Journal of Natural Sciences Research www.iiste.org
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.4, No.24, 2014

Proposed Generalized Method and Algorithms for the Estimation of Parameters and Best Model Fits of Log Linear Model

[1] OKOLI C.N; [2] ONYEAGU S.I AND [3] OSUJI G.A

[1] Department of Statistics, Anambra State University, Nigeria. [2&3] Department of Statistics, Nnamdi Azikiwe University, Awka.Nigeria

Correspondence E-mail- cecilia.okoli@yahoo.com or donfadino@yahoo.com

ABSTRACT

The paper is on proposed generalized method and algorithms developed for estimation of parameters and best model fits of log linear model for q-dimensional contingency table. For purpose of this work, the method was used to provide estimates of parameters of log –linear model for five- dimensional contingency table. In estimating these parameters and best model fit, computer programs in R were developed for the implementation of the algorithms. The iterative proportional fitting procedure was used to estimate the parameters and goodness of fits of models of the log linear model. A real life data was used for illustration and the result obtained showed the best model fit for five dimensional contingency table is [BSGM, BGAM]. This showed that the best model fit has sufficient evidence to fit the data without loss of information. This model has highest p-value and the least likelihood ratio estimate. This model also revealed that state of origin is independent of age given Bscgrade and mode of admission.

Keywords: contingency table , categorical data, hierarchical log –linear models, Parameters, proposed generalized method, algorithms, Iterative proportional fitting procedure.

1. INTRODUTION

Contingency tables are formed when a population is cross-classified according to a series of categories or factors (Agresti, 1996). Each cell count of the contingency tables gives the number of units observed under a particular cross-classification. Additionally, q- contingency table is a contingency table created by cross-classification of more than two categorical variables. A categorical variable is one for which the measurement scale consists of a set of categories (Everitt, 2006). Categorical data are data consisting of counts of observation falling in different categories. Categorical data in contingency tables are collected in many investigations. In order to understand the type of structures in prevailing data appropriate log linear models are fitted. Log linear analysis is a technique that makes no distinction between dependent and independent variables and it used to examine association or interaction among categorical variables (Jeansonne , A,2002). In log linear analysis, expected values of the observations are given by linear combination of number of parameters. The saturated model in log linear model analysis is a model that incorporates all possible effects, such as one-way effect, two way interactions effect, and three-way interactions e.t.c. A saturated model imposes no constraints on the data and always reproduced the observed counts. The parsimonious model in log linear model analysis is incomplete model that achieves satisfactory level of goodness of fit. The log linear model is called hierarchical whenever the model contains higher-order effects incorporates lower-order effects composed of the variables (Everitt, 1977). The reason for including lower-order terms is that the statistical significance and practical interpretation of higher –order terms depend on how the variables are coded. This is undesirable, but with hierarchical models, the same results are obtained, irrespective of how the variables are coded (Bishop et al, 1975).

2. MATERIALS AND METHOD

Material: The aim of this paper is to use proposed generalized method for estimation of parameters and algorithms developed implemented in R computer programs for estimating the parameters and best model fit of log linear model for five-dimensional contingency table. The data on B.Sc grade, State of origin, gender, age and mode of admission were collected from degree results of 1291 graduated students from Micro- Biology department, Anambra State University, Uli, Nigeria.

The variables and Categories

Variable I: B Sc grade were classified as: Pass; Third class; Lower division; Upper division, First class

Variable 2: State of Origin – Indigene and Non-indigene

Variable 3: Gender – Male and Female

Variable 4: Age - under 26 and 26 and 26 & over

Variable 5: Mode of admission – Jamb and Pre –science

2i. GENERALIZED Method

For any $\|[j]\| = t \quad (t \neq 0)$ and $q < \infty \quad (q \in \mathrm{N})$, we define

$$\mu_{\{j_t\}(i_{j_t})} := \log\left\{ \frac{\sum_{i_{j_q}/i_{j_t}} n_{(i_{j_t},+)}}{I_{i_{j_q}} / i_{j_t}} \right\} - \sum_{r=0}^{t-1} \sum_{\{j_s\}\in \mathrm{P}(\{j_t\}:|\{j_s\}|=r)} \mu_{\{j_s\}(i_{j_s})} \quad : \{j_t\} \in \mathrm{P}([q]:|\{j_t\}| = t) \ \ t = 1,2,\ldots,q$$

Where $[q] = \{1,2,\ldots,q\}, \ \mathrm{P}([q])$ is the power set of $[q]$

$\|[\underline{j}]\|$ is the level of interaction/cardinality of the set /length of element in the set. j is a member of collections of sequence $[q]$.

Estimation of parameters for q-dimensional contingency table

For q=5, the saturated log linear for 5-dimensional contingency table and its parameter effects is given by

$$\begin{aligned}\log m_{i_1i_2i_3i_4i_5} = \mu &+ \mu_{1(i_1)} + \mu_{2(i_2)} + \mu_{3(i_3)} + \mu_{4(i_4)} + \mu_{5(i_5)} + \mu_{12(i_1i_2)} + \mu_{13(i_1i_3)} + \mu_{14(i_1i_4)} + \mu_{15(i_1i_5)} + \mu_{23(i_2i_3)} + \mu_{24(i_2i_4)} \\ &\mu_{25(i_2i_5)} + \mu_{34(i_3i_4)} + \mu_{35(i_3i_5)} + \mu_{45(i_4i_5)} + \mu_{123(i_1i_2i_3)} + \mu_{124(i_1i_2i_3)} + \mu_{125(i_1i_2i_5)} + \mu_{134(i_1i_3i_4)} + \mu_{135(i_1i_3i_5)} \\ &+ \mu_{145(i_1i_4i_5)} + \mu_{234(i_2i_3i_4)} + \mu_{235(i_2i_3i_5)} + {}_{245(i_2i_4i_5)} + \mu_{345(i_3i_4i_5)} + \mu_{1234(i_1i_2i_3i_4)} + \mu_{1235(i_1i_2i_3i_5)} + \mu_{1245(i_1i_2i_4i_5)} \\ &+ \mu_{1345(i_1i_3i_4i_5)} + \mu_{2345(i_2i_3i_4i_5)} + \mu_{12345(i_1i_2i_3i_4i_5)}\end{aligned} \quad (2.1)$$

Conventionally, taking $\mu\phi(\phi) = \mu$ **and**

$\|[\underline{j}]\| = 0\left(i.e.[\underline{j}] = \phi\right);$

$$\hat{\mu} = \log\left(\frac{\sum_{i_1 i_2 i_3 i_4 i_5} n_{i_1 i_2 i_3 i_4 i_5}}{I_{i_1} I_{i_2} I_{i_3} I_{i_4} I_{i_5}}\right)$$

$$\left\|\left[\underline{j}\right]\right\| = 1;$$

$$\hat{\mu}_{[1](i_1)} = \log\left(\frac{\sum_{i_2 i_3 i_4 i_5} n_{(i_1,+)}}{I_{i_2} I_{i_3} I_{i_4} I_{i_5}}\right) - \hat{\mu} \qquad \hat{\mu}_{[2](i_2)} = \log\left(\frac{\sum_{i_1 i_3 i_4 i_5} n_{(i_2,+)}}{I_{i_1} I_{i_3} I_{i_4} I_{i_5}}\right) - \hat{\mu}$$

$$\hat{\mu}_{[3](i_3)} = \log\left(\frac{\sum_{i_1 i_2 i_4 i_5} n_{(i_3,+)}}{I_{i_1} I_{i_2} I_{i_4} I_{i_5}}\right) - \hat{\mu} \qquad \hat{\mu}_{[4](i_4)} = \log\left(\frac{\sum_{i_1 i_2 i_3 i_5} n_{(i_4,+)}}{I_{i_1} I_{i_2} I_{i_3} I_{i_5}}\right) - \hat{\mu}$$

$$\hat{\mu}_{[5](i_5)} = \log\left(\frac{\sum_{i_1 i_2 i_3 i_4} n_{(i_5,+)}}{I_{i_1} I_{i_2} I_{i_3} I_{i_4}}\right) - \hat{\mu}$$

$$\left\|\left[\underline{j}\right]\right\| = 2;$$

$$\hat{\mu}_{[12](i_1 i_2)} = \log\left(\frac{\sum_{i_3 i_4 i_5 i_6} n_{(i_1 i_2,+)}}{I_{i_3} I_{i_4} I_{i_5} I_{i_6}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[2](i_2)}$$

$$\hat{\mu}_{[13](i_1 i_3)} = \log\left(\frac{\sum_{i_2 i_4 i_5 i_6} n_{(i_1 i_3,+)}}{I_{i_2} I_{i_4} I_{i_5} I_{i_6}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[3](i_3)}$$

$$\hat{\mu}_{[14](i_1 i_4)} = \log\left(\frac{\sum_{i_2 i_3 i_5 i_6} n_{(i_1 i_4,+)}}{I_{i_2} I_{i_3} I_{i_5} I_{i_6}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[4](i_4)}$$

$$\hat{\mu}_{[15](i_1 i_5)} = \log\left(\frac{\sum_{i_2 i_3 i_4 i_6} n_{(i_1 i_5,+)}}{I_{i_2} I_{i_3} I_{i_4} I_{i_6}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[5](i_5)}$$

$$\hat{\mu}_{[16](i_1 i_6)} = \log\left(\frac{\sum_{i_2 i_3 i_4 i_5} n_{(i_1 i_6,+)}}{I_{i_2} I_{i_3} I_{i_4} I_{i_5}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[6](i_6)}$$

$$\hat{\mu}_{[23](i_2 i_3)} = \log\left(\frac{\sum_{i_1 i_4 i_5 i_6} n_{(i_2 i_3,+)}}{I_{i_1} I_{i_4} I_{i_5} I_{i_6}}\right) - \hat{\mu} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[3](i_3)}$$

$$\hat{\mu}_{[24](i_2 i_4)} = \log\left(\frac{\sum_{i_1 i_3 i_5 i_6} n_{(i_2 i_4,+)}}{I_{i_1} I_{i_3} I_{i_5} I_{i_6}}\right) - \hat{\mu} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[4](i_4)}$$

$$\hat{\mu}_{[25](i_2 i_5)} = \log\left(\frac{\sum_{i_1 i_3 i_4 i_6} n_{(i_2 i_5,+)}}{I_{i_1} I_{i_3} I_{i_4} I_{i_6}}\right) - \hat{\mu} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[5](i_5)}$$

$$\hat{\mu}_{[34](i_3 i_4)} = \log\left(\frac{\sum_{i_1 i_2 i_5} n_{(i_3 i_4,+)}}{I_{i_1} I_{i_2} I_{i_5}}\right) - \hat{\mu} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[4](i_4)}$$

$$\hat{\mu}_{[35](i_3 i_5)} = \log\left(\frac{\sum_{i_1 i_2 i_4} n_{(i_3 i_5,+)}}{I_{i_1} I_{i_2} I_{i_4}}\right) - \hat{\mu} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[5](i_5)}$$

$$\hat{\mu}_{[45](i_4 i_5)} = \log\left(\frac{\sum_{i_1 i_2 i_3} n_{(i_4 i_5,+)}}{I_{i_1} I_{i_2} I_{i_3}}\right) - \hat{\mu} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[5](i_5)}$$

$\left\| \left[\underline{j} \right] \right\| = 3;$

$$\hat{\mu}_{[123](i_1 i_2 i_3)} = \log\left(\frac{\sum_{i_4 i_5} n_{(i_1 i_2 i_3,+)}}{I_{i_4} I_{i_5}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[12](i_1 i_2)} - \hat{\mu}_{[13](i_1 i_3)} - \hat{\mu}_{[23](i_2 i_3)}$$

$$\hat{\mu}_{[124](i_1 i_2 i_4)} = \log\left(\frac{\sum_{i_3 i_5} n_{(i_1 i_2 i_4,+)}}{I_{i_3} I_{i_5}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[12](i_1 i_2)} - \hat{\mu}_{[14](i_1 i_4)} - \hat{\mu}_{[24](i_2 i_4)}$$

$$\hat{\mu}_{[125](i_1 i_2 i_5)} = \log\left(\frac{\sum_{i_3 i_4} n_{(i_1 i_2 i_5,+)}}{I_{i_3} I_{i_4}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[5](i_4)} - \hat{\mu}_{[12](i_1 i_2)} - \hat{\mu}_{[15](i_1 i_5)} - \hat{\mu}_{[25](i_2 i_5)}$$

$$\hat{\mu}_{[134](i_1 i_3 i_4)} = \log\left(\frac{\sum_{i_2 i_5} n_{(i_1 i_3 i_4,+)}}{I_{i_2} I_{i_5}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[13](i_1 i_3)} - \hat{\mu}_{[14](i_1 i_4)} - \hat{\mu}_{[34](i_3 i_4)}$$

$$\hat{\mu}_{[135](i_1 i_3 i_5)} = \log\left(\frac{\sum_{i_2 i_5} n_{(i_1 i_3 i_5,+)}}{I_{i_2} I_{i_5}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[13](i_1 i_3)} - \hat{\mu}_{[15](i_1 i_5)} - \hat{\mu}_{[35](i_3 i_5)}$$

$$\hat{\mu}_{[145](i_1 i_4 i_5)} = \log\left(\frac{\sum_{i_2 i_3} n_{(i_1 i_4 i_5,+)}}{I_{i_2} I_{i_3}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[14](i_1 i_3)} - \hat{\mu}_{[15](i_1 i_5)} - \hat{\mu}_{[45](i_4 i_5)}$$

$$\hat{\mu}_{[234](i_1 i_3 i_4)} = \log\left(\frac{\sum_{i_1 i_5} n_{(i_2 i_3 i_4,+)}}{I_{i_1} I_{i_5}}\right) - \hat{\mu} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[23](i_2 i_3)} - \hat{\mu}_{[24](i_1 i_4)} - \hat{\mu}_{[34](i_3 i_4)}$$

$$\hat{\mu}_{[235](i_1 i_3 i_5)} = \log\left(\frac{\sum_{i_1 i_4} n_{(i_2 i_3 i_5,+)}}{I_{i_1} I_{i_4}}\right) - \hat{\mu} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[23](i_2 i_3)} - \hat{\mu}_{[25](i_1 i_5)} - \hat{\mu}_{[35](i_3 i_5)}$$

$$\hat{\mu}_{[245](i_1 i_4 i_5)} = \log\left(\frac{\sum_{i_1 i_3} n_{(i_2 i_4 i_5,+)}}{I_{i_1} I_{i_3}}\right) - \hat{\mu} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[24](i_2 i_4)} - \hat{\mu}_{[25](i_1 i_5)} - \hat{\mu}_{[45](i_4 i_5)}$$

$$\hat{\mu}_{[345](i_3i_4i_5)} = \log\left(\frac{\sum_{i_1i_2} n_{(i_3i_4i_5,+)}}{I_{i_1}I_{i_2}}\right) - \hat{\mu} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[34](i_3i_4)} - \hat{\mu}_{[35](i_3i_5)} - \hat{\mu}_{[45](i_4i_5)}$$

$\left\|\left[\underline{j}\right]\right\| = 4;$

$$\hat{\mu}_{[1234](i_1i_2i_3i_4)} = \log\left(\frac{\sum_{i_5} n_{(i_1i_2i_3i_4,+)}}{I_{i_5}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[12](i_1i_2)} - \hat{\mu}_{[13](i_1i_3)} - \hat{\mu}_{[14](i_1i_4)} - \hat{\mu}_{[23](i_2i_3)}$$

$$-\hat{\mu}_{[24](i_2i_4)} - \hat{\mu}_{[34](i_3i_4)} - \hat{\mu}_{[123](i_1i_2i_3)} - \hat{\mu}_{[124](i_1i_2i_4)} - \hat{\mu}_{[134](i_1i_3i_4)} - \hat{\mu}_{[234](i_2i_3i_4)}$$

$$\hat{\mu}_{[1235](i_1i_2i_3i_5)} = \log\left(\frac{\sum_{i_4} n_{(i_1i_2i_3i_5,+)}}{I_{i_4}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[12](i_1i_2)} - \hat{\mu}_{[13](i_1i_3)} - \hat{\mu}_{[15](i_1i_5)} - \hat{\mu}_{[23](i_2i_3)}$$

$$-\hat{\mu}_{[25](i_2i_5)} - \hat{\mu}_{[35](i_3i_5)} - \hat{\mu}_{[123](i_1i_2i_3)} - \hat{\mu}_{[125](i_1i_2i_5)} - \hat{\mu}_{[135](i_1i_3i_5)} - \hat{\mu}_{[235](i_2i_3i_5)}$$

$$\hat{\mu}_{[1245](i_2i_3i_4i_5)} = \log\left(\frac{\sum_{i_3} n_{(i_1i_2i_4i_5,+)}}{I_{i_3}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[12](i_1i_2)} - \hat{\mu}_{[14](i_1i_4)} - \hat{\mu}_{[15](i_1i_5)} - \hat{\mu}_{[24](i_2i_4)}$$

$$-\hat{\mu}_{[25](i_2i_5)} - \hat{\mu}_{[45](i_4i_5)} - \hat{\mu}_{[124](i_1i_2i_4)} - \hat{\mu}_{[125](i_1i_2i_5)} - \hat{\mu}_{[145](i_1i_4i_5)} - \hat{\mu}_{[245](i_2i_4i_5)}$$

$$\hat{\mu}_{[1345](i_2i_3i_4i_5)} = \log\left(\frac{\sum_{i_2} n_{(i_1i_3i_4i_5,+)}}{I_{i_1}}\right) - \hat{\mu} - \hat{\mu}_{[1](i_1)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[13](i_1i_3)} - \hat{\mu}_{[14](i_1i_4)} - \hat{\mu}_{[15](i_1i_5)} - \hat{\mu}_{[34](i_3i_4)}$$

$$-\hat{\mu}_{[35](i_3i_5)} - \hat{\mu}_{[45](i_4i_5)} - \hat{\mu}_{[134](i_1i_3i_4)} - \hat{\mu}_{[135](i_1i_3i_5)} - \hat{\mu}_{[145](i_1i_4i_5)} - \hat{\mu}_{[345](i_3i_4i_5)}$$

$$\hat{\mu}_{[2345](i_2i_3i_4i_5)} = \log\left(\frac{\sum_{i_1} n_{(i_2i_3i_4i_5,+)}}{I_{i_1}}\right) - \hat{\mu} - \hat{\mu}_{[2](i_2)} - \hat{\mu}_{[3](i_3)} - \hat{\mu}_{[4](i_4)} - \hat{\mu}_{[5](i_5)} - \hat{\mu}_{[23](i_2i_3)} - \hat{\mu}_{[24](i_2i_4)} - \hat{\mu}_{[25](i_2i_5)} - \hat{\mu}_{[34](i_3i_4)}$$

$$-\hat{\mu}_{[35](i_3i_5)} - \hat{\mu}_{[45](i_4i_5)} - \hat{\mu}_{[234](i_2i_3i_4)} - \hat{\mu}_{[235](i_2i_3i_5)} - \hat{\mu}_{[245](i_2i_4i_5)} - \hat{\mu}_{[345](i_3i_4i_5)}$$

2ii. ITERATIVE PROPORTIONAL FITTING PROCEDURE

The iterative proportional fitting (IPF) in Deming and Stephan (1940) is used to estimate model parameters and best model fit of log linear model. This is to ensure that the expected values are obtained iteratively for model whose expected values are not directly obtainable from marginal totals of observed values.

Consider 5- factor model, without 5-factor interaction given as

$$\log m_{ijklm} = \mu + \mu_{1(i)} + \mu_{2(j)} + \mu_{3(k)} + \mu_{4(l)} + \mu_{5(m)} + \mu_{12(ij)} + \mu_{13(ik)} + \mu_{14(il)} + \mu_{15(im)} + \mu_{23(jk)} + \mu_{24(jl)}$$
$$\mu_{25(jm)} + \mu_{34(kl)} + \mu_{35(km)} + \mu_{45(lm)} + \mu_{123(ijk)} + \mu_{124(ijl)} + \mu_{125(ijm)} + \mu_{134(ikl)} + \mu_{135(ikm)}$$
$$+ \mu_{145(ilm)} + \mu_{234(jkl)} + \mu_{235(jkm)} +_{245(jkm)} + \mu_{345(klm)} + \mu_{1234(ijkl)} + \mu_{1235(ijkm)} + \mu_{1245(ijlm)}$$
$$+ \mu_{1345(iklm)} + \mu_{2345(jklm)} \tag{2.2}$$

This model is hereby used to illustrate IPF algorithm for estimating expected frequencies (m_{ijkl}).

Totals $\widehat{m}_{ijkl.}, \widehat{m}_{ijk.m}, \widehat{m}_{ij.lm}, \widehat{m}_{i.klm}$ and $\hat{m}_{.jklm}$ are characterized to be equal to the corresponding observed marginal totals $n_{ijkl.}$, $n_{ijk.m}$, $n_{ij.lm}$, $n_{i.klm}$ and $n_{.jklm}$ respectively.

To start IPF procedure, we start initial values $\widehat{m}_{ijklm}(0) = 1$ and proceed by adjusting these proportionally to satisfy the first marginal constraint $(\widehat{m}_{ijkl.} = n_{ijkl.})$, calculated from:

$$\widehat{m}_{ijklm}(1) = \frac{\widehat{m}_{ijklm(0)} n_{ijkl.}}{\widehat{m}_{ijkl.(0)}} \tag{2.3}$$

Revise expected values $\widehat{m}_{ijklm}(1)$ to satisfy the second marginal's constraint $\widehat{m}_{ijk.m} = n_{ijk.m}$, using

$$\widehat{m}_{ijklm}(2) = \frac{\widehat{m}_{ijklm}(1) n_{ijk.m}}{\widehat{m}_{ijk.m}(1)} \tag{2.4}$$

Revise expected values $\hat{m}_{ijklm(2)}$ to satisfy the marginal constraint $\hat{m}_{ij.lm} = n_{ij.lm}$, using

$$\widehat{m}_{ijklm}(3) = \frac{\widehat{m}_{ijklm}(2) n_{ij.lm}}{\widehat{m}_{ij.lm}} \tag{2.5}$$

Revise expected values $\widehat{m}_{ijklm}(3)$ to satisfy the marginal constraint $\widehat{m}_{l.klm} = n_{l.klm}$, using

$$\widehat{m}_{ijklm}(4) = \frac{\widehat{m}_{ijklm}(3) n_{l.klm}}{\widehat{m}_{i.klm}(3)} \tag{2.6}$$

Journal of Natural Sciences Research
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.4, No.24, 2014

Complete the cycle by adjusting $\hat{m}_{ijklm}(4)$ to satisfy the fifth marginal constraint $\hat{m}_{.jklm} = n_{.jklm}$, using

$$\hat{m}_{ijklm}(5) = \frac{\hat{m}_{ijklm}(4)\, n_{.jklm}}{\hat{m}_{.jklm}(4)} \qquad (2.7)$$

This five –step is repeated until convergence to desired accuracy is attained.

2iii. ALGORITHMS FOR ESTIMATION OF PARAMETERS AND MODEL FITS IN LOG LINEAR MODELS FOR Q-DIMENSIONAL CONTINGENCY TABLES

The step by step approach for the estimation of parameters that will enable us to develop our R-programs for estimating parameters, expected values and goodness of log linear models are as follows:

Step 1 Given a set of observed data with variables say (2, 3, 4 ,5,… , q) . Identify the dimension of the variable.

Step 2 Estimate the parameters of the model by the method of Iterative proportional fitting (IPF)

Step 3 Estimate the expected values by the method of IPF

Step 4 Compute the model estimates using IPF

Step 5 Using R- programs, the chi squared (χ^2), log –likelihood ratio (G^2) and all the parameters of the model(s) involved are estimated.

3. DATA ANALYSIS AND RESULTS

When the data was analyzed, we obtained this table of output

TABLE 1 SUMMARY OF THE RESULTS OF GOODNESS OF FITS FOR 5-DIMENSIONAL CONTINGENCY TABLE DATA

Model	G^2	χ^2	d.f	p-value
[B,S,G,A,M]	172.9839	176.3949	71	<0.001
[S,G,A,BM]	141.2813	139.1354	67	<0.001
[S,A,M,BG]	139.1010	141.8967	67	<0 .001
[B,G,M,BA]	166.3272	171.4614	67	0
[B,G,M,SA]	172.9646	176.5658	70	0
[B,G,A,SM]	171.2029	174.1919	70	0
[B,S,M,GA]	172.9732	170.3752	70	0
[B,S,A,GM]	172.9719	176.5359	70	0
[B,S,G,AM]	172.8484	175.9629	70	0
[B,M,GA,BS]	127.9595	131.5539	63	<0.001
[S,A,BM,BG]	125.7791	134.4006	63	< 0.001
[S,G,BM,BA]	153.0054	174.6249	63	<0.001

[A,M,BM,SG]	154.6406	163.1955	66	<0.001
[G,BM,SA]	159.6427	173.5571	66	0
[G,A,BM,SM]	157.8861	172.8167	66	0
[S,BM,GA]	159.6514	173.3912	66	0
[S,A,BM,GM]	159.6501	173.3092	66	0
[S,G,BM,AM]	159.5265	172.8867	66	0
[S,BM,BG,BA]	119.1225	132.3240	59	0
[A,BM,BG,SG]	120.7577	128.7868	62	0
[BM,BG,SA]	125.7598	134.5223	62	0
[A,BM,BG,SM]	124.0031	133.9196	62	0
[S,BM,BG,GA]	125.7684	134.4744	62	<0.001
[S,A,BM,BG,GM]	125.3519	134.4098	62	0
[S,BM,BG,AM]	125.6436	134.2637	62	0
[BM,BG,BS,BA]	87.41988	93.03104	55	0.003528851
[A,BM,BG,BS,SG]	91.75409	94.07371	58	0.003132519
[BM,BG,BS,SA]	94.05723	96.53581	58	0.001922623
[BM,BG,BS,SM,A]	93.04289	96.43197	58	0.002388153
[BM,BG,BS,GA]	94.0658	96.45887	58	0.001919054
[BM,BG,BS,GM,A]	93.64933	95.98648	58	0.002095519
[BM,BG,BS,AM]	93.94103	96.38938	58	0.001971264
[A,M,BSG]	104.3455	105.8493	58	0.000432178
[G,M,BSA]	131.8655	127.2042	58	0
[G,A,BSM]	100.7154	99.0451	58	0.0004329178
[S,A,BGA]	124.9574	123.9418	58	0
[S,A,BGM]	113.5894	125.3055	58	0
[S,G,BAM]	145.6230	155.5872	58	0
[B,M,SGA]	167.5483	167.5230	67	0
[B,A,SGM]	163.4812	161.3538	67	0
[B,G,SAM]	170.8800	174.4700	67	0
[M,BSGA]	83.86091	78.14037	39	0
[A,BSGM]	41.96151	40.78206	39	0.3437295
[G.BSAM]	83.35372	79.69283	39	0.000464587
[S,BGAM]	85.02937	87.22491	39	0
[B,SGAM]	161.2704	158.5060	60	0
[BSGA,BSGM]	21.47697	20.70749	20	0.3695523
[BSGA,BGAM]	43.93290	42.81167	20	0.00156376
[BSGA ,SGAM]	77.58305	72.17188	32	0
[BSGM,BSAM]	24.60369	23.95200	20	0.2170156
[BSGM,BGAM]	13.40143	13.44084	20	0.8595041
[BSGM,SGAM]	39.75069	38.61374	32	0.16305041
[BSAM,BGAM]	22.76011	22.69919	20	0.3007267
[BSAM,SGAM]	73.74410	70.67186	32	0
[BGAM,SGAM]	73.90678	76.54197	32	0
[BSGAM]	0	0	O	1

P-value for G^2

4. CONCLUSION

We proposed generalized method and algorithms developed for estimation of the parameters and best model fit of log linear model for q-dimensional contingency table. Five-dimensional contingency table was considered for this paper. In estimating these parameters and best model fit, computer programs in R were developed for the implementation of the algorithms. The Iterative proportional fitting procedure was used to estimate the parameters and models of the log linear model.

The results of the real life data analysis showed the best model fit for 5-dimensional contingency table is $[\text{BSGM, BGAM}]$. This showed that the best model fit has sufficient evidence to fit the data without loss of information. This model also revealed that state of origin is independent of age given BScgrade and mode of admission.

We also discovered that the results of the goodness of fit showed that the best model adequately fit the data set having highest p-value and the least likelihood ratio estimate. The value of the best of fit statistics are: Likelihood ratio (G^2) = 13.40143; d.f =20 ; p –value = 0.8595041

Pearson Chi-square (χ^2) = 13.44084; d.f =20 ; p –value = 0.859041

REFERENCES

[1] Agresti, A. (1996). An Introduction to Categorical Data Analysis. John Wiley & Sons. New York. Pg 297-309

[2] Bishop et al. (1975). Discrete multivariate analysis: Theory and Practice. Cambridge, Mass MIT press Pg 18-37.

[3] Deming , W.E and Stephan F.F (1940) . On least square adjustment of a frequency tables when the expected marginal totals are known. Annuals of Mathematical stat. 11, 427-444.

[4] Everitt, B.S (2006) The Cambridge dictionary of Statistics. Third edition, Cambridge university press. Pg 63-78

[5] Everitt, B.S (1977) . The Analysis of contingency tables. London: Chapman and Hall.pg 74-91

[6] Jeansome, A. (2002). Log –linear models. Retrieved October 25, 2009 from

http:// userwww.edu/-efc/classes/bio1710/ log-linear/log%20linear%20models.htmS

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Academic conference: http://www.iiste.org/conference/upcoming-conferences-call-for-paper/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar