Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754
P-ISSN : 2580-569X
Vol. 6, No. 2, Desember 2022
pp. 107-117

MODEL REGRESSION COX PROPORTIONAL HAZARD WITH BAYESIAN METHOD FOR SURVIVAL ANALYSIS OF COVID-19 PATIENT CASES AT RSUD Dr. PIRNGADI KOTA MEDAN

Aurelia Anandara[1], Hamidah Nasution[2], Ismail Husein[3]
[1,2]Department of Mathematics-Faculty of Mathematics and Natural Sciences / State University of Medan
[3]Department of Mathematics/ Universitas Islam Negeri Sumatera Utara Medan

Article Info

Article history:

Keywords:

Bayesian, Covid-19, Cox proportional hazard, survival analysis.

ABSTRACT

Survival analysis is a statistical procedure for analyzing data by observing the response variable in the form of event time data from the beginning of recording to the end of the event. Survival analysis is used in many fields of medicine. The cox proportional hazard model aims to look at the factors of the recovery rate on patient survival. In this study using Bayesian data with Lognormal distribution in Covid-19 patients at Dr. Pirngadi, Medan City. The predictor variables are Age(X_1), gender(X_2), employment status(X_3), and other diagnose(X_4). Based on the research, the cox proportional hazard model was obtained $\hat{h}(t) = h_0(t)\exp(-0{,}05354X_1 - 0{,}9176X_2 - 0{,}1507X_3 + 0{,}1999X_4)$ with the influential variables based on the credible interval it is known that the age(X_1) and gender$(X_2)$are significant variables. Among the two variables that have the most influence is the age(X_1) because it obtains a larger coefficient, namely $e^{((-0.05354))}$.

Corresponding Author:

Hamidah Nasution,
Department of Mathematics,
State University of Medan, Indonesia
Email: aureliaanndra25@gmail.com

1. INTRODUCTION

Covid-19 is a virus that attacks the respiratory tract with various symptoms that can be experienced, either moderate or severe, that can cause death. According to the Ministry of Health, currently there are around 6 million people in Indonesia who have been infected with Covid-19 with a death rate of around 160 thousand fatalities. Individuals who are vulnerable to Covid-19 include the elderly, people with a history of serious illnesses, smokers, men, and people with blood type A (Siagian 2020).

In the fight against the Covid-19 pandemic, several things can be done by increasing antibodies. Because the immune system is dynamic, that is, it can go up and down. Several factors can affect the body's immunity such as age, hormones, nutrition, emotions, and lifestyle (Amalia 2021).

Survival analysis is currently widely used in the medical field. Survival analysis is a statistical procedure for analyzing data by observing the response variable in the form of event time data from the beginning of recording to the end of the event. Survival analysis aims to estimate the hazard function and survival function based on predictor variables to be interpreted in order to see the relationship between the suspected variables and survival time. Survival analysis used the cox proportional hazard model to see the factors of the recovery rate on the patient's survival ability. In this study, a parametric approach was used in the form of a Bayesian approach with Lognormally distributed data.

Compared to classical statistics, Bayesian has a prior distribution which makes data inference more accurate. Bayesian uses a direct probability distribution on its parameters which makes the level of confidence greater than classical statistics (Preatin 2007). In this Covid-19 case data, the initial distribution that is formed is a Lognormal distribution. According to (Aristizabal 2012) the Lognormal Distribution provides a more accurate estimate of location parameters than all the original data, especially when the population slope is low and the distribution distribution is high.

Based on previous research by (Putri 2017) with the research title "Survival Analysis on the Recovery Rate of Diabetes Mellitus Patients with a Bayesian Approach (Case Study at the Regional General Hospital Dr. Saiful Anwar Malang)". This study uses the cox proportional hazard regression model and the Bayesian approach. With the conclusion that the most influential factor is the age variable.

The purpose of this study itself was to obtain a proportional hazard cox regression model with a Lognormal distribution for the recovery rate of Covid-19 patients and to determine the factors that most influence the recovery rate of Covid-19 patients.

2. RESEARCH METHODE

2.1. Covid-19

Coronavirus Diseases 2019 or commonly known as Covid-19 is a new type of infectious disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). On Covid-19 stil it is not yet known which animal is the source of transmission (KEMENKESRI 2020). With three general symptoms, namely fever, dry cough, and shortness of breath. According to Hidayani (2020) several risk factors in Covid-19 patients, namely age, gender, nosocomial infection from sufferers and hospital staff, comorbid diseases.

2.2. Survival Analysis

Survival itself means the survival of an individual. Survival analysis according to (Kleinbaum 2012) is a statistical procedure for analyzing data by paying attention to the response variable in the form of event time data from the beginning of recording to the end of the event. Time is defined as the observation period of an event expressed in days, weeks, months or years. Meanwhile, events are events that occur such as recurrence, healing or death.

2.3. Survival Function

The survival function $S(t)$ is the probability that the subject or individual survives longer than time t with $(t > 0)$ or can be interpreted as the probability probability that the random variable T is more than time t (Harlan 2017). Suppose T is a non-negative random variable which represents the time until failure occurs. So the cumulative function is declared:

$$F(t) = P(T \leq t) = \int_0^t f(t)dt \tag{1}$$

Based on the cumulative function $F(t)$ from T above, the survival function is defined as follows:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) \tag{2}$$

2.4. Hazard Function

The hazard function h(t) or also called the conditional failure rate is the probability of failure occurring in the time interval from t to t + Δt, with a condition where individuals survive until the beginning of the interval, divided by the width of the interval, expressed as follows (Harlan 2017):

$$h(t) = \frac{f(t)}{S(t)} \tag{3}$$

2.5. Cox Proportional Hazard

The cox proportional hazard model is the most frequently used multiplicative hazard model. Cox proportional hazard is a regression to analyze data on an event that is influenced by covariates with the assumption that it is multiplicative at the baseline hazard. It is called cox proportional hazard because between individuals it has a proportional hazard rate ratio and a constant ratio value (Fernandes 2016). The cox proportional hazard model can model response variables with one or more predictor variables, this model is also able to estimate regression coefficients, hazard ratios, and survival curves. According to (Kleinbaum 2012) the general form of the cox proportional hazard model for the i individual is:

$$\hat{h}(t|X) = h_0(t)\exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right) \quad (4)$$

2.6. Proportional Hazard Assumption Test

According to (Fernandes 2016) The proportional hazard assumption is the ratio at two levels of the hazard function where for level one it is proportional to the hazard function, while at level two where the ratio for both is constant and does not depend on time. Examination of assumptions is divided into two, namely by looking at the graph and numerical or calculation. Checking numerically is more difficult. Meanwhile, a graphical examination can be done by looking at the plot on the log-log survival graph, namely the graph $log\left(-\,log\left(\hat{S}(t)\right)\right)$. Where it is said to be fulfilled if the two log-log survival graphs are parallel. The following are the steps in checking assumptions with a log-log survival chart:

1. Find the estimation of survival function $\hat{S}(t, X)$.
2. Find the result of survival log-log, , $log(-\,log(\hat{S}(t, X)))$.
3. Making the x-axis graph is survival time, while the y-axis is $log(-\,log(\hat{S}(t, X)))$.
4. If each predictor is searched for, the two plots are parallel and two individuals with different curve covariates on a parallel graph then the assumptions are met.

Numerical testing can be carried out using the goodness of fit method using the Schoenfeld residual. According to (Kleinbaum 2012) the fulfillment of numerical testing using the goodness of fit method assumptions are fulfilled if there is no correlation between rank survival time and Schoenfeld residual which is carried out by testing the hypothesis which first calculates the value of the Pearson correlation coefficient. Hypothesis testing is done by looking at the value of the p-value. Hypothesis: H_0: There is no correlation between rank survival time and Schoenfeld residual H_1 : There is a correlation between rank survival time and Schoenfeld residual. If the $p - value < \alpha$ then the decision is rejected H_0, and otherwise H_0 is accepted, where $\alpha = 0,05$.

2.7. Goodness Of Fit Test

The goodness of fit test or model feasibility test is used to measure the accuracy of the sample regression function on the estimated actual value in statistics. The goodness-of-fit test can also be used for survival data analysis which was carried out at the beginning of the study to estimate the distribution of long-time data (Ghozali 2018). In this study, the Anderson-darling test will be used. The Anderson-darling test uses a certain distribution to estimate the critical value with the weakness that the critical value must be calculated for each distribution. With the following equation:

$$A_n^2 = [-\frac{1}{n}\{\sum_{i=1}^{n}(2_i - 1)[\ln F(x_i) + \ln(1 - F(X_{n+1-i}))]\} - n \quad (5)$$

Hypothesis:

H_0 : Survival time according to the predicted distribution

H_1 : Survival time not according to the predicted distribution

If $A_n^2 > \alpha_{n,1-\alpha}$, then the decision is rejected H_0, otherwise the H_0 is accepted.

2.8. Bayesian Analysis

Bayesian is a statistical approach which is a method of using sample data from a population, where it is necessary to know the shape of the initial distribution or priors. The advantage of using Bayesian is that it contains information on prior states in data processing. Using the principle of direct distribution of the parameters. In the Bayesian approach that must be considered, namely the parameter θ. The parameter θ which has a probability $P(\theta)$is the initial level of confidence before the observations are made, it can be called the prior distribution (Katianda 2021). Then the general theorem of the Bayesian approach is:

$$p(\theta|t) = \frac{p(t|\theta)p(\theta)}{p(t)} \quad (6)$$

There are several prior distributions divided into several types, namely:

1. Conjugate or non-conjugate priors based on the likelihood model pattern of the data. It is said to be Conjugate if the prior distribution pattern has a conjugate shape with a density function of its likelihood builder, whereas it is said to be non-conjugate if the prior distribution does not pay attention to the forming pattern of its likelihood function (Laidat 2022).
2. Informative and non-informative priors based on the availability of prior information on the data distribution pattern. It is said to be informative if information about the parameters is known with the prior influencing the results of the posterior distribution and is subjective, while it is said to be non-informative if the information is not available with no significant influence on the posterior distribution and is objective (Diana 2016).

2.9. Markov Chain Monte Carlo(MCMC)

Markov Chain Monte Carlo is a method of constructing random variables based on the characteristics of the Markov chain. By combining the iteration method estimates where the value affects the previous step. One of the most common MCMC techniques is Gibbs Sampling (Katianda 2021). MCMC is a stochastic procedure $\{\theta_1, \theta_2, \dots, \theta_K\}$ to get the equation:

$$f(\theta^{(K+1)}|\theta^{(K)}, \dots, \theta^{(1)} = f(\theta^{(K+1)}|\theta^{(K)}) \quad (7)$$

In constructing $f(\theta|x)$a Markov chain must be prepared first. Here are the steps (Banner 2020):

1. Find the initial value (θ^0).
2. Generating samples with K iterations.
3. Analyze the convergence of sample data. If it is less convergent then it needs to build up more data.
4. Use $[\theta^{(\beta+1)}, \theta^{(\beta+2)}, \dots, \theta^K]$ for posterior sample analysis.
5. Make a posterior distribution plot.

2.10. Gibbs Sampling

Gibbs Sampling is one of the MCMC approaches that can be used even though the data has a difficult distribution and is not explained clearly with the condition that the distribution is univariate. The Gibbs Sampling Algorithm is used to take samples based on high-dimensional complex distributions.

$$p(\alpha|x, \beta_2, \dots, \beta_p, \gamma), p(\gamma|x, \beta_2, \dots, \beta_p, \alpha), p(\beta_1|x, \beta_2, \dots, \beta_p, \alpha, \gamma), \dots, p(\beta_p|x, \beta_1, \dots, \beta_{p-1}, \alpha, \gamma) \quad (8)$$

Or it could also be written as $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(p)}$ (Candrawengi 2020). The steps of the Gibss Sampling algorithm are as follows:

1. Determine the initial value for each parameter $\theta^{(0)} = \theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)}$. $\theta^{(0)}$ is an arbitrary value according to the terms of each distribution of p is the number of parameters.
2. Build the component from each parameter after specifying the initial value, namely:
   $\theta_1^{(r)}$ dari $f_1(\theta_1|\theta_2^{(r-1)}, \theta_3^{(r-1)}, \dots, \theta_p^{(r-1)}, x)$
   $\theta_2^{(r)}$ dari $f_2(\theta_2|\theta_1^{(r)}, \theta_3^{(r-1)}, \dots, \theta_p^{(r-1)}, x)$
   $\theta_3^{(r)}$ dari $f_3(\theta_3|\theta_1^{(r)}, \theta_2^{(r)}, \theta_4^{(r-1)} \dots, \theta_p^{(r-1)}, x)$
   $\vdots$
   $\theta_p^{(r)}$ dari $f_p(\theta_p|\theta_1^{(r)}, \theta_2^{(r)}, \theta_3^{(r)} \dots, \theta_p^{(r-1)}, x)$
3. Obtain the mean, median, standard deviation, quartiles, MC error of the posterior distribution.

2.11. Convergent Check

In the calculation of the MC Error, convergence will be achieved if the MC Error is less than 1% standard deviation (Ntzoufras 2009).

2.12. Parameter Testing

Parameter testing aims to see the results of the response variable on the predictor variable by evaluating and comparing the number of quartiles of the posterior distribution. The hypothesis is:

1. $H_0: \beta_1 = 0$ (the-j predictor variable has no effect on the response variable)
2. $H_1: \beta_1 \neq 0$ (the-j predictor variable has an effect on the response variable)

The test used for the Bayesian approach is a credible interval with a lower limit of 2.5% and an upper limit of 97.5% with a value of $\alpha = 97,5\%$. $H_0$is rejected if the credible interval has no result of 0 which means that the response variable significantly influences the predictor variable (Ntzoufras 2009).

2.13. Distribution Lognormal

Lognormal distribution is defined as the distribution of a random variable whose logarithm is normally distributed. The following is the probability density function of the random variable:

$$f(x;\mu;\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\log(x) - \mu)^2\right) \quad (9)$$

Where $\mu, \sigma$is the parameter of the Lognormal distribution. With $x > 0, -\infty < \mu < \infty, \sigma > 0$, where μ is the scale parameter, σ is the shape parameter. If x is a random variable that has a Lognormal distribution, then $y = ln\, X$ has a normal distribution, that is, with the mean and standard deviation of the random variable $y = ln\, X$. The Lognormal distribution is converted into a standard normal distribution, which is:

$$f(t) = \frac{p(Z = \frac{\ln t - \mu}{\sigma})}{\sigma(t)} \quad (10)$$

With the cumulative function $F(t)$ in the Lognormal distribution it has the following equation:

$$F(t) = \int_0^t \frac{1}{t\sigma\sqrt{2\pi}} \exp\left\{-\frac{[\ln t - \mu]^2}{2\sigma^2}\right\} dt \quad (11)$$

Based on the equation above, the survival function of the Lognormal distribution is obtained, namely:

$$S(t) = P[Z > \frac{\ln t - \mu}{\sigma}] \quad (12)$$

And for the hazard function of the Lognormal distribution, namely:

$$h(t) = \frac{P(Z = \frac{\ln t - \mu}{\sigma})}{\sigma(t)\{P[Z > \frac{\ln t - \mu}{\sigma}]\}} \quad (13)$$

In the proportional hazard cox equation, a Lognormal model can be formed, namely:

$$\hat{h}(t) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4) \quad (14)$$

2.14. METHODS

This research was conducted within a period of approximately two months. Place of research at the Regional General Hospital Dr. Pirngadi which is located on Jalan Prof H M Yamin SH N0. 47 Medan City. This type of research is quantitative which is a case study. This study used secondary data that was previously available and compiled based on documentation (Hardani 2020). This data was obtained based on the patient's medical records on the Covid-19 case at Dr. Pirngadi, Medan City. The population of Covid-19 patients in 2021 is 1,218 people. By calculating the sample size using the slovin technique in this study. Based on the calculation, the research sample is 100 people. Based on reference scientific articles related to this research, the procedure is obtained as follows:

1. Gather an overview of the characteristics of Covid-19 patients at the Dr. Pirngadi City of Medan regarding age, gender, employment status, and other diagnose.
2. Identify events or events, regarding the presence of censored and uncensored data which can be defined as follows:

a. $\delta: 0$ is censored data, for example the patient experienced an event until he was declared dead, was forced home, or moved to another hospital.
b. $\delta: 1$ is uncensored data, for example the patient experiences an event where the condition improves or recovers.

3. Make descriptive analysis on patient data.
4. Make a proportional hazard assumption test to examine survival time t by looking at the $log(-\, log\, \hat{S}(t, X))$ curve and using the goodness of fit method.
5. Conducted the Anderson Darling Test to determine the distribution of survival time data (long stay of Covid-19 patients) at Pirngadi Hospital, Medan City.
6. Determine the hazard function and survival function in the case of Covid-19 patients.
7. Determine the model and estimate the proportional hazard cox regression parameters using Markov Chain Monte Carlo (MCMC) simulation with Gibbs Sampling.
8. Interpret the survival model so as to obtain the factors that have the most influence on the recovery rate of cases of Covid-19 patients.
9. Summing up the results that have been obtained as a whole in the analysis of the data that has been processed.

3. RESULT AND ANALYSIS

a. Characteristics of Covid-19 Patients

1. The period of time of hospitalized patients

Based Table 1, it is known that there were 100 patients with a median of 8 days, and an average length of stay of 9.66 days, which means that patients were treated for about 9 to 10 days and a standard deviation of 5.42. The shortest time is 1 day and the longest is 35 days, this can be seen in Table 1.

Table 1. Period Of Time Of Hospitalized Patients

Description	N	Min	Max	Mean	Median	Sd
Period Of Hospitalization	100	1	35	9,66	8	5,42

2. Age

Based on the description in the Table 2, it can be seen that there were 100 patients with a median of 56.5 years and an average age of 55.8 years, which means that the patients were aged around 55 to 56 years and the standard deviation was 14.9 years. The youngest is 16 years old and the oldest is 83 years old, as shown in Table 2.

Table 2. Mean Age of Patients

Description	N	Min	Max	Mean	Median	Sd
Age	100	16	83	55,8	56,5	14,9

3. Gender

Based on the diagram Figure 1, it is known that the number of female patients is more than that of male patients with a percentage of 62% or 62 female patients and 38% or 38 male patients, as shown in Figure 1.

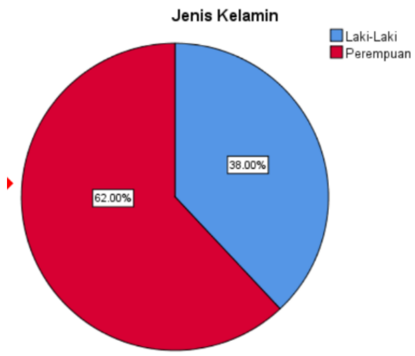

Figure 1. Comparison of The Number Of Male And Female Patients

4. Employment Status

Based on the diagram Figure 2, it is known that the number of patients who work is more than patients who do not work, with the percentage of working patients as much as 53% or 53 patients and patients who do not work as much as 47% or 47 patients, as shown in Figure 2.

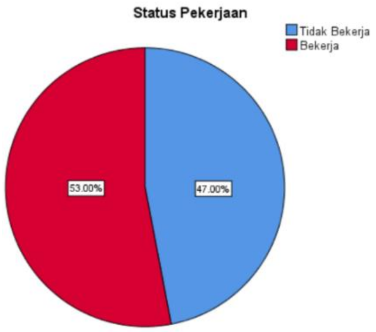

Figure 2. Comparison of Patient Employment Status

5. Other Diagnose

Based on the diagram Figure 3, it is known that the number of patients who do not have a diagnosis is more than patients who have a diagnosis, with the percentage of patients without a diagnosis of 66% or 66 patients and patients who do not work as much as 34% or 34 patients, as shown in Figure 3.

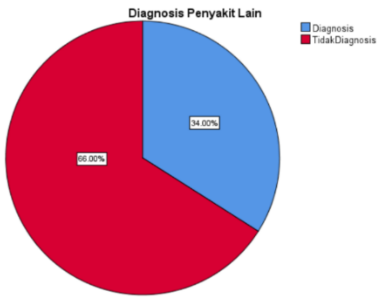

Figure 3. Percentage of Diagnosed and Undiagnosed Patients

b. Asumsi Proportional Hazard

This research uses log-log survival graphs and numerical calculations using the goodness of fit method. Statistical testing using the goodness of fit method is used to strengthen the proportional hazard assumption in addition to using graphs from log-log survival. The results of statistical tests using the goodness of fit method based on the variables of age, employment status, and diagnoses can be described in Figure 4, Figure 5, Figure 6.

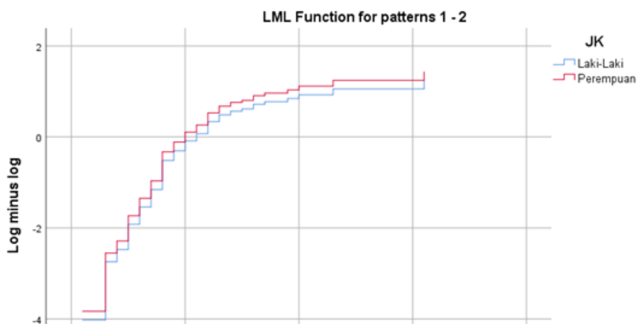

Figure 4. The Results Of Statistical Tests Using the Goodness of Fit Method Based on Variables of Age

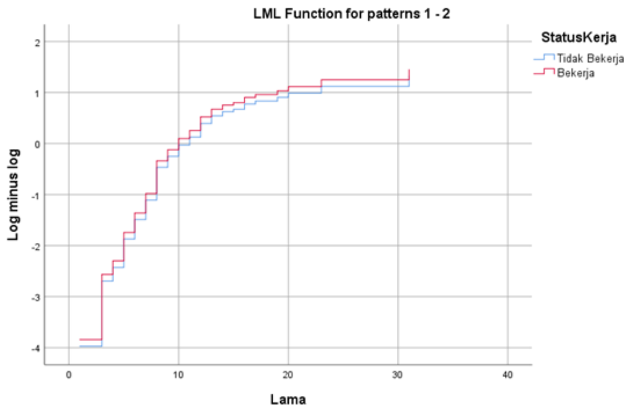

Figure 5. The Results Of Statistical Tests Using the Goodness of Fit Method Based on Variables of Employment Status

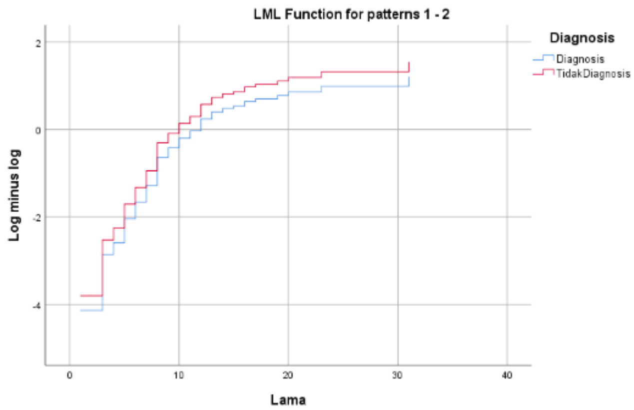

Figure 6. The Results of Statistical Tests Using the Goodness of Fit Method Based on Variables of Other Diagnose

It can be seen from the graph above that the variable gender, employment status, other diagnose variables do not intersect or align with each other, which means that the proportional hazard assumption is fulfilled.

c. Metode Goodness Of Fit

Based on the Table 4, it can be seen that for each $p - value > 0,05$ on the variable that has the hypothesis to accept H_0, it means that there is a correlation between survival time rank and Schoenfeld residual or the proportional hazard assumption with numerical calculations is fulfilled, can be described in Table 4.

Table 4. Schoenfeld Residual Processing Results

No.	Variables	Correlation	*p-Value*	Decision
1	Age	-0,0748	0,817	Accept H_0
2	Gender	0,3431	0,315	Accept H_0
3	Employment Status	0,1219	0,734	Accept H_0
4	Other Diagnose	-0,2565	0,330	Accept H_0

d. Estimation The Period of Time Of Hospitalized Patients

The lognormal distribution gets the highest rank among the others. With the smallest statistical value. So the lognormal distribution that best fits the survival time in order to get optimal results compared to the other ranks, as shown in show on Table 5.

Table 5. Computational Result Of The Lognormal Distribution

No.	Distribution	Test Statistics	Critical Value	Rank1	Conclusion
1	Exponential	12,663	2,5018	6	Reject H_0
2	Gamma	1,5135	2,5018	2	Accept H_0
3	Logistic	2,9009	2,5018	4	Reject H_0
4	Lognormal	1,2486	2,5018	1	Accept H_0
5	Normal	3,5466	2,5018	5	Reject H_0
6	Weibull	1,5733	2,5018	3	Accept H_0

e. Survival Function and Hazard Function

The survival function and hazard function from the Lognormal distribution can be generated using the variables in the Table 6.

Tabel 6. Survival Function and Hazard Function from the Lognormal Distribution

Parameters	Mean	Median	2,5 %	97,5 %
μ	2,101	2,101	1,99	2,215
θ	2,978	2,956	2,244	3,862

While survival function and hazard function from the Lognormal distribution based on the equation are obtained as follows Tabel 7:

Tabel 7. Survival Function and Hazard Function from the Lognormal Distribution

No	Survival Function	Hazard Function
1	0,76	0,38
2	0,68	0,46
⋮	⋮	⋮
35	0,31	0,50

Based on the Tabel 7, it was found that the survival function experienced a decrease in the chances of the patient's recovery rate from time to time. In accordance with the value from the first to the 35th day on the survival function which decreases every day. Whereas for the hazard function, the value from the first day to the 35th increases slowly, which means that the chance for patients to get the disease increases slowly every time. This means that the longer the patient is exposed to the disease the lower the survival rate.

f. Factors Affecting Patient Recovery Rate

The results of parameter estimation using the Bayesian approach which uses the prior distribution which is based on data information on several predictor variables using the Bayesian method with 100,000 iterations. To see the significant variables can be seen based on the credible interval, namely 2.5% and 97.5%. In the table above, which is obtained from the results of the WinBugs program, there are 2 predictor variables that do not contain a zero value, namely the age variable and the gender variable, which means that these variables are significant. Meanwhile, the variables of employment status and disease diagnosis are not significant, as shown in show on Table 8.

Tabel 8. Results of Parameter Estimation using the Bayesian approach

Parameters	Mean	Median	2,5 %	97,5 %
Age (X1)	-0,05354	-0,05349	-0,068	-0,0394
Gender (X2)	-0,9176	-0,9153	-1,45	-0,3896
Employment Status (X3)	-0,1507	-0,1528	-0,7146	0,4234

Other Diagnose	0,19999	0,1978	-0,3559	0,7816

g. Convergent Check

MC Error value is less than 1%, the convergent parameter for each parameter is less than 1% standard deviation, which means that for each parameter it is convergent, which means that the target distribution of the algorithm generated by the sample corresponds to the posterior distribution, as shown in show on Table 8.

Tabel 8. Target Distribution of Algorithm Generated by Sample

Parameters	sd	sd (1%)	MC Error	Hypothesis
b0	0,3836	0,003836	0,004647	Convergent
b1	0,271	0,00271	0,001951	Convergent
b2	0,007339	0,00007339	0,00891	Convergent
b3	0,2889	0,002889	0,001782	Convergent
b4	0,2847	0,002847	0,001806	Convergent

h. Cox Proportional Hazard Model and Hazard Ratio (HR)

For the hazard ratio of each predictor variable, it can be concluded that: The age variable means that the chance of recovery for a patient one year younger is 1.006 compared to a patient one year older. The gender variable, the chance of recovery for male patients, is 2.506 times that of the chance of recovery for female patients. Because occupational status is a variable, patients who do not have a job have a 1.16 times greater chance of recovery than working patients. With diagnostic variables for other diseases, the chance of recovery is 1.22 times greater for patients who do not have other diagnoses than for patients who do. The value of the survival function decreases monotonically, which means that the chances of the patient's recovery rate from time to time will decrease, and the value of the hazard function increases monotonically, which means that the patient's chance of getting the disease increases slowly over time. Tabel of hazard ratio of each predictor variable as shown on Tabel 9.

Tabel 9. Hazard Ratio of Each Predictor Variable

Parameters	Hazard Ratio	Median
Age (X1)	0,994	1,006
Gender (X2)	0,399	2,506
Employment Status (X3)	0,860	1,162
Other Diagnose	1,22	0,819

4. CONCLUSION

1. Based on the analysis results, the cox proportional hazard model is obtained, namely:
   $\hat{h}(t) = h_0(t)\exp\left(-0{,}5354X_1 - 0{,}9176X_2 - 0{,}1507X_3 + 0{,}1999X_4\right)$
2. Between the two variables, the variable that influences the most is the age variable (X_1) because it has a larger coefficient than the gender variable (X_2) which is equal to e^((-0.05354)).

REFERENCES

[1] Abdullah, M. N., (2021): Pendekatan Bayesian untuk Analisis Survival pada Kasus Demam Berdarah Dengue Pasien RSUD Dr. Soetomo Surabaya, Jurnal Sains dan Seni ITS, 9(2), D138–D145.

[2] Amalia (2021): Analisis gejala klinis dan peningkatan kekebalan tubuh untuk mencegah penyakit covid-19, ARTIKEL, 1(8606).

[3] Amelia, L., dan Syakurah, R. A., (2020): Analysis of public search interest towards immune system improvement during the COVID-19 pandemic using google trends, Int J Public Heal Sci, 9(4), 414–20.
[4] Aristizabal, R. J., (2012): Estimating the parameters of the three-parameter lognormal distribution.
[5] Banner (2020): The use of Bayesian priors in Ecology: The good, the bad and the not great, Methods in Ecology and Evolution, 11(8), 882–889.
[6] Candrawengi (2020): Pemodelan Survival pada Kejadian Demam Berdarah Dengue di Surabaya Timur dengan Pendekatan Bayesian, Jurnal Sains dan Seni ITS, 9(1), D8–D14.
[7] Diana, E. N., (2016), Pendekatan Metode Bayesian untuk Kajian Estimasi Parameter Distribusi Log-Normal untuk Non-Informatif Prior, PhD thesis, Institut Teknologi Sepuluh Nopember.
[8] Fernandes, A. A. R., (2016): Pemodelan Statistika Pada Analisis Realibilitas dan Survival, Universitas Brawijaya Press, Malang.
[9] Ghozali, I., (2018): Aplikasi Analisis Multivariate SPSS 25, 9th Edition, Universitas Diponegoro, Semarang.
[10] Harlan, J., (2017): Analisis Survival, Gunadarma, Depok.
[11] Hidayani, W. R., (2020): Faktor Faktor Risiko Yang Berhubungan Dengan COVID 19: Literature Review, Jurnal untuk masyarakat sehat (JUKMAS), 4(2), 120– 134.
[12] Katianda (2021): Estimasi Parameter Model Regresi Linier dengan Pendekatan Bayes, EKSPONENSIAL, 11(2), 127–132.
[13] KEMENKESRI (2020): Pedoman Pencegahan dan Pengendalian Covid-19 Revisi-5, Kementerian Kesehatan RI, Jakarta.
[14] Kleinbaum (2012): Survival Analysis, Springer Science+Business Media, LLC, New York.
[15] Laidat (2022): Estimasi Parameter Distribusi Binomial Negatif Menggunakan Metode Inferensi Bayesian, JURNAL DIFERENSIAL, 4(01), 34–43.
[16] Linn (1985): Effects of unemployment on mental and physical health., 75(5), 502–506.
[17] Marzuki (2021): Covid-19: Seribu Satu Wajah, Yayasan Kita Menulis, Medan.
[18] Ntzoufras, I., (2009): Covid-19: Seribu Satu Wajah, John Willey and Sons Inc, USA.
[19] Preatin (2007): Analisis Survival dengan pendekatan Bayesian untuk memodelkan ketahanan program KB pada individu Ibu di Indonesia tahun 2007, Surabaya: ITS.
[20] Putri, A. A., (2017): Analisis Survival Pada Laju Kesembuhan Pasien Diabetes Melitus Dengan Pendekatan Bayesian (Studi Kasus di Rumah Sakit Umum Daerah Dr. Saiful Anwar Malang), Proceeding of International Conference on Green Technology., 8(1), 268-272.
[21] Siagian, T. H., (2020): Mencari kelompok berisiko tinggi terinfeksi virus corona dengan discourse network analysis, Jurnal Kebijakan Kesehatan Indonesia: JKKI, 9(2), 98–106.
[22] Sugiyono (2016): Metode Penelitian Kuantitaif, Kualitatif dan R dan D, CV Alfabeta, Bandung.
[23] WHO (2021), : Vaccines and immunization : what is vaccination?, https://www.who.int/news-room/. Accessed Februari 20,2022. URL: https://www.who.int/news-room/