Regression Control Chart for Monitoring Road Accident Fatalities

Aako, Olubisi Lawrence* and Ajibode, Ilesanmi Akanbi
Department of Mathematics and Statistics, The Federal Polytechnic, Ilaro, Ogun State.
*olubisi.aako@federalpolyilaro.edu.ng, *08062151141, ilesanmi.ajibode@federalpolyilaro.edu.ng

*Corresponding author

Abstract:

This paper introduces the regression control chart for monitoring the road accident fatalities considering the effect of the external covariate, the number of road accident crashes. The recorded occurrence of road accident fatalities and crashes in a state in the western part of Nigeria was monitored using the conventional individual control chart and regression control chart which integrates two robust methods as the basis of analysis: the linear regression model and the Shewhart control chart theory. The designed conventional individual control chart detects month when road accident fatalities is out-of-control while the regression control chart could not detect any out-of-control. It is therefore recommended that regression control chart be used when the effect of external covariate(s) are being considered.

Keywords: Control chart; Fatalities; Monitoring; Prevalent; Regression; Road Accident

1. Introduction

Road traffic accident is an unexpected phenomenon that occurs as a result of the operation of vehicles. More than 90% of road traffic deaths occur in low- and middle-income countries. Road traffic injury death rates are highest in the African region which includes Nigeria. The increasing rate of fatality of road traffic accident in Nigeria is quite alarming and worrisome considering the considerable economic losses to individuals, their families, and to nations as a whole suffered as a result of traffic accident fatalities (WHO, 2020). Thus, road accident has become a major concern and analyzing accident data has been an important look out to the analysts and significant efforts have been made over the past few decades aimed at developing statistical models for monitoring accident fatalities.

Some authors have worked on road accident data; Adekeye and Aluko (2012) designed Cusum scheme for monitoring road accident fatalities, Braimah et al. (2014) studied the application of count data cumulative sum (CUSUM) control chart in monitoring road traffic crashes. Faweya et al. (2017) examined the daily Road Traffic Crashes (RTC) casualties in Osun state of Nigeria, Comparison of two quality control schemes identified as exponential Weighted Moving Average

(EWMA) and Poisson Cumulative sum (CUSUM) control chart techniques were made using data collected from Federal Road Safety Corps (FRSC) Osun-State. Ihueze and Onwurah (2018) analyzed road traffic crashes in Anambra State, Nigeria with the intention of developing accurate predictive models for forecasting crash frequency in the State using autoregressive integrated moving average (ARIMA) and autoregressive integrated moving average with explanatory variables (ARIMAX) modeling techniques.

Most of the authors do not consider the effect of the relationship between a response of variable and one or more explanatory variables. In practice, there are situation when the process output is affected by external covariates. The process output may appear to be in the out-of-control status even though there is actually no operation problem with the process. In this study, regression control chart procedure is used to monitor road accident fatalities with number of accident crashes as covariate. A regression control chart which is the combination of the conventional control chart and regression analysis was first proposed by Mandel (1969) who used regression control chart to monitor the man-hours required to process pieces of mail at 74 post offices of the United States. Shu et al. (2004) provided an overall review of regression control chart techniques. The formulation of the method was presented, and its computation was illustrated.

The essential idea of the regression control chart is to chart the process output after it has been adjusted for the effect of the covariate and has proved useful in a wide variety of application dues to its simplicity and efficiency (Shu et al., 2004). A least square regression program must process the data prior to constructing of the control chart with the availability of modern computation power the execution time of such chart is often on the order of few seconds. Moreover, it often has better sensitivity and appropriate for a number of purposes than the standard control chart which directly monitor the process output without taking the influence of the covariate into consideration (Mandel, 1969).

2. The Regression Control Chart

Let X be the external covariate and Y the output characteristic of interest. The linear regression model relating X to Y take the form;

$$Y_i = \beta_0 + \beta_1 X_i \qquad (1)$$

It is assumed that Y_i are independent and identically distributed (i.i.d.) normal variables with mean zero and variance σ^2. In practice, the values of σ^2, β_0, and β_1 are unknown and must be estimated using historical data collected from the process. We assume that samples of n paired observations (X_i, Y_i) are available in a set of historical data. The regression line is estimated, scatter plot is drawn to investigate if the paired values of the two variables move in the same direction and coefficient of correlation, r is calculated to compare the values of the dependent variable for different values of the independent variable. After which the regression control chart is drawn using the estimated value of the response variable to place parallel control limits at $\pm 3\sigma_e$ from the regression line. The area between the limits is considered the allowable variation from the independent variable. The σ_e, standard deviation of the regression control chart is the standard error estimate of the regression line.

Mandel (1969) stated the following elements which distinguished the regression control chart has from the conventional control chart.
1. It is a model that controls a varying average rather than a constant average. The central line is the regression line.
2. The control limits are parallel to the regression line rather than to the horizontal axis.
3. The computation for the construction of the regression control chart is time consuming compared to the conventional control chart, but with the help of modern high speed computers, the problem of computation is solved. 4. The regression control chart is appropriate for a number of applications, which the conventional control chart does not readily, applies. It provides the basis of measuring the gains or loss in the response variable, for predicting and forecasting the response variable and scheduling the covariate resources.

3. Data Analysis

3.1 Data Source

The Federal Road Safety Commission (FRSC) is the government agency in Nigeria saddle with the responsibilities for road safety administration. For this study, the monthly record of road crashes and fatalities for the period, 2013 to 2018 used was obtained from FRSC Ogun State Sector Command. It covers 72 months. The total number of crashes recorded for the period was 2, 539 while the total number of fatalities was 1,593 which represents 62.74% of total number of crashes. Regression control chart was used to monitor road accident fatalities taken number of crashes as covariate. Tables 1 and 2 below present the summary of monthly record of road crashes and fatalities for the period, 2013 to 2018.

Table 1

Summary of Number of Road Traffic Crashes

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16.00	26.75	33.00	35.26	40.00	78.00

Table 2

Summary of Number of Road Accident Fatalities

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8.00	15.00	21.50	22.12	27.00	44.00

3.2 Designing of Regression Control Chart for Accident Fatalities

The relationship between road accident fatalities (y) and number of crashes (x) is examined using a scatter plot (figure 1) to show the relationship between the dependent variable *y* against independent variable *x*. There exists positive relationship between the two variables; the coefficient of correlation (*r*) is 0.46. This is a justification that regression control chart can be used in monitoring accident fatalities taken into consideration the effect of auxiliary variable; number of crashes.

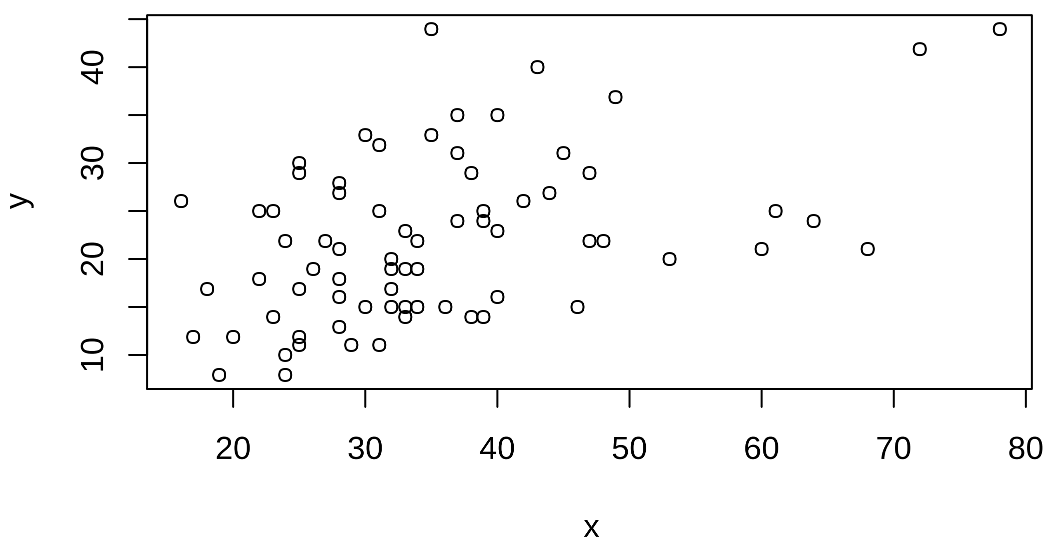

Figure 1: Scatter plot of number of accident fatalities (y) and number of crashes (x)

The regression line (central line) was developed, two and three sigma control limits were set parallel to the regression line as follows;

Equation of the line: $fatalities$ =11.23465 + 0.30882 × Crashes (2)

Standard error of estimate: $\sigma_e = 7.5983$

Three sigma limits: $Control\ limits = (11.23465 + 0.30882 \times \text{Crashes}) \pm 3 \times 7.5983$ (3)

Two sigma limits: $Control\ limits = (11.23465 + 0.30882 \times \text{Crashes}) \pm 2 \times 7.5983$ (4)

Multiple R-squared for the regression analysis is 0.212 which shows that number of crashes account for 21.2% variation in number of accident fatalities. The F-statistic p-value is 4.714e-05 which shows that, there is significant relationship between number of crashes and number of accident fatalities.

The control limits are parallel to the regression line rather than to the horizontal axis. Three solid lines are drawn on the scatter plot, the central line (line of best fit), upper control limit and lower control limit. And two broken lines were also drawn, the upper warning limit and lower warning limit. The lines are expected to slant upward or downward as shown in figure 2.

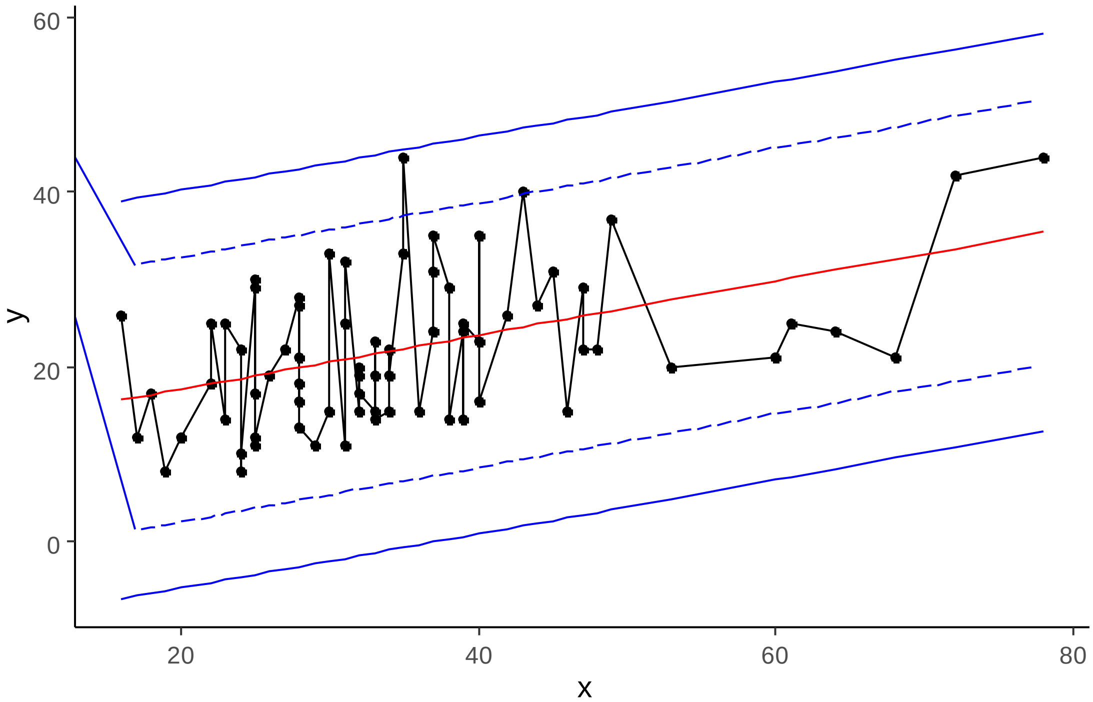

Figure 2: Regression control chart for monitoring accident fatalities taken crashes as covariate

The representation of the conventional control chart for monitoring number of accident fatalities at two and three sigma limits is shown in Figure 3. The conventional control chart is univariate, while the regression control chart is bivariate.

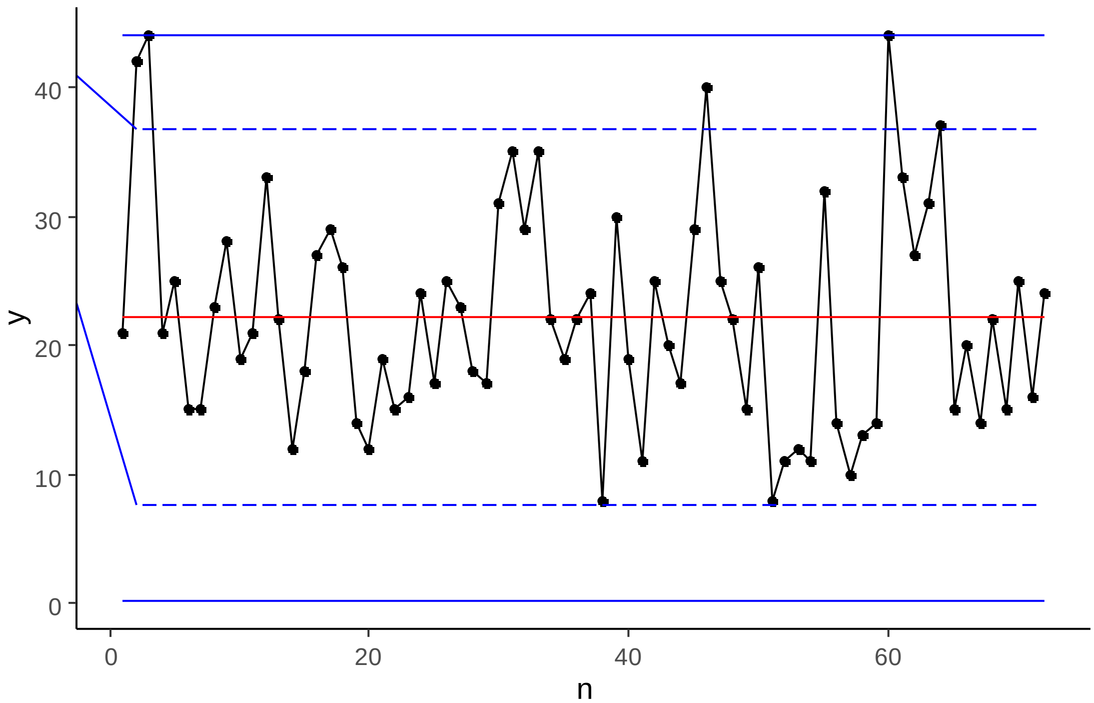

Figure 3: Conventional control chart for monitoring accident fatalities

4. Discussion of Results

The relationship between number of accident fatalities and number of accident crashes was first examined. The scatter plot revealed that the number of accident crashes has a positive effect on the number of accident fatalities. The correlation coefficient which measures the degree of association between the two variables is 0.46. The R-squared revealed that number of accident crashes account for 21.2% variation in number of accident fatalities. Hence, the regression control chart can be used to investigate the variation in number of accident fatalities.

To construct the chart, line of regression equation was first developed and the standard error of estimate was calculated. The control limits are drawn at two sigma ($\pm 2\sigma_e$) (dash line) and three sigma ($\pm 3\sigma_e$) (solid line) parallel lines as shown in figure 2. At three sigma level all values are within UCL and LCL indicating the presence of common cause variation. At two sigma level, there is only one point outside the limits. A comparison is made with conventional control chart in figure 3, at three sigma level all values are within the control limits but two of the values are on the upper control limit indicating serious warning of presence of assignable cause of variation. At two sigma level, there are five points outside the limits and two points on the lower warning limit. This is an indication that conventional control chart can raise false alarm where there is none because it only considers the variation from one characteristic and disregards the effects of variation from the other auxiliary variables.

When investigating the relationship between two variable characteristics, there may be situations that the control chart of one variable characteristic is in control; however, regression control chart is not in control due to correlation and the influence of other variable characteristics. Also, there may be situations that control chart of one variable characteristic may show values outside the control limits indicating special cause, however, regression control chart may indicate control because of justified variation contributed from the other variable characteristic (Hayati, 2017). Generally, special causes are investigated by many authors without considering any other contributing factors (Hayati, 2017). Monitoring the number of accident fatalities without considering some other factors makes such investigation inadequate. The variation from the number of road crashes is a significant source of variation affecting number of accident fatalities; therefore, regression control chart is the appropriate method where the number of road crashes is included in the analysis.

5. Conclusion

A regression control chart is an effective way to account for acceptable variation for the other source. Although regression control chart is an extension of conventional control chart, it offers many advantages over conventional control chart under certain conditions. In this paper, regression control chart was developed to monitor number of accident fatalities considering the effects of variation from number of road accident crashes. The chart was compared with Conventional control chart which monitors only the number of accident fatalities and the variation from other variables ignored. At three sigma level, the regression control chart showed that all values are within the control limits, while the Conventional control chart has two points on the upper control limit. At two sigma level, the regression control chart has one point outside the control limits, while the Conventional control chart has five points outside the control limits and two points on the lower control limit. Hence, the regression control chart justified the effects of variation from the covariate in monitoring the number of accident fatalities. It is recommended that regression control chart, a robust method for process analysis and quality improvement, be used extensively in service, manufacturing, research and healthcare processes.

References

Adekeye, K. S. and Aluko, O. S. (2012). Design of Cusum Scheme for Monitoring Road Accident Fatalities. *Open Journal of Statistics*, 2, 213-218. doi:10.4236/ojs.2012.22026.

Braimah, O.J., Osanaiye, P.A., Eshimokhai, S.A. and Iyoha, A. (2014). Application of Count Data Cumulative Sum (Cusum) Control Chart in Monitoring Road Traffic Crashes. *Sop Transactions on Statistics and Analysis* 1(1), 38-47.

Faweya Olanrewaju, Ogunwale Olukunle Daniel, Adeniran Adefemi Tajudeen. Comparative Analysis on EWMA and Poisson Cusum Chart in the Assessment of Road Traffic Crashes (RTC) in Osun State Nigeria. *American Journal of Theoretical and Applied Statistics.* Vol. 6, No. 2, 2017, pp. 95-99. doi: 10.11648/j.ajtas.20170602.14.

Hayati, F. (2017). Regression control chart for two related variables. *Int. J. Modelling in Operations Management, 6(4), 262-279.* DOI: 10.1504/IJMOM.2017.10011283

Ihueze, C. C. and Onwurah, U. O. (2018). Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. *Accident Analysis and Prevention*, 1(12), 21–29.

Mandel, B. (1969). The regression control chart, *Journal of Quality Technology 1*, 1–9.

Shu, L., Tsung, F. and Tsui, K. (2004). Run-length performance of regression control charts with estimated parameters, *Journal of Quality Technology, 36*, 280–292.

World Health Organization (2020). "Fact Sheets - Road traffic injuries".