

Risk Factors of Road Traffic Accidents and Its Severity in North Shewa Zone, Amhara Region, Ethiopia

Bezarede Mekonnen

Department of Statistics, College of Natural and Computational Science, Debre Berhan University, Debre Berhan, Ethiopia

Email address:

bezaredemekonnen70@gmail.com

To cite this article:

Bezarede Mekonnen. Risk Factors of Road Traffic Accidents and Its Severity in North Shewa Zone, Amhara Region, Ethiopia. *American Journal of Theoretical and Applied Statistics*. Vol. 7, No. 4, 2018, pp. 163-166. doi: 10.11648/j.ajtas.20180704.15

Received: June 8, 2018; **Accepted:** July 7, 2018; **Published:** July 23, 2018

Abstract: Ethiopia is a country with the largest number of road traffic accidents and fatality rate. Pedestrians and the disabled, children and the aged in particular, are the major victims of these accidents. The severity of road traffic crashes is also likely to be much greater in Africa than anywhere else. The major objective of this study was to identify the major factors determining road traffic accidents and its severity in North Shewa Zone. The study is based on secondary data obtained from the police records of the accident cases in North Shewa Zone Police station from February 2013 - September 2016 G. C. The ordinal logistic regression analysis is applied to examine the association between severity levels of road traffic accidents and human related variables, environmental and road related variables and vehicle related variables. Ordinal logistic regression analysis revealed that age of driver, driver-vehicle relationship, speed, alcohol use, chewing khat, educational level, driving experience and vehicle service year were found to be significant predictors for severity levels of road traffic accident.

Keywords: Ordinal Logistic Regression, Road Traffic Accidents, Proportional Odds Model

1. Introduction

Road accidents are a major world economic and social problem as shown by the report of loss of lives and properties in many countries around the world. Reporting indicated the number of fatalities from road accidents per year of about 1.3 million and 50 million injuries were recorded [1] or an average of 3000 deaths/day and 30,000 injuries/day. The World Health Organization (WHO) [2] estimated the economic costs derived from road accidents reached 518 billion USD per year in high income countries and 65 billion USD per year in medium and low income countries.

Sub-Saharan Africa alone with only 4% of the global vehicle registered accounts for 10% of the total road fatalities. Conversely, the high income nations, with 60% of the total global vehicle fleet contribute only 14% of the annual road deaths. Two countries, South Africa and Nigeria, account for most of the reported deaths in Sub-Saharan Africa. The South African figure of over 9,000 has been consistent over time, while Nigeria with 6,185 deaths has declined from a high of over 9,200 in the early 1990s. Ethiopia, Kenya, Uganda, Tanzania and Ghana are the other countries that experience high numbers of road deaths [3].

2. Related Work

There has been considerable the number of researches conducted, focused on predicting motor vehicle crashes on transportation facilities. Many factors can combine to produce circumstances that lead to a motor vehicle crash there is rarely a single cause of such an event. Three categories of factors contribute to crashes: human factors, roadway environment factors, and vehicle factors. Human factors involve the actions taken by, or the condition of, the driver of the motor vehicle, including speeding and violating traffic laws, as well as being affected by alcohol or drugs, inattention, decision errors, and age. Roadway environment factors that contribute to, or are associated with, crashes include the roadway design (for example, medians, narrow lanes, the lack of shoulders, curves, access points, or intersections); roadside hazards (for example, poles, trees, or embankments adjacent to the road); and roadway conditions (for example, rain, ice, snow, or fog). Vehicle factors include any vehicle-related failures that may exist in the automobile or design of the vehicle. In general, human factors are considered to be the most prevalent factors contributing to

crashes, followed by roadway environment and vehicle factors [4].

3. Methodology

3.1. Ordinal Logistic Regression Model

Logistic regression model can be classified as multinomial, ordinal and binary. In this investigation Ordinal logistic regression model was used. The ordinal logistic regression procedure empowers one to select the predictive model for ordered dependent variables.

It describes the relationship an ordered response variable and a set of explanatory variables. The explanatory variables may be continuous or discrete (or any type). Ordinal response models have major importance in social sciences as well as demography, many social and health science phenomena. The responses are discrete or qualitative (which is ordered in nature) rather than continuous or quantitative in nature. Many such analyses involve an outcome or dependent variable that is ordinal and in these studies the logistic regression model has become the statistical model of choice. [5]

3.2. Proportional Odds (PO) Model

Proportional Odds Model is used as a tool to model the ordinal nature of a dependent variable by defining the cumulative probabilities differently instead of considering the probability of an individual event. It considers the probability of that event and all events that are ordered before it. When response categories are ordered, logits can directly incorporate the ordering. The cumulative probabilities are the probability that the response Y falls in category i or below, for each possible i the i^{th} cumulative probability is $pr(Y \leq i) = p_1 + p_2 + \dots + p_i$. The proportional odds model assumes that the cumulative logits can be represented as parallel linear functions of independent variables. That is, for each cumulative logit the parameters of the models are the same, except for the intercept. Consequently, according to the proportional odds assumption, odds ratio is the same for all categories of the response variable. The PO model, however, has some appealing features. At first, it is well invariant under several categories, as only the signs of the regression coefficients change. Secondly, it is invariant under collapsibility of the ordered categories, as the regression coefficients do not change when response categories are collapsed or the category definitions are changed. Thirdly, it produces the most easily interpretable regression coefficients, as $\exp(\beta)$ is the homogenous odds ratio over all cut-off points summarizing the effects of the explanatory factor X on the response Y in one single frequently used measure. Due to these reasons, the PO model is by far the most used regression model for ordinal data. Let Y takes categorical response variable with c ordered categories and assume; $Pr(Y=1)$ is P_1 , $Pr(Y=2)$ is $P_2, \dots, Pr(Y=i)$ is P_i ; for $i=1, \dots, c$.

The Cumulative probability reflects the ordering with;

$$pr(Y \leq 1) \leq pr(Y \leq 2) \leq pr(Y \leq 3) \leq \dots pr(Y \leq c) = 1 \quad \text{and}$$

cumulative probability of the first $c-1$ of Y is $pr(Y \leq i) = \pi_i, \text{ where } i = 1, 2, \dots, c-1$.

Then the odds of the first $c-1$ cumulative probability are

$$\text{odds}(pr(Y \leq i)) = \frac{pr(Y \leq i)}{1 - pr(Y \leq i)} = \left[\frac{\pi_i}{1 - \pi_i} \right] \quad i = 1, 2, \dots, c-1 \quad (1)$$

The proportional odds model models of the first $c-1$ cumulative probability as;

$$\text{logit}[pr(Y \leq i)] = \log(pr(Y \leq i)) = \log \left[\frac{pr(Y \leq i)}{1 - pr(Y \leq i)} \right] = \log \left[\frac{\pi_i}{1 - \pi_i} \right] \quad (2)$$

And the relationship between the cumulative logits of Y is:

$$\log \left[\frac{\pi_i}{1 - \pi_i} \right] = \log \left[\frac{\pi_i}{\pi_{i+1} + \dots + \pi_c} \right]; \quad i = 1, \dots, c-1.$$

Consider the collection of p explanatory variables denoted by $X = (X_1, X_2, \dots, X_p)$ where X is vector. The relationship between the predictor and response variables is not a linear function in logistic regression; instead, the logistic regression function is used, which is the logit transformation of π .

$$\pi_i = \frac{\exp(\alpha_i + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(\alpha_i + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}, \quad i = 1, \dots, c-1. \quad (3)$$

Then the logit or log-odds of having $pr(Y \leq i) = \pi_i$ is modeled as a linear function of the explanatory variables as:

$$\text{Log} \left[\frac{Pr(Y \leq i)}{1 - Pr(Y \leq i)} \right] = \log \left[\frac{\pi_i}{1 - \pi_i} \right] = \alpha_i + \beta_1 x_1 + \dots + \beta_p x_p \quad (4).$$

Equivalent with

$$\log \left[\frac{\pi_i}{1 - \pi_i} \right] = \alpha_i + \sum_{j=1}^p \beta_j x_j; \quad 0 \leq \pi_i \leq 1. \text{ Therefore} \\ \text{logit}[pr(Y \leq i)] = \alpha_i + \sum_{j=1}^p \beta_j x_j; \quad i = 1, \dots, c-1, j = 1, \dots, p. \quad (5)$$

The model assumes a linear relationship for each logit and parallel regression lines. Equation (4) is called proportional odds model and it estimates simultaneously multiple equations of cumulative probability. An equation is solved for each category of the dependent variable except the reference category. In this model each logit has its own α_i term called the threshold value and their values do not depend on the values of the independent variable for a particular case. Logistic regression coefficients are indicates the direction and strength of the relationship between independent variable and the log odds of dependent variable. However, these logistic regression coefficients are a little bit more complicated to intuitively gauge, as they present the influence of a unit change in the independent variable on the log odds of the dependent variable. The influence determines the rate of increase or decrease in the log odds of dependent variable. This means that the effect of the independent

variable is the same for different logit functions, that's also the reason why the model is called the proportional odds model [5-7].

4. Result

Table 1. Parameter Estimates.

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[Severity = 1]	-2.240	0.493	20.596	1	0.000	-3.207	-1.272
	[Severity = 2]	-0.224	0.474	0.224	1	0.636	-1.152	0.704
	[Severity = 3]	1.069	0.476	5.051	1	0.025	0.137	2.001
	[Age= 1]	0.562	0.286	3.865	1	0.049	0.002	1.122
	[Age= 2]	0.559	0.271	4.246	1	0.039	0.027	1.091
	[Age= 3]	0 ^a			0			
	[Relation= 1]	0.593	0.196	9.150	1	0.002	0.209	0.978
	[Relation= 2]	0 ^a			0			
	[Speed= 1]	0.909	0.227	16.099	1	0.000	0.465	1.353
	[Speed= 2]	0 ^a			0			
Location	[Alcohol= 1]	0.857	0.218	15.456	1	0.000	0.430	1.284
	[Alcohol= 2]	0 ^a			0			
	[Khat= 1]	1.040	0.227	21.040	1	0.000	0.596	1.484
	[Khat= 2]	0 ^a			0			
	[Education= 1]	0.564	0.273	4.276	1	0.039	1.099	0.029
	[Education= 2]	0.681	0.271	6.324	1	0.012	1.212	0.150
	[Education= 3]	0.097	0.308	0.100	1	0.752	0.701	0.506
	[Education= 4]	0 ^a			0			
	[Experience= 1]	0.871	0.332	6.886	1	0.009	1.522	0.221
	[Experience= 2]	0.447	0.341	1.723	1	0.189	1.115	0.221
	[Experience= 3]	0 ^a			0			
	[Service= 1]	-0.541	0.213	6.432	1	0.011	-0.959	-0.123
	[Service= 2]	0 ^a			0			

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

Interpretation of the Results

When the proportional odds model is used in the analysis of ordinal data, the coefficients of the explanatory variables in the model is interpreted as logarithm of the ratio of the odds of response variable. The interpretation of parameters corresponding to different variables which are found significant in the model as shown in Table 1 is described in the following section and comparison is made with the reference category.

In this study 13 variables are considered as possible risk factor for severity levels of road traffic accident. Of 13 variables the 8 variable have statistically significant effect on severity levels of road traffic accident by applying multiple ordinal logistic regression. The results displayed in table 1 show that age of drivers, vehicle-driver relation, speed, alcohol, khat, Educational level and Driving experience were found to be significant predictors for severity levels of road traffic accident.

The increment to log odds of severe outcome (severe outcome stands for fatal, severe injury and slight injury) for age of driver 18-30 is 0.562. The estimated odds ratio (OR = 1.75) indicates that drivers whose age are 18-30 are 1.75 times more likely to commit fatal accident than those drivers whose age are 51 and above (reference group). By the same token, drivers whose age is 18-30 are 1.75 times more likely to attain severe outcome (to commit fatal, severe injury and slight injury) than those drivers whose age are 51 and above holding other predictors effects are constant.

The increment to log odds of severe outcome for employed driver is 0.593. The estimated odds ratio (OR = 1.81) indicates that employed drivers are 1.81 times more likely to commit fatal accident than owner drivers (reference group). By the same token, employed drivers are 1.81 times more likely to attain severe outcome (to commit fatal, severe injury and slight injury) than owner drivers.

The log odd of severe outcome for vehicle service year is decreased by -0.541. The estimated odds ratio (OR =0.58) indicates that vehicles whose service are from 0-5 years are 0.58 times less likely to commit severe outcome than those vehicles whose service are above 5 years. Similarly, vehicles whose service are from 0-5 years are 0.58 times less likely to attain severe outcome (to commit fatal, severe injury and slight injury) than those vehicles whose service are above 5 years holding other predictors effects are constant.

The increment to log odds of severe outcome for speed is 0.909. The estimated odds ratio (OR = 2.48) indicates that drivers who drives above the speed limit are 2.48 times more likely to commit fatal accident than drivers who didn't drives above the speed limit. By the same token, drivers who drives above the speed limit are 2.48 times more likely to attain severe outcome (to commit fatal, severe injury and slight injury) than drivers who didn't drives above the speed limit.

The increment to log odds of severe outcome for alcohol is 0.857. The estimated odds ratio (OR = 2.36) indicates that drivers who uses alcohol are 2.36 times more likely to commit fatal accident than drivers who didn't use alcohol.

Similarly, drivers who uses alcohol are 2.36 times more likely to attain severe outcome (to commit fatal, severe injury and slight injury) than who didn't use alcohol holding all other variables constant.

The increment to log odds of severe outcome for khat is 1.04. The estimated odds ratio ($OR = 2.83$) indicates that drivers who uses khat are 2.83 times more likely to commit fatal accident than drivers who didn't use khat. Similarly, drivers who uses khat are 2.83 times more likely to attain severe outcome (to commit fatal, severe injury and slight injury) than who didn't use khat holding all other variables constant.

The increment to log odds of severe outcome for education is 0.564. The estimated odds ratio ($OR = 1.76$) indicates that drivers who can read and write are 1.76 times more likely to commit fatal accident than whose education is diploma and above. Similarly, drivers who can read and write are 1.76 times more likely to attain severe outcome (to commit fatal, severe injury and slight injury) than whose education is diploma and above holding all other variables constant.

The increment to log odds of severe outcome for driving experience is 0.871. The estimated odds ratio ($OR = 2.39$) indicates that drivers whose driving experience is 0-5 years are 2.39 times more likely to commit fatal accident than whose driving experience is 11 years and above. Similarly, drivers whose driving experience is 0-5 years are 2.39 times more likely to attain severe outcome (to commit fatal, severe injury and slight injury) than whose driving experience is 11 years and above holding all other variables constant.

5. Conclusion

This study was intended to identify the risk factors for severity levels of road traffic accident. Thirteen predictors were selected for the study and ordinal logistic regression model was developed to assess the relation between severity levels of road traffic accident and the selected variables.

Based on the results, the multiple ordinal logistic regression analysis was applied to select the most important determinants for severity levels of road traffic accident. The results of the proportional odds model show that age of driver, driver-vehicle relationship, speed, alcohol use, chewing khat, educational level, driving experience and vehicle service year were found to be significant predictors for severity levels of road traffic accident.

Specifically the study showed that severity levels of road traffic accident is more likely for young drivers, low

educational background, less than 5 year driving experience, employed drivers, speedy drivers, and for those drivers who uses alcohol and khat. Furthermore severity levels of road traffic accident are more likely for aged vehicles.

References

- [1] W. H. Organization, Global status report on road safety: time for action, World Health Organization, 2009.
- [2] M. M. Peden, World report on road traffic injury prevention, World Health Organization. 2004.
- [3] Greg Chen): Road traffic safety in African countries – status, trend, contributing factors, countermeasures and challenges, international journal of injury control and safety promotion, Vol. 17:4, pp. 247-255, 2010.
- [4] GAO (Government Accountability Office). Research Continues on a Variety of Factors that Contribute to Motor Vehicle Crashes. GAO-03-436, Government Accountability Office, Washington, D. C. 2003.
- [5] Harrell, Jr., F. E, Regression modeling strategies, with applications to linear models, logistic regression, and survival analysis. Springer-Verlag, New York. 2001.
- [6] Agresti, A. Categorical Data Analysis. Wiley Interscience, New York, second edition, 2002.
- [7] Hosmer, W. D. and S. Lemeshow (2000), Applied Logistic Regression. 2nd Edn., John Wiley and Sons, New York
- [8] K. K. W. Yau. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong, Accident Analysis and Prevention Vol. 36, pp. 333–340. 2004.
- [9] C. O'Donnell, D. Connor, Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice, Accident Analysis and Prevention Vol. 28, pp. 739–753, 1996.
- [10] K. M. Kockelman, Y.-J. Kweon, Driver injury severity: an application of ordered probit models, Accident Analysis and Prevention Vol. 34, pp. 313–321, 2002.
- [11] A. S. Al-Ghamdi, Using logistic regression to estimate the influence of accident factors on accident severity, Accident Analysis and Prevention Vol. 34, pp. 729–741, 2002.
- [12] L. Aarts, I. van Schagen, Driving speed and the risk of road crashes: a review, Accident Analysis and Prevention Vol. 38, pp. 215–224, 2006.
- [13] M. D. Keall, et al., The contribution of alcohol to night time crash risk and other risks of night driving, Accident Analysis and Prevention Vol. 37, pp. 816–824, 2005.