Determination of the optimallogarithmic model for the data of the contingency tables with a practical application in the field of road accidents

Aseel moobder dawood Almarsoumy
Middle technical university ,Rusafa Institute of Administration
Almarsoumy66@yahoo.com

Osama Abdulazeez kadhim Al-Quraish
Directorate of Baghdad Education Karkhlll , Ministry of Education, iraq
ousamastat@gmail.com
orcid : 0000-0002-1701-2679

Abstract
Road accidents are random or accidental incidents that are subject to intermittent distributions ,the most important distributions are (binomial distribution , Poisson distribution , negative binomial distribution) , these incidents are categorical according to the nature of the accident (fatal, fatalwith wounded , wounded only), in this research , we considered the impact of the type of accident (collision , overturned , run over) and the scene (inside the city , outside the city) and it's impact on the nature of the incident individually on the one hand and on the other hand using the contingency tables of class (r×c×k) and estimate the logarithmic model for each case and selecting the optimal model by adopting the maximum likelihood ratio using road accident data in the holy province of Karbala .
key words:Logarithmic model, contingency tables,Road accidents, maximum likelihood ratio

1- Introduction[6][1][12]

Mainstream data is generally interested in studying variables in which data can be categorical intocontingency tables containing responses, for example(Successful / failed, responsive / unresponsive, etc.) Itcan also be defined as standard data according to its relationship with the corresponding groups **,** Marginal groups also indicate the treatment that the researcher relies on in his study, while theresponse indicates a description of the nature of the worker in other words, describing what happensto the worker during the experiment(Success / failure, injured / uninfected, etc.) The data from these experiments are called counted data**.** This type of data appears in many phenomena such as road accidents, floods, environmental studies**,** medical and biological experimentsAnd that these data follow different statistical distributions such as **(**Poisson, Binomial, multinomial)also contingency tables can be defined as repetitions arranged in tables or are the mathematical method used to represent the categorical data. These tables contain two or more variables and these variables are quantitative variablessuch as height, weight, temperature, or be specific variables like hair color or eye color **,** The data categorical as groups or classes can be expressed, and each variable contains a number of levels or classifier Another definition, the most important of which is contingency tables, is the definition that depends on information theoryThis indicates that the contingency tables are essentially the same from a multi-valued community with different possibilities for groups linked to each other as a random sample assumption of

independent observations and each view can be categorical(m) from classification criteria for example (R) for the first factor and (C) for the second factor and (K) for the third factor ... etc**.**
Many researchers addressed the study of methods of data categorical in different ways. Therefore, in this paper, we considered the use of the logarithmic model by modeling the Poisson probability distribution by means of the zigzag method and converting it into the logarithmic linear formula for the contingency tables of the rank (r × c × k) to study the effect of each of the two search variables. (Independent variables) in the approved variable (as a result of traffic accidents)**.**
In analyzing the data categorical for the contingency tables of the rank (r × c × k), there are many different interactions, as estimating the saturated model (which contains all the bilateral interactions and the triple reaction) may be insignificant or not representative of the data accurately as well as the need to know the relationship Between the outcome of road accidents, the type of accident, and the location of the accident, it is therefore necessary to reduce these interactions in the estimated models and choose the best ones by comparison using the maximum possible ratio of each**.**
Several studies and methods have emerged that dealt with the study of categorical data, but few of these studies dealt with estimating the linear logarithmic model of the categorical data and that the main goal of this research is to determine the optimal logarithmic model for road accident data in the holy Karbala governorate for the purpose of showing the impact of the search variables traffic type (overturned, Collision, run-over) and the location of the traffic accident (inside the city, outside the city) and its impact with its ramifications on the dependent variable as a result of the traffic accident (deadly, fatalwith wounded, injured only) as the comparison between the estimated models is done by calculating the maximum possible ratio**.**

2- Liner algorithmic model (3×3)[5][2][3][11]

A linear logarithmic model is defined as the method by which the dependent variable is modeled as a response variable when the values of that variable are in the form of numbered values or in the form of averages, as it is defined as one of the types of models that fall under the umbrella of linear-logarithmic regression models and that the logarithmic linear analysis test is an extension To the square quay test, since in the quay square test, it calculates the expected frequency of cells, and by taking the natural logarithm of the distribution formula, it turns into a linear formula.
And that the linear logarithmic model under the 3-dimensional (3x3) contingency table called the saturated model can be expressed mathematically as follows:

$$\boldsymbol{\log(E_{ijm}) = \lambda + \lambda_1(i) + \lambda_2(j) + \lambda_3(m) + \lambda_{12}(ij) + \lambda_{13}(im) + \lambda_{23}(jm) + \lambda_{123(ijm)}} \quad \ldots\ldots(1)$$

As that:
λ:Represents the general mean of the model
$\lambda_1(\mathrm{i})$: represents the primary effect of the first variable of level (i)
$\lambda_2(\mathrm{j})$: represents the primary effect of the second level variable (j)
$\lambda_3(\mathrm{m})$: represents The primary effect of the third level variable (m)
$\lambda_{12}(\mathrm{ij})$: represents the primary effect between the effect of level (i) for the first variable and level (j) for the second variable
$\lambda_{13}(\mathrm{im})$: represents the primary effect between the effect of level (i) of the first variable and the level (m) of the third variable.
$\lambda_{23}(\mathrm{jm})$: represents the primary effect between the effect of level (j) of the second variable and the level (m) of the third variable.
$\lambda_{123}(\mathrm{ijm})$: represents the primary effect between the effect of level (i) of the first variable and level (j) of the second variable and level (m) of the third variable.

The model above includes the presence of the main effects and interactions of the first and second degrees, and when the model contains a common interaction between rows, columns and cells $\lambda_{123}(\mathrm{ijm})$ for the first-order effects and interactions must be available for the model to be acceptable.

And that the statistical assumptions associated with the above model can be written as follows:

$$H_0: \lambda_{12} = 0 , H_0: \lambda_{13} = 0, \quad H_0: \lambda_{23} = 0 \quad , \quad H_0: \lambda_{123} = 0$$
$$H_0: \lambda_{12} \neq 0 \quad , \quad H_0: \lambda_{13} \neq 0 \quad , \quad H_0: \lambda_{23} \neq 0 \quad , \quad H_0: \lambda_{123} \neq 0 \qquad \text{......(2)}$$

Whereas the null hypothesis indicates that there is no agreement between any of the three variables, i.e., the existence of complete independence between the variables, at which point the linear logarithmic model presented in formula (1) is as follows:

$$\log(E_{ijm}) = \lambda + \lambda_1(i) + \lambda_2(j) + \lambda_3(m) \qquad \ldots\ldots(3)$$

Likewise in the model that assumes $\lambda_{12} = 0$ and $\lambda_{123} = 0$ it is according to the following formula:

$$\log(E_{ijm}) = \lambda + \lambda_1(i) + \lambda_2(j) + \lambda_3(m) + \lambda_{13}(im) + \lambda_{23}(jm) \qquad \ldots\ldots(4)$$

Where in the case $\lambda_{123} = 0$means that the interaction between the first and second variables is equal for all levels of the third variable and that the position of $\lambda_{12} = 0$ means that is equal to zero , And that the model presented in formula (3) means that there is no interaction between the first and second variables at each level of the third variable.

And the scientist (Bishop) indicated that when continuing to delete boundaries λ we get a model that contains only one term that is called the non-comparative model this is not important in practice because the contingency tables are interested in studying comparison models for categorical data .

For the purpose of estimating the parameters given in formula (1) for the linear logarithm model, we find in the first stage the expected frequencies for the cells of the contingency table (3 x 3) we calculate the expected frequencies E_{ijm}in light of the fulfillment of the independence conditions between the three variables of the cells of the contingency table according to the following formulas:

$$E_{ijm} = \frac{n_{i..}\ n_{.j.} n_{..m}}{n_{...}} \qquad \ldots\ldots(5)$$

As that:

$n_{i..}$: Indicates the marginal totals in the contingency table for the first variable (rows variable).

$n_{.j.}$: indicates the marginal totals in the contingency table for the second variable (column variable).

$n_{..m}$: indicates the marginal totals in the contingency table for the third variable (cell variable).

$n_{...}$: the sum of the frequencies of the contingency table.

$$\widehat{K}_{ijm} = \log(E_{ijm}) \qquad \ldots\ldots(6)$$

$$\left.\begin{aligned} \hat{\lambda} &= \bar{k}_{..} \\ \hat{\lambda}_1(i) &= \bar{k}_{i..} - \bar{k}_{..} \\ \hat{\lambda}_2(j) &= \bar{k}_{.j.} - \bar{k}_{..} \\ \hat{\lambda}_3(m) &= \bar{k}_{..m} - \bar{k}_{..} \end{aligned}\right\} \qquad \ldots\ldots(7)$$

For the purpose of applying good conformity tests to the linear logarithm model and its accuracy in correctly representing data, we use the following measures:

- Statistics the square quiz test χ^2 and write as follows:

$$\ldots\ldots(8)\chi_c^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{m=1}^{l}\frac{(n_{ijm}-E_{ijm})^2}{E_{ijm}}$$

Whereas, n_{ijm} represents the observed frequency of the ijm cell in the contingency table, Where the calculated Chi-squared value χ_c^2 is compared With the chi-square value of χ_t^2at the degree of freedom)df =($r\times c \times l - r - c - l + 2$) and the level of significance (0.05) If the calculated value of the Chi-squared is greater than the value of the tabular Chi-squared, the null hypothesis is rejected, that is, there is no independence between the variables.

- Its statistic is the test of maximum likelihood ratio, L^2, and it is expressed mathematically as follows:

$$\ldots\ldots(9)L^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{m=1}^{l} n_{ijm}\ln\left(\frac{n_{ijm}}{E_{ijm}}\right)$$

3-Estimation of the logarithmic model of the contingency table in the case of Poisson distribution[10][3][4][7]

Poisson distribution was known by the famous French mathematician and physicist Simon Dunes Poisson (1781-1840), and it is an intermittent probability distribution that expresses the possibility of a number of rare or unexpected events in which the probability of success is weak, within a specific time period and for a large number of Attempts are called Poisson law for small numbers.

The Poisson distribution is close to the binomial distribution when the probability of success (θ) in the binomial distribution is very small and approaches zero, and the sample size from which the Poisson distribution is measured is large .

Let y be the random variable that represents the number of times the event will occur in the time period, then this variable (y) will follow the Poisson distribution with a parameter of (θ), that the probability mass function of the Poisson distribution is written in the As follows:

$$f(y,\theta) = \frac{\theta^y \; e^{-\theta}}{y!} \quad y= 0,1,2,\ldots\ldots \qquad \ldots\ldots.(10)$$

As that:

θ: represents the probability of success and its value ranges from zero to one ($0< \theta < 1$).

Suppose each cell of the contingency table follows the Poisson distribution with the parameter (θ_{ij})meaning that:

$$n_{ij} \sim poi(\theta_{ij})$$

Whereas, the total sum of the cells of the contingency table $n_{..}$ follows the Poisson distribution with the parameter $\theta_{..}$ that is:

$$n_{..} = \sum\sum n_{ij} \sim poi(\theta_{..}) \text{ Where } \theta_{..} = \sum\sum \theta_{ij}$$

Now, let's derive the common distribution of the reconciliation table cells by $n_{..}$as follows:

$$f(n_{11}, n_{12}, n_{21}, n_{22}/n_{..}) = \frac{f(n_{11}, n_{12}, n_{21}, n_{22})}{f(n_{..})}$$

$$= \frac{\frac{e^{-\theta_{11}}\theta_{11}^{n11}}{n11\,!}\frac{e^{-\theta_{12}}\theta_{12}^{n12}}{n12\,!}\frac{e^{-\theta_{21}}\theta_{21}^{n21}}{n21\,!}\frac{e^{-\theta_{22}}\theta_{22}^{n22}}{n22\,!}}{\frac{e^{-\theta_{..}}\theta_{..}^{n_{..}}}{n_{..}\,!}}$$

Depending on the definition of the Maximum likelihood functionWe find the combined probability of the contingency table cells and estimation of Poisson distribution parameters as follows:

$$L = \prod_{i=1}^{2}\prod_{j=1}^{2} P(n_{ij}\theta_{ij})$$

$$L = \prod_{i=1}^{2}\prod_{j=1}^{2}\frac{e^{-\theta_{ij}}\theta_{ij}^{nij}}{nij\,!}$$

$$L = \frac{e^{-\sum_{i=1}^{2}\sum_{j=1}^{2}\theta_{ij}}\theta_{ij}^{\sum_{i=1}^{2}\sum_{j=1}^{2}n_{ij}}}{\prod_{i=1}^{2}\prod_{j=1}^{2}nij\,!} \quad ......(11)$$

By entering the natural logarithm of both sides of formula (11), we get:

$$\mathrm{LnL} = \sum_{i=1}^{2}\sum_{j=1}^{2}n_{ij}\mathrm{Ln}(\theta_{ij}) - \sum_{i=1}^{2}\sum_{j=1}^{2}\theta_{ij} - \mathrm{Ln}\left(\prod_{i=1}^{2}\prod_{j=1}^{2}nij\,!\right) \quad ......(12)$$

Taking the derivative of the parameter θ_{ij} with respect to the cell (i, j) and equalizing the derivative to zero we get:

$$\frac{n_{ij}}{\theta_{ij}} - 1 = 0$$

$$\hat{\theta}_{ij} = n_{ij} \quad ......(13)$$

And from it we find that:

$$\hat{\theta}_{..} = \sum_{i=1}^{r}\sum_{j=1}^{s}\hat{\theta}_{ij} = \sum_{i=1}^{r}\sum_{j=1}^{s}n_{ij} = n_{..} \quad ......(14)$$

Therefore, estimating the parameters of the logarithmic model of the contingency table is as follows:

$$\hat{P}_{ij} = \hat{\lambda}_i = \frac{\hat{\theta}_{ij}}{\hat{\theta}_{..}} = \frac{n_{ij}}{n_{..}} \quad ......(15)$$

Through what has been reached, these results can be generalized to multiple contingency tables, i.e. mathematical formulas for testing statistical assumptions and estimating the logarithmic model remain as they are.

4-Optimal logarithm selection[8][3][9]

The process of selecting the optimal logarithmic model from among models that contain interactions between variables and cells within the contingency table is summarized by relying on a set of interconnected sequential models by calculating its maximum likelihood ratio(L^2) contained in formula (9) to compounds that merge In the chain stages, the chain models gradually decrease, as follows:

$$\log(E_{ijm}) = \lambda + \lambda_1(i) + \lambda_2(j) + \lambda_3(m) + \lambda_{12}(ij) + \lambda_{13}(im) + \lambda_{23}(jm) \quad ...... (16)$$

The model contained in formula (16) is called the logarithmic model with comprehensive bilateral interactions(ij), (im) , (jm).

$$\log(E_{ijm}) = \lambda + \lambda_1(i) + \lambda_2(j) + \lambda_3(m) + \lambda_{12}(ij) + \lambda_{13}(im) \quad ......(17)$$

The model described in formula (17) is called the logarithmic model with binary interactions associated with the dependent variable(ij), (im) .

$$\log(E_{ijm}) = \lambda + \lambda_1(i) + \lambda_2(j) + \lambda_3(m) + \lambda_{12}(ij) \quad ......(18)$$

The model contained in formula (18) is called the logarithmic model with basic two-way interactions, meaning it includes the interaction between the row variable and the column variable (ij) .

$$\log(E_{ijm}) = \lambda + \lambda_1(i) + \lambda_2(j) + \lambda_3(m) \quad ......(19)$$

The model contained in formula (19) is called the general logarithmic model, meaning that it contains the basic variables (row variable, column variable, cell variable)

5- Practical application

Road accidents data was adopted in Karbala Governorate for the year 2018 as the dependent variable Y represents the result of the accident (deadly, fatal with wounded, injured only) and that the first independent variable X represents the type of accident (collision, rollover, run over), and the second independent variable represents a location The accident (inside the city, outside the city), as the numbers of these incidents are collected by police stations in all districts and districts of Karbala Governorate and these data are sent to the Police Directorate of the Department of Criminal Statistics, as these data meet international standards and indicators as the number of incidents in Karbala Governorate (587) 2018 accident, according to type More than that due to the number of collisions (245) accidents, the number due to the overturned (21) and the number due to run over (321), the number of collisions that led to fatal accidents inside and outside the city (28) accidents, and collisions that led to Fatal accidents with the injured (10) accidents The number of collisions that led to the injured only reached (207) accidents, and the number of overturned accidents that led to fatal accidents reached (6) accidents, and the number of overturned accidents that led to fatal accidents with the injured reached (7) Accidents The overturned accidents led to injured only (8) accidents. As for the number of runaway accidents that led to fatal accidents (66), accidents led to accidents His death with injuries amounted to (16) accidents, and the number of runaway accidents that resulted in injuries was only (239).

Where the variables of the research sample were categorical as follows:

(Y): The variable represented (the result of the accident), which is categorical into:

- fatal accident number (1)
- Fatal accident with the wounded, number (2)
- Injured accident only number (3)

(X): Represents the first independent variable (type of accident), which is categorical into:

- Collision number (1)
- Overturned number (2)
- Run over number (3)

(Z) : Represents the second independent variable (the accident site), which is categorical into:

- Inside the city number (1)
- Out of town number (2)

The study sample data was summarized in Table (1) and it was as follows:

Table (1)
Preparing road accidents in Karbala Governorate categorical according to type, results and location of the accident

		Y						
		Injured accident only		Fatal accident with the wounded			fatal accident	
	The results ofthe accident / The type of accident	outside the city	Inside the city	outside the city	Inside the city	outside the city	Inside the city	Total
X	Collision	85	122	3	7	9	19	245
	Overturned	7	1	6	1	5	1	21
	Run over	104	135	4	12	12	54	321
	Total	196	258	13	20	26	74	587

Source: Preparation of researchers, as data were obtained from the Ministry of Interior - Holy Karbala Police Directorate - Criminal Statistics Division.
Through the statistical program SPSS V23 and the statistical program the results were :obtained as below:

table(2)

Estimating the logarithmic model given in formula (16).

Parameter EstimatesLog Linear PoissonModel										
					95% Confidence Interval					
Parameter	Estimate	Std. Error	Z	Sig.	Lower Bound	Upper Bound	χ^2_c	L^2	d.f	Sig.
Constant	4.613	.098	47.307	.000	4.422	4.804	2.722	2.631	4	.62
[y = 1.00]	-1.917-	.240	-7.978-	.000	-2.387-	-1.446-				
[y = 2.00]	-3.140-	.402	-7.815-	.000	-3.928-	-2.353-				
[x = 1.00]	-.137-	.140	-.981-	.326	-.411-	.137				
[x = 2.00]	-2.626-	.370	-7.094-	.000	-3.352-	-1.901-				
[Z = 1.00]	.315	.126	2.499	.012	.068	.563				
[y = 1.00] * [x = 1.00]	-.711-	.247	-2.878-	.004	-1.196-	-.227-				
[y = 1.00] * [x = 2.00]	1.503	.584	2.575	.010	.359	2.647				
[y = 2.00] * [x = 1.00]	-.325-	.415	-.782-	.434	-1.138-	.489				
[y = 2.00] * [x = 2.00]	2.926	.632	4.627	.000	1.687	4.166				
[y = 1.00] * [Z = 1.00]	.923	.262	3.525	.000	.410	1.436				
[y = 2.00] * [Z = 1.00]	.666	.427	1.558	.119	-.172-	1.503				
[x = 1.00] * [Z = 1.00]	-.012-	.177	-.066-	.947	-.360-	.336				
[x = 2.00] * [Z = 1.00]	-2.647-	.659	-4.020-	.000	-3.938-	-1.356-				

From the results in the table above, we notice that the estimated logarithmic model presented in the theoretical side in the formula (16) called the logarithmic model with comprehensive binary interactions (ij), (im), (jm) was not significant, because the value of the level of significance Sig. The model has greater than the significance level (0.05), and this indicates that one or two of the binary interactions in the logarithmic model were not influential in the dependent variable.

table (3)
Estimating the logarithmic model given in formula (17).

Parameter EstimatesLog Linear PoissonModel										
					95% Confidence Interval					
Parameter	Estimate	Std. Error	Z	Sig.	Lower Bound	Upper Bound	χ_c^2	L^2	d.f	Sig.
Constant	4.636	.084	55.089	.000	4.472	4.801	27.740	27.001	6	.000
[y = 1.00]	-1.794-	.225	-7.967-	.000	-2.235-	-1.353-				
[y = 2.00]	-2.795-	.341	-8.201-	.000	-3.464-	-2.127-				
[x = 1.00]	-.144-	.095	-1.514-	.130	-.330-	.042				
[x = 2.00]	-3.397-	.359	-9.451-	.000	-4.101-	-2.693-				
[Z = 1.00]	.275	.095	2.901	.004	.089	.461				
[y = 1.00] * [x = 1.00]	-.714-	.245	-2.917-	.004	-1.193-	-.234-				
[y = 1.00] * [x = 2.00]	.999	.558	1.792	.073	-.094-	2.092				
[y = 2.00] * [x = 1.00]	-.326-	.414	-.788-	.431	-1.138-	.485				
[y = 2.00] * [x = 2.00]	2.570	.578	4.444	.000	1.437	3.704				
[y = 1.00] * [Z = 1.00]	.771	.247	3.123	.002	.287	1.255				
[y = 2.00] * [Z = 1.00]	.156	.369	.423	.672	-.567-	.878				

From the results in the table above, we notice that the estimated logarithmic model presented in the theoretical side in the formula (17) called the logarithmic model with binary interactions associated with the dependent variable (ij), (im) was significant due to the value of the level of significance Sig. The model has less than the significance level (0.05) and this indicates that the binary interactions (ij), (im) in the logarithmic model were effective in the dependent variable.

table (4)
Estimating the logarithmic model given in formula (18).

Parameter EstimatesLog Linear PoissonModel										
					95% Confidence Interval					
Parameter	Estimate	Std. Error	Z	Sig.	Lower Bound	Upper Bound	χ_c^2	L^2	d.f	Sig.
Constant	4.561	.082	55.573	.000	4.400	4.722	35.356	37.532	8	.000
[y = 1.00]	-1.287-	.139	-9.254-	.000	-1.559-	-1.014-				
[y = 2.00]	-2.704-	.258	-10.471-	.000	-3.210-	-2.198-				
[x = 1.00]	-.144-	.095	-1.514-	.130	-.330-	.042				
[x = 2.00]	-3.397-	.359	-9.451-	.000	-4.101-	-2.693-				
[Z = 1.00]	.404	.084	4.796	.000	.239	.569				
[y = 1.00] * [x = 1.00]	-.714-	.245	-2.917-	.004	-1.193-	-.234-				
[y = 1.00] * [x = 2.00]	.999	.558	1.792	.073	-.094-	2.092				
[y = 2.00] * [x = 1.00]	-.326-	.414	-.788-	.431	-1.138-	.485				
[y = 2.00] * [x = 2.00]	2.570	.578	4.444	.000	1.437	3.704				

From the results in the table above, we notice that the estimated logarithmic model presented in the theoretical side in the formula (18), which is called the logarithmic model with basic two-factor interactions, i.e. the interaction between the row variable and the column variable (ij) was significant due to the fact that the value of the level of significance Sig. The model has less than the significance level (0.05) and this indicates that the basic binary interactions (ij) in the logarithmic model were influencing the dependent variable.

table (5)

Estimating the logarithmic model given in formula (19).

Parameter EstimatesLog Linear PoissonModel										
					95% Confidence Interval					
Parameter	Estimate	Std. Error	Z	Sig.	Lower Bound	Upper Bound	χ^2_c	L^2	d.f	Sig.
Constant	4.599	.079	58.570	.000	4.445	4.753	111.467	67.946	12	.000
[y = 1.00]	-1.513-	.110	-13.696-	.000	-1.729-	-1.296-				
[y = 2.00]	-2.622-	.180	-14.541-	.000	-2.975-	-2.268-				
[x = 1.00]	-.270-	.085	-3.185-	.001	-.436-	-.104-				
[x = 2.00]	-2.727-	.225	-12.116-	.000	-3.168-	-2.286-				
[Z = 1.00]	.404	.084	4.797	.000	.239	.569				

From the results in the table above, we notice that the estimated logarithmic model included in the theoretical side in formula (19) which is called the general logarithmic model was significant due to the fact that the value of the level of significance Sig. The model has less than the significance level (0.05).

Through the results presented in Table (2), Table (3), Table (4) and Table (5) and by comparison using the criterion of L^2 and the level of significance of the estimated models by comparing the level of significance of the model with the level of significance (0.05), if the value The level of significance of the model is less than the significance level (0,05), the model is significant and vice versa, and these results can be summed up in the following table:

table (5)

Comparison of the estimated logarithmic models

Sig.	L^2	χ^2_c	d.f	Formula	Sequence
0,62	2,631	2,722	4	16	1
0.00	27,001	27,740	6	17	2
0.00	37,532	35,356	8	18	3
0.00	67,946	111,467	12	19	4

From Table (6), we note that the first model was not significant, because the value of the level of significance (Sig) is greater than the significance level (0,05). Therefore, it is excluded from the estimated models although it achieved the lowest percentage of L^2 and is equal to (2,631) compared to the models. The other models (second, third, and fourth) were significant. As for the comparison between the three moral models, we find that the logarithmic model (17) is better than the rest of the models because it achieved the lowest value of its statistic, ratioL^2.

6- Conclusions

From the results shown in Table (6), it is clear that the lowest value of maximum likelihood ratio for the moral models with a significance level of (0.05) was for the logarithmic model in the formula (17) as it reached (27.001), from which we infer:

1- Rejecting the null hypothesis that there is autonomy between the two reactions

$H_0: \lambda_{12} = 0$:

From the results in Table (3), the parameters associated with the main variables are all significant except (x = 1) which represents a fatal accident and its effect on the dependent variable (y) as a result of the accident. As for the bilateral interactions between the search variables i.e. the dependent variable (the result of accident y) and the independent variables (Type of accident x) were all significant except for the interaction between (y = 1) and (x = 2) which represents a collision and a fatal accident with the wounded, as well as the interaction between (y = 2) and (x = 1) which represents a coup and a fatal accident That is, it has no effect on the logarithmic model in the formula (19).

2- rejected the null hypothesis$H_0: \lambda_{13} = 0$:

Which states that there is independence between the dependent variable (the result of the accident) and the independent variable (the location of the accident) except for the interaction between (y = 2) and (z = 1) which represents the inversion and within the city was not significant in the logarithmic model contained in the formula (19).

7- Recommendations

Based on what was stated in the theoretical side and the applied side of the research and in light of the conclusions reached by the researchers, the researchers recommend the following:

1- Giving importance to the study of categorical data due to its wide field of application in various fields, especially data that contains multiple classes.

2- The necessity of relying on the probability distributions that correspond to the data categorical for the contingency tables and finding estimation of the models based on different methods.

3- Using other models in studying and analyzing categorical data, such as logistical models and discriminatory analysis.

References

1- Al-Abdi, Fadel Abdul-Abbas, 2012, "Method of Analyzing Linked (Un independent) Classified Data", Basra Research Journal ((Operations)), No. 38, Part 4, University of Basra, College of Education, Iraq.

2- Alan Agresti , 2007 , "An Introduction to categorical data analyses " , Cambridge University press, pp.(1-434).

3- Al-Janabi, Dawiya Salman Hasan and others, 2012, "Analysis of categorical data and its applications", first edition, Al-Jazeera Office for Printing and Publishing, Baghdad-Al-Waziriya, Iraq.

4- Amiri, S., Clarke, B., Clarke, J., and Koepke, H. A. (2015), “A General Hybrid Clustering Technique,” arXiv:1503.01183.

5- Bai, L., Liang, J., Dang, C., and Cao, F. (2011), “A Novel Attribute Weighting Algorithm for Clustering High-Dimensional Categorical Data,” *Pattern Recognition*, 44, 2843–2861.

6- Bennink, M., Croon, M. A., Kroon, B., and Vermunt, J. K. (2016). “Micromacro multilevel latent class models with multiple discrete individual-level variables.” Advances in Data Analysis and Classification, 10(2): 139–154.

7- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2016). Stan: A probabilistic programming language. Journal of Statistical Software, 76(1):1–32.

8- Ehlers , Rene ,2002 , " Maximum Likelihood Procedures for categorical data " , Submitted in partial fulfillment to the requirement for the degree magister scientiae , University of Pretoria , Department of statistics .

9- George Matysiak and Krzysztof Olszewski,(2019)." A Panel Analysis of Polish Regional Cities Residential Price Convergence in the Primary Market", *SSRN Electronic Journal*, 10.2139/ssrn.3408797.

10- Jingchen Hu, Jerome P. Reiter, and Quanli Wang. (2018)." Dirichlet Process Mixture Models for Modeling and Generating Synthetic Versions of Nested Categorical Data." Bayesian Analysis. 13, Number 1, pp. 183–200.

11- Stuart R . Lipsitzand others. (2015). " Testing for independence in $J \times K$ contingency tables with complex sample survey data Biometrics. 2015 Sep; 71(3): 832–840.

12- Winkelmann, R (2008). "Economic Analysis of Count Data", 5 th ed., Springer, Verlag Berlin Heidelberg, Germany.