تشخيص اسباب حوادث السيارات في منطقة كرميان باستخدام نموذج الانحدار اللوجستي متعدد الحدود

ا.م.د. مجد محمود فقي حسين

جامعة السليمانية

E-Mail : mohammad.faqe@univsul.edu.iq

الملخص

تهدف هذه الورقة إلى التعرف على تطبيق لنموذج الانحدار اللوجستي متعدد الحدود (MLR) والذي يعد أحد الأساليب المهمة لتحليل البيانات الفئوية. يتعامل هذا النموذج مع متغير استجابة اسمي / ترتيبي واحد له أكثر من فئتين ، سواء كان متغيرًا اسميًا أو ترتيبيًا. تم تطبيق هذا النموذج في تحليل البيانات في العديد من المجالات ، على سبيل المثال ، التعليم و الصحة والعلوم الاجتماعية والسلوكية. الانحدار اللوجستي متعدد الحدود هو امتداد بسيط للانحدار اللوجستي الثنائي الذي يسمح بأكثر من فئتين من المتغير التابع أو متغير النتيجة. مثل الانحدار اللوجستي الثنائي ، يستخدم الانحدار اللوجستي متعدد الحدود أقصى تقدير للاحتمالية لتقييم احتمالية العضوية الفئوية. تم تطبيق هذه الطريقة في حوادث السيارات ، وتحتوي البيانات على سبعة متغيرات توضيحية مثل (رخصة القيادة ، مكان الحادث ، نوع الطريق ، سرعة السيارة ، الطقس ، نوع المركبة ، ووقت الحادث). حيث يكون نوع الحادث هو متغير الاستجابة والمتغيرات السبعة الاخرى عبارة عن متغيرات مستقلة ، بعد تحليل الجانب العملي للبيانات نستنتج بان هناك بعض من المتغيرات معنوية و مهمة في النموذج ، وتحديداً (رخصة القيادة ، نوع الطريق ، وسرعة السيارة)، كما أثبت النموذج انه لديها القدرة في التنبؤ ووصل إلى الدقة التي أظهرها بنسبة 67٪ في نموذجنا.

كلمات مفتاحية : الانحدار اللوجستي متعدد الحدود (متعدد الجوانب) (MLR) ، نيوتن رافسون ، جودة الملائمة ، AIC، BIC.

Diagnosing the Causes of Car Accidents in the Garmian Region Using Multinomial Logistic Regression Model

Assistant Prof. Dr.Mohammad M. Faqe Hussein
University of Sulaimani
College of Administration and Economics
Statistics and Informatics Dept.
E-Mail : mohammad.faqe@univsul.edu.iq

Abstract:

This paper aims to identify an application of the multinomial logistic regression model

(MLR), which is one of the most important methods for categorical data analysis. This model handles a nominal/ordinal response variable that has more than two categories, whether nominal or ordinal. This model has been applied to data analysis in many fields such as education, health, behavioral and social sciences. Multinomial logistic regression (MLR) is an extension of binary logistic regression that allows for more than two categories of response or outcome variables. Like binary logistic regression, multinomial logistic regression uses maximum likelihood estimation to assess the likelihood of categorical membership. This method was applied to car accidents and the data includes seven variables such as (driver's license, accident location, road type and vehicle speed, weather, vehicle type and accident time). When the accident type is the dependent variable and the seven other variables are independent variables, after analyzing the data We conclude that the model contains many significant variables, most notably (driver's license, road type, and the speed of the car). Each distinct type is significant in the multinomial logistic regression (MLR) model, and the model is also predictive and has achieved the 67% accuracy shown in our model.
Keywords: Multinomial (Polychotomous) Logistic Regression (MLR), Newton-Raphson, Goodness of fit, AIC, BIC.

1 Methodology

1.1 Multinomial (Polychotomous) Logistic Regression [3],[4]

In statistics multinomial logistic regression (MLR) is a technique that generalizes logistic regression to multiclass problems, with two or more likely discrete outcomes. It is a model used to predict the probabilities of the various possible outcomes of a categorical response given a set of independent variables (which can be real, binary, categorical, etc.). In fact, the multinomial logistic regression (MLR) model is a relatively simple generalization of the binary model, and both models respond primarily to logistic regression. In several ways, logistic regression is the normal complement of OLS when the answer is in a categorical variable. If some discrete variables appear between independent variables, they are treated by introducing one or more (0, 1) dummy variables. But when the dependent variable belongs to this type of data, the MLRs, which are not suitable for logistic regression, offer a good alternative.

1.2 Assumptions of Multinomial Logistic Regression [1],[2],[5]

In addition to the above:

1) The response variable should be measured at the nominal level with three more than or equal values.
2) The model has one or more explanatory variables that are (continuous, ordinal, or nominal). However, ordinal explanatory variables must be treated as either continuous or categorical.
3) The model should have explanatory variables and the response variables should have mutually exclusive and exhaustive categories.
4) The data should not show any multicollinearity. Multicollinearity occurs when you have two or more explanatory variables that are highly correlated with each other.
5) There must be a linear relationship between all continuous explanatory variables and the logit transform of the response variable.
6) The data should not have outliers, high leverage, or high influencing points for the scale/continuous variables.

1.3 The Model [6],[8]

Generalizing to a multinomial dependent variable (MLR) requires us to make some notational revisions. Let P indicate the number of discrete categories of the response variable, where P is greater or equal than 2 . Now, consider Z as a random variable that can take on one of P's possible values. Each Z_i is a multinomial random variable if each observation is explanatory; Once again, we aggregated the data into populations, each denoting one single combination of explanatory variable settings. As with the binomial logistic regression model, the column vector contains elements that represent the number of observations in population i, and such that $M = \sum_{i=1}^{N} n_i$, the sample size. Since each observation records one of the possible values for the explanatory variable Z, let y be a matrix with N rows (one for each population) and $P - 1$ column. If P is equal to 2, this reduces to the column vector used in the binomial logistic regression model (BLRM). For each population, y_{ij} represents the observed counts of the $p^{th.}$ value of Z_i. Similarly, ψ is a matrix of the exact dimensions as y, where each element ψ_{ij} is the probability of observing the $p^{th.}$ value of the response variable for any given observation in the $i^{th.}$ population. The design matrix of explanatory variables,X remains the same—it contains N rows and $K + 1$ columns where K is the number of explanatory variables and the first element of each row, $x_{i0} = 1$,the intercept. Let α be a matrix with $K + 1$ rows and $P - 1$ column, such that each element γ_{kp} contains the parameter estimate for the $k^{th.}$ covariate and the $p^{th.}$ value of the explanatory variable. For the MLRM, we modify the linear component to the log of the odds of a $p^{th.}$ observation compared to the $P^{th.}$ observation. We will consider the $P^{th.}$ category to be the omitted or baseline category, where logits of the first $P - 1$ categories are constructed with the baseline category in the denominator.

$$\log\left(\frac{\psi_{ip}}{\psi_{iP}}\right) = \log\left(\frac{\psi_{ip}}{1-\sum_{j=1}^{J-1}\psi_{ip}}\right) = x_{i0}\gamma_{0p} + x_{i1}\gamma_{1p} + \cdots + x_{ik}\gamma_{kp} \qquad i = 1,2,3,\dots,N\,,P = 1,2,3,\dots,p-1$$

$$= \sum_{k=0}^{K} x_{ik}\,\gamma_{kp} \ldots\ldots(1)$$

Solving for, we have

$$\theta_{ij} = \frac{\mathrm{Exp}^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}}{1 + \sum_{p=1}^{P-1}\mathrm{Exp}^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}} \qquad j < J$$

$$\theta_{iJ} = \frac{1}{1+\sum_{p=1}^{P-1}\mathrm{Exp}^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}} \qquad \ldots\ldots(2)$$

1.3 Parameter Estimation [7],[9],[11]

For each population, the response variable follows a multinomial distribution with P levels. Thus, the joint probability density function is:

$$f(y|\gamma) = \prod_{i=1}^{N}\left[\frac{n_i!}{\prod_{p=1}^{P} y_{ip}!} \times \prod_{j=1}^{J} \theta_{ip}^{y_{ip}}\right] \quad ....(3)$$

Then P = 2, this reduces to $f(y|\gamma) = \prod_{i=1}^{N}\left[\frac{n_i!}{y_i!(n_i-y_i)!} \times \theta_i^{y_i}(1-\theta_i)^{n_i-y_i}\right]$. The probability function is algebraically equivalent to Eq. 3, the only difference being that the likelihood function expresses the unknown values of γ in terms of known fixed constant values for y. Since we want to maximize Eq. 3 concerning,γ, the factorial terms not containing any of the θ_{ij} terms can be treated as constants. Thus, the kernel of the log-likelihood function for multinomial logistic regression models is:

$$L(\gamma|y) \simeq \prod_{i=1}^{N}\prod_{p=1}^{P} \theta_{ip}^{y_{ip}} ....(4)$$

Replacing the P$^{th.}$ terms, Eq. 4 becomes:

$$= \prod_{i=1}^{N}\prod_{j=1}^{P-1} \theta_{ip}^{y_{ip}} \cdot \theta_{iP}^{n_i-y_{ip}}$$

$$= \prod_{i=1}^{N}\prod_{p=1}^{P-1} \theta_{ip}^{y_{ip}} \cdot \frac{\theta_{iP}^{n_i}}{\theta_{iP}^{\sum_{j=1}^{P-1} y_{ip}}}$$

$$= \prod_{i=1}^{N}\prod_{p=1}^{P-1} \theta_{ip}^{y_{ip}} \cdot \frac{\theta_{iP}^{n_i}}{\prod_{p=1}^{P-1} y_{iP}^{y_{ip}}} ....(5)$$

Since $b^{x+y} = b^x b^y$, the total in the exponent in the denominator of the last term becomes a product over the first P-1 terms of p. Proceed by summarizing the terms raised to the power of y_{ij} for each p up to P - 1:

$$= \prod_{i=1}^{N}\prod_{p=1}^{P}\left(\frac{\theta_{ip}}{\theta_{iP}}\right)^{y_{ip}} \cdot \theta_{iP}^{n_i} \quad ....(6)$$

Now, substitute for θ_{ip} and θ_{iP} using Eq. 1 and Eq. 2:

$$\prod_{i=1}^{N}\prod_{p=1}^{P-1} e^{(\sum_{k=0}^{K} x_{ik}\gamma_{kp})^{y_{ip}}} \cdot \left(\frac{1}{1+\sum_{p=1}^{P-1} e^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}}\right)^{n_i}$$

$$\prod_{i=1}^{N}\prod_{p=1}^{P-1} e^{y_{ip}\sum_{k=0}^{K} x_{ik}\gamma_{kj}} \cdot \left(1+\sum_{p=1}^{P-1} e^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}\right)^{-n_i} ....(7)$$

Taking the natural log of Eq. 7 gives us the log likelihood function for the multinomial logistic regression model:

$$l(\gamma) = \sum_{i=1}^{N}\sum_{p=1}^{P-1}(y_{ip}\sum_{k=0}^{K} x_{ik}\gamma_{kp}) - n_i\log(1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}) ....(8)$$

As with the binomial model, we want to find the values for γ which maximize Eq. 8. We will do this using the Newton-Raphson method, which involves calculating the first and second derivatives of the log-likelihood function.

$$\frac{\partial l(\gamma)}{\partial\gamma_{kp}} = \sum_{i=1}^{N} y_{ip}x_{ik} - n_i \cdot \frac{1}{1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}} \cdot \frac{\partial}{\partial\gamma_{kp}}\left(1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}\right)$$

$$= \sum_{i=1}^{N} y_{ip}x_{ik} - n_i \cdot \frac{1}{1+\sum_{j=1}^{J-1} e^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}} \cdot Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \frac{\partial}{\partial\gamma_{kp}} \sum_{k=0}^{K} x_{ik}\gamma_{kp}$$

$$= \sum_{i=1}^{N} y_{ij}x_{ik} - n_i \cdot \frac{1}{1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kj}}} \cdot Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \cdot x_{ik}$$

$$= \sum_{i=1}^{N} y_{ij}x_{ik} - n_i\theta_{ij}x_{ik} \quad \ldots\ldots..(9)$$

Note that there are $(P-1)\times(K+1)$ equations in Eq. 9, which we want to set equal to zero and resolve for each γ_{kp}. Although technically it's a matrix, we can think γ of it as a column vector by appending each of the additional columns below the first. In this way we can form the second partial derivatives matrix as a square matrix of order$(P-1)\times(K+1)$. For each γ_{kp} we have to use Eq. 9 in relation to all other γ_{kp} . We can express the general form of this matrix as follows:

$$\frac{\partial'' l(\gamma)}{\partial\gamma_{kp}\,\partial\gamma_{k'p'}} = \frac{\partial}{\partial\gamma_{k'p'}} \sum_{i=1}^{N} y_{ip}x_{ik} - n_i\theta_{ip}x_{ik}$$

$$= \frac{\partial}{\partial\gamma_{k'j'}} \sum_{i=1}^{N} -n_i x_{ik}\psi_{ip}$$

$$= -\sum_{i=1}^{N} n_i x_{ik} \frac{\partial}{\partial\gamma_{k'p'}}\left(\frac{Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}}{1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kj}}}\right) \ldots\ldots\ldots\ldots(10)$$

Applying the quotient rule of $\left(\frac{f}{g}\right)'(b) = \frac{g(b)\cdot f'(b) - f(b)\cdot g'(b)}{[g(b)]^2}$, note that the derivatives of the numerator and denominator differ depending on whether or not $p' = p$:

$$f'(b) = g'(b) = Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \cdot x_{ik'} \qquad p' = p$$

$$f'(b) = 0 \qquad g'(b) = Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \cdot x_{ik'} \quad p' \neq p \qquad \ldots\ldots\ldots\ldots(11)$$

Thus, when $p' = p$, the partial derivative in Eq. 10 be:

$$= \frac{\left(1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kj}}\right)\cdot Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \cdot x_{ik'} - Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \cdot Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \cdot x_{ik'}}{\left(1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kj}}\right)^2}$$

$$= \frac{Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \cdot x_{ik'}\left(1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} - e^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}\right)}{\left(1+\sum_{p=1}^{P-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}}\right)^2}$$

$$= \theta_{ij}x_{ik'}(1-\theta_{ip}) \ldots\ldots\ldots..(12)$$

and when $p' \neq p$, they are:

$$= \frac{0 - Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}} \cdot Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp'}} \cdot x_{ik'}}{(1 + \sum_{j=1}^{J-1} Exp^{\sum_{k=0}^{K} x_{ik}\gamma_{kp}})^2}$$

$$= -\theta_{ij} X_{ik'} \psi_{ip'}$$

We can express now the matrix of second fraction derivatives for the MLR as follows:

$$\frac{\partial^2 l(\gamma)}{\partial \gamma_{kj}\, \partial \gamma_{k'p'}} = -\sum_{i=1}^{N} n_i x_{ik} \theta_{ip}(1 - \theta_{ip}) X_{ik'} \qquad p' = p$$

$$= -\sum_{i=1}^{N} n_i x_{ik} \theta_{ip} \psi_{ip'} X_{ik'} \qquad p' \neq p \ \ \ldots\ldots(13)$$

1.4 Newton- Raphson Method [6],[12],[13]

To show the iterative technique of Newton's method as it uses the (MLR), we need an expression for $\gamma^{(1)} - \gamma^{(0)} = [-l''(\gamma^{(0)})]^{-1}$. Let μ be a matrix with N rows and $P - 1$ columns, the exact dimensions as y and π, with elements equal to $n_i\theta_{ij}$. Then, $l'(\gamma) = X^T(y - \mu)$ expresses a matrix with $K + 1$ rows and $P - 1$ columns, the same dimensions as γ. By matrix multiplication, the elements of this matrix are equivalent to those derived in $l(\gamma) = \sum_{i=1}^{N} y_i\left(\sum_{k=0}^{K} x_{ik}\gamma_k\right) - n_{i.} \cdot \log\left(1 + Exp^{\sum_{k=0}^{K} x_{ik}\gamma_k}\right)$. The expression for the matrix of second partial derivatives is somewhat different from that derived in the binomial case since the equations in $\frac{d}{dx} Exp^{u(x)} = e^{u(x)} \cdot \frac{d}{dx} u(x)$ differ depending on whether or not $p' \neq p$. For the diagonal elements of the matrix of second partial derivatives, i.e., where $p' \neq p$, let W be a square matrix of ordering N, with elements $n_i\theta_{ij}(1 - \theta_{ij})$ on the diagonal and 0 everywhere else. Then, $l''(\gamma) = X^T(y - \mu)$ generates a K +1×K +1 matrix. However, we can only use this formulation for diagonal elements. For the off-diagonal elements, where $p' \neq p$, define **W** as a diagonal matrix with elements$n_i\psi_{ip}\psi_{ik}$, and use the negative of the expression in $l''(\gamma) = X^T(y - \mu)$. Using this dual formulation for W, each step of the Newton's method can proceed as in the BLRM, using $\gamma^{(1)} = \gamma^{(0)} + [X^TWX)]^{-1} \cdot X^T(y - \mu)$

1.5 Goodness of fit[9],[11]

Test the significance of parameters jointly with the test H or likelihood ratio using a hypothesis test

$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$

Ha : There is a least of one $\alpha_k \neq 0$, with k = 1,2,3,.., p

So, test statistics $H = -2\ln[Likelihood_0 - Likelihood_a]$

Where $Likelihood_0$ is Likelihood without explanatory variables, and $Likelihood_a$ is Likelihood with explanatory variables. Test criteria rejected H_0 if $H > \chi^2_{(df,\alpha)}$ or p-value < α=0.05. Therefore, partial test using a Wald test with a hypothesis

$H_0 : \alpha_j = 0$

$Ha : \alpha_j \neq 0$ with $j = 1,2,3,.., p$

So, the Wald test are : $\text{Wald} - \text{Test} = \left(\frac{\widehat{\alpha_i}}{\text{Standard Error}(\widehat{\alpha_i})}\right)^2$ …………(14)

We rejected H_0 if $W > \text{Chi} - \text{square}_{(df,\alpha)}$ or P-value < α(0.01 or 0.05). The next step is to calculate the precision of the classification. The measure for comparing classification processes are established on the information criterion. [3],[11]:

1.6.1 AIC : Akaike information criterion:

To compare the quality of statistical models, use AIC. When you chose the best model, it considered doing a hypothesis test to find the association between the variables and after that we get the good output. We define the AIC as follows:

$$\text{AIC1} = 2 \times Z - 2 \times (\log - \text{likelihood}) \quad ……..(15)$$

Where:

- Z is the parameter numbers (Intercept+ the number of variables) in the model.
- The log –likelihood obtains from statistical results and is a measure the fitted model.

Use the second order of AIC if ($\frac{n}{z} < \approx 40$(small sample)), use the second-order AIC:

$$\text{Akaike information criterion} = \text{AIC1} + \frac{2\times Z(Z+1)}{(n-Z-1)} \quad ……..(16)$$

Where:

- n = Number of observations,
- Z= Parameters number of the model,
- Log-likelihood obtains from statistical results and is a measure the fitted model.

1.6.2: BIC: Bayesian Information Criterion

The BIC criterion is a measure of model selection from a predictable set of models; the model with the lower BIC will be selected. It is closely linked to the AIC. It probably increases the likelihood of suitable models by adding parameters, but this can lead to over fitting. Both BIC and AIC try to solve this problem by introducing a penalty term for the number of parameters in the model; the penalty period is longer for BIC than for AIC. The BIC is formally defined as

$$\text{Bayesian information criterion} = Z \times \ln(n) - 2 \times \text{Ln}(L) \quad ……..(17)$$

Where:

L = the maximum value of the model's probability function

n = the sample size

Z = the parameters number.

1.6.3 The Chi-Square Test[11],[14]

The χ^2 test for goodness of fit is created to test whether observed frequencies differ significantly from expected frequencies. Here are the assumptions:

1. The data must to be grouped or binned, with each group or bin including five or more observations. Some data is already grouped into data categories. The bin sizes depend on the data as a general rule, and bins should contain at least five observations.

2. The data must come from a univariate distribution whose cumulative distribution function must be known. The null hypothesis states that the data follow a specific distribution, and the alternative is that the data do not.

We represent a quantity that summarizes our data. This quantity is the chi-square test. This statistic needs that the data be grouped into "bins.

$$\chi^2 = \sum_{i=1}^{k} \frac{(\text{Oberve}_i - \text{Expect}_i)^2}{\text{Expect}_i} \quad \text{.........(18)}$$

Where O_i=observed frequency for bin j

E_i=Expected frequency for bin j

The χ^2 test statistic in Equation (18) is then approximately chi-square distributed with degrees of freedom in the number if m is the number of population parameters that need to be estimated. The value of chi-square depends on how the data is binned. The larger the sample size, the better the approximation. The null hypothesis is rejected if the statistic in Equation (18) exceeds a critical value determined by the significance level.

1.6 The Data Description

The data we used in this article about car accidents in the Garmian region contains seven variables such as (driver's license, accident location, road type, speed of the car, weather, accident time, vehicle type). The type of accident is the dependent variable and the others are independent variables. We receive the data from the General Directorate for Traffic in the Garmian administration. The data were analyzed using the statistics package for the Social Science 26 software. Uses MLR for data analysis, the dependent variable have three levels; we took all reference categories so we can see the differences between all categories.

Table (1): Summary of case processing

	Accidents	N	Marginal %
Type of accident	Collisions (Crashes) accidents	398	66.33%
	Run-over accidents	110	18.33%
	rollovers(Coup) accident	92	15.33%

Factors	Levels	N	Marginal %
Driver's license	They do not have a driver's license	119	19.8%
	They have a driver's license	481	80.2%
Accident location	Outside	373	62.2%
	Inside	227	37.8%
Road type	Outside	278	46.3%
	Inside	322	53.7%
Speed of car	Slow	275	45.8%
	Fast	325	54.2%
Weather	Normal	573	95.5%
	Not-Normal	27	4.5%
Vehicle type	Small	374	62.3%
	Big	125	20.8%
	Median	101	16.8%
Accident time	Sun	390	65.0%
	Night	210	35.0%
Total		600	

Table (2) : Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	χ^2	df	P-value
Intercept Only	492.099	500.893	488.099			
Final	390.121	460.471	358.121	129.979	14	.000

A first method of assessing goodness-of-fit is to consider variables that we added to improve the model statistically significantly compared to the intercept alone (no added variables). We can see from the table that the table contains a Likelihood Ratio Chi-square test "**P-Value**" column that ($\chi^2_{(0.05,14)} = 23.68$, $\mathbf{P-Value.} = 0.000 < 0.05$), which means that the whole model statistically significantly predicts the dependent variable better than the intercept-only model. We see that the final model significantly improves in fit over a null model.

The Akaike Information Criterion and Bayesian Information Criterion assess a model by how close its fitted values tend to be to actual expected values, which is summarized by a given expected distance between the two. The best model is the one whose fitted values tend to be closest to the actual outcome probabilities. In our model, AIC and BIC and -2log probability are very close.

Table(3): Test of Goodness-of-Fit

	χ^2	df	P-Value
Pearson	197.818	180	.172
Deviance	207.063	180	.081

Two measures provide in the goodness-of-fit table used to retrieve how well the model fits the data, as shown in the table above. The first line, labeled Pearson, shows the Pearsonχ^2 . From the table above we see that ($\chi^2_{(0.05,180)} = 124.34, \mathbf{P-Value} = 0.172 > 0.05$) is not statistically significant. Based on this measure, the model fits the data well. The second statistic is the variance, and if the test shows no significance ($\chi^2_{(0.05,180)} = 124.34\,, \mathbf{P-Value} = 0.081 > 0.05$), we assume the model is performing well fits the data (α =0.05).

Table(4): Pseudo R-Square

Cox and Snell	0.195
Nagelkerke	0.236
McFadden	0.124

The Nagelkerke R-square indicates that 23.6% of the total variation in car accidents due to the variations among the eight predictor variables.

Table (5): Test of Likelihood Ratio

Variables	Model Fitting Criteria/ Reduced Model			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	χ^2	df	P-Value
α-Intercept	399.235	460.792	371.235	13.114	2	.001
Driver's license	405.996	467.553	377.996	19.876	2	.000
Accident location	386.258	447.815	358.258	.138	2	.934
Road type	451.453	513.010	423.453	65.333	2	.000
Speed of car	404.516	466.073	376.516	18.395	2	.000
Weather	388.882	450.439	360.882	2.761	2	.251
Accident time	386.971	448.528	358.971	.850	2	.654

Vehicle type	388.590	450.147	360.590	2.469	2	.291

The table above contains the likelihood-ratio tests of the total contribution of each explanatory variable to the model. The conventionally used threshold, the predictors (driver's license, road type, and car speed) are significant in the model because their values (0.000, 0.000, and 0.000, respectively) are less than (α =0.05).

Table (6) :Parameter Estimates

Accident type		β	Standard Error	Wald test	df	P-Value	Exp(β)	95% Confidence Interval	
								Lower limit	Upper limit
Run-over accidents	α -Intercept	-2.779	.784	12.580	1	.000			
	Driver's license	-.466	.277	2.816	1	.093	.628	.364	1.081
	Accident location	.054	.252	.047	1	.829	1.056	.644	1.731
	Road type	1.248	.297	17.670	1	.000	3.482	1.946	6.229
	Speed of car	.174	.223	.609	1	.435	1.190	.768	1.844
	Weather	.459	.553	.689	1	.407	1.582	.536	4.673
	Accident time	.179	.230	.607	1	.436	1.196	.763	1.875
	Vehicle type	.113	.145	.613	1	.434	1.120	.843	1.487
rollovers(Coup) accident	α -Intercept	-1.364	.727	3.519	1	.061			
	Driver's license	-1.276	.290	19.400	1	.000	.279	.158	.493
	Accident location	.117	.354	.109	1	.741	1.124	.561	2.251
	Road type	-1.847	.351	27.627	1	.000	.158	.079	.314
	Speed of car	1.131	.277	16.704	1	.000	3.098	1.801	5.329
	Weather	.817	.504	2.633	1	.105	2.264	.844	6.077
	Accident time	.162	.267	.367	1	.544	1.175	.697	1.982
	Vehicle type	-.208	.171	1.484	1	.223	.812	.581	1.135

The table above shows the comparison between the first category (Run-over accidents) with the other categories where the variable road type is significant in the model, the p-values of the variable is (0.000 < α = 0.05) and the coefficients (driver's license, road type, and speed of the car) in the second set are significant since their values are (0.025, 0.000, and 0.004) by comparing them all to (α = 0.05).

Table (7) :Parameter Estimates

Accident type		β	Standard Error	Wald test	df	P-Value	Exp(β)	95% Confidence Interval	
								Lower limit	Upper limit
Collisions (Crashes) accidents	α -Intercept	2.779	.784	12.580	1	.000			
	Driver's license	.466	.277	2.816	1	.093	1.593	.925	2.744
	Accident location	-.054	.252	.047	1	.829	.947	.578	1.552
	Road type	-1.248	.297	17.670	1	.000	.287	.161	.514
	Speed of car	-.174	.223	.609	1	.435	.840	.542	1.301
	Weather	-.459	.553	.689	1	.407	.632	.214	1.867

	Accident time	-.179	.230	.607	1	.436	.836	.533	1.311
	Vehicle type	-.113	.145	.613	1	.434	.893	.672	1.186
rollovers(Coup) accident	α -Intercept	1.415	.963	2.161	1	.142			
	Driver's license	-.810	.360	5.055	1	.025	.445	.219	.901
	Accident location	.063	.404	.024	1	.877	1.065	.482	2.351
	Road type	-3.094	.427	52.638	1	.000	.045	.020	.105
	Speed of car	.957	.333	8.239	1	.004	2.603	1.354	5.002
	Weather	.359	.654	.301	1	.583	1.431	.397	5.158
	Accident time	-.017	.325	.003	1	.957	.983	.520	1.858
	Vehicle type	-.321	.208	2.395	1	.122	.725	.483	1.089

The table above shows the comparison between the first category (collision accidents) with the other categories where the variable road type is significant in the model, the p-values of the variable (0.000 < α = 0.05) and the coefficients (driver's license, road type, and speed of the car) in the second set are significant since their values are (0.025, 0.000, and 0.004) by comparing them all to (α = 0.05).

Table (8) :Parameter Estimates

Accident type		β	Standard Error	Wald test	df	P-Value	Exp(β)	95% Confidence Interval	
								Lower limit	Upper limit
Collisions (Crashes) accidents	α-intercept	1.364	.727	3.519	1	.061			
	Driving license	1.276	.290	19.400	1	.000	3.582	2.030	6.319
	The place of the accident	-.117	.354	.109	1	.741	.890	.444	1.781
	Road type	1.847	.351	27.627	1	.000	6.339	3.184	12.622
	The speed of car	-1.131	.277	16.704	1	.000	.323	.188	.555
	Weather	-.817	.504	2.633	1	.105	.442	.165	1.185
	Incident time	-.162	.267	.367	1	.544	.851	.505	1.435
	Vehicle type	.208	.171	1.484	1	.223	1.231	.881	1.720
Run-over accidents	α-Intercept	-1.415	.963	2.161	1	.142			
	Driving license	.810	.360	5.055	1	.025	2.248	1.110	4.557
	The place of the accident	-.063	.404	.024	1	.877	.939	.425	2.075
	Road type	3.094	.427	52.638	1	.000	22.073	9.568	50.922
	The speed of car	-.957	.333	8.239	1	.004	.384	.200	.738
	Weather	-.359	.654	.301	1	.583	.699	.194	2.518
	Incident time	.017	.325	.003	1	.957	1.018	.538	1.924
	Vehicle type	.321	.208	2.395	1	.122	1.379	.918	2.071

The table above shows the comparison between the first category (collision accidents) with the other categories where the variables (driver's license, road type and speed of the car) are significant in the model, the p-values of the variables (0.000, 0.000 and 0.000 respectively) and the coefficients (driver's license, road type, and speed of the car) in the second set are significant since their values are (0.025, 0.000, and 0.004) by comparing them all to (α = 0.05).

Table(9): Show the Classification

Observed	Predicted			
	Collisions (Crashes) accidents	Run-over accidents	Rollovers(Coup) accident	Percent Correct
Collisions (Crashes) accidents	382	1	15	96%
Run-over accidents	108	0	2	0.0%
Rollovers(Coup) accident	72	0	20	21.7%
Overall Percentage	93.7%	0.2%	6.2%	67.0%

The classification table shows that multinomial logistic regression accurately predicts the cases with seven predictor variables (67%). The MLR used the classification to determine which group membership was a good prognostic. Where the first group correctly predicted by the model is 96% of the time [since 382 out of 398 cases in the first level of collision accidents were predicted by the model; 382/(382 + 1 + 15) = 0.96%. Furthermore, the second group that the model correctly predicted is 0/(108 + 0 + 20) = 0.0%, the last category that the model correctly predicted is 20 / (72 + 0 + 20) = 21.7%

1.7 Conclusion:

In this paper we conclude that:

This paper proposes a multinomial logistic regression (MLR) model for analyzing car accidents. This methodology provided a powerful technique for modeling the relationship between the response variable on eight categorical predictors.

1. We have reviewed the model results and carried out some tests (Pearson and Deviance Chi-Square tests) to make sure that the model is fit for the data according to statistical terms.
2. In this paper used MLR to identify the variables that influenced car accidents. The analysis results indicate that the Driving license, road type, and the speed of the car are significant and more affected on car accidents than the other variable.
3. MLR model also has proved its ability to predict and has reached the precision exhibited 67% in our model.
4. There are some variables are significant in all two reference groups in the model; specifically (driver's license, road type, and the speed of car) were significant in each category in the MLR.

1.8 Recommendations:

1. The media must permanently educate citizens and motorists about the importance of safety and how to adhere to traffic laws and regulations.
2. Not to be negligent in the application of the law by traffic police.
3. Make the periodic inspection of the vehicle mandatory.
4. We are focusing on following up on driving education centers and the programs used in them and not granting a driving license until after passing the test.
5. They are stimulating the departments concerned with improving the road network for vehicles to reduce traffic accidents.
6. Instructing the concerned departments to combat drug and alcohol traffickers
7. Include training curricula in leadership training centers and raise awareness on how to use a mobile phone while driving

Reference:

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

[2] A. Erkan and N. Aydin, (2016), "**Examination by Multinomial Logistic Regression Model of the Factors Affecting the Types of Domestic Violence against Women: A Case of Turkey**", International Journal of Scientific & Technology Research Vol. 5, Issue 11. pp. 67-74.

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

[4] Abdalla M. El-Habil, (2012), "**an Application on Multinomial Logistic Regression Model**", Pakistan Journal of Statistics and Operation Research 8(2).

[5] B. Madhu, N. C. Ashok and S. Balasubramanian, (2014), "**A Multinomial Logistic Regression Analysis to Study the Influence of Residence and Socio-Economic Status on Breast Cancer Incidences in Southern Karnataka**", International Journal of Mathematics and Statistics Invention (IJMSI). Vol. 2, Issue 5. pp. 1-8.

[6] Dobson, A.J. (2002). "**An Introduction to Generalized Linear Models**", 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

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

[8] Hendry, David F.; Nielsen, Bent, (2007), "**Econometric Modeling: A Likelihood Approach**", Princeton: Princeton University Press.

[9] Powers, D.A. and Xie, Y.(2000),"**Statistical Methods for Categorical Data Analysis**". San Diego, CA: Academic Press

[10] R. Rana, R. Singhal, (2015), "**Chi-square Test and its Application in Hypothesis Testing, Statistical Section**", Central Council for Research in Ayurvedic Sciences, Ministry of AYUSH, GOI, New Delhi, India, Vol. 1, Issue. 1, pp. 69-71.

[11] Schwab, J. A. (2002)," **Multinomial logistic regression: Basic relationships and complete problems**". http://www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/

[12] Schafer, J.L. (2001)," **Lecture Notes for Statistics 544: Categorical Data Analysis I**", Fall 2001. Penn State Univ. http://www.stat.psu.edu/~jls/

[13] Ward, Michael Don; Ahlquist, John S., (2018), "**Maximum Likelihood for Social Science: Strategies for Analysis**", New York: Cambridge University Press.

[14] Website: The chi-squared test for goodness of fit is explained in

http://www.itl.nist.gov/div898/handbook/eda/section3/eda35f.htm