

---

# Validating regression models to assess factors for motorcycle accidents in Tanzania

Vicent Paul Nyakyi\*, Dmitry Kuznetsov, Yaw Nkansah-Gyekye

School of Computational and Communication Science and Engineering, Nelson Mandela African Institution of Science and Technology, Arusha, Tanzania

## Email address:

nyakyiv@nm-aist.ac.tz (V. P. Nyakyi), dmitry.kuznetsov@nm-aist.ac.tz (D. Kuznetsov),

yaw.nkansah-gyekye@nm-aist.ac.tz (Y. Nkansah-Gyekye)

## To cite this article:

Vicent Paul Nyakyi, Dmitry Kuznetsov, Yaw Nkansah-Gyekye. Validating Regression Models to Assess Factors for Motorcycle Accidents in Tanzania. *Science Journal of Applied Mathematics and Statistics*. Vol. 2, No. 5, 2014, pp. 97-101. doi: 10.11648/j.sjams.20140205.12

---

**Abstract:** There are several ways in which regression model can be validated. One of the ways is a collection of new data to check model predictions (Snee, 1977). In Nyakyi *et al.* (2014) a multiple regression model was formulated, analyzed, and discussed to assess motorcycle accidents in Tanzania. The model included several factors which are wrong overtaking, legal status of not owning license, rough road, high speed, mechanical defect, personal status, experience of a driver, and tarmac road. All these factors were considered to be the causes of motorcycle accidents in Tanzania, specifically in Arusha and Kilimanjaro regions. This paper presents the analysis and discussion of the accidents problem using new data collected through questionnaire; the study's results are then compared with the results, predicted by the multiple regression models, for validation. Questionnaires were designed to access the extent to which identified factors cause motorcycle accidents, and the data is obtained from motorcycle stake holders who in this study are considered to have reliable information about motorcycle accidents in Tanzania. The results of the validation indicate that there is a good correlation between the data obtained from the questionnaire and the data produced by the regression model.

**Keywords:** Questionnaires, Motorcycle Accidents, Fitting, Factors and Validation

---

## 1. Introduction

The proportion of citizens involved in road accidents each year now exceeds the nation's annual population growth. The numbers of causative factors are continuously increasing every day as the need of transport increases. Studies show that 3.4 % of Tanzania's GDP is lost due to traffic accidents each year and the problem is increasing as since 2007, road deaths and injuries have increased by an average of 10% (Kamal, 2010).

According to Mwakapasa (2011), cited in (Nkwame, 2010), motorcycles form a common means of transport in developing countries especially in Africa including Tanzania. Nkwame (2010) further emphasized that motorcyclists form a notable proportion of people who are affected by road traffic accidents.

Road traffic injuries (RTIs) are responsible for a significant proportion of overall injury morbidity and mortality; 90% of mortalities are seen in low and middle income countries (Peden *et al.*, 2005). Motorcycle users are vulnerable on the road and represent an important group to

target for reducing road traffic injuries. Riding motorized two-wheeled vehicles carries a higher risk of being involved in a fatal traffic accident than from using any other common mode of transport. It has been estimated that, per 100 million person travelling hours, 440 motorized two-wheeled vehicle rider fatalities occur, compared to 75 and 25 fatalities for bicyclists and car drivers, respectively (Koornstra *et al.*, 2003).

Many studies have been done about the same and pointed out several factors that lead to motorcycle accidents. The study by Nyakyi *et al.* (2014) formulated a multiple regression model that analyzed the factors that lead to motorcycle accidents in Kilimanjaro and Arusha regions. The study analyzed more than six factors which are mechanical defects, legal status of not owning license, legal status of owning license, rough road, tarmac road, driving experience, wrong overtaking, high speed, and a personal status factor (alcohol or drugs intake). The results from the study stated that the number of motorcycle accidents in both Arusha and Kilimanjaro regions have strong relationship with experience of drivers, tarmac road, and personal status.

This could be attributed to drivers being more confident and driving at high speed on tarmac roads. However, the relationship between the number of motorcycle accidents, legal status of drivers who do not have their own license and rough road did not show significant effect. The study also observed that there is a strong relationship between the number of motorcycle accidents, high speed, and wrong overtaking in Arusha region.

By its nature, a model is more abstract than what it really represents; whatever modeling paradigm or solution technique is being used, the performance measures extracted from a model will only have some bearing on the real system represented if the model is a good representation of what is really happening (Rothenberg, 1989). What constitutes a good model is subjective, but from a modeling point of view the criteria for judging the goodness of models will be based on how accurate measures extracted from the model correspond to the measures which would be obtained from the represented situation. This paper presents the real data collected using a questionnaire on the factors that mostly lead to motorcycle accidents in Kilimanjaro and Arusha regions. The data are then compared to the data that were produced by the model for its validation.

## 2. Models

This paper sought to validate the model developed by Nyakyi *et al.* (2014) in which a regression model to assess the causes of motorcycle accidents in Arusha and Kilimanjaro regions in Tanzania were formulated and analyzed. The model for Kilimanjaro region was represented by the equation

$$Y_1 = 181.02 + 3.513x_2 - 3.838x_3 + 15.098x_6 - 0.213x_7$$

and the model for Arusha region was represented by the equation

$$Y_2 = 134.263 + 1.913x_1 + 0.628x_5 + 8.421x_6 - 1.530x_7$$

where the variables are described in Table 1.

Table 1. Variable and their description

Variable	Description
X <sub>1</sub>	Wrong overtaking
X <sub>2</sub>	Legal status not own license
X <sub>3</sub>	Rough road
X <sub>4</sub>	Legal status own license
X <sub>5</sub>	High speed
X <sub>6</sub>	Personal status
X <sub>7</sub>	Experience of drivers
X <sub>8</sub>	Tarmac road
X <sub>9</sub>	Mechanical defect

## 3. Methodology

The aim of the questionnaire was to extract/get information on what people think on the causes of motorcycle accidents in Tanzania. The study samples were

mainly the motorcycle drivers and very few local citizens who were considered to have enough information on the reasons for motorcycle accident in Tanzania. A set of questions was then prepared, distributed to and collected from the identified samples and data was analyzed as shown in the following section.

## 4. Data Collected from Questionnaires

The data from the questionnaires were presented as in Figures 1, 2, 3, 4, 5 and 6. The analysis of each questionnaire was done in comparison with the results from the formulated models.

### 4.1. How Many Motorcycle Accidents Have You Ever Had in the Past Three Years

From the questionnaire the statistics of number of drivers and number of accidents each driver had was analyzed as summarized in Figure 1.

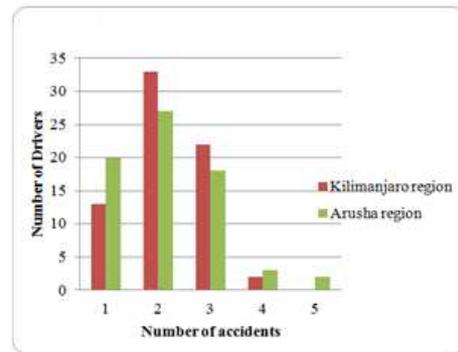


Figure 1. Histogram for Motorcycle Accidents per driver

From Figure 1, it can be observed that the highest bar graph in Kilimanjaro region has more than 30 drivers who have been involved in at least two accidents while in Arusha less than 30 drivers have been involved in at least two accidents. The observed high number of accidents involved in Kilimanjaro was also predicted by the models formulated in this study. It can also be noted that every interviewed motorcycle driver has been involved at least once in accidents.

### 4.2. For How Long Have You been Driving a Motorcycle

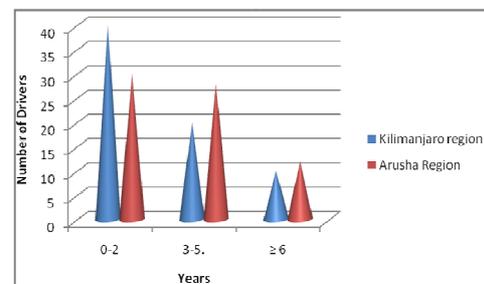


Figure 2. Experience of Drivers in Years

This question was intended to identify the experience of drivers so as to justify the formulated models as it was

observed from the models that lack of experience could be among the factors which contribute to motorcycle accidents. The collected information was presented in Figure 2.

According to the data, Kilimanjaro region has a large number of inexperienced drivers of about 40 out of 70 which is equivalent to 57% compared to Arusha region with about 30 drivers out of 70 which is equivalent to 43%. The high frequency of drivers without an experience of driving in Kilimanjaro region can also lead to high incidences of motorcycle accidents compared to Arusha region where it was observed that about 28 drivers out of 70 have an experience within the range of 3-5 years of driving as shown in Figure 2.

**4.3. What Do You Think are Causes of Motorcycle Accidents**

This question was intended to identify the driver’s understanding on this problem of motorcycle accidents and also to let him/her identify the major and popular causes of motorcycle accident. The collected information was presented as shown in Figure 3 and Figure 4.

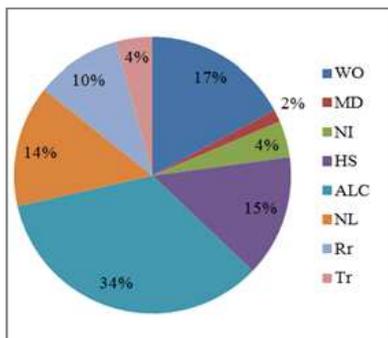


Figure 3. Pie chart for Kilimanjaro Region

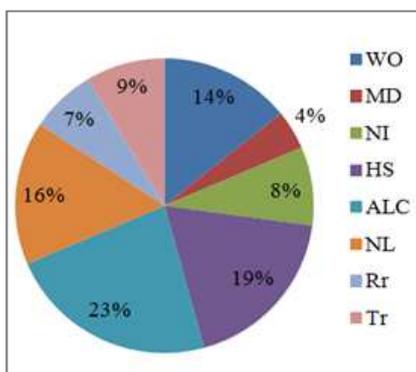


Figure 4. Pie chart for Arusha Region

Key: WO = Wrong Overtaking,  
 MD = Mechanical Defects,  
 NI = Night,  
 HS= High Speed  
 ALC = Alcohol and drug use,  
 NL = No license,  
 Rr= Rough road and  
 Tr = Tarmac road.

From Figure 3 and Figure 4 it can be observed that among all the listed causes of motorcycle accidents, alcohol has

shown a large percentage of about 34% in Kilimanjaro region and 23% in Arusha region. This trend was also predicted by the formulated models for Kilimanjaro and Arusha regions. This could be attributed to the culture in which people who live in these two regions (especially the Chagga) start taking local beer at a very tender age and thus grow with this spirit which become even worse at the age between 18 and 40 which is the same age interval that populates motorcycle drivers. The other reason is the weather condition (cold) in these regions. It was also observed that, the number of drivers who do not have license in the two regions is high as shown in Figure 3 and Figure 4. From Figure 3 and Figure 4, the percentage of drivers with no license was 14% and 16% respectively in Kilimanjaro and Arusha regions. This tendency has been observed in Figure 2 where the number of drivers with experience between 0-2 years was large compared to others in the two regions. However, from the formulated models when holding other variables constant the number of drivers not having their own license in Kilimanjaro was observed to increase by a factor of about 3.513. From this point of view it is obvious that the drivers with no license can contribute to the motorcycle accidents due to the reason that they have less experience in driving motorcycles.

Figure 3 and Figure 4 shows that wrong overtaking has a large contribution to accidents in the two regions with percentages of 17 and 14 in Kilimanjaro and Arusha regions respectively. However, from the formulated models it was also noted that in Arusha the number of motorcycle accidents will increase by 1.913 which is a high increase. From these observations one may conclude that wrong overtaking could be one among other factors which contribute to motorcycle accidents.

Furthermore, it was noted that mechanical defects cause motorcycle accident by 2% and 4% in Kilimanjaro and Arusha regions respectively. However, these results agreed with the formulated models where by the mechanical defects have been fused due to its low effect of contribution in motorcycle accidents. This is due to the fact that most of the motorcycles are new and have been increasing in recent years as it has been reported by Mpinga (2012). It was also indicated in Figure 3 and Figure 4 that during the night the number of motorcycle accidents was low in Kilimanjaro region with 4% while in Arusha region it was noted to be 8%. This tendency of increase in Arusha may be attributed by the fact that, from the feasibility study Arusha region has a large population compared to Kilimanjaro region; also more people live away from the town center and therefore leave the town before sunset and few remain in town during night time. The night motorcycle accidents were also high in Arusha region as it is the leading tourism city and with Tanzanite mining activities.

**4.4. What Do You Think are the Common Causes of Motorcycle Accidents in Your Region**

This question was intended to identify the opinions of motorcycle accidents in their regions as in question 3 the

given factors were guided by a researcher. The driver's opinions are presented in Figure 5.

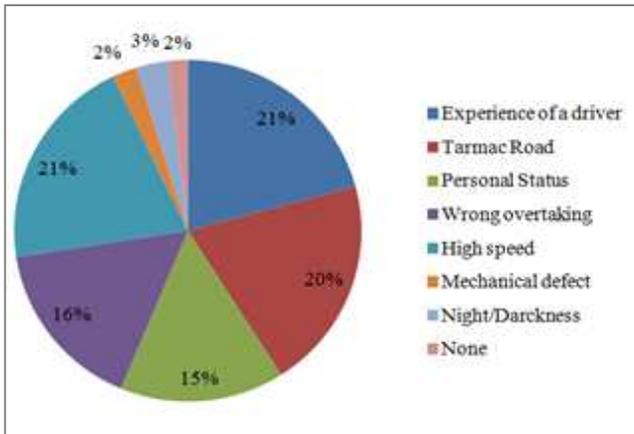


Figure 5. Causes of motorcycle accident.

From the interviewed motorcycle riders as in Figure 5 it was noted that very few drivers of about 5% of the total drivers who were interviewed said that the common causes of motorcycle accidents are mechanical defects and darkness. About 16% of them were blaming car drivers by saying that sometimes car drivers are the causers of the motorcycle accidents by interrupting into other motor driver's sites especially when overtaking. About 21% said that accidents are mostly caused by lack of experience of the motorcycle drivers while about 21% said accidents are caused by high speed. It was also noted that other drivers said 20% and 15% of motorcycle accidents are caused by tarmac road and personal status (taking in alcohol) respectively. About 2% of motorcycle drivers did not mention any reason which can contribute to motorcycle accidents.

From the interviewed motorcycle riders as in Figure 5 it was noted that very few drivers, about 5% of the total drivers who were interviewed, said that the common cause of motorcycle accidents are mechanical defects and darkness. About 16% of them were blaming car drivers by saying that sometimes car drivers are the main causers of the motorcycle accidents by interrupting into other motor driver's sites especially when overtaking. About 21% said that accidents are mostly caused by lack of experience of the motorcycle drivers while about 21% said accidents are caused by high speed. Other drivers said 20% and 15% of motorcycle accidents are caused by tarmac road and personal status (taking in alcohol) respectively. About 2% of motorcycle drivers did not mention any reason that contributes to motorcycle accidents.

**4.5. Opinions of Motorcycle Driver's on the Activity to Reduce the Number of Motorcycle Accidents**

From the questionnaire motorcycle drivers gave out their opinions on what could be done to reduce the number of motorcycle accidents. The results have been presented in Figure 6.

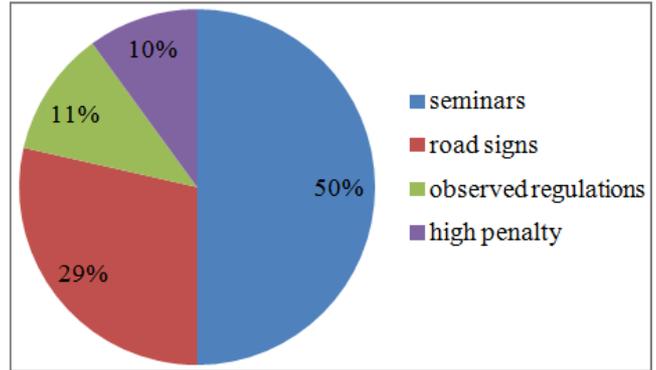


Figure 6. Opinions from the motorcycle drivers to reduce accidents

As from Figure 6 it was observed that about 50% of opinions from the motorcycle drivers mentioned that seminars should be provided to motorcycle drivers to reduce the number of accidents. It was also noted that about 29% requested for posters with road signs to be increased in order to reduce the problem of accidents. About 11% mentioned the regulations to be observed by motorcycle drivers while about 10% said that high penalty should be offered for those who abide the road regulations.

**5. Discussion and Results**

**5.1. Discussions**

This study is different from the other previous studies presented by different scholars in the literature section because it has included the experience and personal status (alcohol and drug) of drivers and the use of questionnaire approach for validation.

Using the data obtained from the questionnaire it was found that in both Kilimanjaro and Arusha regions the factors that dominate as the causes of motorcycle accident are experience of the driver, tarmac roads, personal status, wrong overtaking, and high speed. The factor of personal status which entails drug taking including alcohol shows a very strong relation with motorcycle accidents in both Kilimanjaro and Arusha regions; this may be due to the existing culture in these two regions where people start drinking alcohol at a very tender age.

**5.2. Results**

The data obtained by the questionnaire makes a good correlation when related with the data produced by the regression model developed above as shown using the graph fitting in Figure 7 below.

The best line fit in the Figure above makes the model developed above reliable in assessing the causes of motorcycle accident in Kilimanjaro and Arusha regions as it produces similar results as that produced by the regression model.

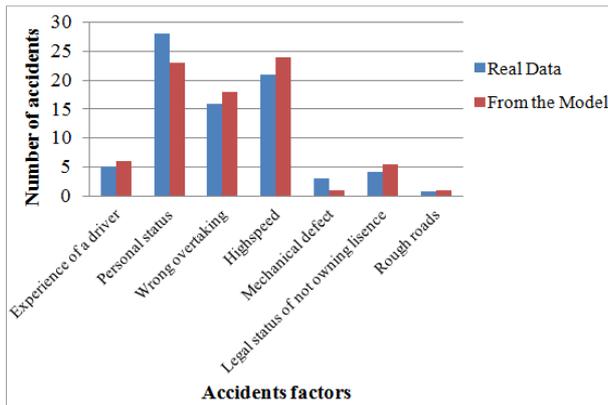


Figure 7. Goodness of fitting of regression model

## 6. Conclusion

Motorcycle accidents seem to be a threat to the society especially the youth. The causes of these accidents are tagged as factors that are in one way very difficult to monitor, but in another way rooted in the daily cultural practice. In Arusha and Kilimanjaro regions motorcycle accidents are the highest as compared to accidents from other means of transport. It was identified from motorcycle's stake holders within the study is that what is happening is the reality and through this information we might have the reliable information to use in solving motorcycle accident problem in Kilimanjaro, Arusha and Tanzania in general.

## References

- [1] Kamal, S. (2010). Serious Injury and Injurious Burden in Urban Bangladesh. *Journal of Applied Science and Technology*. 7:71-79
- [2] Koornstra, M., Broughton, J., Esberger, R., Glansdorp, C., Köppel, W., Taylor, F., Cauzard, J., Evans, A., Hantula, L., Piers, M. and Vanlaar, W. (2003). *Transport safety Performance in the EU: a statistical overview*. European Transport Safety Council, Brussels, Belgium, p. 32.
- [3] Mpinga, M. (2012). Tanzania Traffic Police Report. *The Guardian News Paper* December 25, 2012
- [4] Mwakapasa, E. G. (2011). *Attitudes towards and practice of Helmet use Among Commercial*
- [5] Nkwame, M. (2010, June 19). Motorcycle accidents claim 181 lives in four months. Tanzania: *Daily News*.
- [6] Nyakyi, V., Kuznetsov, D. and Nkansah-Gyekye, Y. (2014). Mathematical Model to Assess Motorcycle Accidents in Tanzania. *Journal of Mathematical Theory and Modeling*. Vol.4, No.9, 112-119
- [7] Peden, M., McGee, K. and Sharma, G. (2005). *The injury Chart Book*. A graphical overview of the global burden of injuries. Geneva: World Health Organization, 2000:5.
- [8] Rothenberg, J. (1989). The nature of modeling Rand.
- [9] Snee, R. D. (1977). *Validation of regression models: methods and examples*. *Technometrics* 19:415-428.