International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 23
December 2018

The Assessment and Comparison of Performance Analysis Using Exponentially Weighted Moving Average and Cumulative Sum Scheme.

Bature, Tajudeen Atanda
Alao, Nurudeen Abiodun
Danjuma, Habiba

Department of Science & Technology Education, Al-Hikmah University, Ilorin
Department of Statistics, Kwara State Polytechnic, Ilorin
Department of Statistics, Federal Polytechnic, Bali

Email: tajudeenatanda1@gmail.com

ABSTRACT

This research work is designed for monitoring some prominent road traffic offences in Kwara state of Nigeria .A historical value of the most prominent road traffic offences were used to construct the prediction interval using the non-parametric and bootstrap methods at two different significant levels. The prediction intervals obtained can be used as tool to identify future values that suggest changes in the frequencies the road traffic offences considered in the study. The interval can be used to predict changes in the future values. The two approaches are compared and random sample of sizes 35 and 25 from the studied data were used to evaluate the robustness in variation of the two approaches. The results indicate that the two approaches exhibit similar patterns. However, the bootstrap approach prediction interval is more stringent than the non-parametric approach.

Keywords: Cumulative Sum Scheme (CSS), Exponentially Weighted Moving Average (EWMA), Non-parametric and Bootstrap methods, Road Traffic.

1.0 INTRODUCTION

Quality control is the use of techniques and activities to achieve sustain and improve the quality of a product or service. Quality can be viewed from the different approaches of the quality gurus like Philip B. Crosby defined quality as conformance to requirement. Joseph M. Juran described quality as ‘fitness for purpose’, that is a product or service should do exactly what it is designed to do. W. Edwards Deming stated that quality should be aimed at the views of the customer’s present and future needs.

The primary mandate of Federal Road Safety Commission (FRSC) among others is to curtailed road accidents on our roads which can be achieved by thorough monitoring of road traffic offences which can be viewed as rendering services to the citizens of Nigeria and the more the road traffic offences are curtailed, the better the customers (citizens) are satisfied.

Road traffic offences can be seen as a defect to the norm of the country; the proportion of defects in a process (service-oriented or manufacturing) that is attributable to a specific cause can be used to monitor the maintenance action or evaluate intervention programme of FRSC. If the historical data are assumed to be a representative of the process when it is under control, prediction intervals can be used to identify future values that suggest a change in the root cause composition of defects.

The main objective of this research paper is to design and compare the performance of CUSUM and EWMA chart.

2.0 LITERATURE REVIEW

The literature deals with the collection of other relevant research works that will provide bases for present research the collection, collation and coordination of related literature are of paramount importance in every study. Review of related literatures in statistical research is sometimes very elaborate and very necessary due to its vast areas of applications.

Dugans (1974), Lucas (1976),Hawkins (1992), stated that the Cumulative Sum Control (CUSUM) chart is much more efficient than the usual X-bar control chart for detecting small variations in the average.

Growlers (1987, 1989) and Lucas & Succucci (1990) presented the exponentially weighted moving average control chart, so good choice to detect small change in process average. Various modification and supplemental criteria have also suggested that control charts based on moving average is also an effective chart in detecting chart in detecting small process change.

Most of CUSUM applications for continuous variable are base on the assumption that the process is normally distributed, the reason being that the parameters to be determined are usually based on a very large and a fairly uniform sample size so that the assumptions of normality and constant variability are reasonably appropriate, Ewan (1983).

Lucas (1985) suggested that the poison CUSUM should be used when it is administratively convenient to record the number of counts in a given sampling interval. When the number of counts follows a poison distribution, the time between events CUSUM scheme follows Exponential distribution.

3.0 Research Methodology

This section describes the statistical tools used to meet the study objectives. The principles of Pareto analysis, the Bootstrap and Non-Parametric prediction Intervals techniques are discussed.

3.1 DATA SOURCES

The data used for this study is secondary data from the road traffic offences record office, Zone RS8 Command Headquarter of the Federal Road Safety Commission (FRSC), Offa Road, GRA, Ilorin, Kwara State. It consists of the thirty eight classes of road traffic offences. This command cover three states and are under Zone RS8.These states and their various sector commands are Kwara state sector command (RS8.1), Ekiti state sector command (RS8.2) and Kogi state sector command (RS8.3).The data is record of road traffic offences in these three states for a period of four years, 2009 - 2013 (forty eight months).

3.2 THE PARETO ANALYSIS.

Pareto Analysis is a statistical technique in decision making that is used for the selection of a limited number of tasks that produce significant overall effect. Pareto analysis is a relatively simple methodology that is used when trying to determine which tasks or factors have the most impact (Cervone, 2009). It ranks the data/factors in the descending order from the highest frequency of occurrences to the lowest frequency of occurrences. The total frequency is summed to 100 percent. It uses the Pareto Principle (also known as the 80/20 rule) the idea that 20% of the work generates 80% of the benefit of doing the whole job. Or a large majority of road traffic offences (80%) are caused by a few key ones (20%).

3.2 PREDICTION INTERVALS

The point and interval estimations of mean provide good information on the unknown parameter μ of a normal or non-distribution from which a large sample is drawn. A prediction interval represents the range where a single new observation is likely to fall given specified settings of the predictor. Prediction intervals account for the variability around the mean response inherent in any prediction. Like confidence intervals, prediction intervals have confidence level and can be a two –sided range, or an upper or lower bound. Unlike confidence intervals however, prediction intervals predict the spread for individual observation rather than the mean. Sometimes, other than the population mean, the experimenter may be interested in predicting the possible values of future observations. For instance, in a quality control case, the experimenter may need to use the observed data to predict a new observation. In the scenario, prediction interval can be constructed for observed values such that the Lower and Upper limit in the prediction interval represent the Lower and the Upper Control Limit respectively.

3.2.1 Non-Parametric Prediction Interval

Non-Parametric prediction interval is a powerful statistical tool that uses only few assumptions. Based on the historical values, precise probabilities for the future values are defined on the lower and upper probabilities.

Let $P = (P_{i1}, P_{i2}, \ldots, P_{in})$ denote the data set, the distribution of $P_{i,n+1}$ is derived by Daniel and Norman (1997) and presented as follows:

For any two arbitrary functions *f(p)* and *g(p)*

$$\Pr[f(p) < P_{i,n+1} < g(p)] = E\{[f(p) < P_{i,n+1} < g(p)]\} \quad \text{----------} \ (I)$$

Suppose we let *f(p)* = *p*(*r*) and *g(p)* = *p(s)* where $1 \le r < s \le n$. Furthermore, *p(k)* denote a larger order statistics among the classes of *P*. Then (I) becomes:

$\Pr[p(r) < P_{i,n+1} < p(s)] = E[V(s) - V(r)]$ ---------- (II)

where V(r) and V(s) are the r^{th} and s^{th} statistics from sample of size n independent and identically Uniform [0, 1] observation. (Mood, Graybill and Boes; 1974.)

From David (1981), expression in (II) is equivalent to

$(S - r)/(n+1)$ .Therefore, a 100% prediction interval for $P_{i,n+1}$ is any interval of the form:

$[p(r), p(s)]$ ----------- (III)

With r and s chosen such that

$\frac{(S-r)}{(n+1)} = 1-\alpha$ ------------ (IV)

The Lower prediction interval is the r^{th} observation on the order statistic and upper prediction interval is the s^{th} observation on the order statistics.

3.3.2 The Bootstrap Method

Principle

The bootstrap is a recent developed technique for making certain type of statistical inferences. It is only recently developed because it requires modern computer power (i.e. statistical packages like R program, latest SPSS version) to simplify the often rigorous and intricate calculations of traditional statistical theory. Bootstrap techniques are re-sampling methods for assessing uncertainty. They are useful when inference is to be based on complex procedure for which theoretical results are unavailable or not useful for the sample size met in practice.

The basic idea of the bootstrapping method is that, in absence of any other information about the distribution, the observed sample contains all the available information about the underlying distribution, and hence re-sampling the sample is the best guide to what can be expected from re-sampling from the distribution.

Suppose that a sample X of size n is given as X = $(X_1, X_2, \ldots X_n)^T$ is used to estimate a parameter θ of the distribution and let θ = w(X) be a statistic that estimates θ. For the purpose of statistical inference on θ, we are interested in the sampling distribution of θ so as to assess the accuracy of our estimator or to set confidence intervals for our estimate of θ.

If the true distribution P were known, we could draw samples **X (k)**, k = 1, 2, …, K from P and use Monte Carlo methods to estimate the sampling distribution of our estimate θ. Since P is unknown and we cannot sample from it, the bootstrapping idea suggests re-sampling the original sample instead. This distribution from which the bootstrap samples are drawn is the empirical distribution.(*Efron, B. (1979). Bootstrap methods: another look at the jackknife. Annals of Statistics 7,1-26.)*

The procedure for bootstrap estimate of the sampling distribution of θ involves computations using Monte as follow:

3.3 Random Number Generator

This is statistical software that statistical that can be used to generate random number. The numbers are generated with a uniform distribution – that is, no number within the specified range is any more or less likely to appear than any other number.

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 23
December 2018

The mechanism is as follow:

Generate a random number between [1] **and** [48] **= 35 Get number**

- Specify the range of the number
- Click on "Get number".

4.0 RESULTS AND DISCUSSIONS

4.1 PARETO ANALYSIS OF ROAD OFFENCES DATA

The Pareto chart of the number of people who committed the road traffic offences is plotted. The right side of the Pareto chart is frequency of the offenders and the left side is the cumulative percentage of the offenders.

The Pareto chart is shown in figure 1 below.

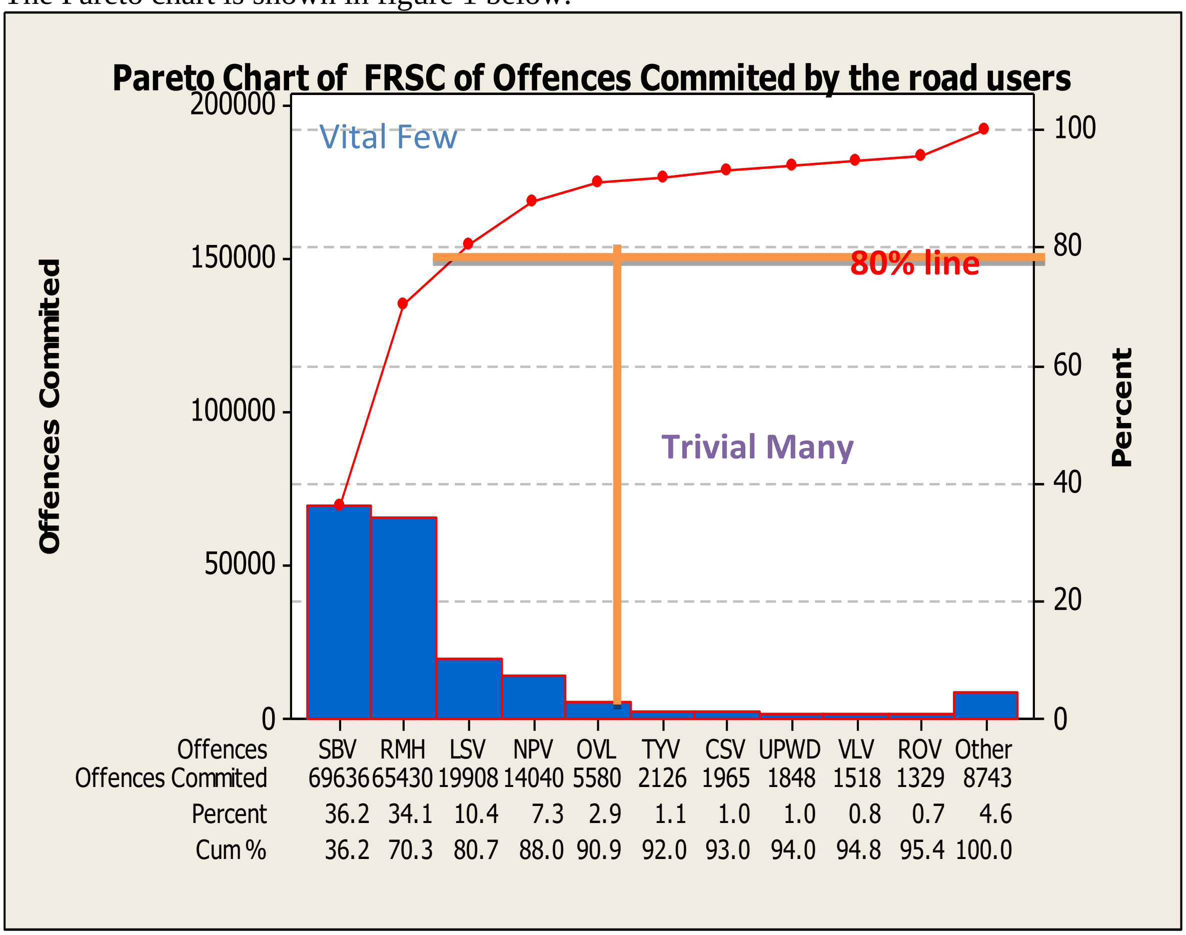

Fig. 1 *Pareto chart showing the FRSC road traffic offences committed by the road users.*

From the Pareto chart in Figure 1 above, the vital few that contributed 80% 0f the offences committed include: Seat Belt Violation (SBV), Riding Motorcycle without crash Helmet (RMH),

Light/Sign Violation (LSV), Plate Number Violation (PNV) and Overloading (OVL) while the remaining offences are the trivial few that contributed just 20% to the offences.
Therefore, prediction intervals will be constructed for these five prominent offences as reflected by Pareto chart.

4.2 DETERMINING THE PREDICTION INTERVAL FOR THE PROMINIENT OFFENCES

From the result obtained from Pareto chart, the proportion P_i of the number of the people that committed offence i within a month will be use to construct the prediction interval.

$$P_i = \frac{X_i}{\sum_i^n X_i} \quad i = 1, 2, \ldots, 30.$$

Where X_i and $\sum_i^n X_i$ denote the number of people that committed offence i within a month and total number of people that committed all the offences with a month.
The prediction interval for P_i which will capture the value of some specified probability $1-\propto$ can be constructed. Such intervals are very useful FRSC, most especially for process management whenever the prediction interval falls outside the prediction interval, necessary measure /intervention will be needed.

4.2.1 Determining the Prediction Interval Using the Non Parametric Method

For the studied data, the number of n months considered is 48 (n=48). Therefore using the equation derived by David (1981) which is given by:

$$\frac{(s-r)}{(n+1)} = 1-\alpha \qquad (1 \leq r < s \leq n)$$

Where **r** and **s** are the rth and sth ordered statistics from a sample of size n independently and identically distributed Uniform [0, 1] observation.
The prediction interval for future values $p_{i,\, n+1}$ is the value corresponding to [r, s] from the studied data.
When α = 0.05, then r =1 and s = 47 i.e. $\frac{(47-1)}{(49)} \approx 0.95$.
Using the data on SBV, for instance, the Lower and Upper prediction interval is given as [0.1345, 0.5535].
Similarly, using the data on RMH, the Lower and Upper prediction interval is given as [0.0765, 0.4616].
when α = 0.10, r =2 and s = 46 i.e. $\frac{(46-2)}{(49)} \approx 0.90$.
Using the data on SBV and RMH, for instance, the Lower and Upper prediction intervals are given as [0.1820, 0.4939] and [0.1635, 0.4490] respectively.
Following this procedure, the prediction intervals for the five offences are computed and presented in table 1

Table1: The Prediction Interval Using the Non Parametric Method.

OFFENCES	α	Lower	Upper	Ranges

SBV	**0.05**	0.1345	0.5535	0.4190
	0.10	0.1820	0.4939	0.3119
RMH	**0.05**	0.0765	0.4616	0.3851
	0.10	0.1635	0.4490	0.2855
LSV	**0.05**	0.0032	0.1900	0.1868
	0.10	0.0070	0.1686	0.1616
NPV	**0.05**	0.0245	0.1391	0.1146
	0.10	0.0259	0.1325	0.1066
OVL	**0.05**	0.0023	0.0945	0.0922
	0.10	0.0039	0.0945	0.0906

From Table 1 above, we are 95% and 90% confident that the proportion of the road traffic offences will be within the intervals as presented above.

4.2.2 Determining the Prediction Interval Using the Bootstrap Method

Using the stepwise procedure for bootstrapping,(Using SPSS) the following results are obtained and presented in Table2 below.

Table2: Prediction Interval Using Bootstrap Method

OFFENCES	α	**Lower**	**Upper**	**Ranges**
SBV	**0.05**	0.2023	0.3791	0.1768
	0.10	0.2205	0.3622	0.1417
RMH	**0.05**	0.2371	0.4045	0.1674
	0.10	0.2502	0.3964	0.1438
LSV	**0.05**	0.0542	0.1111	0.0569
	0.10	0.0564	0.1072	0.0508
NPV	**0.05**	0.0492	0.0794	0.0302
	0.10	0.0503	0.0782	0.0279
OVL	**0.05**	0.0224	0.0621	0.0397
	0.10	0.0233	0.0603	0.0370

From Table 2, we are 95% and 90% confident that the proportion of the road traffic offences will be within the intervals as presented above

4.3 EVALUATION OF THE TWO PREDICTION INTERVALS.

To evaluate the two methods discussed in this project work, the parameters of the intervals are varied and the differences in the prediction intervals based on the approaches are computed.

4.3.1 The Non Parametric Prediction Intervals.

The limit of the prediction intervals for the non-parametric method on varying the sample sizes, say n = 48, 35 and 25 at varying level of significance say α = 0.05 and 0.10 are computed. It should be noted that the value of n chosen are obtained using a statistical software called random number generator where each sample has equal chance of been selected. When n = 35 and α = 0.05, then the value of **r** and **s** that satisfies the equation

$$\frac{(s-r)}{(n+1)} = 1-\alpha$$

is given by r = 1 and s =35 i.e. $\frac{(35-1)}{(35+1)} \approx 0.95$.

When n = 35 and α = 0.10, then value of **r** and **s** that satisfied the equation

$$\frac{(s-r)}{(n+1)} = 1-\alpha$$

S is given by r = 2 and s =34 i.e. $\frac{(34-2)}{(35+1)} \approx 0.90$

When n = 25 and α = 0.05, then value of **r** and **s** that satisfied the equation

$$\frac{(s-r)}{(n+1)} = 1-\alpha$$

is given by r = 1 and s =25 i.e. $\frac{(25-1)}{(25+1)} \approx 0.95$.

When n = 25 and α = 0.05, then value of **r** and **s** that satisfied the equation

$$\frac{(s-r)}{(n+1)} = 1-\alpha$$

is given by r = 1 and s =24 i.e. $\frac{(24-1)}{(25+1)} \approx 0.90$.

The above values are used in the computation of the non- parametric prediction intervals and are presented in Table3

Table3: Non Parametric Prediction Interval at varying sample sizes.

Offences	n	$\alpha = 0.05$			$\alpha = 0.10$		
		Lower	**Upper**	**Range**	**Lower**	**Upper**	**Range**
SBV	**48**	0.1345	0.5535	0.4190	0.1820	0.4939	0.4190
	35	0.1345	0.5959	0.4614	0.1820	0.5536	0.3716
	25	0.1345	0.5959	0.4614	0.1345	0.4650	0.3305
RMH	**48**	0.0765	0.4616	0.3851	0.1635	0.4490	0.2855
	35	0.0765	0.4854	0.4089	0.1635	0.4616	0.2981
	25	0.0765	0.4854	0.4089	0.0765	0.4490	0.3725
LSV	**48**	0.0032	0.1900	0.1868	0.0070	0.1686	0.1616
	35	0.0032	0.1900	0.1868	0.0070	0.1686	0.1616
	25	0.0032	0.1900	0.1868	0.0032	0.1686	0.1616
NPV	**48**	0.0245	0.1391	0.1146	0.0259	0.1325	0.1066
	35	0.0245	0.1062	0.0817	0.0259	0.1047	0.0788
	25	0.0245	0.1062	0.0817	0.0245	0.0918	0.0673
OVL	**48**	0.0023	0.0945	0.0922	0.0039	0.0945	0.0906
	35	0.0023	0.1432	0.1409	0.0039	0.0982	0.0943

	25	0.0023	0.1432	0.1409	0.0032	0.0982	0.0950

4.3.2 The Bootstrap Prediction Interval

The limits of the prediction interval for the bootstrap method on varying the sample sizes (n = 48, 35 and 25) at significant level of 5% and 10% are computed. The values of the intervals for the different sample sizes for the offences under consideration are presented in Table4.

Table 4: Bootstrap Prediction Intervals at Varying Sample sizes

Offences	n	$\alpha = 0.05$			$\alpha = 0.10$		
		Lower	**Upper**	**Range**	**Lower**	**Upper**	**Range**
SBV	**48**	0.2023	0.3791	0.1768	0.2205	0.3622	0.1417
	35	0.3263	0.4932	0.1669	0.3201	0.4832	0. 1631
	25	0.3481	0.5111	0.1630	0.3563	0.4941	0.1378
RMH	**48**	0.2371	0.4045	0.1674	0.2502	0.3964	0.1438
	35	0.2072	0.3743	0.1671	0.2183	0.3671	0.1488
	25	0.1794	0.3341	0.1547	0.1992	0.3223	0.1231
LSV	**48**	0.0542	0.1111	0.0569	0.0564	0.1072	0.0508
	35	0.0693	0.1372	0.0679	0.0741	0.1312	0.0571
	25	0.0531	0.1463	0.0932	0.0681	0.1442	0.0761
NPV	**48**	0.0492	0.0794	0.0302	0.0503	0.0782	0.0279
	35	0.0461	0.0872	0.0411	0.0473	0.0901	0.0428
	25	0.0413	0.0872	0.0459	0.0432	0.0772	0.0340
OVL	**48**	0.0224	0.0621	0.0397	0.0233	0.0603	0.0370
	35	0.0221	0.7401	0.7180	0.0261	0.0822	0.0561
	25	0.0332	0.0823	0.0491	0.0383	0.0743	0.0360

5.0 Finding and Conclusion

From Figure1, the Pareto Chart shows that only five of the road traffic offences contributed 83% to the overall offences. In summary, the percentage of each of the five most prominent road traffic offences which include: Seat Belt Violation (SBV), Riding Motorcycle without crash Helmet (RMH), Light/Sign Violation (LSV), Number Plate Violation (NPV) and Overloading (OVL) are 32%, 30%, 11%, 8% and 2% respectively while the 25 remaining offences are the trivial many that contributed just 17% to the overall road traffic offences.

From Table1 and Table2, the values of the lower limits for the Non parametric and Bootstrap prediction intervals generally increase in values as α assumes higher values (i.e. as α increases from 0.05 to 0.10) and generally decrease in the values of the upper limits in the same direction for the five offences under study. The ranges of the prediction limits also decreases as α level assume higher values for the two methods. Table1 and table 2 also reveal that the non-parametric and Bootstrap method have the same pattern.

Table3 and Table4 reveal that the changes in sample sizes (n= 25, 35 & 48) does not result in any consistent pattern in the values of prediction limits and the ranges of the prediction intervals. However, it was observed that the prediction Intervals for the bootstrap method is more stringent than that of Non-Parametric approach. However, the prediction intervals for the two methods are generally similar.

5.2 CONCLUSION

From the results obtained in the data analysis, the following conclusions can be drawn:
The most prominent road traffic offences which are the "Vital Few" are: Seat Belt Violation (SBV), Riding Motorcycle without crash Helmet (RMH), Light/Sign Violation (LSV), Number Plate Violation (NPV) and Overloading (OVL) while the remaining 25 offences are the "Trivial Many". Prediction interval for these prominent offences can be is a very useful tool to road safety agencies managers. This is because it can be used to monitor the proportion of offences by Road Safety agencies. Thus, the results in the research can be used for appropriate intervention before many casualties are recorded which could lead to accident and eventually loss of lives. Since the prediction intervals for the two methods have same similar pattern, bootstrap method may be preferred as the intervals are more stringent and easier to operate.

REFERENCES

[1] Adekunle J.A. (2010). Road traffic accident deaths and socio-economic development in Nigeria. *Int.Rev.Bus.Soc.Sci.* 1(5):47-50.

[2] Adekeye, K.S., Lamidi, M.A. and Osanaiye, P.A.(2010). Prediction intervals: A tool for monitoring outbreak of some prominent diseases; *Global Journal of Mathematics and Statistics*, 2(1), 41-46.

[3] Cutter, S.L.(1993).Living with Risks: Geography of Technological Hard. Great Britain: Edward Arnold

[4] David, H. A. (1980) Order Statistics, John Willey and Sons, New York.

[5] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7, 1-26.

[6] Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap, Chapman & Hall,

[6] Federal Road Safety Commission (Establishment) Act 2007 No. 22 A 583.

[7] Federal Road Safety Commission Annexure1 (further instruction pdf). Retrieved on 08-18-2014.

[8] Gururaj, G. (2004). Alcohol and Road Traffic Injuries in South Asia*: Challenges for Prevention,* JCPSP, 14(12):713-718.

[9] Hawng, G.J.T. and Ding, A. A. (2012): Predictor intervals for Artificial Neural Networks, *Journal of American Statistical Association*. 92(438), 748-757.

[10] Jacob, G. Aeron, T.A. and Astrop, A. (2000).Estimating Global road fatalities, London England: *Transport Research Laboratory Report 445.*

[11] Karuppusami, G. and Gandhinathan, R., (2006): Pareto analysis of critical success factors of total quality management. *The TQM Magazine*, 18(4), 372-385.

[12] Komba, E. (2006). Traffic fatalities and Economic growth Accident Anal.37:167-178.

[13] Kual, A. Sinha, N.D., Sukka, S.K. and Sinha, A.K (2010). Epidemiological Study of Road Accident Cases from West Nepal, India J. Community Med. 35(1): 115-121.

[14] Mayr, A. (2012): Prediction intervals for Future BMI values of individual children a non-parametric approach by quantile boosting, BMC, Medical research methodology Journal, 12(6).

[15] Mind Tools. "Pareto Analysis: Selecting the Most Important Changes to Make." www.mindtools.com/pages/article/newTED.

[16] Mohan, D (2007). Road Safety in less-motorized environments: Future Concerns. International Journal of Epidemiology. 31:327-532.

[17] Odero, W (1998). Alcohol-related road traffic Injuries in Eldoret, Kenya, East Africa Medical Journal. 75:708-711.

[18] Pludemmann, A. Parry, C.D.H., Donson, H. and Sukhai, A. (2004). Alcohol use and trauma in Cape Town, Durban and Port Elizabeth, South African: Injuries Control and Safety Promotion, 11(4): 265-267.

[19] Pratte, D. (1998): Road to Ruin: Road Traffic Accident in the developing World.NEXUS:13:46-62.

[20] World Health Organization (2002). A 5 year WHO Strategy for road traffic Injury prevention. Retrieved from //http:whq.bdoc.who.int/hq/2001/WHO.

[21] World Health organization (2004). World Report on Road Traffic Injury Prevention

[22] Rodriquez .A and Ruiz E. (2009); Bootstrap Predictions in state-space, *Journal of Time Series Analysis*, Blackwell, 30(2))

[23] Tanvir, Ahmed, Raj Narayan Acharjee, M.D. Abdur Rahim, Noman Sikder, Taslima Akther, Mohd. Rifat Khan, MD.Fazle Rabbi and Anup Saha, (2013): An Application of Pareto Analysis and Cause and Effect Diagram for Minimizing Defect Percentage in Sewing Section of a Garment

Factory in Bangladesh. *International Journal of Modern Engineering Research.(www.ijmer):* 3(6), 3700 – 3715.