A MULTILEVEL MODEL FOR PREDICTING ROAD TRAFFIC FATALITIES IN GHANA

Christian A. Hesse, BSc, MPhil
John B. Ofosu, BSc, PhD, FSS
Department of Mathematics and Statistics, Faculty of Social Studies, Methodist University College Ghana
Francis T. Oduro, BSc, MPhil, PhD
Department of Mathematics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi Ghana

Abstract

A Multilevel Random Coefficient (MRC) model for predicting road traffic fatalities in Ghana is proposed. In this model, the number of road traffic fatalities and the regional groups are conceptualized as a hierarchical system of road traffic fatalities and geographical regions of Ghana, with fatalities and regions defined at separate levels of this hierarchical system. Instead of estimating a separate regression equation for each of the 10 regions in Ghana, a multilevel regression analysis was applied to estimate the values of the regression coefficients for each region based on data given. The result shows that there is significant intercept variation in terms of the dependent variable y across the 10 regions. It was estimated that about 58% of the variation in y is a function of the region to which it is observed, thus, validating the application of the multilevel model. Using the random slope model M_2, it was found that, from 2001 to 2012 in Greater Accra region, all the 12 estimated road traffic fatality figures are within 10% of the actual figure. Out of the 22 calculated figures, from 1991 to 2012, 15 are within 10% of the actual figure and 19 are within 20% of the actual value.

Keywords: Road Traffic, accident, morbidity and mortality, multilevel models

Introduction

Smeed (1949) gave a regression model for estimating road traffic fatalities. Hesse et al. (2014) derived a modified form of Smeed's regression formula for estimating road traffic fatalities in Ghana, where the regression coefficient α and β are fixed unknown parameters.

Similar to a Bayesian model, where the parameters are considered as random variables, this paper seeks to develop a Multilevel Random Coefficient (MRC) model for predicting road traffic fatalities in Ghana. In this model, the number of road traffic fatalities and the regional groups are conceptualized as a hierarchical system of road traffic fatalities and geographical regions of Ghana, with fatalities and regions defined at separate levels of this hierarchical system. One can think of MRC models as ordinary regression models that have additional variance terms for handling non-independence due to group membership.

Ghana is divided into the following ten administrative/geographical regions:

1. Greater Accra Region,
2. Ashanti Region,
3. Western Region,
4. Eastern Region,
5. Central Region,
6. Volta Region,
7. Northern Region,
8. Upper East Region,
9. Upper West Region,
10. Brong Ahafo Region.

Instead of estimating a separate regression equation for each of the 10 regions in Ghana, a multilevel regression analysis is applied to estimate the values of the regression coefficients for each region based on data given. This paper illustrates the estimation of the regression coefficient using the Linear & Nonlinear Mixed Effects (nlme) package in *R* (Pinhiero & Bates, 2000). This class of models is also often referred to as mixed-effects models (Snijders & Bosker, 1999). The key to understanding MRC models is to understand how nesting fatalities within geographical regions can produce additional sources of variance (non-independence) in data (Hox, 1998).

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

$$D_{ij}/P_{ij} = \nu_j \left(N_{ij}/P_{ij}\right)^{\beta_j} u_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

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

Taking logarithms, to base e, of both sides of Equation (1), we obtain

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

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 $\varepsilon_{ij} = \ln u_{ij}$. Thus, y_{ij} a value of the random variable Y_{ij}. For each region j, we assume that Y_{ij} has the normal distribution mean $\alpha_j + \beta_j x_{ij}$ and variance σ^2. Table 1 shows the observable number of road traffic fatality D_{ij}, the number of registered vehicles N_{ij}, and the estimated population size P_{ij}, for each region in Ghana, from 1991 to 2011.

Table 1: Regional distribution of the number of road traffic fatalities, registered vehicles and eestimated 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 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*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
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 first variance term, τ_0, that distinguishes a MRC model from a regression model is a term that reflects the degree to which regions differ in their intercepts. The second variance term, τ_1, that distinguishes a MRC model from typical regression reflects the degree to which slopes between independent and dependent variables vary across regions. A third variance term is common to both MRC and regression models. This variance term, σ^2, reflects the degree to the actual value of *y* differs from its predicted value within a specific region.

Unconditional means model M_0

In this section, we examine if there will be significant intercept variation (τ_0). In this case, the general assumption is that, there is significant variation in σ^2 (Bryk & Raudenbush, 1992). If τ_0 does not differ by more than chance levels, there may be little reason to use random coefficient modeling since simpler Ordinary Least Squares (OLS) modeling will suffice. Note that if slopes randomly vary even if intercepts do not, there may still be reason to estimate random coefficient models (Snijders & Bosker, 1999).

First of all, we estimate an unconditional means model. An unconditional means model does not contain any predictors, but includes a random intercept variance term for groups. This model essentially estimates how much variability there is in mean *Y* values (i.e., how much variability there is in the intercept) relative to the total variability. The model is:

$$\left.\begin{aligned} Y_{ij} &= \alpha_j + \varepsilon_{ij}, \\ \alpha_j &= \gamma_0 + e_{\alpha j} \end{aligned}\right\} \quad \text{..(3)}$$

In combined form, the model is:

$$Y_{ij} = \gamma_0 + e_{\alpha j} + \varepsilon_{ij}. \quad \text{..(4)}$$

The dependent variable, Y_{ij}, has been expressed in terms of a common intercept γ_0, and two error terms: the between-group error term, $e_{\alpha j}$, and the within-group error term, ε_{ij}. The model essentially states that any *Y* value can be described in terms of an overall mean plus some error associated with group membership and some individual error. We wish to determine two estimates of variance;

1. τ_0 associated with $e_{\alpha j}$ reflecting the variance in how much each groups' intercept varies from the overall intercept (γ_0),
2. σ^2 associated with ε_{ij} reflecting how much each individuals' score differs from the group mean.

The unconditional means model and all other random coefficient models that we will consider are estimated using the lme (linear mixed effects) function in the nlme package of *R* (Pinheiro & Bates, 2000). In the unconditional means model, the fixed portion of the model is γ_0 (an intercept term) and the random component is $e_{\alpha j} + \varepsilon_{ij}$. The observed variance within region *j* is given by $s_j^2 = \frac{1}{18}\sum_{i=1}^{19}\left(y_{ij} - \overline{y}_{.j}\right)^2$, where $\overline{y}_{.j}$ is the mean of the j^{th} region. The observed within-region variance, or pooled within-region variance is

$$MSW = s_{\text{within}}^2 = \frac{1}{180}\sum_{j=1}^{10}\sum_{i=1}^{19}\left(y_{ij} - \bar{y}_{.j}\right)^2 = \frac{1}{180}\sum_{j=1}^{10} 18 s_j^2 = \frac{1}{10}\sum_{j=1}^{10} s_j^2 \quad \text{.......(5)}$$

If the model in Equation (4) holds, then the expectation of S_{within}^2 is equal to σ^2. That is $E\left(S_{\text{within}}^2\right) = \sigma^2$. Thus,

$$\hat{\sigma}^2 = s_{\text{within}}^2 . \quad \text{..........(6)}$$

The observed between-region variance (variance of the group means) is given by

$$s_{\text{between}}^2 = \frac{1}{9}\sum_{i=1}^{19}\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2 , \quad \text{..........(7)}$$

where $\bar{y}_{..}$ is the overall mean. The total observed variance is

$$MST = s_{\text{total}}^2 = \frac{1}{189}\sum_{j=1}^{10}\sum_{i=1}^{19}\left(y_{ij} - \bar{y}_{..}\right)^2 . \quad \text{..........(8)}$$

It can be shown that $MST = MSW + MSA,$ where $MSA = 19 S_{\text{between}}^2$. The expectation of the between-region variance is given by

$$E\left(S_{\text{between}}^2\right) = \tau_0 + \frac{\sigma^2}{19} . \quad \text{..........(9)}$$

Thus, $$\hat{\tau}_0 = s_{\text{between}}^2 - \frac{\hat{\sigma}^2}{19} . \quad \text{..........(10)}$$
$$= \frac{MSA - MSW}{19} .$$

Intraclass Correlation Coefficient (ICC)

As with the completely randomized single-factor experiments, it is useful to determine how much of the total variance is between-groups. This can be accomplished by calculating the Intraclass Correlation Coefficient (ICC). Using this model we can estimate the ICC value ρ by the equation (Hox, 2010 and Snijders & Bosker, 1999)

$$\hat{\rho} = \frac{\hat{\tau}_0}{\hat{\tau}_0 + \hat{\sigma}^2} . \quad \text{..........(11)}$$

where $\hat{\tau}$ and $\hat{\sigma}^2$ are point estimates of τ and σ^2 respectively. The standard error of this estimator, where n = 19 and a = 10, is given by

$$S.E.\left(\hat{\rho}\right) = \left(1-\rho\right)\left(1+(n-1)\rho\right)\sqrt{\frac{2}{n(n-1)(a-1)}} \quad \text{..........(12)}$$

We now begin the analysis using nlme package in *R*. First the data set, i.e. the regional distribution of road traffic fatalities in Table 2, is copied on the clipboard and loaded for analysis as shown in Listing (1) (Bliese, 2013).

```
> fatalities<-read.table(file="clipboard",sep="\t",header=T)
```
....Listing (1)

In the model, the fixed formula is *y*~1 as applied in Listing (2). The random formula is random=~1|GRP (Bartko, 1976 and Bliese, 2000). This specifies that the intercept can vary as a function of group membership

```
> Null.Model<-lme(y~1,random=~1|Regions,data=fatalities,
  +control=list(opt="optim"))
```
……….............................Listing (2)

The purpose of the unconditional means model is to estimate the between-group and within-group variance in the form of τ_0 and σ^2, respectively. The option control=list(opt="optim") in the call to lme instructs the program to use R's general purpose optimization routine (Shrout & Fleiss, 1979). The VarCorr function provides estimates of variance for an lme object (Bliese, 2000).

```
>       VarCorr(Null.Model)
        Regions = pdLogChol(1)
                      Variance        StdDev
  (Intercept)    0.1891104      0.4348683..
  Residual       0.1389485      0.3727579
```
...............Listing (3)

Thus, from Listing 3, the point estimates of τ_0 and σ^2 are $\hat{\tau}_0 = 0.1891104$ and $\hat{\sigma}^2 = 0.1389485$. Thus,

$$\hat{\rho} = \frac{0.1891104}{0.1891104 + 0.1389485} = 0.5764526.$$

The ICC has values that lie in the range [0, 1]. It describes how strongly observations between regions resemble each other. If there is full agreement in every region, then $\sigma^2 = 0$ and the ICC = 1. If there is no agreement, then $\sigma^2 = 1$ and the ICC = 0. The closer the ICC value to 1, the stronger the resemblance of observations between regions.

Estimating group-mean reliability

The reliability of group means often affects one's ability to detect emergent phenomena. In other words, a prerequisite for detecting emergent relationships at the aggregate level is to have reliable group means (Bliese, 1998). By convention, estimates around 0.70 are considered reliable. Group mean reliability estimates are a function of the ICC and group size (Bliese, 2000). ICC(2) is among regions variance (*MSA*) minus within regions variance (*MSW*) over among regions variance (*MSA*).

$$ICC(2) = \frac{MSA - MSW}{MSA} \quad \text{...(13)}$$

The GmeanRel function from the multilevel package in *R* calculates the ICC, the group size, and the group mean reliability for each group (Bryk & Raudenbush, 1992). When we apply the GmeanRel function to our Null.Model based on the 10 regions in the fatalities data set, we are interested in two things. First, we are interested in the average reliability of the 10 regions. Second, we are interested in determining whether or not there are specific regions that have particularly low reliability. The result of the function GmeanRel(Null.Model) shows that the reliability of all the 10 regions of Ghana are greater than 0.70, where the overall group-mean reliability is acceptable at 0.9627688.

Determining whether τ_0 is significant.

If it is assumed that the within-region deviations ε_{ij} are normally distributed, then we can test the hypothesis that ICC is 0, which is the same as the null hypothesis that there are no regional differences, or the true between-region variance is 0. The test statistic is $F = \frac{MSA}{MSW}$, which has the *F*-distribution with 9 and 180 degrees of freedom. The estimate of *MSA*, *MSW* and the ICC value can also be computed from an ANOVA model, given in Listing (4) (Bliese, 2013).

\> tmod<-aov(y~as.factor(Regions),data=fatalities............Listing (4)

The results of Listing (4) can be summarized in Table 3. We reject the null hypothesis at 0.05 level of significance if the observed *F* value is greater than $F_{0.05,9,180} = 1.9322$. From Table 5.14, since the observed value 26.8592 is greater than 1.9322, we reject the null hypothesis and hence the ICC is significantly different from 0. Thus, intercept variance (τ_0) estimate of 0.1891104 is significantly different from zero.

Table 3: Analysis of variance table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio	F-crit
Among regions	33.5884	9	3.73205	26.8592	1.9322
Within regions	25.0107	180	0.13895		
Total	58.5991	189			

Random intercept model: M_1

At this point of the analysis, there are two sources of variation that we can attempt to explain in subsequent modeling – within-region variation (σ^2) and between-region intercept variation (τ_0). In this section, we begin to build a model that predicts these two sources of variation. The first step towards modeling between-group variability is to let the intercept vary between regions. This reflects that some groups tend to have, on average, higher responses *Y* and others tend to have lower responses. The form of the model is:

$$\left.\begin{aligned} y_{ij} &= \alpha_j + \beta_j x_{ij} + \varepsilon_{ij}, \\ \alpha_j &= \gamma_0 + \gamma_1 \bar{x}_j + e_{\alpha j} \\ \beta_j &= \delta_0, \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(14)$$

where γ_0, γ_1 and δ_0 are constants to be estimated and $\bar{x}_j$ is the mean of the observations, x_{ij}, in the j^{th} region of Ghana When we combine the three rows into a single equation, we get an equation that looks like a common regression equation with an extra error term $(e_{\alpha j})$. This error term indicates that *y* intercepts (i.e., means) can randomly differ across groups. The combined model is:

$$Y_{ij} = \gamma_0 + \delta_0 x_{ij} + \gamma_1 \bar{x}_j + e_{\alpha j} + \varepsilon_{ij}, \quad j = 1,\ 2,\ \ldots,\ 10. \ \ldots\ldots\ldots(15)$$

Essential assumptions are that all residuals, $e_{\alpha j}$ and ε_{ij}, are mutually independent and have zero means given the values x_{ij} of the explanatory variable (Hox, 2010). For the $e_{\alpha j}$, just as for the ε_{ij}, it is assumed that they are drawn from normally distributed populations. The population variance of the fatality-level residuals, ε_{ij}, is assumed to be constant across the regions, and is denoted by σ^2; the population variance of the regional-level residuals $e_{\alpha j}$ is denoted by τ_0. Thus, model M_1 has four parameters: the regression coefficients γ_0 and γ_1, and the variance components σ^2 and τ_0. The residual variance, i.e., the variance conditional on the value of *X*, is

$$\text{Var}\left(Y_{ij} \middle| x_{ij}\right) = \text{var}\left(e_{\alpha j}\right) + \text{var}\left(\varepsilon_{ij}\right) = \tau_0 + \sigma^2, \ \ldots\ldots\ldots\ldots\ldots(16)$$

while the covariance between two different units (i and i', with $i \neq i'$) in the same region is

$$\text{Cov}\left(Y_{ij}, Y_{i'j} \middle| x_{ij}, x_{i'j}\right) = \text{var}\left(e_{\alpha j}\right) = \tau_0. \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots(17)$$

The fraction of residual variability that can be ascribed to fatality-level is given by $\sigma^2 / \left(\sigma^2 + \tau_0\right)$, and for regional-level this fraction is $\tau_0 / \left(\sigma^2 + \tau_0\right)$.

Of the covariance or correlation between *Y*-values of two units in the same region, a part may be explained by their *X*-values, and another part is unexplained. This unexplained, or residual, correlation between them is the residual intraclass correlation coefficient,

$$\rho_1\left(Y \middle| X\right) = \frac{\tau_0}{\sigma^2 + \tau_0}. \quad \text{...(18)}$$

This parameter is the correlation between the Y-values of two randomly drawn units in one randomly drawn region, controlling for variable X. If model M_1 is valid while the intraclass correlation coefficient is 0, i.e. $e_{\alpha j} = 0$ for all regions j, then the grouping is irrelevant for Y-variable conditional on X, and one could have used ordinary linear regression. If the residual intraclass correlation coefficient, or equivalently, τ_0, is positive, then the hierarchical linear model is a better analysis method than ordinary least squares regression analysis. Using the data in Table 2, model M_1 is specified in the R package lme as shown in Listing (5).

```
Model.1<-lme(y~x+G.x,random=~1|Regions,data=fatalities,
control=list(opt="optim"))
```
……..Listing (5)

The result of the application of the R functions summary(Null.Model) and summary(Model.1), which presents the parameter estimate and standard errors for both models (M_0 and M_1), are simplified in Table 4.

Table 4: Intercept-only model and model with explanatory variables

Model	M_0: intercept only		M_1: with predictor	
Fixed effect	Coefficient	Standard Error	Coefficient	Standard Error
γ_0 = Intercept	-9.669	0.140	-10.076	0.743
δ_0 = coffiecient of x_{ij}			0.459	0.037
γ_1 = coefficient of $\overline{x}_j$			-0.545	0.166
Random part	Parameters	Standard Error	Parameter	Standard Error
τ_0 = var($e_{\alpha j}$)	0.189	0.209	0.209	0.145
σ^2 = var(ε_{ij})	0.139	0.086	0.076	0.063
Deviance	198.201		94.554	

In this table, the intercept-only model estimated the intercept as -9.688842, which is simply the average y values of all regions and fatalities. The variance of the fatality-level residual error, symbolized by σ^2, is estimated as 0.1389485. The variance of the regional-level residual errors, symbolized by τ_0 is estimated as 0.1891104. The deviance reported in Table 4 is a measure of model misfit; when we add explanatory variable to the model, the deviance is expected to go down (Hox, 2010).

In the second model, where the explanatory variable was included, the regression coefficients for all three variables are significant. Notice that the x-scores are significantly positively related to y-scores. Furthermore after controlling the fatality-level relationship, average x-scores are negatively related to the average y-score in a region. The interpretation of this model indicates that the slope at the regional-level significantly differs from the slope at the fatality-level. A unit increase at the regional-level is associated with a –0.085 (–0.545 + 0.460) decrease in average *y*-score. The coefficient of –0.545 reflects the degree of difference between the two slopes.

The within-region and between-region regression coefficients would be equal if, in Equation (15), the coefficient of $\overline{x}$ would be 0, i.e. $\gamma_1 = 0$. This null hypothesis can be tested using the test statistics

$$T = \frac{\text{estimate}}{\text{standard error}},$$

which has the *t*-distribution with 9 degrees of freedom. The value of *T* based on the given data is $t = -0.544840/0.1658 = -3.286$, which is significant at 0.05 level.

The within-region deviation about this regression equation, ε_{ij}, have a variance of 0.0759 (standard deviation 0.2755). Within each region, the effect (regression coefficient) of x_{ij} is equal to 0.459, so the regression lines are parallel. Regions differ in two ways; they may have different mean x-values, which affects the expected results y_{ij} through the term $0.545\overline{x}_j$; this is an explained difference between the regions; and they have randomly differing values for e_{0j}, which is an unexplained difference. These two ingredients contribute to the region-dependent intercept, given by $-10.076 + e_{\alpha j} - 0.545\overline{x}_j$.

The application of the function coef(Model.1), in R, gives the estimate of the regional-level residual $\hat{e}_{\alpha j}$ and the corresponding values of α_j and β_j for each region, which are summarized in Table 5. The values of $\overline{x}_j$, computed from Table 2, and the corresponding values of $\hat{\nu}_j = e^{\hat{\alpha}_j}$ are also given in Table 5.

Table 5: Estimate of the values of $e_{\alpha j}$, α_j, β_j and $\hat{\nu}_j$ for each region

Regions	Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong Ahafo
$\hat{e}_{\alpha j}$	0.43873	0.24502	0.00278	0.58195	0.36494	-0.14103	-0.59011	-0.42436	-0.59757	0.11865
$\overline{x}_j$	-2.35474	-3.92579	-4.85053	-5.24053	-5.41474	-5.53579	-4.59368	-4.24947	-3.91211	-4.99790
$\hat{\alpha}_j$	-8.35385	-7.69156	-7.42995	-6.63827	-6.76036	-7.20038	-8.16278	-8.18457	-8.54161	-7.23378
$\hat{\beta}_j$	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906	0.45906
$\hat{\nu}_j$	0.0002355	0.0004567	0.0005932	0.0013093	0.0011588	0.0007463	0.0002851	0.0002789	0.0001952	0.0007218

The estimated values of α and β can be used to estimate the number of road traffic fatalities in each region. For instance, in Greater Accra region, where $\overline{x} = -2.35474$, the estimated values for α and β are $\alpha_1 = -10.076 +$

$0.43873 - 0.545\bar{x} = -8.35385$ and $\beta_1 = 0.45906$, respectively. Therefore, the estimate for ν_1 is

$$\hat{\nu}_1 = e^{-8.35385} = 0.0002355. \quad \text{..(19)}$$

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

$$D_{i1}/P_{i1} = 0.0002355\left(N_{i1}/P_{i1}\right)^{0.45906}, \quad \text{…………………………......(20)}$$

where D_{i1} is the number of road traffic fatalities in the i^{th} year, N_{i1} number of registered vehicles in the i^{th} year and P_{i1} is the estimated population size in the i^{th} year, for Greater Accra region.

Random slope model *M*$_2$

In the random intercept model of M_1, the group differ with respect to the average value of the dependent variable: the only random group is the random intercept. But the relation between explanatory and dependent variables can differ between regions in more ways. We therefore continue our analysis by trying to explain the third source of variation, namely, variation in the slope, τ_1. The model that we test is:

$$\left.\begin{aligned} y_{ij} &= \alpha_j + \beta_j x_{ij} + \varepsilon_{ij}, \\ \alpha_j &= \gamma_0 + \gamma_1 \bar{x}_j + e_{\alpha j} \\ \beta_j &= \delta_0 + e_{\beta j} \end{aligned}\right\} \quad \text{……………………………..…….......(21)}$$

The intercepts α_j as well as the regression coefficients, or slopes, β_j are region-dependent. When we combine the three rows into a single equation in the form

$$y_{ij} = \gamma_0 + \delta_0 x_{ij} + \gamma_1 \bar{x}_j + e_{\beta j} x_{ij} + e_{\alpha j} + \varepsilon_{ij}, \quad j = 1,\ 2,\ ...,\ 10. \text{ ...(22)}$$

It is assumed that the regional-level residuals $e_{\alpha j}$ and $e_{\beta j}$ as well as the fatality-level residuals ε_{ij} have mean 0, given the value of the explanatory variable *X*. Thus, δ_0 is the average regression coefficient just like γ_0 is the average intercept. The first part of Equation (21), $\gamma_0 + \delta_0 x_{ij} + \gamma_1 \bar{x}_j$, is called the fixed part of the model. The second part $e_{\beta j} x_{ij} + e_{\alpha j} + \varepsilon_{ij}$, is called the random part (Hox, 2010). The term $e_{\beta j} x_{ij}$ can be regarded as random interaction between group (region) and *x*. This model implied that the regions are characterized by two random effects: their intercept and their slope. Thus, *x* has a random coefficient. These two regional effects are usually correlated. The assumption is that, for different regions, the pairs of random effect $\left(e_{\alpha j},\ e_{\beta j}\right)$ are independent and identically distributed, that they are independent of the fatality-level residuals ε_{ij}, and that all ε_{ij} are independent and identically distributed. The variances of the fatality-level residuals ε_{ij}, is again denoted σ^2; the variances covariance of the regional-level residuals $\left(e_{\alpha j},\ e_{\beta j}\right)$ is denoted as follows (Snijders & Bosker, 1999):

$$\left.\begin{aligned} &\mathrm{var}(e_{\alpha j}) = \tau_0, \\ &\mathrm{var}(e_{\beta j}) = \tau_1, \\ &\mathrm{cov}(e_{\alpha j},\ e_{\beta j}) = \tau_{01}. \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(23)$$

Thus, from Equations (21) and (22),

$$\mathrm{var}\left(Y_{ij}\middle|x_{ij}\right) = \tau_0 + 2\tau_{01}x_{ij} + \tau_1 x_{ij}^2 + \sigma^2, \ \ldots\ldots\ldots\ldots(24)$$

and, for two different years i and i' $(i \neq i')$,

$$\mathrm{cov}\left(Y_{ij},\ Y_{i'j}\middle|x_{ij}, x_{i'j}\right) = \tau_0 + \tau_{01}(x_{ij} + x_{i'j}) + \tau_1 x_{ij} x_{i'j}. \ \ldots\ldots\ldots(25)$$

The slope β_j is normally distributed random variable with mean δ_0 and variance τ_1. The variance term associated with $e_{\beta j}$ is τ_1. Since 95% of the probability of a normal distribution is within two standard deviations from the mean, it follows that approximately 95% of the regions have slopes between $\delta_0 - 2\sqrt{\tau_1}$ and $\delta_0 + 2\sqrt{\tau_1}$.

Fig. 1 presents 10 regression lines for the 10 regions of Ghana using the data in Table 2. The figure demonstrates regression lines that characterize, according to this model, the population of geographical regions in Ghana. In *R* this model is designated as shown in Listing (6).

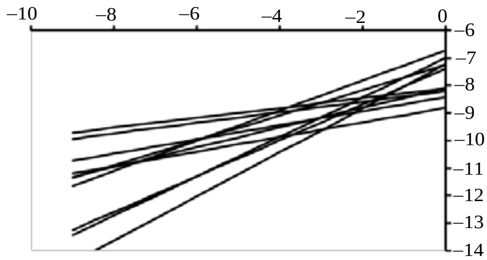

Fig. 1: Ten random regression lines from Table 2

```
> Model.2<-lme(y~x+G.x,random=~x|Regions,
    data=fatalities,control=list(opt="optim")) ……………......Listing (6)
```

The summary of the results of Listing (6) is obtain by the *R* function summary(Model.2). The *R* function VarCorr(Model.2) provides estimates of variances (Bliese, 2013).

Table 6 presents the parameter estimate and standard errors for the models M_0, M_1 and M_2. The within-region regression in model M_2 is 0.4459 and between-region regression coefficient is -0.3384 + 0.4459 = 0.1075.

All the standard errors of the estimated parameters in model M_2 are smaller than the corresponding values of model M_1. Moreover, the deviance, which measures the model misfit, is much lower in M_2 as compare to that of M_1. Thus, estimate parameters based on model M_2 is preferred.

In the null model M_0, the variance estimate from the within-region residual, σ^2, was 0.1389. and the variance estimate for the intercept, τ_0, 0.1891. The variance estimates from the model M_2, with one predictors, are $\hat{\sigma}^2 = 0.0630$ and $\hat{\tau}_0 = 0.1545$. That is, the variance of the within-region residuals decreased from 0.1389 to 0.0630 and the variance of the between-region intercepts decreased from 0.1891 to 0.1545.

Table 6: Comparison of models M_0, M_1 and M_2

Model	M_0: intercept only		M_1: with predictor		M_2: with predictor	
Fixed effect	Coeff.	Standard Error	Coeff.	Standard Error	Coeff.	Standard Error
γ_0 = Intercept	-9.687	0.1401	-10.0756	0.7426	-9.2341	0.2065
δ_0 = coffiecient of x_{ij}			0.4591	0.0374	0.4459	0.0707
γ_1 = coefficient of $\overline{x}_j$			-0.5448	0.1658	-0.3384	0.0516
Random part	Parameter	Standard Error	Parameter	Standard Error	Parameter	Standard Error
$\tau_0 = \text{var}(e_{\alpha j})$	0.189	0.2085	0.209	0.146	0.1545	0.1243
$\tau_1 = \text{var}(e_{\beta j})$					0.0382	0.0618
$\tau_{01} = \text{cov}(e_{\alpha j}, e_{\beta j})$					0.0766	
$\sigma^2 = \text{var}(\varepsilon_{ij})$	0.139	0.086	0.076	0.063	0.0630	0.0576
Deviance	198.201		94.554		64.749	

$$\text{Variance explained} = 1 - \frac{\text{varince with predictor}}{\text{variace without predictor}} \quad ....(26)$$

The y-values explained $1-(0.0630/0.1389)$ or 0.55 (55%) of the within-region variance in σ^2, and regional-mean values $\overline{x}$ explained $1-(0.1545/0.1891)$ or 18% of the between-region intercept variance τ_0. Should the value of 0.0382 for the random slope variance be considered to be high? The slope standard deviation is $\sqrt{0.0382} = 0.195$, and the average slope is $\delta_0 = 0.4459$. The values of 'average slope $\pm$ two standard deviations' range from 0.0559 to 0.8359. This implies that the effect of x is clearly positive in all regions. Table 7 gives the slope of the least square regression line for each of the 10 regions of Ghana based on the data in Table 2.

Table 7: Slope of the least square regression line for each region in Ghana

Greater Accra	Ashanti	Western	Eastern	Central	Volta	Northern	Upper East	Upper West	Brong-Ahafo
0.267	0.360	0.258	0.179	0.193	0.549	0.717	0.654	0.804	0.458

It can be seen from Table 7 that all the 10 regions have slopes between 0.0559 and 0.8359. Thus, the normality assumption of the slope is validated. The correlation between random slope and random intercept is

$$\rho_{\alpha\beta} = \frac{0.0766}{\sqrt{0.1545 \times 0.0382}} = 0.997.$$

The standard deviation of the *x*-values is about 1.05, and the mean is –4.5. Hence fatalities with *x* values among the bottom fewer percent or the top few percent have *x* values of about – 6.6 and –2.4, respectively. Substituting these values in the contribution of the random effect gives $e_{\alpha j} - 6.6 e_{\beta j}$ and $e_{\alpha j} - 2.4 e_{\beta j}$. It follows from Equations (5.42) and (5.43) that when $x = -6.6, \ -2.4$,

$$\mathrm{Var}\left(Y_{ij} \middle| x_{ij} = -6.6\right) = 0.1545 + 2 \times 0.0766 \times (-6.6) + 0.0382 \times (-6.6)^2 + 0.0630 = 0.8704,$$

$$\mathrm{Cov}\left(Y_{ij},\ Y_{i'j} \middle| x_{ij} = -6.6, x_{i'j} = -2.4\right) = 0.1545 + 0.0766(-6.6 - 2.4) + 0.0382 \times 6.6 \times 2.4 = 0.0702$$

$$\mathrm{Var}\left(Y_{ij} \middle| x_{i''j} = -2.4\right) = 0.1545 + 2 \times 0.0766 \times (-2.4)^2 + 0.0382 \times (-2.4) + 0.0630 = 0.0699.$$

and therefore

$$\rho\left(Y_{ij},\ Y_{i'j} \middle| x_{ij} = -6.6, x_{i'j} = -2.4\right) = \frac{0.0702}{\sqrt{0.8704 \times 0.0699}} = 0.2846.$$

Thus, the highest value of *x* and the least value of *x* in the same region are positively correlated over the population of regions. The positive correlation corresponds to the result that the value of *x* for which the variance given by (5.42) is minimal, is outside the range from -6.6 to -2.4. For the estimates in Table 5.17, this variance is

$$\mathrm{Var}\left(Y_{ij} \middle| x_{ij} = x\right) = 0.1545 + 0.1532x + 0.0382x^2 + \sigma^2.$$

Equating the derivative with respect to *x* to 0, shows that the variance is minimal when $x = -0.1532/0.0382 = -4.01$, which is within the range –6.6 to –2.4. In Table 8, the model M_2 represents within each region, denoted *j*, a linear regression equation

$$Y_{ij} = -9.2341 + 0.4459 x_{ij} - 0.3384 \overline{x}_j + e_{\beta j} x_{ij} + e_{\alpha j} + \varepsilon_{ij}, \quad \ldots\ldots\ldots\ldots\ldots(27)$$

where $e_{\alpha j}$ and $e_{\beta j}$ are region-dependent deviations each with mean 0 and variances 0.1545 and 0.0630, respectively. The application of the R code coef(Model.2) gives the intercept and the coefficients of *x* and $\overline{x}$ as shown in Table 8.

Table 8: Intercept and coefficients of x and $\overline{x}$

No.	Regions	Intercept	x	$\overline{x}$
1	Greater Accra	-9.506844	0.3083572	-0.3384525
2	Ashanti	-9.402255	0.3614688	-0.3384525
3	Western	-9.319224	0.4053849	-0.3384525
4	Eastern	-9.704008	0.2109577	-0.3384525
5	Central	-9.575697	0.2758323	-0.3384525
6	Volta	-9.270846	0.4259363	-0.3384525
7	Northern	-8.806641	0.6594775	-0.3384525
8	Upper East	-8.839118	0.6439825	-0.3384525
9	Upper West	-8.530726	0.7993004	-0.3384525
10	Brong Ahafo	-9.385768	0.3686119	-0.3384525

The estimate of regional-level residuals $\hat{e}_{\alpha j}$ and $\hat{e}_{\beta j}$ and the corresponding values of α and β for each region are given in Table 9. Based on Table 9, the estimate of the number of road traffic fatalities, $\hat{D}_{ij}$, of the j^{th} geographical region of Ghana in the i^{th} year, can be obtained from the formula

$$\hat{D}_{ij} / P_{ij} = \hat{\nu}_j \left(N_{ij} / P_{ij} \right)^{\hat{\beta}_j}, \quad j = 1, 2, ..., 10 \quad \text{.........................(28)}$$

N_{ij} is the number of registered vehicles in the i^{th} year recorded in the j^{th} region while P_{ij} represents the population size in the i^{th} year recorded in the j^{th} region.

Table 9: Estimate of regional-level residuals and the values of α and β

Regions	$\hat{e}_{\alpha j}$	$\hat{e}_{\beta j}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}_j = e^{\hat{\alpha}}$
Greater Accra	-0.273	-0.138	-8.709877	0.3083572	0.0001649
Ashanti	-0.168	-0.084	-8.073562	0.3614688	0.0003117
Western	-0.085	-0.041	-7.677551	0.4053849	0.0004631
Eastern	-0.470	-0.235	-7.930339	0.2109577	0.0003597
Central	-0.342	-0.170	-7.743066	0.2758323	0.0004337
Volta	-0.037	-0.020	-7.397244	0.4259363	0.0006129
Northern	0.427	0.214	-7.251897	0.6594775	0.0007088
Upper East	0.395	0.198	-7.400873	0.6439825	0.0006107
Upper West	0.703	0.353	-7.206664	0.7993004	0.0007416
Brong Ahafo	-0.152	-0.077	-7.694218	0.3686119	0.0004555

For instance, in Greater Accra region, where $\overline{x} = -2.35474$, the estimated values for α and β are

$$\hat{\alpha} = -9.2341 + 0.33845 \times (2.35474) - 0.273 = -8.710,$$

$\hat{\beta} = 0.4459 - 0.138 = 0.308.$

Therefore, the estimate for ν_j is

$$\hat{\nu}_j = e^{-8.710} = 0.0001649. \quad \ldots\ldots(29)$$

Equation (5.46), for Greater Accra region, therefore becomes

$$D_{i1}/P_{i1} = 0.000164948\left(N_{i1}/P_{i1}\right)^{0.3083572}. \quad \ldots\ldots(30)$$

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 (30), are given in Table 10. The percentage differences between the calculated and actual values are also given. It can be seen that, from 2001 to 2012 in Greater Accra region, all the 12 calculated figures are within 10% of the actual figure. Out of the 22 calculated figures, from 1991 to 2012, 15 are within 10% of the actual figure and 19 are within 20% of the actual value.

Table 10: Comparison of actual fatalities and fatalities estimated from Equation (28) for Greater Accra region

i	Year	D_{i1}	$\hat{D}_{i1}$	Error	Error %	i	Year	D_{i1}	$\hat{D}_{i1}$	Error	Error %
1	1991	126	120.1	5.9	4.7	**12**	2002	239	262.5	-23.5	9.8
2	1992	164	125.4	38.6	23.5	**13**	2003	240	262.6	-22.6	9.4
3	1993	115	134.7	-19.7	17.1	**14**	2004	299	290.1	8.9	3.0
4	1994	155	147.7	7.3	4.7	**15**	2005	306	304.3	1.7	0.6
5	1995	190	161.6	28.4	14.9	**16**	2006	325	319.8	5.2	1.6
6	1996	191	179.1	11.9	6.2	**17**	2007	370	336.0	34.0	9.2
7	1997	174	192.4	-18.4	10.6	**18**	2008	385	350.6	34.4	8.9
8	1998	258	207.1	50.9	19.7	**19**	2009	420	378.5	41.5	9.9
9	1999	172	223.7	-51.7	30.1	**20**	2010	424	384.8	39.2	9.3
10	2000	196	241.6	-45.6	23.2	**21**	2011	425	403.8	21.2	5.0
11	2001	239	249.1	-10.1	4.2	**22**	2012	435	442.4	-7.4	1.7

Conclusion

The method of least squares was used by Hesse et al. (2014) to derive a modified form of Smeed's regression formula for estimating road traffic fatalities in Ghana, where the regression coefficients, α and β, were fixed unknown parameters.

In this paper, we considered a similar study with data from the 10 geographical/ administrative regions of Ghana. The difference with the modified Smeed's regression formula is that we assume that each region has a different intercept coefficient α_j, and a different slope coefficient β_j. Since the parameters are assumed to vary across the various regions, they are considered to be random variables, which are given as a probability model.

The number of road traffic fatalities and the regional groups are conceptualized as a hierarchical system of road traffic fatalities and regions, with fatalities and regions defined at separate levels of this hierarchical system. Instead of estimating a separate regression equations for each of the 10 regions in Ghana, a multilevel regression analysis was applied to estimate the values of α and β for region based on data given.

Prior to this analysis, the Intra-region Correlation Coefficient (ICC) was determined to be significantly different from zero (0), and thus the intercept variance (τ_0) estimate of 0.1891104 is significantly different from zero. These results show that there is significant intercept variation in terms of the dependent variable *y* across the 10 regions. It was estimated that about 58% of the variation in *y* is a function of the region to which it is observed. Thus, a multilevel model, which allows for random variation in *y* among regions, is better than a model that does not allow for this random variation.

To determine the values of the regression coefficients, firstly the random intercept model M_1, which allows variability of the regression intercept between regions, with fixed slope, was developed. It was found that, the application of this model leads to serious under-estimation of the number of road fatalities in Greater Accra region. The analysis was therefore continued by trying to explain another source of variation due to regional distribution of regression slope, using the random slope model, M_2.

The within-region regression in model M_2 is 0.4459 and between-region regression coefficient is 0.1075. From Table 6, it can be seen that, the standard errors of the estimated parameters in model M_2 are smaller than the corresponding values of model M_1. Moreover, the deviance, which measures the model misfit, is much lower in M_2 as compare to that of M_1. Thus, estimate parameters based on model M_2 is preferred.

Using model M_2, from 2001 to 2012 in Greater Accra region, all the 12 calculated figures are within 10% of the actual figure. Out of the 22 calculated figures, from 1991 to 2012, 15 are within 10% of the actual figure and 19 are within 20% of the actual value.

References:

Bartko, J. J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin, 83*, 762-765.

Bliese, P. (2013). Multilevel modeling in R (2.5): A brief introduction to R, the multilevel package and the nlme package. University of North Texas, cran.r-project.org/doc/contrib./Bliese_Multilevel.pdf.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and Analysis. In K. J. Klein & S. W. Kozlowski (Eds.), *Multilevel Theory, Research, and Methods in Organizations* (pp. 349-381). San Francisco, CA: Jossey-Bass, Inc.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

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.

Hox, J. J. (1998). Multilevel modeling. When and why. In I. Balderjahn, R. Mathar and M Schader (Eds.), (classification, data analysis, and data highways), 147 – 154. New York: Springer Verlag.

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

Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

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

Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.