A Multinomial Logistic Regression Model for Analyzing Attitudes towards Political Activities: A Case Study in Erbil/ Kurdistan Region of Iraq

Khwazbeen Saida Fatah

Department of Mathematics, College of Science, University of Salahaddin, Kurdistan Region, Iraq, khwazbeen@yahoo.com, khwazbeen@uni-sci.org.

Abstract:

Regression models are considered as the most commonly used statistical analysis techniques to describe the functional relationships between a dependent variable (either continuous or categorical) and a set of independent variables based on samples from a particular population. In this paper, a Multinomial Logistic Regression Model is proposed to investigate the variations of multi ethnic-religious people towards political attitudes. This model is applied specifically to a case study conducted in Erbil in Kurdistan Region of Iraq, where multi-ethnic and religious groups live in. Results for statistical analysis show that the variations of ethnicity or religion have no much effect on the political attitudes for the majority of citizens of this region.

Keywords: Regression Analysis, Logistic Regression, Multinomial Logistic Regression, Survey, Analysis of Political Attitudes.

1. Introduction

In practical life, there are several situations where it is necessary to predict the average value of one random variable Y (a dependent or response variable, either categorical or continuous) given a specific value of another variable X (independent or predictor variable). Regression models may be one of the most commonly used statistical analysis techniques for investigating and modelling the relationships between a dependent variable and one or more independent variables based on samples from a particular population in which the dependent variable represents the phenomenon that is needed to be investigated. Therefore, the purpose of the regression analysis is to evaluate the relative impact of a predictor variable (or more than one variable) on a particular outcome, or the regression methodology models the distribution of a variable, called response, with the help of one or more predictor variables. If only one predictor is studied for its relationship with the response variable, then it is a simple regression analysis. In multiple regressions more than one predictor is studied for their relationship with the response variable (Wani, 1971; Wackerly et al, 2002). Applications of regression analysis are numerous including fields of engineering, physical and chemical sciences, managements, life and biological sciences and social sciences.

However, when the response variable is a qualitative variable (dichotomic or polychotomic), as for example, in political science, it is not possible to use regression methodology to analyse whether a person will or will not participate in political activities, which is considered as a categorical variable. There is a good reason that linear regression models cannot be applied directly to categorical response variable; one of the assumptions of linear regression is that the relationship between variables is linear, but when the outcome variable is categorical, this assumption is violated. Another assumption is that Y follows a normal distribution and therefore should be between $(-\infty, +\infty)$ and hence the result is not logical when the response variable is not continuous. Therefore, the model is transformed into logistic regression based on logistic function and then the data is transformed using logarithm transformation. This transformation is a way of expressing a non linear relationship in a linear way (Berry and Feldman, 1985; Berry, 1993; Field, 2009). It expresses the multiple regression equation in logarithmic term (called the logit) and thus overcomes the problem of violating the assumption of linearity. Such statistical models are used to establish the best fit for explaining the relationship between a dependent variable Y, with either two categories, which is represented by 0 or 1 (called dichotomic), or with more than two categories, and a set of independent predictors, $X_1, X_2, \ldots, X_m$. When the variable Y is dichotomic, the model is referred to as Logistic Regression (LR) and when the response variable Y is qualitative with more than two categories or it is polychotomic, it is called Multinomial Logistic Regression (MLR).

Cornfield et al. (1961) is the first to use Logistic Regression Model (LRM). In 1967, Walter and Duncan used this methodology to estimate the probability of the occurrence of a process as a function of other variables. The use of LRM increased during the 1980s, and at present it constitutes one of the most widely used methods in research studies for various fields of life studies such as Health Sciences, Political and Sociological Sciences (Colin and Robert, 1984; Allenby and Leuk, 1994; Balabanis and Vassileoiu, 1999; Henn et al., 2002; Hosme and Lemeshow, 2000; Tabachnick et al., 2001; Li and Marsh, 2008; Al-Habib, 2012).

In this study, the main objective is to explore the usage of MLR models in analysing attitudes of people in Kurdistan Region of Iraq towards political participation such as in voting, party membership and other mainstream of political activities. The issue of political participation has received a great deal of attention in recent years from academic and politicians across most liberal democratic countries. This reflects concerns about the decline of participation in political activities (Li and Marsh, 2008). In this context, unsurprising, in Kurdistan Region of Iraq, despite the efforts for consolidation of the main principles of democracy, there has been no attention paid to the scientific studies in addition to investigations and statistical analysis for such investigations that will be used to predict future behaviour of Kurdistan Region population, which is known to be a multi-ethnic and religious population. Thus, the study has two main objectives, first, in order to investigate the behaviour of multi ethnic and religious people living in Kurdistan Region of Iraq towards participation in political activities, a survey was conducted. The survey covered areas of Erbil city where their citizens are from the main ethnic groups, Kurdish, Turkmen and Chaldean-Assyrian and belong to three main religions, Muslims, Christian and Yazidi. The second objective is to fit a MLR model for analysing the political attitudes for people living in Erbil and then tests the significance of the analysis.

This paper is organized as follows: following an introduction to regression analysis and regression models in section 1, construction of the Logistic Regression (LR) and Multinomial Logistic Regression (MLR) are presented in section 2 with the definition of Logit functions and Odds and Odds Ratios. In section 3, the sample survey and the proposed model for the study are described. In section 4, data analysis for the implementation of proposed model to the case study is displayed for assessing the proposed model, the Log-Likelihood Statistic is introduced in section 5. Finally, in section 6, the main conclusion and recommendations are stated.

2. Construction of the Logistic Regression Models

The simple and multiple linear regression methods are used to model the relationship between a quantitative response variable and one or more explanatory variables. In simple linear regression, the mean of the response variable is modelled as a linear function of the explanatory variable

$$E(Y/X) = \alpha + \beta X. \qquad (1)$$

When the dependent variable Y is categorical with more than one category, it is also possible to relate $P(Y)$ and X through a linear relationship $P(Y) = \alpha + \beta X$. Unfortunately, this always is not a good model. As long as $\beta \neq 0$, extreme values of X will give values of $\alpha + \beta X$ that are inconsistent with the fact that $0 \leq P(Y) \leq 1$.

When Y is a dichotomic dependent variable with response 0 when the event does not occur and response 1 when the event does occur while the predictor variable X is either qualitative or quantitative, it is inappropriate to assume that they are normally distributed and thus the data cannot be analysed using regression methods described so far. In order to avoid this problem, it is common to define a binding function between $(-\infty, +\infty)$ where a normal distribution for the dependent variable $P(Y)$ is obtained.

In case when the response variable has only two outcomes, success or failure, if their values are assumed to be 1 and 0 respectively, then the mean is the proportion of 1, or it is the probability of success, i.e. $P = \text{Prob}(Y = 1)$. With n-independent observations, the distribution for number of successes is Binomial. Therefore, the interest is to study how $P(\text{success})$ depends on an explanatory variable X , where it can be either categorical or quantitative; LR is the most common statistical method for describing such relationships (Thompson, 2008).

Thus, the linear regression defined by (1) is transformed in to another expression:

$$P(Y) = \alpha + \beta X \qquad (2)$$

To construct LR model, which is based on Logistic Function, the mathematical form of logistic function is described. The logistic function f(y), is described as

$$f(y) = \frac{1}{1+e^{-y}}, -\infty < y < +\infty. \text{ where } 0 < f(y) < 1. \qquad (3)$$

The fact that the logistic function f(y) ranges between 0 and 1 is the primary reason for its popularity in modelling probabilities. The model is designed to describe a probability, which is always some number between 0 and 1. Thus, for the logistic model, an estimate is obtained which can never be above 1 or below 0. While this is not always true for other possible models, which is why the logistic model is often the first choice when a probability is to be estimated.

Therefore, in order to obtain a logistic model from the logistic function $f(Y)$ (or $P(Y)$) , where Y is described as $Y = \alpha + \beta X$, α, β are constant terms representing unknown parameters with only one predictor variable .

Then, it follows that

$$f(y) = \frac{1}{1+e^{-(\alpha+\beta X)}} \qquad (4)$$

Where $f(y)$ is the probability of Y occurring, e is the base of natural logarithms, and the other coefficients form a linear combination similar to that in simple regression. Therefore, the terms α

and β in this model represent unknown parameters that are needed to be estimated based on data obtained from the sample observations.

To define LR model in terms of probability of an event occurring, the probability being modelled in (4) can be denoted by the conditional probability statement; it is described by

$$P(event\ y\ ocuuring/X) = \frac{1}{1+e^{-(\alpha+\beta X)}} \quad (5)$$

Hence, in LR the probability that a given individual will fall into one outcome group or the other is estimated. The difficulty with LR analysis is to develop an equation for predictor variables that classifies objects into groups with the greatest accuracy and the real focus is on some function of the proportion of the targeted outcome P , the outcome is coded 1 when it is occurring and 0 when it is not occurring and P values ranges between 0 and 1.

Just like linear regression, it is possible to extend this model to include more than one predictor. Hence, the equation in (5) becomes

$$P(Y/X_1, X_2, \dots, X_n) = \frac{1}{1+e^{-(\alpha+\beta_1 X_1+\beta_2 X_2+\dots+\beta_n X_n)}} \quad (6)$$

β_is are the coefficients associated to the independent variables X_is.

2.1 The Odds

When a binary outcome variable is modelled using LR, it is assumed that the logit transformation of the outcome variable has a linear relationship with the predictor variables. This makes the interpretation of the regression coefficients somewhat tricky. Therefore, LR models use the concept of odds and odds ratios and try to interpret the logistic regression results using odds rather than proportions as a measure of the probability.

The odds are defined as the ratio of the probability of success coded 1 over the probability of failure coded 0. If P is the probability of an event occurring (targeted outcome or success), then $(1 - P)$ will be the probability of that event not occurring, or it is the ratio of the number of occurrences to the number of non occurrences. The odds of an event can be defined as:

$$\text{Odds} = \frac{\text{P(event occurring)}}{\text{P(event not occurring)}} \quad , \quad \text{or}$$

$\text{Odds} = \text{P}/\ (1 - \text{P})$, where

$\text{P(event occurring)} = \alpha + \beta X$,

$\text{P(event not occurring)} = 1 - \text{P(event occurring)} = 1 - (\alpha + \beta X)$

When Odds = 1 it indicates equal probabilities of occurrence and non occurrence (0.5).

When Odds < 1, it indicates that occurrence is less likely than non occurrence.

An Odds >1 indicates that occurrence is more likely than non occurrence.

In LR analysis, it is important to distinguish probabilities from odds not only for accuracy in reporting findings, but also for the interpretation of the LR coefficients. When findings are explained as odds, they can be converted to probabilities using the following:

$$Probability = \frac{Odds}{1 - Odds}$$

This transformation from probability to odds is a monotonic transformation, meaning the odds increase as the probability increases or vice versa. Probability ranges from 0 and 1 but Odds range from 0 and positive infinity. To overcome this difficulty, a transformation of the dependent variable will make LR reasonable even if data include or not include extraordinary odds.

Logistic models use natural log (ln) of the odds as the focus of analysis; it is called logits, which is an alternative way to write the logistic model. To get the logit from the logistic model, a transformation of the model is made. The basic idea behind logits is to use a logarithmic function to restrict the probability values to (0,1)

Thus, the log-odds is modelled as a linear function of the explanatory variable where

$$\text{logit}(P(Y)) = \ln\left(P / (1 - P)\right) = \ln\left\{\frac{\frac{1}{1+e^{-(\alpha+\beta X)}}}{1-\frac{1}{1+e^{-(\alpha+\beta X)}}}\right\} = \ln\left\{ e^{(\alpha+\beta X)}\right\} = \alpha + \beta X$$

where α indicate the intercept while represents the regression coefficient, the change in the logarithm of the odds of the event with a 1-unit change in the predictor X . This implies that a one unit-change in the predictor variable X brings about β change in logit(p).

Therefore, the general LR model with one predictor is defined as

$$\text{logit}\left(P(Y)\right) = \alpha + \beta X \qquad (6)$$

when the probability of success is transferred using the logit transformation, a quantity logit(P) is obtained, which is linear in the explanatory variable .

Although this model looks similar to a simple linear regression model, the underlying distribution is Binomial and the parameters α, β cannot be estimated in exactly the same way as for simple linear regression. Instead, the parameters are usually estimated using the Method of Maximum Likelihood (Hosmer and Lemeshow, 2000).

2.2 The Odds Ratio

An alternative probability measure called the Odds Ratio is used in LR analysis; it is a popular measure of the strength of association between success event and failure which compares the odds of an event occurring to the odds of that event not occurring.

Thus, $\text{Odds Ratio} = \frac{\text{Odds(event occurring)}}{\text{Odds(event not occurring)}}$

or $\text{Odds Ratio} = \{P(Y = 1) / (1 - P(Y = 1))\} / \{P(Y = 0) / (1 - P(Y = 0))\}$

Since the Odds of success $= \frac{P}{1-P}$ $= e^{(\alpha+\beta X)} = e^{\alpha} e^{\beta X}$, or

p = probability of success = $\frac{e^{(\alpha+\beta X)}}{1+e^{(\alpha+\beta X)}}$

Because the explanatory variable X increases by one unit from X to X + 1, the odds of success change from $e^{\alpha}e^{\beta X}$ to $e^{\alpha}e^{\beta(X+1)}$ the odds ratio (OR) is therefore $e^{\alpha}e^{\beta X}e^{\beta} / e^{\alpha}e^{\beta X} = e^{\beta}$. The odds ratio e^{β} has a simpler interpretation in the case of a categorical explanatory variable with two categories; in this case it is just the odds ratio for one category compared with the other.

Therefore, for the logit(P(Y)), which is the logarithm of the corresponding odds, if an variable has a coefficient β, then a unit increase in X increases the log odds by an amount equal to β. This means that the odds themselves are increased by a factor of e^{β}, hence the logarithms of all the odds can be converted back to probability by the formula $P = \frac{e^{Y}}{1+e^{Y}}$.

Thus, Odds Ratio is an indicator of the change in odds resulting from a unit change in the predictor. It represents the constant effect of a predictor on the likelihood that one outcome will occur. Since the effect of the predictor(s) on the probability of the response has different values depending on the value of the predictor, the odds ratio is a single summary score of the effects.

Since β performs the same function as regression coefficient in linear regression in that they indicates the direction and the strength of the relationship between the

independent and dependent variables Therefore:
when $\beta = 0\ (e^{\beta} = 1) \rightarrow \Pr(\text{success})$ is the same as at each value of x.
$\beta > 0\ (e^{\beta} > 1) \rightarrow \Pr(\text{success})$ increase as x increases.
$\beta < 0\ (e^{\beta} < 1) \rightarrow \Pr(\text{success})$ decreases as x increases.

Note that the odds ratio is sometimes called the odds multiplier, because it is the value the odds are multiplied by, when the value of the explanatory variable is increased by one unit.

When n independent predictor variables X_is, or factors, (equation (6)) are involved in deciding the eventual outcome, the odds ratio for each one is calculated separately. The joint effect of all the independent variables put together may be expressed mathematically as

$$\text{Odds} = P_i / (1 - P_i)\ , i = 1,2,\dots,n \qquad (7)$$

Similarly, logarithmic transformation is applied and then the linear relationship between the response variable with two categories and n independent predictors is defined by

$$\text{logit}(P_i) = \ln\left(P_i / (1 - P_i)\right) = \alpha + \sum_{j=1}^{n} \beta_j X_{i,j} \qquad (8)$$

The model described by (8) is called the Binomial LR model; the essential assumptions of logistic regression are independence between the successive observations and the existence of a linear relationship between logit(P(Y)) and the predictors $X_1, X_2, \dots, X_n$.

The Multinomial Logistic Regression (MLR) model is a simple extension of the Binomial LR model; most of multivariate analysis techniques require the basic assumptions of normality and continuous data, involving independent and/or dependent variables. MLR is used when these assumptions tend to be violated.

Thus, MLR model can be understood as a simple extension of LR that allows each category of an unordered response variable to be compared to an arbitrary reference category providing a number of logit regression models. A binary LR model compares one dichotomy (for example, passed-failed, Yes, No, etc.) whereas the MLR model compares a number of dichotomies. This procedure outputs a number of logistic regression models that make specific comparisons of the response categories. This means that there is a variable for all categories except one, so if there are M categories, there will be M-1 dummy variables. All but one category has its own dummy variable. Each category's dummy variable has a value of 1 for its category and a 0 for all others. One category, the reference category, does not need its own dummy variable, as it is uniquely identified by all the other variables being 0. Therefore, MLR can estimate a separate binary LR model for each of those dummy variables. The result is M-1 binary LR models. When there are M categories of the response variable, the model consists of (M -1) logit equations which are fitted simultaneously.

The most significant factor to be considered here is that each one tells the effect of the predictors on the probability of success in that category, in comparison to the reference category. This means that each model has its own intercept and regression coefficients; it explains the effect of each category differently.

The general MLR model is described below:

$$logit\left\{\frac{P(Y=i)}{P(Y=j)}\right\} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \qquad (9)$$

where i is identified category and j is the reference category.

The models described by (9) are the proposed models of choice behaviour for M categories of the response variable using (M-1) logit models. These models provide (M-1) estimates for the effect that each predictor has on the response. This means that it provides the effect of X_1 on the choice between categories 1,2 and the effect of X_1 on the choice between categories 1,3 and so on, and the effect of X_1 across all categories. Similarly, it provides the effect of X_2 through all logit models.

3. Sample survey and the proposed model

In this study, to investigate and model the political attitudes of multi-ethnic and religious people in Kurdistan Region of Iraq, a sample survey was conducted. A questionnaire was designed to identify participants from both genders (male and female) for different ages with their religions (Muslim, Christian or Yazidi) and ethnics (Kurdish, Turkmen or Chaldean Assyrian) and then to express to which extent they have interest in practising political principles. The respondent targeted were citizens from different ethnics and religions living specifically in Erbil city; they were selected randomly from different institutions in public and private sectors and various residential areas. In this study, 500 questionnaires were distributed; only a list of 411 (82.2%) completed questionnaires were received. In the survey, all participants were asked “Do you participate in the political activities” with only three options; “Yes”, “No” and “To Some Extent”.

Hence, a MLR model is specifically designed to investigate the relationship between a response variable representing the main question “Do you participate in Political Activities”, which has three unordered category answers, and a set of predictors represented by two categorical variables, ethnicity and religion. In order to model which of three responses (there are three categories in the unordered response variable) is likely to be chosen by a respondent, two logit models are constructed; one comparing category “No” with the reference category “Yes” and one comparing category “To Some Extent” with the reference category “Yes”. The two predictors are X_1, X_2 representing ethnic and religion respectively.

The general MLR model of choice behaviour between three response categories and the two predictors can therefore be represented using the following logit models.

$$logit\left\{\frac{P(Y=i)}{P(Y=j)}\right\} = \alpha + \beta_1 * X_1 + \beta_2 * X_2 \quad (10)$$

Where $logit$ refers to the log odds of a person choosing one response ($i = "No"\ or\ i = "To\ Some\ Extent"$) compared to another response (the reference category ($j = "Yes"$); a choice that is influenced by the independent variables.

X_1 represents the ethnic and X_2 represents the religion. α, β_1, β_2 are the main parameters for the MLR models. Each parameter indicates the effect of a predictor on the likelihood that one outcome will occur.

The MLR models described by (10) provide two estimates for the effect that each explanatory variable has on the response category “Yes”, which is selected as a reference category, and then provides estimates of the overall significance.

From (10) two logits model are obtained; they are described below

$$logit\left\{\frac{P(Y="No")}{P(Y="Yes")}\right\} = \alpha + \beta_1 * \text{ethnic} + \beta_2 * \text{religion} \quad (11a)$$

$$logit\left\{\frac{P(Y="To\ Some\ Extent")}{P(Y="Yes")}\right\} = \alpha + \beta_1 * \text{ethnic} + \beta_2 * \text{religion} \quad (11b)$$

To apply this model and demonstrate its technique, the data obtained from the analysis of the respondent questionnaire is used. Below is the statistical analysis for the data obtained from the conducted survey.

4. Data Analysis and Results

In this study, the logistic regression model is describing the relationship between the categorical dependent variable which is the main question: "Do you participate in Political Activities" with only three different category answers, "Yes", "To Some Extent" and "No" and two categorical predictors. The first variable is the Religion with three categories (Muslim, Christian and Yazidi), the second one is the National Ethnicity which consists of three categories; Kurdish, Turkmen and Chaldean-Assyrian. It is to be noted that within this population there are people with other ethnics and religions but they are minorities and their participation in this survey were very limited, thus they were excluded from this survey. The following tables display the data analysis for the sample survey and the implementation of the proposed model.

Table 1 shows the frequency table for the categorical variable "Do you participate in Political Activities" with both predictors, the Religion and the Ethnicity, for all 411 respondents to this questionnaire. The data show that the majority of the population (77.9%) do participate in political activities but (7.5%) do not participate while (14.6%) they do participate in some way. This is a good indicator that the population of this region are interested in participation in political life.

Table 1- Summary for Responses and the Two Predictor Variables

		N	*Percentage*
Do you participate in the political activities?	Yes	320	77.9%
	To Some Extent	60	14.6%
	No	31	7.5%
Religion	Muslim	314	76.4%
	Christian	63	15.3%
	Yazidi	34	8.3%
National Ethnicity	Kurdish	295	71.8%
	Turkmen	55	13.4%
	Chaldean_Assyrian	61	14.8%
Valid		411	100.0%
Total		411	

Table 2 displays the cross tabulation of the two predictors Religion against National Ethnicity for all respondents; they were as follows: 295 (71.8%) respondents were from Kurdish ethnicity of which 261 (83.1%) were Muslims and 34 (100%) Yazidi. Turkmen were 55 (13.4%) participants of which 53 (16.9%) were Muslims and 2 (3.2%) were Christian. Finally, Chaldean Assyrian constitutes (14.8%) of which (96.8%) were Chaldean Assyrian from Chaldean Assyria Christian and the other (3.2%) were Turkmen. Respondents who declined to participate, in the survey, stated reasons such as insufficient experiences in conducting surveys or lack of interest. The following table, Table 2, displays the data representing the percentage for participants with different ethnics and religions. The data analysis shows that the majority of Kurdish are Muslims or Yazidi but all Chaldean_Assyrian are Christian while majority of Turkmen are Muslims.

Table 2: Religion * National Ethnicity Cross tabulation

Religion		National Ethnicity			Total
		Kurdish	Turkmen	Chaldean Assyrian	
Muslim	Count	261	53	0	314
	% within Religion	83.1%	16.9%	0.0%	100.0%
Christian	Count	0	2	61	63
	% within Religion	0.0%	3.2%	96.8%	100.0%
Yazidi	Count	34	0	0	34
	% within Religion	100.0%	0.0%	0.0%	100.0%
Total	Count	295	55	61	411
	% within Religion	71.8%	13.4%	14.8%	100.0%

In Table 3 the percentage for responses to the main question in the questionnaire "Do you participate in Political Activities?" for each ethnic and religion groups are displayed. For example, Kurdish Muslims 80.8% say Yes with only 6.1% No while 13% decide on To Some Extent. In the case of Kurdish Yazidi and Turkmen Muslims are almost similar. For Chaldean Assyrian, 65% selects Yes and 21.3% selects to Some Extent but 13.1% says No.

Table 3: Participation in Political Activities for different Ethnics and Religions

Ethnicity	Religion	Do you participate in Political Activities?					
		Yes		To Some Extent		No	
		Count	Row N %	Count	Row N %	Count	Row N %
Kurdish	Muslim	211	80.8%	34	13.0%	16	6.1%
	Christian	0	0.0%	0	0.0%	0	0.0%
	Yazidi	26	76.5%	4	11.8%	4	11.8%
Turkmen	Muslim	42	79.2%	8	15.1%	3	5.7%
	Christian	1	50.0%	1	50.0%	0	0.0%
	Yazidi	0	0.0%	0	0.0%	0	0.0%
Chaldean_Assyrian	Muslim	0	0.0%	0	0.0%	0	0.0%
	Christian	40	65.6%	13	21.3%	8	13.1%
	Yazidi	0	0.0%	0	0.0%	0	0.0%

Thus, the data analysis provides good evidences that, Kurdistan Region population with different religions and ethnics are interested in practising political activities.

To apply the MLR model proposed for this study, a set of respondents (n=411) with a categorical response variable (Y) with 3 categories taking on values 0,1 and 2 and two predictor variables X_1 representing Ethnicity and X_2 representing the Religion are considered.

Table 4 shows parameter estimates for logit model, the statistical software SPSS is used.

Table 4: Parameter Estimates for MLR Logit Model

Participation in the Political Activities?[a]		B	Std. Error	Sig.	Exp(B)
To Some Extent	Intercept	-2.828	1.604	0.078	
	Kurdish	0.957	1.512	0.527	2.603
	Turkmen	1.124	1.450	0.438	3.077
	Chaldean Assyrian	0[b]	.	.	.
	Muslim	0.046	0.568	0.935	1.047
	Christian	1.705	1.572	0.278	5.499
	Yazidi	0[b]	.	.	.
No	Intercept	13.370	0.387	0.000	
	Kurdish	-15.242	0.662	0.000	2.402E-7
	Turkmen	-15.301	0.929	0.000	2.263E-7
	Chaldean Assyrian	0[b]	.	.	.
	Muslim	-0.707	0.596	0.236	0.493
	Christian	-14.979	0.000	.	3.123E-7
	Yazidi	0[b]	.	.	.

a. The reference category is: Yes.
b. This parameter is set to zero because it is redundant.

Remark: The reference category is selected to be “Yes” category, the selection for each of Christian and Yazidi are mostly set to be “Yes”, therefore these variables are shown to be redundant.

In Table 5, values for β indicate the direction and the strength of the relationship between the independent and dependent variables and e^{β} represents the change in the odds of the outcome by increasing dependent by 1 unit –odds ratio. The model parameters shown in the table are explained as follows: for example, holding religion constant, the odds for someone who is Kurdish of selecting the response "To Some Extent” rather than “Yes” is 2.603 times (160.3% higher than) the log odds for someone who is Chaldean Assyrian or Yazidi. This explains that there is a positive relation between the response and the predictor variable and the probability that Y equals an event “To Some Extent” is twice as likely (2.603 times) as the value of X is increased one unit.

Holding ethnicity constant, the odds for someone who is Muslim and replies “No” rather than “Yes” is decreased by .707 this equates to an odd ratio $e^{(-.707)} = .493$. This means that the odds of selecting non reference categories rather than reference category is higher than for someone Christian or Yazidi. Muslims are therefore much more likely to choose “Yes” rather than “No” after controlling ethnicity. Kurdish are therefore much more likely to select “Yes” than “No” after controlling Religion. The same result is obtained in the case when Christians are considered.

Table 5: Parameter Estimates

Participation in the Political Activities?		B	Exp(B)
To Some Extent	Intercept	-2.828	
	Kurdish	0.957	2.603
	Turkmen	1.124	3.077
	Chaldean_Assyrian	0	.
	Muslim	0.046	1.047
	Christian	1.705	5.499
	Yazidi	0	.
No	Intercept	13.370	
	Kurdish	-15.242	2.402E-7
	Turkmen	-15.301	2.263E-7
	Chaldean_Assyrian	0	.
	Muslim	-.707	0.493
	Christian	-14.979	3.123E-7
	Yazidi	0	.

Similarly, parameters between non-reference categories of the dependent variable with non-reference categories are explained in which log odds is defined by:

logit $\left(\frac{{P_i}^j}{{P_i}^1}\right) = (\alpha^{(1)} - \alpha^{(j)}) + (\beta^{(j)}{}_1 - \beta^{(1)}{}_1)X_1 + (\beta^j{}_2 - \beta^1{}_2)X_2$ where α^j, $\beta^j{}_1, \beta^j{}_2$ are unknown parameters for the population $j = 2,3,$

Hence, holding Religion constant, the odds for someone who is Kurdish responding to his participation in political life “To Some Extent” rather than “Yes” is 0.846 times the odds for someone who is Turkish. It is calculated as exp(.957-1.124) = exp(-.167) = 2.603/3.077 = .846. Hence 15.4% lower.

Similarly, if another reference category other than “Yes” id selected then the same procedure is followed for analysing data.

5. Assessing the Model: The Log-Likelihood Statistic

Likewise multiple regression, in LR the observed and predicted values can be used to assess the fit of the model. The measure statistic used is called the Log-Likelihood (LL). It is based on summing the probabilities associated with the predicted and actual outcomes (Field, 2009).

$$\text{Log-Likelihood(LL)} = \sum_{i=1}^{n}[Y_i * \ln(P(Y_i) + (1 - Y_i) * \ln(1 - P(Y_i))] \quad (12)$$

The LL defined by (12) is analogous to the residual sum of squares in multiple regressions in the sense that it is an indicator of how much unexplained information is after the model has been fitted.

It, therefore, follows that the large values of the LL statistic indicate poorly fitted model, because the larger the values the more unexplained observations there are. Then Chi Square statistic is used which is define as

$$\chi^2 = 2[LL\ (new) - LL(baseline)] \qquad (13)$$

$LL\ (new)$ indicates the log likelihood for the fitted model and $LL(baseline)$ refers to the model that gives the best prediction when there are values other than the values of the outcomes or the model when only the constant is included. When one or more predictors are added to the model then there will be an improvement (new model). The degrees of freedom (df) for χ^2 will be df for the new model – df of the model with only constant.

As in linear regression, to know not only how well the model fits the data, but also the individual contribution of predictors, the estimated regression coefficient β and their standard errors are used to compute t-statistic. In LR there is analogous statistic known as the Wald statistic, which has special distribution known as the Chi-Square distribution. It is defined as follows

Wald = β / SE(β), SE denotes the standard error for β.

Wald statistic tests whether the β coefficient for that predictor is significantly differ from zero. If the coefficient is significantly differs from zero then it can be assumed that the predictor is making a significant contribution to the prediction of the out come Y_i . Wald statistic is identical to t-test; it is usually used to ascertain whether a variable is a significant predictor of the outcome.

As for LR, a MLR model can assess the effect of individual or groups of predictor variables on the response by comparing the Deviance statistics (-2LL) for two nested models The resulting statistic is tested for significance using the chi-square distribution with the number of degrees-of-freedom equal to the difference in the number of terms between the two models.

In this study, SPSS is used for assessing the proposed model. Table 6 show that -2LL for models, the null model and the final model. The LL is a measure of how much unexplained variability there is in the data. The difference in LL is (41.266-32.060 = 9.206). This change is significant, which means that the final model explains a significant amount of the original variability, in other words it is a better fit than the null model.

Table 6: Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	41.266			
Final	32.060	9.206	8	.325

In Table 7, the individual parameter estimates are shown for response categories of the dependent variable "Do you Participate in Political Activities" comparing the two responses "To Some Extent" and category "No", taking reference category "Yes". It explains that responses "To Some Extent" with "Yes" and "No" with "Yes" appear to be similar as both predictors are significant except one category which is Muslim.

Table 7: Parameter Estimates

Do you participate in the Political Activities? a		B	Std. Error	Wald	df	Sig.
To Some Extent	Intercept	-13.370	1.619	68.189	1	0.000
	Kurdish	15.242	1.683	82.002	1	0.000
	Turkmen	15.301	1.789	73.161	1	0.000
	Chaldean_Assyrian	0^b	.	.	0	.
	Muslim	0.707	0.596	1.407	1	0.236
	Christian	14.979	1.572	90.789	1	0.000
	Yazidi	0^b	.	.	0	.
No	Intercept	-16.198	0.449	1299.428	1	0.000
	Kurdish	16.198	0.838	373.808	1	0.000
	Turkmen	16.425	1.119	215.460	1	0.000
	Chaldean_Assyrian	0^b	.	.	0	.
	Muslim	0.754	0.769	0.960	1	0.327
	Christian	16.684	0.000	.	1	.
	Yazidi	0^b	.	.	0	.

a. The reference category is: Yes.

b. This parameter is set to zero because it is redundant.

6. Conclusions and recommendations

In this study, a multinomial logistic regression (MLR) model for analysing political attitudes was proposed. This methodology provided a powerful technique for modelling the relationship between the dependence of unordered categorical response variable on two categorical predictors. The proposed model was implemented using data from a survey conducted on Erbil city, which is a multi-ethnic and religious governorate in Kurdistan Region of Iraq. It was used to analyse the political attitudes for the population based on two main factors, ethnicity and religion. This study was the first to model the analysis of the behaviour of population towards the participation in political activities in this region. The main aim for the conducted survey was to analyse the political attitudes for people living in Erbil in Kurdistan Region of Iraq. The survey was constitutes of two parts, the first one was concerned with identification, ethnic and religion of the participants. The second part investigates whether they participate in political activities with categorical answers of three categories, "Yes", "To Some Extent" and "NO". The statistical analysis for the survey results explained the relationship between the dependent categorical variable representing the main question "Do you participate in political activities" with three categories and the independent predictors, the ethnic and religion. The most interesting result was that the majority of the population have interest in participating political activities; this is a good indicator towards setting up principles of democracy in the region.

The data analysis, using SPSS displayed the relation of practising the political activities and both factors, ethnic and religion especially for those groups whose participation was more than others, like Kurdish and Muslims. The model fit ascertains a good fit model for almost all groups where their participation was high.

Finally, as it is the first study concerns modelling the political attitudes for Kurdistan Region population, the survey only covered the main categories for each of the national ethnicity and religion and it did not consider the minorities in the region; is recommended that this study has to be extended to cover a larger population of Erbil, the sample size to be bigger than 411, to include ethnic groups and religions with samples of bigger sizes, hence the estimate will be more accurate and the results will be more reliable. Moreover, it is recommended that other factors will be included such as age and gender and then the model will be defined by four independent predictors and the analysis will be more detailed.

References

Al-Habib, A. (2012) "An Application on Multinomial Logistic Regression Model", *Pakistan Journal of Statistics & Operation Research*, 8 (2), 271-298

Allenby, M. and Leuk, J. (1994) 'Modeling household purchase behaviours with logistic normal regression', *Journal of the American Statistical Association*, 89(428), 1218–1231;

Balabanis, G. and Vassileoiu, S. (1999) "Some additional predictors of home shopping through the Internet", *Journal of Marketing Management*, 15(1), 361–385;

Berry, D. (1993) *Understanding Regression Assumptions*, Sag university paper series on quantitative applications in the social sciences, Beverly Hills, CA:Sage;

Berry, D. And Feldman, S. (1985) *Multiple Regression in Practice*, Sag university paper series on quantitative applications in the social sciences, Beverly Hills, CA:Sage;

Colin, B. And Robert, J. (1984) "Calculation of polychotomous logistic regression parameters using individualized regressions", *Biometrica*, 7(1), 11-18;

Cornfield, J., Gordon T. and Smith W. (1961) " Quantal response curves for experimentally uncontrolled variables, *Bulletin of the International Statistical Institute*, 38 (1), 97—115;

Field, A. (2009) *Discovering Statistics Using SPSS*, (2nd Edition), London, SAGE Publications;

Henn, M., Weinstein, M. and Wring, D. (2002) "A generation apart? Youth and political participation in Britan", *British Journal of Politics and International Relations*, 4(1), 167-192;

Hosmer, W. and Lemeshow, S (2000) *Applied logistic regression,* 2^{nd} edition, New York, Wiley- USA: John Wiley and Sons;

Li, Y. and Marsh, D. (2008) “New forms of political participation: searching for expert citizens and every day makers”, *British Journal of Politics and International Relations*, 38(1), 247-272;

Tabachnick, G., Fidell, S. and Osterlind, J (2001) *Using multivariate statistics*, US, Allyn and Bacon Boston;

Thompson, B. (2008) Foundations *of Behavioral Statistics*, USA, New Age International Limited;

Wackerly, D. Mendenhall, W. And Scheaffer, R. (2002) Mathematical Statistics with Applications, Sixth Edition, USA, Duxbury-Thomson Learning;

Walter S. And Duncan D (1967) “ Estimation of the probability of an event as a function of several variables”, *Biometrika*, 54(1), 79-167;

Wani, J. (1971) Probability and Statistical Inference, USA, Meredith Corporation