



# International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531  
(Print & Online)

<http://gssrr.org/index.php?journal=JournalOfBasicAndApplied>



---

## Factorial Design and Model for the Effect of National Road Safety Strategy Two (NRSS II).

Richard Ayamah<sup>a\*</sup>, K. F. Darkwah<sup>b</sup>

<sup>a</sup>*Mathematics/ICT Department, St. Joseph's College of Education, Bechem-B/A, Ghana.*

<sup>b</sup>*Mathematics Department, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.*

<sup>a</sup>*r.ayamah@yahoo.com*

<sup>b</sup>*fordarkk@gmail.com*

### Abstract

One of the biggest challenges facing the road transportation industry in Ghana is the scale and magnitude of road traffic injuries. This led to the introduction of the second National Road Safety Strategy (NRSS II) for the period 2006-2010, after the implementation of NRSS I for the period 2001-2005, to reduce Road Traffic Accident (RTA) and its fatality rate to the barest minimum. Factorial design was used to investigate the effect of NRSS II in the reduction of RTA, and also find a model for adequate description of the various observations in the data for the number of RTA recorded by Motor Traffic and Transport Unit (M.T.T.U) of the Ghana Police Service. Two-factor factorial design with mixed factors (with region as the random factor and accident nature as the fixed factor) was used for the presentation and analysis of the data of the yearly number of RTA for the period 2006-2009 recorded by the regional branches of M.T.T.U in the Northern and Ashanti regions of Ghana. It was observed from our analysis that the campaign on the NRSS II helped in the reduction of the seriousness of RTA.

---

\* Corresponding author.

E-mail address: r.ayamah@yahoo.com.

A mixed-effect model was determined and it was found to be adequate with variability of 92.91% of the number of RTA explained by region, accident nature and the region-accident nature interaction at significant level of .05. Our research shows that socio-economic factors of a region have a great impact on the severity and frequency of RTA.

**Keywords:** National Road Safety Strategy; Road Traffic Accident; Two-factor factorial design; random factor; fixed factor; mixed factors; mixed-effect model.

## **1. Introduction**

Road transport continues to be the most prevalent mode of transportation worldwide and is undoubtedly vital for the socio-economic development in most countries. The various Governments have over the past decades continued to pursue comprehensive road infrastructure development and transport service programmes aimed at improving accessibility and mobility in their respective countries as part of poverty reduction strategy, as in reference [1].

However, road traffic accidents (RTAs) are a major public health concern, resulting in a yearly estimated 1.2 million deaths and 50 million injuries worldwide. Africa has the highest average rising trends in road traffic fatalities at a rate of 67 per 10,000 registered vehicles and population risk of 28 per 100,000 populations as compared to other continents, as in reference [2]. Also, in the developing world, RTAs are among the leading cause of death and injury. In Ghana for instance, more than 1,600 deaths and over 10,000 injuries are recorded yearly as a result of road traffic accidents, leading to the Government spending about GH¢ 221,000,000, which is about 1.6 percent of the Nation's Gross National Product, as in reference [3]. This justifies the need for a comprehensive and sustainable investment in road safety management.

Various studies have addressed the different aspects of RTA's; with most of the literature focusing on predicting or establishing the critical factors influencing injury severity, as in reference [4]. Numerous data mining-related studies have been undertaken to analyze RTA data locally and globally, with results frequently varying depending on the socio-economic conditions and infrastructure of a given location, as in reference [2]. In [5], Pendharkar and Ossenbruggen for instance used a logistic regression model to identify the prediction factors of crashes and crash-related injuries, and their study showed that village sites were less hazardous than residential or shopping sites. In [6], Getnet also investigated the potential application of data mining tools (using the J48 algorithm and PART algorithm) techniques to develop models supporting the identification and prediction of major driver and vehicle risk factors that cause RTAs. Chang and Chen too conducted data mining research focusing on building tree-based models to analyze freeway accident frequency, as in reference [7]. In [8], Sohn and Hyungwon conducted a research using three data mining techniques (neural network, logistic regression, and decision tree) to select a set of influential factors and to construct classification models for RTA severity in Korea.

In their study [9], Akgurgor and Yildiz used factorial design analysis method to investigate the sensitivity of the accident prediction model and evaluated that average daily traffic (ADT), lane width ( $W$ ), width of paved

shoulder (P), median (H) and their interactions have significant effects on number of accident, as in reference [9]. Also, Chhotu and Chandra used fractional factorial design for split-plot (SP) survey data collection and analysis in order to investigate how road users value safety and how much they are willing to pay for a better safety, as reference [10]. In this research, two-factor factorial design was employed to investigate how effective NRSS II was in the reduction of RTA based on the data for the period 2006-2009 and extending to 2010 to inform policy strategy for the next face for the implementation of NRSS III which was to take off in 2011.

## 2. Materials and Methods

“In many experiments where two or more factors are being investigated, neither factor is considered extraneous; each is of major concern to the experimenter. When this occurs; the experiment is called a factorial experiment to emphasize the fact that interest is centered on the effect of two or more factors on a measured response” [11].

In two-factor factorial design, only two factors, say factors A and B are involved. In Table 1 for instance, the first column is headed by factor A and comprises random levels counted as (1, 2, ..., a) row headings; the second column is headed by factor B and has b fixed levels counted as (1, 2, ..., b) column headings; the third column, headed by Totals ( $y_{i..}$ ,  $i = 1, 2, \dots, a$ ) shows the sum of the observations (replications) for the  $i$ th level of factor A and the last column, headed by Means ( $\bar{y}_{i..}$ ), is the average of the observations for the  $i$ th level of factor A. The last but one row, headed by Totals ( $y_{.j}$ ,  $j = 1, 2, \dots, b$ ), is the sum of the observations for the  $j$ th level of factor B and the last row, headed by Means ( $\bar{y}_{.j}$ ), is the average of the observations for the  $j$ th level of factor B.

As there are a levels of factor A and b levels of factor B, ab is the total number of cells for treatment combinations. A treatment combination (cell) is a level of factor A applied in conjunction with a level of factor B. Also, if there are n observations (replicates) in each cell, the total number of observations (replications) in the experiment is given by abn. Furthermore, the kth observation taken at the  $i$ th level of factor A and  $j$ th level of factor B is denoted by  $y_{ijk}$ , where  $i = 1, 2, \dots, a$ ;  $j = 1, 2, \dots, b$  and  $k = 1, 2, \dots, n$ . For example,  $y_{12n}$  is the  $n$ th observation taken at the first level of factor A and at the second level of factor B.

We list below, some useful symbols, some of which are used in Table 1

$$y_{ij.} = \sum_{k=1}^n y_{ijk} \qquad \bar{y}_{ij.} = \frac{y_{ij.}}{n}$$

$$y_{i..} = \sum_{j=1}^b \sum_{k=1}^n y_{ijk} \qquad \bar{y}_{i..} = \frac{y_{i..}}{bn}$$

$$y_{.j} = \sum_{i=1}^a \sum_{k=1}^n y_{ijk} \qquad \bar{y}_{.j} = \frac{y_{.j}}{an}$$

$$y_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk} \qquad \bar{y}_{...} = \frac{y_{...}}{abn}$$

**Table 1:** General data layout for two-factor factorial design.

	<i>Factor B</i>				<i>Totals</i>	<i>Means</i>
<i>Factor A</i>	<i>1</i>	<i>2</i>	...	<i>b</i>	$y_{i\cdot}$	$\bar{y}_{i\cdot}$
<i>1</i>	$y_{111}, y_{112}, \dots, y_{11n}$  $y_{11\cdot}$  $\bar{y}_{11\cdot}$	$y_{121}, y_{122}, \dots, y_{12n}$  $y_{12\cdot}$  $\bar{y}_{12\cdot}$	...	$y_{1b1}, y_{1b2}, \dots, y_{1bn}$  $y_{1b\cdot}$  $\bar{y}_{1b\cdot}$	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
<i>2</i>	$y_{211}, y_{212}, \dots, y_{21n}$  $y_{21\cdot}$  $\bar{y}_{21\cdot}$	$y_{221}, y_{222}, \dots, y_{22n}$  $y_{22\cdot}$  $\bar{y}_{22\cdot}$	...	$y_{2b1}, y_{2b2}, \dots, y_{2bn}$  $y_{2b\cdot}$  $\bar{y}_{2b\cdot}$	$y_{2\cdot}$	$\bar{y}_{2\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>a</i>	$y_{a11}, y_{a21}, \dots, y_{a1n}$  $y_{a1\cdot}$  $\bar{y}_{a1\cdot}$	$y_{a21}, y_{a22}, \dots, y_{a2n}$  $y_{a2\cdot}$  $\bar{y}_{a2\cdot}$	...	$y_{ab1}, y_{ab2}, \dots, y_{abn}$  $y_{ab\cdot}$  $\bar{y}_{ab\cdot}$	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
<i>Totals (y<sub>·j</sub>)</i>	$y_{\cdot 1}$	$y_{\cdot 2}$	...	$y_{\cdot b}$	$y_{...}$	$\bar{y}_{...}$
<i>Means (ȳ<sub>·j</sub>)</i>	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	...	$\bar{y}_{\cdot b}$		

**Assumptions:**

The observations in the (i,j)th cell constitute a random sample size *n* from a population that is assumed to be normally distributed with mean ( $\mu_{ij}$ ) and variance  $\sigma^2$ . All the sample observations in the *ab* cells are assumed to have the same variance  $\sigma^2$ . Also, it is assumed that the populations from which *n* independent identically distributed observations are taken are combinations of factors and that equal number of observations (*n*) is taken at each factor combination (cell).

**2.1 Model Formulation**

According to Milton and Arnold, model formulation entails the development of prediction equations (statistical models) by statistical or mathematical method from experimental data. In this paper we use the two-factor factorial design with mixed-effect factors, as in reference [11].

Let  $\mu$  = the overall mean effect estimated by  $\bar{y}_{...}$ ;

$\alpha_i$  = independent random treatment effect of factor **A**, (independent of  $E_{ijk}$  and normally distributed with zero mean and variance  $\sigma_A^2$ ) and is estimated by  $(\bar{y}_{i..} - \bar{y}_{...})$ ;

$\beta_j$  = fixed treatment effect of factor **B** (such that  $\sum_{j=1}^b \beta_j = \mathbf{0}$ ) and estimated by  $(\bar{y}_{.j.} - \bar{y}_{...})$ ;

$(\alpha\beta)_{ij}$  = the interaction effect (which assumes a random status independent and normally distributed with zero mean and variance  $(\sigma_{AB}^2)$ ) and is estimated by  $(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})$ ;

$E_{ijk}$  = the measure of the deviations of the observed value,  $y_{ijk}$ , in the  $(i,j)$ th cell from  $\mu_{ij.}$ , estimated by  $y_{ijk} - \bar{y}_{ij.}$ .

The model is given by:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk} \quad , \quad (1)$$

where  $i = 1, 2, \dots, a$ ;  $j = 1, 2, \dots, b$  ; and  $k = 1, 2, \dots, n$ .

Neglecting  $E_{ijk}$ ,

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \quad , \quad (2)$$

Using the estimated terms,

$$\hat{y}_{ijk} = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) = \bar{y}_{ij.} \quad . \quad (3)$$

This means, each of  $k$ th observation in the  $(i,j)$ th cell is estimated by the average  $(\bar{y}_{ij.})$  of the  $n$  observations (replicates) in that cell.

The significance of the model is checked by using hypothesis testing based on the  $F$ -ratios from the output of analysis of variance (ANOVA) and the result could further be verified by the coefficient of determination ( $R^2$ ). That is if  $R^2 = r\%$ , then  $r\%$  of the variability of the response in the model is explained by factor A, factor B and the interaction between factors A and B.

Before the conclusions from the analysis of variance are adopted, the adequacy of the model is checked. The primary diagnostic tool for model adequacy checking is residual plots, of which one of them is the normal probability plot. In [12], a residual is essentially an error in the fit of a model.

The residuals for two-factor factorial model are given by:

$$E_{ijk} = y_{ijk} - \bar{y}_{ij}.$$

(4)

For the normal probability plot of residuals, if the underlying error distribution is normal, then the plot exhibit some kind of linearity, hence the adequacy of the model. However, a very common defect that often shows up on this plot is the occurrence of an outlier which can seriously distort the analysis of variance. Mostly, the cause of the outlier is such human error as calculation error, date coding error or copying error. However, a suspected outlier could be checked by examining the standardized residuals value ( $d_{ijk}$ ) given by:

$$d_{ijk} = \frac{E_{ijk}}{\sqrt{MSE}}$$

(5)

A residual value ( $d_{ijk}$ ) bigger than 3 in absolute is a potential outlier which can cause a serious distortion to the conclusion drawn from the ANOVA.

## 2.2 Problem Statement

In 2001, the Ministry of Transportation (MoT) and the National Road Safety Commission (NRSC) launched the first National Road Safety Strategy (NRSS I) for period 2001- 2005 which provided a broad framework for coordinated intervention in road safety. The aim of NRSS I was to reverse the upward trends in road traffic accidents and casualties in the country over the period from 2001- 2005, as in reference [13].

In line with the African Ministerial Conference on Road Safety held in Accra in September 2000, the NRSS I set three strategic objectives, namely:

- 5% reduction in fatalities by the year 2005, using 1998 as the base year.
- 20% reduction in fatalities by the year 2010.
- Development of the capacity to influence the quantity and quality of road safety intervention.

This led to the trends in road traffic fatalities in Ghana declining from 31 per 10,000 registered vehicles in 1998 to 23 per 10,000 registered vehicles in 2005, which was a third of Africa's average of 67 per 10,000 registered vehicles, as in reference [2]. This made Ghana's efforts recognized by World Bank as a good example in Africa and other developing countries, as in reference [13]. Hence, the introduction of the second National Road Safety Strategy (NRSS II), for the period 2006-2010. The objective of NRSS II was to build on the objectives of NRSS I to reduce RTA fatalities on a year-on-year basis and achieve a total of less than 1000 by the year 2015. This was in line with the African objective of 50% reduction in RTA fatalities from 2005 to 2015 set in the 4th African Road Safety Conference held in February 2007 in Accra, [1].

After five years running of NRSS II, this paper assesses its effectiveness using factorial design. For this purpose, data was collected from the Ashanti and Northern regional offices of the Motor Traffic and Transport Unit (M.T.T.U) of Ghana Police Service. The data is based on the yearly number of RTA for the period 2006- 2009 and is shown in Table 2 below. The first column shows the two regions (Ashanti and Northern) as factor A, the second column shows the years as experimental units and the third column shows the accident nature (minor, major and fatal cases and their respective codes in parenthesis) as factor B.

**Table 2:** Number of RTA recorded in Northern and Ashanti Regions for the period 2006-2009.

REGION	YEAR	ACCIDENT NATURE		
		MINOR CASES (1)	MAJOR CASES (2)	FATAL CASES (3)
NORTHERN REGION	2006	86	48	50
	2007	128	47	50
	2008	33	39	46
	2009	77	28	43
ASHANTI REGION	2006	1174	564	290
	2007	633	362	212
	2008	951	558	253
	2009	961	576	194

**2.3 Computational Procedure**

The data was analyzed using MINITAB software installed on a personal computer (laptop) of Lenovo brand with processing speed of 2.00 GHz and installed memory (RAM) of 1.5 GB.

Two-factor factorial design with mixed factors was the experimental design used. Under this, the test statistic *F*-ratio was used for the hypothesis testing of no interaction between accident nature and region (i.e. to find out whether accident nature depends on region or not); and also for the hypothesis testing of no difference among the treatment combination means (i.e. to find out whether there is a difference in variability among the cell means or not). This was followed by the Tukey’s test, under the multiple comparison tests, to verify the results of the hypothesis tests.

Graphical analysis of interaction plot test was used to confirm that NRSS II was effective in the reduction of the seriousness of RTA for the period 2006-2010. Also, a residual plot, precisely normal probability plot, was used to check the adequacy of the model.

### 3. Results

Using the general two-factor factorial design format (Table 1 above), the data layout and computation for the yearly number of RTA for the period 2006-2009 is as shown in Table 3 below. The first column is headed by Region which serves as factor A and comprises two ( $a = 2$ ) random levels, Northern and Ashanti regions. The second column shows accident nature which serves as factor B with three ( $b = 3$ ) fixed levels as the number of minor cases, major cases and fatal cases of RTA in each year. The four-year data for each accident nature in the two regions constitute the observations. The sum of observations and their averages are shown in the various cells of Table 3. The last two columns are respectively the sum of all the observations in each region and their averages. Also, the last two rows are respectively the sum of all the observations under each accident nature and their averages.

**Table 3:** RTA data, in two- factor factorial design presentation, for the effect of NRSS II.

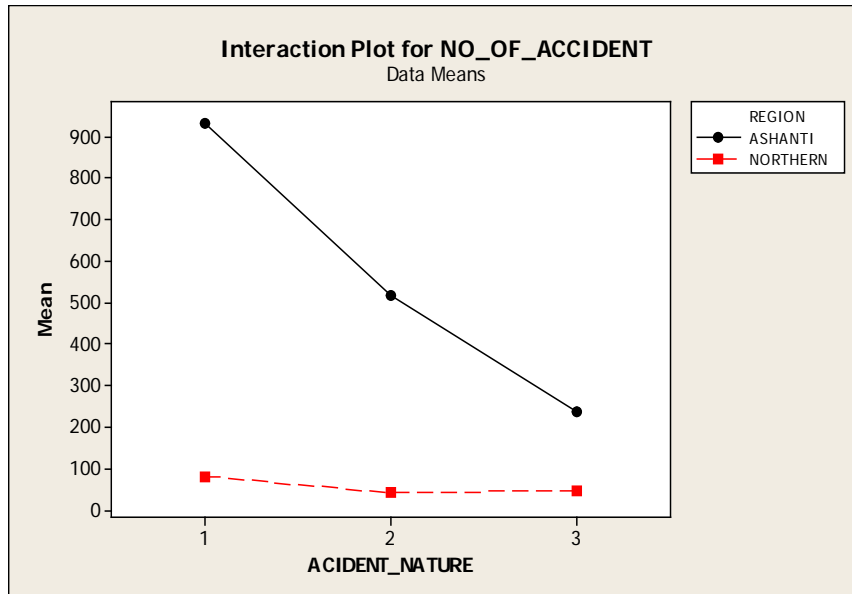
Region (Factor A)	Nature of Accident (Factor B)			Totals	Means
	Number of Minor Cases (j=1)	Number of Major Cases (j=2)	Number of Fatal Cases (j=3)	$y_{i.}$	$\bar{y}_{i.}$
Northern Region (i=1)	86, 128, 33, 77  $y_{11.} = 324$  $\bar{y}_{11.} = 81$	48, 47, 39, 28  $y_{12.} = 162$  $\bar{y}_{12.} = 40.5$	50, 50, 46, 43  $y_{13.} = 189$  $\bar{y}_{13.} = 47.25$	$y_{1.} = 675$	56.250
Ashanti Region (i=2)	1174, 633, 951, 961  $y_{21.} = 3719$  $\bar{y}_{21.} = 929.75$	564, 362, 558, 576  $y_{22.} = 2060$  $\bar{y}_{22.} = 515$	290, 212 253, 194  $y_{23.} = 949$  $\bar{y}_{23.} = 237.25$	$y_{2.} = 6728$	560.667
Totals ( $y_{.j}$ )	4043	2222	1138	$y_{...} = 7403$	$\bar{y}_{...} = 308.458$
Means ( $\bar{y}_{.j}$ )	505.375	277.75	142.25		

#### 3.1 Descriptive Analysis of Data

To assist in interpreting the results of this study, a graph of the mean responses per cell (i.e. cell means) is constructed and this is shown in Fig. 1 below. The vertical axis is scaled with the cell means ( $\bar{y}_{ij.}$ ) of the number of RTA and the horizontal axis is scaled with the accident nature (i.e. minor cases, major cases and fatal cases).



In the body of the plot, the black line indicates the plot of the cell means against the accident nature of Ashanti region while the red-dashes line indicates that of Northern region. Also, from the left, the first marker in each line corresponds to minor cases; the second marker corresponds to major cases; and the third marker corresponds to fatal cases recorded in the respective regions.



**Fig. 1:** The interaction plot of accident nature (minor=1, major= 2 and fatal=3) against the cell means of the number of RTA.

Since the two plotted lines are not parallel, there is significant interaction between accident nature and region. In general, the seriousness of RTA is on the decrease in the two regions. Right from the minor cases through the major cases to the fatal cases, the steepness of the lines shows that the seriousness of RTA in Ashanti region decreases at a higher rate than that in the Northern region. However, looking at the line for Northern region, there is a decrease in the seriousness of RTA from minor cases to major cases but a slight increase from major cases to fatal cases. The plot confirms the fact that the NRSS II has been effective in the reduction of the seriousness of RTA for the period 2006 -2009, only that the effect is not the same across the two regions.

### 3.2 Inferential Analysis of Data

#### 3.2.1 Hypothesis Testing

Using significance level ( $\alpha$ ) of .05 and with two-factor factorial design with mixed factors, the MINITAB output for the raw data displayed in Table 2 is as shown in Table 4.

The statements of the null hypothesis ( $H_0$ ) and the alternate hypothesis ( $H_1$ ) are as follows:

- i. For the hypothesis of no interaction:  $H_0 : \sigma_{AB}^2 = 0$   
 $H_1 : \sigma_{AB}^2 \neq 0$

Therefore, the test statistic, given by:  $F_{AB} = \frac{MS_{AB}}{MS_E} = 20.58$  (shown in Table 4).

From the  $F$  distribution table,  $F_{\alpha(a-1)(b-1),ab(n-1)} = F_{0.05,2,18} = 3.55$ . Since calculated  $F$  for accident nature-region interaction ( $F_{AB}$ ) is equal to  $20.58 > F_{0.05,2,18} = 3.55$ ,  $F_{AB}=20.58$  is significant at  $\alpha = .05$ . Therefore  $H_0 : \sigma_{AB}^2 = 0$  is rejected, indicating that there is statistical evidence that there is interaction between region (factor A) and accident nature (factor B).

ii. The analysis is continued by testing the hypothesis of no difference among the treatment combination means

as:  $H_0 : \sigma_{Tr}^2 = 0$   $H_1 : \sigma_{Tr}^2 \neq 0$

The test statistics ( $F$ - ratio) is given by:  $F_{Tr} = \frac{MS_{Tr}}{MS_E} = 47.18$  (shown in Table 4).

From the  $F$ -distribution table,  $F_{\alpha(ab-1),ab(n-1)} = F_{0.05,5,18} = 2.77$ . Since the calculated  $F$  for treatment combination means difference ( $F_{Tr}$ ) is equal to  $47.18 > F_{0.05,5,18} = 2.77$ , the treatment combination variance is significant at  $\alpha = .05$ . Hence,  $H_0 : \sigma_{Tr}^2 = 0$  is rejected, indicating that there is statistical evidence that there is a difference in variability among the cell (treatment combination) means.

**Table 4:** MINITAB ANOVA output of the number of RTA using two-factor factorial design with mixed factors.

Source of variation	Degrees of freedom (DF)	Sum of squares	Mean squares (MS)	Calculated F	Probability Values (P)
Treatment combination	5	2502009	500402	47.18	0.00
Region (Factor A)	1	1526617	1526617	143.94	0.00
Accident nature (Factor B)	2	538755	269378	25.40	0.00
Region- Accident nature interaction	2	436637	218318	20.58	0.00
Error	18	190911	10606		
Total	23	2692920			
Co-efficient of determination ( $R^2$ )	92.91%				

### 3.2.2 Multiple Comparisons Test

The test statistic for Tukey's (multiple comparisons) test is given by:

$$T_{\alpha} = q_{\alpha[a,ab(n-1)]} \sqrt{\frac{MS_E}{n}}, \Rightarrow T_{0.05} = q_{0.05}(2,18) \sqrt{\frac{10606.181}{4}}.$$

From the studentized range statistic ( $q$ ) distribution table,  $q_{0.05}(2,18) = 2.97$ . This implies that  $T_{0.05} = 2.97 \sqrt{\frac{10606.181}{4}} = 152.935$ . Therefore, from the Tukey's test, the pair-wise comparisons for minor, major and fatal cases given by  $|\bar{y}_{21.} - \bar{y}_{11.}| = 848.75$ ,  $|\bar{y}_{22.} - \bar{y}_{12.}| = 474.5$  and  $|\bar{y}_{23.} - \bar{y}_{13.}| = 190$  respectively. They are all greater than  $T_{0.05} = 152.935$ , hence the number of RTA recorded in Northern and Ashanti regions under minor, major and fatal cases are significantly different.

### 3.3 Model for Number of RTA

The general effect model given by:

$$y_{ijk} = u + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}$$

Where  $i = 1, 2, \dots, a$ ;  $j = 1, 2, \dots, b$  and  $k = 1, 2, \dots, n$ .

#### 3.3.1 Estimation of Model Parameters

Estimating the parameters of the model, the fixed factor effects ( $\mu$  and  $\beta_j$ ) are respectively estimated as  $\hat{\mu} = \bar{y}_{...} = 308.458$  and  $\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$  such that ( $\hat{\beta}_1 = 196.917$ ,  $\hat{\beta}_2 = -30.708$  and  $\hat{\beta}_3 = -166.208$ ), using the values of  $\bar{y}_{.j.}$  in Table 3.

The variance components,  $\sigma_A^2, \sigma_{AB}^2$  and  $\sigma^2$  (for the random effects,  $\alpha_i$  and  $(\alpha\beta)_{ij}$ ) are also respectively estimated as:

$\hat{\sigma}_A^2 = \frac{MS_A - MS_{AB}}{bn} = 126334.238, \hat{\sigma}_{AB}^2 = \frac{MS_{AB} - MS_E}{n} = 51928.028$  and  $\hat{\sigma}^2 = MS_E = 10606.181$  Since none of the estimated variance components values is zero, as confirmed by the corresponding significant  $P$ - values ( $P=0.00$ ) in Table 4, the terms in the effect model are effective. Also, from Table 4 above, co-efficient of determination ( $R^2$ ) = 92.91% means that about 92.91% of the variability in the number of RTA is explained by the region (factor A), accident nature (factor B) and the interaction between region and accident nature.

#### 3.3.2 Model Adequacy Checking

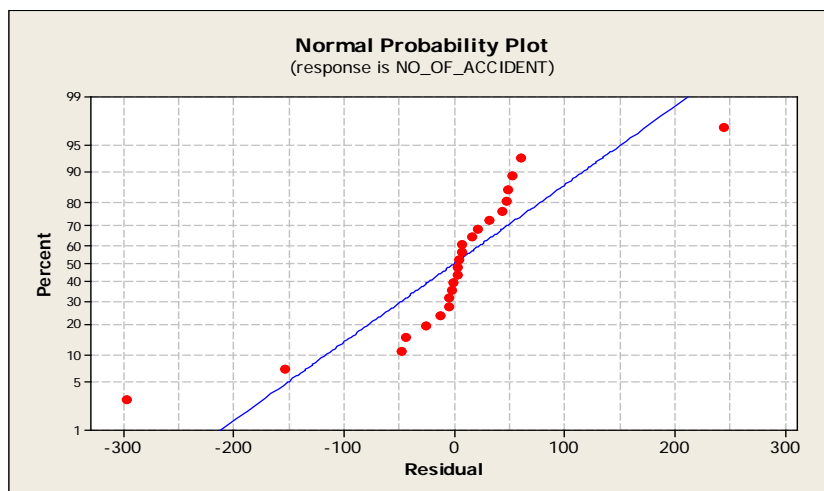
With reference to equation (4), the MINITAB output of residuals of the number of RTA is shown in Table 5 below. In Table 5, the first column deals with region (Northern and Ashanti), the second column is headed by

accident nature (minor cases, major cases and fatal cases) with sub-headings being actual values, predicted values (cell means) and residuals under each accident nature.

**Table 5:** MINITAB output of residuals for number of road traffic accident (R T A).

Region	Accident nature								
	Minor cases (1)			Major cases (2)			Fatal cases (3)		
	Actual values	Predicted values	Residual	Actual values	Predicted values	Residual	Actual values	Predicted values	Residual
Northern	86	81	5.00	48	40.50	7.50	50	47.25	2.75
	128	81	47.00	47	40.50	6.50	50	47.25	2.75
	33	81	-48.00	39	40.50	-1.50	46	47.25	-1.25
	77	81	-4.00	28	40.50	-12.50	43	47.25	-4.25
Ashanti	1174	929.75	244.25	564	515.00	49.00	290	237.25	52.75
	633	929.75	-296.75	362	515.00	-153.00	212	237.25	-25.25
	951	929.75	21.25	558	515.00	43.00	253	237.25	15.75
	961	929.75	31.25	576	515.00	61.00	194	7.25	-43.25

The normal probability plot of the residuals (in Table 5) is as shown in Figure 2 below. The vertical axis is scaled with the percentages of the normal probabilities of accident nature while the horizontal axis is scaled with the residuals. The scatter plot is the red dots while the blue diagonal line is establishing the linearity of the plot.



**Fig. 2:** Normal probability plot of residuals of number of RTA.

Visual examination of Table 5 and the normal probability plot of Fig. 2 reveal three extreme residuals (244.25, -296.75 and -153.00). However, taking the standardized value of the biggest residual among the three (i.e. -296.75) and with reference to equation (5), we have:

$$d_{212} = \frac{E_{212}}{\sqrt{MS_E}} = \frac{-296.75}{\sqrt{1060.18}} = -2.882$$

Since this standardized value ( $d_{212} = -2.882$ ) is less than 3 in absolute, the error effect is negligible and therefore has no effect or influence on the model adequacy, likewise the other two outliers. The linearity of the plot in fig. 2 also proves the normality or the independent assumption of the data, hence the confirmation of the model adequacy.

Therefore, the mixed-effect model:

$$y_{ijk} = u + \alpha_i + \beta_j + (\alpha\beta)ij,$$

Adequately describes the observations of the yearly number of RTA

#### 4. Discussion

From the graphical analysis of the interaction plot in fig. 1, Ashanti region is far ahead of Northern region in terms of the number of RTA recorded under each accident nature. This is confirmed by the corresponding cell means in Table 3. We ascribe this difference to be the result of such physical factors as relative differences in vehicular population, number of motorable roads and the human population in the two regions.

The interaction plot further revealed that there was a general reduction in the seriousness of RTA in both Northern and Ashanti regions, only that the seriousness reduction rate in Ashanti region was higher than that in the Northern region. Hence, the effect of NRSS II in Ashanti region was higher than that in the Northern region. Also, since two-factor factorial design with mixed factors was employed in this study, it is reasonable to generalize that the effect of NRSS II (2006-2010) on RTA was not the same across all the ten regions in Ghana.

Since the calculated  $F$  for accident nature-region interaction ( $F_{AB}=20.58$ ) is significant, there is interaction between region and accident nature at significance level of .05, giving the general indication that accident nature(factor B) is dependent on region (factor A). This could also mean that the effect of NRSS II depends on region. This further necessitated the test for the differences between treatment combination (cell) means, which was also significant at  $\alpha = .05$  with  $F_{Tr}= 47.18$ . This means that there was enough statistical evidence that there is at least a difference between the number of RTA recorded in the Northern region under minor, major and fatal cases, and that of Ashanti region, respectively, at  $\alpha = .05$ . This also gives a clear indication that the effect of NRSS II in the reduction of RTA in the Northern region is different from that in the Ashanti region.

The parameters of the mixed-effect model:

$$y_{ijk} = u + \alpha_i + \beta_j + (\alpha\beta)ij + E_{ijk}$$

Were well estimated with the fixed factor effect terms ( $u$  and  $\beta_j$ ) and the random factor effect terms ( $\alpha_i$ ,  $(\alpha\beta)ij$  and  $E_{ijk}$ ) all being fitted in the model.

Also, the coefficient of determination ( $R^2$ ) = 92.91% proved that about 92.91% of variability in the mixed-effect model of the number of RTA is explained by the region (factor A), accident nature (factor B) and region-accident nature interaction.

The residual plot, precisely the normal probability plot, showed linearity. The error effect was neglected since the standardized residual value ( $d_{212}$ ) = -2.882 is less than absolute 3. Hence, the model could basically be used for accurate and adequate description of the various observations in the data for the number of RTA recorded by the M.T.T.U of the Ghana Police Service for the period 2006-2010.

Last but not the least, our study showed that even before the end of the programme, the seriousness of RTA was reducing.

## 5. Conclusion

Two-factor factorial design was successfully used for the RTA data layout, as desired. The test of no interaction being significant at  $\alpha = .05$  means that there is statistical evidence that accident nature depends on region. Also, the test of cell means difference was significant at  $\alpha = .05$ , giving the indication that the effect of NRSS II is not the same across the two regions, Northern and Ashanti.

There was also an indication from the steepness of the lines in the interaction plot that the seriousness reduction rate of RTA in Ashanti region is higher than that in the Northern region, meaning the effect of NRSS II in Ashanti region is greater than that in Northern region. This further gives the indication that the campaign on NRSS II has not gotten the same impact on RTA across all the regions in Ghana. All the same, the NRSS II generally helped in the reduction of the seriousness of RTA.

The following mixed-effect model is appropriate and adequate in the description of the observations in the data for the number of RTA for the period 2006-2010:

$$y_{ijk} = u + \alpha_i + \beta_j + (\alpha\beta)ij + E_{ijk}$$

The uniqueness of this study, among other published work, lies in the fact that this is the first time two-factor factorial design with mixed factors is being used to determine the effectiveness of a road safety programme.

Following the successful use of the two-factor factorial design in this study, it is recommended that  $2^k$  and  $3^k$  factorial design be used in some of the subsequent RTA studies based on the fact that the  $k$  factors involved are of equal levels.

## References

- [1] Ministry of Transportation. "Launching of National Road Safety Strategy II," Accra, Ghana, 2006.
- [2] WHO. "Road Safety Conference," Lome, Togo, 2006.
- [3] Building and Road Research Institute. "Launching of National Road Safety Strategy II," Accra, Ghana, 2006.
- [4] M. Chong, A. Abraham and M. Paprzycki.(2005, Dec). "Traffic Accident Analysis Using Machine learning Paradigms". *Informatica*. [On-line]. 29(1), pp. 89-98. Available: [www.informatica.si/PDF/29-1/09](http://www.informatica.si/PDF/29-1/09). [Jul. 2, 2013].
- [5] P. J. Ossenbruggen, J. Pendharkar and J. Ivan. (2001, Jul). "Roadway safety in rural and small urbanized areas". *Accidents Analysis and Prevention*. [On-line]. 33(1), pp. 485-498. Available: [www.sciencedirect.com/science/article](http://www.sciencedirect.com/science/article). [Jul. 10, 2013].
- [6] M. Getnet. (2009, Sep). "Applying data mining with decision tree and rule induction techniques to identify determinant factors of drivers and vehicles in support of reducing and controlling road traffic accidents". *International Journal of Tomography and Statistics*. [On-line]. 22(1), pp.103-112. Available: [www.softcomputing.net/tomography.pdf/article](http://www.softcomputing.net/tomography.pdf/article). [Jan. 24, 2014].
- [7] L. Chang and W. Chen. (2005, Oct). "Data mining of tree-based models to analyze freeway accident frequency". *Journal of Safety Research*. [On-line]. 36(1), pp. 365-375. Available: [www.sciencedirect.com/science/article](http://www.sciencedirect.com/science/article). [Jun 2013].
- [8] S. Hyungwon and S. Sohn (2001, Jan). "Pattern recognition for a road traffic accident severity in Korea". *Ergonomics*. [on-line]. 44(1), pp. 107-117. Available: [www.ncbi.nlm.nih.gov/pubmed/11214896](http://www.ncbi.nlm.nih.gov/pubmed/11214896). [Aug. 8, 2013].
- [9] A.P. Akgungor and O. Yildiz (2007, Jan). "Sensitivity Analysis of an Accident Prediction Model by the Fractional Factorial Method". *Accident Analysis and Prevention*. [On-line]. 39(1), pp. 63-68. Available: [www.ncbi.nlm.nih.gov/pubmed/16930519](http://www.ncbi.nlm.nih.gov/pubmed/16930519). [Aug. 8, 2013].
- [10] A.K. Chhotu and C. S. Kumar (2014, Apr). "Willingness to Pay for Better Safety on State Highways". *International Journal of Civil Engineering Research*. [On-line]. 5(4), pp. 407-410. Available: [www.ripublication.com/ijcer.htm/ijcerv5n4spl\\_16.pdf](http://www.ripublication.com/ijcer.htm/ijcerv5n4spl_16.pdf). [Jul. 6, 2014].
- [11] J. S. Milton and J. C. Arnold. *Introduction to Probability and Statistics*. New York: McGraw-Hill Inc, 1995, pp. 604-655.
- [12] R. E. Walpole, H.R. Myers, S. L. Myers and V. Ye. *Probability and Statistics for Engineers and Scientists*. New Jersey: Pearson Education Inc, 2007, pp. 573-609.

[13] National Road Safety Commission. "Road Safety Situation in Ghana", Presented at the Launching of National Road Safety Strategy II, Accra, Ghana, 2006.

## Appendix

### CALCULATION OF SUM OF SQUARES

Manually, the required sums of squares, with reference to table c above are:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{abn} = 2692919.958, \quad SS_A = \frac{1}{bn} \sum_{i=1}^a y_{i..}^2 - \frac{y_{...}^2}{abn} = 1526617.041$$

$$SS_B = \frac{1}{an} \sum_{j=1}^b y_{.j.}^2 - \frac{y_{...}^2}{abn} = 538755.083, \quad SS_{Tr} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b y_{ijk}^2 - \frac{y_{...}^2}{abn} = 2502008.708$$

$$\text{From } SS_{Tr} = SS_A + SS_B + SS_{AB}, \quad SS_{AB} = SS_{Tr} - SS_A - SS_B = 436636.584$$

$$\text{From } SS_T = SS_{Tr} + SS_E, \quad SS_E = SS_T - SS_{Tr} = 190911.25$$

Hence, the required mean squares are:

$$MS_A = \frac{SS_A}{a-1} = 1526617.041, \quad MS_B = \frac{SS_B}{b-1} = 269377.542, \quad MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} = 218318.292$$

$$MS_E = \frac{SS_E}{ab(n-1)} = 10606.181, \quad MS_{Tr} = \frac{SS_{Tr}}{ab-1} = 500401.742$$

### CALCULATION OF CO-EFFICIENT OF DETERMINATION ( $R^2$ )

With reference to table 4 above, the sum of squares model given by:

$$SS_{Model}(SS_{Tr}) = SS_A + SS_B + SS_{AB} = 1526617 + 538755 + 436637 = 2502009$$

$$\text{Hence, } R^2 = \frac{SS_{Model}}{SS_T} = \frac{2502009}{2692920} = 0.9291$$