Survival Analysis On The Rate Of Diabetes Mellitus Patient Recovery With Bayesian Methode

Andiani Afifah Putri*, Suci Astutik**
***** Departement of Mathematics, Brawijaya University of Malang
** Departement of Mathematics, Brawijaya University of Malang

Article Info

Article history:

Received Jul 12th, 2017
Revised Aug 20th, 2017
Accepted Oct 26th, 2017

Keyword:

Cox Proportional Hazard
Diabetes Mellitus
Bayesian Metode

ABSTRACT

In survival analysis there is a survival model, the Cox model that used for knowing the risk factor, which influence the desease in time. Beside, Cox model also used to determine the rate of recovery or the ability of somebody to survive from the desease. The model of regression cox proportional hazard used for data where the assumption of proportional hazard has fulfiled. The estimation of regression cox model parameter, that is using Bayesian method because the result will be more maximaze than if we use the classic method. In this research, Bayesian method applied in the case of diabetes mellitus patient in the public hospital of Dr. Saiful Anwar Malang from January to December 2015, that has 174 sample. Based on Credible Interval 2.50% and 97.50% known age is the significant of independent variable. Meanwhile, the non significant of independent variable which are the gender, employment status and others diagnose.

Corresponding Author:

Suci Astutik,
Departement of Mathematics,
Brawijaya University of Malang,
Jl. Veteran No. 16 Malang, Jawa Timur, Indonesia 65145
Email: suci_sp@ub.ac.id

I. INTRODUCTION

Indonesia is one of the countries which has the high occurence of diabetes diasease includes in 10 largest countries with diabetes sufferers in the world. Diabetes itself is a non contagious diasease where there is an increase in blood sugar levels in the body and has been considered as a global diasease by the World Health Organization (WHO). According to International Diabetes Federation (IDF) in the data center and information in 2014, Ministry of Health of Republic of Indonesia, in 2014 there were 328 million people affected by diabetes diasease, this number will continue to increase each year and is expected to reach 592 million people by the year 2035. There have been a lot of studies conducted related to diabetes. The study that is likely to be used is survival analysis.

Survival analysis in statistic which used to analyze the time until an event occurs that has a goal to estimate the probability of life chance, relapse from remission, death, and the others events until the certain periode. In survival analysis there is a survival model, the Cox model that used for knowing the risk factor, which influence the desease in time. Beside, Cox model also used to determine the rate of recovery or the ability of somebody to survive from the desease. Comparison parameter estimation in the hazard model between maximum likelihood estimation and bayesian method that the result not much different from the accuracy of estimation, but, from the mean of probability, bayesian method is better than maximum likelihood estimation [1].

In this research, survival analysis with Bayesian method that having a lognormal distribution 3 parameters is expected to be able to explain the survival model and the factors that affect the rate of recovery of diabetes mellitus desease. Independent variables that used are gender, age, profession status and others diagnose.

2. RESEARCH METHOD

2.1. Survival Analysis

Survival analysis in statistic which used to analyze the time until an event occurs that has a goal to estimate the probability of life chance, relapse from remission, death, and the others events until the certain periode. In survival analysis there is a survival model, the Cox model that used for knowing the risk factor, which influence the desease in time. Beside, Cox model also used to determine the rate of recovery or the ability of somebody to survive from the desease.

Comparison parameter estimation in the hazard model between maximum likelihood estimation and bayesian method that the result not much different from the accuracy of estimation, but from the mean of probability, bayesian method is better rather than maximum likelihood estimation [1]. In this research, survival analysis with Bayesian method that having a lognormal distribution 3 parameters is expected to be able to explain the survival model and the factors that affect the rate of recovery of diabetes mellitus desease. Independent variables that used are gender, age, profession status and others diagnose. a model that used in survival analysis is a cox proportional hazard model. The cox proportional hazard equation is

$$h(t,X)=h_0(t)\exp(\beta_1X_1+\beta_2X_2+\cdots+\beta_pX_p)$$
$$=h_0(t)\exp\sum_{i=1}^{p}\beta_iX_i \quad (1)$$

where $h_0(t)$ is a baseline hazard function, β is a values of the parameters, X is a value of independent variable, and p is a number of independent variables.

Hazard ratio is defined as the hazard for one individual divided by the hazard for a different individual [2]. The definition of hazard ratio is

$$HR=\frac{h(t,\boldsymbol{X}^*)}{h(t,\boldsymbol{X})}=\frac{h_0(t,\boldsymbol{X}^*)\exp\left(\sum_{j=1}^{p}\beta_jX_{ij}^*\right)}{h_0(t,\boldsymbol{X}^*)\exp\left(\sum_{j=1}^{p}\beta_jX_{ij}\right)}=\exp\left(\sum_{i=1}^{p}\beta_i(X_i^*-X_i)\right) \quad (2)$$

With $\boldsymbol{X}^*=(X_1^*,X_2^*,\ldots,X_p^*)$ and $\boldsymbol{X}=(X_1,X_2,\ldots,X_p)$ denote the set of X's for two individuals.

2.2. Bayesian Analysis

In Bayesian, a θ parameter viewed as a random variables which has a distribution, called a prior distribution. A prior distribution used to determine the posterior distribution that form a Bayesian estimator. Bayesian model based on the posterior model that using data of the past as a prior information and observasion data as a likelihood function. A prior distribution divided into some of type there are (1) Conjugate dan non conjugate prior. (2) Informative dan non informative prior [3]. a θ prior or $f(\theta)$ will determine a posterior distribution $f(\theta|x)$ based on bayes theorem :

$$f(\boldsymbol{\theta}|\boldsymbol{x})=\frac{f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(x)}\propto f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \quad (3)$$

Markov chain monte carlo (MCMC) numeric simulation is used in the bayesian method. this simulation can approached by gibbs sampling method. the technique of gibbs sampling based on the arrangement of markov chain that has been converging on stationary distribution named posterior distribution $f(\theta|y)$.

On Bayesian method, simulation used is numeric simulation of Markov Chain Monte Carlo (MCMC). Markov Chain Monte Carlo (MCMC) can use approach one of of them is Gibbs Sampling method. Gibbs Sampling technique is based on the compilation of Markov Chain that converges on stationary distribution namely posterior distribution $f(\theta|y)$. The steps in the process Gibbs Sampling algorithm is:

1. Determine the initial value of for each parameter
$\boldsymbol{\theta}^{(0)}$is arbitrary value within the limits of the conditions of each distribution and is the amount of parameters.
2. The simulation process after the initial value is determined
3. Form $\boldsymbol{\theta}^{(r)}$ and keep it as a set of values that is raised on the iteration of the algorithm.
4. Obtain a summary of the results of the posterior distribution [4].

2.3. Data

Data used in this study was secondary data on the characteristics of hospitalized patients with diabetes mellitus disease in the Saiful Anwar General Hospital Malang started on January 5, 2015 until December 25, 2016. The variables used were: Variable response: the period of time of hospitalized patients with diabetes mellitus. Variable predictors: (1) X_1 = Gender (male = 0 and female = 1) (2) X_2 = Age (years) (3) X_3 = Employment Status

(Unemployment= 0 and Working = 1) (4) X_4 = other diagnose (There is other diagnose = 0. There is no other diagnose = 1).

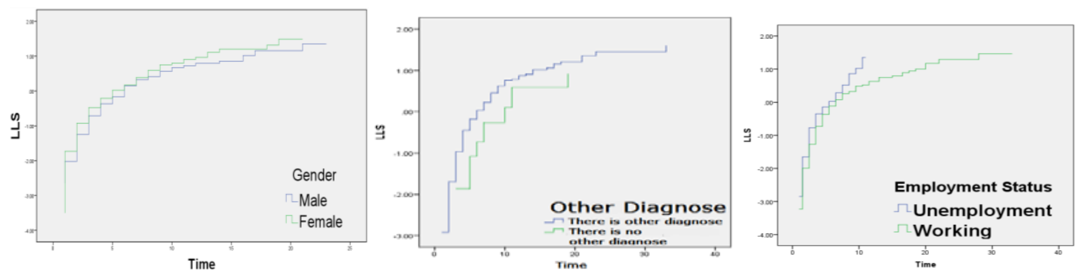
2.4. Method

Data analysis method used were as follows: (1) Collect data on inpatients with diabetes mellitus at RSSA Malang. (2) Identify events, presence of censored and uncensored data. (3) Describe the research data by using descriptive analysis. (4) Check the assumption of *proportional hazards.* (5) Describe the data distribution of *survival* time (period of time of hospitalized patients with diabetes mellitus) at RSSA Malang. (6) Determine the model and suspect of *Cox* regression parameter using *Markov Chain Monte Carlo* (MCMC) simulation with *Gibss Sampling.* (7) Determine the function and *hazard* function in patients with diabetes mellitus. (8) Determine the factors that affect the rate of recovery of patients with diabetes mellitus.

3. RESULTS AND ANALYSIS

3.1. Proportional Hazard Assumptions

This study used a curve of ln $(-\ln(S(t,x))$ to find out whether the proportional hazard assumption have been fulfilled or not. Based on Figure 1 shows that the three variables have a parallel curve, then it can be concluded that the variables have met the assumption of *proportional hazard.*



(a) (b) (c)

Figure 1. Curve of ln $(-\ln(S(t,x))$ on variables of survival time (a)Gender (b)Other diagnose, and (c)The employment status

3.2. Survival Function and Hazard Function

The function survival and hazard function of patients were formed based on the result of variables of lognormal distribution of three parameters with Bayesian method on the period of time of hospitalized patients with diabetes mellitus at RSSA Malang. Survival function and hazard function using Bayesian methods with non-conjugate priors, as many as 50,000 iterations and burn-in amounted to 1334 as shown in Table 1.

Table 1. Survival function and *hazard* function Day

	S(t)	h(t)
1	0.089	0.582
2	0.247	0.582
3	0.364	0.288
4	0.452	0.176
5	0.520	0.120
6	0.575	0.087
7	0.619	0.067
8	0.657	0.053
9	0.688	0.043
10	0.715	0.035

Based on Table 1 on the survival function S(t) shows that the length of the patient suffering from diabetes mellitus from the first day until the tenth day experience an increase means that the opportunity of the patient's recovery rate are higher. On the hazard function h (t) from the first day until the tenth day experience decline means that the opportunity of patients to suffer diabetes mellitus are lower. This shows that the longer patient suffers from diabetes then the higher patient's survival against diabetes and the opposite.

3.3. Factors Affecting the Patient's Recovery Rate

The results of the estimation of survival model parameter using non-informative priors with 60,000 iterations shows in Table 2.

Table 2. Results Estimation of Survival Analysis Parameter

Variables	Mean	2.50%	Median	97.50%
(X 1) Gender	0.267	-1.89	0.2706	2.431
(X 2) Age	-1.492	-3.864	-1.286	-0.281
(X 3) Employment Status	2.103	-0.061	2.111	4.258
(X 4) Other Diagnose	1.285	-0.547	1.292	3.098

Based on the interval of Credible Interval in Table 2. the variables which significantly affect the recovery rate of patients with diabetes mellitus is age. While the gender, employment status, and other diagnose have no significant effects.

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

Cox proportional hazards regression model got from the posterior of survival model parameters can be written as follows: $\hat{h}(t)= h0(t)\exp\,(0.26X1-1.49X2+2.10X3+1.28X4)$. Interpretation of survival models using hazard ratios can be seen in Table 3.

Table 3. Independent Variable of Hazard ratio

Variable	Hazard Ratio	$\frac{1}{Hazard\ Ratio}$
Gender	1.29	0.78
Age	0.23	4.35
Employment status	8.17	0.13
Other diagnose	3.60	0.28

The opportunity of recovery in patients with diabetes mellitus when the patient age is one year younger is 4.35 times greater than patients who aged one year older. However, there are three opportunities that provide no significant effect, namely the opportunity to recover in women patients with diabetes mellitus is smaller by 1.29 times than the chances of recover in male patients with diabetes mellitus. The chances to recover in patients with diabetes mellitus who have smaller jobs is 8.17 times the chances of a recovery in patients with diabetes melitus who have no job. The chances to recover in patients with diabetes melitus who have no other diagnose is smaller by 3.60 times the chances of recover in patients with diabetes mellitus who have other diagnose.

4. CONCLUSION

Based on the result of analysis using Bayesian methods, the factors that significantly affect the rate of recovery in patients with diabetes mellitus is age. While the gender, employment status, and other diagnose variables do not significantly affect the rate of patients recovery with diabetes mellitus.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the Ministry of Research, Technology, and Higher Education of Republic Indonesia that given us the fundamental grants fund.

REFERENCES

[1] Kneib T, Fahrmeir L. A mixed model approach for structured hazard regression. Department of Statistics, University of Munich, Munich. 2004. 386 paper 400.

[2] Kleinbaum DG, Klein M. Survival analysis. New York: Springer Science+Business Media. LLC. 2012.
[3] Box GEP, Tiao GC. Bayesian Inference in Statistical Analysis. Canada: Addison-Wesley Publishing Company, Inc.. 1973.
[4] Ntzoufras I. Bayesian Modeling Using WinBUGS. USA: John Willey and Sons, Inc. 2009.