Zanco Journal of Human Sciences (ZJHS)
ISSN (Print):2218-0222
ISSN (Online): 2412-396X

Vol.28 Issue 1 2024
ID No. 1453 (PP 258 -277)

https://doi.org/10.21271/zjhs.28.1.17

Research Article

Estimating Parameters and Predicting the Risk of MiscarriageInfertility, and Psychological Disorders Among Chemical Attack Survivors of Kurdistan by Using Multinomial Logistic Regression Analysis

Nazeera Sedeek Kareem Al Barznji*

* Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil
nazeera.kareem@su.edu.krd

Reiceved ǁ 31/07/2023
Accepted ǁ 06/09/2023
Published ǁ 15/02/2024

Keywords:
Multinomial Logistic Regression,
Parameter Estimates Wald statistic,
Classification,
chemical gas,
Kurdistan.

Abstract

Since the Ba'ath regime took over Iraq, they oppressed the Kurdish nation in every way. In 1987, on the pretext of the Iran-Iraq war, they bombed the two cities (Sardasht and Piran shar) in East Kurdistan After that year, 40 villages and cities in South Kurdistan were bombed include of them was Halabja on 3/16/1988. In addition to 5,000 martyrs and 10,000 wounded, most of the people of Halabja suffer from various diseases, to determine the level of risks of chemical gas to survivors suffering from miscarriage, infertility and mental disorders, this study was recorded by interviewing (500) infected people recorded in the Ministry of Martyrs and Anfal in Kurdistan.The results found that the most risk on [Abortion disease, Psychological disease (female), Infertility disease(male), and transmission of disease], and the level (0.70-0.90) is the most affected by the chemical gases that cause Miscarriage, infertility, and Psychological diseases In this paper IBM SPSS software used for Analyzing the data.

About the Journal
ZANCO Journal of Humanity Sciences (ZJHS) is an international, multi-disciplinary, peer-reviewed, double-blind and open-access journal that enhances research in all fields of basic and applied sciences through the publication of high-quality articles that describe significant and novel works; and advance knowledge in a diversity of scientific fields.
https://zancojournal.su.edu.krd/index.php/JAHS/about

1. Introduction

Multinomial logistic regression (MLR) analysis is the best method for predicting how chemical gas will affect the risk of Miscarriage, Infertility, and Psychological Disorders of chemical attack survivors in Kurdistan. Whenever a response variable contains more than two nominal or unsorted subcategories, "logistic regression analysis" one of the most beneficial models available. In order to estimate the likelihood, multiclass problems are used to binary logistic regression. Verhulst, a French mathematician, created the binary logistic function in the 19th century to describe population expansion and the progression of autocatalytic biochemical reactions. Verhulst published his recommendations after Quetelet modified them between 1838 and 1847. For the decades up to the early 1830s, the logistic model was in excellent agreement with the actual trajectory of the populations of France, Belgium, Essex, and Russia, In 1920, Pearl and Reed made a fresh discovery about the logistic function while researching the rise of the American population. The multinomial logistic regression is generalization of logistic regression, depend on the maximum likelihood method (ML) to estimate parameters of the model, the advantage of the multinomial logistic regression model it do not need normally distributed (explanatory) variables and it used to prediction.

2. Objective

Due to the tragic spread of miscarriage, infertility and Psychological disorders in the chemical-bombed areas, we aim to predict the ratio of the effects of the chemical gas on the bodies of the victims by multinomial logistic regression.

3. Theoretical Aspect

3.1 Logistic function

A technique for fitting a regression is logistic regression. ,

$$Y = f(G)$$

$$f(G) = \frac{1}{1+e^{-G}} \quad \dots (1)$$

$$f(x) = \frac{e^{A+B\,x}}{1+e^{A+B_1\,x}} = \frac{e^{(A+B\,x)}\,e^{-(A+B\,x)}}{[1+e^{(A+B\,x)}]e^{-(A+B\,x)}}$$

$$f(x) = \frac{1}{[1+e^{-(A+B\,x)}]} \quad \dots (2)$$

Where: A. coefficient of determination B. slope of the regression [Greene, W. H. (2012)]

3.2 The Multinomial Logistic Model:

To obtain the logistic model from the multinomial logistic model, consider the following linear relationship: Baltas and Doyle(2001), Darroch and Ratcliff (1972) & [Schafer J.L. (2006)]

$$G = A + B_1x_1 + B_2x_2 + \cdots + B_mx_i$$

$$f(x) = \frac{1}{[1+e^{-(A+B_1x_1+B_2x_2+\cdots+B_mx_i)}]}$$

$$f(x) = \frac{1}{[1+e^{-(A+\sum B_m x_i)}]} \quad \dots (3)$$

We denote the probability that a binary dependent variable Y=1 (P) in the logistic function as follows:

$$P = P(\,Z=1) = \frac{1}{[1+e^{-(A+\sum B_m x_i)}]}$$

Consider the following steps for each of the m potential outcomes to arrive at the multinomial logistic model:

$$ln\frac{Pr\ (Z_i = 1)}{Pr\ (Z_{i_i} = m)} = A + B_1x_1$$

⋮ ⋮

$$ln\frac{Pr\ (Z_i = m-1)}{Pr\ (z_i = m)} = A + B_mx_m$$

… (4)

“If we exponentiation both sides, and solve for the probabilities”, we get:

$$Pr(Z_i = 1) = Pr(Z_i = m)e^{A+B_1x_i}$$

$$Pr(Z_i = m-1) = Pr(Z_i = m)e^{A+B_{m-1}x_i}$$

$$1 = 1 + \sum_{m-1}^{m} e^{A+B_mx_i}\ Pr(Z = m)$$

⋮ ⋮

$$Pr(Z = m) = \frac{1}{1 + \sum_{m-1}^{m} e^{A+B_mx_i}}$$

… (5)

This will help us determine the other probability:

$$Pr(Z_i = 1) = \frac{e^{A+B_1x_i}}{1 + \sum_{m-1}^{m} e^{A+B_mx_i}}$$

$$Pr(Z_i = 2) = \frac{e^{A+B_2x_i}}{1 + \sum_{m-1}^{m=1} e^{A+B_mx_i}}$$

$$Pr(Z_i = m\ ) = \frac{e^{A+B_mx_i}}{1 + \sum_{m-1}^{m=1} e^{A+B_mx_i}}$$

… (6)

3.3Model Fitting Information

3.3.1The Maximum Likelihood Estimators: the (MLR) analysis used to obtain the maximum likelihood estimate of (it estimated by the Newton-Raphson iteration method or by estimation method of Fisher-Scoring iteration because both methods have the same result**.** Given a sequence of n data point (Carpenter, 2008)

$$D = (x_i, c_j)_{j<n},\ \text{With } x_j \in R^d \text{ and } c_j \in \{0, \dots, m-1\}.$$

… (7)

Assuming that the stochastic part e of the utility function is distributed with a double exponential distribution, the likelihood of observing actual choices, given input vector x and model parameter vector B, can be expressed by

$$L(Y|X,B) = \prod_{i<n} P(m|x_i, A, B) = \prod_{i=1}^{n}\left[\frac{\exp\ (A + B_mx_i)}{1 + \sum_{m'<k-1} \exp\ (B_{m'}x_i)}\right]$$

… (8)

In a model with a parameter matrixβ, the data's log likelihood is

$$\log l(B) = \log P(D\,|B) = log \prod_{i<n} P(m|x_i, A, B)$$

$$= \sum_{i<n} log\ p(m|\,x_I, A, B)$$

$$= \sum_{i<n} log \frac{\exp\ (A, B_{m,}x_i)}{1+\sum_{m'<k-1} \exp\ (A, B_{m'}x_i)}$$

“The maximum likelihood (ML) estimate $\hat{\beta}$is the value of B maximizing the likelihood of the

$$\hat{B}_{MLE} = \arg\max \log l(B) = \arg\max \log P(D|C)$$

… (9)

We can do gradient descent on B

… (10)

$$\frac{\partial}{\partial B_{m,i}} log\ L(B) = \sum_{i<n} \frac{\partial}{\partial B_{m,i}} log\ P(m|x_i, A.B)$$

$$= \sum_{i<n} \{ \frac{\partial}{\partial \beta_m} A, B_m . x_i - \frac{\partial}{\partial \beta_{c,i}} log(1 + \sum_{m<k-1} exp(A, B_{m'} . x_i)) \}$$

$$= \sum_{j<n} \{ x_{j,i} I(m) - \frac{1}{1 + \sum_{m<k-1} exp(A, B_{m'} . x_j)} \frac{\partial}{\partial B_{m,i}} (1 + \sum_{c'<k-1} exp(A, B_{m'} . x_j)) \}$$

$$= \sum_{j<n} \{ x_{j,i} L(m) - P(m|x_m, A, B) x_i \}$$

$$\frac{\partial}{\partial B_m} log\, l(B) = \sum_{j<n} \left\{ x_{j,i} I(m) - \frac{1}{1 + \sum_{m<k-1} exp(A, B_{m'} . x_j)} \sum_{c'<k-1} \frac{\partial}{\partial \beta_{m,i}} exp(A, B_m . x_i) \right\}$$

$$= \sum_{j<n} \left\{ x_{j,i} L(m) - \frac{1}{1 + \sum_{m<k-1} exp(A, B_m . x_i)} \sum_{c'<k-1} \left(exp(B_{c'} . x_j) \frac{\partial}{\partial B_{c,i}} B_{c'} . x_i \right) \right\}$$

$$= \sum_{j<n} \{ x_{j,i} [L(m) - P(c|x_i, A, B] \} \quad \ldots \quad (11)$$

$$\sum_{i<n} \{ x_i [P(m|x_i, A, B] \} = 0 \quad \ldots \quad (12)$$

The discrepancy between both the outcome and the model prediction is the residual for training

3.3.2 Likelihood Ratio Test:

Is used whether coefficients for predictor factors are statistically significant overall under beta for the model's covariates equal zero.
(s). *Karl (2016), Osibanjo (2015) and Barznji (2018)*

$$LR = -2ln \left[\frac{Likelihood\ without\ the\ variable}{likelihood\ with\ the\ variable} \right] = -2ln \left(\frac{L\ \ at\ H_0}{L\ at\ MLE(s)} \right)$$

$$= -2L\ H_0 + 2L\ (MLE) \quad \ldots \quad (13)$$

When n is big, $LR \sim x^2$ has a degree of freedom equals the number of parameter estimates.
Note: Likelihood ratio test is assumed as a method to Alternative to Wald test

3.4 Model Fitting Criteria Estimation Fit

The likelihood value, which is sums of squares values used to the fundamental indicator estimation technique fits data. With a value of -2 times the log of the likelihood (-2LL), often known, L.R measures model estimation fits the data. Since likelihood = 1 and -2LL = 0, the minimal value of -2LL is 0, which amounts to a great combination. So, the higher the quality fit, the lower the -2LL number. The -2LL value can be used to compute measurements similar to the R2 measures in multiple regressions or to examine the change in fit between two equations.

3.4.1Akaike Information Criterion (AIC):

$$AIC = \frac{-2 \ln \hat{L}(M_k) + 2p}{N} \quad \ldots (14)$$

Where

p is the model's parameter

L denotes “the maximum likelihood” of the model $\hat{L}(M_k)$. It is thought that a model fits more accurately if its AIC is less.

3.4.2Bayesian Information Criterion (BIC)

established straightforward and precise, is a third measure of fit, particularly if there are roughly 40 observations (Raftery, 1995). This is how BIC is described:

$$BIC_k = D(M_k) - df_k LnN \quad \ldots (15)$$

"Where: $D(M_k)$ is the deviance of the model M_k
df_k is the degrees of freedom for deviance"

3. 5Goodness-of-Fit Fagerland Morten W [1], David W Hosmer, Anna M Bofin(2008)

There are two techniques to evaluate a multinomial logistic regression model's goodness-of-fit to Examining model-estimating fit is one method, while evaluating predicted accuracy is another "like the classification matrix in discriminate analysis". Although the two methods analyze model fit from different angles, they produce comparable outcomes

3. 5.1 Chi-Square Pearson goodness of fit

A goodness of fit test may be used:

$$\mathcal{X}^2 = \sum_{i=1}^{N} (O_i - E_i)^2 / E_i \quad \ldots (16)$$

Where:
$\mathcal{X}^2$= Pearson's cumulative test statistic, which asymptotically approaches a Chi-square distribution.
O_i = The observations.
N = Total of observations
$E_i E_i$ = The expected (theoretical) count *David, Paul &Richard (2007)*& Pearson, K. (1900)

3.5.2 Deviance Goodness of Fit

It measured as follows:

$$D(M_k) = 2\sum_{i=1}^{n}\sum_{j=1}^{i} S(y_i = j) log\left[\frac{S(y_i = j)}{Q_{ij}}\right] \quad \ldots (17)$$

Where $\widehat{Q}_{ij}$ is a predicted value for $i = 1 .2.3. \ldots . n$ [Isaac O. Ajao1; Adebisi A. Ogunde(2017)] [Hosmer D.W. and S. Lemeshow (1980)].

3.6 Pseudo R2

3.6.1 McFadden's adjusted R2

The McFadden's adjusted $\mathbf{R}^2$measure, compares a full model with all parameters (C_{full}) the model contains only intercept ($C_{intercept}$) and defined as:

$$\bar{R}^2_{McF} = 1 - \frac{ln\,\hat{L}(C_{full}) - K^*}{n\hat{L}(C_{intercept})} \quad \ldots (18)$$

K: The number of parameters [D. Mcfadden. (1974)]

3.6.2 Cox and Snell Pseudo R2

The calculation of the percentage of variation decreased including additional variables in forms the basis of an alternate goodness.

The equation as follows:

$$R^2_{McF} = 1 - \left[\frac{L(0)}{L(\hat{\beta})}\right]^{\frac{2}{w}} \quad \ldots (19)$$

"Where $L(\widehat{B})$ is the likelihood of the current is model; $L(0)$ is the likelihood of the initial model Cox, D.R. & Snell, E.J. (1989)

3.6.3 Nagelkerke Paseudo R2

The Nagelkerke measure adjusts the C and S measure

$$R_N^2 = \frac{R_{cs}^2}{\max(R_{cs}^2)}$$ … (20)

Where $\max(R_{cs}^2) = 1 - \{l(0)\}^{\frac{2}{w}}$ [*Bo Hu, Jun Shao and Mari Palta (2006)]*

3.7 Parameter Estimates

It is a subfield of statistics where the estimation of a distribution's parameters is done using sample data.

3.7.1 Odds and Odds Ratio (Exp (B))

-Odds

The algebraic expression of the linear regression expression is identical to the possibility that a case corresponds to the predictor variables This demonstrates connecting. Due to its range between zero and infinity, the logit provides an adequate criterion on is readily.

-Odds ratios(OR): Can be understood as the impact of changing X by one unit on the odds ratio that would result from holding the other model variables unchanged. An independent continuous variable's odds ratio is given by the following formula:

$$\text{odds ratio} = \frac{\frac{P(1)}{1-P(1)}}{\frac{P(0)}{1-P(0)}}$$ …

(21)

$$OR = \frac{odds(x+1)}{odds(x)} = \frac{\left(\frac{P(x+1)}{1-P(x+1)}\right)}{\left(\frac{P(x)}{1-P(x)}\right)} = \frac{\frac{e^{A+B(x+1)}}{1+e^{A+B(x+1)}}}{\frac{e^{A+B(x)}}{1+e^{A+B(x)}}} = \frac{\frac{e^{A+B(x+1)}}{1}}{\frac{e^{A+B(x)}}{1}} = \frac{e^{A+B}}{e^{A}} = e^{B}$$

$$lnOR = ln\frac{e^{A+B}}{e^{A}} = e^{B} = odds\ ratio$$

(22)

"The regression coefficient (B) in the population model is the log (OR)"

3.7.2 Standard error

$$SE = \frac{\sigma}{\sqrt{n}}$$

3.7.3Wald statistic or the Z-test Criterion: Hauck Jr, W. W., & Donner, A. (1977).

The Wald test compares an estimated $\hat{B}$ with an assumed A value. The estimated B was discovered as the maximum argument of the unconstrained probability function. In specifically, the curvature of the log-likelihood function weighs the squared difference $\hat{B} - A$.Test on a single (individual)

$$W = \frac{(B-A)^2}{varB}$$

…. (23)

3.8 The Classification table allows the standardization process to identify and classify key words by providing: Standard abbreviations for each word;

classification table of the multinomial logistic regression model. The predicted categories (columns) have been calculated using the fixed part of the model. The values, except for the number of sample plots, are row percent's. The overall classification efficiency is 60.8%

The classification table shows the practical results of using the multinomial logistic regression model. For each case, the predicted response category is chosen by selecting the category with the highest model-predicted probability. Cells on the diagonal are correct predictions. it is another way to predict accuracy

4.Practical Aspect

4.1Summery about Kurdistan and Chemical Attack

Greater Kurdistan locations in the western Asia ,it is a land of the Kurdish nationalist, Kurdistan indelibly consists four regions“ Northern Kurdistan in Turkey, Southern Kurdistan Iraq , Eastern Kurdistan in Iran, Western Kurdistan in Syria and another locations like ,Armenian and Azerbaijan ”Kurdistan Region (Southern Kurdistan)[Kaya, Z... (2020)] first acquired autonomous situation in a 1970.
In the modern times, the Battle of 1915 is the first battle zone most notorious for Nazis chemical gas use, they used (chlorine gas), where 6000 soldiers were killed. After the First World War more sophisticated chemical agents such as, Mustard gas were developed the Italian army against Ethiopia during the 1936-1937 and used in China by Japanese armed forces during the Second World War. It seems that parties of war continuously disregarded international laws and regulations, which were in place as far back as 1675. Even though new treaties such as Geneva Convention took place in 1925 prohibiting the use of chemical gases years later, Before the Iran-Iraq War, Iraq had a sizable chemical assault program that dated back to the 1960s. On June 28, 1987, Iraqi aircraft on Sardasht, and piranshar a cities in Iranian Kurdistan (Eastern Kurdistan) **Namo Kurdistani** (2018), dropped chemical weapons. For the first time in human history, Iran-Iraq War: Chemical weapons were deployed by Iraq against Iranian soldiers, targeting a city full of civilians with gas attacks. Along with aerial bombs, artillery, rocket launchers, tactical rockets, glider sprayers, and missile launchers, it also used chemical weapons against its own Kurdish people. In 1987–1988, he used 40 Kurdish villages as testing sites for chemical strikes on thousands of innocent civilians. On March 16, 1988, the worst of these attacks wreaked havoc on the city of Halabja, leaving more than 5000 people dead and 10,000 injured [(Hiltermann, J. R. (2007)]. . The first international medical mission to visit Halabja, [Medecins Sans Frontieres] (MSF), sent a Belgian-Dutch team that confirmed the use of mustard gas and perhaps cyanide. In all of Kurdistan, between 50,000 and 100,000 people are believed to have died, according to Human Rights Watch/Middle East. Organic phosphorous nerve gases that attack the eyes and respiratory systems include sarin, tabun, and VX. The nose, throat, and lungs' mucous membranes are all affected by mustard gas, a blistering chemical.[Amire Qadr. (2015)]According to medical, genetic reports, chemical gas affect chromosomes change it to abnormality which damage the romosomes causes mutation to know the seriousness of these diseases are transmitted through the genes to future generations, [Paul A. D’Agostino (2008)]

Figure(1) Greater Kurdistan

figure 2: Chemical Attack areas in Eastern Kurdistan [Omar Lhafdoozi.(2006) &Namo Kurdistani (2018)]

In this figure, the red areas are the chemically bombed areas of East Kurdistan

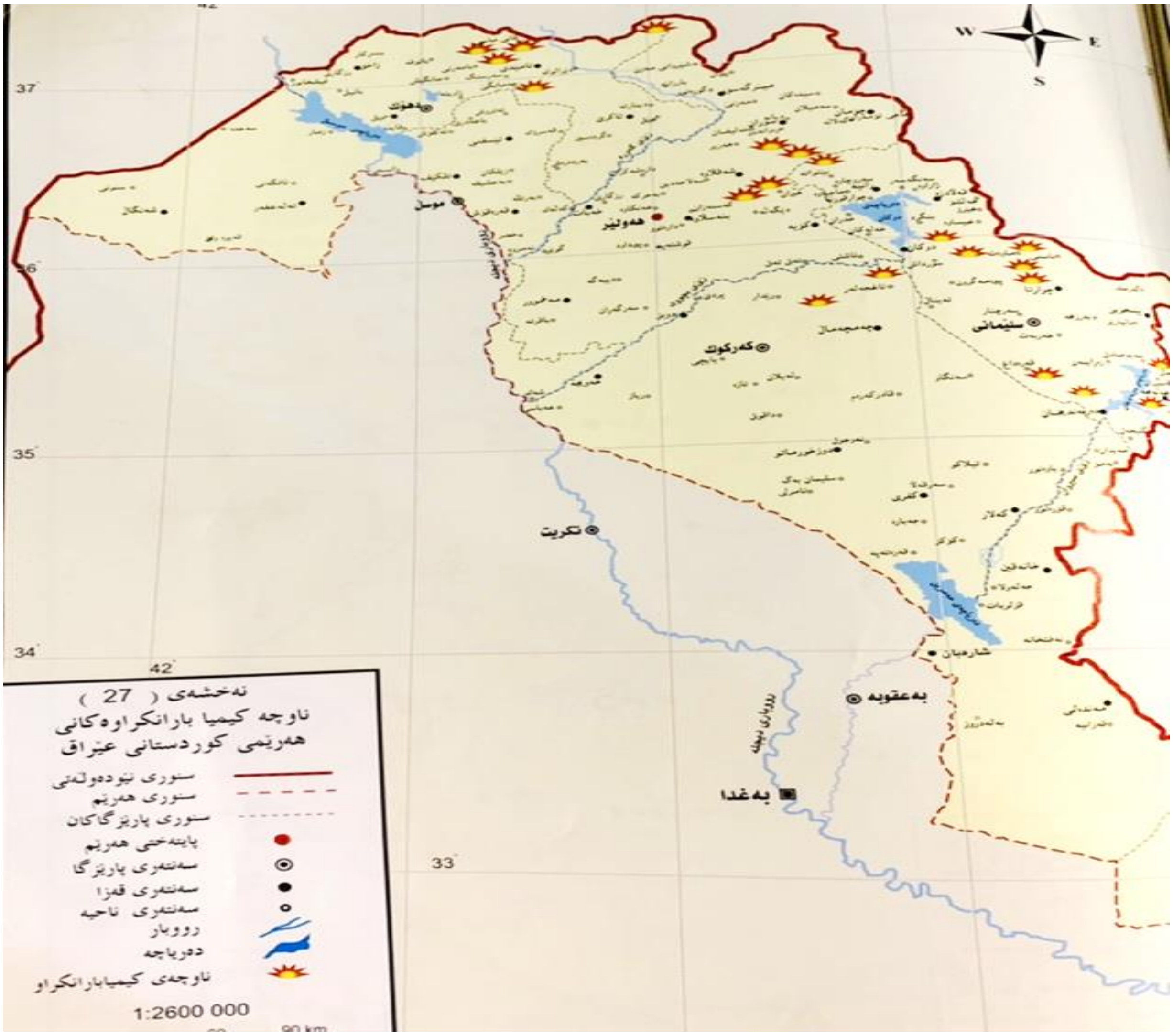


Figure 3: Chemical Attack areas in Southern Kurdistan Dr.Haddad,H.Y,(2009)]
In this figure, the red areas are the chemically bombed areas of South Kurdistan (Kurdistan Region)

4.2 Description of the Data

4.2.1Data collection methods

This study included the sample of size (500) patients has chosen by random sampling which was obtained from several sources

1-list of patients, which are recorded by ministry of Martyr and Anfal. .

2-List of Questionnaire.

3-Taking information from the records of some doctors who specialize in [Abortion, Infertility and Psychological diseases] in the areas of chemical attack.

Table (1) Percentage of the Members Presented during a Chemical Attack and **data about number of patients by Chemical Attack regions**

Percentage of the Members presented during a Chemical Attack		**Percent**
		86.1%
Region	**Locations**	**Number of patients**
Southern Kurdistan	Duhok	50
	Erbil	100
	Sulaimania	115
	Karkuk	55

	Halabja	150
Eastern Kurdistan	Sardasht and piran shar	30
Total		500

This table shows patients sample in this paper, (Number of patients with their locations whom infected by chemical gas in Southern Kurdistan and Eastern Kurdistan contain 500 cases

4.3 Description of the variables

Table (2) Shows description of the variables

The Response (Dependent) Variable	
Y	Levels of the effect of chemical gas on the survivors of the chemical attack on Kurdistan [levels of the percentages of the risk chemical gas (0.28-0.48), (0.49-0.69) and(0.70-0.90)]
The predictor variables (Independent variables)	
X_1	Gender
X_2	Miscarriage Disease
X_3	Infertility Disease
X_4	Psychological Disease
X_5	Transmission of Disease
X_6	Blood Group

This table defines the types of variables

The upper part of the table symbol(Y) represents the response variables or (non-independent variables).

The lower part of the table shows the six independent variables[X_1 =Gender, X_2 =Miscarriage Disease,

X_3= Infertility Disease, X_4=Psychological Disease, X_5=Transmission of Disease, X_6 =Blood Group]

Table (3) Descriptive Statistical Analysis for the Response Variable and the predictor (Independent) Variable

Levels of the percentages of the risk chemical gas	Variables					
	Y= Response Variable	predictor variables				
		X_2 =Miscarriage 88.4%		X_3=Infertility 61.0%	X_4=Psychological diseases 20.7%	
0.28-0.48	13.6 %	11.6%		13.2%	20. 4%	
0.49-0.69	36.4%	40.0%		39.0%	25.0%	
0.70-0.90	50 %	48.4%		47.8%	54.6%	
Total	100%	100.0%		100.0%	100.0%	
		X_1 :Gender Types			X_5 : Hereditability of diseases	
		female	61.8%		Yes	85.0%
		male	38.2%		No	15.0%
		Total	100.0		Total	100.0%
		Blood groups				
		A-	7.2%			
		A+	23.4%			
		AB-	2.0%			
		AB+	5.4%			
		B-	0.8%			
		B+	7.45			
		O-	1.2%			
		O+	52.6%			
		Total	100.05			

This table contains the analyses of the Response Variable levels percentages of the effect of the chemical gas on the body of the survivors of the chemical attack on Kurdistan and the predictor variables percentages of diseases X_2= Miscarriage, X_3=Infertility and X_4=Psychological diseases **as** results of the chemical attack as follows**:**

First: in the response variable levels results the largest percentage is (50%) at the level [0.70-0.90], followed by is (36.4%) at the level [0.49-0.69], and (13.6 %) percentage at the level [0.28-0.48]**.**

Second: by the results of predictor variables X_2 = Miscarriage, has the largest percentage (88.4%) followed by X_3=Infertility has the 61.0% percentage and Psychological diseases has the smallest percentage 20.7%.

Third: all disease has the largest percentages at the level of (0.70-0.90) followed (0.49 - 0.69) then (0.28-0.48) has a smallest percentage.

Fourth: the X_1 = Gender with ,X_5 = the Hereditability of diseases in the patients, the female patients has (61.8%) which it is the largest percentage of Gender)but in the Hereditability of diseases is (85.0%) of families has genetic and the finally predictor variables percentage of blood groups of the chemical attack patients shows blood groups [(O^+) has largest percentage (52.6)followed by the percentage of the (A^+) has(23.4)the lowest percentage is in(B^-)with (0.8%) it refers persons with (O^+)blood groups more affected by all chemical gas.

4.5 Advanced Statistical Analysis: Multinomial Logistic Regression (MLR)

4.5.1Model Fitting Information

Table (4) shows the model fitting information

Model Fitting Information						
Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig. (P- value)
Intercept Only	303.593	312.022	299.593			
Final	309.338	385.201	273.338	26.254	16	0.011

This table : shows both methods of (AIC) model fitting criteria for intercept only(the null model) is (303.593) and the AIC for final model is (309.338) but the model fitting criteria by Bayesian Information (BIC) for Intercept only (the null model) (312.022)and BIC for final model is (385.201) they the which it refer to final model by (AIC) . tests for Intercept Only is (299.593) and for final model is(273.338) . Since, the smaller the value is the better fit. We see model fit is significant $\chi^2_{(16)}= 26.254$, has [P-value =0.011< 0.05], which indicates that our full model predicts significantly better, or more accurately, than the null model indicating a good model fit confirm to the effect of chemical gas to get all three disease

4.5.2Goodness-of-Fit

Table (5) Methods of Goodness-of-Fit

	Chi-Square	d,f	Sig.
Pearson	102.151	94	0.266
Deviance	117.480	94	0.077

This table: represents the further evidence Goodness-of-Fit test by both methods the Pearson Chi-Square (102.151) with [P-value =0.266> 0.05] and Deviance Chi-Square (117.480) with [P-value =0.077> 0.05] interpret lack of significance it refer to good fit**.**

4.5.3Pseudo R2 Measures

Table (6) Pseudo R^2 Measures

Cox and Snell	0.812

This table: shows Pseudo R-Square “it is the proportions of variations n the dependent variable explained by the independent variables 81.2% according to Cox &Snell R^2 value, R-square statistic is best type”(it refers to the effect of chemical gas caused to 81.2% of all diseases)

4.5.4Parameter Estimation

Table(7) shows Parameter Estimates (an iterative MLE method)

Y		B coefficients for the models	Std. Error	Wald test	Sig. (p-value)	Exp (B) odds ratios (OR)
0.49 - 0.69	Intercept	3.518	1.317	7.140	0.008	1.800
	X_1= Gender	-1.287	0.592	-2.17	0.030	0.276
	X_5=Transmission of disease	-0.0.95	0.433	4.768	0.029	0.909
	X_2= Miscarriage disease	0.517	0.011	-3.06	0.000	1.676
	X_3 =Infertility disease (female) (male)	-0.552 0.1793	0.492 0.317	1.874 0.870	0.012 0. 572	0.576 1.196
	X_4=Psychological disease (female) (male)	0.188 -1.754	0.094 0.181	2.00 -9.72	0.045 0.000	1. 206 0.173
0.70 - 0.90	Intercept	0.562	0.202	2.77	0.006	
	X_1= Gender	-0.011	0.006	-1.78	0.075	0.989
	X_5= Transmission of disease	0.496	0.193	2.56	0.010	1.643
	X_2= Miscarriage disease	0.942	0.225	4.18	0.000	2.565
	X_3 =Infertility disease (female) (male)	0.661 -0.283	0.216 0.040	-3.06 -0.71	0.002 0.477	0.516 0.754
	X_4=Psychological disease (female) (male)	0.975 0.588	0.236 0.228	4.12 0.338	0.000 0.010	2.650 1.800

An important feature of the multinomial logit model is that it estimates *k-1* models, where *k* is Based on the results obtained in this table, the response variable has three levels so it estimates two models**,** and by Sig. (p-value) it can be shown that in level (0.49 - 0.69) there is one predictor variable X_3 =Infertility disease (male) is not significant, It means all predictor affected by chemical gas except Infertility disease in male .but in in level (0.70 - 0.90), X_3 =Infertility disease (male) and X_1(gender) are not significant, It means all predictor affected by chemical gas except Infertility disease in male, and X_1(gender)

The odds ratio or [Exp(B)] of a coefficient indicates how much the risk of the chemical gas. An odds ratio > 1 indicates that the risk increases in level (0.49 - 0.69) like [X_2= Miscarriage disease= 1.676, X4=Psychological disease (female)= 1. 206, X_3 =Infertility disease(male)= 1.196] but in level (0.70 - 0.90). X4=Psychological disease (female)= 2.650, X_2= Miscarriage disease=2.565 , X_5= transmission of disease=1.643

The prediction equation with significant factors can be written are:

For level (0.49- 0.69) the prediction equation is:

$\hat{y}$ = Log (odds) = 3.518-1.287 (gender) -0.095 (Transmission of disease) + 0.517 (Miscarriage disease) -0.552 (Infertility disease(female)) + 0.188 (Psychological disease (female))- 1.754 Psychological disease (male)

For level (0.70- 0.90) the prediction equation is:

$\hat{y}$ = Log (odds) = 0.562 -0.496 (Transmission of disease) + 0.942 (Miscarriage disease) -0.661 (Infertility disease(female) + 0.975 (Psychological disease (female))-0.588 (Psychological disease female)-

As with any regression, the positive coefficients indicate a positive relationship with the dependent variable. It means This means that chemical gases are very effective for infection:

Miscarriage disease and Infertility disease (male) and Psychological disease (female) at level**(0.49- 0.69)**

However, transmission of disease, miscarriage disease and infertility disease(female) and Psychological disease (female) and (male)at level **(0.70- 0.90)**

4.5.5Classification

Table (8) Classification

	Classification				
Observed		Predicted			
		0.28-0.48	0.49-0.69	0.70-0.90	Percent Correct
0.28-0.38		0	9	59	0.0%
0.49-0.69		0	27	155	74.8%
0.70-0.90		0	29	221	88.4%
Overall Percentage		0.0%	73.0%	87.0%	79.6%

This table shows the value of the accuracy the results used to emphasize which level has a best predicted by the model. Where the first level (0.70-0.90) correctly predicted by the model with (88.4%) the second level (0.49-0.69) correctly predicted by the model with (74.8%) and the third level (0.28-0.48) correctly predicted by the model with (0.0%**)** but the Overall Percentage correctly predicted by the model with79.6% it is a good prediction to the effect of chemical gas

5.1 Conclusions

After applying the multinomial logistic regression model on the sample of size (500) patients of chemical attack in (Southern Kurdistan and Eastern Kurdistan) this sample was chosen by stratify random sampling which were recorded by the Ministry of Martyr and Anfal in Kurdistan and several other sources. Our focus is on Miscarriage, Infertility and Psychological diseases. We noted some of the patients having more than one disease. The conclusions reached are as follows:

The conclusions reached are as follows:

--The proportion of people who were presented at the time of the chemical attacks were (86.1%).

--The proportion of (female 61.8%) and (male 38.2%)

--The proportion of (Miscarriage 88.4%)و (Infertility 61.0%) and (Psychological diseases (20.7%)

--all disease in the level of (0.70-0.90) has the largest percentage followed (0.49 -0.69) then (0.28-0.48) has a lowest percentage

--**In blood group** (O^+ with percentage 52.6%) has largest percentage, followed by (A^+ with23.4%), but the lowest percentage was (B^- with 0.8%) so this refers people with (O^+) blood groups were more affected by the chemical gas.

-**Model Fitting Information** (AIC), (BIC), and Likelihood Ratio Chi-Square test, the Full Model (containing all the predictors)has a significant improvement fit over the Null Model by $\chi^2_{(16)}$= 26.254, has [P-value =0.011< 0.05] it emphasizes the existence of the effect of chemical gas on every explanatory variables.

--**Goodness-of-Fit** test the Pearson Chi-Square with [P-value =0.266> 0.05] and and Deviance Chi-Square [P-value =**.077**> 0.05] which interpret lack of significance, which refers to a good fit, it refers to the previous explanation to the effect chemical gas on causing diseases.

--**Pseudo R-Square** 81.2% according to Cox &Snell R^2 value, R-square statistic is best type" (it refers to the effect of chemical gas caused to 81.2% of all diseases)

--**Parameter estimation**. by Sig. (p-value) all predictor variables affected by chemical gas except Infertility disease in male but in in level (0.70 - 0.90) X_3 =Infertility disease (male) and X_1(gender) are not significant, It means all predictor affected by chemical gas except Infertility disease in male, and (gender).and by the odds ratio or [Exp(B)] the risk increases in level (0.49 - 0.69) for [X_2= Miscarriage disease and X_4=Psychological disease (female, X_3 =Infertility disease(male)= 1.196] but in level (0.70 - 0.90). X4=Psychological disease (female), X_2= Miscarriage disease, X_5= transmission of disease. The final formulated model for both levels contains the significant predictor variables as with any regression, the positive coefficients indicate a positive relationship with the dependent variable. It means This means that chemical gases are very effective for infection: Miscarriage disease and Infertility disease (male) and psychological disease (female) at level **(0.49- 0.69)**

However, transmission of disease, miscarriage disease and infertility disease(female) and psychological disease (female) and (male)at level **(0.70- 0.90)**

-**Classification** The highest classification opportunity that classifies the level (0.70-0.90) correctly predicted by the model in (MLR) is (88.4%) with finally correct prediction for model by overall percent of testing (79.6%).

6. Recommendations

1- We recommend the opening of a hospital for the treatment of patients afflicted by chemical attack in Kurdistan.

2- More research needs to be done on other diseases that are affected by chemical gas of chemical attacks.

3- Kurdistan Regional Government should start collecting many various data regarding the chemical attacks done on Kurdistan, due to the lack of proper data available now.

References`:

[1] Agresti, A. (2007). An Introduction to Categorical Data Analysis. John Wiley & Sons, Inc

[2] Barznji, N.S.K, (2018)"*Using Logistic Regression Analysis and Linear Discrimination Analysis to identify the risk factors of diabetes*" Vol.22, No.6,(pp 248-268), ID No.2115)

[3] Bo Hu, Jun Shao and Mari Palta (2006) *"PSEUDO-R² MODEL IN LOGISTIC REGRESSION"* University of Wisconsin-Madison Statistica Sinica 16), 847-860

[4] Cox, D.R. & Snell, E.J. (1989) " *Analysis of Binary Data*" Second Edition. Chapman [5]Darroch, J.N. & Ratcliff, D. (1972)" *Generalized iterative scaling for log-linear models*"The Annals of Mathematical Statistics. 43 (5): pp.1470–1480.

[6]David E. B, Paul F. V & Richard D. D (2007)." *Stats, Modeling the World*"pp. 606-627

[7] D. Mcfadden. (1974) Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, Frontiers in Econometrics, chapter Four, pages 104–142. Academic Press, New York,.

[8] Dr. Haddad, H.Y Abdurrahman.S.M & Khoshnaw H.M,(2009)"*Atlas of Iraqi Kusrdistan Region, Iraq and the World*" first edition, Erbil

[9]Fagerland Morten W [1], David W Hosmer, Anna M Bofin(**2008**) " *Multinomial goodness-of-fit tests for logistic regression models*"Affiliations expand MID: 18203120 DOI: 10.1002/sim.3202

[10]Garson, D. (2009). "*Logistic Regression with SPSS*". North Carolina State University, Public administration Program

[11] Greene, W. H. (2012)" *Econometric Analysis*" (Seventh ed.). Boston: Pearson Education. pp. 803–806. ISBN 978-0-273-75356-8

[12] Hauck Jr, W. W., & Donner, A. (1977)" *Wald's test as applied to hypotheses in logit analysis*"Journal of the American statistical association, 72(360a), 851-853.

[13] Hosmer D.W. and Lemeshow (1980)" *A goodness-of-fit test for the multiple logistic regression model*" Communications in Statistics A10:1043- 1069.

[14] Isaac O. Ajao1; Adebisi A. Ogunde(2017)"*Measuring Deviance Goodness-of-Fit For Poisson Regression Model with Application to Count Data in Medicine*" Department of Mathematics and Statistics,The Federal Polytechnic, Ado-Ekiti,Ado-Ekiti, Nigeria. Edited Proceedings of 1st International Conference Vol. 1, 2017

[15]Jason Brownlee (2019)"*Probabilistic Model Selection with AIC, BIC, and MDL*"Last Updated on August 28, 2020
[16] Kaya, Z... (2020)" *Mapping Kurdistan Territory, Self-Determination and Nationalism".* Cambridge University Press. pp. 2, 137, 177, 197.
[17] Maria. T. O'Shea.(2004)" *Trapped between the map and reality: geography and perceptions of Kurdistan"* 258pp., Routledge, (see p. 77)
[18]McHugh ML(2009) "*The odds ratio: calculation, usage, and interpretation"Biochem* http://dx.doi.org/10.11613/BM.2009.011. [Google Scholar]
[19] Namo Kurdistani (2018)"*Iran review. 25th anniversary of Sardasht Chemical Bombardment*" Web.http://www.iranreview.org/content/Documents/25th-Anniversary-of-Sardasht-Chemical-Attack.htm
[20] Omar Lhafdoozi.(2006*)" 28th June Sardasht Chemical Bombardment*" UN Report 2006. p153
[21]Pearson, K. (1900)." *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*" (PDF). Philosophical Magazine. Series 5. 50 (302)
[22]Sandro Sperandei(2014)"*Understanding logistic regression analysis* "Biochem Med (Zagreb). Feb; 24(1): 12–18Published online 2014 Feb 15. doi: 10.11613/BM.2014.003
[23] Schafer J.L. (2006). "*Multinomial logistic regression models*". STAT 544-ecture19

[24]VilleHallikainen, Mikko Kalle Tapani Hyppönen, Juha Olavi Hyvönen (2007)"*Establishment and Height Development of Harvested and Naturally Regenerated Scots Pine near the Timberline in North- East Finnish Lapland"* 41(1) Follow journal DOI:10.14214/sf.308 License CC BY-SA

The photos reflect the conscience of the oppressors of the Ba'ath regime in their oppression in Kurdistan.

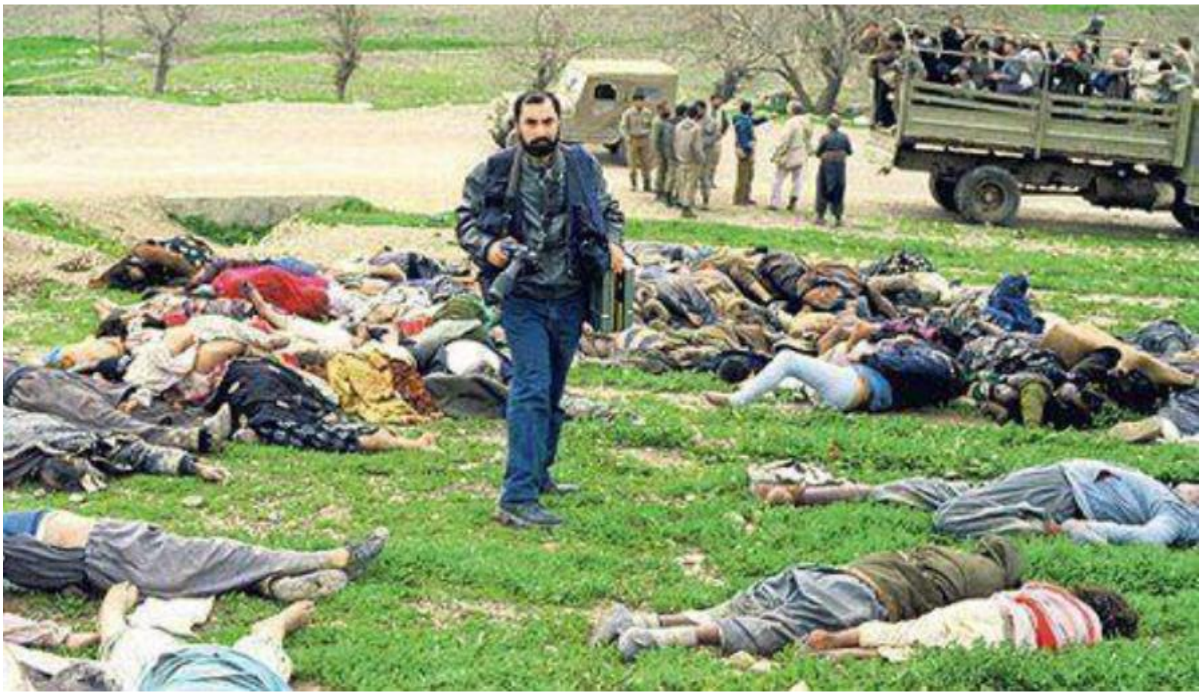
Hallabja during Chemical Attack

Grave yard of Halabja Victims in Halabja, Kurdistan

Mass martyrs' cemetery contains 1500 martyrs

Martyrs are a father with two children in a house of Halabja chemical attack

Martyr (Omar Khawar) and his baby in Halabja chemical attack



The martyrs of the chemical attack were martyred en masse on the road between Halabja and Iran

خەملاندنی پارامێتەرەکان و پێشبینی کردنی مەترسییەکانی لەبارچوون و نەزۆکی و نەخۆشی دەروونی لە نێوان ڕزگاربووانی هێرشی کیمیایی لە کوردستان بە بەکارهێنانی شیکاری پاشەکشەی لۆجستیکی فرەژمارەیی

نەزیرە سەدیق کەریم بەرزنجی

بەشی ئامار و زانیارییەکان، کۆلێژی بەڕێوەبردن و ئابووری، زانکۆی سەلاحەددین-هەولێر

nazeera.kareem@su.edu.krd

پوختە

لەو کاتەی ڕژێمی بەعس دەستی بەسەر عێراقدا گرتووە، بە هەموو شێوەیەک ستەمیان لە نەتەوەی کورد کردووە. ساڵی 1987 بە بیانووی شەڕی ئێران و عێراق هەردوو شاری (سەردەشت و پیران شار)یان لە ڕۆژهەڵاتی کوردستان بۆردومان کرد دوای ئەو ساڵە 40 گوند و شار لە باشووری کوردستان بۆردومان کران لەوانەش هەڵەبجە بوو لە 16/3/ . ١٩٨٨. جگە لە پێنج هەزار شەهید و 10 هەزار بریندار، زۆربەی خەڵکی هەڵەبجە بەدەست نەخۆشی جۆراوجۆرەوە دەناڵێنن، بۆ دیاریکردنی ئاستی مەترسییەکانی غازی کیمیایی بۆ ئەو کەسانەی کە بەدەست لەبارچوون و نەزۆکی و تێکچوونی دەروونییەوە دەناڵێنن، ئەم توێژینەوەیە بە چاوپێکەوتن لەگەڵ (500) تووشبوو لەوکەسانەی لە وەزارەتی شەهیدان و ئەنفال لە کوردستان تۆمارکراون.

لە ئەنجامەکاندا دەرکەوتووە کە زۆرترین مەترسی لەسەر [نەخۆشی لەباربردن، نەخۆشی دەروونی (مێ) ،نەخۆشی نەزۆکی(نێر)، و گواستنەوەی نەخۆشی]، و ئاستی (0.70-0.90) زۆرترین کاریگەری لەسەرە بەهۆی ئەو گازە کیمیاییانەی کە دەبنە هۆی لەباربردنی منداڵ، نەزۆکی، و نەخۆشییە دەروونییەکان.لەم توێژینەوەیەدا نەرمەکاڵای IBM SPSS کە بۆ شیکردنەوەی داتاکان بەکارهاتووە.

وشەی سەرەکییەکان: پاشەکشەی لۆجستیکی فرەژمارەیی، خەملاندنی پارامێتەرەکان (ئاماری والد، ئاماریپۆلێنکردن) ، گازی کیمیایی، کوردستان .

تقدير المعلمات والتنبؤ بمخاطر الإجهاض والعقم والاضطرابات النفسية بين الناجيات من الهجمات الكيميائية في كردستان باستخدام تحليل الانحدار اللوجستي متعدد الحدود

نظيرة صادق كريم البرزنجي

قسم الإحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة صلاح الدين-اربيل

nazeera.kareem@su.edu.kr

ملخص

منذ أن استولى نظام البعث على العراق، اضطهد الأمة الكردية بكل الطرق. وفي عام 1987، بحجة الحرب العراقية الإيرانية، قاموا بقصف مدينتي (سردشت وبيران شار) في شرق كردستان. وبعد ذلك العام، تم قصف 40 قرية ومدينة في جنوب كردستان، منها حلبجة بتاريخ 3/16/ 1988. بالإضافة إلى 5000 شهيد و10000 جريح، يعاني معظم أهالي حلبجة من أمراض مختلفة، ولتحديد مستوى مخاطر الغاز الكيماوي على الناجيات اللاتي يعانين من الإجهاض والعقم والاضطرابات النفسية، تم تسجيل هذه الدراسة من خلال مقابلة (500) مصابة من الأشخاص المسجلون في وزارة الشهداء والمؤنفلين في كردستان.

توصلت النتائج إلى أن الأكثر خطورة على [مرض الإجهاض، المرض النفسي (الأنثوي)، مرض العقم (الذكر)، وانتقال المرض]،وبمستوى (0.70-0.90) هو الأكثر تأثراً بالغازات الكيميائية المسببة للإجهاض، - العقم، والأمراض النفسية وفي هذا البحث تم استخدام برنامج IBM SPSS لتحليل البيانات.

الكلمات المفتاحية: الانحدار اللوجستي متعدد الحدود ، تقديرات المعامِلات إحصائية والد، تصنيف ، غاز كيميائي ، كردستان.