Parameter: Journal of Statistics
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 6 No. 1 2026, 43 - 50
DOI: *https://doi.org/10.22487/27765660.2026.v6.i1.17901*
eISSN: 2776-5660

ANALYSIS OF THE RELATIONSHIP BETWEEN REGION, AGE, SEX, AND MARITAL STATUS USING A FOUR-WAY LOG-LINIER MODEL IN BENGKULU PROVINCE IN 2022

Putri Damayanti[1], Novelisa Syendra Wanti[2], Gita Sonia[3], Rada Nur Aisa[4], Etis Sunandi[5], Pepi Novianti[6*]

[1,2,3,4,5,6]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Bengkulu, Bengkulu, Indonesia

***e-mail**: [1]*pd337768@gmail.com*, [2]*novelisasw@gmail.com*, [3]*soniagita252@gmail.com*, [4]*nuraisarada@gmail.com*, [5]*esunandi@unib.ac.id*, [6*]*pie_novianti@unib.ac.id*

ABSTRACT

This study examines the association among age group, gender, marital status, and district/city of residence for individuals aged 10–24 years in Bengkulu Province in 2022 using a four-way log-linear model. Secondary data were obtained from the Central Bureau of Statistics (BPS) and arranged into a four-way contingency table. Model parameters were estimated using the maximum likelihood method, while model adequacy was evaluated using the Pearson Chi-square goodness-of-fit test. The results showed that the independence model as well as all models containing only two-way and three-way interactions failed to adequately fit the observed data. In contrast, the saturated model adequately represented the observed frequency distribution and was therefore used as the reference model to describe the complete interaction structure among the four variables. These findings indicate that the associations among age group, gender, marital status, and district/city involve higher-order interactions that cannot be explained by lower-order log-linear models. Therefore, a comprehensive interaction structure is required to accurately represent the demographic relationships in Bengkulu Province.

Keywords: *Bengkulu Province, categorical data, contingency table, four-way log-linear model, goodness-of-fit.*

Cited: Damayanti, P., Wanti, N. S., Sonia, G., Aisa, R. N., Sunandi, E., & Novianti, P. (2026). Analysis of the Relationship Between Region, Age, Sex, and Marital Status Using a Four-Way Log-Linier Model in Bengkulu Province in 2022. *Parameter: Journal of Statistics*, 6(1), 43-50. https://doi.org/10.22487/27765660.2026.v6.i1.17901

[*]Received: November 2025; Reviewed: May 2026; Published: June 2026

INTRODUCTION

Categorical data is data whose measurement scale consists of a set of nominal and ordinal categories (Akhyak, 2024). Multidimensional categorical data analysis is an important statistical approach for examining complex relationships among categorical variables. This approach allows researchers to understand the structure of interdependence among variables in greater depth than simple univariate or bivariate analyses. Categorical data consists of observational data containing variables that comprise categories and are presented in the form of observed frequencies. A categorical variable is a variable whose data falls on a nominal or ordinal scale (Amelia, 2022). In practice, categorical data is widely encountered in various fields of research, particularly in the social sciences, public health, and behavioral sciences.

A contingency table is one of the most effective and informative ways to present categorical data. A contingency table consists of rows and columns that represent the categories of the observed categorical variables. The intersections of the rows and columns form cells that show the frequencies of the combinations of the observed variable categories (Ani, Suci, & Eni, 2020). Analysis of a contingency table allows researchers to compare proportions across groups while assessing the strength of the relationship between the categorical variables under study. In addition, contingency tables also facilitate the visualization of data distribution patterns, thereby providing an initial overview before further statistical testing is conducted.

Log-linear models have wide-ranging applications across various fields of science, from socioeconomics, biology, and medicine to natural language processing and genetics. These models are extremely useful tools due to their flexibility in handling various high-dimensional categorical data structures. Log-linear analysis involves two main stages: finding an appropriate model and interpreting the selected model. Log-linear models can be used to analyze relationships between categorical variables without distinguishing between response and predictor variables, thereby providing a more symmetrical perspective on the relationships between variables. In some cases, conflicting conclusions often arise when an analysis considers only two factors (variables) compared to an analysis involving three variables (Jaka, 2022). This highlights the importance of considering more complex interaction effects in multivariate analysis.

The log-linear model is a statistical model used for frequency data in contingency tables, in which the expected values of each cell are modeled logarithmically as a linear combination of the effects of the variables and their interactions. Therefore, this model is effective for testing the presence of relationships or associations between categorical variables (Christensen, 2025). The logarithmic transformation ensures that predicted frequency values are always positive and allows for the interpretation of model parameters in terms of odds ratios and multiplicative effects. The four-dimensional log-linear model is an extension of the three-dimensional log-linear model—or the trivariate model—and thus falls under the category of high-dimensional log-linear models (Azen et al., 2024). The use of high-dimensional log-linear models allows researchers to identify more complex and comprehensive interaction patterns among variables that cannot be detected through low-dimensional analysis. This ability to detect high-order interactions is crucial for understanding complex social and demographic phenomena.

Bengkulu Province, as one of Indonesia's provinces, has diverse demographic characteristics, comprising 1 city and 9 regencies, each with distinct socioeconomic patterns. The population composition of Bengkulu Province shows significant variation in terms of age structure, marital status, gender, and distribution across regencies and cities. These variations reflect population dynamics influenced by various factors such as urbanization, migration, educational levels, and regional economic conditions. Understanding these patterns is essential for supporting targeted development planning, particularly in the fields of education, reproductive health, and community empowerment. An in-depth analysis of these demographic characteristics can provide important information for policymakers in designing development programs that meet the needs of communities in each region. With a comprehensive understanding of the demographic structure, local governments can allocate resources more efficiently and design more targeted policy interventions.

Log-linear models have been widely used in analyzing relationships among categorical variables; for example, studies by Sihotang and Zuhri (2020) and Salsabila et al. (2024) applied log-linear models to three-dimensional contingency tables to identify patterns of association among variables.

However, the application of four-dimensional log-linear models in population data analysis remains relatively limited, particularly in uncovering simultaneous interactions among region, gender, age group, and marital status. This limitation highlights the need to apply four-dimensional log-linear models to gain a more comprehensive understanding of the association structures among demographic variables. Therefore, this study aims to analyze the relationships among districts/cities, gender, age groups, and marital status of the population in Bengkulu Province in 2022. This study employs log-linear model analysis on incomplete four-dimensional contingency tables to identify the variables that are mutually associated among the four observed variables. The results of this study are expected to contribute theoretically to the field of statistics, particularly regarding the application of multidimensional log-linear models, as well as provide practical information useful for formulating population policies in Bengkulu Province.

MATERIALS AND METHODS

Linear Log Model

The log-linear model is a statistical model for frequency data in contingency tables, where the expected cell values are modeled in logarithmic form as a linear combination of the effects of variables and their interactions. Thus, this model is effective for testing associations among categorical variables (Christensen, 2025). The log-linear model is a special case of Generalized Linear Models for data that follows a Poisson distribution. The log-linear model is used to analyze the relationship between categorical response variables that form a contingency table and to determine which variables tend to cause dependency (Aravkin et al., 2024).

4-Way Linear Log Model

The four-dimensional log-linear model is an extension of the three-dimensional log-linear model or trivariate model, and can thus be categorized as a high-dimensional log-linear model. In this study, the four-dimensional log-linear model is used to analyze four categorical variables, namely District/City (W), Age Group (X), Gender (Y), and Marital Status (Z). Variable W consists of i categories, variable X consists of j categories, variable Y consists of k categories, and variable Z consists of l categories. With a total of n individual observations, the observation for cell (i, j, k, l) is denoted as n_{ijkl}. In general, the expected value for cell (i, j, k, l) is denoted by $\mu_{i,j,k,l}$. For a four-dimensional log-linear model with variables W, X, Y, Z, the general form of the independent model equation for the four categorical variables is (Azen et al., 2011):

$$\log \mu_{i,j,k,l} = \lambda + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z \quad (1)$$

For i=i ... I, j=1 ... J, k=1 ... K, l=1 ... L

with its degrees of freedom being (Christensen, 2025):

$$db = db\left(log\mu_{i,j,k,l}\right) - (db(\mu) + db\left(\lambda_i^W\right) + \left(\lambda_j^X\right) + \left(\lambda_k^Y\right) + (\lambda_l^Z))$$

$$db = IJKL - ((1) + (I-1) + (J-1) + (K-1) + (L-1)$$

$$= IJKL - I - J - K - L + 3$$

If there is interaction among the four variables, then the model becomes:

$$\log \mu_{i,j,k,l} = \lambda + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_l^Z + \lambda_{ij}^{WX} + \lambda_{ik}^{WY} + \lambda_{il}^{WZ} + \lambda_{jk}^{XY} + \lambda_{jl}^{XZ} + \lambda_{kl}^{YZ} + \lambda_{ijk}^{WXY} + \lambda_{ijl}^{WXZ} + \lambda_{ikl}^{WYZ} + \lambda_{jkl}^{XYZ} + \lambda_{ijkl}^{WXYZ} \quad (2)$$

Description:

$\mu_{i,j,k,l}$: expected frequency in cell-i, j, k, l;
λ : grand mean parameter;
λ_i^W : the main effect of the variable or the influence of factor W category-i;
λ_j^X : the main effect of the variable or the influence of factor X category – j;
λ_k^Y : the main effect of the variable or the influence of factor Y category – k;
λ_l^Z : the main effect of the variable or the influence of factor Z category – l;
λ_{ij}^{WX} : the interaction effect of variables W and X or the influence of the cell- ij interaction factor ;
λ_{ik}^{WY} : the interaction effect of variables W and Y or the influence of the cell -ik interaction factor;
λ_{il}^{WZ} : the interaction effect of variables W and Z or the influence of the cell -il interaction factor;
λ_{jk}^{XY} : the interaction effect of variables X and Y or the influence of the cell -jk interaction factor;
λ_{jl}^{XZ} : the interaction effect of variables X and Z or the influence of the cell -jl interaction factor;
λ_{kl}^{YZ} : the interaction effect of variables Y and Z or the influence of the cell -kl interaction factor;
λ_{ijk}^{WXY} : the interaction effect of variables W, X, Y or the influence of the cell-ijk interaction factor;

λ_{ijl}^{WXZ} : the interaction effect of variables W, X, Z or the influence of the cell -ijl interaction factor;
λ_{ikl}^{WYZ} : the interaction effect of variables W, Y, Z or the influence of the cell -ikl interaction factor;
λ_{jkl}^{XYZ} : the interaction effect of variables X, Y, Z or the influence of the cell -jkl interaction factor;
λ_{ijkl}^{WXYZ} : the interaction effects of variables $W, X, Y,$ and Z or the influence of the cell-$ijkl$ interaction factor.

Estimation of Expected Frequencies

The expected frequency estimate is the estimate that occurs from the values of each category for each variable. The way to find the expected frequency estimate for a four-dimensional log-linear model is using the maximum likelihood method. For instance, n_{ijkl}represents the observed data of variables W, X, Y, and Z, which follow a Poisson distribution, and μ_{ijkl} represents the expected value. Then, the probability density function is(Hammond et al., 2024) :

$$f(n_{ijkl};\ \mu_{ijkl}) = \frac{e^{-\mu_{ijkl}}\mu_{ijkl}{}^{n_{ijkl}}}{n_{ijkl}!} \quad (3)$$

Whereas the function of the likelihood is:

$$L(\mu) = \prod_{i=1}^{W}\prod_{j=1}^{X}\prod_{k=1}^{Y}\prod_{l=1}^{Z}\frac{e^{-\mu_{ijkl}}\mu_{ijkl}{}^{n_{ijkl}}}{n_{ijkl}!} \quad (4)$$

Consequently, the log-likelihood function can be expressed as:

$$\log L(\mu) = \sum_{i=1}^{W}\sum_{j=1}^{X}\sum_{k=1}^{Y}\sum_{l=1}^{Z} n_{ijkl}\ln\{\mu_{ijkl}\} - \sum_{i=1}^{W}\sum_{j=1}^{X}\sum_{k=1}^{Y}\sum_{l=1}^{Z}\mu_{ijkl} - \sum_{i=1}^{W}\sum_{j=1}^{X}\sum_{k=1}^{Y}\sum_{l=1}^{Z}\ln\{n_{ijkl}!\} \quad (5)$$

The likelihood function employed is the one that involves only the model parameters. Terms not involving the model parameters are treated as constants. Consequently, the log-likelihood function in Equation (5) becomes:

$$\log L(\mu) = \sum_{i=1}^{W}\sum_{j=1}^{X}\sum_{k=1}^{Y}\sum_{l=1}^{Z} n_{ijkl}\ln\{\mu_{ijkl}\} - \sum_{i=1}^{W}\sum_{j=1}^{X}\sum_{k=1}^{Y}\sum_{l=1}^{Z}\mu_{ijkl} \quad (6)$$

The Goodness-of-Fit Test

The estimated model parameters are first obtained, followed by performing a goodness-of-fit test, which can be conducted using either the Chi-Square statistic or the Likelihood Ratio statistic. The Chi-Square statistic also serves to assess the presence of a significant association among the measured variables. Specifically, it is applied to test the null hypothesis that the expected population frequencies conform to a specified model, utilizing either the Likelihood Ratio Test (G^2) or the Pearson Chi-Square Test (χ^2) (Bilder & Loughin, 2024):

- Hypotheses:
  H_0: Model fits the data ($n_{ijkl} \approx \hat{\mu}_{ijkl}$)
  H_1: Model does not fit the data ($n_{ijkl} \neq \hat{\mu}_{ijkl}$)
- significance level:
  $\alpha = 5\%$
- Test statistic:
  either the Likelihood Ratio Test, or the Pearson Chi-Square Test (χ^2) is employed:
  Likelihood ratio test: $G^2{:}2\sum_{i=1}^{W}\sum_{j=1}^{X}\sum_{k=1}^{Y}\sum_{l=1}^{Z} n_{ijkl}log\left(\frac{n_{ijkl}}{\hat{\mu}_{ijkl}}\right)$ (7)
  Pearson Chi-Square test: $\chi^2{:}\ \sum_{i=1}^{W}\sum_{j=1}^{X}\sum_{k=1}^{Y}\sum_{l=1}^{Z}\frac{(n_{ijkl}-\hat{\mu}_{ijkl})^2}{\hat{\mu}_{ijkl}}$ (8)
- Decision rule: reject H_0 if $G^2 > \chi^2_{\alpha,df}$ or $\chi^2 > \chi^2_{\alpha,df}$ or $P_{value} < \alpha$.

Model Evaluation

After fitting the log-linear models, each model was evaluated using the Pearson chi-square goodness-of-fit test. The goodness-of-fit test was used to determine whether the expected frequencies generated by the model were consistent with the observed frequencies in the contingency table. A model was considered to adequately fit the data if the null hypothesis of the goodness-of-fit test was not rejected. The model that best represented the observed data was then selected for further interpretation.

Data Sources and Research Variables

Secondary data were obtained from the Badan Pusat Statistik (BPS) of the Republic of Indonesia. The dataset consists of the population aged 10 years and above in Bengkulu Province, disaggregated by age group, gender, marital status, and district/city. These categorical variables were analyzed using a four-way log-linear model. The study focuses on individuals in the age range 10–24 years, covering 9 districts and 1 city. All statistical analyses were performed using **R software**. The four-way log-linear models were fitted using the ***loglm()*** function from the ***MASS*** package based on the maximum likelihood estimation method. The goodness of fit of each log-linear model was evaluated using the Pearson chi-square statistic, and the model that best fit the observed data was identified based on the goodness-of-fit test results.

Table 1. Research Variables

Symbol	Variable	Category
W	District/City	9 districts, 1 city (Bengkulu Province)
X	Age Group	10–14 years, 15–19 years, 20–24 years
Y	Gender	Male, Female
Z	Marital Status	Married, Not Married

Data Analysis Steps

The data analysis for this study comprised the following steps:

1. Organizing data into a four-dimensional contingency table based on age, gender, marital status, and region.
2. Fitting four-way log-linear models, starting from the independence model and continuing to two-way interaction, three-way interaction, and saturated models.
3. Evaluating the goodness of fit of each model using the Pearson chi-square statistic.
4. Selecting the model that best fit the observed data based on the goodness-of-fit test results.
5. Interpreting the results of the analysis to explain the main effects and interactions between variables.

RESULTS AND DISCUSSION

Data Description

The data used in this study is cross-tabulated data on the population of Bengkulu Province by district/city, age group, gender, and marital status. The data was obtained from publications by the Central Statistics Agency (BPS) and reprocessed by the researchers. The variables used consist of four categories, namely district/city (North Bengkulu, Kaur, Seluma, Muko-Muko, Lebong, Kepahiang, Rejang Lebong, South Bengkulu, Bengkulu City, and Central Bengkulu), age group (10–14 years, 15–19 years, and 20–24 years), gender (male and female), and marital status (unmarried and married). Thus, there are $10 \times 3 \times 2 \times 2 = 120$ combinations of data cells that form the basis of the four-way log-linear model analysis. In general, the number of unmarried residents is higher than that of married residents. From 120 data points on the population aged 10–24 years in 10 cities/districts in Bengkulu Province, it was found that the average number of unmarried males was 8,310 and the number of married males was 550, while the number of unmarried females was 7,023 and the number of married females was 1,324. For the three age groups (10–14, 15–19, 20–24 years), the highest number of married males was found in North Bengkulu Regency, while the highest number of married females was found in Rejang Lebong Regency for the 10–14 and 15–19 age groups, and in North Bengkulu Regency for the 20–24 age group. Meanwhile, the largest number of unmarried residents, both male and female, are found in Bengkulu City. This data is used to construct a four-way log-linear model to determine the pattern of dependence or interaction between the four variables, as well as to identify significant relationships between region, age, gender, and marital status.

Goodness-of-Fit Test

The data used in this study are presented in the form of a contingency table. In this table, there are cells with a frequency of 0, which is the number of married females in the 10-14 age group in Central Bengkulu Regency. Therefore, the goodness of fit test in this study was only conducted using Pearson's Chi-Square statistic, without considering the Likelihood Ratio value. The first log-linear model tested was the fully independent (saturated references) model. The test was conducted to see the significant relationship between variables. The results of this test are presented in the following table:

Table 2. Full Independence Model Analysis (*Saturated References)*

Model	Log Linier Model	DF	χ^2	${\chi^2}_{\alpha,df}$	Description
1	(*W, X, Y, Z)*	106	123233	131.0315	H_0 Rejected

Based on the table above, the log linear model (*W, X, Y, Z)* with DF 106 shows that $(\chi^2) > ({\chi^2}_{\alpha,df})$ or 123233 > 131.0315 it can be concluded at a 5% significance level that there is sufficient evidence to reject H_0 meaning that the model (*W, X, Y, Z)* does not fit the actual situation.

Because the four-way model was not appropriate, an analysis was also conducted on other interaction factors involving two-way and three-way combinations between variables, such as interactions between district and gender, between age group and marital status, and between all three variables simultaneously. Of the 112 interaction models analyzed, none adequately fitted the observed data based on the goodness-of-fit test. This indicates that lower-order interaction models were insufficient to explain the observed frequency distribution in the contingency table. Therefore, the relationships among district/city, age group, gender, and marital status are likely to involve more complex interaction structures.

Table 3. Two-way and Three-Way Model Analysis

Model	Log Linier Model	DF	χ^2	${\chi^2}_{\alpha,df}$	Description
2	*(WX, Y, Z)*	88	123514.5	110.898	H_0 Rejected
3	*(WY,X, Z)*	97	123133.4	120.9896	H_0 Rejected
4	*(XY,W, Z)*	104	122653.0	128.8039	H_0 Rejected
5	*(WZ,X,Y)*	97	113626	120.99	H_0 Rejected
6	*(XZ,W,Y)*	104	*24954.04*	128.8039	H_0 Rejected
7	*(YZ,W,X)*	105	96557.73	129.918	H_0 Rejected
8	*(WX, YZ)*	87	96802.24	109.7733	H_0 Rejected
9	*(WY, XZ)*	95	24991.93	118.7516	H_0 Rejected
10	*(WZ, XY)*	95	113226.6	118.7516	H_0 Rejected
⋮	⋮	⋮	⋮	⋮	⋮
108	*(WYZ,WX,XY)*	58	87016.95	*76.7778*	H_0 Rejected
109	*(WYZ,WX,XZ)*	58	4105.54	76.7778	H_0 Rejected
110	*(WXZ,WY,XY)*	48	16406.89	65.17077	H_0 Rejected
112	*(WXZ,WY,YZ)*	94	11695.42	117.6317	H_0 Rejected
113	*(WXY,WZ,YZ)*	49	87713.2	66.33865	H_0 Rejected

Based on the table above, the results of the two-way and three-way log-linear model analysis, namely from model 2 to model 113, show that the value of χ^2 is greater than ${\chi^2}_{\alpha,df}$ at a significance level of $\alpha = 0.05$. This means that all of the two-way and three-way models tested do not correspond to the actual situation, or in other words, these models are not yet able to describe the relationship between variables well.

This discrepancy indicates that the interactions formed between two or three variables are still insufficient to explain the data variation in the four-way contingency table used. Thus, the relationship between the variables of district/city, age group, gender, and marital status is likely to be more complex and involve higher-level interactions.

Therefore, the analysis was continued using a saturated model, which is the most complete log-linear model because it includes all main effects and interactions between variables up to the four-way level. This model was used as a comparison to see the extent to which simpler models deviate from ideal conditions, as well as to assess whether the entire data structure can be perfectly explained by the combination of these four variables.

Table 4. Saturated Model Analysis

Model	Log Linier Model	DF	χ^2	${\chi^2}_{\alpha,df}$	Description
114	(*WXYZ)*	0	0	0	H_0 Not Rejected

Based on Table 4, the saturated log-linear model (*WXYZ*) has a Pearson Chi-square statistic of 0 with 0 degrees of freedom, indicating a perfect fit to the observed data. Therefore, the null hypothesis cannot be rejected, indicating that there is no significant difference between the observed and expected frequencies under the saturated model.

Furthermore, the goodness-of-fit results for all partial models involving two-way and three-way interactions indicate that these models do not adequately fit the observed data, as the null hypothesis was rejected for each model. This suggests that lower-order interaction models are insufficient to explain the observed frequency distribution. Consequently, the relationships among district/city, age group, gender, and marital status are likely to involve higher-order interactions that cannot be represented adequately by simpler models.

The goodness-of-fit results indicate that only the saturated model adequately fits the observed data. Therefore, rather than identifying it as the best model the saturated model is used as the reference model to describe the complete interaction structure among the four study variables. The general form of the corresponding four-way log-linear model is given as follows:

$$logm_{ijkl} = \mu + \lambda_i{}^{W} + \lambda_j{}^{X} + \lambda_k{}^{Y} + \lambda_l{}^{Z} + \lambda_{ij}{}^{WX} + \lambda_{ik}{}^{WY} + \lambda_{il}{}^{WZ} + \lambda_{jk}{}^{XY} + \lambda_{jl}{}^{XZ} + \lambda_{kl}{}^{YZ} + \lambda_{ijk}{}^{WXY} + \lambda_{ijl}{}^{WXZ} + \lambda_{ikl}{}^{WYZ} + \lambda_{jkl}{}^{XYZ} + \lambda_{ijkl}{}^{WXYZ}$$

Description:

$\mu_{i,j,k,l}$: expected frequency in cell-i,j,k,l;
λ : grand mean parameter;
λ_i^W : the main effect of the variable or the influence of factor W category-i;
λ_j^X : the main effect of the variable or the influence of factor X category – j;
λ_k^Y : the main effect of the variable or the influence of factor Y category – k;
λ_l^Z : the main effect of the variable or the influence of factor Z category – l;
λ_{ij}^{WX} : the interaction effect of variables W and X or the influence of the cell- ij interaction factor ;
λ_{ik}^{WY} : the interaction effect of variables W and Y or the influence of the cell -ik interaction factor;
λ_{il}^{WZ} : the interaction effect of variables W and Z or the influence of the cell -il interaction factor;
λ_{jk}^{XY} : the interaction effect of variables X and Y or the influence of the cell -jk interaction factor;
λ_{jl}^{XZ} : the interaction effect of variables X and Z or the influence of the cell -jl interaction factor;
λ_{kl}^{YZ} : the interaction effect of variables Y and Z or the influence of the cell -kl interaction factor;
λ_{ijk}^{WXY} : the interaction effect of variables W, X, Y or the influence of the cell-ijk interaction factor;
λ_{ijl}^{WXZ} : the interaction effect of variables W, X, Z or the influence of the cell -ijl interaction factor;
λ_{ikl}^{WYZ} : the interaction effect of variables W, Y, ,Z or the influence of the cell -ikl interaction factor;
λ_{jkl}^{XYZ} : the interaction effect of variables X, Y, Z or the influence of the cell -jkl interaction factor;
λ_{ijkl}^{WXYZ} : the interaction effects of variables W, X, Y, and Z or the influence of the cell-$ijkl$ interaction factor.

The goodness-of-fit analysis indicates that lower-order log-linear models, including those with only main effects, two-way interactions, and three-way interactions, were not sufficient to adequately represent the observed data. In contrast, the saturated model provided an adequate fit, suggesting that the observed frequency distribution cannot be fully explained by simpler interaction structures. These findings indicate that the joint distribution of age group, district/city, gender, and marital status is more complex than can be represented by lower-order models. However, further interpretation of the specific interaction effects would require examination of the estimated model parameters.

CONCLUSION

This study applied a four-way log-linear model to examine the relationships among district/city, age group, gender, and marital status of the population aged 10–24 years in Bengkulu Province. The goodness-of-fit analysis showed that the independence model and all lower-order models containing only two-way and three-way interactions did not adequately represent the observed data. Only the saturated model provided an adequate fit and was therefore used as the reference model to describe the complete interaction structure among the four categorical variables.

These findings indicate that the demographic relationships among district/city, age group, gender, and marital status involve higher-order interactions that cannot be represented adequately by simpler log-linear models. Future studies are recommended to incorporate additional demographic and socioeconomic variables, such as educational attainment, employment status, and income level, to obtain a more comprehensive understanding of population characteristics and their interaction patterns.

REFERENCES

Aisyah Amalia, R. I. (2022). Aplikasi Log-Linier pada Hubungan Usia Ibu dan Paritas Terhadap Kelahiran Prematur. *Jurnal Publikasi Kesehatan Masyarakat Indonesia*, 9;108-117.

Akhyak, A. M. (2024). *Revitalisasi Penggunaan Media Serta Metode Belajar Dalam Pembelajaran Matematika dan Teknik.* Tulungangung: Akademia Pustaka.

Ani, B. A., Suci, A. D., & Eni, S. (2020). *Analisis Data Kategorik Menggunakan R: Teori dan Aplikasinya pada Berbagai.* Malang: UB Press.

Aravkin, A., Choromanska, A., Deng, L., Heigold, G., Jebara, T., Kanevsky, D., & Wright, S. J. (2024). *Log-Linear Models, Extensions, and Applications* (MIT Press (ed.); Neural Inf).

Azen, R. U. of W.-M., Walker, & Cindy M.University of Wisconsin-Milwaukee. (2011). *Categorical Data Analysis for the Behavioral and Social Sciences* (L. Taylor and Francis Group (ed.)).

Bilder, C. R., & Loughin, T. M. (2024). Analysis of categorical data with R. In *Analysis of Categorical Data with R*.

Christensen, R. (2025). *Log-linear Models and Logistic Regression* (Springer (ed.); 4th ed.).

Hammond, C., van der Heijden, P. G. M., & Smith, P. A. (2024). Generating contingency tables with fixed marginal probabilities and dependence structures described by loglinear models. *Journal of Statistical Computation and Simulation*, *94*(12), 2797–2812.

Jaka, N. (2022). Perkembangan Analisis Data Kategorik dan Tantangan di Era Big Data. *Pidato Pengukuhan Guru Besar Rapat Terbuka Senat Universitas Islam Indonesia*. Yogyakarta: Universitas ISlam Indonesia.

Salsabila, A., Fitrianto, A., & Aliu, M. A. (2024). Association of Poverty Categories, Educational Characteristics, and Area of Residence in Indonesia Using a Three-Way Log-Linear Model. *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, *17*(1), 624-634.

Sihotang, Siti Fatimah, and Zuhri Zuhri. "Analisis Model Log Linier Tiga Dimensi Untuk Data Kualitatif Dengan Metode Forward." *Journal of Mathematics Education and Science* 6.1 (2020): 2528-4363.

Zhang, J., Yang, Y., & Ding, J. (2023). Information criteria for model selection. *Wiley Interdisciplinary Reviews: Computational Statistics*, *15*(5).