APPLICATION OF LOG-LINEAR MODELS IN ANALYSIS OF STUDENTS' MATHEMATICS ANXIETY

NOR HAFIZAH BINTI ABD MUJIB

A dissertation submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Science (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

JANUARY 2017

To my beloved mom, dad, husband and kids.

ACKNOWLEDGEMENT

"Say, If the sea were ink for [writing] the words of my Lord, the sea would be exhausted before the words of my Lord were exhausted, even if We brought the like of it as a supplement." (Qur'an 18:109)

Alhamdulillah, all praises be to Allah, to Whom I am grateful for guidance in this journey to seek His knowledge. The more I learn, the more I realized how much His 'words' that I don't know. Through this research I can't be more amazed that how perfectly and meticulously He has designed even as tiny as a cell to sustain life.

I am thankful to my supervisor, Associate Prof. Dr. Ismail Bin Mohamad for all his patience, guidance and support during the execution of this project. Through his expert guidance, I was able to overcome all the obstacles that I encountered in these enduring ten months of my project. In fact, he always gave me immense hope every time I consulted with his over problems relating to my project.

I would like to express my gratitude to those who have encouraged and guided me to complete this thesis. I am most indebted to my parents, my husband, thanks for the love, friendship and for standing together through thick and thin. To my lovely kids, siblings and family, thank you for their endless love, encouragement and prayer. My deep appreciation goes to my best friends; Liza and Kak Aili, for your warmth and unwavering support, you guys is such a great friend. I never feel alone walking through this road with your company.

I also wish to show my appreciation to those who directly or indirectly involved in finishing this thesis. Without them, this thesis would not have been the same as presented here.

Finally I would like to express my gratitude to Ministry of Education Malaysia for the scholarship.

ABSTRACT

A log-linear modeling for three-dimensional contingency tables is used with categorical variables. In this study, hierarchical log-linear models are used to fit the observed frequencies and to determine the suitable model for the expected frequencies of pre-university students' mathematics anxiety data. For this reason, 546 students were selected to complete mathematics anxiety scale questionnaire. Data on gender, course, grade and level of mathematics anxiety were examined for presence of association using log-linear models. Estimating log-linear model parameters is carried out using Maximum Likelihood method, where the Newton-Raphson iteration method is used to find the numerical estimates of the parameters. Selection of the best model is conducted using deviance of the models. The significance of the model is determined using Goodness of Fit Test. We also determine the odds ratio between gender, course, *Sijil Pelajaran Malaysia* (SPM) Additional Mathematics grade and anxiety level to establish the risk among the groups. The final model shows that gender, course and SPM Additional Mathematics score play a role in determining the students' level of mathematics anxiety.

ABSTRAK

Model log-linear untuk jadual kontigensi tiga dimensi digunakan dengan pemboleh ubah kategorikal. Dalam kajian ini, model log-linear berhierarki telah digunakan untuk disuaikan dengan kekerapan cerapan dan untuk menentukan model yang sesuai dengan kekerapan jangkaan bagi data keresahan terhadap matematik untuk pelajar pra-universiti. Bagi tujuan ini, 546 pelajar telah dipilih untuk menjawab soal selidik mengenai keresahan terhadap matematik. Data berkenaan jantina, kursus, gred dan tahap keresahan terhadap matematik telah dikaji menggunakan model log-linear jika terdapat hubungan antara pemboleh ubah tersebut. Jangkaan parameter model log-linear diperolehi dengan menggunakan kaedah Kebolehjadian Maksimum, di mana kaedah pelelaran *Newton-Raphson* digunakan untuk mencari jangkaan berangka bagi parameter tersebut. Pemilihan model terbaik telah dijalankan menggunakan devian untuk model itu. Model yang signifikan telah diperolehi dengan menggunakan Ujian Kebagusan Penyuaian. Kami juga menentukan nisbah kemungkinan antara jantina, kursus, gred Matematik Tambahan Sijil Pelajaran Malaysia (SPM) dan tahap keresahan terhadap matematik untuk menentukan risiko di kalangan kumpulan. Model terakhir menunjukkan bahawa jantina, kursus dan gred Matematik Tambahan SPM memainkan peranan dalam menentukan tahap keresahan pelajar terhadap matematik.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

$\prod$	-	product
$\sum$	-	summation
+	-	plus
–	-	minus
×	-	multiply
%	-	percent
π	-	pi
β	-	beta
λ	-	lambda
χ^2	-	chi-squared
∂	-	delta
Ω	-	omega
e	-	exponential
P_0	-	Poisson
!	-	factorial
df	-	degree of freedom
log	-	natural logarithm
s.e	-	Standard Error
MLE	-	Maximum Likelihood Estimator
OR	-	Odds Ratio
SPM	-	*Sijil Pelajaran Malaysia*

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Log-linear modelling is among the most famous methods used to analyze frequency data. Since its first appearance, the associated methods have been developed. Log-linear modeling is being widely used in analyzing multivariate frequency tables or is called multivariate cross-classifications. Specifically, log-linear modeling is applied to identify the main effects or interactions required to express the joint distribution in the cross-classification.

Fitting log-linear models involves choices concerning the parameters that are significantly different from zero. The main candidates to be included in a model are significant parameters while non-significant parameters cost degree of freedom, but this investment fails to make contribution to the explanation of the data (Von Eye & Mun, 2013).

According to McEvoy & Richards (2001), during early 1970s, log-linear modeling procedures were developed by statisticians such as Goodman (1970) and Mosteller (1968) and have been listed into the family of general linear models (GLMs). This includes modeling procedures for example ordinary least squares regression and logistic regression (McCullagh & Nelder, 1989; Nelder & Baker, 1972). Log-linear models are generalized linear models commonly used to model the frequency in contingency table. A log-linear model consists of a number of

parameters which are combined together to determine an estimated frequency for each cell in the contingency table. The modeling is called log-linear model because the equation of a model in this form calculates the logarithm of the expected values from a linear combination of parameters.

The explanatory variables are usually categorical. The data are summarized as cross-classified table if there are only a few explanatory variables. Frequency or count in each cell of the table is the response variable. Explanatory variables are the variable used to describe the table. The log-linear model is used for the generalized linear models which can be appropriate for the study design that have some constraints on the cell frequencies (Dobson & Barnett, 2008).

Mathematics anxiety is a person's negative reaction to situations dealing with numbers, mathematics calculations and is defined as "a feeling of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problem in a wide variety of ordinary life and academic situations" (Richardson & Suinn, 1972). Mathematics anxiety reactions can go to severe from mild, from minor frustration to overwhelming emotional and physiological trouble (Ashcraft, 2002).

1.2 Background of the problem

The data of categorical variable that are collected in the form of their frequencies and summarized in contingency tables are called categorical data. A question then is how one can retrieve information from such categorical data. In statistics, linear model is used in different ways according to the context. Log-linear model is a mathematical model that takes the form of a function whose logarithm is a polynomial function of the parameters of the model.

The log-linear model is used to study behaviour selection in mathematics anxiety since the advantage of log-linear model is the ability to easily blend multiple features. Using the log-linear model can enhance decision making ability of behaviour selection (Och, 2003).

Three dimensional contingency tables are quite crucial to model because it involves many variables and parameters. The analysis of three way tables can be more complex compared to two dimensional contingency tables. The higher dimensional of contingency tables increases the complexity of the model as well as when explaining the results of analysis and interpretation.

Mathematics anxiety among Malaysian students is the main factor that influences mathematics achievement (Puteh & Khalin, 2016; Zakaria & Nordin, 2008; Davrajoo, 2007). Mathematics anxiety is actually related to many constraints in the process of learning mathematics. Previous studies from Hembree (1990), Ma (1999) and Ho *et al.* (2000) showed that students who experienced mathematics anxiety will tend to avoid mathematics courses. Therefore they will not be applying jobs that are related to mathematics skills.

1.3 Statement of the problem

Contingency tables can be modelled by many methods in generalized linear models. For two-way contingency tables, it is quite a straight forward modeling but it is more difficult or tedious to model the three dimensional and multi-dimensional contingency table using other methods other than log-linear model. Thus, the research is aimed to analyze the three dimensional contingency table by using log-linear model.

Mathematics anxiety among college students is thought to be related to some explanatory variables which can be summarized in a contingency table. Log-linear model is a way to determine this relationship and identify the effect structure of this relationship.

1.4 Objective of the study

The main objectives of this research are:

i. To identify factors associated to mathematics anxiety based on log-linear model.

ii. To determine the most parsimonious log-linear model in categorical data analysis for three dimensional contingency tables using deviance of the models.

iii. To determine the odds ratio between mathematics anxiety factors.

1.5 Scope of the study

This research will focus on the statistical methods involving the analysis of explanatory variables for three dimensional contingency tables. The data will be analyzed which the response and variables are categorical data whereby they will be measured on nominal scales. Observations of the data will be referred to the frequencies in the cells of the contingency tables that are summarized in a cross-classified table.

1.6 Significance of the study

Analyzing the three dimensional contingency tables by using log-linear models is the main objective proposed in this research. Moreover, the study will survey the level of mathematics anxiety among pre-university students. The usefulness of the suitable model chosen is hoped to be beneficial to the education area. Hopefully, the outcome result can provide some useful information regarding this issue to improve the students' mathematics performance in college.

1.7 Thesis Outline

This dissertation contains five chapters. Chapter 1 deals with the introduction to the research. It discusses the background of the problem, problem statement, objectives of the study, scope of study and significance of the study. Chapter 2 is literature review. It discusses the categorical data, background of log-linear models, the advantages and application of log-linear models and previous study on mathematics anxiety. In chapter 3, the dissertation is motivated by discussing the general methodology to model frequency in contingency tables, probability distribution and log-likelihood functions. Log-linear models for three dimensional are presented and how to select the parsimonious model that represent the relationships between the variables. The results will be presented in Chapter 4. Finally, chapter 5 presents the conclusion and some recommendation for future study.

REFERENCES

Abidin, R. A. Z., Alwi, F., & Jaafar, N. F. M. (2010). A case study on mathematics anxiety for mathematics of finance course. *Online Submission*, *7*(11), 1-11.

Agresti, A. (2007). *An introduction to categorical data analysis*: Second Edition: Wiley: New Jersey.

Agresti, A. (2013). *Categorical data analysis* Third Edition: John Wiley & Sons, Inc. Publication, Hoboken, New Jersey, *USA*.

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current directions in psychological science*, *11*(5), 181-185.

Bangar, Y., Dohare, A., Kolekar, D., Avhad, S., & Khan, T. (2015). Seasonal variation in morbidity pattern in cattle by log-linear model approach. *Journal of Applied Animal Research*, *43*(3), 283-286.

Bishop, Y. M., & Fienberg, S. E. (1969). Incomplete two-dimensional contingency tables. *Biometrics*, 119-128.

Bolarinwa, I., & Bolarinwa, B. (2013). Log linear modeling. *International Journal of Advanced Scientific and Technical Research*, *3*(1), 587-595.

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

Chambers, R. L., & Welsh, A. H. (1993). Log-linear models for survey data with non-ignorable non-response. *Journal of the Royal Statistical Society. Series B (Methodological)*, 157-170.

Christensen, R. (2006). *Log-linear models and logistic regression*: Springer Science & Business Media.

Davrajoo, E. (2007). *Mathematics anxiety and its relationship with form four students' achievement in the klang district, malaysia* (Unpublished doctoral dissertation, Universiti Putra Malaysia).

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and brain functions*, *8*(1), 1.

Dobson, A. J., & Barnett, A. (2008). *An introduction to generalized linear models*. Third Edition., Chapman & Hall, United States of America.

Everitt, B.S. (1992). *The analysis of contingency tables*. Second Edition., Chapman & Hall, London.

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage.

Fienberg, S. E., & Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, *137*(11), 3430-3445.

Goodman, L. A. (1970). The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association* 65(329), 226–256.

Grades for SPM subjects. (2016, August 15). Retrieved from http://jpnkelantan.moe.gov.my/v1/images/perkhidmatan/KEPUTUSAN%20SPM%202014.pdf

Haque, M. M., Chin, H. C., & Debnath, A. K. (2012). An investigation on multi-vehicle motorcycle crashes using log-linear models. *Safety science*, *50*(2), 352-362.

Hayati, S. (2007). *Log-Linear Model and Iterative Proportional Fitting Algorithm in the Analysis of Categorical Data.* (Unpublished Master's Dissertation). Universiti Teknologi Malaysia, Johor.

Haynes, A. F., Mullins, A. G., & Stein, B. S. (2004). Differential models for math anxiety in male and female college students. *Sociological Spectrum*, *24*(3), 295-318.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for research in mathematics education*, 33-46.

Ho, H.-Z., Senturk, D., Lam, A. G., Zimmer, J. M., Hong, S., Okamoto, Y., Wang, C.-P. (2000). The affective and cognitive dimensions of math anxiety: A cross-national study. *Journal for research in mathematics educationuij*, 362-379.

Hopko, Derek R., Rajan Mahadevan, Robert L. Bare, and Melissa K. Hunt. (2003). The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment* 10, no. 2: 178-182.

Jansen, B. R., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences, 24*, 190-197.

Jeansonne, A. (2002). Loglinear models. (2016, April 14). Retrived from http://userwww.sfsu.edu/~efc/classes/biol710/loglinear/Log% 20Linear, 20.

Kalayci, N., & Basaran, M. A. (2014). A combined approach using multiple correspondence analysis and log-linear models for student perception in quality in higher education. *Procedia Economics and Finance*, *17*, 55-62.

Kamal, I. S. M., Mahmud, Z., Sulong, S., & Azid, N. N. N. (2014). Examining Potential Risk Factors to Acute Pancreatitis Disease: A Comparison of Loglinear Models in a Malaysian Case Study. *Journal of Medical Sciences, 14*(4), 153.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and psychological measurement*, *30*(3), 607-610.

Lawal, B. (2003). *Categorical data analysis with SAS and SPSS applications*. London: Lawrence Erlbaum Associates.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for research in mathematics education*, 520-540.

Maloney, E. A., Waechter, S., Risko, E. F., & Fugelsang, J. A. (2012). Reducing the sex difference in math anxiety: The role of spatial processing ability. *Learning and Individual Differences*, *22*(3), 380-384.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models.* New York: Chapman and Hall.

McEvoy, P., & Richards, D. (2001). Using log-linear models to analyse categorical data. *Nursing Times Research, 6*(5), 867-875.

Mohd Rustam, M. R. (2013). *Pengaruh orientasi matlamat pencapaian dan keresahan Matematik terhadap pencapaian Matematik pelajar*. (Unpublished master's thesis). Universiti Teknologi Malaysia, Skudai, Johor.

Moreira, E. E., Paulo, A. A., Pereira, L. S., & Mexia, J. T. (2006). Analysis of SPI drought class transitions using loglinear models. *Journal of Hydrology*, *331*(1), 349-359.

Moreira, E. E., Coelho, C. A., Paulo, A. A., Pereira, L. S., & Mexia, J. T. (2008). SPI-based drought category prediction using loglinear models.*Journal of Hydrology*, *354*(1), 116-130.

Moreira, E. E. (2015). SPI drought class prediction using log-linear models applied to wet and dry seasons. *Physics and Chemistry of the Earth, Parts A/B/C*. doi: http://dx.doi.org/10.1016/j.pce.2015.10.019

Mosteller, F. (1968). Association and estimation in contingency tables. *Journal of the American Statistical Association*, *63*(321), 1-28.

Nelder, J. A., & Baker, R. J. (1972). Generalized linear models. *Encyclopedia of statistical sciences*.

Och, F. J. (2003). Minimum error rate training in statistical machine translation. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*-Volume 1.

Olmuş, H., & Erbaş, S. (2012). Analysis of traffic accidents caused by drivers by using Log-linear models. *PROMET-Traffic&Transportation, 24*(6), 495-504.

Prima V., Tutut, H., Muhamad, N.A.W., Ahmad O. & Suriya K.S. (2010). Exploring Mathematics Anxiety Among Engineering Students. *Procedia Social and Behavioral Sciences* (8). (pp.482 – 489)

Puteh, M., & Khalin, S. Z. (2016). Mathematics Anxiety and Its Relationship with the Achievement of Secondary Students in Malaysia. *International Journal of Social Science and Humanity*, *6*(2), 119.

Radelet, M. L., & Pierce, G. L. (1991). Choosing those who will die: Race and the death penalty in Florida. *Fla. L. Rev.*, *43*, 1.

Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: psychometric data. *Journal of counseling Psychology, 19*(6), 551.

Sanusi, W., & Ibrahim, K. (2012). Application of loglinear models in estimating wet category in monthly rainfall. *Sains Malaysiana, 41*(11), 1345-1353.

Sekaran, U. (2010). *Research methods for business: A skill building approach*. 5th ed., John Wiley & Sons.

Skinner, C., & Shlomo, N. (2008). Assessing identification risk in survey microdata using log-linear models. *Journal of the American Statistical Association, 103*(483), 989-1001.

Tabachnick, B.G. & L.S. Fidell. (1996). *Using multivariate statistics*. 3rd Edition. Harper Collins. New York, New York, USA.

Tanaka, N., Kinoshita, T., Asada, T., & Ohashi, Y. (2003). Log-linear models for assessing gene–age interaction and their application to case-control studies of the apolipoprotein E (apoE) gene in Alzheimer's disease. *Journal of human genetics, 48*(10), 520-524.

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. *Southeast Asian journal of tropical medicine and public health, 36*(5), 1283.

Von Eye, A., & Mun, E. Y. (2013). Effects in a Table. *Log-Linear Modeling: Concepts, Interpretation, and Application*, 13-22.

Wahid, S. N. S., Yusof, Y., & Razak, M. R. (2014). Math anxiety among students in higher education level. *Procedia-Social and Behavioral Sciences, 123*, 232-237.

Yenilmez, K., Girginer, N., & Uzun, O. (2007). Mathematics anxiety and attitude level of students of the Faculty of Economics and Business Administrator; The Turkey Model. In *International Mathematical Forum* (Vol. 2, No. 41, pp. 1997-2021).

Zakaria, E., & Nordin, N. M. (2008). The effects of mathematics anxiety on matriculation students as related to motivation and achievement. *Eurasia Journal of Mathematics, Science & Technology Education*, *4*(1), 27-30.

Zakaria, E., Zain, N. M., Ahmad, N. A., & Erlina, A. (2012). Mathematics anxiety and achievement among secondary school students. *American Journal of Applied Sciences*, *9*(11), 1828.

Zettle, Robert D. and Susan J. Raines. (2000). The relationship of trait and test anxiety with mathematics anxiety. *College Student Journal* 34:246–58.

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*(1), 1.