ENTHUSIASTIC
INTERNATIONAL JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 1, Issue 1, April 2021, pp. 7-12

Survival Analysis Based on Average Response Time of Maritime Search and Rescue (SAR) Incidents in 2019 Using Kaplan-Meier Method and Log-Rank Test

Duhania Oktasya Mahara [a,1,*], Muhammad Hasan Sidiq Kurniawan[b,2]

[a] Departement of Statistics, Universitas Islam Indonesia, Jl.Kaliurang km.14,5, Yogyakarta 55584, Indonesia
[b] Departement of Statistics, Universitas Islam Indonesia, Jl.Kaliurang km.14,5, Yogyakarta 55584, Indonesia
[1] duhaniaoktasya@gmail.com*; [2] hasan.sidiq@uii.ac.id
* corresponding author

ARTICLE INFO

Article history
Received February 13, 2021
Revised March 25, 2021
Accepted April 15, 2021

Keywords
Response time average for ship accidents
Survival analysis
Kaplan-Meier method
Log-Rank test

ABSTRACT

Indonesia is the largest archipelagic country in the world (based on area and population), which makes it as one of countries with the most significant maritime activities. Therefore, there has been a high rate of maritime accidents in Indonesia. The National Search and Rescue Agency (BASARNAS) as a non-ministerial government agency with the primary task of Search and Rescue (SAR) operation deals with several types of accidents, including maritime accidents. Response time as the time to receive news about the accidents until the SAR unit comes to the rescue is very crucial in this matter. Average response time is stipulated based on BASARNAS's regulations to estimate information about the survival probability of the victims. This research concerns with the survival analysis using Kaplan-Meier Method and Log-Rank Test. The researchers categorized maritime accidents into three categories: 'Low', 'Medium', and 'High'. This classification aims to find out whether the survival function of each category has the same or different function and to investigate whether there are differences from the given responses or not. The survival analysis with Kaplan-Meier method revealed that the three categories had different survival functions. The survival analysis was followed by a Log-Rank Test. The final result shows that there is no difference in the responses given by the three categories when maritime accidents occur.

1. Introduction

Two-third of Indonesia's territory is water and Indonesia has a water area of 3.25 million km^2 [1]. This fact makes Indonesia as the largest archipelagic country in the world (based on area and population) [2]. It also contributes to the high rate of maritime activities in Indonesian waters, and thus leading to the high rate of maritime accidents. The National Search and Rescue Agency (BASARNAS) is a non-ministerial government agency with a primary task of search and rescue (SAR) operations. The SAR operations at BASARNAS are divided into five groups of accidents, namely Aircraft accidents, Maritime, Disaster, Hazardous Condition for Human, and Special Handling (LPK).

p-ISSN 2798-253X
e-ISSN 2798-3153

The ability to react or the ability to respond to each of the occurring accident greatly affects the success of this SAR operation. A measure ability or response ability can be measured through the data of response time for each accident. The determining factor for this response time is the time to receive the news until the SAR Unit moves to the scene. Using the average time response data from each SAR office that has been determined in accordance with the regulations of the National Search and Rescue Agency, the research can estimate information about the survival probability or survival tests using Survival Analysis or Survival Test Analysis which has a variable of outcome time unit.

In this research, the researcher categorized the maritime accidents into three levels, namely low, medium, and high. This research aims to know the survival function of each category and to find out whether there are different treatments in response based on the maritime accidents in each province/different SAR offices based on the average response time of the number of maritime accidents in Indonesia in 2019.

2. The Proposed Method

2.1. Survival Analysis

Survival analysis is one of statistical methods that aims to analyze data with the outcome variable of time unit. The time unit data refers to the time of the study to the occurrence of an event and can also be referred to as survival time [3]. This survival analysis can be applied in many sectors, such as biology, medicine, sociology, engineering, and other fields [4].

2.2. Survival Function and Hazard Function

Survival function is the probability of an individual to survive longer than t. The survival function is a monotonous fall to time t.

$$S(t)=P(T>t)$$

$$= \int_t^{\infty} f(x)dx \quad (1)$$

Meanwhile, hazard function is the level of occurrence of an event. If it is known that a research object survives until time t, this function is not probability. Thus, the possibility of the value is greater than 1. Hazard function is denoted by the h letter with the hazard function being $h(t) \geq 0$.

2.3. Kaplan-Meier Method

Kaplan-Meier method is one of survival methods in statistics, which uses an exact time rather than the time interval of follow-up. Kaplan-Meier method or sometimes known as the product-limit was discovered in 1958 by Kaplan and Meier. This method is used to deal with missing data problem [5]. If there is no censored observation, the survival function is as written in the following equation:

$$S(t)= \frac{\text{number of survived object on } T \geq t}{n}, \; t \geq 0 \quad (2)$$

Kaplan-Meier estimator functions as ladder that drops when an event occurs. The Kaplan-Meier estimator is a non-parametric function, which does not assume an infinite number of parameters. The number of parameters or quantities that will be estimated in Kaplan-Meier is as many points in time when the event occurs. In this method, grouping is not carried out at certain intervals. The Kaplan-Meier method can be used to handle censored data with relatively simple calculations, which can also be applied to small sample data. This method also provides a graphical representation of the survival function [6].

2.4. Log-Rank Test

This test is carried out by grouping samples into certain intervals. The Log-Rank test is often used to analyze samples which are grouped into two or more related groups [7]. There are several steps that must be taken in analyzing using the Log-Rank Test method. The first step is determining the hypothesis. The hypothesis for this test is written in the followings:

$H_0: S_i(t)=S_j(t)$ (There is no difference between each handling group)

$H_1: S_i(t)\neq S_j(t)$ (There is a difference between each handling group)

p-ISSN 2798-253X
e-ISSN 2798-3153

The next step is determining the statistics test, which aims to determine the average similarity of events between two groups:

$$W=\frac{(O_1-E_1)^2}{E_1}+\frac{(O_2-E_2)^2}{E_2} \quad (3)$$

where:
O_1, O_2 : the number of uncensored observations for each group.
$E_1=\sum_{t\in T} e_{1t}$ dan $E_2=\sum_{t\in T} e_{2t}$: the expected value of the event occurring in each group.
H_0 is rejected if the result $W > \chi^2_{\alpha,df=1}$.

3. Method

3.1. Data Source

This research obtained secondary data, taken from the National Search and Rescue Agency data recapitulation, related to maritime accident data handled by BASARNAS during 2019 per province/SAR office, which includes the average response time.

3.2. Reseach Variabel

This research analyzes the variables presented in Table 1.

Table 1. Units of Research Variables

Variable	Description
Location	Province or SAR office location
Total	Total number of Maritime accidents
Survive	Survivors
Die	Died Casualties
Missing	Missing Casualties
Mean of response time	Average Response time during 2019
Low accident rate	Maritime accident with 0 to 12 casualties
Medium accident rate	Maritime accident with 13 to 26 casualties
High accident rate	Maritime accident with > 26 casualties

3.3. How to Collect Data

This study used secondary data, which were not obtained directly from the research object or subjects, but were taken from the archive of the National Search and Rescue Agency.

3.4. Data Analysis Method

This research used Microsoft Excel 2013 and R Studio as the research software and the Kaplan-Meier and Log-Rank Test to analyze the data. The research analysis was conducted using the following steps:

1. Inputting the data
2. Conducting a survival analysis using Kaplan-Meier method (Making a plot)
3. Interpreting the plot results.
4. Making assumptions from the results.
5. Conducting a Log-Rank Test to confirm existing assumptions.
6. Doing hypothesis test.
7. Drawing conclusions from the results of the analysis.

4. Results and Discussion

There has been a varying rate of maritime accidents from one area to another in Indonesia. The different number of maritime accidents in each location is presented by the following distribution map technique.

p-ISSN 2798-253X
e-ISSN 2798-3153

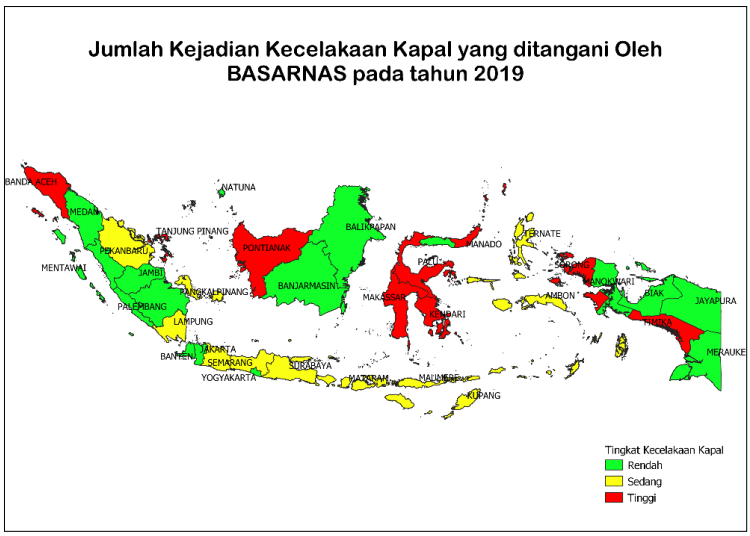

Fig.1. Map of the distribution number of maritime accidents in 2019

The map in Fig.1 visualizes the distribution of maritime accidents handled by BASARNAS in 2019, which have been categorized and differentiated by color. The green color shows the province/SAR office location with a low rate of maritime accident, the yellow color indicates the province/SAR office location with a medium rate of maritime accident, and the red color is a province/SAR office location with high rate of maritime accident.

4.1. Kaplan-Meier

This nonparametric method for survival analysis is based on three categories, namely Low Rate of Maritime Accident, Medium Rate of Maritime Accident, and High Rate of Maritime Accident. The following figure illustrates the outputs of the analysis results and plots of each category.

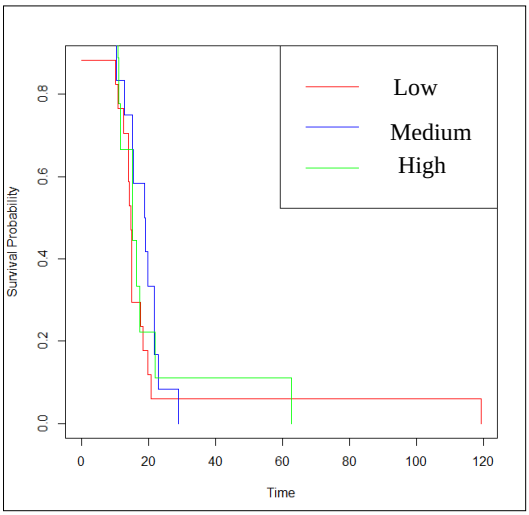

Fig.2. Kaplan-Meier plot of three categories

From the output as shown in Fig.2 above, it can be seen that the survival function of three categories has a different S(t) line. It starts from the time span of 20-40 minutes for Low Rate of Maritime Accident that gets a faster response time of approximately 20 minutes. Then, the Medium Rate of Maritime Accident and High Rate of Maritime Accident get the slower response time. Therefore, it can be assumed that the Low Rate of Maritime Accident, Medium Rate of Accident, and High Rate of Maritime Rate has a significantly different survival function. It can be assumed that these differences are attributed to differences in handling and responding to the call (giving or not). To ensure this assumption, the researcher conducted the next step of Log-Rank Test.

4.2. Log-Rank Test

This following are the results of the Log-Rank analysis of average response time of maritime accidents that were handled by BASARNAS in 2019. In this Log-Rank test, the maritime accident rate is divided into three categories based on their levels: Low, Medium, and High. This category is based on the total number of maritime accidents for a year by each province or SAR Office location. The result is described below:

Hypothesis Test:

i. Hypothesis

H_0:$S_1(t)=S_2(t)=S_3(t)$ (There is no difference in handling to giving response between maritime accident categories)

H_1: at least one category has a different $S(t)$ (There is a time difference in handling to giving response between maritime accident categories)

ii. Significance level

$\alpha = 0.05$

iii. Statistics test

$$W=\frac{(O_1-E_1)^2}{E_1}+\frac{(O_2-E_2)^2}{E_2}+\frac{(O_3-E_3)^2}{E_3}$$

iv. Critical area

If p-value < α, H_0 is rejected or if $W > \chi^2_{\alpha,df=1}$ H_0 is rejected

v. Decision

Because the obtained p-value = 0.6 > α = 0.05, H_0 is not rejected

vi. Conclusion

With a significance level of 5% or 0.05, H_0 is not rejected. In other words, there was no difference in handling to giving response time between all maritime accidents that were handled by BASARNAS in 2019.

5. Conclusion

From the discussion, the following can be concluded:

1. Using Kaplan-Meier analysis, the survival function for each maritime accident rate showed no different between one to another. This fact was shown by the Fig. 2 that the line of survival function between categories coincide with each other. However, those conclusion is solely based on Kaplan-Meier graphic. Therefore, it is necessary to use statistical test to determine whether there are any different survival function on each category.
2. Using Log-Rank test, it is clear that there is a difference of survival time between maritime accident categories. In other words, there is a different response time for each category. Looking at Fig. 2, we can conclude that the high rate of accident have slow response time. This fact is indicated by the highest value of survival function. On this basis, it is necessary to improve the average survival time to minimize the casualties of maritime accidents.

Acknowledgment

The authors gratefully acknowledge Mr. Muhammad Hasan Sidiq Kurniawan, S.Si., M.Sc for his continuous support and advice throughout the research. The authors would also like to express gratitude to the family and any parties for their valuable support and contribution in this research.

References

[1] K. T. Pursetyo, W. Tjahjaningsih and H. Pramono, "Comparative Morphology Of Blood Cockles in Kenjeran and Sedati," *Jurnal Ilmiah Perikanan dan Kelautan*, vol. 7, no. 1, pp. 31-33, 2015.

[2] S. T. Hutagalung, "Penetapan Alur Laut Kepulauan Indonesia (Alki): Manfaatnya dan Ancaman bagi Keamanan Pelayaran di Wilayah Perairan Indonesia," *Jurnal Asia Pasific Studies*, vol. 1, no. 1, 2017.

[3] D. G. Kleinbaum and M. Klein, Survival Analysis: A Self-Learning Text, London: Springer, 2012.

[4] D. Collet, Modelling Survival Data in Medical Research, USA: Chapman& Hall/CRC, 2003.

[5] J. Lawless, Statistical Models and Methods for Lifetime Data, New York: John Wiley & Sons, 2003.

[6] A. Cook, Kaplan Meier Estimate of (S(t)), Singapore: National University of Singapore, 2008.

[7] J. P. Moeschberger, Survival Analysis Techniques for Censored and Truncated Data, New York: Springer Science and Business Media, 2003.

p-ISSN 2798-253X
e-ISSN 2798-3153