The Application of Log-Linear Model to Selected Poison Patients

Fatin N.S.A., Norlida M.N. and Siti Z.M.J.[*]

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia

Log-linear model is a technique used to analyze the cross-classification categorical data or the contingency table. It is used to obtain the parsimony models that describe the interaction between the categorical variables in contingency tables. Log-linear models are commonly used in evaluating higher dimensional contingency tables that involves more than two categorical variables. This study focuses on analyzing data of poisoned patients from 2012 to 2014 using log-linear model. There are two model analyzed; model for demographic data of patients and model of poisoning information. For the first model, the variables involved are gender, age, race and state. Variables for the second model are circumstance of exposure, type of exposure, location of exposure, route of exposure and types of poison. Both log-linear models are developed to investigate the association between variables in the model. As a result of this study, the best model for demographic data and poisoning information are the model with three-ways interaction. For the best model of demographic data, there is an association between gender, age and race, race, gender and state as well as age, race and state. Meanwhile, the best model for poisoning information reveals that there is relationship between circumstance of exposure, route of exposure and type of poison, location of exposure, route of exposure and type of poison, circumstance of exposure, type of exposure and route of exposure, circumstance of exposure, location of exposure and route of exposure, circumstance of exposure, type of exposure and type of poison and also type of exposure, location of exposure and type of poison.

Keywords: log-linear; demographic; gender; age; race; state; circumstance of exposure; type of exposure; location of exposure; route of exposure; types of poison

I.   INTRODUCTION

The National Poison Center, based at Universiti Sains Malaysia, Penang was established in 1994. Its function is to deal with poison-related inquiries from the general public and healthcare practitioners. All inquiries about poisoning cases to the center are recorded. However, clearly not all inquiries are reported to the center and thus Malaysian poisoning data derived from an analysis of poisoning inquiries to the center falls short of representing the actual occurrence of poisoning in Malaysia. Most published data on poisoning in Malaysia are based on occupational-related incidences or poisoning admissions at health facilities (Rajasuriar *et al.,* 2007).

According to Rajasuriar *et al.,* (2007), there were 21714 poisoning admissions reported with 779 deaths from 1999 to 2001. The case-fatality rate was 35.88 per 1000 admissions. According to the paper, most admissions (89.7%) and deaths (98.9%) occurred in adults. Some 55.1% of all admissions were female, which most likely involving pharmaceutical agents while male poisoning admissions were more often due to chemical substances. In addition, according to Rajasuriar *et al.,* (2007), the prevalence of poisoning and death was highest among Indians compared to all other races in Malaysia.

Overall, most poisoning admissions were due to

*Corresponding author's e-mail: szulaikha.szmj@usm.my

pharmaceutical agents. The largest number of fatalities were due to pesticide. Pesticide was also the most common substance reported in cases of intentional self-harm. Most cases of poisoning admissions occurred due to accidental exposure (47.0%), followed by intentional self-harm (20.7%) (Rajasuriar *et al.*, 2007).

Recent studies of log-linear models in health analysis include Bucchi *et al.,* (2019), Voumard *et al.,* (2019), Gautam *et al.,* (2019), Kamranrad *et al.,* (2019), Lin et al. (2019) and Wiseman *et al.,* (2019). The last two have some relevance to poisoning: Mercury for Wiseman *et al.,* (2019) and Asbestos for Lin *et al.*, (2019).

The objectives of our present study are to build a suitable log-linear model to analyze the data on poisoned patients among Malaysians and to analyze the relationship between the variables of the log-linear model.

II.   MATERIALS AND METHODS

The data employed in this study was obtained from the National Poison Centre from 2012 to 2014. The data is a secondary type data which was obtained from questionnaire surveys on poisoning by the National Poison Centre's pharmacist. There were 9949 respondents with nine variables representing the background of each respondent. The nine variables were circumstances of exposure, types of exposure, gender, race, age, location of exposure, state, route of exposure and types of poison. There are two models analyzed this study: model for demographic data of patients and model for poisoning information. The variables involved in the first model are gender, age, race and state. Variables for the second model are circumstance of exposure, type of exposure, location of exposure, route of exposure and types of poison. Both log-linear models were developed to investigate the association between variables in the model.

Consider a three-dimensional contingency table with X, Y and Z as its variables and let μ_{ijk} denote expected frequencies. Suppose all expected frequencies are positive i.e. $\mu_{ijk} > 0$, and let $n_{ijk} = \log(\mu_{ijk})$. A dot in a subscript signifies the average with respect to that index; for example,

$$n_{.jk} = \frac{\sum_i n_{ijk}}{I}$$, where I is number of cells. Then the general log-linear model for a three-way contingency table is

$$\ln \mu_{ijk} = \mu + \lambda_{X(i)} + \lambda_{Y(j)} + \lambda_{Z(k)} + \lambda_{XY(ij)} + \lambda_{XZ(ik)} + \lambda_{YZ(jk)} + \lambda_{XYZ(ijk)}$$

Here the singly subscript terms is for main effects, the doubly-subscript terms are for to two-factor interactions, and the triply-subscript terms is for three-factor interactions (Christensen, 1990).

Background studies on log-linear models can be found in Adarabioyo (2014), Ghosh and Bhattacherjee (2009), Knoke and Burke (1980), Christensen (1990), Muller and Mayhall (1971), Kuo *et al.,* (2012) and Panagiotako and Pistavos (2004). Some studies of poisoning cases in Malaysia can be found in Adnan *et al.,* (2013) and Rajasuriar *et al.,* (2007).

III.   RESULTS AND DISCUSSIONS

The background of the respondents in this study are divided into circumstance of exposure, types of exposure, gender, age, race, location of exposure, state, route of exposure and type of poison experienced by the patients (Table 1).

Table 1. Characteristics of respondents

Variable	Category	Frequency (%)
Circumstance of exposure	Intentional	5487 (55.2%)
	Unintentional	4395 (44.2%)
	Others	67 (0.7%)
Type of exposure	Acute	9389 (94.4%)
	Acute on chronic	437 (4.4%)
	Others	123 (1.2%)
Gender	Male	5325 (53.5%)
	Female	4624 (46.5%)
Age	Young (<10 years old)	2222 (22.3%)
	Adult (10-50 years old)	6578 (66.1%)
	Old: (>50 years old)	1149 (11.5%)
Race	Malay	3789 (38.1%)
	Chinese	1777 (17.9%)
	Indian	3328 (33.5%)
	Others	1055 (10.6%)
Location of exposure	Home	9302 (93.5%)
	Workplace	304 (3.1%)
	Others	343 (3.4%)
State	West Coast (Peninsular Malaysia)	7912 (79.5%)
	East Coast	1112 (11.2%)

	(Peninsular Malaysia) East Malaysia	925 (9.3%)
Route of exposure	Ingestion Inhalation Others	9366 (94.1%) 308 (3.1%) 275 (2.8%)
Types of Poison	Pesticide Pharmaceutical Others	2863 (28.8%) 3546 (35.6%) 3540 (35.6%)

Based on Table 1, majority of the respondents are intentionally exposed to poisons, with 55.2%. Most of the respondents experienced acute type of exposure which are 94.4%.

Most of the respondents are male. They comprised 53.5% of the total respondents, while female respondents are only 46.5%. Most of the respondents admitted with poisoning case are adults of age 10 to 50 (66.1%). Malay respondents are the majority category for race which are 38.1%.

For the location of poisons exposure, home is the most common place for the respondents who experienced poisoning exposure (93.5%). Most of the respondents (79.5%) are from West Coast of Peninsular.

Most of the respondents (94.1%) were exposed to poisons are through ingestion. As for types of poison, most of the respondents experienced pharmaceutical type of poison that is 35.6%.

There are four variables in conducting log-linear analysis for demographic data which are gender, age, race and state. In order to obtain the best-fitted as well as parsimony log-linear model, two methods of model selection will be used; single step method and backward elimination. A single step method is used to get an idea for determination an initial model. In determining the appropriate model, a full saturated model is used as the first model to start with. The backward elimination method is conducted in order to determine the unsaturated model that would provide the best fit to the data.

Considered a saturated model by letting μ_{hijk} be a measurement in the cell (h, i, j, k) consistent with h^{th}-gender, i^{th}–age, j^{th}– race, k^{th}– state, where $h = 0,1$; $i = 0,1,2$; $j = 0,1,2,3$ and $k = 0,1,2$. For the purpose of estimating and hypothesis testing, a cell count, μ_{hijk} is assumed to follow a multinomial distribution. Thus, the saturated model can be formulated as follows:

$$\ln \mu_{hijk} = \mu + \lambda_{C(h)} + \lambda_{D(i)} + \lambda_{E(j)} + \lambda_{G(k)}$$
$$+ \lambda_{CD(hi)} + \lambda_{CE(hj)} + \lambda_{CG(hk)} + \lambda_{DE(ij)}$$
$$+ \lambda_{DG(ik)} + \lambda_{EG(jk)} + \lambda_{CDE(hij)} + \lambda_{CDH(hik)}$$
$$+ \lambda_{CEG(hjk)} + \lambda_{DEG(ijk)} + \lambda_{CDEG(hijk)}$$

where C: Gender (0=male, 1=female); D: Age (0=young, 1=adult, 2=old); E: Race (0=Malay, 1=Chinese, 2=Indian, 3=Others) and G: State (0=West Coast, 1=East Coast, 2=East Malaysia). For the saturated model, both goodness-of-fit test G^2 and χ^2 are zero as well as the p-value.

The test for K-way and higher order effects are zero helps to determine and initial log-linear model. The hypothesis for this test is K-way and higher order effects are not equal to zero. From Table 2, it can be seen that $K = 3$ is statically significant at level of 0.05 significance. The maximum interaction for the model can be determined by the formula $(2^K - 1)$. Based on the formula, the maximum interaction for the model is $(2^3 - 1) = 7$. The significantly p-value showed by 3-way $(K = 3)$ effect indicates that 3-way and higher order effects are not equal to zero. Thus, the 3-way interactions or 4-way interactions are probable to be included in the model.

Table 2. Test that K-way and higher order effects are zero

K	**df**	G^2	**P**	χ^2	**P**	i
4	12	7.221	0.843	7.209	0.844	4
3	40	72.097	0.001	73.677	0.001	5
2	63	2924.480	0.000	3567.153	0.000	2
1	71	18858.977	0.000	31938.203	0.000	0

For the test of K-way effects are zero, the investigation is start from the bottom of the table as in Table 3. Obviously, the fourth order interaction term is not statistically significant even at 5% level for both G and χ^2. Thus, fourth order interaction terms would not be included in the model. In contrary, the significantly p-value for 2-way effects $(K = 2)$ and 3-way effects $(K = 3)$ indicates that both two factor and three factor interactions would be included in the model respectively.

Table 3. Test that K-way effects are zero

K	df	G	P	χ^2	P	i
1	8	15934.497	0.000	28371.051	0.000	0
2	23	2852.384	0.000	3493.475	0.000	0
3	28	64.875	0.000	66.468	0.000	0
4	12	7.221	0.843	7.209	0.844	0

For the case of $K = 1$, the main effect terms for both the G^2 and χ^2 are statistically significant at 0.05 level of significant. This means that the main effect terms are included in the model. The model is hierarchical. So, because of the two interactions and three interactions terms are included in the model, therefore the main effect terms should be incorporated in the model too. In determining the good fitted model, it can be concluded that the interaction in the model is up to three interaction terms.

From the result from backward elimination method in Table 4, the final model is the model with three interaction terms. In order to test the hypothesis as the model does not accurately fit the observed data, the goodness-of-fit that is G^2 statistics is used.

Table 4. Summary result of backward elimination method

Step	Model	Omitting interaction	df	G^2	p
1	[CDE] [CDG] [CEG] [DEG]		12	7.221	0.843
2	[CDE] [CEG] [DEG]	[CDG]	16	8.861	0.919

As can be seen in Table 4, the likelihood-ration, G^2 statistics for the final model is not statistically significant at 0.05 level of significant. Therefore, the model does not accurately fit the observed data. For this reason, it can be concluded that the best and parsimony unsaturated model is the model with three interaction terms of [CDE] [CEG] [DEG]. The best-fitted log-linear model can be formulated as follows:

$$\ln \mu_{hijk} = \mu + \lambda_{C(h)} + \lambda_{D(i)} + \lambda_{E(j)} + \lambda_{G(k)}$$
$$+ \lambda_{CD(hi)} + \lambda_{CE(hj)} + \lambda_{CG(hk)} + \lambda_{DE(ij)}$$
$$+ \lambda_{DG(ik)} + \lambda_{EG(jk)} + \lambda_{CDE(hij)}$$
$$+ \lambda_{CEG(hjk)} + \lambda_{DEG(ijk)}$$

There are five variables that will be used in conducting the log-linear analysis for poisoning information which circumstance of exposure, types of exposure, location of exposure, route of exposure and types of poison.

Consider a saturated model by letting μ_{pqrst} be a measurement in the cell (p, q, r, s, t) consistent with p^{th}-circumstance of exposure, q^{th}-types of exposure, r^{th}-location of exposure, s^{th}-route of exposure, t^{th}-types of poison, where $p = 0,1,2$; $q = 0,1,2$; $r = 0,1,2$; s$= 0,1,2$ and $t = 0,1,2$. For the purpose of estimating and hypothesis testing, a cell count, μ_{pqrst} is assumed to follow a multinomial distribution. Thus, the saturated model can be formulated as follows:

$$\ln \mu_{hijk} = \mu + \lambda_{A(p)} + \lambda_{B(q)} + \lambda_{F(r)} + \lambda_{H(s)} + \lambda_{J(t)}$$
$$+ \lambda_{AB(pq)} + \lambda_{AF(pr)} + \lambda_{AH(ps)} + \lambda_{AJ(pt)} + \lambda_{BF(qr)}$$
$$+ \lambda_{BH(qs)} + \lambda_{BJ(qt)} + \lambda_{FH(rs)} + \lambda_{FJ(rt)} + \lambda_{HJ(st)}$$
$$+ \lambda_{ABF(pqr)} + \lambda_{ABH(pqs)} + \lambda_{ABJ(pqt)} + \lambda_{AFH(prs)}$$
$$+ \lambda_{AFJ(prt)} + \lambda_{AHJ(pst)} + \lambda_{BFH(qrs)} + \lambda_{BFJ(qrt)}$$
$$+ \lambda_{BHJ(qst)} + \lambda_{FHJ(rst)} + \lambda_{ABFH(pqrs)} + \lambda_{ABFJ(pqrt)}$$
$$+ \lambda_{ABHJ(pqst)} + \lambda_{AFHJ(prst)} + \lambda_{BFHJ(qrst)}$$
$$+ \lambda_{ABFHJ(pqrst)}$$

The two goodness-of-fit tests, G^2 and χ^2 show that the saturated model fit the data well since the value for both test are equal to zero.

In order to get an idea in determining an initial log-linear, the hypothesis of K-way and higher order effects are zero is tested. From the output displayed in Table 5, notice that $K = 3$ is statistically significant at 0.05 level of significance. Therefore, the potential maximum interaction for the model is seven, which is obtained from the formula of $(2^K - 1) = (2^3 - 1) = 7$. By hypothesis testing of this test, 3-way and higher order effects are not equal to zero. This test is useful in finding out the possible number of interactions in the model.

Table 5. Test that K -way and higher order effects are zero

K	df	G^2	p	χ^2	p	i
5	32	1.721	1.000	0.954	1.000	2
4	112	15.849	1.000	11.684	1.000	5
3	192	335.939	0.000	442.209	0.000	6
2	232	3899.463	0.000	17895.681	0.000	2
1	242	61226.005	0.000	300325.314	0.000	0

For the test of K -way effects are zero, the hypothesis test could be as K -way effects are not equal to zero. Based on the result in Table 6, the fourth $(K = 4)$ and the fifth $(K = 5)$ order interaction terms are not statistically significant at 0.05 level of significance. Thus, both four and five-way interaction terms would not be included in the model. For the two tests of G^2 and χ^2, both statistics show that the main effects, 2 -way and 3 -way effects are significant at 5% level. Therefore, the model would have a maximum of three-way interaction effects only.

Table 6. Test that K -way effects are zero

K	df	G^2	p	χ^2	p	i
1	10	57326.542	0.000	282429.634	0.000	0
2	40	3563.524	0.000	17453.472	0.000	0
3	80	320.090	0.000	430.524	0.000	0
4	80	14.128	1.000	10.731	1.000	0
5	32	1.721	1.000	0.954	1.000	0

According to the backward elimination method, the result of K -way effects are zero and it can be concluding the highest interaction terms that would be in the model is three interactions. Thus, this method is carried to determine the unsaturated model that fit the data best.

Table 7. Summary result of backward elimination method

Step	Model	Omitting interaction	df	G^2	p
1	[ABFH] [ABFJ] [ABHJ] [AFHJ] [BFHJ]		32	1.721	1.000
2	[ABFH] [ABFJ] [AFHJ] [BFHJ]	[ABHJ]	48	3.233	1.000
3	[ABFH] [ABFJ] [AFHJ] [BHJ]	[BFHJ]	64	5.958	1.000
4	[ABFH] [ABFJ] [BHJ] [AHJ] [FHJ]	[AFHJ]	80	9.035	1.000
5	[ABFJ] [BHJ] [AHJ] [FHJ] [ABH] [AFH] [BFH]	[ABFH]	96	10.125	1.000
6	[BHJ] [AHJ] [FHJ] [ABH] [AFH] [BFH] [ABF] [ABJ] [AFJ] [BFJ]	[ABFJ]	112	16.608	1.000
7	[BHJ] [AHJ] [FHJ] [ABH] [BFH] [ABF] [ABJ] [AFJ] [BFJ]	[AFH]	120	25.063	1.000
8	[AHJ] [FHJ] [ABH] [BFH] [ABF] [ABJ] [AFJ] [BFJ]	[BHJ]	128	35.795	1.000
9	[AHJ] [FHJ] [ABH] [BFH] [ABF] [ABJ] [BFJ]	[AFJ]	136	46.447	1.000
10	[AHJ] [FHJ] [ABH] [BFH] [ABJ] [BFJ] [AF]	[ABF]	144	57.940	1.000

Table 7 shows the summary result of backward elimination procedure for poisoning information. Obviously, the final model in step 10 is the best-fitted model with three interaction terms. The goodness-of-fit test, G^2 statistics is used to test whether the best model is fitted to the observed data. The hypothesis would be the model does not accurately fit the observed data. It is clearly seen that the G^2 is not statistically significant at 5% level of significance for the final model in step 10. This indicates that the model accurately fits the observed data. Thus, the best model obtained is the model with three interaction terms of [AHJ] [FHJ] [ABH] [BFH] [ABJ] [BFJ] [AF]. The best fitted log-linear model can be expressed as:

$$
\begin{aligned}
\ln \mu_{hijk} = {}& \mu + \lambda_{A(p)} + \lambda_{B(q)} + \lambda_{F(r)} + \lambda_{H(s)} + \lambda_{J(t)} \\
& + \lambda_{AB(pq)} + \lambda_{AF(pr)} + \lambda_{AH(ps)} + \lambda_{AJ(pt)} + \lambda_{BF(qr)} \\
& + \lambda_{BH(qs)} + \lambda_{BJ(qt)} + \lambda_{FH(rs)} + \lambda_{FJ(rt)} + \lambda_{HJ(st)} \\
& + \lambda_{ABH(pqs)} + \lambda_{ABJ(pqt)} + \lambda_{AHJ(pst)} + \lambda_{BFH(qrs)} \\
& + \lambda_{BFJ(qrt)} + \lambda_{FHJ(rst)}
\end{aligned}
$$

V. SUMMARY

This study has focused on the analysis of data of poisoned patients from 2012 to 2014 in Malaysia using two log-linear models. The best log-linear model of demographic data indicated that there is an association between gender, age and race, race, gender and state as well as age, race and state. The best log-linear model for poisoning information indicated that there is relationship between circumstance of exposure, route of exposure and type of poison, location of exposure, route of exposure and type of poison, circumstance of exposure, type of exposure and route of exposure, circumstance of exposure, location of exposure and route of exposure, circumstance of exposure, type of exposure and type of poison and also type of exposure, location of exposure and type of poison.

VI. ACKNOWLEDGEMENT

IV. REFERENCES

Adarabioyo, M. I. 2014, Application of log-linear model to determinants of child mortality in Nigeria. International Journal of Scientific Research and Innovative Technology. 1(1).

Adnan, L. H. M., Kamaldin, J., Mohamad, N., Salatore, S. A. & Suhaimi, R. 2013, The risk of accidental chemical poisoning cases among children (≤12 years old) admitted to Hospital Universiti Sains Malaysia: 5-years review. Journal of Clinical Toxicology. 2(5).

Bucchi, L., Baldacchini, F., Mancini, S., Ravaioli, A., Giuliani, O., Vattiato, R., Falcini, F., Giorgi Rossi, P., Campari, C., Canuti, D., Di Felice, E., Sassoli de Bianchi, P., Ferretti, S., Bertozzi, N. & Biggeri, A. 2019, Estimating the impact of an organized screening programme on cervical cancer incidence: A year 26-year study from northern Italy. International Journal of Cancer. 144(5), 1017-1026.

Christensen, R. 1990, Log-linear Model. Springer-Verlag, New York.

Gautam, N., Kakchapati, S., Shrestha, S. & Wanishsakpong, W. 2019, Patterns and trends of malaria in 25 risk districts of Nepal from 2001 to 2017. Clinical and Experimental Vaccine Research. 8(1), 77-85.

Ghosh, A. K. & Bhattacherjee, A. 2009, Loglinear model for assessment of risk factors of occupational injuries in underground coal mines. Journal of Geology and Mining Research. 1(2), 25-33.

Kamranrad, R., Amiri, A. & Niaki, S. T. A. 2019, Phase-I monitoring of log-linear model-based processes (a case study in health care: Kidney patients). Quality and Reliability Engineering International.

Knoke, D. & Burke, P. J. 1980, Log-linear Models. Quantitative Applications in Social Science, Saga Univ.

Kuo, L. H., Wei, H. W., Lin, L., Huang, H. C. & Yang, H. J. 2012, Use of general log-linear analysis procedure by fitting a log-linear model with age group and gender as factors for preschool aboriginal teachers in Taiwan. EDUCIT. 22, 127-132.

Lin, R. T., Chang, Y. Y., Wang, J. D. & Lee, L. J. H. 2019, Upcoming epidemic of asbestos-related malignant pleural mesothelioma in Taiwan: A prediction of incidence in the

next 30 years. Journal of the Formosan Medical Association. 118(1P3), 463-470.

Muller, T. P. & Mayhall, J. T. 1971, Analysis of contingency table data on torus mandibularis. American Journal of Physical Anthropology. 34, 149-154.

Panagiotakos, D. B. & Pistavos, C. 2004, Interpretation of epidemiological data using multiple correspondence analysis and log-linear analysis. Journal of Data Science. 2, 75-86.

Rajasuriar, R., Awang, R., Hashim, S. B. H. & Rahmat, H. R. B. H. 2007, Profile of poisoning admissions in Malaysia. Human & Experimental Toxicology. 26, 73-81.

Voumard, M., Giannakis, S., Carratala, A. & Pulgarin, C. 2019, E. coli-MS2 bacteriophage interactions during solar disinfection of wastewater and the subsequent post-irradiation period. Chemical Engineering Journal. 359, 1224-1233.

Wiseman, C. L. S., Parnia, A., Chakravartty, D., Archbold, J., Copes, R. & Cole, D. 2019, Total, methyl and inorganic mercury concentrations in blood and environmental exposure sources in newcomer women in Toronto, Canada. Environmental Research. 261-271.