Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.8, 2014
www.iiste.org

A Bayesian Model for Predicting Road Traffic Fatalities in Ghana

Christian A. Hesse [1*] John B. Ofosu [2] Francis T. Oduro[3]

[1, 2] Department of Mathematics and Statistics, Faculty of Social Studies, Methodist University College Ghana. P. O. Box DC 940, Dansoman - Accra Ghana

[3] Department of Mathematics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi Ghana

* E-mail of the corresponding author: akrongh@yahoo.com

Abstract

Bayesian model for predicting the annual regional distribution of the number of road traffic fatalities in Ghana is derived, using road traffic accident statistics data from the National Road Safety Commission, Ghana Statistical Service and Driver and Vehicle Licensing Authority. The data span 1991 to 2009. Since the parameters are assumed to vary across the various regions, they are considered to be random variables with probability distributions. The Markov Chain Monte Carlo (MCMC) sampling techniques were used to draw samples from each of the posterior distribution, thereby determining the values of the unknown parameters for each region based on a given data.

The study has shown that population and number of registered vehicles are predominant factors affecting road traffic fatalities in Ghana. The effect of other additional factors on road traffic fatality such as human error (due to the driver, passenger and/or pedestrian), vehicle (its condition and maintenance), environmental/weather and nature of the road cannot be ruled out.

Key Words: Road traffic accidents, Fatality, Bayesian and Estimation

1. Introduction

Hesse et al. (2014) proposed a regression model for estimating road traffic fatalities in which the formula

$$D/P = \alpha\left(N/P\right)^{\beta} u \quad \ldots\ldots(1)$$

gave a fairly good fit to the data from Ghana. In Equation (1), α and β are fixed unknown parameters which are estimated using the method of least squares and D is the number of road traffic fatalities, N is the number of registered vehicles and P is the population size. Based on a random sample, inferences were made about α and β. The results obtained by Hesse et al. (2014) were consistent with other reported studies by Smeed (1949), Bener and Ofosu (1991), Jacobs and Bardsley (1977), Fouracre and Jacobs (1977), Ghee et al. (1997) in which an expression of the form

$$D/N = \hat{\alpha}\left(N/P\right)^{\hat{\beta}} \quad \ldots\ldots(2)$$

was used for the estimation of fatalities. The parameters α and β are safety and hazard indices, respectively, of a given country (Hakkert et al., 1976).

In this study, we consider the situation where, before the sample is taken, some information about α and β is known. It is assumed that, this information about α and β can be expressed in the form of a probability distribution. Thus, it is assumed that the unknown parameters α and β are values of some random variables with probability density functions $p(\alpha)$ and $q(\beta)$, respectively (Ofosu et al., 2014).

As observed from the results of Hesse et al. (2014), the number of road traffic fatalities recorded in Ghana is predominantly influenced by the number of registered vehicles and the population size. In this paper, we consider the problems that investigate the relationship between road traffic fatality and any of the 10 geographical regions of Ghana. Ghana is divided into the following ten administrative/geographical regions (regional capitals are in parentheses):

1. Greater Accra Region (Accra),
2. Ashanti Region (Kumasi),
3. Western Region (Sekondi-Takoradi),
4. Eastern Region (Koforidua),
5. Central Region (Cape Coast),
6. Volta Region (Ho),
7. Northern Region (Tamale),
8. Upper East Region (Bolgatanga),
9. Upper West Region (Wa),
10. Brong Ahafo Region (Sunyani).

In this Bayesian analysis, the population consists of the ten regions in Ghana and the number of road traffic fatalities within these regions, each year. The sampling procedure consists of selecting 19 years and the observed number of road traffic fatalities within each region.

This study seeks to address the problem of predicting a future observable road traffic fatality D_{ij} in the i^{th} year recorded in the j^{th} region of Ghana.

2. Estimation of road traffic fatalities in Ghana, using Bayesian analysis

In order to obtain the formula for the estimation of D_{ij}, the number of road traffic fatalities in the i^{th} year recorded in the j^{th} region of Ghana, a relation of the form

$$D_{ij}/P_{ij} = \nu_j \left(N_{ij}/P_{ij}\right)^{\beta_j} u_{ij} \quad \text{.....(3)}$$

is adopted, where ν_j and β_j are parameters to be estimated. In Equation (3), N_{ij} is the number of registered vehicles in the i^{th} year recorded in the j^{th} region, P_{ij} is the population size in the i^{th} year recorded in the j^{th} region and u_{ij} is such that $\varepsilon_{ij} = \ln u_{ij}$ is $N(0,\ \sigma^2)$.

A separate regression model can be fitted to each of the ten geographical regions for the purpose of forecasting the annual fatality. Thus, within the j^{th} geographical region of Ghana, in the i^{th} year, we have the following regression equation:

$$y_{ij} = \alpha_j + \beta_j x_{ij} + \varepsilon_{ij}, \quad j = 1,\ 2,\ ...,\ 10. \quad \text{.....(4)}$$

where $\alpha_j = \ln \nu_j$, $x_{ij} = \ln\left(N_{ij}/P_{ij}\right)$, $y_{ij} = \ln\left(D_{ij}/P_{ij}\right)$ and $\varepsilon_{ij} = \ln u_{ij}$, $i = 1,\ 2,\ \ldots,19.$ Thus, y_{ij} is a value of the random variable Y_{ij}. For each region j, we assume that Y_{ij} has the normal distribution with mean $\alpha_j + \beta_j x_{ij}$ and variance σ^2, where σ^2 is the pooled estimate of the population variance. Thus, the p.d.f. of Y_{ij} is given by

$$f_{Y_{ij}}\left(y_{ij}\,\middle|\,\alpha_j,\ \beta_j\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\left[y_{ij} - \left(\alpha_j + \beta_j x_{ij}\right)\right]^2}{2\sigma^2}\right\} \quad \text{.....(5)}$$

The likelihood function, for 19 years data, in j^{th} region is given by

$$p\left(\boldsymbol{y_j}\,\middle|\,\alpha_j,\ \beta_j\right) = \prod_{i=1}^{19} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\left[y_{ij} - \left(\alpha_j + \beta_j x_{ij}\right)\right]}{2\sigma^2}\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{19}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{19}\left(y_{ij} - \alpha_j - \beta_j x_{ij}\right)^2\right], \quad \text{.....(6)}$$

where $\boldsymbol{y_j} = (y_{1j},\ y_{2j},\ ...,\ y_{19j})$. In Hesse et al. (2014), the least square estimates of α and β, based on the national data, were found to be approximately $\hat{\alpha} = -8.3$ and $\hat{\beta} = 0.3$. Thus, it is assumed that the parameters α_j and β_j are independently and normally distributed, where α_j is $N(-8.3,\ 1^2)$ and β_j is $N(0.3,\ 0.1^2)$. Therefore, the probability density functions of α_j and β_j are, respectively, given by

$$p\left(\alpha_j\right) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\alpha_j + 8.3\right)^2\right\}, \quad p\left(\beta_j\right) = \frac{1}{\sqrt{0.02\pi}} \exp\left\{-\frac{1}{0.02}\left(\beta_j - 0.3\right)^2\right\}. \quad \text{.....(7)}$$

The model has been fully specified, where we have a likelihood as well as a prior on each parameter (Ofosu et al., 2014). The posterior distribution over the parameters conditional on the data is given by

$$\underbrace{p\left(\theta_j\,\middle|\,\boldsymbol{y}_j\right)}_{posterior} = c \times \underbrace{f\left(\boldsymbol{y}_j\,\middle|\,\theta_j\right)}_{likelihood} \times \underbrace{p\left(\theta_j\right)}_{prior}.$$

where $\theta_j = \left(\alpha_j,\ \beta_j\right)$ and c is a constant. Therefore, we can write:

$$\underbrace{p\left(\alpha_j, \beta_j \middle| \boldsymbol{y}_j\right)}_{posterior} = c\, \underbrace{p\left(\boldsymbol{y}_j \middle| \alpha_j, \beta_j\right)}_{likelihood} \underbrace{p\left(\alpha_j\right) p\left(\beta_j\right)}_{prior}.$$

From Equations (6) and (7), this lead to

$$p\left(\alpha_j, \beta_j \middle| \boldsymbol{y}_j\right) = k \exp\left[-\tfrac{1}{2\sigma^2}\sum_{i=1}^{19}\left(y_i - \alpha_j - \beta_j x\right)^2 - \tfrac{1}{2}\left(\alpha_j + 8.3\right)^2 - \tfrac{1}{0.02}\left(\beta_j - 0.3\right)^2\right], \quad \ldots\ldots\ldots(8)$$

where $k = c\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{19}{2}} \times \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{0.02\pi}}$.

In Bayesian data analysis, one way to apply a model to data is to find the maximum a posteriori (MAP) parameter values (Steyvers, 2011). The goal here is to find the parameter estimates that maximize the posterior probability of the parameters given the data. In other words, we find the mode of the posterior distribution. This corresponds to:

$$\theta_{MAP} = \underset{\theta_j}{\arg\max}\; p\left(\theta_j \middle| \boldsymbol{y}\right), \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(9)$$

The Bayesian approach that we are really interested in is posterior sampling. With the MAP approach, we get a single set of parameter values for a model. Therefore, we are characterizing the posterior distribution with the mode of this distribution (Hox, 2010).

The Markov Chain Monte Carlo (MCMC) sampling techniques are used to draw samples from the posterior distribution. These samples can be used to calculate a number of things, such as means, variances and other moments of the distribution. We can also check whether there are any correlations between parameters. Because the conditional distributions in Equation (8) do not correspond to any analytic expression, the bivariate component-wise Metropolis-Hastings sampler is therefore used to get samples from the posterior distribution (Martinez and Martinez, 2002).

Table 1 shows the observable number, D_{ij}, of road traffic fatality in the i^{th} year recorded in the j^{th} region, the number, N_{ij}, of registered vehicles in the i^{th} year recorded in the j^{th} region and P_{ij}, the population size in the i^{th} year recorded in the j^{th} region of Ghana from 1991 to 2011.

Table 1: Regional distribution of the number of road traffic fatalities, registered vehicles and estimated population size from 1991 to 2011

	Greater Accra 1			Ashanti 2			Western 3			Eastern 4			Central 5		
Year	D_{i1}	N_{i1}	P_{i1}	D_{i2}	N_{i2}	P_{i2}	D_{i3}	N_{i3}	P_{i3}	D_{i4}	N_{i4}	P_{i4}	D_{i5}	N_{i5}	P_{i5}
1991	126	81382	1934520	183	21394	2641258	65	4485	1443424	183	3476	1852699	98	2226	1321216
1992	164	85027	2019639	153	22353	2731061	90	4686	1489614	204	3632	1878637	122	2326	1348961
1993	115	97240	2108503	168	25563	2823917	108	5359	1537282	207	4153	1904938	97	2660	1377289
1994	155	119066	2201277	161	31301	2919930	49	6562	1586475	186	5086	1931607	123	3257	1406212
1995	190	144805	2298133	174	38068	3019208	104	7981	1637242	192	6185	1958650	128	3961	1435743
1996	191	183331	2399251	175	48196	3121861	105	10104	1689634	196	7830	1986071	130	5014	1465893
1997	174	210101	2504818	220	55233	3228004	111	11580	1743702	181	8974	2013876	131	5747	1496677
1998	258	242341	2615030	283	63709	3337756	127	13356	1799500	291	10351	2042070	146	6628	1528107
1999	172	282373	2730091	178	74233	3451240	104	15563	1857084	294	12061	2070659	165	7723	1560198
2000	196	314963	2905726	280	82800	3612950	111	17359	1924577	295	13453	2106696	185	8615	1593823
2001	239	349917	2995804	350	91989	3710500	146	19285	1963069	296	14946	2150937	206	9571	1643232
2002	239	377880	3088673	351	99341	3810683	146	20827	2002330	297	16140	2196106	207	10336	1694172
2003	240	396783	3184422	360	104310	3913572	146	21868	2042377	298	16947	2242225	208	10853	1746691
2004	299	433482	3283139	565	113957	4019238	158	23891	2083224	325	18515	2289311	234	11857	1800838
2005	306	472736	3384917	314	124277	4127757	154	26054	2124889	299	20191	2337387	183	12930	1856664
2006	325	518494	3489849	340	136306	4239207	155	28576	2167386	305	22146	2386472	190	14182	1914221
2007	370	568681	3598034	376	149500	4353665	156	31342	2210734	305	24289	2436588	190	15555	1973562
2008	385	580546	3709574	416	152619	4471214	169	31996	2254949	294	24796	2487756	150	15879	2034742
2009	420	634779	3824570	440	166876	4591937	180	34985	2300048	320	27112	2539999	220	17362	2097819
2010	424	691909	4010054	454	181895	4780380	157	38134	2376021	259	29553	2633154	167	18925	2201863
2011	425	755421	4134366	460	198592	4909450	190	41634	2423541	260	32265	2688450	190	20662	2270121
	Volta 6			Northern 7			Upper East 8			Upper West 9			Brong-Ahafo 10		
Yea	D_{i6}	N_{i6}	P_{i6}	D_{i7}	N_{i7}	P_{i7}	D_{i8}	N_{i8}	P_{i8}	D_{i9}	N_{i9}	P_{i9}	D_{i10}	N_{i10}	P_{i10}

1991	92	2008	1382575	41	5653	1412935	23	4037	834245	13	3651	513584	96	3738	1444102
1992	50	2098	1408844	30	5906	1452497	32	4218	843422	8	3814	525396	61	3906	1481648
1993	59	2399	1435612	17	6755	1493167	14	4824	852700	16	4362	537481	100	4467	1520171
1994	27	2938	1462888	31	8271	1534976	20	5907	862079	3	5341	549843	69	5469	1559695
1995	80	3573	1490683	38	10059	1577955	21	7184	871562	13	6496	562489	86	6652	1600248
1996	85	4524	1519006	40	12735	1622138	26	9095	881149	14	8224	575426	87	8422	1641854
1997	43	5184	1547867	35	14594	1667558	14	10423	890842	6	9425	588661	100	9651	1684542
1998	91	5980	1577277	61	16834	1714250	26	12023	900641	16	10871	602200	120	11132	1728340
1999	72	6968	1607245	76	19615	1762249	30	14009	910548	22	12667	616051	124	12971	1773277
2000	89	7772	1635421	78	21878	1820806	48	15625	920089	25	14129	576583	130	14468	1815408
2001	135	8634	1676307	79	24306	1873609	34	17360	931130	26	15697	587538	149	16074	1857162
2002	135	9324	1718214	80	26249	1927944	34	18747	942304	26	16951	598701	150	17359	1899877
2003	140	9791	1761170	90	27562	1983854	45	19685	953611	35	17799	610077	154	18227	1943574
2004	167	10696	1805199	131	30111	2041386	68	21505	965055	37	19446	621668	202	19913	1988277
2005	122	11665	1850329	97	32838	2100586	79	23453	976635	30	21207	633480	192	21716	2034007
2006	169	12794	1896587	112	36016	2161503	82	25723	988355	34	23259	645516	244	23818	2080789
2007	170	14032	1944002	113	39502	2224187	83	28213	1000215	35	25511	657781	245	26123	2128647
2008	179	14325	1992602	95	40327	2288688	59	28801	1012218	36	26043	670279	155	26668	2177606
2009	180	15663	2042417	113	44094	2355060	65	31492	1024364	40	28476	683014	259	29160	2227691
2010	143	17073	2118252	114	48062	2479461	45	34326	1046545	54	31039	702110	169	31784	2310983
2011	144	18640	2171208	123	52474	2551365	54	37477	1067476	56	33888	715450	297	34701	2364136

Table 2, computed from Table 1, shows the values of $x_{ij} = \ln\left(N_{ij}/P_{ij}\right)$ and the corresponding values of $y_{ij} = \ln\left(D_{ij}/P_{ij}\right)$ for the ten regions of Ghana.

Table 2: Value of $y_{ij} = \ln\left(D_{ij}/P_{ij}\right)$ and $x_{ij} = \ln\left(N_{ij}/P_{ij}\right)$ from 1991 – 2009

	Greater Accra 1		Ashanti 2		Western 3		Eastern 4		Central 5		Volta 6		Northern 7		Upper East 8		Upper West 9		Brong Ahafo 10	
Year	x_{i1}	y_{i1}	x_{i2}	y_{i2}	x_{i3}	y_{i3}	x_{i4}	y_{i4}	x_{i5}	y_{i5}	x_{i6}	y_{i6}	x_{i7}	y_{i7}	x_{i8}	y_{i8}	x_{i9}	y_{i9}	x_{i10}	y_{i10}
1991	-3.17	-9.64	-4.82	-9.58	-5.77	-10.01	-6.28	-9.22	-6.39	-9.51	-6.53	-9.62	-5.52	-10.45	-5.33	-10.50	-4.95	-10.58	-5.96	-9.62
1992	-3.17	-9.42	-4.81	-9.79	-5.76	-9.71	-6.25	-9.13	-6.36	-9.31	-6.51	-10.25	-5.51	-10.79	-5.30	-10.18	-4.93	-11.09	-5.94	-10.10
1993	-3.08	-9.82	-4.70	-9.73	-5.66	-9.56	-6.13	-9.13	-6.25	-9.56	-6.39	-10.10	-5.40	-11.38	-5.17	-11.02	-4.81	-10.42	-5.83	-9.63
1994	-2.92	-9.56	-4.54	-9.81	-5.49	-10.39	-5.94	-9.25	-6.07	-9.34	-6.21	-10.90	-5.22	-10.81	-4.98	-10.67	-4.63	-12.12	-5.65	-10.03
1995	-2.76	-9.40	-4.37	-9.76	-5.32	-9.66	-5.76	-9.23	-5.89	-9.33	-6.03	-9.83	-5.06	-10.63	-4.80	-10.63	-4.46	-10.68	-5.48	-9.83
1996	-2.57	-9.44	-4.17	-9.79	-5.12	-9.69	-5.54	-9.22	-5.68	-9.33	-5.82	-9.79	-4.85	-10.61	-4.57	-10.43	-4.25	-10.62	-5.27	-9.85
1997	-2.48	-9.57	-4.07	-9.59	-5.01	-9.66	-5.41	-9.32	-5.56	-9.34	-5.70	-10.49	-4.74	-10.77	-4.45	-11.06	-4.13	-11.49	-5.16	-9.73
1998	-2.38	-9.22	-3.96	-9.38	-4.90	-9.56	-5.28	-8.86	-5.44	-9.26	-5.58	-9.76	-4.62	-10.24	-4.32	-10.45	-4.01	-10.54	-5.05	-9.58
1999	-2.27	-9.67	-3.84	-9.87	-4.78	-9.79	-5.15	-8.86	-5.31	-9.15	-5.44	-10.01	-4.50	-10.05	-4.17	-10.32	-3.88	-10.24	-4.92	-9.57
2000	-2.22	-9.60	-3.78	-9.47	-4.71	-9.76	-5.05	-8.87	-5.22	-9.06	-5.35	-9.82	-4.42	-10.06	-4.08	-9.86	-3.71	-10.05	-4.83	-9.54
2001	-2.15	-9.44	-3.70	-9.27	-4.62	-9.51	-4.97	-8.89	-5.15	-8.98	-5.27	-9.43	-4.34	-10.07	-3.98	-10.22	-3.62	-10.03	-4.75	-9.43
2002	-2.10	-9.47	-3.65	-9.29	-4.57	-9.53	-4.91	-8.91	-5.10	-9.01	-5.22	-9.45	-4.30	-10.09	-3.92	-10.23	-3.56	-10.04	-4.70	-9.45
2003	-2.08	-9.49	-3.62	-9.29	-4.54	-9.55	-4.89	-8.93	-5.08	-9.04	-5.19	-9.44	-4.28	-10.00	-3.88	-9.96	-3.53	-9.77	-4.67	-9.44
2004	-2.02	-9.30	-3.56	-8.87	-4.47	-9.49	-4.82	-8.86	-5.02	-8.95	-5.13	-9.29	-4.22	-9.65	-3.80	-9.56	-3.46	-9.73	-4.60	-9.19
2005	-1.97	-9.31	-3.50	-9.48	-4.40	-9.53	-4.75	-8.96	-4.97	-9.22	-5.07	-9.63	-4.16	-9.98	-3.73	-9.42	-3.40	-9.96	-4.54	-9.27
2006	-1.91	-9.28	-3.44	-9.43	-4.33	-9.55	-4.68	-8.97	-4.91	-9.22	-5.00	-9.33	-4.09	-9.87	-3.65	-9.40	-3.32	-9.85	-4.47	-9.05
2007	-1.84	-9.18	-3.37	-9.36	-4.26	-9.56	-4.61	-8.99	-4.84	-9.25	-4.93	-9.34	-4.03	-9.89	-3.57	-9.40	-3.25	-9.84	-4.40	-9.07
2008	-1.85	-9.17	-3.38	-9.28	-4.26	-9.50	-4.61	-9.04	-4.85	-9.52	-4.94	-9.32	-4.04	-10.09	-3.56	-9.75	-3.25	-9.83	-4.40	-9.55
2009	-1.80	-9.12	-3.31	-9.25	-4.19	-9.46	-4.54	-8.98	-4.79	-9.16	-4.87	-9.34	-3.98	-9.94	-3.48	-9.67	-3.18	-9.75	-4.34	-9.06

The analysis assumes that the observations, y_{ij}, from each region, are normally and independently distributed. In practice, these assumptions will usually not hold exactly. Thus, we check the normality assumption using Shapiro-Wilk *W* test. We test the null hypothesis as follows.

H_0: observations y_{ij}, from each region come from a normally distributed population

against the alternative hypothesis

H_1: observations from each region are not from a normally distributed population .

The computation of the *W* test statistics for Greater Accra, given a sample of size $n = 19,$ is as follows:

1. Order the observations to obtain an ordered sample $y_{(1)} \le y_{(2)} \le ... \le y_{(19)}$, as shown in Table 3.
2. Compute
$$s^2 = \sum_{i=1}^{19} y_i^2 - \frac{1}{19}\left(\sum_{i=1}^{19} y_i\right)^2 = 0.6466.$$
3. Compute
$$b = \sum_{i=1}^{10} a_{20-i}\left(y_{20-i} - y_i\right) = 0.7955.$$
4. Compute
$$W = b^2/s^2 = 0.9787.$$

We reject H_0 at the 0.05 level of significance if the computed value of W is less than 0.901, the tabulated 5% point of the distribution of the Shapiro-Wilk test statistic. Since 0.9787 is greater than 0.911, we fail to reject H_0 at 0.05 level of significance and conclude that there is enough evidence that the observations from Greater Accra are normally distributed.

Table 3: Computation of the *W* test statistics

i	y_i	y_{20-i}	$y_{20-i} - y_i$	a_{20-i}	$a_{20-i}\left(y_{20-i} - y_i\right)$
1	-9.8166	-9.1167	0.6999	0.4808	0.3365
2	-9.6724	-9.1732	0.4992	0.3232	0.1613
3	-9.6391	-9.1824	0.4567	0.2561	0.1170
4	-9.6041	-9.2238	0.3803	0.2059	0.0783
5	-9.5747	-9.2815	0.2932	0.1641	0.0481
6	-9.5611	-9.3039	0.2572	0.1271	0.0327
7	-9.4931	-9.3113	0.1818	0.0932	0.0169
8	-9.4668	-9.4006	0.0662	0.0612	0.0041
9	-9.4384	-9.4186	0.0198	0.0303	0.0006
10	-9.4363	-9.4363	0.0000	0.0000	0.0000
					$b = 0.7955$

The value of the Shapiro-Wilk W test statistic for each of the remaining nine regions is given in Table 4.

Table 4: Observed values of the *W* test statistic

Test Statistic	Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong-Ahafo
W_0	0.9787	0.9329	0.7402	0.9358	0.9570	0.8788	0.9022	0.9498	0.9187	0.9575

Since, in the Western and Volta regions, the observed values of the test statistic are less than 0.901, we reject H_0 at the 5% significance level and conclude that the samples from these two regions are from non-normally distributed populations. For each of the remaining 8 regions, we fail to reject H_0 at the 5% level of significance and conclude that there is enough evidence that the observations are normally distributed. Since the regression analysis is robust to small departures from the normality assumption, we can further conclude that the normality assumption for the Bayesian analysis is validated.

The posterior distribution is bivariate as it involves the two parameters α and β. Suitable MCMC approach is the Metropolis Hastings (MH) sampler. The following is a high-level description of the MH procedure:

1. Set $t = 1$
2. Generate an initial value for α and β.
3. Repeat
   $t = t + 1$
   Do a MH step on α
      Generate a proposal α^* from a suitable proposal normal distribution;
      Evaluate the acceptance probability a using α^*, α and β;
      Generate a u from a Uniform(0, 1) distribution
      If $u \le a$, accept the proposal and set $\alpha = \alpha^*$,
   Do a MH step on β,

Generate a proposal β^* from a suitable proposal normal distribution
Evaluate the acceptance probability a using β^*, β and α;
Generate a u from a Uniform(0, 1) distribution
If $u \le a$, accept the proposal and set $\beta = \beta^*$,

4. Until $t = T$

The following is the MATLAB code for implementation of component-wise Metropolis sampler for the posterior distribution.

Listing 1: Matlab code for function to evaluate the posterior distribution

```
function y = post(alpha,beta,s,x1,y1)
y = exp(-((alpha+8.3)^2)/2-((beta-0.3)^2)/0.02-sum(y1-alpha-beta*x1)^2/(2*s^2))
```

Listing 2: Implementation of component-wise Metropolis sampler for the posterior distribution

```
% % Metropolis procedure to sample from the posterior distribution
% Component-wise updating. Use a normal proposal distribution

s = std(y1)

% % Initialize the Metropolis sampler
T=5000; % Set the maximum number of iteration
propsigma=[1 0.1]; % standard deviation of proposal distribution
thetamin=[-11.3,0]; % define minimum for alpha and beta
thetamax=[-5.3,1]; % define maximum for alpha and beta
seed=1;rand('state',seed);randn('state',seed); % set the random seed
state=zeros(2,T); % Init storage space for the state of the sampler
alpha=unifrnd(thetamin(1),thetamax(1)); % Start value for alpha
beta=unifrnd(thetamin(2),thetamax(2)); % Start value for beta
t=1; % initialize iteration at 1
state(1,t)=alpha; % save the current state
state(2,t)=beta

% % Start sampling
while t<T; % Iterate until we have T samples
t=t+1

   % % Propose a new value for alpha
   new_alpha = normrnd(alpha,propsigma(1))
   pratio = post(new_alpha,beta,sigma,x1,y1)/...
                      post(alpha,beta,sigma,x1,y1)
   a = min([1 pratio]); % Calculate the acceptance ratio
u = rand; % Draw a uniform deviate from [0 1]
   if u < a; % Do we accept this proposal?
       alpha = new_alpha; % proposal becomes new value for alpha
   end

   % % Propose a new value for beta
   new_beta = normrnd(beta,propsigma(2))
   pratio = post(alpha,new_beta,sigma,x1,y1)/...
            post(alpha,beta,sigma,x1,y1)
   a = min([1 pratio]); % Calculate the acceptance ratio
u = rand; % Draw a uniform deviate from [0 1]
if u < a % Do we accept this proposal?
beta = new_beta; % proposal becomes new value for theta2
end

   % % Save state
   state(1,t) = alpha;
   state(2,t) = beta;
end
```

The estimated values of α, β and ν for each of the ten regions, as the result of implementation of component-wise Metropolis sampler for the posterior distribution, are given in Table 5.

Table 5: Estimated values of α, β and ν for each of the ten regions in Ghana

	Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong Ahafo
α	-8.2632	-7.0636	-5.9856	-5.6653	-5.5969	-6.4298	-7.7806	-8.6744	-7.9141	-7.0833
β	0.5159	0.6386	0.7761	0.6488	0.6700	0.5992	0.5357	0.3118	0.6086	0.4946
ν	0.000258	0.000856	0.002515	0.003464	0.003709	0.001613	0.000418	0.000171	0.000366	0.000839

For instance, in Greater Accra region, the estimated values for α and β are $\alpha_1 = -8.2632$ and $\beta_1 = 0.5159$, respectively. Therefore, the estimate for ν_j is

$$\hat{\nu}_j = e^{-8.2632} = 0.0002578. \quad \ldots\ldots(10)$$

Equation (4.12), for Greater Accra region, therefore becomes

$$D_{i1}/P_{i1} = 0.0002578\left(N_{i1}/P_{i1}\right)^{0.5159}. \quad \ldots\ldots(11)$$

The actual road traffic fatalities for Greater Accra, D_{i1}, from 1991 to 2012, together with the corresponding values of $\hat{D}_{i1}$ calculated from Equation (11), are given in Table 6. The percentage differences between the calculated and actual values are also given.

Table 6: Comparison of actual fatalities and fatalities estimated from Equation (11) for Greater Accra region

i	Year	D_{i1}	$\hat{D}_{i1}$	Error	Error %	No.	Year	D_{i1}	$\hat{D}_{i1}$	Error	Error %
1	1991	126	97	29	22.8	**12**	2002	239	269	-30	12.7
2	1992	164	102	62	38.1	**13**	2003	240	280	-40	16.8
3	1993	115	111	4	3.3	**14**	2004	299	298	1	0.4
4	1994	155	126	29	18.7	**15**	2005	306	316	-10	3.3
5	1995	190	142	48	25.1	**16**	2006	325	336	-11	3.5
6	1996	191	164	27	14.1	**17**	2007	370	358	12	3.2
7	1997	174	180	-6	3.3	**18**	2008	385	367	18	4.6
8	1998	258	198	60	23.4	**19**	2009	420	390	30	7.0
9	1999	172	218	-46	26.9	**20**	2010	424	418	6	1.5
10	2000	196	238	-42	21.5	**21**	2011	425	443	-18	4.4
11	2001	239	255	-16	6.7	**22**	2012	435	469	-35	8.0

It can be seen that, from 2004 to 2012 in Greater Accra region, all the 9 calculated figures are within 10% of the actual figure. Out of the 22 calculated figures, from 1991 to 2012, 12 are within 10% of the actual figure and 17 are within 20% of the actual value.

Table 7 shows the regional distribution of the estimated road traffic fatalities $\hat{D}_{ij}$, together with percentage difference, d_i, between the calculated and actual values of road traffic fatalities for all the ten regions of Ghana, from 1991 to 2012.

Table 7: Regional distribution of the estimated road traffic fatalities, from 1991 to 2012

		Greater Accra		Ashanti		Western		Eastern		Central		Volta		Northern		Upper East		Upper West		Bron Ahafo	
i	Year	$\hat{D}_{i1}$	d_1	$\hat{D}_{i2}$	d_2	$\hat{D}_{i3}$	d_3	$\hat{D}_{i4}$	d_4	$\hat{D}_{i5}$	d_5	$\hat{D}_{i6}$	d_6	$\hat{D}_{i7}$	d_7	$\hat{D}_{i8}$	d_8	$\hat{D}_{i9}$	d_9	$\hat{D}_{i10}$	d_{10}
1	1991	97	22.8	104	43.0	41	36.8	109	40.3	68	30.7	44	51.7	31	25.2	27	17.6	9	28.8	64	33.7
2	1992	102	38.1	109	29.0	43	52.4	113	44.6	70	42.3	46	8.1	32	6.0	28	13.7	10	19.8	66	8.0
3	1993	111	3.3	120	28.7	48	55.7	124	40.2	78	20.0	50	14.9	35	103.6	29	107.3	10	34.4	71	28.7
4	1994	126	18.7	138	14.3	56	15.1	142	23.7	89	27.3	57	111.4	39	26.0	31	55.7	12	299.2	80	15.8
5	1995	142	25.1	158	9.1	66	36.5	162	15.7	103	19.8	65	19.2	44	15.6	33	58.8	14	4.7	89	3.7
6	1996	164	14.1	186	6.4	80	23.9	190	3.3	121	6.8	75	11.7	51	26.3	36	39.1	16	13.2	101	16.6
7	1997	180	3.3	206	6.5	89	19.4	208	15.0	134	2.0	82	90.9	55	57.2	38	171.6	17	189.7	110	10.0
8	1998	198	23.4	228	19.4	101	20.7	229	21.2	148	1.4	90	1.0	60	1.4	40	54.1	19	19.5	120	0.4
9	1999	218	26.9	254	42.9	114	9.8	255	13.4	165	0.1	99	38.1	66	13.0	42	41.1	21	3.7	131	5.4
10	2000	238	21.5	277	0.9	125	12.8	275	6.8	179	3.3	107	20.2	71	8.7	44	8.1	22	11.8	140	7.3
11	2001	255	6.7	299	14.4	137	6.5	297	0.2	194	5.9	115	14.8	76	3.4	46	35.2	24	8.9	149	0.2
12	2002	269	12.7	318	9.5	146	0.3	314	5.7	206	0.4	122	9.9	81	0.8	47	39.6	25	3.8	156	4.2
13	2003	280	16.8	331	8.1	152	4.0	326	9.5	215	3.5	127	9.6	84	6.8	49	8.0	26	25.9	162	5.1
14	2004	298	0.4	353	37.4	163	3.4	348	7.2	231	1.4	135	19.3	89	32.0	50	25.9	28	25.4	171	15.3
15	2005	316	3.3	377	20.1	176	14.0	371	24.1	247	35.0	143	17.5	95	2.5	52	34.0	29	2.3	181	5.9
16	2006	336	3.5	404	18.8	189	22.2	397	30.1	265	39.7	153	9.5	101	10.1	54	34.0	31	8.2	191	21.6
17	2007	358	3.2	433	15.1	204	31.0	425	39.2	285	50.1	163	3.9	107	5.1	56	32.3	33	4.9	203	17.3
18	2008	367	4.6	443	6.4	209	23.4	433	47.4	292	94.8	167	6.7	110	15.7	57	3.4	34	5.7	207	33.6
19	2009	390	7.0	473	7.5	225	24.8	463	44.6	313	42.4	178	1.2	117	3.4	59	9.1	36	9.7	219	15.5
20	2010	418	1.5	507	11.7	242	54.0	495	91.3	337	102.0	190	32.9	125	9.9	62	36.9	38	28.8	233	37.7
21	2011	443	4.4	542	17.8	260	36.9	528	103.2	361	90.2	202	40.5	133	8.2	64	18.9	41	27.0	302	1.6
22	2012	469	8.0	576	22.4	278	43.1	561	110.9	385	98.3	214	45.6	141	11.9	67	20.8	43	24.5	321	5.7

Out of the 220 calculated figures, 1991 to 2012, about 130 are within 20% of the actual figure. In Brong Ahafo region, for instance, 11 out of the 22 calculated figures are within 10% of the actual figures and 17 are within 20% of the actual value.

3. Conclusion

Bayesian analysis of road traffic fatalities in Ghana by geographical/administrative regions has been performed using road traffic accident statistics data from the National Road Safety Commission, Ghana Statistical Service and Driver and Vehicle Licensing Authority. The data span 1991 to 2009.

The formula for predicting the number of road traffic fatalities, D_{ij} for the j^{th} region in the i^{th} year using a Bayesian approach is given by

$$D_{ij}/P_{ij} = \hat{\nu}_j \left(N_{ij}/P_{ij}\right)^{\hat{\beta}_j}, \qquad j = 1,\ 2,\ \ldots,\ 10.$$

where the values of the parameters ν_j and β_j for each region are given in Table 5.

The study has confirmed the fact that predominant factors influencing road traffic fatalities in Ghana are population and number of registered vehicles.

References

Bener, A. and Ofosu, J. B. (1991). Road traffic fatalities in Saudi Arabia. Journal of the International Association of Traffic and Safety Sciences, **15**, 35–8.

Fouracre, P. and Jecobs, G. D. (1977). Further research on road accident rate in developing countries. TRRL report LR 270. Transport and Road Research Laboratory, Crowthorne, Berkshire.

Ghee, C. Silcock, D. Astrop, A. and Jacobs, G. (1997). Socio-economic aspects of road accidents in developing countries. TRL Report 247. Crowthorne: Transport Research Laboratory.

Hesse, C. A., Ofosu, J. B., & Lamptey, B. L. (2014). A Regression Model for Predicting Road Traffic Fatalities in Ghana. *Open Science Repository Mathematics*, Online (open-access), e23050497. doi:10.7392/openaccess.23050497.

Hakkert, A., Livneh, M. and Mahalel, D. (1976). Levels of safety in accidents studies. A safety in accidents studies a Safety index. Australian Road Research Board Proc.

Hox, J. J. (2010). Multilevel analysis: Techniques and applications 2nd edition. Routledge Taylor and Francis Group.

Jacobs, G., and Bardsley, M. (1977). Research on road accidents in developing countries. Traffic engineering & control.

Martinez, W. L. and Martinez, A. R. (2002). Computational statistics handbook with MATLAB. Chapman & Hall/CRC, Boca Raton London New York Washington, D.C.

Ofosu, J. B., Hesse, C. A. and Otchere, F. (2014). Intermediate Statistical Methods. *EPP Books Services, Accra*.

Smeed, R. (1949). Some statistical aspects of road safety research. J. Roy Stats. Soc. Series-A Vol. 12, No. 1 pp. 1-23.

Steyvers, M. (2011). Computational statistics with MATLAB. University of California, Irvine, psiexp.ss.uci.edu/research/teachingP205C/205C.pd.