CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES
https://cajmtcs.casjournal.org/index.php/CAJMTCS
Volume: 07 Issue: 04 | October 2026 ISSN: 2660-5309

Article

Comparative Analysis of Contingency Models in Medical Data using Log-linear Models and Logistic Regression Models

Laith Adnan Hammad[*1]

1. Open Educational College, Branch Samarra, Salah aL-Din Iraq
* Correspondence: Laithadnanhammad@gmail.com

Citation: Hammad L. A. Comparative Analysis of Contingency Models in Medical Data using Log-linear Models and Logistic Regression Models. Central Asian Journal of Mathematical Theory and Computer Sciences 2026, 7(4), 32-39.

Received: 22nd May 2026
Revised: 16th Jun 2026
Accepted: 11th Jul 2026
Published: 07th Aug 2026

Abstract: This research explored the use of linear logistic studies and binary and multi-logarithmic reference models in analyzing n×m Tables Contingency for the development of relationships between genders. It also examined how each model interprets the relationships between these correlations. This research has applications in the medical field, where the effects of age and gender on weight were studied, and their impact on weight prediction was investigated. Logarithmic models identified weight gain in any age group, while logistic models developed a predictive equation for the effect of age and gender on weight. Each model has its own characteristics logarithmic models focus on studying the relationship between variables, while logistic models construct a predictive equation to interpret the results.

Keywords: Tables Contingency, Log Model, Logistic Model, Weight.

1. Introduction

The study of data categorised into two or more variables in contingency tables is of great importance in data analysis and in obtaining results that aid in understanding the relationship between different variables and taking appropriate action. With the development of statistics, numerous models have emerged that examine the relationship between variables; among the most commonly used are linear logarithmic models and logistic models (binary and multinomial) [1], [2], [3], [4], [5]. Each has its own distinctive characteristics: linear logarithmic models examine the relationship between continuous quantitative variables and transform non-linear relationships into linear ones, whilst logistic models are used to analyse binary variables and predict relationships; both models have wide-ranging applications in various medical, biological, and economic fields [6], [7].

2. Materials and Methods

The materials used in this research comprise medical data on the weight of children aged 6 to 16 years, specialist scientific references in statistics, and the statistical software package R. The methods employed comprised a theoretical study of logarithmic and logistic models, the conversion of the data into contingency tables, the application of the linear logarithmic model and the logistic model to the data [8], [9], [10], [11], [12], the evaluation of performance using statistical indicators, and finally a comparison of the results:

1. Contingency tables [6], [7]

These are tables consisting of several cells, each of which represents the frequency of a particular variable. They consist of two or more variables, each of which has levels, and are classified as follows:

1. Two-way tables: These consist of only two variables, each of which has different levels.
2. Multidimensional tables: These consist of more than two variables, such as three-way tables.

Contingency tables can be classified into two categories based on the presence or absence of zero frequencies, namely:

1. Incomplete tables: These are tables that contain zero frequencies and zero probabilities
2. Complete tables: These are tables that do not contain zero frequencies in any of their cells. These include two-way contingency tables

Two-Dimensional Contingency Tables

These are tables consisting of only two variables; the first variable contains n rows, with each row representing a specific level, and the second variable contains m columns. The general structure of this type of table, in terms of frequencies and proportions in the population, is illustrated in Table (1)

Table 1. A two-way compatibility.

A \ B	**1**	**2**	**…**	**j**	**…**	**m**	**Total**
1	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1m}	$x_{1.}$
	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	p_{1m}	$p_{1.}$
2	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2m}	$x_{2.}$
	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	p_{2m}	$p_{2.}$
.	.	.		.		.	.
.	.	.		.		.	.
i	x_{i1}	x_{i2}	$\cdots$	x_{ij}	$\cdots$	x_{im}	$x_{i.}$
	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	p_{im}	$p_{i.}$
.	.	.		.		.	.
.	.	.		.		.	.
n	x_{n1}	x_{n2}	$\cdots$	x_{nj}	$\cdots$	x_{nm}	$x_{n.}$
	p_{n1}	p_{n2}	$\cdots$	p_{nnj}	$\cdots$	p_{nm}	$p_{n.}$
Total	$x_{.1}$	$x_{.2}$	$\cdots$	$x_{.j}$	$\cdots$	$x_{.m}$	$x_{..}$
	$p_{.1}$	$p_{.2}$	$\cdots$	$p_{.j}$	$\cdots$	$p_{.m}$	P

x_{ij}: the number of observations in the population at level i of variable A and level j of variable B.

p_{ij}: the probability of an observation at level i of variable A and level j of variable B.

x_i : the total number of observations at level i of variable A.

$$x_{i.} = \sum_{j=1}^{m} x_{ij} \qquad i = 1,2,\dots n$$

$$x_{.j} = \sum_{i=1}^{m} x_{ij} \qquad j = 1,2,\dots m$$

xj: the sum of the observations at level j of variable B.

$$p_i = pr(A = i) \quad , p_{i.} = \sum_{j=1}^{m} p_{ij} \qquad i = 1,2, \dots n$$

$$p_j = pr(B = j) \quad p_{.j} = \sum_{i=1}^{m} p_{ij} \qquad j = 1,2, \dots m$$

$$K = X_{..} = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \qquad i = 1,2, \dots n \quad , \quad j = 1,2, \dots m$$

K=H : Total population size

When a random sample of size k is drawn from Table (1), n×m contingency tables are formed, as shown in Table**(2)**

Table 2. N×M contingency tables for a sample of size k drawn from table 1.

A \ B	1	2	…	j	…	m	Total
1	x_{11}	x_{12}	…	x_{1j}	…	x_{1m}	$x_{1.}$
2	x_{21}	x_{22}	…	x_{2j}	…	x_{2m}	$x_{2.}$
.	.	.		.		.	.
.	.	.		.		.	.
.	.	.		.		.	.
I	x_{i1}	x_{i2}	…	x_{ij}	…	x_{im}	$x_{i.}$
.	.	.		.		.	.
.	.	.		.		.	.
n	x_{n1}	x_{n2}	…	x_{nj}	…	x_{nm}	$x_{n.}$
Total	$x_{.1}$	$x_{.2}$	…	$x_{.j}$	…	$x_{.m}$	$k = x_{..}$

Where k, x_j, x_i, and x_{ij} were previously defined in Table (1), but this definition applies to a sample of size k from the population K. The definition of the probabilities given in Table (1) remains valid.

2. Sampling Distributions [7]

To understand the relationship between variables and to derive estimates for a specific model, it is necessary to understand sampling distributions. One such distribution is the independent Poisson distribution, which is used when there are no restrictions on sample size and each cell follows an independent Poisson distribution. Its probability density function (p.d.f.) is**:**

$$f(x_{ij}) = \prod_{i,j} \frac{D_{ij}^{x_{ij}} e^{-m_{ij}}}{x_{ij}!} \quad , \quad x_{ij} = 0,1,2, \dots \qquad \dots\dots(1)$$

Where: x_{ij} is the observed frequency of cell (i, j) and: D_{ij} is the expected frequency of cell (i, j)**.**

$$\text{..................... (2)} \qquad D_{ij} = \frac{x_{i.} \ x_{.j}}{k}$$

Where :

3. The linear logarithmic model in n×m contingency tables [6], [7]

There are two types of this model:

3.1. The full model

This is a full model that includes all the main effects of the variables included in the model and all possible interactions between the variables at different levels. If we have a sample of size k arranged in an n×m contingency table, the model formula is:

$$\mathrm{LogD_{ij}} = \mathrm{v} + \mathrm{v_{1i}} + \mathrm{v_{2j}} + \mathrm{v_{12ij}} \quad \dots\dots\dots\dots\dots\dots\dots (3)$$

And this model achieves:

$$\sum_{i} \mathrm{v_{1i}} = 0 \quad , \quad \sum_{j} \mathrm{v_{2j}} = 0$$

$$\sum_{i} \mathrm{v_{12ij}} = 0 \quad , \quad \sum_{j} \mathrm{v_{12ij}} = 0$$

And the teachers' assessment of him:

$$\hat{\mathrm{v}} = \frac{1}{\mathrm{nm}} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{LogD_{ij}} \quad \dots\dots\dots\dots\dots\dots. (4)$$

$$\widehat{\mathrm{v_{1i}}} = \frac{1}{\mathrm{m}} \sum_{j=1}^{m} \mathrm{LogD_{ij}} - \hat{\mathrm{v}} \quad \dots\dots\dots\dots\dots\dots.. (5)$$

$$\widehat{\mathrm{v_{2j}}} = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} \mathrm{LogD_{ij}} - \hat{\mathrm{v}} \quad \dots\dots\dots\dots\dots\dots.. (6)$$

$$\widehat{\mathrm{v_{12ij}}} = \mathrm{LogD_{ij}} - \frac{1}{\mathrm{n}} \sum_{i=1}^{n} \mathrm{LogD_{ij}} - \frac{1}{\mathrm{m}} \sum_{j=1}^{m} \mathrm{LogD_{ij}} + \hat{\mathrm{v}} \quad \dots\dots\dots\dots (7)$$

3.2. The Independence Model

This is an unsaturated model, meaning that there is no interaction between the variables and they are independent of one another. It is defined by the formula:

$$\mathrm{LogD_{ij}} = \mathrm{v} + \mathrm{v_{1i}} + \mathrm{v_{2j}} \quad \dots\dots\dots\dots\dots\dots\dots (8)$$

The bivariate interaction between the two variables is $\mathrm{v_{12}} = 0$ It satisfies the conditions of the accreditation model itself, and other unsaturated models can be formed according to the formula: when $\mathrm{v_1} = 0$ where $\mathrm{v_{12}} = 0$ where :

$$\mathrm{LogD_{ij}} = \mathrm{v} + \mathrm{v_{2j}} \quad \dots\dots\dots\dots\dots\dots\dots (9)$$

And when it is $\mathrm{v_2} = 0$ where $\mathrm{v_{12}} = 0$ where

$$\mathrm{LogD_{ij}} = \mathrm{v} + \mathrm{v_{1i}} \quad \dots\dots\dots\dots\dots\dots\dots (10)$$

Table 3. Shows the degrees of freedom in a linear logarithmic model for an n×m contingency table.

Borders v	Degrees of freedom
v	**1**
v_1	**n-1**
v_2	**m-1**
v_12	**(n-1) (m-1)**
Total	**Nm**

4. Goodness-of-fit tests [13]

Once a linear logarithmic model has been formulated in terms of dependence or independence, the next stage is to test this model for accuracy using goodness-of-fit tests:

4.1. Chi-squared statistic (χ^2) In n×m contingency tables, it is defined by the formula:

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(\mathrm{x_{ij}} - \mathrm{D_{ij}})^2}{\mathrm{D_{ij}}} \quad \dots\dots\dots\dots\dots\dots.. (11)$$

Where:

x_{ij} : is the observed frequency.

D_{ij} : is the expected frequency. The degrees of freedom of the statistic are (n-1)(m-1).

4.2. Likelihood Ratio Statistic: G^2 In n × m contingency tables, it is defined by the formula:

$$G^2 = 2\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \text{Log} \frac{x_{ij}}{D_{ij}} \quad \ldots\ldots\ldots\ldots\ldots\ldots..(12)$$

and approaches χ^2 with (n-1)(m-1) degrees of freedom.

5. The logistic regression model [13], [14], [15]

This is a type of non-linear regression model used to study the relationship between an independent variable and one or more dependent variables. It is considered an important statistical method as it best represents the relationship and is characterised by a high degree of flexibility. There are several types of this model, but the most common is:

5.1. The Binary Logistic Regression Model

This is one type of logistic regression model and is used when the dependent variable has only two values (success, failure), (smoker, non-smoker), (infected, not infected). That is, when the response occurs, y = 1 with probability p, and when it does not occur, y = 0 with probability 1 – p. It is easy to interpret and well-suited to the data as it provides reliable probabilities for decision-making and follows a Bernoulli distribution.

$$E(y_i) = p \quad ....(13)$$

The logistic function is:

$$p_i = \frac{e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}}{1+e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}} \quad ....(14)$$

$$p(y=1) = \frac{1}{1+e^{-(\beta_0+\sum_{i=1}^{n}\beta_i x_i)}} \quad ....(15)$$

$$p(y=0) = 1 - \frac{e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}}{1+e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}} = \frac{1}{1+e^{\beta_0+\sum_{i=1}^{n}\beta_i x_i}} \quad ......(16)$$

Where β_0, β_i are the parameters to be estimated; to convert the logistic regression model to a linear one, we use the logit function, where $-\infty \leq p \leq \infty$:

$$\text{logit} = \text{Ln}\left(\frac{p_i}{1-p_i}\right) \quad .....(17)$$

Where: p_i is the probability of the event occurring

$1-p_i$ is the probability of the event not occurring

$$\text{logit} = \log_e = \text{Ln Weighting ratio and } \frac{p_i}{1-p_i}$$

5.2. The multinomial logistic regression model

Is similar to the binary model and requires the dependent variable to be categorical; however, it takes into account several mutually exclusive categories, such as levels of education (primary, lower secondary, upper secondary, …) or social statuses (married, widowed, divorced, single); or, in medical trials, where a drug has more than one effect, these are often measured on a scale and categorised as (no effect, low effect, moderate effect, high effect). Consequently, the multinomial model takes more than two values and is regarded as an attempt to develop the binary model. When the behaviour is binary, the linear regression model has

$$y_i = \beta_0 + \beta_i x_i + u_i$$

A probability model between (0, 1) and the regression equation:

$$E(y_i|x_i) = \beta_0 + \beta_i x_i \quad ..... (18)$$

Where $0 \leq E(y\,|x\,) \leq 1$ compare between the tow equation we have equivalent

$$E(y_i|x_i) = \beta_0 + \beta_i x_i = p_i$$

3. Results and Discussion

Practical Aspects

To apply the linear log-normal model and the binary logistic model, data were used from a sample of 121 children of normal weight or overweight, and to examine the effect of age and sex on weight; the sample is shown in Table **(4)**

Table 4. Classification of the sample by gender, weight and age groups.

Age groups \ Gender and weight	Natural man	A fat man	Nature female	An overweight woman	Total
7-8	2	2	10	3	17
9-10	3	2	13	5	23
11-12	5	1	23	9	38
13-14	2	1	11	7	21
15-16	6	5	8	3	22
Total	18	11	65	27	121

Application 1 : The Linear Log-Normal Model

After collecting, classifying and organising the data into contingency tables.

We **first** calculate the expected frequencies for each cell in Table (4) using Formula (2), and the results are shown in Table (5)

Table 5. Expected frequencies.

Age groups \ Gender and weight	Natural man	A fat man	Nature female	An overweight woman	Total
7-8	2.52	1.54	9.14	3.80	17
9-10	3.42	2.09	12.35	5.14	23
11-12	5.66	3.46	20.41	8.47	38
13-14	3.12	1.91	11.28	4.69	21
15-16	3.28	2.0	11.82	4.90	22
Total	18	11	65	27	121

Secondly: We calculate the χ^2 statistic using the formula (11) The calculated χ^2 value was 13.37. To compare this with the table value of χ^2 for 12 degrees of freedom and a significance level of 0.05, we find that the table value is $\chi^2 = 21.03$. We conclude that the calculated value is smaller than the table value, which means that the variables are independent.

Thirdly: We calculate the effects of the factors using equations (4), (5) and (6), and the results are shown in Tables (6) and (7).

Table 6. Shows the effect of age on weight.

Age groups	7-8	9-10	11-12	13-14	15-16
The Factor	0.168	0.004	0.03	1.128	8.67

Table 7. Illustrates the effect of gender on weight.

Age groups	man	woman
7-8	0.240	0.248
9-10	0.056	0.038
11-12	1.815	0.363
13-14	0.831	1.135
15-16	6.71	1.96

We can see from Table (6) that the effect becomes more pronounced as age increases. In the first, second and third age groups, the effect is slight; in the 13–14 age group, the effect was moderate; and in the 15–16 age group, the effect was very significant, meaning that weight increases in this age group. As for Table (7), the effect of gender was similar in the 7–8 age group; negligible in the 9–10 age group; males were more prone to obesity in the 11–12 age group; females were more prone to obesity in the 13–14 age group; whilst in the 15–16 age group, the effect on males was very significant.

Application 2: The Binary Logistic Model

We know the variables (1: obese, 0: normal) and (1: male, 0: female), and we use equations (14), (15), (16) and (17) to find the logarithmic equation.

$$\log(\mathrm{p}) = \ln\left(\frac{\mathrm{p}}{1-\mathrm{p}}\right) = \beta_0 + \beta_1(\text{age }) + \beta_2(\text{Gender})$$

Where p = 1 represents obesity, β_0 is a constant, β_1 represents the effect of age on obesity, and β_2 represents the effect of gender on obesity; the results were as follows**:**

$$\beta_0 = -1.373 \text{ و } \beta_1 = 0.043 \quad \text{ و } \beta_2 = 0.347 \text{ and the equation}$$

$$\log(\mathrm{p}) = -1.373 + 0.043(\text{age }) + 0.347(\text{Gender})$$

$\beta_1 > 0$: each additional year of age increases the likelihood of obesity; $\beta_1 < 0$: each additional year of age reduces the likelihood of obesity; $\beta_2 > 0$: males have a higher likelihood of obesity than females; $\beta_2 < 0$: females have a higher likelihood of obesity than males

$O_{(R\text{-}age)} \cong 1.044$ means that the probability of obesity increases by 4.4 per cent each year, and $O_{(R\text{-}gender)} \cong 1.415$ means that males have a 41.5% higher probability of obesity than females**.**

4. Conclusion and Recommendations

Age and gender have a significant impact on weight. Log-linear models indicate at which stage obesity increases and in which gender, whilst logistic models provide a predictive equation for the relationship between age, gender and weight, showing that the probability of obesity in males increases from 32.6 per cent at the age of seven to 41.6 per cent at the age of 16, whilst for females it rises from 25.5 per cent to 33.5 per cent . The researcher recommends that, when analysing data in contingency tables, if the aim is to determine the relationship between variables, linear logarithmic models should be used; however, if the aim is to interpret the probability of an event occurring or to predict its occurrence, logistic models should be used. The researcher also recommends further applications in the medical, economic and agricultural fields.

REFERENCES

[1] L. A. Goodman, "The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications," *J. Am. Stat. Assoc.*, vol. 65, no. 329, pp. 226–256, 1970, doi: 10.1080/01621459.1970.10481076.

[2] J. A. Nelder, “Log Linear Models for Contingency Tables: A Generalization of Classical Least Squares,” *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 23, no. 3, pp. 323–329, 1974, doi: 10.2307/2347125.

[3] D. R. Cox, “The Regression Analysis of Binary Sequences,” *J. R. Stat. Soc. Ser. B*, vol. 20, no. 2, pp. 215–232, 1958, doi: 10.1111/j.2517-6161.1958.tb00292.x.

[4] C. B. Begg and R. Gray, “Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions,” *Biometrika*, vol. 71, no. 1, pp. 11–18, 1984, doi: 10.1093/biomet/71.1.11.

[5] D. Böhning, “Multinomial Logistic Regression Algorithm,” *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, 1992, doi: 10.1007/BF00048682.

[6] S. S. Al-Obaidi and J. I. Al-Bayati, *Applied Statistics: Methods of Hypothesis Testing*, 3rd ed. Modern Academy for University Books, 2022.

[7] GraphPad Software, “Contingency Table Analysis,” 2025, *San Diego, California, USA*.

[8] S. J. Haberman, “Log-Linear Fit for Contingency Tables,” *J. R. Stat. Soc. Ser. C (Applied Stat.*, vol. 21, no. 2, pp. 218–225, 1972, doi: 10.2307/2346506.

[9] J. A. Nelder and R. W. M. Wedderburn, “Generalized Linear Models,” *J. R. Stat. Soc. Ser. A*, vol. 135, no. 3, pp. 370–384, 1972, doi: 10.2307/2344614.

[10] V. Bewick, L. Cheek, and J. Ball, “Statistics Review 14: Logistic Regression,” *Crit. Care*, vol. 9, no. 1, pp. 112–118, 2005, doi: 10.1186/cc3045.

[11] J. C. Stoltzfus, “Logistic Regression: A Brief Primer,” *Acad. Emerg. Med.*, vol. 18, no. 10, pp. 1099–1104, 2011, doi: 10.1111/j.1553-2712.2011.01185.x.

[12] J. Fox, “The {R} Commander: A Basic-Statistics Graphical User Interface to {R},” *J. Stat. Softw.*, vol. 14, no. 9, pp. 1–42, 2005, doi: 10.18637/jss.v014.i09.

[13] S. H. A. Al-Farhud, “Using Logistic Regression to Study the Factors Affecting Stock Performance,” *J. Al-Azhar Univ. (Natural Sci. Ser.*, vol. 9, pp. 47–68, 2014.

[14] T. A. Al-Khazaali, “Estimation of a Logistic Regression Model for a Multi-Level Dependent Variable,” University of Karbala, Iraq, 2021.

[15] I. S. Hanaish and I. I. Al-Daq, “Using Binary Response Logistic Regression Technique to Diagnose the Most Important Factors Affecting the Causes of Premature Infant Mortality: A Case Study in the City of Misrata for the Years 2019 and 2022,” *Afaq Al-Ma’rifah J.*, vol. 2, no. 8, p. 21, 2025.