Research Journal of Mathematics and Statistics 8(1): 1-7, 2016
DOI:10.19026/rjms.8.2827
ISSN: 2042-2024, e-ISSN: 2040-7505

Submitted: November 17, 2015 Accepted: February 3, 2016 Published: March 25, 2016

Research Article

Logistic Regression Approach to Modelling Road Traffic Casualties in Ghana

[1]Christian A. Hesse, [1]John B. Ofosu and [2]Francis T. Oduro
[1]Department of Mathematical Sciences, Faculty of Informatics and Mathematical Sciences, Methodist University College Ghana, P.O. Box DC 940, Dansoman-Accra, Ghana
[2]Department of Mathematics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Abstract: In this study, we shall derive a logistic regression model for predicting the annual distribution of the proportion of road traffic casualties who die as a result of road traffic accidents in Ghana. Road traffic casualties are defined as road traffic victims who are injured or killed within 30 days of the accident. With 1991 as our reference year, we considered ten independent variables that are represented by each of the 10 years from 1992 to 2001. Using a significance level of 0.05, we show that the logistic regression coefficients for the years 1993, 1998, 1999, 2000 and 2001 are significantly different from zero, while those of the remaining years are not significant. That is, there is little statistical justification for including coefficients for the years 1992, 1994, 1995, 1996 and 1998 in the model. The proposed model was used to estimate the number of road traffic fatalities from the year 2002 to 2011, a period of ten years and the results were compared with the actual fatalities. It was noted that all the calculated figures corresponding to the coefficients that were significantly different from 0 were within 10% of the actual figure and only one of the five coefficients, which were not significant, estimated road traffic fatality within 10% of its actual value.

Keywords: Casualties, fatalities, injury, logistic regression, road traffics

INTRODUCTION

The increasing population size, with a corresponding increase in the number of registered vehicles accompanied by rapidly expanding road network, has resulted in increase in Road Traffic Fatalities (RTFs). Shirley (2006) discovered that safe human behaviour is a major risk factor in accounting for Road Traffic Injuries (RTIs) especially in developing countries where it is estimated that 64 to 95% of casualties are due to improper human activity by a driver, passenger or pedestrian. Unlike many fatal diseases, road traffic accidents kill people from all age groups including young and middle-aged people in their active years. Hesse and Ofosu (2014a) reported that a cumulative total of 17 436 fatalities is recorded over a 10-year period from 2001 to 2010 where the highest fatalities during this period were in the 26-35 year old. Road traffic accidents are responsible for a far higher rate of death among men, by an approximate ratio of 3:1 (Hesse and Ofosu, 2014b).

Road traffic casualty refers to any road traffic accident victim injured or killed within 30 days of the accident. It should be pointed out that the European Economic Commission (EEC) and the World Health Organization (1979) have recommended a definition for road traffic accident fatalities which includes only deaths which occur within 30 days following the accident, since 93-97% of these fatalities take place within a one month period. A number of countries have not yet adopted this definition (World Health Organization, 1979). For example, in some countries, a road traffic fatality is recorded only if the victim dies at the site or is dead upon arrival at the hospital. In order to make comparison of accident statistics between countries reasonable, figures obtained from countries which have not adopted the 30-day fatality definition, should be properly adjusted. No adjustment is required for figures from countries such as Ghana, U.S.A and Great Britain, which have adopted the standard fatality definition.

Table 1, adapted from the National Road Safety Commission (NRSC) of Ghana, shows the annual distribution of road traffic injuries and fatalities in Ghana, from 1991 and 2013. The road traffic accident statistics in 2013 represent a reduction of 15.3% in fatalities over the 2012 figure. The fatality figure of 1 898 in 2013 is the lowest since year 2007. Relative to

Corresponding Author: Christian A. Hesse, Department of Mathematical Sciences, Faculty of Informatics and Mathematical Sciences, Methodist University College Ghana, P.O. Box DC 940, Dansoman - Accra, Ghana

Table 1: Annual distribution of road traffic fatalities and injuries in Ghana from 1991 to 2013

		Casualty		
i	Year	Fatality	Injury	Total
1	1991	920	8773	9693
2	1992	914	9116	10030
3	1993	901	7677	8578
4	1994	824	7664	8488
5	1995	1026	9106	10132
6	1996	1049	9903	10952
7	1997	1015	10433	11448
8	1998	1419	11786	13205
9	1999	1237	10202	11439
10	2000	1437	12310	13747
11	2001	1660	13178	14838
12	2002	1665	13412	15077
13	2003	1716	14469	16185
14	2004	2186	16259	18445
15	2005	1776	14034	15810
16	2006	1856	14492	16348
17	2007	2043	14373	16416
18	2008	1938	14531	16469
19	2009	2237	16259	18496
20	2010	1986	14918	16904
21	2011	2199	14020	16219
22	2012	2240	13001	15241
23	2013	1898	10611	12509
Total		36142	280527	316669
Percentage		11.41	88.59	100.00

the year 2001, the 2013 figure for fatalities (1 898) recorded an increase of 14.3%, indicating an upward trend. A cumulative total of 316 669 casualties were recorded over the 23-years period, where fatalities formed 11.4% of this figure.

According to NRSC of Ghana report, the number of road traffic crashes in 2013 (i.e., 9 200) represents a decrease of 23.9 and 18% over the 2012 and 2001 figures, respectively. The number of fatal crashes and their resulting fatalities in the previous year also saw a decrease. Compared to the 2012 figures, fatal crashes decreased in 2013 by 17% and fatalities by 15.3%. There was also a decrease of 17.9% in the overall number of casualties in 2013 compared with 2012. Relative to the year 2001, the 2013 figures for fatal accidents and fatalities recorded corresponding increases of 24.7 and 44.5%, respectively, whilst overall casualties recorded a decrease of 15.6%.

In the logistic regression analysis of this data, road traffic casualty is considered as the response or dependent variable of interest and year as predictors. The response has two categories: fatality and injury. The general objective of this analysis is to describe the way in which casualty distribution of road traffic fatalities varies by year and use this variation to predict future distribution. Logistic regression was proposed, as an alternative to ordinary least squares, in the late 1960s and early 1970s (Cabrera, 1994), and it became routinely available in statistical packages in the early 1980s. Since that time, the use of logistic regression has increased in the social sciences (e.g., Chuang, 1997; Janik and Kravitz, 1994; Tolman and Weisz, 1995) and in educational research, especially in higher education (Austin *et al*., 1992).

Other studies have been conducted in the area of road traffic casualties in Ghana. Hesse *et al*. (2014a) derived a Bayesian model for predicting the annual regional distribution of the number of road traffic fatalities in Ghana. The study showed that population and number of registered vehicles are predominant factors affecting road traffic fatalities in Ghana. Similar conclusions were arrived at when a least square regression method (Hesse *et al*., 2014b) and multilevel random coefficient method (Hesse *et al*., 2014c) were used to derive models for predicting road traffic fatalities in Ghana.

MATERIALS AND METHODS

Let n_i denote the number of road traffic casualties in the i^{th} year in Ghana and let y_i denote the number of Road Traffic Fatalities (RTFs) in the i^{th} year in Ghana. We view y_i as a value of a random variable Y_i that takes the values 0, 1, …, n_i If we assume the n_i observations for each year are independent, and they all have the same probability p_i of dying as a result of RTAs, then Y_i has the binomial distribution with parameters p_i and n_i [i.e. Yi - $B(ni, pi)$]. The probability mass function of Y_i is given by:

$$f(y_i) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}, \qquad y_i = 0, 1, \ldots, n_i \tag{1}$$

It can be shown that the expected value and variance of Y_i are (Ofosu and Hesse, 2010):

$$E(Y_i) = n_i p_i \text{ and } \operatorname{var}(Y_i) = n_i p_i (1 - p_i). \tag{2}$$

The odds_i is the ratio of the probability to its complement, or the ratio of favourable to unfavourable cases. Thus:

$$\text{odds}_i = \frac{p_i}{1 - p_i}. \tag{3}$$

We take logarithms, calculating the logit or log-odds:

$$\pi_i = \text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right), \tag{4}$$

If the logit of the underlying probability p_i is a linear function of the predictors, then we can write

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik},$$
$$= x_i \beta, \qquad i = 0, 1, \ldots, k \tag{5}$$

where,

x_i : The transpose of a vector of covariates
β : A vector of regression coefficients

Exponentiating Eq. (5) we find that the odds for the i^{th} unit are given by:

$$\frac{p_i}{1-p_i} = \exp\{x_i\beta\}. \qquad (6)$$

Solving for the probability p_i in the logit model gives:

$$p_i = \frac{\exp\{x_i\beta\}}{1+\exp\{x_i\beta\}}. \qquad (7)$$

Maximum likelihood estimation: The p.d.f. of Y_i is:

$$f(y_i) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i}, \qquad y_i = 0, 1, ..., n_i$$

The likelihood function is given by:

$$\begin{aligned} l(\beta) &= \prod_{i=0}^{k}\binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i} \\ &\propto \prod_{i=1}^{k}\left(\frac{p_i}{1-p_i}\right)^{y_i}(1-p_i)^{n_i} \\ &\propto \prod_{i=0}^{k} e^{x_i\beta y_i}\left(\frac{1}{1+e^{x_i\beta}}\right)^{n_i} = \prod_{i=0}^{k} e^{x_i\beta y_i}\left(1+e^{x_i\beta}\right)^{-n_i} \end{aligned} \qquad (8)$$

The maximum likelihood estimates of $\beta_0, \beta_1, ..., \beta_k$ are the values of $\beta_0, \beta_1, ..., \beta_k$ which maximize the likelihood function. They are also the values of $\beta_0, \beta_1, ..., \beta_k$ which maximize:

$$L(\beta) = \ln l(\beta) = \sum_{i=0}^{k} x_i\beta y_i - \sum_{i=0}^{k} n_i \ln\left(1+e^{x_i\beta}\right) \qquad (9)$$

The first derivative of $x_i\beta$ with respect to β_j is x'_{ij}, thus:

$$\begin{aligned} \frac{\partial L}{\partial \beta_j} &= \sum_{i=0}^{k} y_i x'_{ij} - \sum_{i=0}^{k} n_i\left(\frac{e^{x_i\beta}}{1+e^{x_i\beta}}\right)x'_{ij} = \sum_{i=0}^{k} y_i x'_{ij} - \sum_{i=0}^{k} n_i p_i x'_{ij} \\ &= \sum_{i=0}^{k} y_i x'_{ij} - \sum_{i=0}^{k} \mu_i x'_{ij} = \sum_{i=0}^{k}(y_i - \mu_i)x'_{ij} \end{aligned} \qquad (10)$$

where, $\mu_i = E(Y_i) = n_i p_i$, p_i depends on the covariates x_i and β is a vector of $(k+1)$ parameters. Setting each partial derivative in (10) to zero, and replacing $\beta_0, \beta_1, ..., \beta_k$ by $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ we obtain the maximum likelihood estimates of $\beta_0, \beta_1, ..., \beta_k$. The methods of solution are iterative in nature and have been programmed into logistic regression software. The interested reader may consult the text by McCullagh and Nelder (1989) for a general discussion of the methods used by most programs. The second derivatives used in computing the standard errors of the parameter estimates, $\hat{\beta}$, are:

$$\frac{\partial L^2}{\partial \beta_j \beta_l} = -\sum_{i=0}^{k} n_i x'_{ij}\frac{\partial}{\partial \beta_l}\left(\frac{e^{x'_{ij}\beta}}{1+e^{x'_{ij}\beta}}\right) = -\sum_{i=0}^{k} n_i p_i(1-p_i)x'_{ij}x'_{il}.$$

The comparison of observed to predicted values using the likelihood function is based on the following expression:

$$\begin{aligned} D &= -2\ln\left[\frac{\text{(likelihood of the fitted model)}}{\text{(likelihood of the saturated model)}}\right] \\ &= 2\left[\ln(\text{likelihood of the saturated model}) - \ln(\text{likelihood of the fitted model})\right] \end{aligned} \qquad (11)$$

The log-likelihood of the fitted model can be written as:

$$\begin{aligned} L(\hat{\beta}) &= \sum_{i=0}^{k}\left[y_i \ln \hat{p}_i + (n_i - y_i)\ln(1-\hat{p}_i)\right] \\ &= \sum_{i=0}^{k}\left[y_i \ln\left(\frac{\hat{y}_i}{n_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - \hat{y}_i}{n_i}\right)\right]. \end{aligned} \qquad (12)$$

For the saturated model, we replace $\hat{y}_i$ in Equation (12) by y_i. Equation (11) then becomes:

$$D = 2\sum_{i=0}^{k}\left\{y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i - \hat{y}_i}\right)\right\}, \qquad (13)$$

where,
y_i : The observed
$\hat{y}_i$: The fitted value for the i^{th} observation

In particular, to assess the significance of an independent variable, we compare the value of *D* with and without the independent variable in the equation. The change in *D* due to the inclusion of the independent variable in the model is:

$$\begin{aligned} G &= D\text{ (model without the variable)} - D\text{ (model with the variable)} \\ &= -2\ln\left[\frac{\text{(likelihood of the model without variable)}}{\text{(likelihood of the fitted model)}}\right]. \end{aligned}$$

It can be shown that, when the variable is not in the model, the maximum likelihood estimate of β_0 is $\ln(m_1/m_0)$, where $m_1 = \sum_{i=0}^{k} y_i$ and $m_0 = \sum_{i=0}^{k}(n_i - y_i)$.

Thus:

$$G = -2\left\{\frac{\ln\left[\left(\frac{m_1}{n}\right)^{m_1}\left(\frac{m_0}{n}\right)^{m_0}\right]}{\sum_{i=0}^{k}\left[y_i \ln \hat{p} + (n_i - y_i)\ln\left(1-\hat{p}\right)\right]}\right\} =$$

$$2\left\{\sum_{i=0}^{k}\left[y_i \ln \hat{p}_i + (n_i - y_i)\ln\left(1-\hat{p}_i\right)\right] - \left[m_1 \ln m_1 + m_0 \ln m_0 - n \ln n\right]\right\}, \quad (14)$$

where, $n = m_1 + m_0$. If the hypothesis that $\beta_j = 0,\ i = 1,\ 2,\ \ldots, k$ is true, then G has the chi-square distribution with k degrees of freedom (Hosmer *et al*., 2013).

RESULTS AND DISCUSSION

In this section, we illustrate the use of statistical packages in *R* to fit logistic regression models as a special case of a generalized linear model with family binomial and link logit. We first begin the analysis using nlme package in *R*. First, the data set, on road traffic casualties from 1991 to 2001 in Ghana, is loaded for analysis as shown in listing (1).

Listing (1):

```
>    rtf<-data.frame(matrix(c(1,920,8773,2,914,9116,3,901,7677,4,824,7664,5,
1026,9106,6,1049,9903,7,1015,10433,8,1419,11786,9,1237,10202,10,1437,12310,11,1660,13178),11,3,byrow=TR
UE))
>    names(rtf)<-c("year","Fatality","Injury")
>    rtf$Casualty<-rtf$Fatality+rtf$Injury
>    rtf$year<-factor(rtf$year,labels=c("1991","1992","1993","1994","1995",
     "1996","1997","1998","1999","2000","2001"))
>    rtf$Y<-cbind(rtf$Fatality,rtf$Injury)
>    rtf
year Fatality Injury Casualty Y.1 Y.2
1 1991 920 8773 9693 920 8773
2 1992 914 9116 10030 914 9116
3 1993 901 7677 8578 901 7677
4 1994 824 7664 8488 824 7664
5 1995 1026 9106 10132 1026 9106
6 1996 1049 9903 10952 1049 9903
7 1997 1015 10433 11448 1015 10433
8 1998 1419 11786 13205 1419 11786
9 1999 1237 10202 11439 1237 10202
10 2000 1437 12310 13747 1437 12310
11 2001 1660 13178 14838 1660 13178
```

Listing (2) shows the fit of the logistic regression model to the data using the glm() function in *R*.

Listing (2):

```
> logistic<-glm(Y~year,family = binomial, data = rtf)
```

The results of the application of the *R* function ‘summary (logistic)’, which presents the parameter estimate and standard errors for the model, are simplified in Table 2.

The fitted logistic equation, for the i^{th} year, is therefore given by:

$$\ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = -2.25506 - 0.04490x_{i1} + 0.11258x_{i2} \ldots + 0.18333x_{i10}, \quad (15)$$

where $x_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{ohterwise,} \end{cases}$ which gives the odds for the i^{th} year as:

$$\frac{\hat{p}_i}{1-\hat{p}_i} = \exp\{-2.25506 - 0.04490x_{i1} + 0.11258x_{i2} ... + 0.18333x_{i10}\}.$$

Thus,

$$\hat{p}_i = \frac{\exp\{-2.25506 - 0.04490x_{i1} + 0.11258x_{i2} ... + 0.18333x_{i10}\}}{1 + \exp\{-2.25506 - 0.04490x_{i1} + 0.11258x_{i2} ... + 0.18333x_{i10}\}}. \tag{16}$$

For instance, when $i = 0$, $\hat{p}_0 = \frac{\exp\{-2.25506\}}{1+\exp\{-2.25506\}} = 0.09491$, which gives the estimate of the proportion of road traffic casualties who died in the year 1991. Note that, in computing for the value of $\hat{p}_0$, $x_{i1} = x_{i2} = ... = x_{i10} = 0$. Note further that, in computing for $\hat{p}_i$, $i > 0$, the predictor x_{ij}, takes the value one (1) for $i = j$ while the remaining 9 predictors assume the value zero (0). Thus, from Table 2, when $i = 5$, :

$$\hat{p}_5 = \frac{\exp\{-2.25506 - 0.04490(0) + ... + 0.01006(1) + ... + 0.18333(0)\}}{1 + \exp\{-2.25506 - 0.04490(0) + ... + 0.01006(1) + ... + 0.18333(0)\}} = \frac{\exp\{-2.25506 + 0.01006\}}{1 + \exp\{-2.25506 + 0.01006\}} = 0.09578.$$

The remaining values of $\hat{p}_i$ are given in Table 2. The method for specifying the design variables involves setting all of them equal to 0 for the reference year (1991), and then setting a single design variable equal to 1 for each of the other groups.

The significance of the logistic regression relationship can be assessed by using the null deviance to test the hypotheses:

H_0: $\beta_j = 0, \ \forall \ j = 1, 2, ..., 10$ against
H_1: not all the $\beta_i = 0$

at 0.05 level of significance. The test statistic is:

$$G = 2\left\{\sum_{i=0}^{10}\left[y_i \ln \hat{p}_i + (n_i - y_i)\ln(1-\hat{p}_i)\right] - \left[12402 \times \ln(12402) + 110148 \times \ln(110148) - 12284 \times \ln(122550)\right]\right\}.$$

When H_0 is true, G has the chi-square distribution with 10 degrees of freedom (Hosmer *et al*., 2013). We reject H_0 at significance level 0.05 if the computed value of G is greater than $\chi^2_{0.05,10} = 18.31$. From the *R* function 'summary(logistic)', the value of the test statistic is $G_0 = 74.182$. Since 74.182, the calculated value of G, is greater than 18.31, the test is significant at the 5% level. We therefore reject the null hypothesis in this case and conclude that at least one of the 10 coefficients is different from zero.

Since the analysis indicates that the null hypothesis should be rejected at the 5% level, it means that some of the coefficients are significantly different from zero. But as to which of the coefficients are significantly different from zero, the analysis does not specify. Before concluding that any or all of the coefficients are nonzero, we may look at the univariable Wald test statistics (Hosmer *et al*., 2013):

$$W_j = \frac{\beta_j}{se(\beta_j)}. \tag{17}$$

These are shown in the seventh row of Table 2, labeled *z*. Under the hypothesis that the i^{th} coefficient is zero, W_i has the standard normal distribution. The eighth row of Table 2 shows the p-values which are computed under this hypothesis. The coefficients for the years 1993, 1998, 1999, 2000 and 2001 are different from zero, at 0.05 level

Table 2: Parameter estimates for logistic model of road traffic fatalities in Ghana from 1991 to 2001

j	0	1	2	3	4	5	6	7	8	9	10
Years	1991 Intercept	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Estimates $\hat{\beta}_j$	-2.25506	-0.04490	0.11258	0.02494	0.07179	0.01006	-0.07502	0.13810	0.14517	0.10721	0.18333
Standard errors	0.03465	0.04904	0.04941	0.05045	0.04781	0.04749	0.04777	0.04462	0.04591	0.04448	0.04335
$Odds_i$	0.10487	0.10026	0.11736	0.10752	0.11267	0.10593	0.09729	0.12040	0.12125	0.11673	0.12597
$\hat{p}_i$	0.09491	0.09113	0.10504	0.09708	0.10126	0.09578	0.08866	0.10746	0.10814	0.10453	0.11188
z	-65.07200	-0.91600	2.27900	0.49400	1.50200	0.21200	-1.57100	3.09500	3.16200	2.41000	4.22900
p-value	2×10^{-16}	0.35992	0.02269	0.62102	0.13315	0.83223	0.11629	0.00197	0.00157	0.01593	2.35×10^{-5}

Table 3: Design variables for year 2004

Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Variables for 2004	$x_{31} = 0$	$x_{32} = 0$	$x_{33} = 2$	$x_{34} = 0$	$x_{35} = 0$	$x_{36} = 0$	$x_{37} = 0$	$x_{32} = 0$	$x_{39} = 0$	$x_{3,10} = 0$
$\hat{\beta}_j$	-0.04490	0.11258	0.02494	0.07179	0.01006	-0.07502	0.13810	0.14517	0.10721	0.18333

Table 4: Comparison of actual fatalities and fatalities estimated from Equation (16)

j	1	2	3	4	5	6	7	8	9	10
Year	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
$\hat{\beta}_j$	-0.04490	0.11258	0.02494	0.07179	0.01006	-0.07502	0.13810	0.14517	0.10721	0.18333
D	1665	1716	2186	1776	1856	2043	1938	2237	1986	2199
$\hat{D}$	1318.9	1879.1	1831.3	1707.3	1580.1	1359.0	2000.0	2274.2	1944.0	2131.6
Error	346.1	-163.1	354.7	68.7	275.9	684.0	-62.0	-37.2	42.0	67.4
Error %	20.8	9.5	16.2	3.9	14.9	33.5	3.2	1.7	2.1	3.1

of significance, while those of the remaining years (1992, 1994, 1995, 1996 and 1998) are not significantly different from zero.

According to Hosmer *et al.* (2013), the decision to include variables in a model cannot be base entirely on tests of statistical significance. The choice of variables in the model may be influenced by other considerations. It is possible for the coefficient of some variables to be zero at certain level of significance, but when taken collectively, considerable confounding can be present in the data (Rothman *et al.*, 2008; Maldonado and Greenland, 1993; Greenland, 1989; Miettinen, 1976).

The purpose of analysing these data is not the determination of the parameters. Interest is centered on how good the model is in estimating future road traffic fatality values using these estimates. At this stage, we wish to use the model in Eq. (15) to estimate the number of road traffic fatalities from the years 2002 to 2011, a period of ten years. To do this, a single design variable x_{ij}, for year i, is set equal to 2 when $i = j$ and then all remaining variables are set equal to 0, where i represents any of the years from 2002 to 2011. We use $x_{i1}, x_{i2}, ..., x_{i10}$ in Eq. (15) as our design variables for the years 2002, 2003, and 2011, respectively. For instance, in year 3 (i.e., the year 2004), the design variables together with the corresponding parameter estimates are given in Table 3.

Thus, a point estimate of the proportion of road traffic casualties who died in 2004 is given by Eq. (16):

$$\hat{p}_{2004} = \frac{\exp\{-2.25506 + 0.02494(2)\}}{1 + \exp\{-2.25506 + 0.02494(2)\}} = 0.09929.$$

From Table 1, the total number of road traffic casualties in 2004 is 18 445. Thus a point estimate of the total number of road traffic fatalities in 2004 is (to the nearest whole number):

$$\hat{D}_{2004} = 18\,445 \times 0.099\,29 = 1831.$$

The actual road traffic fatalities D together with the values of $\hat{D}$ calculated from Eq. (16) are given in Table 4. The percentage differences between the calculated and actual values are also given in Table 4.

It can be seen that all the calculated figures, $\hat{D}$, corresponding to the coefficients, β_j, that were significantly different from 0, are within 10% of the actual figure and only one (i.e., 0.07179) of the five coefficients, that were not significantly different from 0, estimated D within 10% (i.e., 3.9%) of its actual value.

CONCLUSION

Logistic regression analysis of road traffic fatalities in Ghana has been performed using road traffic accident data from the National Road Safety Commission. The data span from 1991 to 2001. The formula for predicting the proportion of road traffic casualties who die in the i^{th} year using a logistic regression approach is:

$$\ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + ... + \hat{\beta}_{10} x_{i10}, \quad i = 0, 2, ..., 10,$$

where, the values of the parameters $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_{10}$ are given in Table 2. Using the model to estimate the number of road traffic fatalities from 2002 to 2011 in Ghana, it was noted that of the 10 calculated figures, 6 are within 10% of the actual figure.

REFERENCES

Austin, J.T., R.A. Yaffee and D.E. Hinkle, 1992. Logistic Regression for Research in Higher Education. In: Smart, J.C. (Ed.), Higher Education: Handbook of Theory and Research. Springer, New York, 8: 379-410.

Cabrera, A.F., 1994. Logistic Regression Analysis in Higher Education: An Applied Perspective. In: Smart, J.C. (Ed.), Higher Education: Handbook of Theory and Research. Vol. 10, Agathon Press, pp: 225-256.

Chuang, H.L., 1997. High school youth's dropout and re-enrollment behaviour. Econ. Educ. Rev., 16(2): 171-186.

Greenland, S., 1989. Modeling variable selection in epidemiologic analysis. Am. J. Public Health, 79(3): 340-349.

Hesse, C.A. and J.B. Ofosu, 2014a. A moving average analysis of the age distribution and the pattern of road traffic fatalities in Ghana, from 2001 - 2010. Math. Theory Modeling, 4(3): 41-50.

Hesse, C.A. and J.B. Ofosu, 2014b. Epidemiology of road traffic accidents in Ghana. Eur. Sci. J., 10(9): 370-381.

Hesse, C.A., J.B. Ofosu and F.T. Oduro, 2014a. A bayesian model for predicting road traffic fatalities in Ghana. Math. Theory Modeling, 4(8): 1-9.

Hesse, C.A., J.B. Ofosu and B.L. Lamptey, 2014b. A regression model for predicting road traffic fatalities in Ghana. Open Sci. Reposit. Math., Online (open-access): e23050497. DOI: 10.7392/openaccess.23050497.

Hesse, C.A., J.B. Ofosu and F.T. Oduro, 2014c. A multilevel model for predicting road traffic fatalities in Ghana. Eur. Sci. J., 4(8): 81-99.

Hosmer, D.W., S. Lemeshow and R.X. Sturdivant, 2013. Applied Logistic Regression. 3rd Edn., Wiley, Hoboken, New Jersey.

Janik, J. and H.M. Kravitz, 1994. Linking work and domestic problems with police suicide. Suicide Life-Threat., 24(3): 267-274.

Maldonado, G. and S. Greenland, 1993. Interpreting model coefficients when the true model form is unknown. Epidemiology, 4(4): 310-318.

McCullagh, P. and J.A. Nelder, 1989. Generalized Linear Models. 2nd Edn., Chapman and Hall, London, NY, Washington.

Miettinen, O.S., 1976. Stratification by a multivariate confounder score. Am. J. Epidemiol., 104(6): 609-620.

Ofosu, J.B. and C.A. Hesse, 2010. Introduction to Probability and Probability Distributions. EPP Books Services, Accra.

Rothman, K.J., S. Greenland and T.L. Lash, 2008. Modern Epidemiology. 3rd Edn., Wolters Kluwer Health/Lippincott Williams and Wilkins, Philadelphia.

Shirley M., 2006. Road traffic accidents: A challenging epidemic. Sultan Qaboos Univ. Med. J., 6(1): 3-5.

Tolman, R.M. and A. Weisz, 1995. Coordinated community intervention for domestic violence: The effects of arrest and prosecution on recidivism of woman abuse perpetrators. Crime Delinquency, 41(4): 481-495.

World Health Organization, 1979. Road Traffic Accidents Statistics. Regional Office for Europe, Copenhagen.