Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
**Vol. 18, 2023, page 27 – 32**

# Statistical Analysis of Violent Crime in Malaysia

**Nur Nabilah Fatini Mohd Zaki, Muhammad Fauzee Hamdan***
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
*Corresponding author: mfauzee@utm.my

**Abstract**
The increasing importance of security in today's society has led to a focus on reducing criminal incidents. Crime is a significant concern, impacting institutions and organizations. This study analyzedcrime cases in Malaysia from 2006 to 2017 by using a clustering approach. Multivariate observations were divided into groups to identify patterns in the data set. The 13 states of Malaysia, including KualaLumpur and Borneo states Sabah and Sarawak, registered 117,388 violent crime cases. The resultingset of crimes was used to create a hierarchical tree using the dendrogram agglomerative hierarchical clustering method, showing trends in different types of crimes. The study also aimed to determine the potential of clustering for detecting and fixing the current problem.

**Keywords:** Mann- Kendall test; dendrogram; clustering approach; trend analysis

## 1. Introduction

Crime is a social injustice that involves illegal acts that violate state law and are disapproved of by society. Crimes include murder, robbing, injuring others, operating a vehicle while intoxicated, abusing children, and neglecting taxes. Criminal law prohibits and punishes such behavior, with self-defense actions never illegal. Data analytics and technology can help identify patterns in criminal activity. The legal definition of crime is complex and debatable, with crime varying depending on time and place. The need to control crimes is growing, putting pressure on the investigation department. A system to evaluate crime and use data analytics to make their job easier. This study investigates crime trends in Malaysia and clusters states with similarity based on different types of crime cases.

Crime analysis is a systematic investigation of crime and disorder issues, focusing on sociodemographic, geographic, and temporal aspects. It aims to assist law enforcement agencies in criminal detection, reduction, prevention, and assessment. The increasing intensity and flexibility of crimes cause significant costs to society, monetary loss, social loss, and danger to communityharmony. The computing era can help in crime prediction and implementing proper protections to limit harm to property and loss of life. Crime analysis is essential for law enforcement agencies to identify and capture criminals, reduce and prevent crime, and formalize guidelines for crime prevention.

Clustering algorithms are employed to analyse crime data, identifying similarities between crimes in the remaining database and those with an increasing size trend. This study analysed the number of cases in 14 states in Malaysia from 2006 to 2017, proposing the use of clustering and trend analysis to determine crime analysis in Malaysia. The results can contribute to a better conclusion of the study's objective and provide valuable lessons for decision-making processes. Crime data analytics can provide information about crime statistics and reports for all states in Malaysia, aidinglaw enforcement agencies in proactive decision-making to provide adequate security and safety for citizens.

## 2. Literature Review

Trend analysis is a technique used to create future predictions based on historical data. It involves comparing data over a specific time frame and detecting uptrends, downtrends, and standstills. Numerical data is used to compute trend analysis and can show if crime is going up or down in the areas targeted. Numerous elements, such as socioeconomic circumstances, social inequality, and unemployment rates, are related to crime patterns (Hale, 1008; Talha, 2008; Fougère, et al.,2009; Newburn, 2016).

In addition to these definitional issues, it has been shown that different countries' reporting of criminal cases differs significantly (Anna Alvazzi del Frate,2003). Trend analysis management tools are widely available and one of the simplest is to plot the data points and determine whether a trend exists graph. This would ensure law enforcement function, which includes hard investigation for identifying and analysing criminals and their patterns of crime and interruption (S. Sathyadevan, M. Devan, and S. Surya Gangadharan,2014).

Crime is an illegal, punishable act committed against any person or organization with the intent to cause harm. It includes robberies, looting, sexual harassment, rape, kidnapping, and homicides, which are increasing in frequency from rural to metropolitan areas. Crime analysis is a key component of contemporary police tactics that focus on problem-solving proactively and making decisions based on intelligence (Hinkle et al., 2020).

It uses formulas, databases, and computer systems to pinpoint high-risk states and criminals, identify criminal patterns, and identify criminal trends (Fyfe et al., 2017; Taylor et al., 2007). Depending on types of surroundings, the pattern of the crime problem can be identified in many ways,such as the environment's trait of being densely populated, economy, social and political considerations, and demographic factors (Appiahene-Gyamfi et al.,2002; Perreault et al.,2008; Savoie et al.,2006). As pronounced by Becker (1968), a major and important impact on increasing rates of crime is an increase in inequality in income (Madden & Chiu 1998; Kelly 2000; Maria & Meloni 2000; Fajnzylber et al. 2002; İmrohoroğlu et al. 2006).

Cluster analysis is the division of data or objects into groups based on distance and similarity. According to Sousa and Gama (2014), hierarchical clustering algorithms are effective for identifying natural clusters in data analysis, often without prior knowledge of the data structure. The two forms of hierarchical clustering are agglomerative and divisive, respectively (S. Everitt, 2011). A dendrogram is a diagram used to show the results of a hierarchical cluster analysis, showing the distance between clusters and which clusters have been joined at each stage.

Agglomerative hierarchical clustering methods start with one item in each cluster and iteratively merges the two clusters that are the most similar to one another in terms of the similarity metric until all the objects are included in a single cluster (Turi, 2001). In complete linkage, the maximum distance between any two data points one from the first cluster and one from the second is the distance between two clusters.

## 3. Methodology

### 3.1 Trend Analysis

A crime trend is a significant shift in a few crime categories' characteristics across a certain time period and area. To find and describe trends, analysts employ trendlines and channels. The Mann-Kendall test is used to check for a linear monotonic trend in a given time series of data. The null hypothesis, $H_0$ states that there is no monotonic trend, and alternative hypotheses, $H_a$ is an upward monotonic trend or a downward monotonic trend.

### 3.2 Mann Kendall Test

The first step in using Mann-Kendall test is by listing the data in the order which they were collected over time. Next, determine the sign

$$\frac{n(n-1)}{2}$$

possible differences $x_j - x_k$, where $j > k$. The Mann-Kendall test for a time series $x_1, x_2, \ldots, x_n$ of length n is to compute the indicator function $sgn(x_i - x_j)$ such as:

$$sgn(x_i - x_j) = \begin{cases} 1, & x_i - x_j \\ 0, & x_i - x_j \\ -1, & x_i - x_j \end{cases}$$

Then, compute the mean and variance of the above quantity. Mean $E[s]$ is given by

$$E[s] = \sum_{i=1}^{n-1} \sum_{j=j+1}^{n} sgn(x_i - x_j)$$

While Variance $VAR(S)$ is given by

$$VAR(S) = \frac{1}{18}(n(n-1)(2n+5) - \sum_{k=1}^{p} q_k(q_k-1)(2q_k+5))$$

Afterward, compute the MK test statistic $Z_{MK,}$ as follows

$$Z_{MK,} = \frac{S-1}{\sqrt{VAR(S)}} \text{ if } S > 0$$

$$= 0 \; if \; S = 0$$

$$\frac{S+1}{\sqrt{VAR(S)}} if \; S < 0$$

Test the null hypothesis with

$H_0$: No trend
$H_1$: Upward monotonic trend

At the Type I error $\alpha$, where 0 < $\alpha$ < 0.05. Then $H0$ is rejected and $H0$ is accepted if $ZMK \geq Z1-\alpha$. Next to test $H_0$ above

$H_1$: Downward monotonic trend

At the Type I error rate $\alpha$, $H0$ is rejected and $H\alpha$ is accepted if $ZMK \leq Z1-\alpha$. To test $H0$ above

$H_1$: Upward or downward monotonic trend

At the Type I error rate $\alpha$, $H0$ is rejected and $H\alpha$ is accepted if $|ZMK| \geq Z_{1-\alpha/2}$, where the vertical bars denote absolute value

### 3.3 Hierarchical Analysis

The first step in performing Agglomerative clustering is assigned each data point as a single cluster. Next, determine the distance measurement and calculate the distance matrix. Choose the suitable linkage to merge the clusters. Update the distance matrix and repeat the process until every data point becomes one cluster, with a dendrogram, then we choose our cut-off value to acquire the number of the cluster.

### 3.4 Distance Measure

Distance measurements are used to compare two things in a cluster analysis, with bigger distance measures indicating less similarity and smaller distance measures indicating dissimilarity. Euclidean distance is computed as

$$D_{ij} = \sqrt{\sum_{i=1}^{n}(x_{ki} - x_{kj})^2}$$

### 3.5 Inter- cluster Distance Measure

Farthest neighbour clustering is a technique used to join components in one cluster into larger clusters by placing each component in its own cluster at the start. The formula of complete linkage clustering is as follows:

$$d_{AB} = max^{i\epsilon A}_{i\epsilon B}(d_{ij})$$

## Results and discussion

### Trend Analysis

The incidence of violence crime cases in Malaysia over the last 12 years ago, reported 117,388 number of cases from years 2006 to 2017. By using Mann Kendall test, for murder case, the result shows $Z_{MK}$= -2.6743 which is less than significance value $Z_a = 1.64$. Therefore $H_0$ is being rejected means the result is said to be statistically significant. It is also found that alternative hypothesis, *S* = -40 in *n*=12. A negative value of S indicates that there is a decreasing trend shown. For rape cases $z$ = -2.6119, $n$ = 12, $p$ -value = 0.009004, S=-39, $var(S)$ = 211.6667, $r$ = - 0.5954. The result above shows $Z_{MK}$= -2.6119 which is less than significance value $Z_a = 1.64$. From the Mann Kendall analysis, it is found that alternative hypothesis, *S* = -39. A negative value of S indicates that there is a decreasing trend shown.

For armed robbery cases, the result above shows $Z_{MK}$ = -3.0858 which is less than significance value $Z_a = 1.64$. Therefore $H_0$ is being rejected means the result is said to be statistically significant. From Mann Kendall analysis it shows *n*=12, and found that alternative hypothesis, *S* = -46.A negative value of *S* indicates that there is a decreasing trend shown. For causing hurt, the result shows the result above shows $Z_{MK}$ = -2.9486 which is less than significance value $Z_a = 1.64$ . Therefore $H_0$ is being rejected means the result is said to be statistically significant. A Mann Kendall analysis shows a negative value of S= -44, which indicates that there is a decreasing trend. This concludes that all 4 types of crimes indicate the decreasing number of crime cases throughout the years.

### Clustering Analysis

The dendrogram created by using SPSS software. From the figure below all 4 types of crime cases create 3 number of clusters, which each cluster shows similarity between states.

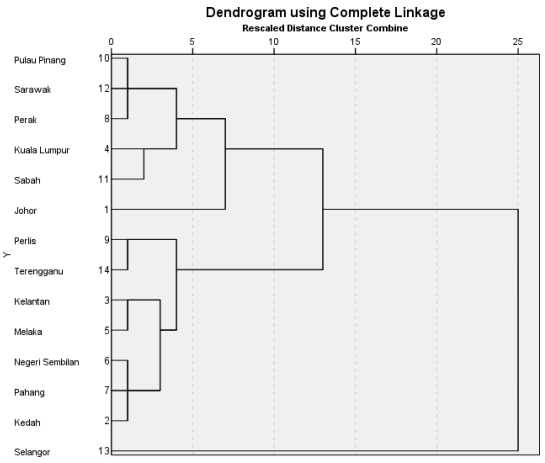


Figure 1: Dendrogram for Murder cases

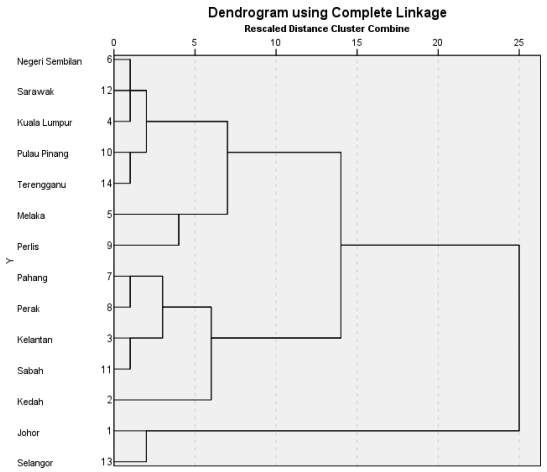


Figure 2: Dendrogram for Rape cases

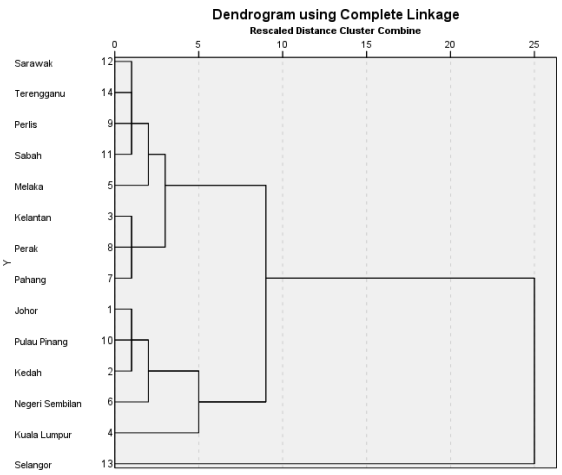


Figure 3: Dendrogram for Armed Robbery cases

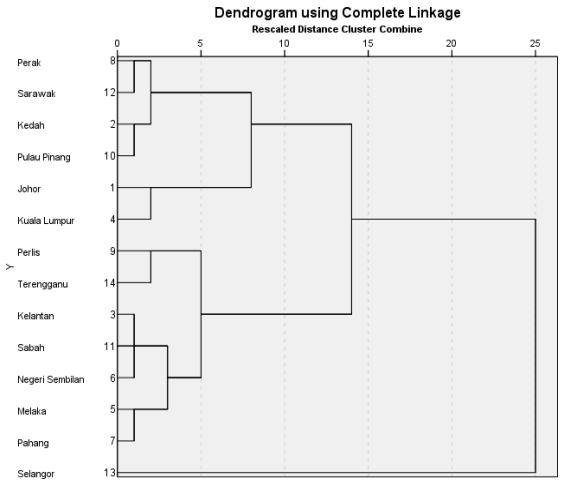


Figure 4: Dendrogram for Causing Hurt cases

The dendrogram in Figure 1 shows 3 clusters, with Pulau Pinang, Sarawak, Perak, Kuala Lumpur, Sabah and Johor grouped together in cluster 1. These states are grouped because of the socioeconomic report, which focuses on economic growth. The second cluster consists of 7 members which are Perlis, Terengganu, Kelantan, Melaka, Negeri Sembilan, Pahang and Kedah. These states have lower population in Malaysia. Selangor is the state that differs the most from the others due to its most dangerous city. There are 3 clusters created from the dendrogram in Figure 2, which is cluster 1 consists of Negeri Sembilan, Sarawak, Kuala Lumpur, Pulau Pinang, Terengganu, Melaka, Perlis, and Pahang. These are the states in Malaysia which have the highest levels of human development, according to the index. Cluster 2 consists of 5 members that are Pahang, Perak, Kelantan, Sabah, and Kedah which contain the higher poverty rate. Selangor and Johor form a different cluster group due to their significantly more dangerous than Malaysia's other urban areas.

The dendrogram in Figure 3 above shows three clusters. Sarawak, Terengganu, Perlis,Sabah, Melaka, Kelantan, Perak, and Pahang form cluster 1 which recorded the highest rates of drug addiction. Cluster 2 consists of Kuala Lumpur, Pulau Pinang, Kedah, Negeri Sembilan, and Johor withthe greatest proportion of foreign employees within their states. Selangor stands alone and form cluster 3. There are 3 clusters created from the dendrogram in Figure 4. Perak, Sarawak, Kedah, Pulau Pinang, Johor, and Kuala Lumpur form cluster 1. It contains the highest number of populations in the states. Perlis, Terengganu, Kelantan, Sabah, Negeri Sembilan, Melaka, and Pahang, which are all from middle-income states, form cluster 2. The only member of cluster 3 is Selangor.

## Conclusion

The purpose of this study is to develop the dendrogram and determine number of clusters in each crime. The cluster shows the similarity of states that assist authorities in planning an effective way to against the criminal activity. Besides, results from Mann Kendall test, indicates the downward trend based on the value of $Z_{MK} < Z_a$ make the result to reject the $H_0$ which makes the states monotonic trend. Results also show the trend states in each case is decreasing since the value of *S* is negative. across the year 2006 to 2017. It shows the good indicator throughout these 12 years. All the strategy from previous years can be updated according to this era to produce less criminal activity in Malaysia. This study can be improved by using cluster validation to confirm the number of clusters in every crime cases.

## References

[1] Anna Alvazzi del Frate, A. A. (2003). *The Voice of Victims of Crime: Estimating the true level of ...*

The Voice Of Victims Of Crime: Estimating The True Level Of Conventional Crime. https://www.researchgate.net/publication/242547931

[2] Appiahene-Gyamfi, J. (2002). An Analyses of the Broad Crime Trends and Patterns in Ghana. Journal of Criminal Justice, 30(3): 229- 243. doi:10.1016/S0047-2352(01)00138-6

[3] Brown, E. (2022). *Specialized knowledge: Understanding crime analyst's roles and ...* Specialized knowledge: Understanding crime analyst's roles and responsibilities and the impact of their work. https://journals.sagepub.com/doi/full/10.1177/17488958221095980

[4] Fougère, D., Pouget, J., & Kramarz, F.. (2009). *Youth unemployment and crime in France - JSTOR*. Youth Unemployment and Crime in France. https://www.jstor.org/stable/40282795

[5] Gama, J., & Martins Sousa, L. P. (2014). *The application of hierarchical clustering algorithms for recognition*. The Application of Hierarchical Clustering Algorithms for Recognition Using Biometrics of the Hand. https://www.researchgate.net/publication/301691212_The_Application_of_Hierarchical_Clustering_Algorithms_for_Recognition_Using_Biometrics_of_the_Hand

[6] Hinkle, J. C., Weisburd, D., Telep, C. W., & Petersen, K. (2020, June 15). *Problem-oriented policing for reducing crime and disorder: An updated ...* Problem-oriented policing for reducing crime and disorder: An updated systematic review and meta-analysis. https://onlinelibrary.wiley.com/doi/full/10.1002/cl2.1089

[7] İmrohoroğlu (2006). *Inequality and violent crime - JSTOR*. Inequality and violent crime. https://www.jstor.org/stable/10.1086/338347

[8] Omran, M. G. H., Engelbrecht, A. P., & Salman, A. (2007, January 1). *An overview of clustering methods*. Intelligent Data Analysis. https://content.iospress.com/doi/10.3233/IDA-2007-11602

[9] S. Everitt, B. (2011). *An introduction to classification and clustering - 2011 - wiley series.* Cluster Analysis. https://onlinelibrary.wiley.com/doi/10.1002/9780470977811.ch1

[10] S. Sathyadevan, M. Devan, and S. Surya Gangadharan, "Crime analysis and prediction using data mining," in Networks Soft Computing (ICNSC), 2014 First International Conference on, Aug 2014, pp. 406– 41