eta_{..} \qquad (20)$$

$$\lambda_{3j}^{AB} = \eta_{3j} - \eta_{3.} - \eta_{.j} + \eta_{..} \qquad (21)$$

$$\lambda_{4j}^{AB} = \eta_{4j} - \eta_{4.} - \eta_{.j} + \eta_{..} \qquad (22)$$

where

$$\eta_{ij} = \log \mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

$$for\ \ i = 1, 2, 3, 4$$

$$\eta_{1.} = \frac{\sum_{j=1}^{3} \eta_{1j}}{3} = \frac{\eta_{11} + \eta_{12} + \eta_{13}}{3}$$

$$\eta_{2.} = \frac{\sum_{j=1}^{3} \eta_{2j}}{3} = \frac{\eta_{21} + \eta_{22} + \eta_{23}}{3}$$

$$\eta_{3.} = \frac{\sum_{j=1}^{3} \eta_{3j}}{3} = \frac{\eta_{31} + \eta_{32} + \eta_{33}}{3}$$

$$\eta_{4.} = \frac{\sum_{j=1}^{3} \eta_{4j}}{3} = \frac{\eta_{41} + \eta_{42} + \eta_{43}}{3}$$

$$\eta_{.j} = \frac{\sum_{i=1}^{4} \eta_{ij}}{4} = \frac{\eta_{1j} + \eta_{2j} + \eta_{3j} + \eta_{4j}}{4},$$

$$j = 1, 2, 3$$

$$\eta_{..} = \frac{\sum_{i=1}^{4} \sum_{j=1}^{3} \eta_{ij}}{12}$$

$$= \frac{\eta_{11}+\eta_{12}+\eta_{13}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$

Then, we solve the equation (19), (20), (21), and (22). By some manipulations in algebra, we found that

$$\sum_{i=1}^{4} \lambda_{ij}^{AB} = \lambda_{1j}^{AB} + \lambda_{2j}^{AB} + \lambda_{3j}^{AB} + \lambda_{4j}^{AB} = 0$$

Next, for parameter λ_{ij}^{AB} $for\ j = 1, 2,\ and\ 3$

$$\lambda_{i1}^{AB} = \eta_{i1} - \eta_{i.} - \eta_{.1} + \eta_{..} \quad (23)$$

$$\lambda_{i2}^{AB} = \eta_{i2} - \eta_{i.} - \eta_{.2} + \eta_{..} \quad (24)$$

$$\lambda_{i3}^{AB} = \eta_{i3} - \eta_{i.} - \eta_{.3} + \eta_{..} \quad (25)$$

where

$$\eta_{ij} = log\,\mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$
$$for\ j = 1, 2, 3$$

$$\eta_{.1} = \frac{\sum_{i=1}^{4}\eta_{i1}}{4} = \frac{\eta_{11}+\eta_{21}+\eta_{31}+\eta_{41}}{4}$$

$$\eta_{.2} = \frac{\sum_{i=1}^{4}\eta_{i2}}{4} = \frac{\eta_{12}+\eta_{22}+\eta_{32}+\eta_{42}}{4}$$

$$\eta_{.3} = \frac{\sum_{i=1}^{4}\eta_{i3}}{4} = \frac{\eta_{13}+\eta_{23}+\eta_{33}+\eta_{43}}{4}$$

$$\eta_{i.} = \frac{\sum_{j=1}^{3}\eta_{ij}}{3} = \frac{\eta_{i1}+\eta_{i2}+\eta_{i3}}{3},$$
$$for\ i = 1, 2, 3, 4$$

$$\eta_{..} = \frac{\sum_{i=1}^{4}\ \sum_{j=1}^{3}\eta_{ij}}{12}$$
$$= \frac{\eta_{11}+\eta_{12}+\eta_{13}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$

Then, we solve the equation (23), (24), and (25). We found that

$$\sum_{j=1}^{3} \lambda_{ij}^{AB} = \lambda_{i1}^{AB} + \lambda_{i2}^{AB} + \lambda_{i3}^{AB} = 0$$

Now, we investigate the parameters λ_i^A and λ_j^B. It should satisfy the condition that

$$\sum_{i=1}^{4} \lambda_i^A = \sum_{j=1}^{3} \lambda_j^B = 0$$

By using the parameter in equation (19), (20), (21), and (22), we found that

$$\lambda_1^A = \eta_{1.} - \eta_{..}$$
$$= \frac{\sum_{j=1}^{3}\eta_{1j}}{3} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{21}+\eta_{22}+\eta_{23}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$
$$= \frac{3(\eta_{11}+\eta_{12}+\eta_{13})}{12} - \frac{\eta_{21}+\eta_{22}+\eta_{23}+\eta_{31}+\eta_{32}+\eta_{33}+\eta_{41}+\eta_{42}+\eta_{43}}{12} \quad (26)$$

$$\lambda_2^A = \eta_{2.} - \eta_{..}$$
$$= \frac{\sum_{j=1}^{3}\eta_{2j}}{3} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{21}+\eta_{22}+\eta_{23}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$
$$= \frac{3(\eta_{21}+\eta_{22}+\eta_{23})}{12} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{31}+\eta_{32}+\eta_{33}+\eta_{41}+\eta_{42}+\eta_{43}}{12} \quad (27)$$

$$\lambda_3^A = \eta_{3.} - \eta_{..}$$
$$= \frac{\sum_{j=1}^{3}\eta_{3j}}{3} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{21}+\eta_{22}+\eta_{23}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$
$$= \frac{3(\eta_{31}+\eta_{32}+\eta_{33})}{12} - \frac{\eta_{21}+\eta_{22}+\eta_{23}+\eta_{11}+\eta_{12}+\eta_{13}+\eta_{41}+\eta_{42}+\eta_{43}}{12} \quad (28)$$

$$\lambda_4^A = \eta_{4.} - \eta_{..}$$
$$= \frac{\sum_{j=1}^{3}\eta_{4j}}{3} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{21}+\eta_{22}+\eta_{23}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$
$$= \frac{3(\eta_{41}+\eta_{42}+\eta_{43})}{12} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{31}+\eta_{32}+\eta_{33}+\eta_{21}+\eta_{22}+\eta_{23}}{12} \quad (29)$$

By adding the equation (26), (27), (28), and (29), we found that

$$\sum_{i=1}^{4} \lambda_i^A = \lambda_1^A + \lambda_2^A + \lambda_3^A + \lambda_4^A = 0$$

Next, we want to show that $\sum_{j=1}^{3} \lambda_j^B = 0$. By using the parameters in equation (23), (24), and (25) we found that

$$\lambda_1^B = \eta_{.1} - \eta_{..}$$
$$= \frac{\sum_{i=1}^{4}\eta_{i1}}{4} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{21}+\eta_{22}+\eta_{23}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$
$$= \frac{2(\eta_{11}+\eta_{21}+\eta_{31}+\eta_{41})}{12} - \frac{\eta_{12}+\eta_{22}+\eta_{32}+\eta_{42}+\eta_{13}+\eta_{23}+\eta_{33}+\eta_{43}}{12} \quad (30)$$

$$\lambda_2^B = \eta_{.2} - \eta_{..}$$
$$= \frac{\sum_{i=1}^{4}\eta_{i2}}{4} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{21}+\eta_{22}+\eta_{23}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$
$$= \frac{2(\eta_{12}+\eta_{22}+\eta_{32}+\eta_{42})}{12} - \frac{\eta_{11}+\eta_{21}+\eta_{31}+\eta_{41}+\eta_{13}+\eta_{23}+\eta_{33}+\eta_{43}}{12} \quad (31)$$

$$\lambda_3^B = \eta_{.3} - \eta_{..}$$
$$= \frac{\sum_{i=1}^{4}\eta_{i3}}{4} - \frac{\eta_{11}+\eta_{12}+\eta_{13}+\eta_{21}+\eta_{22}+\eta_{23}+\cdots+\eta_{41}+\eta_{42}+\eta_{43}}{12}$$
$$= \frac{2(\eta_{13}+\eta_{23}+\eta_{33}+\eta_{43})}{12} - \frac{\eta_{11}+\eta_{21}+\eta_{31}+\eta_{41}+\eta_{12}+\eta_{22}+\eta_{32}+\eta_{42}}{12} \quad (32)$$

Using the value in the equation (30), (31), and (32) we found that

$$\sum_{j=1}^{3} \lambda_j^B = \lambda_1^B + \lambda_2^B + \lambda_3^B = 0$$

Therefore, the parameters λ_i^A, λ_j^B, and λ_{ij}^{AB} for data in Table 3 satisfy the condition of the log linear saturated model for two dimensional contingency tables.

The particular characteristic of the model in equation (16) is the additional parameter λ_{ij}^{AB} included in the model, so there is an interaction between two variables in the log linear saturated model. When all λ_{ij}^{AB} equals zero, the log linear saturated model could become the log linear independence model.

It is interesting to note that the parameters in equation (16) could be interpreted in terms of odds ratio. This condition can be described easily for 2 *X 2* contingency tables. The log odds ratio for this table is described by

$$\log\theta = \log\frac{(\mu_{11}\mu_{22})}{(\mu_{12}\mu_{21})}$$

$$= \log\mu_{11} + \log\mu_{22} - \log\mu_{12} - \log\mu_{21} \qquad (33)$$

By using the definition of the log linear saturated model for two dimensional contingency tables and using some manipulations in algebra, we could modify equation (33) to become

$$\log\theta = \lambda_{11}^{AB} + \lambda_{22}^{AB} - \lambda_{12}^{AB} - \lambda_{21}^{AB} \qquad (34)$$

As we can see from equation (34), the log odds ratio was determined by λ_{ij} which is the log odds ratio is represented as a function of the association parameters. It could become the direct relationship between the odds ratio and the association paratemers in the log linear saturated model. For zero constraints which is equation (17), we found that

$$\lambda_{11}^{AB} = \lambda_{22}^{AB} = -\lambda_{12}^{AB} = -\lambda_{21}^{AB}$$

Therefore, the equation (34) could be simplified, so we have

$$log\theta = 4\lambda_{11}^{AB}$$

It is interesting to note that the model in equation (16) could become the independence model when the parameters$\lambda_{ij}^{AB} = 0, for\ i = 1,2,\ldots,I\ and\ j = 1,2,\ldots J$. So the log odds ratio between variable A and B equals zero. Therefore, the odds ratio between variable *A* and *B* equals one. It indicates that there is no association between variable *A* and variable *B*.

CONCLUSION

For two dimensional contingency tables and by using the multinomial sampling design, we could use the independence model and the saturated model of log linear model in order to investigated the association patterns between two variables. In the log linear independence model, the logit for the binary response does not depend on the level of the explanatory variables. Meanwhile, there is a direct relationship between the odds ratio and the association paratemers in the log linear saturated model. The association parameters could be interpreted in terms of odds ratio.

REFERENCE

Agresti A. 2007. *An Introduction to Categorical Data Analysis*. New York : John Wiley and Sons.

Agresti A. 2002. *Categorical Data Analysis*. New York : John Wiley and Sons.

Agresti A. 1990. *Categorical Data Analysis*. New York : John Wiley and Sons.

Agresti A. 1984. *Analysis of Ordinal Categorical Data*. New York : John Wiley and Sons.

Christensen R R. 1990. *Log linear Models*. New York : Springer Verlag

Guntur R. 2012. Models for Contingency Tables [*Thesis for Master of Mathematical Science*]. Adelaide : School of Mathematical Science.

Powers DA and Xie Y. 2000. *Statistical Method for Categorical Data Analysis*. San Diego : Academic Press

Preisser JS, Koch GG. 1997. Categorical Data Analysis in Public Health. *Annual Review Public Health* 18: 51 – 82.

Tiensuwan M, Yimprayoon P, Lenbury Y. 2005. Application of Log Linear Models to Cancer Patients : a Case Study of Data from the National Cancer Institute. *Asian J Trop Med Public Health* 36(5) :

Tunaru R. 2001. Models for Association Versus Causal Models for Contingency Tables. *Royal Statistical Society* 50(3): 257–269.