The Application in Log Linear Regression Model

Zhaoyang Qin[1,a,*]

[1] Northern Illinois University, College of Liberal Arts and Sciences, 1425 Lincoln Hwy, DeKalb, United States
a. z1857466@students.niu.edu
*corresponding author

Keywords: log linear regression model, tree diameter, the coarse root biomass, lead-contaminated house dust, children's blood-lead levels

Abstract: As we all know, the log linear regression model is a model that the logarithm of the independent variables or dependent variables or both fit a linear relationship. In most cases, the log linear regression model is appropriate for multivariable rather than univariate analysis. Although the general linear regression is heard more often, the degree of fitting of log linear regression may be better to some extent. The model is widely used in plenty of fields of research from the past to present. In this paper, I will present parts of the applications of the model and compare if the model has good fitting degrees in some specific regions.

1. Introduction

The log linear regression model [1] analysis is a method that can effectively create a relationship between several variables. In this model, the analysis always adds a hypothesis test in order to decrease the errors. In the model, the variables are treated the same. The advantage of this model is that it solves some problems that the associations of multi-variables are not linear, especially the exponential relationships, which is easy and clear to understand and learn. Several assumptions are needed before making this model. First is that the observations should be independent and random. Secondly, observed frequencies are normally distributed about expected frequencies over repeated samples [2]. Thirdly, the logarithm of E(y), where y is the response variable, is a linear combination of xi, the independent variables. Finally, the data are supposed to be categorial. There is no doubt that the log-linear regression model is widely used in lots of fields in the world.

2. Applications

1. The log linear regression can be used in sports. It can be used in sports to analysis the popularity of a certain game. In "Demand for Football and Intramatch Winning Probability: An Essay on The Glorious Uncertainty of Sports", Falter other three researchers [3] analyses three factors which have influence on the attendance: the socioeconomic variables the purely football related variables and the "incentives variables". In this process, they change the attendance into the logarithm of the attendance. Therefore, the author could use the linear regression to make the

model. As a result, they found that socioeconomic variables and coefficient of the unemployment rate have a conspicuous positive effect on the attendance. And the mean wage (economic context) has a obvious negative influence on the attendance.

2. The log linear regression model can be applied to the management. In "Forecasting Patent Filings at the European Patent Office (EPO) with a Dynamic Log Linear Regression Model: Applications and Extensions", Hingley and Park [4] wanted to research a model of European Patent Office(EPO) filings and predict the EPO filings in the long-run. In the research, the author find a log linear relationship: the logarithm of ratio of the number of EPO filings filed by a source country and the number of workers in the source country is the log linear combination of the lags of one and two year, the ratio of R&D expenditures and the number of workers, the ratio of the "trend" level of output and the number of workers and a linear relationship with a business cycle indicator. One of the advantages of this model is that it can be applied to year-to-year distinctions. And they found that the level of this prediction depends on explanatory variables for GDP and R&D.

3. The log linear regression model can be applied to biography. In "The Allometry Of Coarse Root Biomass: Log-Transformed Linear Regression or Nonlinear Regression". Lai, Yang, Lin, Kerkhoff and Ma [5] wanted to study the relationship between tree diameter and the coarse root biomass. At first, they cite a power-law relationship which is $y=a*(x^b)$, so that the author made it to a log linear regression to change the multiple into plus to make the regression easily. And then, the author tried to compare it with a nonlinear model and concluded that the relationship between tree diameter and the coarse root biomass satisfy a log linear regression model. Additionally, if researchers do not correctly use the non-linear model, there will be an error to estimate the related variables.

4. In "Improved Log-Linear Model Estimators of Abundance in Capture－Recapture Experiments", Rivest and Lévesque [6] firstly made log linear regression models to estimate a specific population. Based on this data analysis, they created a new log-linear regression between animal heterogeneity and a behavioral response. Besides, they made an assessment about the models using a general formula. Further, they tried to make a modification about reducing the asymptotic biases and evaluate the modifications. In the end, they successfully reduced the asymptotic biases and the MSE of the regression models.

5. The log linear regression can also be used in Psychometrika. In "Defining a Family of Cognitive Diagnosis Models Using Log-Linear Models with Latent Variables", Henson, Templin and Willse [7] tried to create a family of cognitive diagnosis models. They discussed the log-linear regression model between attribute mastery and the probability by using the latent variables since it is a common model to analyse the problem. Based on this model, other new methods to solve the problem had also been argued.

6. The log-linear model was once used in diseases in animals. In "Seasonal Variation in Morbidity Pattern in Cattle by Log-Linear Model Approach", Bangar, Dohare, Kolekar, Avhad [8] and Khan analyzed the seasonal variations in pattern of various diseases in cattle of Pune division of Maharashtra state of India between January 2010 and December 2010,the 11years.In this process, they used the log-linear regression model based on the data of more than 1500 cattle that was random selected on livestock owners. And then they found that 342 cattle were infected the diseases. As for the season variations, they found that winter is the highest season that the cattle was infected the digestive disease. But for the reproductive diseases, the infection in summer is higher than winter. In the end, they suggested farmers should learn about the season variations of different kinds of diseases and prevent the diseases in advance.

7. The linear-regression model can also be used in the traffic safety. In "Analysis on Reckless Driving Behavior by Log-Linear Model", Jang [9] analyzed what factors could influenced on a driver's reckless driving behavior. In this analysis, he built a log-linear regression model to make the relationship between reckless driving behavior and the driver's characteristics(age, marital status, education level) based on categorical analysis on the data of the drivers who were on traffic accidents recorded in the Jeonbuk Province Police Department and a questionnaire on drivers taken in traffic safety education at the Jeonbuk Division of Road Traffic Safety Authority. By analysis, he found that the traffic accidents are the consequence of both the accidental and habitual reckless behaviour. As a result, he found that first, if the drivers have two years or more safety education, the probability that he suffered traffic accidents or reckless behaviour is higher. And if drivers have less households' members, the same probability is higher.

8. The log linear model can also be used in education. In "Use the General Log-linear Analysis Procedure by Fitting a Log-linear Model with Age Group and Gender as Factors for Preschool Aboriginal Teachers in Taiwan", Kuo, Wei, Lin, Huang, & Yang[10] wanted to analyse the relationship between the number of preschool Taiwanese aboriginal teacher and age group and gender. At first, they found there is no relationship between age and gender in order to eliminate interference. Then, they used the regression analysis and found a log linear model to fit the association of ratio of teachers and age, gender. As a result, they concluded that between the age of 35 and 39, the ratio of preschool Taiwanese aboriginal teacher was the highest,21.5% and teachers who were 55 or above was lowest, which is just more than 3%.As for the gender, female teachers made a significance difference.

9. The log linear model can also be used in magazine. In "An Approximate Log-Linear Model for Predicting Magazine Audiences", Dananer [11] firstly introduced a log-linear model created by Danaher in 1988 about the prediction of magazine exposure. One of the innovations about this paper is that it provides an approximation of this model which decreased much amount of computation time and cancelled the need for a large stored array. Then, the author made a comparison between the approximation and the previous model and one of Leckenby and Kishi's Dirichlet-multinomial models for equal-insertion schedules. The conclusion is that the accuracy of the approximation is between the previous model and Leckenby and Kishi's Dirichlet-multinomial models.

10. The log linear model can be used in analyzing social behavior of fishes. In "Log-linear model Based Behavior Selection Method for Artificial Fish Swarm Algorithm." First, Huang and Chen [12] stressed the importance of analysis on behavior on fishes: global exploration ability and convergence speed. And then, they created a new log-linear model to increase the decision making of the behavior selection. The model researched the relationship between the behavior selection and the state factor of fishes. Based on this model and further studying, they concluded that adaptive movement behavior with setting according to diversity of fish swarm could be able to enhance the performance of artificial fish swarm and the inhibition behavior could quicken the convergence speed.

11. The log linear model could also be used in health and prevention of illness. In "A Log-Linear Model for Predicting Risk Factors for Rabies Positivity in Raccoons in Virginia, 1984-1987", Torrence and Elizabeth [13] developed a log-linear model to study the epidemiology of rabies in raccoons in Virginia and decide the risk to predict positivity for rabies. They found 11 explanatory variables and 8 of them were announced to be risk factors in the prediction. In order to make this model valid, they used the data of 1988 and part of 1989 to make the hypothesis test at 99% level. In the end, they concluded that three second order variables were obvious to influence the risks. Between the year 1984 and 1987, Virginia witnessed a small rabies case.

12. The log linear model can be used in agriculture. In "Worldwide Pattern of Multilocus Structure in Barley Determined by Discrete Log-Linear Multivariate Analyses", Zhang, Maroof, Allard [14] made a discrete log linear model to analyze the Hordeum vulgare L, based on the data from the electrophoretic assay for seven enzyme loci of 1,032 accessions of cultivated barley. In their analysis, they used three steps to accomplish their goal. Firstly, find and cancel the inconsequential effects. And then, they tried to make the model the best one to explain the frame of the genetic system. Finally, evaluate all terms in the model to make it perfect. From this model, they concluded that cultivated barley is layered into structured complexes of loci and in different areas, Multilocus structure are not the same. With the increasing distance of Southwest Asia, such difference increased as well.

13. The log linear model can also be used in congenital malformations. In "Using Log-Linear Models to Test for Associations Among Congenital Malformations", Beaty, Yang, Khoury, Harris and Liang [15] made a log-linear model to test the relationship between birth defects or congenital malformations, based on the data from more than 15000 babies between 1968 and 1986. As a result, in some defects, they found that there are significance association, but among others, there is no heavy relationship. And then, they tested a simpler log linear model, and presented some pairwise association among anal malformations. Other than the log-linear model, they also made some cluster analysis to show the relationship of three specific malformations.

14. The log-linear model can be used in concern about natural environment. In "Use of Loglinear Models to Assess Factors Influencing Concern for The Natural Environment.", LAKHAN, D. LAVALLE [16] used the log-linear regression model to analyze the interactions based on the multi-dimensional data. By the interviews over 1500 people, they collected the data and applied it to the log-linear model and decided which factors among education, age, residential location, and gender have significant concern about the natural environment. As a result, they found that education is the factor that had direct relationship with people's concern about the environment. And then, they found that age had associations with education, which means to some extent, age has influenced on the environmental concern. As for age, the younger the interviewer was, the more environmental concern he had. For other variables, like residential location and gender, they had little difference on the people's environmental concern.

15. The log linear model can also be used in analyzing human errors. In "The Application of The Loglinear Model to Quantify Human Errors", Lin and Hwang [17] focused on creating a log linear model to study the human errors when people are in a emergency, which offers a way to find the factors resulting in those errors. By using this model, they found that the external performance shaping factor (PSF), internal PSF, and stressor PSF had significance influence on human errors. What's more, the first-order relationship between external PSF and stressor PSF, and the second order association with the three variables also effected human errors to some extent. In the end, they successfully used the log linear model to get the maximum likelihood probability of these human errors.

16. The log linear model can also be used in children's health. In "Log-Additive versus Log-Linear Analysis of Lead-Contaminated House Dust and Children's Blood-Lead Levels", Rust, Burgoon, Lanphear and Eberly [18] made a comparison between log linear model and the log additive model. They used two methods to collect dusts, BRM and the wipe method and found that the log-linear model was better than the log additive model to describe the diversity in children's blood-lead levels. The log additive model can only describe the limited rise in the probability when the blood levels are more than 10 μg/dl, while the log linear model can explain the obvious increase. In the end, they concluded that the log linear model was the regular model to promote the dust standard.

3. Conclusions

The log-linear model can be applied to many fields, such as sports, health, natural environment, disease, agriculture and human behavior. In these areas, a few fittings are better. For instance, when studying the relationship between tree diameter and the coarse root biomass, the log linear model is superior than the non-linear model. And while finding the association between lead-contaminated house dust and children's blood-lead Levels, the log linear model, which is a better choice can explain far more information than log additive model. But flaw can exist when using the log linear model. When predicting the magazine audiences, researchers found that the accuracy of the approximation of the log linear model is between the previous log model and Leckenby and Kishi's Dirichlet-multinomial models. In a word, the application of the log linear model is wide range.

References

[1] Christensen, R. (1997), Log-Linear Models and Logistic Regression (2nd ed.). Springer.

[2] Field, A. (2005), "observed frequencies are normally distributed about expected frequencies over repeated samples" Discovering statistics using SPSS (2nd ed.)

[3] Falter, Jean-Marc, Perignon, Christophe (2000), Demand for football and intramatch winning probability: an essay on the glorious uncertainty of sports.

[4] Peter Hingley, Walter Park (2016), Forecasting Patent Filings at the European Patent Office (EPO) with a Dynamic Log Linear Regression Model: Applications and Extensions

[5] Jiangshan Lai, Bo Yang, Dunmei Lin, Andrew J. Kerkhoff, Ma (2013), The allometry of coarse root biomass: Log-transformed Linear Regression or nonlinear regression

[6] Louis-Paul Rivest and Tina Lévesque (2001), Improved log-linear model estimators of abundance in capture-recapture experiments

[7] Robert A. Henson, Jonathan L. Templin, John T. Willse (2009), Defining a Family of Cognitive Diagnosis Models Using Log-Linear Models with Latent Variables

[8] Bangar, Y.C. Dohare, A.K. Kolekar, D.V. Avhad, S.R. et al (2015), Seasonal variation in morbidity pattern in cattle by log-linear model approach

[9] Tae Youn Jang, (July 2006), "categorical analysis on the data of the drivers who were on traffic accidents recorded in the Jeonbuk Province Police Department and a questionnaire on drivers taken in traffic safety education at the Jeonbuk Division of Road Traffic Safety Authority", Analysis on reckless driving behavior by log-linear model.

[10] Lung-Hsing Kuo, Huei-Mei Wei, Lydia Lin, Hsun-Chih Huang, Hung-Jen Yang (2012), Use the general log-linear analysis procedure by fitting a log-linear model with age group and gender as factors for preschool aboriginal teachers in Taiwan

[11] Peter J. Danaher (1989), "approximation", "one of Leckenby and Kishi's Dirichlet-multinomial models for equal-insertion schedules", An Approximate Log-Linear Model for Predicting Magazine Audiences

[12] Huang Zhehuang, Chen Yidong (2015), "global exploration ability and convergence speed", "adaptive movement behavior with setting according to diversity of fish swarm", Log-Linear Model Based Behavior Selection Method for Artificial Fish Swarm Algorithm

[13] Torrence, Mary Elizabeth (1990), A log-linear model for predicting risk factors for rabies positivity in raccoons in Virginia, 1984-1987

[14] Qifa Zhang, M. A. Saghai Maroof, R. W. Allard (1990), "from the electrophoretic assay for seven enzyme loci of 1,032 accessions of cultivated barley", Worldwide pattern of multilocus structure in barley determined by discrete log-linear multivariate analyses

[15] Beaty TH, Yang P, Khoury MJ, Harris EL, Liang KY (1991), Using log-linear models to test for associations among congenital malformations

[16] V. CHRIS LAKHAN, PLACIDO D. LAVALLE (2002), Use of Loglinear Models to Assess Factors Influencing Concern for the Natural Environment

[17] Ya-Lih Lin, Sheue-Ling Hwang (1992), "The external performance shaping factor (PSF), internal PSF, and stressor PSF", The application of the loglinear model to quantify human errors

[18] Steven W.Rust, David A.Burgoon, Bruce P.Lanphear, Shirley Eberly Log-Additive versus Log-Linear Analysis of Lead-Contaminated House Dust and Children's Blood-Lead Levels