Descriptive Analysis Using R for Age Trend in Dengue Cases

Nor Farisha Muhamad Krishnan¹, Zuriani Ahmad Zukarnain², Marhainis Jamaludin², Noorihan Abdul Rahman²

¹ Faculty Computer and Mathematical Sciences, Universiti Teknologi MARA
Kampus Shah Alam, 40450 Shah Alam, Selangor
² Faculty Computer and Mathematical Sciences, Universiti Teknologi MARA
Kampus Machang, Bukit Ilmu, 18500 Machang, Kelantan

Authors’ Email Address: *¹farisha.krishnan@gmail.com, ²zurianiaz@uitm.edu.my, ²marhainisf@uitm.edu.my, ²noorihan@uitm.edu.my

Received Date: 21 April 2020
Accepted Date: 18 June 2020
Published Date: 31 July 2020

ABSTRACT

Dengue is a viral infection disease transmitted by Aedes mosquitoes which can lead to fatality. The case study is conducted for Kota Bharu city which is the capital city of Kelantan State in Malaysia. Different areas in Kota Bharu recorded different number of cases. The aim of this study is to see whether there is a trend in age range for those who are infected by dengue. Data on dengue cases from 2015 until 2019 were obtained from Department of Health in Kelantan. We have done descriptive analysis on range of age among the infected victims using R programming. The data visualization presented that victims from the age range between 10 years old and 35 years old were the most common in area Kubang Kerian and Panji with 20.3% and 19.7% respectively. Nevertheless, gender was not a factor for dengue cases. The mean age is 30.17 years and 28.38 years old for both ranges correspondingly. Dengue outbreak was also affected by the age of victims and has a significant health problem primarily in adolescents and young adults. Public awareness and proper vector control are needed to keep the dengue cases low and to prevent outbreaks.

Keywords: Age, Areas, Dengue cases, R programming

INTRODUCTION

Dengue is a mosquito-borne viral infection which spread by dengue virus (DENV). DENV serotypes have four types (DENV1, DENV2, DENV3 and DENV4) which means that it is possible for an individual to be infected by dengue for four times. The symptoms of dengue fever infection are suspected high fever (40°C), headache, muscle and joint sore, rashes on the skin and high fever. The virus belongs to the Flaviviridae family which is transmitted through the bite of infected Aedes aegypti and Aedes albopictus female mosquitoes (Asnis & Crupi, 2005). Dengue cases have grown dramatically around the world. In fact about half of world’s population are at risk due to dengue fever. Dengue can be found in urban, semi-urban areas and even rural areas and are beginning to affect certain countries. Usually dengue is influenced by rainfall, temperature, relative humidity and unplanned rapid urbanization. It has been estimated about 50 - 100 million infections per year arise across almost 100 countries which may spread further and this causing in high economic burden for both individuals and governments (Guo et al., 2017). Generally, dengue just
affected 9 countries before 1970 but now has spread globally specifically in WHO regions of Africa, the America, the Eastern Mediterranean, South-East Asia and Western Pacific. However America, South-East Asia and Western Pacific regions are the most extremely exaggerated, and Asia are almost 70% affected (Chau et al., 2012; World Health Organisation, 2014). The common area of dengue outbreak were reported in Western Pacific Region, mainly after 2010 and these cases were identified in China, Singapore and Malaysia (Guo et al., 2017).

In Malaysia, dengue also has become a major problem since the trend of dengue keep increasing for the past few decades. In fact The World Health Organization (WHO) stated that the incidence rise dramatically over the last 50 years and the virus keep spreading to new countries whether rural or urban. One of the countries that experience this situation is Malaysia and the trend continued to increase since 2001 until 2014 (Mudin, 2015). There was 1466 cases were reported in Negeri Sembilan in year 2010 which can be divided into 1342 dengue fever (DF) and 124 were dengue haemorrhagic fever (DHF). Dengue affected all age groups particularly young adults and males as well as mostly reported to occur in urban area and Seremban district (Ahmad Nizal et al., 2012). The study also mentioned that most of the infections happened in the first half of the year starting with DENV2 and DENV3 serotypes that being the most predominant. Many preventive measures had been taken by Ministry of Health (MOH) to cope and control the dengue transmission in the community. The actions include larvaciding, space spraying of insecticide to kill adult mosquitoes, mass premises inspection for breeding places, organizing a campaign, enforcement of Destruction of Disease-Bearing Insects Act, 1975 and so on (Abu Hassan Shaari et al., 2015). Selangor and Wilayah Persekutuan Kuala Lumpur or Putrajaya were identified to be the most infected area of dengue outbreaks because the areas were known as the busiest location in Malaysia. Apart from that, the selected regions have the great number of population which would cause high physical contact and promoted the dengue epidemics (Ibrahim et al., n.d.).

Dengue cases continue to rise gradually despite several measures have been taken to reduce dengue. All individuals’ knowledge regarding dengue and healthy lifestyle is essential for developing an intervention for future treatment. A study was conducted in Kubang Kerian which is a dengue hotspot region in Kelantan identified that most people have different knowledge about dengue signs and symptoms based on their education level, dengue fever experience, and awareness of dengue from public campaign. Different people have different ways of treatment seeking behaviour according to their financial and history of hospitalization for dengue treatment (Nur Fatini H, Mangantig E, 2017). It can be concluded that early recognition of dengue signs and symptoms are vital for immediate treatment in order to reduce complications and fatality. Kota Bharu district was recorded the highest number of cases with 291 cases on August 2019. The Health Director of Kelantan Dr Zaini Husin said that the areas which contributed to the highest number of cases were Kubang Kerian, Panji, Sering, Banggu and Limbat (Abdullah, 2019). *Aedes aegypti* and *Aedes albopictus* are the mosquitoes which cause dengue. A study on distribution of Aedes larvae was done in Kelantan, Terengganu and Sabah for three period of time (October 2008, November 2008 and June 2009). Kg. Paya Rambai, Kelantan has the highest mean number of larvae discovered for both indoors and outdoors (Wan-Norafikah et al., 2012).

The purpose of this study is to see whether there is a trend in age range for those who are infected by dengue. From the youngest to the eldest, anyone is prone to dengue because dengue spread through *Aedes Aegypti*, a domestic, day-biting mosquito that prefers to bite humans. Different range of age group contribute to the increasing of dengue cases. Since this study will display which category of age that rated high in dengue cases, immediate actions can be taken to reduce dengue outbreaks. Prior to that, the demographic analysis will take part in order to obtain the important information regarding the dengue
victims in Kelantan and provide valuable contribution to identify which age group is more likely contribute to dengue fever.

LITERATURE REVIEW

A study has been done in Malaysia which compared the urban-rural difference and determined the age specific dengue seroprevalence in Malaysia. The cases were at great level in urban areas since 2001 and quick stabilizing in rural regions at same high levels by 2008. At least 80% of Malaysian populations of 65 years of age from both urban and rural location were infected (C. H. Chew et al., 2016). In comparison, dengue cases, too, have increased in Brazil since 2001 and a research had been done to compare the number of dengue cases in two groups. Those groups are patients <15 years old and patients >15 years in São Luís, Maranhão, from 2002 to 2011. It can be concluded that children under the age of 15 was the highest age group in dengue cases patients (Luís, 2017). Similar study was conducted in San Lazaro Hospital (SLH) concluded that the ages of dengue cases ranged from 2 to 37 years old (median=17 years) and the peak rate was 15-19 years old. Dengue is regularly considered as a paediatric disease (Velasco et al., 2014). A cross – sectional study in Vietnam proved that dengue fever is becoming more serious medical issue and had affected all age groups and province globally (Huy et al., 2019). It concluded that the rate of mortality was 0.8% and the patients were mostly the elderly. Since there is no dengue vaccination, it suggested to communicate and educate people about prevention and vector control continuously.

Another study in Central Nepal was conducted to determine the occurrence of dengue in clinically suspected patients has the highest number of dengue cases that was observed in 21 to 30 years age group. This study also showed that the cases rate was higher during winter and in urban locations (Nepal, 2014). The the lowest dengue cases recorded was among children age below 10 years old. To ensure resource intensive control programs are targeted efficiently, the understanding on changes in ecology and epidemiology of dengue is vital. Thus, the previous study conducted in Singapore presented that the age group that prone to dengue fever was found to be between 21 to 40 years old and the Malay race was found to be most protected against dengue compared to others ethnicity. The findings also explained that dengue outbreak mostly occur in urban region (Yung et al., 2016).

Similar study was done in India using cross-sectional study for all diagnosed cases of dengue fever at Mahatma Gandhi Hospital for a certain period (1st September to 31st October 2015). The highest number of dengue cases were reported young males within the age group 21 to 40 years as they were exposed to outdoor activity (Singh Rathore et al., 2015). Dengue has been a health threat and endemic and huge number of dengue cases were reported in monsoon and post monsoon period (September and October) every year (Singh Rathore et al., 2015; Yousaf et al., 2018). The study in Faisalbad explained that males between 16 – 30 years of age are more likely to expose to dengue fever. These people mostly spend their time outdoor so they are more prone to be affected. Thus, dengue control program should be done in order to manage surveillance rate (Yousaf et al., 2018).

The increasing rate of dengue cases has significant impact on global healthcare services. Mostly individuals within the mean age of 30 years are prone to dengue fever (Guo et al., 2017; Liew et al., 2016). The dengue patients were frequently reported clinical symptoms such as fever, malaise, headache, and asthenia. There is a study in Kuala Lumpur indicated that prevalence of dengue was higher in age group between 20 to 29 years. The infection was more to male patients and in Malay population. In fact, the highest infection recorded in the month of June and the lowest was in August (Chew & Rahman, 2012). A quite similar study was done in Klang Valley as it has high dengue cases with a total number of 61,455 cases were registered.
in year 2014. The dengue patients were mostly between 17 years to 37 years of age and come from urban populations (Woon, Yuan Liang; Wan Ng. Chiu; Nani Mudin, Rose; Suli, 2019). Overall, it can be concluded that age group influence the number of dengue cases.

METHODOLOGY

Study Area

This study took place in Kelantan which has 10 districts in total. Kelantan has a tropical climate, with temperatures between 21°C and 32 °C and intermittent rain throughout the year. However, the study only focused on Kota Bharu with an area 409 km². Kota Bharu is located in the north-eastern part of Peninsular Malaysia, and lies near to the mouth of the Kelantan River at 6°8′N 102°15′E. In 2010, the district’s population is estimated to be 491,237. Kota Bharu is chosen to be part of this study because it is the capital city of Kelantan which has a large population number. Besides Kota Bharu has the highest dengue cases (57.4%) compared to other districts as shown in Figure 1.

![Bar Chart of Dengue Cases based on District for 5 Years](image)

**Figure 1: The dengue cases in Kelantan for the past 5 years from 2015 – 2019**

Kota Bharu can be categorized into several areas for instance Badang, Banggu, Beta, Kadok, Kemumin, Ketereh, Kota, Kota Bharu, Kubang Kerian, Limbat, Panji, Pendek, Peringat, Salor and Sering. It is found that Kubang Kerian and Panji that contribute to the highest dengue fever cases as shown in figure 3. It shows that 20.3%, the highest of dengue cases in Kubang Kerian followed by Panji with 19.7% of cases. It can be generalized of what health Director of Kelantan had said that Kubang Kerian and Panji contribute the most to dengue cases.
Data Collection

This study used secondary data obtained from Department of Health in Kelantan. The data is from 2015 until 2019 which include all district in Kelantan with the parameters age, districts, area, race, identification number, and notification number. Basically, the data was provided by Vector Borne Disease Control in Kelantan that represents the accumulated number of dengue cases reported in Kelantan.

Table 1: Description of the parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Types</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Numeric</td>
<td>The age of dengue victims</td>
</tr>
<tr>
<td>Districts</td>
<td>Categorical</td>
<td>All districts in Kelantan</td>
</tr>
<tr>
<td>Area</td>
<td>Categorical</td>
<td>The chosen area is in Kota Bharu</td>
</tr>
<tr>
<td>Identification Number</td>
<td>Numeric</td>
<td>The number that used to identify whether the patients is the same patients and will get to know the recurring cases</td>
</tr>
<tr>
<td>Notification Number</td>
<td>Numeric</td>
<td>The unique identification, thus the duplicate data can be avoided</td>
</tr>
</tbody>
</table>

Analysis Method

Extracting knowledge from data in context of huge databases is the goal for term Knowledge Discovery in Databases (KDD). Pre-processing is one of the steps in KDD process which come after data selection. Terms of KDD and data mining are different. KDD refers to the overall process of determining the valuable information from data whilst data mining is an application of algorithms to extract patterns from the data without additional steps of the KDD process. KDD involves the evaluation and interpretation of the patterns to build decision of what qualifies as knowledge. It additionally incorporates the decision of encoding plans, pre-processing, examining, and projections of the data prior to the data mining step (Fayyad et al., 1996; Gullo, 2015). The KDD process is as the following:
Since the study used secondary data, the data should undergo pre-processing like data cleaning which need to remove duplicate and dealing with missing value. Some datasets may have duplicate data which may affect the finding. To identify duplication, a unique identification is needed. Removing the duplicate data will produce the better output. Elimination methods for existing duplicate in data cleaning work on the origin of computing the degree of similarity between nearby records in a sorted database (Low et al., 2001).

Missing value in datasets can cause great problems to machine learning systems and data mining which are not prepared to handle them, loss of information, increased the errors, and the findings might be bias. In fact missing data are the source of weakness in learning systems and may degrade the quality of data (Abu-soud, 2019; Dong & Peng, 2013). There is a study on missing data treatment that would analyses the missing data mechanisms and treatment rules. Popular and conventional treatment were introduced and compared. The listed methods are case deletion (CD), Maximization Likelihood Methods (ML), Mean/Mode Imputation (MMI), All Possible Values Imputation (APV), Regression Methods (RM), K-Nearest Neighbor Imputation (KNN), Multiple Imputation (MI), and Bayesian Iteration Imputation (BII) [28]. There are 3 categories of missing data. The highest level of randomness is called Missing completely at random (MCAR). It means that the probability of missing on attribute does not depend on any value of attribute. Next is missing at random (MAR) which is the probability of missing data on any attribute does not depend on its own particular value, but on the values of other attributes. The third one is not missing at random (NMAR). Missing data depends on the values that are missing (Peng & Lei, n.d.).

Though missing values are representative of messiness in real world data, removing them might reduce the amount of the data or probably the worst may happen in machine learning. In order to treat missing data, it can be manually inspected as ‘NaN’, ‘NA’, ‘None’, ‘’ , ‘?’ and others. This may work in some cases when the proportion of missing values is relatively low (<10%) (Angelov, 2017). It is very common to find missing value in almost every dataset. They may not cause any trouble for Knowledge Discovery in Databases (KDD) process if the rate of missing is less than 1%. It is still manageable if the rate between 1% and 5% but sophisticated methods need to be handle when the rate is 5-15% and more than 15% will cause severely impact in any interpretation because the discoveries may be bias (Acuña & Rodriguez, 2004; Peng & Lei, n.d.).

On the first stage, the duplicate data was checked using notification number and no data was recurring. Then analysis proceeded to check the missing data. There were only 5 missing data in variable age, and they were removed as they did not affect the data. There were recurring data in patients after undergoing the process using the IC which is a unique identification for patients. This shows that an individual can infected by dengue more than once. However, we just ignored as it does not give any effect to analysis since it is still categorised as dengue victims.
In order to have a better analysis on graph, R programming was used for this study and several packages needed to be installed. R language is an open source software that used for statistical computing and get a better graphics. R can be obtained for free of charge from Comprehensive R Achieve Network (CRAN) (Lo, 2006). This study only focused on age trends and locality especially in Kota Bharu. First of all, the data which is in excel sheet need to be import into R and package readxl is required.

```r
> data <- read_excel("DengueData.xlsx")
```

Installing the package ggplot2 in order to build any type of graphs like bar chart, pie chart, histogram, boxplot and so on. A bar graph is built to determine which locality has the highest dengue cases.

```r
> ggplot(data, aes(x)) + 
+   geom_bar(aes(y=(..count..)/sum(..count..)), fill="cyan4") +
+   geom_text(aes(y = ((..count..)/sum(..count..)), label =
+   scales::percent((..count..)/sum(..count..))),
+   stat = "count", vjust = -0.35) +
+   scale_y_continuous(labels = percent) +ggtitle("Bar Chart")
```

```r
> ggplot(data, aes(x = )) + geom_histogram(binwidth = 10,
+   aes(fill=..count..), fill="yellow", col="black") +
+   geom_vline(xintercept=mean(data$x), lwd=1, linetype=2,
+   color="red") + ggtitle("Histogram")
```

**RESULTS AND DISCUSSION**

After the data had undergo the pre-processing, the pattern of dengue victims’ age can be analysed through histogram as shown in figure 3 which take place in Kota Bharu. The mean of age is 28.92. This shows that dengue cases are high among the adolescents. Based on the figure 4, it supported that age range between 15 – 30 years old for those areas which have the highest dengue cases.

![Histogram of Age for Five Years](image)

*Figure 4: The histogram for age of dengue victims for 5 years in Kota Bharu from 2015 until 2019*
Then the study continued to determine the age trend in Kubang Kerian and Panji since both areas had quite high cases. In both areas the dengue cases was rather similar. Even the distribution of age for dengue victims is precised to Kubang Kerian and Panji, the range of age is still between 10 to 35 years old with mean value 30.17 years and 28.38 years respectively. This shows that age give significant effect in increasing dengue cases in Kelantan. According to previous study, this research can conclude that the adolescents contribute to the highest dengue cases.

Figure 5: The histogram for age of dengue victims in 5 years in Kubang Kerian from 2015 until 2019

Figure 6: The histogram for age of dengue victims in 5 years in Panji from 2015 until 2019
CONCLUSION

The results of this study describe the demographic trends of dengue infections in Kota Bharu between year 2015 and 2019. There was an increase in older age groups reporting dengue fever. Dengue fever in people > 15 years of age is now more common than in children in this district. Even when the analysis was specified to Kubang Kerian and Panji, the range of age still the same which is between 10 years and 35 years patients. This means that dengue cases were high among the children and especially adolescent. Briefly, age trend contributes to the dengue cases to keep rising. However, the current analysis was not able to detect factor to quantify a relationship between age and dengue fever. Future work needs to be done to determine why people age between 15 – 30 years are the highest in being infected with dengue.

REFERENCES


Woon, Yuan Liang; Wan Ng, Chiu; Nani Mudin, Rose; Suli, Z. (2019). Health facility use by dengue patients in the Klang Valley, Malaysia: a secondary analysis of dengue surveillance data. 10(2), 39–45. https://doi.org/10.5365/wpsar.2019.10.1.001

