

**A STATISTICAL ANALYSIS OF INFLUENTIAL FACTORS ON CRIME IN
BADULLA DISTRICT**

K.M.S.K. Gunathilake* and N. Satkunanathan

Department of Mathematics and Statistics, University of Jaffna, Jaffna, Sri Lanka

**shakilakalhari4596@gmail.com*

Crime is one of the critical social problems that Sri Lanka has been experiencing in recent years. It seriously affects communal harmony and socioeconomic status. Therefore, studying and identifying factors influencing crime is vital in policy-making and development. This study focuses on identifying factors influencing the crime rate in the Badulla District. For this study, the number of crimes was treated as a count response variable, while occupation, crime type, area and age group were treated as explanatory variables. The required data covering the period from January 2020 to December 2020 were collected from the Crime Unit of Badulla Police Station. Chi-squared test of independence was carried out to determine the significant relationship between crime type with other explanatory variables. Further, a log-linear model was applied as a linear modelling approach to identify the most significant variables, as it is appropriate for modelling counts in the contingency table. Model parameters were estimated by using the maximum likelihood method. The likelihood ratio test was also used to check the model's overall fit. Based on the results of the chi-squared test, it was noted that there is a significant relationship between crime type and occupation ($p=0.0008$), area ($p=0.0223$) and age group ($p=0.0182$). Therefore, based on the dispersion parameter and likelihood ratio test results, log-linear model fits well for the number of crimes. Moreover, the results of a log-linear model revealed that area and age group are significantly related to the number of crimes. While controlling other explanatory variables, it was observed that the expected number of crimes in rural areas was about twice the number in urban areas. Furthermore, the expected number of crimes committed by 25-54 in rural and urban areas was higher than in other age groups while controlling other explanatory variables.

Key words: Chi-Squared test, Contingency table, Likelihood ratio test, Log-Linear model