

Log-Linear Applications To The Relationship Of Maternal Age And Parity To Preterm Birth

Aisyah Amalia, Rachmah Indawati

Faculty of Public Health Universitas Airlangga

Correspondence email: aisyah.amalia-2018@fkm.unair.ac.id

ABSTRACT

The log-linear model is a special model of the general linear model of Poisson distributed data, and also the development of cross-tabulation analysis of two or more categorical variables. The purpose of this study was to determine the relationship or interaction between the variables of maternal age, parity, and preterm birth. This study uses medical record data at the Haji General Hospital in East Java Province, namely patient data for pregnant women who gave birth between January 1st 2020 to 31st December 2021. The number of samples was 147 respondents. The data analysis method used is log-linear regression analysis. The log-linear model is used as an alternative solution to show if there is a relationship between several variables in a multidimensional contingency table, with the ability to modify the interaction between two or more variables. The resulting log-linear model is: $\log \mu_{ijk} = 4.083 - 0.693 (X) - 0.638 (Y) - 3.795 (Z) + 2.143 (YZ)$. The resulting model states that there is no simultaneous interaction between preterm birth, maternal age, and parity, but there is a partial interaction between maternal age and parity where preterm birth is significant in the model (YZ, X).

Keywords: Log- linear, 3 dimensional models, Premature, Age, Parity

INTRODUCTION

Data nature in statistics is divided into two types, i.e., qualitative and quantitative data. Qualitative data are characterized based on categories and not in a numerical form. Categorical data are widely discovered in the research. One of the qualitative data analysis methods is the independency test. The independency test aims to discover whether there is a relationship between variables. However, the independency test does not reveal which category is causing the dependency. Therefore, a log-linear model is employed.

The log-linear model is a specific model from general linear models used in Poisson distributed data. The log-linear model is developed from the cross-tabulation analysis of two or more categorical variables analyzed using the natural logarithm of each cell's contents in table (1). The log-linear model shows the relationship between categories with a nominal or ordinal measure (2,3). The log-linear model does not differentiate between dependent variables (responses) and independent variables (predictors). However, variables in the model serve as the response variables (4,5). This log-linear model can create a contingency table with several dimensions. Log-linear model dimensions are the number of influencing variables. The two-dimensional model is called a bivariate log-linear model, the three-dimensional model is called a trivariate log-linear model, and the multidimensional model is the most complex. Log-linear aims to analyze the relationship pattern between a group of categorical variables consisting of two or more and predict the parameters depicting the relationship pattern between categorical variables. The log-linear model utilization is interpretable and has flexibility in the modeling related to the variance analysis and regression.

Newborn health plays a vital role in a country's development. Next-generation quality is affected by infant health from the womb. Infant death is a health problem in developing countries which is a concern for all countries. In 2030, one of the primary SDGs in the health and welfare sector is reducing the infant mortality rate (IMR) by 25 per 1,000 live births and neonatal mortality rate (NMR) by 12 per 1,000 live births. Preterm birth affects 43% of overall newborn deaths worldwide. Live birth before completing 37 weeks of gestation is defined as preterm birth (6). The WHO official website states that there are 15 million preterm births annually. Indonesia ranks fifth as the country with a high preterm birth rate, with 657,700 births. The number of low birth weight (LBW) babies is a rough reflection of preterm birth in Indonesia. Based on the 2013 Riskesdas report, the proportion of preterm birth and low birth weight (LBW) babies was 10.2%. In 2018, this number increased to 29.5%. In Indonesia, neonatal death is the most significant factor contributing to the high infant death rate. Based on the 2007 Riskesdas report, the causes of neonatal death in 0-7-day-olds were 25.9% due to respiratory issues and 32.3% due to prematurity (7).

Identifying risk factors before pregnancy is an intervention to prevent preterm birth. Factors affecting preterm birth are categorized into three: fetal, pregnancy, and maternal factors. Fetal factors include twin birth and stillborn. Pregnancy factors include a history of miscarriage, pregnancy complications, history of preterm birth, and premature rupture of membranes. Maternal factors include maternal age, parity, spacing of pregnancies, and maternal smoking behavior. Two in three preterm births occur in expecting mothers without risk factors (8). Therefore, preterm birth is a multifactorial process, and some events occur spontaneously without clear risk factors.

Applications using a log-linear analysis can be found in various case studies. There are several references to log-linear analysis implementation, including a study by Regina Ester Kewinay (2020) that analyzed log-linear to the effect of student attitude and motivation on discipline (9) and a study to discover the relationship between age, income, and health insurance ownership (10). Based on predecessor studies, this study was conducted using the three-dimensional log-linear method to discover the effect of maternal age and parity on preterm birth. The study employed a three-dimensional log-linear model analysis to reveal the correlation pattern between several categorical variables. The variables used in this analysis included preterm birth, maternal age, and parity acquired from medical records of Haji General Hospital (called RSUD)

Haji of East Java Province. The study results are expected to understand the relationship and interaction patterns between variables used to solve preterm birth issues.

Log-Linear Model

The log-linear model has been widely utilized to analyze data with nominal, ordinal, or categorical data. Data analysis with three-variable observation (trivariate) using a categorical type is focused on model creation by testing the interaction of constituent factors in the model, either with two or three factors. Several dimensions are categorized into log-linear models, i.e., two-dimensional, three-dimensional, and four-dimensional log-linear models.

The log-linear approach arranges numbers in cells into a contingency table presented in a list of rows and columns. The contingency table analysis is a relatively simple data compilation technique for observing the relationship between variables in a table.

Three-Dimensional Log-Linear Model

The possible three-dimensional models are as follows:

1. Independent model

The independent model in log-linear is as follows:

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

note:

- μ_{ijk} = expected frequency in each cell
- μ = intercept / general average
- λ_i^X = the effect of X variable in category i
- λ_j^Y = the effect of Y variable in category j
- λ_k^Z = the effect of Z variable in category k

Equation (1) or the independent model means that variables 1, 2, and 3 are in the model, although no interactions occur between the three.

Where:

$$\lambda_{ij}^{XY} = \lambda_{ik}^{XZ} = \lambda_{jk}^{YZ} = \lambda_{ijk}^{XYZ} = 0$$

2. Joint independence/Partial independence model

The joint independence model in log-linear is as follows:

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$$

note:

- μ_{ijk} = frekuensi yang diharapkan dalam setiap sel
- μ = expected frequency in each cell
- λ_i^X = the effect of X variable in category i
- λ_j^Y = the effect of Y variable in category j
- λ_k^Z = the effect of Z variable in category k
- λ_{ij}^{XY} = the interaction effect between X and Y variables

With $\lambda_{ik}^{XZ} = \lambda_{jk}^{YZ} = \lambda_{ijk}^{XYZ} = 0$, equation (2) of the model above indicates that the dependence between the X and Y variables on the Z variable is significant in the model. Another possibility is that the dependence between the X and Z variables on the Y variable is significant in the model, or the dependence of the Z and Y variables on the X variable is significant in the model. Therefore, the joint independence model occurs if an interaction occurs between two variables within three available variables (XY, Z) or (XZ, Y) or (YZ, X).

3. Conditional independence model

The conditional independence model in log-linear is as follows:

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$$

note:

μ_{ijk}	= expected frequency in each cell
μ	= intercept / general average
λ_i^X	= the effect of X variable in category i
λ_j^Y	= the effect of Y variable in category j
λ_k^Z	= the effect of Z variable in category k
λ_{ij}^{XY}	= the interaction effect between X and Y variables
λ_{ik}^{XZ}	= the interaction effect between X and Z variables

With $\lambda_{jk}^{YZ} = \lambda_{ijk}^{XYZ} = 0$, equation (3) above states the dependence between X and Y variables and between X and Z (XY, XZ). The possible interactions are dependence between X and Y variables and between Y and Z (XY, YZ) and dependence between X and Z variables and between Y and Z variables (XZ, YZ).

4. Homogeneous interaction model

The homogeneous interaction model in log-linear is as follows:

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

note:

μ_{ijk}	= expected frequency in each cell
μ	= intercept / general average
λ_i^X	= the effect of X variable in category i
λ_j^Y	= the effect of Y variable in category j
λ_k^Z	= the effect of Z variable in category k
λ_{ij}^{XY}	= the interaction effect between X and Y variables
λ_{ik}^{XZ}	= the interaction effect between X and Z variables
λ_{jk}^{YZ}	= the interaction effect between Y and Z variables

With $\lambda_{ijk}^{XYZ} = 0$, Equation (4) above states a significant interaction between two factors in the model, although there is no interaction or dependence between the three factors (XY, XZ, YZ).

5. Saturated model

The saturated model in log-linear is as follows:

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

note:

μ_{ijk}	= expected frequency in each cell
μ	= intercept / general average
λ_i^X	= the effect of X variable in category i
λ_j^Y	= the effect of Y variable in category j
λ_k^Z	= the effect of Z variable in category k
λ_{ij}^{XY}	= the interaction effect between X and Y variables
λ_{ik}^{XZ}	= the interaction effect between X and Z variables

λ_{jk}^{YZ} = the interaction effect between Y and Z variables
 λ_{ijk}^{XYZ} = the interaction effect between X, Y, and Z variables

Table 1. Three-Dimensional Log-Linear Model

Log-Linear Model	Symbol
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$	(X, Y, Z)
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$	(XY, Z)
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}$	(XZ, Y)
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$	(YZ, X)
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$	(XY, XZ)
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$	(XY, YZ)
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$	(XZ, YZ)
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$	(XY, XZ, YZ)
$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$	(XYZ)

METHOD

The study type is non-reactive research with a cross-sectional study design. The study population was all expectant mothers who gave birth and were recorded in the medical records of RSUD Haji of East Java Province from 1 January to 31 December 2020 and 2021. The study samples were expectant mothers who gave birth in RSUD Haji of East Java Province, amounting to 147 samples.

Samples were collected using the random sampling technique based on predetermined criteria. The study instrument was a data collection sheet. The variables used in the study were preterm birth (X), categorized into 0: yes and 1: no, age (Y), categorized into 0: with risk and 1: no risk, and parity (Z), categorized into 0: with risk and 1: no risk. The study was performed from March to April 2022 in RSUD Haji of East Java Province. The analysis carried out in the study was the log-linear regression test to produce a log-linear model in discovering the relationship or interaction between variables.

RESULT AND DISCUSSION

Descriptive Analysis

The data acquired from medical records of RSUD Haji of East Java Province amounted to 147 samples, comprising three variables, i.e., preterm birth (yes and no categories), maternal age (with and without risk categories), and parity (with and without risk categories). The following is the contingency table.

Table 2. The Relationship of Maternal Age, Parity, and Preterm Birth

Preterm Birth	Maternal Age	Parity		Total
		With risk (<2 and/or >3)	No Risk (2-3)	
Yes	With risk (<20 and/or >35 years)	3	13	16
	No risk (20-35 years)	1	32	33
No	With risk (<20 and/or >35 years)	6	34	40
	No risk (20-35 years)	1	57	58
Total		11	136	147

K-Way Test

The first step was examining the full model fit with the first hypothesis to observe whether the effect of order k or more equals zero. The test started from the highest to lowest order with the following hypotheses:

H_0 : the effect of order k or more = 0

H_1 : the effect of order k or more $\neq$ 0

H_0 is rejected if the opportunity value of G^2 (likelihood ratio chi-square) $< \alpha$.

Then, testing whether the k^{th} effect is zero used the following hypotheses:

H_0 : the effect of order k = 0

H_1 : the effect of order k $\neq$ 0

H_0 is rejected if the opportunity value of G^2 (likelihood ratio chi-square) $< \alpha$.

Table 3. K^{th} Effect Test Results for Factors Affecting Preterm Birth

	K	df	Likelihood Ratio	
			Chi-Square	Sig.
K-way and Higher Order Effects ^a	3	1	.042	.838
	2	4	10.800	.029*
	1	7	161.458	.000*
K-way Effects ^b	1	3	150.658	.000*
	2	3	10.758	.013*
	3	1	.042	.838

Based on K-way and Higher Order Effects (the effect test of k^{th} or more is equal to zero) results, $K=1$ and $K=2$ produced P-values $< \alpha$ (0.05), and hence, H_0 was rejected. It indicates that orders $K=1$ and $K=2$ enter the model. Meanwhile, $K \geq 3$ showed P-values $> \alpha$, and hence, could not reject H_0 .

Meanwhile, K-way Effects (the effect of k^{th} is equal to zero) results showed that $K=1$ and $K=2$ resulted in P-values $< \alpha$ (0.05), and hence, H_0 was rejected. It indicates that orders $K=1$ and $K=2$ enter the model. Meanwhile, $K \geq 3$ showed P-values $> \alpha$, and hence, could not reject H_0 . Therefore, the log-linear mode acquired is as follows:

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

Partial Association Test

The k^{th} effect test was continued with the partial association test to examine the effect of each variable and the interaction effect of each variable. The general hypotheses are as follows:

H_0 : No interaction effect between variables

H_1 : There is an interaction effect between variables

H_0 is rejected if the Partial Chi-Square value $> \chi^2(\text{df}, \alpha)$ or significance value $< \alpha$ (0.05).

Table 4. Partial Association Test Results for Factors Affecting Preterm Birth

Effect	Df	Partial Chi-Square	Sig.	Effect	Df	Partial Chi-Square	Sig.
Preterm*Maternal Age	1	1.120	.290	Preterm Birth	1	16.650	.000*
Preterm*Parity	1	.235	.6289	Maternal Age	1	8.414	.004*
Maternal Age*Parity	1	9.776	.002*	Parity	1	125.594	.000*

The table above shows an interaction effect between maternal age and parity, demonstrating that maternal age and parity variables affected the model. Meanwhile, preterm birth, maternal age, and parity variables each affected the model.

Goodness of Fit Test

The Goodness of Fit test aims to determine whether the model fits. In other words, it examines whether there is a difference between the observation and model prediction results. The hypotheses employed are as follows:

H_0 : fit model

H_1 : unfit model

Table 5. Goodness of Fit Test

Criteria	Value	Df	Sig.
Likelihood Ratio	1.210	3	0.751
Pearson Chi-Square	1.206	3	0.752

The table above shows that the likelihood ratio significance value was $0.751 > \text{significant alpha } (0.05)$. Thus, it concluded that the model was fit.

Best Model Selection

The best model selection aims to create or produce the best model having correct relationship patterns or interactions between variables in the data. There are several log-linear model selection methods, e.g., elimination, backward, forward, and others. Fundamentally, the backward method completes a model based on the hierarchy process, i.e., from the complete to the simplest model. Meanwhile, the forward method adds each parameter of the variable effect, starting from the simplest to the complete models.

General model (0)

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

The model selection method acquired the best log-linear model, i.e., the joint independence model [YZ, X]. This model states a dependency between Y and Z variables, while the X variable remains exist or is significant in the model. Therefore, the best model is as follows:

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$$

Where: μ_{ijk} = expected frequency in each cell
 μ = intercept / general average
 λ_i^X = the effect of X variable (preterm birth)
 λ_j^Y = the effect of Y variable (maternal age)
 λ_k^Z = the effect of Z variable (parity)
 λ_{jk}^{YZ} = the interaction effect between Y (maternal age) and Z (parity) variables

Table 6. Parameter estimates for each variable and interaction

Parameter	Estimate	Z	Sig.
Constant	4.083	33.749	.000*
[preterm=0]	-0.693	-3.962	.000*
[preterm=1]	0	-	-
[maternal age=0]	-0.638	-3.541	.000*
[maternal age=1]	0	-	-
[parity=0]	-3.795	-5.308	.000*
[parity=1]	0	-	-
[maternal age=0]*[parity=0]	2.143	2.671	.008*
[maternal age=0]*[parity=1]	0	-	-
[maternal age=1]*[parity=0]	0	-	-
[maternal age=1]*[parity=1]	0	-	-

The table above shows an overall effect, a primary effect relationship from each variable, and a correlation effect between variables. The analysis results showed estimated values of each parameter, as presented in the table above. The parameter values are interpreted by observing the sign of each estimated parameter. If the estimated parameter value is positive, the parameter is considered to have a positive association/relationship pattern with other parameters. Conversely, if the estimated parameter value is negative, the parameter is considered to have a negative associative pattern with other parameters. The following is the log-linear model for the (X, Y, Z, YZ) model.

$$\log \mu_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$$

$$\log \mu_{ijk} = 4.083 - 0.693 (X) - 0.638 (Y) - 3.795 (Z) + 2.143(YZ)$$

The above model shows a partial interaction between maternal age (Y) and parity (Z). It also shows the dependency requirement between maternal age (Y) and parity (Z). Furthermore, it shows independency between maternal age (Y) and preterm birth (X) and between parity (Z) and preterm birth (X). The dependency requirement indicates that maternal age and parity variables are co-dependent. Table 5 above shows that maternal age and parity variables have a positive association. The researchers assumed that the parity risk is low if the mother is young. Meanwhile, as the mother gets older, the parity risk gets higher. It shows that age determines the parity risk.

It contrasts the relationship between maternal age and preterm birth variables and between parity and preterm birth variables. They show independence, where maternal age and parity did not relate to preterm birth. Preterm birth is a multifactorial process. Therefore, identifying risk factors before pregnancy or during pre-conception is an intervention to prevent preterm birth. A combination of obstetric, psychological, genetic, sociodemographic, and medical factors affects preterm birth events (11). However, expectant mothers with risk factors are not guaranteed for preterm birth. Most preterm births are spontaneous and caused by various factors, where each expectant mother has different issues. Hence, there is no definite identified cause.

CONCLUSION

The analysis concludes that K-way and partial association test results demonstrate interaction effects between variables. The best model is joint independence (YZ, X). The acquired model states no simultaneous interactions between preterm birth, maternal age, and parity. However, there is a partial interaction between maternal age and parity, where preterm birth is significant to the (YZ, X) model. The log-linear model produced is $\log \mu_{ijk} = 4.083 - 0.693(X) - 0.638(Y) - 3.795(Z) + 2.143(YZ)$.

REFERENCES

1. Jeansonne, A. Log-Linear Models. 2012.
2. Wohingati, Jesica Aproditta. Analisis Hubungan Jenis Pelanggaran, Lokasi Jembatan Timbang, Dan Golongan Kendaraan Menggunakan Model Log Linier 3 Dimensi. Skripsi. Surabaya: Fakultas Vokasi Institut Teknologi Sepuluh Nopember, 2018.
3. Viestri, Dinar Ariana. Pemodelan Log Linier Dan Regresi Logistik Biner Bivariat Pada Hasil Medical Check-Up Pegawai Negeri Sipil (PNS) Institut Teknologi Sepuluh Nopember (ITS). Skripsi. Surabaya: Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember, 2015.
4. Sihotang, S. F. Analisis Model Log Linier Tiga Dimensi Untuk Data Kualitatif Dengan Metode Forward. Journal of Mathematics Education and Science. 2020;6(1), 62–69.
5. Pontoh, R. S., & Sunengsih, N. Pemodelan Karakteristik Tingkat Pendidikan Anak di Provinsi Jawa Barat Menggunakan Log Linear. Seminar Nasional Pendidikan Matematika. 2016, 1–7.

6. WHO. 2018
7. Badan Penelitian dan Pengembangan Kesehatan Kementrian Kesehatan RI 2008. Laporan Riset Kesehatan Dasar (Riskesdas) 2007. Jakarta.
8. Krisnadi, dkk. Prematuritas. Bandung : Refika Aditama. 2009
9. Kewinay, R. E. Aplikasi Analisis Log-Linier Pada Pengaruh Sikap Dan Motivasi Mahasiswa Terhadap Kedisiplinan. Jurnal Diferensial. 2020;02(02), 37–48.
10. Priscilla, E., Prinssesa, J., Siang, M., Aurelia, J., & Nugraha, E. S. Log Linear Model on Contingency Table to Analyze Relationship between Age , Income , and Health Insurance Ownership. Journal of Actuarial, Finance and Risk Management (JAFRM). 2022;1(1), 17–26.
11. Prawirohardjo, S. (2016). Ilmu Kebidanan. Jakarta, Indonesia: PT Bina Pustaka Sarwono Prawirohardjo