Journal of Actuarial, Finance and Risk Management (JAFRM)
Vol. 1(1), June, 2022.

Log Linear Model on Contingency Table to Analyze Relationship between Age, Income, and Health Insurance Ownership

Evelyn Priscilla [1], Jeslyn Prinssesa [2], Mei Siang Jemima Aurelia [3], Edwin Setiawan Nugraha [4*]

[1,2,3,4] *Study Program of Actuarial Science, School of Business, President University, 17550, Indonesia*

**Corresponding author:edwin.nugraha@president.ac.id*

Abstract— Health insurance is a type of insurance that is important for everyone to have since it has benefits as protection against health risks that may occur in the future. Unfortunately, most people nowadays do not really want health insurance, especially people who are relatively young and have low incomes. Young people feel that they are still strong and do not get sick easily, while people with low incomes cannot afford to buy insurance because of the high premium prices. Therefore, the relationship between age, income, and insurance ownership (other than BPJS) needs to be known to help insurance companies develop new strategies. In this study, we implemented a Log-Linear model on a contingency table using survey data that we took in Jabodetabek, Bali, and Kalimantan areas. The results showed that the Log-Linear model (OI.OA.IA) was efficient enough to determine the relationship between age, income, and insurance ownership with a 95% confidence level. Homogeneous interactions happened so that there is no relationship between age, income, and insurance ownership, but there were relationships between age and income, age and insurance ownership, and income and insurance ownership. This research is expected to assist insurance companies in determining their target market and developing their marketing.

Keywords— Age; Contingency Table; Health Insurance Ownership; Income; Log-linear Model.

I. INTRODUCTION

Health insurance is a type of insurance product that specifically covers the health or care costs of the owners or customers of the insurance if they fall ill or have an accident [1]. Nowadays, diseases continue to develop with age, so it is important to have insurance as a form of early protection against something that might happen in the future. Health insurance, in general, can be in the form of private insurance and government insurance. One example of well-known government insurance in Indonesia is BPJS. This government health insurance is oriented towards providing health insurance to the community as a whole. In its implementation, BPJS uses the principle of "mutual cooperation" so that it can help people from middle to lower economic conditions. While people with middle to upper economic conditions, usually use private insurance because private insurance coverage can be more tailored to their health needs [2].

Insurance is needed to anticipate things that are not desirable. However, in many cases, people do not have insurance because they are hindered by economic conditions. The premium they need to pay to insurance companies is relatively high because they will also get a sizeable claim from insurance if something happens to their health. In addition, awareness and education of insurance ownership are still low. Many people feel that they are still relatively young and strong, so they do not need insurance to protect them [3]. Regarding to these problems, many health insurance companies have difficulty in making regulations regarding their target market and marketing strategy.

Therefore, it is necessary to analyze the association or relationship between age, income, and health insurance ownership. Researchers who have also carried out this analysis are Jessie Lie in 2011 who analyzed the effects between age and income on Individual Health Insurance Premiums in the United States. She used the OLS (Ordinary Least Squares) analysis method and 2SLS (Two-Stages Least Squares) regressions in determining the effect of these variables. From the results of the analysis obtained, age and income do have a relationship with insurance. But she suggested that other researchers could analyze with other methods in order to get more accurate results. In Indonesia, no one has done this analysis. This encourages the authors to determine the relationship between age and income with insurance ownership in the specific area in Indonesia, which are Jabodetabek, Bali, and Kalimantan since the authors live around those areas, by using a more accurate method, namely Log-Linear model and contingency table.

Log-Linear model on contingency table is a statistical method that can be applied to cases of quantitative data. These two methods are chosen since contingency tables are commonly used to determine the relationship between variables, while the log-linear model is usually used to determine the risk or influence of each category of a

variableon other variables [4]. In this study, the type of data used is primary data taken based on survey results. The purpose of this study is to analyze the relationship between age, income, and health insurance ownership in Jabodetabek, Bali, and Kalimantan so that it can be used as a reference for insurance companies to make marketing strategies and set their target market.
Below are several additional information regarding to the variables.

A. *Age Classification*

Age status is one of the benchmarks to determine its relationship with health insurance. As people get older, the level of concern for health is also getting higher. According to the Indonesian Ministry of Health, age can be divided into several categories, which are: Children – youth for people who are younger than 20 years old, adults for people who are 20 until 45 years old, and seniors for people who are older than 45 years old [5].

B. *Provincial Minimum Wage (UMP)*

UMP is the minimum wage made by the government for all cities in a province. Previously, the Provincial Minimum Wage was known as the Level I Regional Minimum Wage. The legal basis for determining the UMP is the Regulation of the Minister of Manpower and Transmigration Number 7 of 2013 concerning Minimum Wages. The UMP is determined by the governor by taking into account the recommendations of the Provincial Wage Council [6].
Decent Living Needs (KHL) is the guide in determining the minimum wage by taking into economic growth and account productivity. The decent living needs component is calculated based on the living needs of workers in meeting basic needs which include the need for 2100kcal per day of food, housing, clothing, education, and so on [7].

II. METHOD

A. *Contingency Table*

A contingency table is a form of compiling data that is simple enough to see the relationship between several variables in one table. In general, the three-way contingency table has (I x J x K) cells, consisting of I rows, J layers, and K columns [8]. The I x J x K contingency table can be presented in the table as shown below.

TABLE 1.
THREE-WAY CONTINGENCY TABLE

Factor A	Factor B	Factor C				Total
		C_1	C_2	...	C_k	
A_1	B_1	n_{111}	n_{112}	...	n_{11k}	n_{11+}
	...	...	...	...	...	...
	B_j	n_{1j1}	n_{1j2}	...	n_{1jk}	n_{1j+}
A_2	B_1	n_{211}	n_{212}	...	n_{21k}	n_{21+}
	...	...	...	...	...	...
	B_j	n_{2j1}	n_{2j2}	...	n_{2jk}	n_{2j+}
...	...	...	...	...	...	...
A_i	B_1	n_{i11}	n_{i12}	...	n_{i1k}	n_{i1+}
	...	...	...	...	...	...
	B_j	n_{ij1}	n_{ij2}	...	n_{ijk}	n_{ij+}
Total		n_{++1}	n_{++2}	...	n_{++k}	n_{ijk}/n_{+++}

Source: Unpublished Negeri Yogyakarta University Thesis

where,
n_{ijk} = the number of observations in row i, layer j, and column k
n_{+++} = the number of all observations
The three variables are said to be independent if

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k} \quad (1)$$

Where,
π_{ijk} = probability of observation in row i, layer j, and column k
π_{i++} = probability of observation in rows i
π_{+j+} = probability of observation in layer j
π_{++k} = probability of observations in column k
The expected frequency for the observed cell values (i,j,k) is μ_{ijk}, so it can be said that $\mu_{ijk} = n\,\pi_{ijk}$. The probability of the cell is the proportion of the sample. Suppose $\pi_{ijk} = \hat{\pi}_{i++}\hat{\pi}_{+j+}\hat{\pi}_{++k}$, each sample proportion is

$$\hat{\pi}_{i++} = \frac{n_{i++}}{n}, \hat{\pi}_{+j+} = \frac{n_{+j+}}{n}, and\ \hat{\pi}_{++k} = \frac{n_{++k}}{n} \quad (2)$$

B. Log-Linear Model

Log-Linear model is a model to obtain a statistical model which states the relationship between categorical variables. Log-Linear model can be divided into 5 models, namely:

1) Independent model

The independent model in Log-Linear means that the factor of A, B, and C are in the model, but there is no interaction between the three variables. The form of the independent model is as follows:

$$log(\mu_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C \quad (3)$$

where,

μ_{ijk} = the expected frequency of each *ijk* cell in the model

λ = overall average effect

λ_i^A = influence level *i* factor A

λ_j^B = influence level *j* factor B

λ_k^C = influence level *k* factor C

$$\lambda = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\log(\mu_{ijk})}{IJK} \quad (4)$$

$$\lambda_i^A = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}\log(\mu_{ijk})}{JK} - \lambda \quad (5)$$

$$\lambda_j^B = \frac{\sum_{i=1}^{I}\sum_{k=1}^{K}\log(\mu_{ijk})}{IK} - \lambda \quad (6)$$

$$\lambda_k^C = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\log(\mu_{ijk})}{IJ} - \lambda \quad (7)$$

This model has the condition,

$$\sum_{i=1}^{I}\lambda_i^A = 0, \sum_{j=1}^{J}\lambda_j^B = 0, and \sum_{k=1}^{K}\lambda_k^C = 0 \quad (8)$$

If the three variables are mutually independent (A,B,C), then the expected frequency estimators of each cell are as follows:

$$\mu_{ijk} = \frac{n_{i++}n_{+j+}n_{++k}}{n_{+++}^2} \quad (9)$$

2) Saturated model

The saturated model is a model that contains all possible parameters and other parameters that cannot be entered. The saturated model can be stated as follows:

$$log(\mu_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \quad (10)$$

where,

λ_{ij}^{AB} = the influence of cell interaction factors - *ij*

λ_{ik}^{AC} = the influence of cell interaction factors – *ik*

λ_{jk}^{BC} = the influence of cell interaction factors – *jk*

λ_{ijk}^{ABC} = the influence of cell interaction factors – *ijk*

$$\lambda_{ij}^{AB} = \frac{\sum_{k=1}^{K}\log(\mu_{ijk})}{K} - \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}\log(\mu_{ijk})}{JK} - \frac{\sum_{i=1}^{I}\sum_{k=1}^{K}\log(\mu_{ijk})}{IK} - \lambda \quad (11)$$

$$\lambda_{jk}^{BC} = \frac{\sum_{j=1}^{J}\log(\mu_{ijk})}{J} - \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}\log(\mu_{ijk})}{JK} - \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\log(\mu_{ijk})}{IJ} - \lambda \quad (12)$$

$$\lambda_{ik}^{AC} = \frac{\sum_{i=1}^{I}\log(\mu_{ijk})}{I} - \frac{\sum_{i=1}^{I}\sum_{k=1}^{K}\log(\mu_{ijk})}{IK} - \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\log(\mu_{ijk})}{IJ} - \lambda \quad (13)$$

$$\lambda_{ijk}^{ABC} = log(\mu_{ijk}) - \frac{\sum_{i=1}^{I}\log(\mu_{ijk})}{I} - \frac{\sum_{j=1}^{J}\log(\mu_{ijk})}{J} - \frac{\sum_{k=1}^{K}\log(\mu_{ijk})}{K} - \frac{\sum_{j=1}^{J}\sum_{k=1}^{K}\log(\mu_{ijk})}{JK} - \frac{\sum_{i=1}^{I}\sum_{k=1}^{K}\log(\mu_{ijk})}{IK} - \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\log(\mu_{ijk})}{IJ} \quad (14)$$

This model has the condition:

$$\sum_{i=1}^{I} \lambda_i^A = \sum_{j=1}^{J} \lambda_j^B = \sum_{k=1}^{K} \lambda_j^C = \sum_{i=1}^{I} \lambda_{ij}^{AB} = \cdots = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \lambda_{ijk}^{ABC} = 0 \quad (15)$$

Its expectation frequency estimator is

$$\mu_{ijk} = n_{ijk} \quad (16)$$

3) *Partial independence model*

The partial independence model is usually called as one interaction of two variables model. The form of one interaction of two variables in log-linear is:

$$log(\mu_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_l^C + \lambda_{ij}^{AB} \text{ or } \lambda_{ik}^{AC} \text{ or } \lambda_{jk}^{BC} \quad (17)$$

The expected frequency estimator for each cell is:

$$\mu_{ijk} = \frac{n_{ij+}n_{++k}}{n_{+++}} \quad (18)$$

4) *Conditional independence model*

The conditional independence model is usually called as two interaction of two variables model. The form of two interactions of two variables in log-linear is:

$$log(\mu_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} \text{ or } \lambda_{ij}^{AB} + \lambda_{jk}^{BC} \text{ or } \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \quad (19)$$

The expected frequency estimator for each cell is:

$$\mu_{ijk} = \frac{n_{ij+}n_{i+k}}{n_{i++}} \quad (20)$$

5) *Homogeneous interaction model*

The form of homogeneous interaction model in log linear is:

$$log(\mu_{ijk}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \quad (21)$$

The expected frequency estimator for each cell is:

$$\mu_{ijk} = \frac{n_{ij+}n_{i+k}n_{+jk}}{n_{i++}n_{+j+}n_{++k}} \quad (22)$$

C. Odds Ratio

The odds ratio is one of various statistics used to measure the particular event risk when a specific factor is present. The odds ratio is the ratio of two odds that is used as a descriptive statistic [9].

$$OR = \frac{odds1}{odds2} = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)} \quad (23)$$

D. Hypothesis Test

Hypothesis testing is a temporary answer to a problem that is presumptive because it still has to be proven. The presumed answer is a temporary truth, which will be verified by data collected through research [10]. In the Log-Linear model on a three-way contingency table, there are several hypotheses used in the analysis. The following tests are performed in the following steps:

1) The goodness of Fit Test

The first step is to apply the goodness of fit test. The Goodness of Fit Test in principle aims to find out whether a data distribution from the sample follows a certain theoretical distribution or not [9]. The formula of the Goodness of Fit Test can be stated as

$$D = 2\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left\{ y_{ijk} log\left(\frac{y_{ijk}}{n_{ijk}\hat{\pi}_{ijk}}\right) + (n_{ijk} - y_{ijk}) log\left(\frac{n_{ijk}-y_{ijk}}{n_{ijk}-n_{ijk}\hat{\pi}_{ijk}}\right) \right\} \quad (24)$$

n_{ijk} = the number of observations in row i, layer j, and column k
y_{ijk} = probability of success in row i, layer j, and column k
$\hat{\pi}_{ijk}$ = probability of observations in row i, layer j, and column k

D1: deviance value model (1)
D2: deviance value model (2)
The difference between the two deviance values will have a Chi-Square distribution. H_0 is rejected if the difference between the two deviances is greater than the value of the Chi-Square table.
The test statistic used in the Goodness of Fit Test is the Pearson's Chi-Square statistic with the following formula:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{(n_{ijk}-\mu_{ijk})^2}{\mu_{ijk}} \quad (25)$$

Apart from the chi-square statistic, the likelihood ratio statistic can also be used. This statistic is a chi-square approach, with the following formula:

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} n_{ijk}\log\left(\frac{n_{ijk}}{\mu_{ijk}}\right) \quad (26)$$

2) Best Model Selection

After doing the goodness of fit test, the next step is to do select the best model. Selecting the best model starts from determining the hypothesis. The model that has the weakest effect will be eliminated, leaving one model that matches the observed data. The steps taken in selecting the best model are as follows [11]:

a. Consider model (2) is the ABC model as the best model.
b. For model (1), take all the possible tested models to be used as the best model.
c. Compare model (1) and model (2) with the following hypotheses:
   H0: model (1) is the best model
   H1: model (1) is not the best model

d. Test statistics used: ${\chi^2}_{calculation} \le {\chi^2}_{table}$ or $G^2 \le {\chi^2}_{table}$ and $p - value(sig) > \alpha$ so failed to reject H_0.
   - If H_0 is rejected, then it is stated that model (2) is the best.
   - If it fails to reject H_0, then find the best model from the possible models by using the largest p-value and the smallest G^2 [12].

E. Data Type and Sources

This study uses quantitative data and primary data which aims to develop and analyze numerical data from survey results in the form of the number of health insurance ownership in Indonesia based on age and income. To obtain this data, the authors conducted a survey containing three variables: age categories, monthly income, and ownership of health insurance other than BPJS to 267 respondents who live in Jabodetabek, Bali, and Kalimantan [13].

F. Population

According to Sekaran and Bougie (2011), they define a population as a collection of individuals or group objects that researchers are interested in analyzing [14]. Therefore, there are several parts of the population that will be selected for research. The population of this research is the author's friends and their parents who lived around Jabodetabek, Bali, and Kalimantan.

G. Sample

Sekaran and Bougie (2011) defined a sample as a subset of population which are going to assess the problems in the research. In determining the sampling, the researchers set the criteria and purposes of the study. Later on, the authors decided to use snowball sampling, which is a sampling method where the survey was chained from a respondent to another, since the authors need respondents from several specific areas and several types of age categories, which can be obtained if the respondents encourage their family to also fill the survey. A total sample of 267 data is used for this research, which satisfies the criteria as below.

1) People aged <20 years, 20-45 years, and > 45 years.
2) People whose income is less than UMP or more than UMP.
3) People who have or do not have health insurance other than BPJS.

H. Sampling Technique

The technique used in this study is simple random sampling techniques. This technique explains when the chance that each element of the population has the same to be selected as a sample. the reduction of the potential for human bias is the goal of a simple random sampling technique when inclusion selection events in the sample occur. The author uses this technique to get a general opinion of the participants. In addition, this technique is also considered fairer in selecting a sample, so every participant has an equal chance to be selected. Therefore, the procedure used to obtain data is to randomly distribute surveys to participants who meet the criteria.

I. Data Collection

This research will gather data from a survey that had been done by the authors. The survey consists of the name of the respondent (optional) and choice of age, income, and health insurance ownership category. The procedure of handling the data will be conducted as follows:

1) Use Google Form as the platform to make a survey and receive the answers from the respondent.
2) Use Microsoft Excel 2016 in transforming the raw data which are obtained from Google Form.

J. Data Analysis

This research uses R software to assist in processing statistical data precisely and quickly to produce various outputs in this research. To analyze the issue quantitatively, the survey data will be transformed into a three-way contingency table to simplify calculations when analyzing data. For the next step, the data will be analyzed using the log-linear model with a hypothesis test. The steps of analyzing data with the log-linear model are as follows.

1) Hypothesis test: Define the null hypothesis and the alternative hypothesis that will be used in this research.
2) Level of significance: Determine the α to know the tolerance limit in accepting the error of the hypothesized result.
3) Expected value: The result is needed to assist in calculating the goodness of fit test.
4) The goodness of fit test: This statistical analysis is used to find out the best model that meets the criteria ($\chi^2_{calculation} \leq \chi^2_{table}$, $G^2 \leq \chi^2_{table}$, and p-value > α).
5) Model selection: Find the best model that has the largest p-value and the smallest G^2 if the model that passes the test is more than one.
6) Odds ratio: Association patterns from the best model can be determined.
7) Deviance: The association for each model can be known.
8) Residual: Strong and weak associations can be known in general.
9) Parameter estimation: Estimate the λ to construct the Log-Linear model.
10) Decision: Failed to reject H0 if there is one model that has $\chi^2_{calculation} \leq \chi^2_{table}$, $G^2 \leq \chi^2_{table}$, and p-value > α.
11) Conclusion: Make a conclusion based on the decision.

III. Result and Discussion

A. Data Preparation

The data is obtained from a survey with a total of 267 respondents. For the survey, there are three variables, namely age categories (<20 years, 20-45 years,> 45 years), monthly income (<UMP,> UMP), and ownership of health insurance other than BPJS (Yes, No). Below is the contingency table of the data.

TABLE 2
THREE-WAY CONTINGENCY TABLE FROM SURVEY DATA

Health insurance ownership	Income	Age			Total
		<20	20-45	>45	
Yes	<UMP	21	26	19	66
	>UMP	12	29	37	78
	Total	33	55	56	144
No	<UMP	25	31	20	76
	>UMP	17	15	15	47
	Total	42	46	35	123
Total		75	101	91	267

Source: Survey data via Google form conducted by the authors

B. Hypothesis Test

The hypothesis test for log-linear model on three-way contingency table can be stated as follows:

H_0: The tested model is the best model

H_1: Saturated model is the best model

The level of significance for this hypothesis test is 95% and the decision is fail to reject H_0 if the tested model has $\chi^2_{calculation} \leq \chi^2_{table}$, $G^2 \leq \chi^2_{table}$, and p-value > 0.05. The p-value will be used in sections 4.4 and 4.5 in determining the goodness of fit test and the best model.

TABLE 3
EXPECTED FREQUENCY ESTIMATION

Ownership (O)	Yes						No					
Income (I)	<UMP			>UMP			<UMP			>UMP		
Age (A)	<20	20-45	>45	<20	20-45	>45	<20	20-45	>45	<20	20-45	>45
OIA	21	26	19	12	29	37	25	31	20	17	15	15
OI.OA.IA	17.66	27.45	20.89	15.34	27.55	35.11	28.34	29.55	18.11	13.66	16.45	16.89
OA.IA	20.24	31.04	24.00	12.76	23.96	32.00	25.76	25.96	15.00	16.24	20.04	20.00
OI.A	18.54	24.97	22.50	21.91	29.51	26.58	21.35	28.75	25.90	13.20	17.78	16.02
O.I.A	21.51	28.97	26.10	18.94	25.50	22.98	18.38	24.75	22.30	16.18	21.78	19.63
OI.OA	15.13	25.21	25.66	17.87	29.79	30.33	25.95	28.42	21.63	16.05	17.58	13.37
OI.IA	21.38	26.49	18.13	18.10	27.46	32.45	24.62	30.51	20.87	10.91	16.54	19.55
OA.I	17.55	29.25	29.78	15.45	25.75	26.22	22.38	24.46	18.61	19.66	21.54	16.39
IA.O	24.81	30.74	21.03	15.64	23.73	28.05	21.19	26.26	17.97	13.36	20.27	23.96

Source: *RStudio*

C. *Expected Frequency Estimation*

The expected value must be known to assist in calculating the goodness of fit test. We are trying to know which model has the closest values to the original data or the saturated model. Model (OI, OA, IA) has the smallest difference value with the observed data or the saturated model (OIA). Below is the expected frequency of all the tested models.

D. *The goodness of Fit Test*

By utilizing the help of *RStudio*, the statistics value of Likelihood Ratio (G^2), statistics value of Pearson' Chi-Squared (X^2), degree of freedom (df), p-value of G^2, and p-value of X^2 can be obtained for all the tested models. The models that passed the goodness of fit test are detailed as shown below.

TABLE 4
GOODNESS OF FIT TEST

Model	G^2	X^2	df	p-value (G^2)	p-value (X^2)
OI.OA.IA	3.59	3.60	2	0.17	0.16
OI.OA	8.57	8.48	4	0.07	0.08
OI.IA	7.35	7.50	4	0.12	0.11

Source: *RStudio*

From the table above, model (OI, OA, IA), model (OI, OA), and model (OI, IA) are models that fit with the data because those three models have $\chi^2{}_{calculation} \leq \chi^2{}_{table}$, $G^2 \leq \chi^2{}_{table}$, and p-value > 0.05 for both G^2 and χ^2 with $\chi^2{}_{table}$ for df 2 is 5.99 and for df 4 is 9.49. Since there are 3 tested models, researchers have to find only one best model.

E. *Best Model Selection*

To get the best model, choose a model that has a relatively small G^2 value and a large p-value among the appropriate model combinations. From those three models, the model (OI, OA, IA) is the best model because this model has the smallest likelihood ratio statistic value and the highest p-value, both for G^2 and X^2. So the decision for this hypothesis is failed to reject H_0 since $\boldsymbol{\chi^2{}_{calculation}(3.60)} \leq \boldsymbol{\chi^2{}_{table}(5.99)}, \boldsymbol{G^2(3.59)} \leq \boldsymbol{\chi^2{}_{table}\ (5.99)}$, and p-value (0.17 and 0.16) > 0.05. The conclusion is the tested model (OI, OA, IA) is the best model.

F. *Odds Ratio*

After the best model has been found, the odds ratio is needed to know the association patterns for models OI, OA, and IA. The odd ratio can be illustrated by presenting estimated marginal and conditional associations.

For conditional association:
1) Between income and age:
Odd ratio = 1.16
2) Between insurance ownership and age:
Odd ratio = 0.67
3) Between insurance ownership and income:
Odd ratio = 0.55

While for the marginal association:
1) Between income and age:
Odd ratio = 1.22
2) Between insurance ownership and age:
Odd ratio = 0.66
3) Between insurance ownership and income:
Odd ratio = 0.52

From the results above, it can be interpreted as below.
1) The OI in conditional association for this model equals 0.55. It means for each level of A, people who have the income >UMP and have the health insurance are 0.55 times more likely than people who have the income <UMP. The OI marginal association of 0.52 does not consider another factor, which is factor A, whereas the conditional association of 0.55 consider about it.
2) The OA in conditional association for this model equals 0.67. It means for each level of I, people who have an older age and have health insurance are 0.67 times more likely than people who have a younger age. The OA marginal association of 0.66 does not consider another factor, which is factor I, whereas the conditional association of 0.67 consider about it.
3) The IA in conditional association for this model equals 1.16. It means for each level of O, people who have an older age and have an income >UMP are 1.16 times more likely than people who have a younger age. The IA marginal association of 1.22 does not consider about another factor, which is factor O, whereas the conditional association of 1.16 consider about it.

G. Deviance

Deviance is used to know the association for model (OI), model (OA), and model (IA). The deviances for the models are given as follows:

TABLE 5
DEVIANCE

	ANOVA 1	ANOVA 2	ANOVA 3
	Deviance	Deviance	Deviance
Model1	8.57	7.35	9.04
Model2	3.59	3.59	3.59

Source: *RStudio*

From the results, it can be interpreted as below.
1) The first ANOVA will be used to find the association between income and age on the selected model.
D1 - D2 = 8.57 – 3.59 = 4.98
4.98 < χ^2 table (5.99)
This shows that there is an association between income and age.

2) The second ANOVA will be used to find the association between ownership and age on the selected model.
D1 - D2 = 7.35 – 3.59 = 3.76
3.76 < $\chi 2$ table (5.99).
This shows that there is an association between ownership and age.

3) The third ANOVA will be used to find the association between income and ownership on the selected model.
D1 - D2 = 9.04 – 3.59 = 5.45
5.45 < $\chi 2$ table (5.99).
This shows that there is an association between income and ownership.

H. Residual

For residual, it has a standard normal distribution. A weak association is indicated by absolute values larger than 2 for the few cells and 3 for many cells. Because the data uses a three-way contingency table, residuals that are larger than 3 have weak associations between each factor.

TABLE 6
RESIDUAL

Health insurance ownership	Income	Age		
		<20	20-45	>45
Yes	<UMP	-3.34	1.45	1.89
	>UMP	3.34	-1.45	-1.89
No	<UMP	3.34	-1.45	-1.89
	>UMP	-3.34	1.45	1.89

Source: *RStudio*

According to the residual data obtained, people who have ages <20 years with any conditions have standard residuals absolute values greater than 3. This indicates that they have weak associations between each factor, while the others have a strong association between each factor.

I. Parameter Estimation

To estimate the parameter, the *glm* function in *Rstudio* will be used to assist in analyzing the data. Each parameter λ for model (OI, OA, IA) can be obtained using *RStudio*. Below is the Log Linear model for model (OI, OA, IA) is:

$\log(\mu_{ijk}) = \lambda + \lambda_i^{A} + \lambda_j^{I} + \lambda_k^{O} + \lambda_{ij}^{AI} + \lambda_{ik}^{AO} + \lambda_{jk}^{IO}$

$\log(\mu_{ijk}) = 2.85 - 0.11(A) + 0.16(I) + 0.74(O) + 0.31(AI) - 0.28(AO) - 0.59(IO)$

IV. CONCLUSION

Homogeneous interaction (model OI.OA.IA) is the best model to know the relationship between each factor for this case. This is proved by no interaction between insurance ownership, age, and income simultaneously; while there are interactions between insurance ownership and income, insurance ownership, and age, as well as income and age. By this result, the insurance company will be able to obtain the relation of age, income, and ownership so that the company can make the best strategy.

REFERENCES

[1] Suryono, A. (2009). Asuransi Kesehatan Berdasarkan Undang-Undang Nomor 3 Tahun 1992. Jurnal Dinamika Hukum, 9(3), 213-221.

[2] Vandawati, Z., & Sabrie, H.Y. (2016). Aspek Hukum Kartu Indonesia Sehat. Yuridika, 31(3), 498-520.

[3] Lie, J. (2011). Effects of Age and Income on Individual Health Insurance Premiums. Undergraduate Economic Review, 7(1), 1-18.

[4] Simonoff, J. S. (2003). Analyzing Categorical Data. New York: Springer.

[5] Amin, M.A. (2017). Klasifikasi Kelompok Umur Manusia Berdasarkan Analisis Dimensi Fraktal Box Counting Dari Citra Wajah Dengan Deteksi Tepi Canny. Mathunesa Jurnal Ilmiah Matematika, 2(6), 33-42.

[6] Pratomo, D. S. (2011). Kebijakan Upah Minimum untuk Perekonomian yang Berkeadilan : Tinjauan UUD 1945. Journal of Indonesian Applied Economics, 5(2), 269-285.

[7] Mahila, S. (2014). Kebutuhan Hidup Layak dan Pengaruhnya Terhadap Penetapan Upah Minimum Provinsi Ditinjau dari Hukum Ketenagakerjaan. Jurnal Ilmiah Universitas Batanghari Jambi, 14(2), 42-51.

[8] Hapsari, G. (2011). Model Log Linear Untuk Tabel Kontingensi Tak Sempurna Berdimensi Tiga [Unpublished Thesis]. Negeri Yogyakarta University.

[9] Nugraha, J. (2013). Pengantar Analisis Data Kategorik: Metode dan Aplikasi menggunakan Program R. Yogyakarta: CV Budi Utama.

[10] Furchan, A. (2004). Pengantar Penelitian dalam Pendidikan. Yogyakarta: Pustaka Pelajar.

[11] Nurman, T. A. (2013). Analisis Data Kategori Dengan Log Linier Menggunakan Prinsip Hirarki (Studi Kasus Jumlah Kecelakaan Lalu Lintas di Kota Makassar Tahun 2011). Jurnal Teknosains, 7(1), 99-110.

[12] Lestyorini, M. (2010). Model Log Linear Multivariat Empat Dimensi [Unpublished Thesis]. Negeri Yogyakarta University.
[13] Abdi, H. (2021, March 8). Tujuan Penelitian Kualitatif dan Kuantitatif, Kenali Perbedaannya. Liputan 6. https://hot.liputan6.com/read/4500971/tujuan-penelitian-kualitatif-dan-kuantitatifkenaliperbedaannya.
[14] Sekaran, U., & Bougie, R. (2011). Research Methods for Business. Great Britain: Wiley.