Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Applications and Applied Mathematics:
An International Journal
(AAM)

Vol. 14, Issue 1 (June 2019), pp. 63 - 75

Construction and Application of Log-linear Models to Assess Academic Performance

Indika R. Wickramasinghe

Department of Mathematics
Prairie View A&M University
Prairie View, Texas, USA
iprathnathungalage@pvamu.edu;

Received: September 14, 2018; Accepted: March 28, 2019

Abstract

Log-Linear Models (LLMs) are important techniques used in categorical data analysis. Though there are some available published work about LLMs, the explanation of model building process and the theoretical background are not adequate. Furthermore, research about the application of the LLM theory and the selection procedure of the best model are handful. Therefore, this manuscript aims to fill that vacuum. At first, the construction of LLM and Hierarchical Log-Linear Models (HLLMs), a branch of LLMs are discussed in connection with both 2×2 and $2 \times 2 \times 2$ contingency tables. Secondly, an application is presented to analyze the collected data set about the academic performance of elementary students. The manuscript also discusses the criteria to select the best model that fits the collected data.

Keywords: Log-linear Models; Contingency tables; Students' success; Linear Models; Hierarchical Models; Academic Performance

MSC 2010 No.: 62P25, 97M10

1. Introduction

Contingency tables, introduced by Pearson (1903) are comprehensive forms to summarize the association between two or more categorical variables. Each subject in the multivariate nature is represented by a tuple of characteristics. After counting the common characteristics of all subjects, this numerical value is considered as the cell count of the table at the respective combination of the characteristic (Bishop et al, 1904). LLMs represent a mathematical function of a combination

of parameters in a linear form. These models are very convenient and most widely used mathematical models to analyze categorical data, represented in the cross-classified nature in the contingency tables (Christensen, 1997). The flexibility and the interpretability are considered as two of the advantages of LLMs.

As the literature indicates, Andersen (1974) provides a comprehensive study about multidimensional contingency tables. Estimating the effect of interactions among the categorical variables has a high importance. Bartlett (1935) discusses how to test for three-factor-interactions in contingency tables. Hu et al. (2009) discusses a class of log-linear models to determine the interactions in high-dimensional genomic data. Further discussions and applications of LLMs, contingency tables, and parameter estimation of models are presented in the literature. Fienberg and Meyer (1983) review the calculation of maximum likelihood estimates (MLE) of LLMs for large multi-way contingency tables. Furthermore, Kelderman and Psychometrika (1992) describe algorithms to obtain MLEs of the parameters in LLMs for analyzing educational and psychological data. In another study, Zhu et al. (2006) utilizes LLM and Structural Equation Modeling (SEM) to analyze caregivers' health data. Though both methods give similar results, the author finds that LLM is more parsimonious and converges more easily. A comprehensive study about graphical forms of LLMs and fundamentals of contingency tables are presented by Gauradha (2017). In addition, Brzezinska (2012) utilizes likelihood ratio approach and after that the author discusses how to use AIC and BIC information criteria for the selection of the best log-linear model.

A branch of LLMs, called Hierarchical Log-Linear Models (HLLMs) represent all the lower and main effect interactions (Howell, 2009) in the model. Therefore, a LLM can be either hierarchical or non- hierarchical in nature. According to the description of Chapman et al. (2008), HLLMs consist of only parameters for which entire implied parameters are also comprised of the model. Applications of these models can be seen in wide variety of disciplines including marketing, social and psychological research (Brzezinska, 2012). Though there are some published research work in LLMs, according to our point of view the studies exploring the theoretical foundation of these models and application of them in a practical scenario are significantly less. When it comes to HLLMs, the available number of literature work is not sufficient. Therefore, the aim of this work is to fulfill the above need. First, the manuscript attempts to explain the theoretical aspect and the formation of LLMs. The second phase focuses on the application of the formed theoretical model to analyze academic performances of elementary students. For this data, multiple linear regression can be used to study the relationship among the variables related to students' academic performance.

One of the main issues of the above approach when explanatory variables are of categorical nature is the introduction of dummy variables. This does not become a major issue for the case of less number of categorical variables, but things are not pleasing when there are larger number of categorical variables and their levels. Moreover, the interpretation and introduction of higher number of variables become cumbersome as well. Analysis of variance (ANOVA) is a powerful technique in analyzing categorical variables, but if there is a doubt about the violation of homogeneity of the variance of groups (homoscedasticity), above technique is out of the consideration. Therefore, LLM is the appropriate technique for this data analysis.

The rest of the manuscript flows as follows. Section 2 describes the underline theory related to contingency tables, and the model building of both LLMs and HLLMs. Section 3 describes the collected data about students' academic performance and focusses on the identification of the relationship among student related variables. After constructing models, one of the main issues is the selection of the best model that fits the data. This is discussed in Section 4. Finally, Section 5 concludes the manuscript.

2. Methodology

2.1. Contingency Tables

Contingency table can be represented as a list f of frequencies$(f(x) \in \mathbb{N}: x \in \mathcal{X})$, where $\mathcal{X}$ is the sample space. For an instance, consider the following simplest case of 2×2 contingency table shown in Table 1.

Table 1: 2×2 contingency table

		Factor A	
		Level 1	**Level 2**
actor B	**Level 1**	$m(x_1)$	$m(x_2)$
	Level 2	$m(x_3)$	$m(x_4)$

Consider a given random contingency table $f \in \mathfrak{R}^p$ with sample space $\mathcal{X}$. Let's assume that there is a random vector $F = (F(x) : x \in \mathcal{X})$ that generates f and the mean of the random vector F is denoted by $m = (m(x): x \in \mathcal{X}) = \big(E\big(F(x)\big): x \in \mathcal{X}\big)$. As we always assume that $m(x)$ is positive value, the logarithmic mean vector of f is defined. This is represented by μ and is defined by $(\log[m(x)] : x \in \mathcal{X})$.

1.2. The Log- Linear Model (LLM) and Modeling Contingency Tables

In log-linear modeling, we make assume that $\mu = (\mu(x): x \in \mathcal{X})$ belongs to a q-dimensional linear space M contained in $\mathfrak{R}^p$. Consider an orthogonal basis of $\mathfrak{R}^p$. As Rapallo (2003) shows, using a finite number of vectors in the orthogonal basis, one can generate $2^p - 1$ linear subspaces of $\mathfrak{R}^p$. Suppose M be any model for μ with the dimension $q \leq p$. Let $P_M f$ be the orthogonal projection of $f \epsilon M = \{\mu \epsilon \mathfrak{R}^p | \exists c \in \mathfrak{R}^q : \mu = \breve{M} c\}$, then

$$P_M f = \breve{M}\big(\breve{M}^T \breve{M}\big)^{\sim} \breve{M}^T f.$$

Following Rapallo (2003), let's consider the case of $p = 8$ and $\mathfrak{R}^8$ can be spanned by the following orthogonal vectors. u_1 = (1, 1, 1, 1, 1, 1, 1, 1)$'$, u_2 = (1, 1, 1, 1,-1, -1,-1,-1)$'$, u_3 = (1,1,-1,-1,1,1,-1-,1)$'$, u_4 = (1,-1- 1,-1, 1,-1,1,-1)$'$, u_5 = (1, 1,-1,-1,-1,-1, 1, 1)$'$; u_6 = (1-1,1,-1,-1,1,-1,1)$'$,u_7 = (1,-1,-1,1, 1,-1,-1,1)$'$ and u_8 = (1,-1,-1,1,-1,1,1,-1)$'$.

Let $X, Y,$ and Z be the selected factors with levels, $i = 1, 2, \dots, I; j = 1, 2, \dots, J;$ and $k = 1, 2, \dots, K$ and n_{ijk}be the cell count belongs to the cell (i, j, k) as represented by the Table 2.

Table 2: $2 \times 2 \times 2$ contingency table

Y\X	$\boldsymbol{X = 1}$			$\boldsymbol{X = 2}$		
	$Z = 1$	$Z = 2$	Total	$Z = 1$	$Z = 2$	Total
$Y = 1$	$f(x_{111})$	$f(x_{112})$	n_{11+}	$f(x_{211})$	$f(x_{212})$	n_{21+}
$Y = 2$	$f(x_{121})$	$f(x_{122})$	n_{12+}	$f(x_{221})$	$f(x_{222})$	n_{22+}
Total	n_{1+1}	n_{1+2}	N	n_{2+1}	n_{2+2}	N

In this case of $2 \times 2 \times 2$, the sample space can be represented as $\mathcal{X} = (\mathrm{x}_{111}, \mathrm{x}_{112}, \dots, \mathrm{x}_{222})$. For the unknown parameter μ, any model M is a linear subspace of $\Re^8$, which is generated by the above orthogonal basis. As mentioned before, there are $(2^8 - 1)$ = 255 possible models and each model, $M_i; i = 1, \dots, 255$ is constructed using the corresponding spans.

Model 1: $M_1 = \mathrm{span}(u_1)$.

Model 2, $M_2 = \mathrm{span}(u_2)$,

…

Model 255: $M_{255} = \mathrm{span}(u_1, u_2, \dots, u_8)$.

Here,

$$\mu = (\mu_{111}, \mu_{112}, \dots, \mu_{222})' \text{ and } \mu_{ijk} = \log(m_{ijk})\,; i, j, k = 1,2.$$

As an illustration purpose, these model constructions can be explained for p = 4 as follows. In this case, there are 15 possible models using the naming conversion of Agresti (1990, p. 144).

Here, we present some of the selected models for the case of p=4. For this case we consider $w_1 = (1,1,1,1)', w_2 = (1,1,-1,-1)', w_3 = (1,-1,1,-1)'$, and $w_4 = (1,-1,-1,1)'$. Let's consider the model M_1 is spanned by vector u_1. Since $\mu = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})'$, where $\mu_{ijk} = \log(m_{ijk})\,; i, j = 1,2$. Then, $\mu = c_1 u_1$ for some $c_1 \epsilon \Re$.

Hence, $\mu = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})' = (c_1, c_1, c_1, c_1)' = \breve{M}_1(c_1)$, where $\breve{M}_1 = (1,1,1,1)'$. Then, the orthogonal projection can be found.

$$P_{M_1} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

By taking $\lambda = c_1$, the model can be rewritten as $\mu_{ij} = \log(m_{11}) = \lambda; i, j = 1,2$. Similarly, the rest of the models $(M_2 - M_4)$ are spanned by a single vector and can be constructed accordingly.

The fifth model is spanned by vectors w_1 and w_2 and can be constructed as follows.

$$\begin{aligned}\mu &= c_1 w_1 + c_2 w_2 \\ &= (c_1 + c_2, c_1 + c_2, c_1 - c_2, c_1 - c_2)' \\ &= \breve{M}_5 (c_1, c_2)', \text{ where } c_{1,} c_2 \epsilon \Re.\end{aligned}$$

Here,

$$\breve{M}_5 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix}$$

and the orthogonal projection becomes,

$$P_{M_5} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

By defining $\lambda = c_1, \lambda_1^X = c_2,$ and $\lambda_2^X = -c_2,$ the model can be represented as follows.

$$\mu_{ij} = \log(m_{ij}) = \begin{Bmatrix} \lambda + \lambda_1^X; i = 1, j = 1 \\ \lambda + \lambda_1^X; i = 1, j = 2 \\ \lambda + \lambda_2^X; i = 2, j = 1 \\ \lambda + \lambda_2^X; i = 2, j = 2 \end{Bmatrix} = \lambda + \lambda_i^X.$$

Using a similar approach, the rest of the models $(M_6 - M_{10})$ can be constructed using the two vectors. After all the models spanned by using two vectors, let's consider the models spanned by three vectors u_1, u_2 and u_3 as follows.

$$\begin{aligned}\mu &= c_1 w_1 + c_2 w_2 + c_3 w_3 \\ &= (c_1 + c_2 + c_3, c_1 + c_2 - c_3, c_1 - c_2 + c_3, c_1 - c_2 - c_3)' \\ &= \breve{M}_{11} (c_1, c_2, c_3)', \text{ where } c_{1,} c_2, c_3 \epsilon \Re.\end{aligned}$$

Here,

$$\breve{M}_{11} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

and the orthogonal projection,

$$P_{M_{11}} = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & \frac{-1}{4} \\ \frac{1}{4} & \frac{3}{4} & \frac{-1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{-1}{4} & \frac{3}{4} & \frac{1}{4} \\ \frac{-1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

By defining $= c_1, \lambda_1^X = c_2, \lambda_2^X = -c_2, \lambda_1^Y = c_3$ and $\lambda_2^Y = -c_3$, the model can be represented as follows:

$$\mu_{ij} = \log(m_{ij}) = \begin{Bmatrix} \lambda + \lambda_1^X + \lambda_1^Y; i = 1, j = 1 \\ \lambda + \lambda_1^X + \lambda_2^Y; i = 1, j = 2 \\ \lambda + \lambda_2^X + \lambda_1^Y; i = 2, j = 1 \\ \lambda + \lambda_2^X + \lambda_2^Y; i = 2, j = 2 \end{Bmatrix} = \lambda + \lambda_i^X + \lambda_j^Y.$$

As mentioned before the rest of the models can be generated accordingly. Finally, let's consider the model spanned by all the four vectors w_1, w_2, w_3 and w_4.

$$\begin{aligned} \mu &= c_1 w_1 + c_2 w_2 + c_3 w_3 + c_4 w_4 \\ &= (c_1 + c_2 + c_3 + c_4, c_1 + c_2 - c_3 - c_4, c_1 - c_2 + c_3 - c_4, c_1 - c_2 - c_3 + c_4)' \\ &= \breve{M}_{15}(c_1, c_2, c_3, c_4)', \text{ where } c_{1,} c_2, c_3, c_4 \epsilon \Re. \end{aligned}$$

Here,

$$\breve{M}_{15} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

and the orthogonal projection

$$P_{M_{15}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

By defining $\lambda = c_1, \lambda_1^X = c_2, \lambda_2^X = -c_2, \lambda_1^Y = c_3$ and $\lambda_2^Y = -c_3$, the model can be illustrated as follows:

$$\mu_{ij} = \log(m_{ij}) = \begin{Bmatrix} \lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}; i = 1, j = 1 \\ \lambda + \lambda_1^X + \lambda_2^Y + \lambda_{12}^{XY}; i = 1, j = 2 \\ \lambda + \lambda_2^X + \lambda_1^Y + \lambda_{21}^{XY}; i = 2, j = 1 \\ \lambda + \lambda_2^X + \lambda_2^Y + \lambda_{22}^{XY}; i = 2, j = 2 \end{Bmatrix} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}.$$

Each of the above models are constructed under the following restrictions.

$$\sum_i \lambda_i^X = \sum_j \lambda_j^Y = 0; \sum_i \lambda_{ij}^{XY} = 0; \sum_j \lambda_{ij}^{XY} = 0.$$

As stated before, using the orthogonal basis of eight vectors the total number of models one can construct is 255. The final model of this collection is named as the full model, and it can be expressed according to the following notation:

$$\log(m_{ijk}) = \log\left(m(x_{ijk})\right) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ},$$

under the following restrictions.

$$\begin{aligned}
&\textstyle\sum_{i=1}^2 \lambda_i^X = \sum_{j=1}^2 \lambda_j^Y = \sum_{k=1}^2 \lambda_k^Z = 0;\ \sum_{i=1}^2 \lambda_{ij}^{XY} = 0 \text{ for } j = 1,2,\\
&\textstyle\sum_{j=1}^2 \lambda_{jk}^{YZ} = 0 \text{ for } k = 1,2;\ \sum_{i=1}^2 \lambda_{ik}^{XZ} = 0 \text{ for } i = 1,2,\\
&\textstyle\sum_{i=1}^2 \lambda_{ijk}^{XYZ} = 0 \text{ for } j,k = 1,2;\ \sum_{j=1}^2 \lambda_{ijk}^{XYZ} = 0 \text{ for } i,k = 1,2,\\
&\textstyle\sum_{k=1}^2 \lambda_{ijk}^{XYZ} = 0 \text{ for } i,j = 1,2.
\end{aligned}$$

2.3. Hierarchical Log Linear Models

As stated before, hierarchical log-linear models (HLLMs) represent all the lower and main effect interactions. For an instance, if the interaction of *X* and *Y* factors is in the model, then the factors *X* and *Y* also should be in the model. Using the generator multigraph, Khamis (1996) shows how to represent hierarchical loglinear models graphically.

According to the definition of hierarchical models in Agresti (1990), if a model contains λ_{ij}^{XY} term, it also contains λ_i^X and λ_j^Y as well. Based on that definition, we can list the following 19 models (in Table 3) for the case of $2 \times 2 \times 2$.

Consider the following full model again. If the interaction terms in the full model are set to zero, we have the main effect model, $\log\left(m(x_{ijk})\right) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$. Usually the hierarchical models are of interest (Khamis, 1996) out of log-linear models.

Table 3. Hierarchical Models

Model #	**Model** $\log(m_{ijk})$
1	λ
2	$\lambda + \lambda_i^X$
3	$\lambda + \lambda_j^Y$
4	$\lambda + \lambda_k^Z$
5	$\lambda + \lambda_i^X + \lambda_j^Y$
6	$\lambda + \lambda_i^X + \lambda_k^Z$
7	$\lambda + \lambda_j^Y + \lambda_k^Z$
8	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$

9	$\lambda + \lambda_i^X + \lambda_k^Z + \lambda_{ik}^{XZ}$
10	$\lambda + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$
11	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$
12	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$
13	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}$
14	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$
15	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$
16	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$
17	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$
18	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$
19	$\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$

3. Modelling Elementary Students' Academic Performance

As stated before, after discussing the model constructions, the aim is to use these models to analyze a dataset. For this purpose, a dataset was collected about academic performances of elementary students in New Mexico State. We investigate the characteristics that influence students' academic success. First grade students' weight, height, number of hours a student watches TV per day, number of hours a student sleeps per day were collected. As a means of quantifying their academic-performance, Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scores were recorded for each student. DIBELS score is a measurement evaluation for the acquisition of early stage literacy skills for K-6 graders. Furthermore, student's Body Mass Index (BMI) was calculated for each individuals using their heights and weights. After a preliminary investigation about the collected data, student's physical health, sleeping duration, and TV time, are some of the identified indicators that can influence on students' academic performance. Furthermore, initial investigations showed that there is a significant difference in the academic performance with students with TV time one hour, BMI value of 16, and Sleeping time of eight hours. Therefore, in this study we analyze the above data to study the impact of the selected three categorical variables on the student's performance. Data analysis was conducted using R Studio (Ver. 1.0.153).

A summary of the descriptive statistics of the collected data is displayed in Table 4. Following values show the mean values of each variable and their standard devastations.

Table 4: Descriptive statistics of gathered data

Sleep Time	TV time	BMI	Avg DIBELS
8.11 (0.74)	3.78 (0.40)	17.51 (3.14)	192.24 (61.75)

For the data analysis purpose let's consider the three factors as follows:

$$X = Number\ of\ hours\ for\ \ the\ sleep\ (Less\ than\ 8\ or\ more\ than\ 8\ )$$
$$Y = Number\ of\ hours\ for\ TV\ (Less\ than\ 1\ or\ more\ than\ or\ equal\ to\ 1\ )$$

$$Z = BMI\ value\ (Less\ than\ 16\ lb\ in^{-2}\ or\ more\ than\ or\ equal\ to\ 16\ lb\ in^{-2})$$

For the coding purposes, we adhere to the following labeling to represent the data in the contingency table.

$$\begin{aligned}
x_{111} &= (Sleep < 8\ hrs, \quad TV < 1\ hrs, BMI < 16)\\
x_{112} &= (Sleep < 8\ hrs, \quad TV < 1\ hrs, BMI \geq 16)\\
x_{121} &= (Sleep < 8\ hrs, \quad TV \geq 1\ hrs, BMI < 16)\\
x_{122} &= (Sleep < 8\ hrs, \quad TV \geq 1\ hrs, BMI \geq 16)\\
x_{211} &= (Sleep \geq 8\ hrs, \quad TV < 1\ hrs, BMI < 16)\\
x_{212} &= (Sleep \geq 8\ hrs, \quad TV < 1\ hrs, BMI \geq 16)\\
x_{221} &= (Sleep \geq 8\ hrs, \quad TV \geq 1\ hrs, BMI < 16)\\
x_{222} &= (Sleep \geq 8\ hrs, \quad TV \geq 1\ hrs, BMI \geq 16)
\end{aligned}$$

Table 5. Sleeping Time, TV watching time, and BMI value

	Sleeping Time < 8			**Sleeping Time ≥ 8**		
TV Time	**BMI <16**	**BMI ≥16**	**Total**	**BMI <16**	**BMI ≥16**	**Total**
Less than 1	6	15	21	16	47	63
More than 1	10	30	40	45	37	82
Total	16	45		61	84	

Table 5 shows transformation of collected data set into the $2 \times 2 \times 2$ contingency table. With the help of the above coding, hierarchical models can be fitted for the above data. For each fitted model, table 6 displays the level of significance and some selected important criteria that helps to select the best model that fits the data.

4. Model Selection

According to the literature, model selection is based on different criterions. Akaike information criterion (AIC), Pearson chi-square test statistic and Bayesian information criterion are some of the popular criterions. According to the AIC (Akaike, 1973), typically the best model has a lower AIC value than the other models. The AIC statistic is calculated according to equation (1):

$$AIC = G^2 - 2 \times df, \tag{1}$$

where

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk} ln\left(\frac{n_{ijk}}{m_{ijk}}\right) \tag{2}$$

and n_{ijk}, m_{ijk} are observed and expected cell counts represented in the contingency table. The $G^2(2)$ statistic (equation 2) is called the Pearson chi-square test statistic or the likelihood ratio statistic. Higher values of G^2 indicate that the model fails to fit the data well. In addition to G^2statistic, Bayesian information criterion (BIC) also used for the model selection (Raftery, 1986) is given by equation (3).

$$BIC = G^2 - df \times \ln(n), \tag{3}$$

where *n* is the sample size. According to the table 6, Model #19 has the lowest values for G^2 and *BIC*. When selecting the appropriate model, saturated model (model #19) is selected in the absence of any other possible model. When fitting the model 19 for the data, it uses the same number of data points as number of parameters. As a result, the saturated model will be the perfect fit for the data, but statistically the use of the model is minimal. Therefore, in practice we use the saturated model if there is no simpler model to select. Furthermore, it is mentioned that the interpretation of the saturated model is harder with the increment of the number of the variables. By considering above and all the values of *AIC*, *G*2, and BIC, model 8 can be considered as the best model that fits the data. According to this model, it consists of all the factors, *X*, *Y*, and the interaction effect of *X* and *Y*. As the model shows, if a student who watches TV less than one hour per day and sleeps less than eight hours per day increases sleeping time for more than 8 hours per day, the log value of the DIBELS score is expect to increase by 1.10. Though the coefficient of *Y* (TV time) is positive, it is not statistically significant. In the meantime, consider a student who sleeps less than eight hours per day and watches TV less than one hour per day, if the student increases sleeping time for more than eight hours and time to watch TV for more than one hour, it is expected to decrease the log value of the DIBELS score by 0.76.

Table 6: Estimated Hierarchical Log Linear Models

Model	**Factors**	**Coefficient (Significance)**	**AIC**	G^2	**BIC**
2	*X*	0.67 ***	84.88	0.00	-31.16
3	*Y*	0.93 ***	68.96	0.00	-31.16
4	*Z*	0.93 ***	68.96	0.00	-31.16
5	*X, Y*	0.67 ***, 0.13	86.08	0.00	-25.96
6	*X, Z*	0.67 ***, 0.93***	51.93	0.17	-25.79
7	*Y, Z*	0.67***, 0.93 ***	51.93	0.00	-25.96
8	*X, Y, X*Y*	1.10***, 0.64 , -0.76 *	82.45	0.00	-20.77
9	*X, Z, X*Z*	0.78 **, 1.03***, -0.16	53.71	0.11	-20.66
10	*Y, Z, Y*Z*	0.28, 1.04***, -0.20	71.80	0.00	-20.77
11	*X, Y, Z*	0.67***, 0.13 , 0.93 ***	53.13	0.14	-20.63
12	*X, Y, Z, X*Y*	0.78**, 0.13, 1.03***,-0.16	54.93	0.08	-15.50
13	*X, Y, Z* *X*Z*	0.78 ***, 0.13 **, 1.03 *** -0.16	54.93	0.08	-15.50
14	*X, Y, Z, Y*Z*	0.67***, 0.28, 1.04 ***,-0.20	54.78	0.09	-15.49
15	*X, Y, Z* *X*Y, X*Z*	1.21 ***, 0.64 *, 1.03 *** -0.76 *, -0.16	51.30	0.57	-9.82
16	*X, Y, Z* *X*Y, Y*Z*	0.79 *, 1.04 ***, 1.20**** -0.76 *, -0.20	51.15	0.61	-9.77

17	*X, Y, Z* *X*Z, Y*Z*	0.28, 1.14 ***, -0.78 ** -0.16, -0.20	56.57	0.04	-10.35
18	*X, Y, Z* *X*Y, X*Z,* *Y*Z*	*1.25 ***, 0.82 *, 1.19 *** *-0.77 *, -0.23, -0.21*	*52.82*	0.42	-4.78
19	*X, Y, Z* *X*Y, X*Z* *Y*Z, X*Y*Z*	0.99 *, 0.51, 0.92 , -0.34 , 0.16 , 0.18 , -0.59	54.173	0.00	0.00

*p-value between 0.01 and 0.05
***p*-value between 0.001 and 0.01
****p*-value between 0 and 0.001

5. Conclusion

Log linear models are useful techniques to analyze the association among the categorical variables. Furthermore, it enable to explore the interaction effects among the considered categorical variables. This manuscript attempted to discuss the building process of log-linear models from the mathematical point of view and to illustrate how to use these models to identify the relationships among the selected categorical variables. In addition, a consideration was given to select the best model when several competitive models are available. Finally, we analyzed a data set to understand the relationship between the students' academic success and the related categorical variables. Though it is difficult to generalize these findings due to the nature of an observational study, our analysis indicates that the number of hours the students sleep and the number of hours they watch TV significantly influence students' academic performance. Though we expected that the BMI value will be a significant factor for the academic performance, this data did not show such a significance. WHO report (2004) states, hindering the growth of the brain and acquisition of cognitive and intellectual capabilities are some of the major issues that children who do not receive enough sleep face. Furthermore, the centers for disease control and prevention (CDC) indicates that the childhood obesity is a complex health issue that creates harmful effect on the body in a various ways. Therefore, it is very vital that all parents pay attention on this matter and further research is required to have a deeper investigate about the above discussed effects.

Acknowledgement

The author is thankful to Dr. P. Hadjicostas, all the anonymous referees and Dr. Aliakbar Montazer Haghighi for their valuable support, comments and suggestions for the improvement of the manuscript.

REFERENCES

Agresti, A. (1990). Categorical Data Analysis, New York: John Wiley & Sons.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. Pages 267–281 in Proceedings of the Second International Symposium on Information Theory.

Andersen, A. H. (1974). Multidimensional contingency tables. Scandinavian Journal of Statistics, 1(3), 115-127.

Bartlett, M. S. (1935). Contingency table interactions. Supplement to the Journal of the Royal Statistical Society, 2(2), 248-252. doi: 10.2307/2983639.

Beck, A. R. (2016). Psychosocial aspects of obesity. NASN Sch Nurse. 31(1):23–27.

Bishop, Y. M. (1969). Full contingency tables, logits, and split contingency tables. Biometrics, 25(2), 383-400. doi: 10.2307/2528796.

Brzezinska, J. (2012). Hierarchical log-linear models for contingency tables. Fol Oecon. 269: 123–129.

Cawley, J., Frisvold D., and Meyerhoefer C. (2013). "The Impact of Physical Education on Obesity among Elementary School Children." Journal of Health Economics, 32(4): 743-755.

Chapman, J.M., Onnie C.M., Prescott N.J, Fisher S.A., Mansfield J.C., Mathew C.G. (2009). Searching for genotype-phenotype structure: using hierarchical log-linear models in Crohn disease. Am J Hum Genet. 84:178–87. doi:10.1016/j.ajhg.2008.12.015.

Christensen, R. (1997). Log-linear models and logistic regression (2nd ed.). New York, NY: Springer. doi: 10.1007/b97647

Fienberg, S. E. and Meyer, M. M. (1983). Loglinear models and categorical data analysis with psychometric and econometric applications. *Journal of Econometrics, 22*, 191–214.

Halfon, N., Kandyce L., Slusser W. (2013). Associations between obesity and comorbid mental health, developmental, and physical health conditions in a nationally representative sample of US children aged 10 to 17. Academic Pediatrics. 13.1:6–13.

Howell, D. C. (2009). Statistical methods for psychology (7th ed.). Belmot, CA: Cengage Learning. 630–655.

Hu, J., Joshi, A., Johnson, V. E. (2009). Log-Linear Models for Gene Association. *Journal of the American Statistical Association, 104*(486), 597–607. http://doi.org/10.1198/jasa.2009.0025.

Kelderman, H. (1992). Psychometrika 57: 437. https://doi.org/10.1007/BF02295431.

Morrison, K. M., Shin S., Tarnopolsky M. (2015). Association of depression and health related quality of life with body composition in children and youth with obesity. Journal of Affective Disorders;172:18–23.

Pearson, K. (1903). Mathematical Contributions to the Theory of Evolution. XI. On the Influence of Natural Selection on the Variability and Correlation of Organs. *Philosophical Transactions of the Royal Society of London, 200* (Ser. A), 1-66. http://dx.doi.org/10.1098/rsta.1903.0001

Raftery, A. E. (1986). *Choosing models for cross-classification*, Amer. Sociol. Rev. 51, 145-146.

Rapallo, F. (2003). Log-linear Models and Toric Ideals, Doctorate of research in Mathematics and applications, University of Genoa, Italy.

WHO technical meeting on sleep and health (2004). Sleeping is required for every living organism, Bonn Germany, 22-24 January.

Zhu, B., Walter S. D., Rosenbaum P. L., Russell D. J., Raina, P. (2006). Structural equation and log-linear modeling: a comparison of methods in the analysis of a study on caregivers' health, BMC Medical Research Methodology, 6:49