SciencePG
Science Publishing Group

Determinants that Lead Drivers into Traffic Accidents: A Case of Arba Minch City, South Ethiopia

Tesfahun Zewde

Statistics Department, Arba Minch University, Arba Minch, Ethiopia

Email address:
tesfahunzewde@yahoo.com

Abstract: This study attempted to assess the factors that lead drivers into traffic accidents at Arba Minch City. 200 drivers were selected using stratified random sampling method and the data have been collected using structured questioners. From sampled drivers 62% of the drivers were involved in one or more accidents. Poisson regression model was the appropriate one compared to the negative binomial regression model for the data. From Poisson regression analysis variables like driver experience, driving after alcohol use, having more licenses, speedy driving, and number of punishments were the causes that lead drivers into traffic accidents in the study area. Road safety professionals should target these factors in their efforts to reduce the occurrence of traffic accidents.

Keywords: Drivers, Traffic Accidents, Poisson Regression Model

1. Introduction

Road accidents are very common all over the world and annual global road crash statistics states that Some 1.25 million people die each year as a result of road traffic crashes, on average 3,287 deaths a day. Unless action is taken, road traffic injuries are predicted to become the fifth leading cause of death by 2030. Residents of developing countries are at much higher risk of Road traffic accident than are residents of high-income countries. A big gap still separates high-income countries from low- and middle-income ones where 90% of road traffic deaths occur in spite of having just 54% of the world's vehicles [1].

The severity of road traffic crashes is also likely to be much greater in Africa than anywhere else, because many vulnerable road users are involved, poor transport conditions such as lack of seat belts, overcrowding, and hazardous vehicle environments. The poor reporting system has also masked the magnitude of the problem in Africa [2].

Ethiopia is one of the fastest developing countries in the world with still having a low road network density and vehicle ownership level. According to federal police commission report the death rate due to car accident is significantly increasing among pedestrians and passengers from time to time in Ethiopia (3).

Evidences noted that human behavior is the most common factor accounting for more than 85% of all traffic accidents [4]. Among the risky human behaviors is driving over the recommended speed. Studies has indicated that an increase of 1 km/h in mean traffic speed results in a 3% increase in the incidence of accident crashes and a 4-5% increase in fatal crashes [5]. Another risky behavior identified for road traffic accident is taking alcohol and driving. Knowledge, belief, attitude on risky driving behaviors and driving experience were also important aspects of risky behaviors identified with evidences [6]. Mobile phoning while driving is becoming one of the riskier behaviors as well [7]. Arba Minch City is one of the City's that is found in the South Nation Nationalities and People's Regional States of South Ethiopia. According to the city traffic police office in the city more than 351 accidents occurred from 2013 to 2015 and most of the accidents occurred by driver's error. Therefore it needs further study to identify the factors that lead drivers into traffic accidents in City. The purpose of this study is to assess the factors that lead drivers into traffic accidents in Arba Minch City.

2. Data Description

For this study a random sample of 200 drivers who had been working in the Arba Minch City from January 2014 to August 2016 were considered. The data were collected using structured questioners.

The response variable is: The number of accidents the driver faced in the study period.

Predictors considered are: age, educational status, Marital status, Driver experience, Alcohol use, Speaking on the phone while driving, Vehicle ownership, Speedy driving, Number of licenses, Driver vehicle relationship, vehicle service year, average driving time per a day, and number of punishments for the violation the law.

3. The Models

3.1. The Poisson Regression Model

Poisson regression is a standard model for analysis of count data. The outcome variable for this study is the number of accidents; which is consists of counts. Thus the data is assumed to follow Poisson distribution. Poisson regression assumes that the dependent variable Y (number of occurrences of an event) follows a Poisson distribution (8). Let Yi be the random variable takes non negative values, $i = 1, 2, \ldots, n$ where n is the number of observations. If Yi is the number of counts for i^{th} occasion and follows a Poisson distributions with the probability function:

$$P(Y_i = y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i}, \quad y_i = 0, 1, 2, ..., n \quad (1)$$

With mean and the variance are equal $E(Yi \mid Xi) = Var (Yi \mid Xi) = \lambda i$, where $\lambda i = exp(Xi^T \beta)$, is the mean number of counts, X_i is the ith row of covariate matrix, and $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ are unknown k-dimensional vector of regression parameters. In a Poisson regression model, the model coefficients are estimated by the maximum likelihood method. The likelihood function L is the product of the terms in equation (1) over all n measured values y_i i.e

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i} \quad (2)$$

The log-likelihood function is given by

$$\ln(L(\beta)) = \sum_{i=1}^{n}[-\lambda_i + y_i \ln \lambda_i - \ln(y_i!)] = \sum_{i=1}^{n}[-\exp(X_i^T\beta) + y_i X_i^T\beta - \ln y_i!] \quad (3)$$

By differentiating equation (3) with respect β

$$\frac{\partial Ln(\beta)}{\partial \beta_j} = \sum_{i=1}^{n}(y_i - \exp(X_i^T\beta)X_i) = 0, \quad j = 1, 2, 3, ..., k \quad (4)$$

Estimating the regression coefficient in the Poisson regression model is not obtained from a direct equation but rather the Newton Raphson method is used for estimating the parameters that are unknown in the model (9).

3.2. Negative Binomial Regression Model

As mentioned above, a limitation of the Poisson distribution is that the mean equals the variance of the distribution. Previous work in the field of accident research has shown that this is not always the case. Suppose a Poisson model is used for modeling accidents and the variance of the data exceeds the estimated mean of the accident data distribution. The data are then said to be over dispersed, and the underlying assumption of the variance being equal to the mean for the Poisson distribution is violated. The negative binomial, which is a discrete distribution, provides an alternative model to deal with over dispersion in count data such as accident frequencies.

Unlike the Poisson distribution, the negative binomial distribution adds a quadratic term to the variance representing over dispersion. The negative binomial model takes the form:

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i!\Gamma(\frac{1}{\alpha})}(\frac{\alpha\lambda_i}{1+\alpha\lambda_i})^{y_i}(\frac{1}{1+\alpha\lambda_i})^{\frac{1}{\alpha}} \quad (5)$$

Where α is the over dispersion parameter and the variance is $\lambda_i + \alpha\lambda_i^2$

If $\alpha = 0$, the negative binomial reduces to Poisson distribution. The larger the value of α, the more variability there is in the data over and above that associated with the mean λ

The Log-likelihood function of the NB model is obtained from the following equation;

$$\ell = \sum_{i=1}^{n}\left[\sum_{i=1}^{y_i-1}\ln(y_i + \frac{1}{\alpha}) - \ln y_i! + \frac{1}{\alpha}[\ln(\frac{1}{\alpha}) - \ln(\frac{1}{\alpha} + \lambda_i)] + y_i[\lambda_i - \ln(\frac{1}{\alpha} + \lambda_i)]\right] \quad (6)$$

In order to estimate β and α as in the Poisson model, the iteration procedure or method of Newton Raphson is applied.

3.3. Model Comparison

When many regression models are available for a given data set, one can compare the performance of alternative models based on some measures of goodness-of-fit. Several measures of goodness-of-fit have been proposed in literature.

3.3.1. Test for Dispersion

The Negative binomial regression model reduces to the Poisson regression model when the dispersion parameter $\alpha \rightarrow$ zero. To assess the adequacy of the Negative binomial regression model over the Poisson regression model, one may test the hypothesis

$$H_0 : \alpha = 0 \quad vs \quad H_a : \alpha \neq 0 \quad (7)$$

This is to test for the significance of the dispersion parameter. The inclusion of the dispersion parameter α in the

regression model is justified when the null hypothesis H_0 is rejected. To carry out the test in (7), one can use the Pearson chi-square statistic

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\lambda}_i)}{\hat{\lambda}_i} \qquad (8)$$

For large samples the distribution of Pearson's statistic is also approximately chi-squared with n-p d.f. If $\dfrac{\chi^2}{n-p}$ is significantly larger than one, over dispersion is indicated. This indicates that negative binomial regression model will be more appropriate than Poisson regression model to data.

3.3.2. Deviance Goodness of Fit

After fitting the various models involve in the research work to the data obtained, it is very important to check the overall fit as well as the quality of the fit of the respective models. The quality of the fit between the observed values (y) and the predicted values $\hat{\lambda}$ can be measured by the various test statistics, but one useful statistics is called the deviance goodness of fit test. For a better model, one must expect a smaller value of the Deviance. Hence the smaller the value of the deviance of a specific model the better the model or the more statistically significant the model becomes

3.3.3. Akaike Information Criterion (AIC)

When several maximum likelihood models are available, one can compare the performance of alternative models based on several likelihood measures which have been proposed in the statistical literature. One of the most popular used measures is AIC. The AIC is defined as

$$AIC = -2\ln L + 2K \qquad (9)$$

Where $\ln L$ is the log-likelihood value of the estimated model and K is the number of estimated parameters (10). A relatively small value of AIC is favorable for the fitted model.

3.3.4. Bayes Information Criterion (BIC)

The BIC is the other most regularly used measure to compare the performance of alternative models based on several likelihood measures. BIC is defined as

$$BIC = -2\ln L + K \ln(n) \qquad (10)$$

Where lnL is the log-likelihood value of the estimated model and K is the number of estimated parameters and n the number of observation. For this measure, the smaller the BIC value, the better the model is.

4. Results and Discussion

The purpose of this study was to assess factors that lead drivers into traffic accidents. The following Table 1 shows the frequencies of each categorical explanatory variable. Almost all of the respondents were male (99%) and 1% were female. Education level of the drivers have three categories from these 12%, 42% and 46% were from 5-8, from 9-12 and above 12 respectively. Marital status of the drivers have three categories from these 50.5%, 42% and 7.5% were single, married and Divorced/widowed respectively. 97.5% of drivers have driving license and 2.5% have no driving license. 25%, 49.5% and 25.5% of the drivers have been driving speed less than 30, 30-40 and above 40 respectively. 37.5% of the drivers sometimes drive after alcohol use and 62.5% do not drive after alcohol use. Vehicle Ownership have two categories, from these 83% were private and 17% were government and other. 56% of the drivers use phones while they are driving and 44% not.

Table 1. Summary of frequencies of categorical variables.

Variables		Frequency	Percent
Sex	female	2	1
	Male	198	99
Education	From 5-8	24	12
	From 9-12	84	42
	Above 12	92	46
Marital Status	Single	101	50.5
	Married	84	42
	Divorced	15	7.5
License Status	Yes	195	97.5
	no	5	2.5
Vehicle Type	Motor	28	14.0
	Bajaj	82	41.0
	Minibus	34	17.0
	Buses	13	6.5
	Cargo	6	3.0
	Automobiles &others	37	18.5
Vehicle ownership	Private	166	83.0
	Government & others	34	17.0
Average Speed	<30km/hr	50	25
	30-40km/hr	52	25.5
	>40km/hr	99	49.5
Driver-Vehicle relationship	Owner	77	38.5
	Hired &others	123	61.5

Variables		Frequency	Percent
Driving after drinking	No	125	62.5
	Sometimes	75	37.5
Speaking on the phone	No	88	44.0
	Sometimes	112	56.0

Table 2 shows that the following information about the continuous variables. The maximum driver age is 65 and the minimum 19. On average the drivers age is 31.65 this may not be representative of the driver's age because mean is not a good representative of the data if the data have extreme observations. In maximum a driver who violated the law had been punished 10 times and on average 2.25. The maximum number of licenses a driver owns is 4 and minimum is 0. On average a driving time for a driver in a day is 8.32 hours with standard deviation 3.92 hours. The average driving experience for a driver is 7.34 with standard deviation 8.77.

Table 2. *Continuous variable information.*

	Minimum	Maximum	Mean	Std. Deviation
Number of licenses	0	4	1.42	.71
Driver age	19	65	31.65	10.32
Vehicle service year	60	21	4.2	4.03
Average driving time in a day	2	15.3	8.32	3.92
The number of Punishment	0	10	2.25	2.06
Driver Experience	1	40	7.34	8.77

The minimum number of the accident is 0 and the maximum is 5 and the mean and variances are 1.25 and 1.435 respectively.

4.1. Analysis for Model Comparison

The Poisson Deviance and Pearson Chi square divided by their respective degrees of freedom are nearly equal to one (Table 3). This can lead to the assumption that the dispersion parameter α is zero. This may indicate that the Poisson regression model is the appropriate one for the data but this should farther approved by AIC and BIC of the two models

Table 3. *Goodness of fit of the model.*

	Poisson regression model		
	Value	df	Value/df
Deviance	196.559	182	1.080
Pearson Chi-square	189.775	182	1.043

The values of AIC and BIC for Poisson regression model is less than that of Negative binomial regression model AIC and BICs respectively (Table 4). These indicate that the Poisson regression model is the appropriate model than the negative binomial regression model for the data. Consequently, the analysis is done by applying Poisson Regression model.

Table 4. *The AIC and BIC results of PR and NBR models.*

	Poisson regression model	Negative Binomial regression
AIC	571.276	573.276
BIC	630.646	654.944

4.2. Analysis of the Data Using Poisson Regression Model

The Omnibus test helps to compare the fitted model against the intercept-only model. The results of the Omnibus test (Chi-square value) shows highly significant at level of significance, $\alpha=0.5$ (Table 5) and the explanatory variables predict the number of accidents well and the model is good fit model.

Table 5. *Summary statistics of the likelihood ratio test.*

Likelihood Ratio Chi-square	df	Sig
55.723	17	0.000

Among the 13 predictor variables provided from the Poisson regression analysis five of them namely; drivers experience, alcohol use, speed, number of license drivers have and number of punishments for the violation the law were found to be significantly determine the number of accidents

Table 6. *Results of analysis of data using poisson regression model.*

Variables		β	S.E	Wald	df	Sig	Exp(β)
Age		-0.009	0.0112	0.570	1	0.450	0.992
Education	5-8	.300	0.2472	1.473	1	0.225	1.350
	9-12	0.042	0.1791	0.054	1	0.816	1.042
	>12	(re)					
Marital Status	Single	-0.097	0.1768	0.302	1	0.582	0.907
	Married	-0.152	0.2528	0.360	1	0.548	0.859
	Divorced	(re)					
Driver experience		-0.050	0.0235	4.558	1	0.033	0.951
Alcohol use	Yes	0.612	0.1819	11.336	1	0.001	1.845
	No(re)						
Speaking on the phone while	speak	0.155	0.1513	1.043	1	0.307	1.167

Variables		β	S.E	Wald	df	Sig	Exp(β)
driving	No(re)						
Vehicle owner	Gov &other	-0.213	0.2425	0.774	1	0.379	0.808
	Private	(re)					
	>40	0.779	0.3472	5.036	1	0.025	2.179
Speed	30-40	0.543	0.3594	2.247	1	0.131	1.721
	<30(re)						
Number of License		0.148	0.044	11.56	1	0.001	1.159
Number of punishment		0.137	0.0412	11.043	1	0.001	1.147
Driver vehicle relationship	Owner	-0.217	0.1697	1.641	1	0.200	0.805
	Hired &other	(re)					
Vehicle service year		0.020	0.0217	0.867	1	0.351	1.020
Driving time in a day		0.013	0.0157	0.730	1	0.393	1.014

Table 6 gives that the following information about the significant variables. The experience of the driver increases in one year would increase the accidents frequency by 0.951. Drivers who drive after having alcohol are expected to have 1.845 more accident than those who do not have alcohol. Drivers who drive above 40 km/hr are expected to have 2.179 more accident frequency than the reference category below 30km/hr. A unit increases in the number of license of the drivers would increase the number of accidents by 1.159. The number punishment increases in one unit would increase number of the accidents by 1.147.

4.3. Discussion

Road traffic accidents have become a major public health and economic problem worldwide. This problem is getting worse in developing countries, like Ethiopia. This study showed that from 200 sampled drivers 62% of the drivers were involved in one or more accidents.

Variables like driving experience, alcohol use, speed, number of licenses and number punishment were statistically significantly affect the traffic accidents at 5% level of significance. Driving experience of the driver is one of the factors that affect the traffic accidents in the study area. A similar study in Tanzania showed that less experienced drivers were found to have high risky driving behaviors and traffic accidents compared than more experienced drivers [11]. Driving after consuming alcohol was one of the causes of road traffic accidents for this study. The similar study showed that alcohol use was one the predictor variable for fatal accidents. And also, the same study showed that younger age drivers were one of the causes of traffic accidents [12]. For this study, age of the respondents was not significant; but from the coefficient of the driver age we have information about the decrease of traffic accidents as increase of age. Education level of the drivers was not significant cause of traffic accidents. However, the other study showed that the more the drivers educated the less the occurrence of the accidents and Speed was the most causes of traffic accidents for this study at 5% level of significance [13]. About 30% of accidents on the roads were caused because of speedy driving [14]. The other significant variable was number of licenses a driver has. This study showed that as the number of driver license increases the traffic accidents also increases. Violating the traffic law was one of the causes of traffic accidents for this study. The more the violation of the traffic law and rules showed that the more the accident is.

5. Conclusion

The main objective of this study was to identify the determinate factors that lead drivers into traffic accidents. From sampled drivers 62% of the drivers were involved in one or more than one accidents. Poisson regression model was employed to analyze the data. The results revealed that variables like driver's experience, alcohol use, driving over speed, number of licenses and number punishment were the main causes of traffic accident. The remaining variables were not statistically significant at 5% level of significance. Road safety professionals should target these factors in their efforts to reduce the occurrence of traffic accidents. Further studies can be made on the area of traffic accident by considering detail and accurate information on various variables.

References

[1]   WHO, 2015: Global status report on road safety 2015.

[2]   Lagarde E: Road traffic injury is an escalating burden in Africa and deserves proportionate research efforts. PLos Medicine 2007, 4: 0967-0971.

[3]   Federal Democratic Republic of Ethiopia: Police Commission Report 2014.

[4]   Peden M, Scurfield R, Sleet D: World report on road traffic injury prevention 2004.

[5]   Finch J, Kompfner P, Lockwood R, Maycock G: Speed limits and accidents Transport Research Laboratory; 1994.

[6]   Waller F, Stewart J, Hansen A: The potentiating effects of alcohol on driver injury. J Am Med Assoc 1996, 255: 522-527.

[7]   US department of transportation: Traffic Safety Facts Washington, DC; 2009.

[8]   Washington SP, Karlaftis MG, Mannering FL (2003) Statistical and econometric methods for transportation data analysis. Chapman Hall/CRC, Boca Raton, FL.

[9]   Bauer, K. et al (2007): Statistical Models Of Accidents On Interchange Ramps And Speed- Change Lanes. U. S. DOT FHWA. Turner-Fairbanks Highway Research Center. Retrieved December 8, 2006 from the World Wide Web http://www.tfhrc.gov/ safety/pubs/97106/index.htm

[10] N. Ismail, H. Zamani, Estimation of Claim Count Data using Negative Binomial, Generalized Poisson, Zero Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression Models. Casualty Actuarial Society E Forum. (Spring 2013).

[11] Deus DK: Risk Factors and Road Traffic Accidents in Tanzania: A Case Study of Kibaha District facultate for sunfunnsvitenskap Og teknologiledelse; 2006.

[12] Qirjako G, Burazeri G, Hysa B, Roshi E: Factors Associated with Fatal Traffic Accidents in Tirana, Albania: Cross sectional Study. Croat Med J 2008, 49(6): 734–740.

[13] Haile Mekonnen Fenta, Demeke Lakew Workie. Analysis of Factors that Affect Road Traffic Accidents in Bahir Dar City, North Western Ethiopia. Science Journal of Applied Mathematics and Statistics. Vol. 2, No. 5, 2014, pp. 91-96. doi: 10.11648/j.sjams.20140205.11.

[14] Akongbota J. (2011, May 12). Reducing accidents on our roads. Retrieved December 21, 2011, from government of Ghana official portal web site: http://www.ghana.gov.gh/index.php?option=com_content&view=article&id=5864:reducingaccidents-on-our-roads&catid=24:features&Itemid=167.