SURVIVAL MODELING ON NON ACTIVE STUDENTS' STUDY OF UNIVERSITAS TERBUKA: A CASE STUDY

····· **Dewi Juliah RATNANINGSIH**
ORCID: https://orcid.org/0000-0003-0526-2328
Post-graduate student at Statistics Study Program
Department of Statistics, Bogor Agricultural University
Bogor, Indonesia

Dr. Asep SAEFUDDIN
ORCID: https://orcid.org/0000-0002-1694-9515
Department of Statistics, Bogor Agricultural University
Bogor, Indonesia

Dr. Anang KURNIA
ORCID: https://orcid.org/0000-0001-9409-2361
Department of Statistics, Bogor Agricultural University
Bogor, Indonesia

Dr. Wayan MANGKU
ORCID: https://orcid.org/0000-0002-7961-9812
Department of Mathematics, Bogor Agricultural University
Bogor, Indonesia

ABSTRACT

Survival analysis is a statistical method, a part of the Generalized Linear Models, used to study the survival time of an individual to an event. The application of survival analysis is indispensable in open and distance education, like Universitas Terbuka (UT), Indonesia. This analysis can determine students' durability of study opportunities within a certain time. The analysis showed that the appropriate survival model to describe the covariate that was supposed to influence the study durability of the UT students of the Non Basic Education Program was Log Logistics. Covariate that influenced the survival of study was course of study, age, marital status, employment status, educational background, number of credits taken, 1st semester's result, Grade Point Average (GPA), and the courses taken per semester.

Keywords: Survival analysis, non active students, application and case studies, open and distance education, Universitas Terbuka.

INTRODUCTION

Survival analysis is a statistical analysis regarding the survival or the length of time of an individual or a unit of life in certain circumstances. The purpose of survival analysis is to determine the relation between the incident and the explanatory variables measured at the time of the study. McCullagh and Nelder (1983) suggested that the main purpose of survival analysis is to analyze the data that is always positive in the measuring scale with a long distance interval between the initial and final data. The data type in survival analysis is the time of a certain point of time until a failure happens.

The data, the time, until the failure happensis called survival time. The survival rate has two characteristics, namely: (a) not negative and has a distribution with a long tail, and (b) several subjects may have a period of time so that the incidence of failure is not known or generally the survival time is unknown. Data that have both characteristics are called censored (Dobson, 2002). Statistical analysis methods will generally produce biased interpretation if there are censored or incomplete data (Cain et al., 2011).

There are several types of data censored in the survival analysis, the right, left and interval censors (Collet, 1997). The right censor occurs when the individuals are alive until the study ends. Left censor occurs if the observed incidence has occurred in an individual before he/she is included in a research period. Meanwhile, the interval censor is a censor in which the survival time is within a certain interval. Right censor consists of 3 types: I, II, and random. In type I right censor, the number of individuals in the beginning is determined and the research time is determined in a certain period of time. Survival time of individuals who do not experience such events cannot be determined with certainty. On the right censor type II, the number of individuals at the beginning of the study is determined, and the study time is also specified until death/failure happens to a certain amount. At random censor, however, it can occur in a clinical trial. Periods of study are determined and individuals entered at different times during those period.

The use of the survival analysis with censored data requires a certain distribution assumption. Some parametric distribution commonly used in the survival analysis is exponential, Weibull, Log Normal, Logistics, Log Logistics, and Gamma (Okello and Abdou, 2014). In survival analysis, the Cox regression model is frequently used. This is a regression model that expresses the degree of hazard in individuals with certain characteristics called covariates (Cox and Oakes, 1984). Cox regression model is included in the generalized linear models (GLM). Selection of the model and estimation of the model parameters depend on response on the variable distribution and link function used.

The application of survival analysis in education, especially in distance higher education, such as Universitas Terbuka has not been done. This study is very important in student-related policy making. Given the fact that fluctuations of the UT students' academic record vary and the high trends of UT students who are potentially non active, the survival study of the student studies is imperative. This analysis will show the magnitude of UT students' better opportunities to survive and continue their studies as well as the potentials of non active students to be. In addition, the level of risk of becoming a non active students can be seen by looking at specific characteristics (covariates).

In this paper, right censored data type I is used because the number of students involved in this study is certain and the period of academic record is determined in advance. This paper will also discuss study survival model of non active students in the Non Basic Education Program of UT by using the right type I censored data according to the response variable data distribution.

SELECTED LITERATURE REVIEW

Distance Higher Education
Distance Higher Education is a system run by distance learning and distance teaching system. In Indonesia, distance learning system is an option considering the geographical conditions of the country, where locations are separated by oceans. UT helps communities in the archipelago get access to higher education.

Characteristics of distance higher education students may cause problems because students are required to cope many aspects of their life, such as family, work, leisure and study time, asides from learning (Schuemer, 1993). Factors that influence the pattern of students' continuity in their studies depend on the examination results of the previous semester and their study programs (Soeleiman, 1991). Ratnaningsih et al. (2008) and Rahayu (2009) suggested the there are factors that lead to non active status. Non active students are those who do not register whentheir academic leave (four consecutive semesters) period expires. Factors cover age, sex, address, employment status, marital status, scholarship, formal education background, field of study background, Performance Index in 1st semester, credits taken in the 1st. semester, Performance Index in the 2nd. semester, credits taken in the 2nd semester, GPA, current credits, the chosen course of study, and the learning units of the open and distance education of UT (ODLU-UT).

Klapproth and Schaltz (2014) suggested that gender, GPA, socioeconomic, and demographic affect student retention. Learning experience and previous academic achievement are decisive factors in university students' study continuity (McCormick and Lucas, 2014). In addition, Thomas et al. (2014) suggested that educational background affects online learning both in informal and formal education; while Islam (2010) and Kadarko (2000) stated that learning readiness and strategies will determine the success of their studies in UT learning programs.

Estimation of Data Distribution

Estimation of data distribution on survival analysis with parametric approach was determined by response variable data distribution. The test often used to determine the most appropriate distribution of data was the Anderson-Darling test (AD) which is designed to detect a mismatch flatness of the curve in many distributions. The AD test is stronger than the Chi - Square test and Kolmogorov – Smirnov (Law and Kelton, 2000).

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln F(X_i) + \ln(1 - F(X_{n+1-i}))] \qquad (1)$$

The value of the A^2 statistic test was compared with the critical value of 1.92 for α by 10%. The smaller Anderson-Darling statistic value indicatesa better data distribution.

Survival Analysis

Survival analysis is a statistical method used to to analyze data of a specific time until the occurrence of an event (Lee, 1992). An event can be death, healing, relapse, failure or other events determined by investigators. Three important elements in determining the survival time (Y) were initial time (time of origin), a clear definition of failure time, and time scale as unit of measurement. The difference between survival analysis and other statistical analysis is the presence of censored data (Gijbels, 2010). The causes of censored data are, among others (Dobson, 2002):

      (1) Loss to follow-up, which occurs when the object moves, dies or refuses to participate.

      (2) Drop out, which occurs when the treatment is stopped for some reasons.

      (3) Termination, which occurs at the end of the study when the objects have not reached the failure event.

The distribution of survival time usually has three functions, namely, probability density function, survivor function and hazard function. Probability density function from y survival time is defined as the limit of the possibility of failure in an interval of y and $(y + \delta y)$ or a chance that a person fails in a short time interval of y to $(y + \delta y)$. That function is formulated as follows.

$$f(y) = \lim_{(y+\delta y)\to 0} \frac{P(individual\ dies\ in\ interval\ (y, y+\delta y))}{\delta y} \qquad (2)$$

Survival function $S(y)$ is defined as an object's probability to survive after y time. The survival function for continuous random variables is stated in the following formulation:

$$S(y) = P(Y \geq y) = 1 - F(y) \qquad (3)$$

Function $S(y)$ is a non increasing function with character $S(y)$ = 1 for y = 0 and $S(y)$ = 0 for y = ∞. Hazard function is a momentary failure rate at short intervals $(y, (y+\delta y))$. The hazard function for continuous random variables is formulated as follows:

$$h(y) = f(y) \times \frac{1}{S(y)} \qquad (4)$$

Cox Proportional Hazard Regression Model

The Cox regression model is a regression model that shows the degree of hazard of individuals with certain characteristics, called covariates (Cox and Oakes, 1984). Jones and Branton (2005) used Cox regression to determine policy settings or dissemination of studies using hazard ratios.

The hazard function is a function of the basic hazard which depends on the time and the function that are dependent on the covariates as as follows:

$$h(y, X) = h_0(y)G(X; \beta) \qquad (5)$$

Because $h(y, X)$ and $h_o(y)$ are positive, $G(X,\beta)$ is also positive. Cox and Oakes (1984) chose $G(X; \beta)$ = exp($\beta^T X$) so that the model becomes:

$$h(y, X) = h_0(y)e^{(\beta^T X)} \qquad (6)$$

If the above equation is multiplied by the log, it will generate the following form.

$$\log h(y, X) = \log h_0(y) + \beta^T X \qquad (7)$$

This form is a linear model with a link log function.

Model Selection Criteria

The criteria in selecting the best common model in survival analysis is the maximum value of the Log Likelihood. Lee (2014) suggested that the goodness of fit of other models that can be used with SAS software are AIC (Akaike Information Criterion) and SBC (Schwarz's Bayesian Criterion). AIC and SBC are functions of the n number of observations, the sum of squared errors (SSE), and a number of explanatory variables where k is included in the intercept. AIC values are formulated as follows:

$$AIC = n \ln\left[\frac{SSE}{n}\right] + 2k \qquad (8)$$

Meanwhile, the SBC values are formulated as follows.

$$SBC = n \ln\left[\frac{SSE}{n}\right] + k \ln n \qquad (9)$$

The selection of the best model is done by examining the three values. A model is said to be good if the model has a maximum value of log likelihood and a minimum value of AIC and SBC.

RESEACRH METHOD

Data for this study were taken from the UT's Non Basic Education Program students's record for 18 semesters (2005.1 to 2013.2). The number of data analyzed was as many as 4890 students in 23 courses.

Response variables were a long survival sudy time of the students measured in semester. Independent variables in this study were factors which are suspected to have affected their survival, which cover:

> The students' addresses which was categorized into two, namely, district (0) and city (1).
> The locations where students enrolled. ODLU-UT locations as many as 37 cities, ranging from Banda Aceh to Jayapura and was coded in accordance with the applicable provisions at UT.
> The study programs offered by UT. These programs were coded according to applicable regulations at UT.
> Prior education before enrolling at UT was categorized into 6 categories, namely, Senior High School, Non Diploma Program 1 (D1), Non Diploma Program 2 (D2), Non Diploma Program 3 (D3), Bachelor (S1), and Masters (S2).
> Age of students at the time of admission. These variables were grouped into: age <35 years (1), age between 35 and 45 years (2), and age > 45 years (3).
> Department of origin; student's home department before enrolling at UT. Grouping method was based on education major code origin where the students were registered on the end of the catalog UT 2014 (Anonymous, 2014).
> Marital status was grouped into married and single.
> Status of work was grouped into working and not working.
> The number of credits (NC) taken during study period at UT. These variables were grouped into 4, namely: NC≤38 (1), 39 <NC <84 (2), 85≤NC≤116 (3), and≥117 NC (4).
> The average number of registered courses per semester. This variablesweregrouped into 3, namely: (1) NRC <5, (2) NRC ≤ 5 ≤ 8, and (3) NRC> 8
> The performance indexesin 1st. semester (PI-1) were grouped into 4, namely: (1) PI-1 ≤ 1.00, (2) 1.00 <PI-1 ≤ 2.00, (3) 2.00 < PI-1 ≤ 3.00, and (4) PI-1> 3.00.
> Grade Point Average (GPA) were grouped into 4, namely: (1) ≤ 1.00 GPA, (2) 1.00 <GPA ≤ 2.00, (3) 2.00 <≤ 3.00 GPA and (4) GPA> 3.00.

Meanwhile, the censored data were coded as "variable censor" and categorized into uncensored (1) and censored (0). Data of durability of the students is called censored if the students do not change his/her active status into non active at the time of observation. In other words, students whose status were active until the end of the observation period or have completed their studies are considered censored. However, if during the period of observation, the students become of non active, they were called complete observation (uncensored). This was possible because its durability during this period is clear.

Phases of data analysis performed in this study are:

> Creating statistical description of the studied variables.
> Examining the distribution of data in the form of a survival variable response.
> Model selection by using the Log Likelihood values, AIC and SBC.
> Estimating parameters and testing the significance of the model based on Maximum Likelihood method.
> Developing a model based on the hazard function and determining the survival rate of a student's study based on the hazard function and determining the odds ratios of the variables that get into the model.

Software used for the analysis was the SAS 9.4 and Minitab 16. Minitab software was used to examine the distribution variable response of the data, while SAS is used to analyze other analyses; and to determine model criteria, SAS is used.

RESULT AND DISCUSSION

Descriptive Data

Descriptive data analysis was conducted based on demographic characteristics and the students' academic programs. The demographic characteristics of students based on the censored data status are presented in Table 1 which shows the characteristics of students. In general, students of 2005.1 to 2013.2 registration periods lived in rural areas/districts (62.07%); are male (69.63%), age between 35 to 45 years (35.24%), married (61.31%), and have a working status (88.88%).

Table 2 shows formal education background of the students prior to admissions (62.82% high school); their academics performances were: (1) 38.28% have collected between 39 and 84 credits, (2) 59.53% got their 1st. semester Performance Index between 1.00 and 2.00, (3) 58.20% had GPA between 1.00 and 2.00, and (4) 63.95% took less than 5 subjects per semester.

Table 1. Distribution of non basic education students' registration period 2005.1-2013.2 based on demographic characteristics

| Variables observed | Categories | Status of Censored | | | | Total |
| | | Censored[1] | | Not Censored[2] | | |
		N	%	N	%	
Place of residence	Villages	1252	41.25	1783	58.75	3035
	Cities	542	29.22	1313	70.78	1855
Sex	Female	587	39.53	898	60.47	1485
	Male	1207	35.45	2198	64.55	3405
Age	< 35 years	327	19.77	1327	80.23	1654
	35 - 45 years	522	30.30	1201	69.70	1723
	> 45 years	945	62.46	568	37.54	1513
MaritalStatus	Married	1388	46.30	1610	53.70	2998
	Not married	406	21.46	1486	78.54	1892
Work status	Working	1617	37.21	2729	62.79	4346
	Not working	177	32.54	367	67.46	544

Explanation:
1) Getting censored means that students have active status during observation or have completed their studies at the time of the research.
2) Not censored meaning that students who have non active status during observation period.

Table 2. Distribution of IOU's non basic education students, registration period 2005.1-2013.2 based on academic characteristics

| Variables observed | Categories | Status of Censor | | | | Total |
| | | Censored | | Not censored | | |
		n	%	n	%	
Formal Education Level	Senior High School	600	19.53	2472	80.47	3072
	D-I	27	40.91	39	59.09	66
	D-II	31	55.36	25	44.64	56
	D-III	1119	68.61	512	31.39	1631
	Bachelor	17	26.15	48	73.85	65

Variables observed	Categories	Status of Censor				Total
		Censored		Not censored		
		n	%	n	%	
Number of Credits taken (System Credit Semester, SCS)	SCS ≤38	23	1.43	1582	98.57	1605
	39<SCS<84	1061	56.68	811	43.32	1872
	85≤SCS≤116	68	16.15	353	83.85	421
	≥117 SCS	642	64.72	350	35.28	992
Performance Index of 1st. Semester	PI-1 ≤ 1.00	93	8.63	985	91.37	1078
	1.00 <PI-1 ≤ 2.00	978	33.60	1933	66.40	2911
	2.00 <PI-1 ≤ 3.00	688	79.81	174	20.19	862
	PI-1 > 3.00	35	89.74	4	10.26	39
Grade Point Average (GPA)	GPA ≤ 1,00	3	1.00	296	99.00	299
	1,00 <GPA ≤ 2,00	302	10.61	2544	89.39	2846
	2,00 < GPA ≤ 3,00	1408	85.23	244	14.77	1652
	GPA> 3,00	81	87.10	12	12.90	93
Number of courses taken	NRC< 5	965	30.86	2162	69.14	3127
	5 ≤ NRC ≤ 8	640	43.36	836	56.64	1476
	NRC> 8	189	65.85	98	34.15	287

Table 1 and Table 2 show each explanatory variable that was examined, students with censorred data had a high percentage condition compared to the censored one. This shows that many UT's students in Non Basic Education Program, on registration period 2005.1 to 2013.2, were non active. This means many students did not continue their studies in the next semester for four consecutive semesters. The number of students who were non active reaches 63.30 %, while 30.60 % and 6.10% are active and completed their studies respectively.

Data Distribution Examination of Response Variable

Examination of data distribution of survival time on the response variable Y was done with Minitab software using Anderson-Darling (AD) test. The following explains a suitability test value (goodness of fit) Anderson-Darling test against multiple distribution was tested. In the process of initial data exploration, the criteria used for distribution examination were the smaller the value of AD, the closer the distribution to the actual data distribution. Based on the statistical AD value, from the seven tested distribution, the smallest AD value distribution were Weibull, Gamma, and Log Logistics. Thus, the appropriate distribution for data response of this study. In this study, the three distributions were tested to find the best model that best described the survival analysis study of the UT's Non Basic Education students.

Table 3. The test result of the appropriateness of data distribution examination of respon variable

Distribution	Score-AD	Score-P
Normal	144.398	<0.005
Log Normal	87.108	<0.005
Exponential	174.875	<0.003
Weibull	53.032	<0.010
Gamma	56.397	<0.005
Logistic	119.270	<0.005
Log Logistic	81.763	<0.005

Survival Model Selection

Before analyzing the best survival models, the survival data plot was performed in order to see data description. Survival plot in this study was done against the time, ie. the survival time data indicator of data censored (0 means censored; 1 means not censored), and a treatment variable group of students was grouped based on or work status, namely: working (1) and not working (0). The following is a plot of survival of the students.
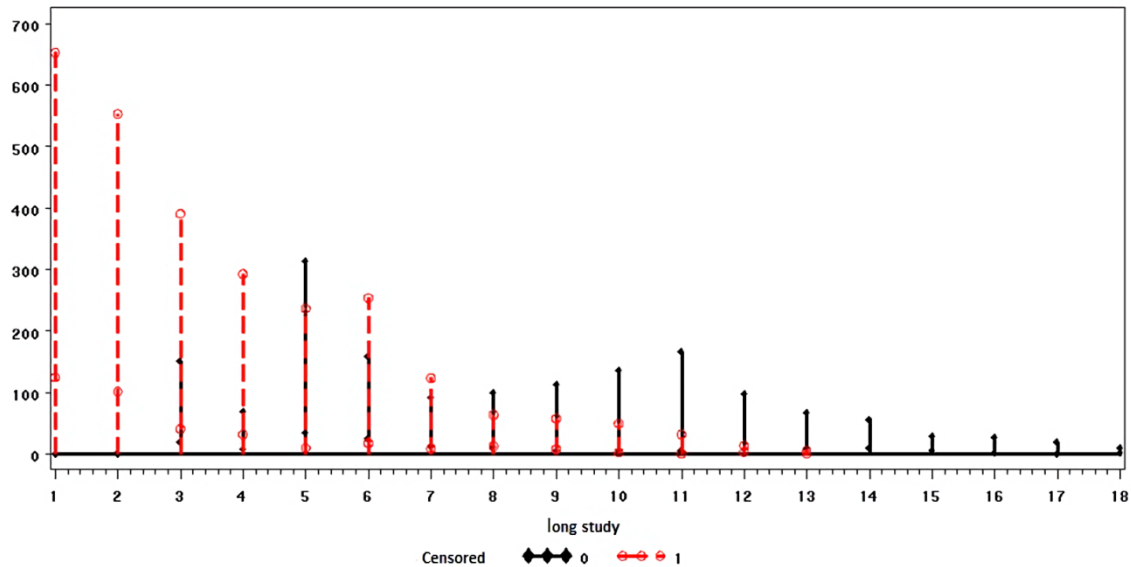

Figure 1. Plot survival data of UT's non basic education program student of 2005.1-2013.2.

To determine the best survival models, SAS macro program was made to facilitate the calculation of the AIC and SBC values for seven possible models. Results of the analysis are presented in Table 4 which shows that the best survival model to describe study survival of UT student was Log-Logistic models. This can be seen from the maximum value of the Log-Likelihood, the smallest value of AIC and SBC. Survival analysis with Log-Logistic models gave a better value for the full model involving all covariates and that involving only the significant covariate. Based on the Wald test at alpha level of 5%, there were two not significant covariates, namely gender and place of residence. Each Pr> Chi-Square is 0.9071 and 0.3365.

Meanwhile, other models that fit the tests indicate that the chosen model for survival (Log Logistics) is the best is reinforced by statistical tests presented in Table 5 which shows that out of the three test, the survival models of Log Logistics is the best model, proved from the probability values of the tests were less than 0.05 (highly significant).

Table 4. Survival model criteria on data of non basic education students' registration period 2005. 1 up to 2013.2

Model	Full Model (all covariates)			Not full Model (significant covariates)		
	Log-Likelihood	AIC	SBC	Log-Likelihood	AIC	SBC
Weibull	-3238.55	6503.10	6587.53	-3238.57	6499.15	6570.59
Log-Logistic	-3064.37	6154.75	6239.18	-3064.83	6151.67	6223.11
Gamma	-3127.16	6282.32	6373.25	-3127.37	6278.74	6356.68

Table 5. Test of equality over strata

Test	Chi-Square	DF	Pr>Chi-Square
Log-Rank	170.076	1	<.0001
Wilcoxon	300.581	1	<.0001
-2Log(LR)	152.455	1	<.0001

Model Parameter of Prediction and Testing Parameter Model

Table 6 shows model parameter estimation based on maximum likelihood in any significant covariates.

Table 6. Maximum Likelihood of estimation results analysis against covariate model parameter of the Log-Logistics survival model covariates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard ratio
Intercept	1	-13.913	0.0689	407.69	<.0001	
Study Program (SP)	1	0.0027	0.0008	10.22	0.0014	1.003
Age	1	0.0059	0.0013	19.48	<.0001	1.006
Marital Status	1	0.0477	0.0205	5.41	0.0200	1.049
Working Status	1	-0.0545	0.0249	4.80	0.0284	0.947
Education	1	0.0452	0.0077	34.18	<.0001	1.046
System Credit Semester	1	0.0067	0.0003	578.39	<.0001	1.007
GPA	1	0.9460	0.0367	663.17	<.0001	2.575
1st Semester Performance Index	1	-0.3813	0.0300	161.56	<.0001	0.683
Courses	1	0.2270	0.0059	1463.06	<.0001	1.255
Scale (σ)	1	0.2789	0.0042			
Location (μ)	1	21.541	0.0104			

Explanation: $\gamma = \exp(\mu) = 8.6204$ and $\beta = \sigma^{-1} = 3.5973$

Study Survival Model of the UT Students and Study Survival Rate

Based on the covariates that significantly affect the durability of the students' study period (Table 6), the durability model of UT student using Cox regression is formulated as follows:

$h(y,X) = h_0(y)exp(0.0027$ SP $+ 0.0056$ age$+ 0.0477$ marital status -0.0545 working status $+ 0.0452$ education $+ 0.0067$ System Credit Semester $+ 0.9460$ GPA $- 0.3813$ Performance Index (1) $+ 0.2270$ Courses)

Meanwhile, from Table 6, the parameter scale (σ) and location (μ) from the Log Logistic model, was 8.6204 and 3.5973 respectively and the hazard model function is:

$$h(y_i) = \frac{\left(\frac{\beta}{\gamma}\right)\left(\frac{y_i}{\gamma}\right)^{\beta-1}}{1+\left(\frac{y_i}{\gamma}\right)^{\beta}} = \frac{\left(\frac{3.5973}{8.6204}\right)\left(\frac{y_i}{8.6204}\right)^{3.5973-1}}{1+\left(\frac{y_i}{8.6204}\right)^{3.5973}} = \frac{0.4173\left(\frac{y_i}{8.6204}\right)^{2.5973}}{1+\left(\frac{y_i}{8.6204}\right)^{3.5973}}$$

The hazard function shows students' probability to become non active at the time of y_i. The survival distribution fuction and hazard function of the survival model of UT students' study survival models are shown in Figure 2.
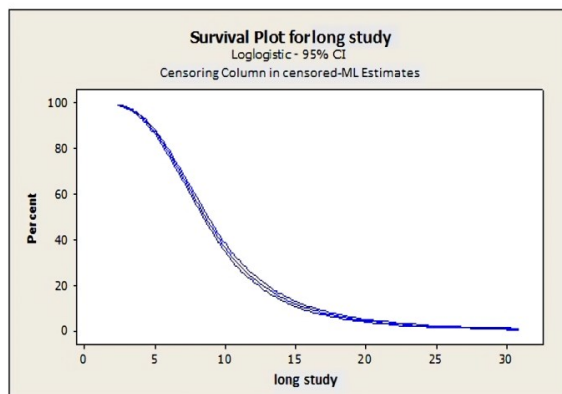

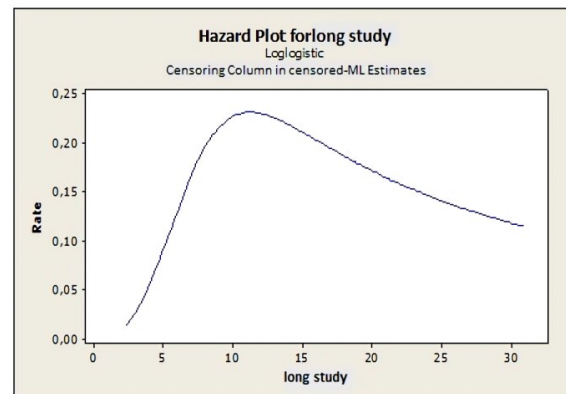**Figure 2a**   **Figure 2b**

Figure 2. Plot survival function and hazard function of study survival of the IOU's non basic education students

Figure 2 shows a survival function plot and the hazard function to describe study survival of UT's Non Basic Education Program students. Figure 2a shows that the chance of the students to survive tends to decrease, meaning there were many students whose status become non active. Figure 2a shows a sharp decline in the fifth semester. Students whose statuses are not active in the fifth semester means that they stopped registering from their second semester. Their low Performance Index in the first semester may cause their decision to not register in the next semester. The previous semester's academic result affected students' resistance (Soeleiman, 1991; Ratnaningsih et al., 2008; McCormick and Lucas, 2014; Klapproth and Schaltz, 2014).

Figure 2b shows a plot of the function of survival hazard ofUT student studies with the Log Logistics model. This figure shows the possibility of failure in study, in this case, changing the status into non active. From the figure it shows that the most possible time for the students to be non active was at the tenth semester of the fifth year. The chances are nearly 0.23; meaning that almost 23% of the total students in the tenth semester will be non active, but then it gradually declined in the next semester.

Odds Ratio of the Survival Rate of the Student Studies
Odds ratio is a measurement used to determine risk level. It compares ratio between the incidences of individuals with the condition of covariates on successful categories with failed categories. Odds ratio in each covariate is examined andthe result is used to support the analysis of study survival in Table 6.

Table 6 (column 7) shows that the risk of becoming non active for UT students of Non Basic Education program was based on significant explanatory variables that can be explained as follows. The highest risk is explained by GPA variables. Students having low GPA have more risks to be non active than 2575 students who have high GPA. The highest risk of becoming non active is the variable of courses taken per semester. Students who took many courses are at risk of becoming non active by as much as 1.255 times compared to those who did not. It is very possible because taking courses with excessive regulation is more difficult for students who manage their studies and exams. This fact is supported by Bean (1982) which suggests that the cessation of re-registered students in higher education institutions is determined by several variables, including: GPA, courses taken, and the main work of students.

Similarly, variables of being married and educational background contributed to high risk. As many as 1.049 and 1.046 Students with married status tended to be at risk of being non active than those who were not by 1.049 times. It is very possible because generally students who are married have difficulty in managing their time between studying and family. This fact is supported by the results of the research by Kadarko (2000) and Islam (2010), which suggests that in general, UT students who are married and working will have difficulty in managing their time, as well as readiness and very low self-learning habits. Similarly, educational background has a very important role in the learning process at UT-based online and face to face. The risk of students having educational background which were not in accordance with their chosen field of study is 1.046 times compared to those with linear educational background. This is stated by Thomas et al. (2014) who said that the student's educational background plays an important role in online learning.

CONCLUSION

Survival analysis can be utilized in education institution like UT in the effort of determining durability of student's study probability. This is needed in policy-making in relation to the student's participation in higher education. Long study distributions (response variable) of Non Basic Education Program of UT students at the registration period 2005.1 to 2013.2 were suitable for Gamma, Weibull, and Log Logistics. The data used is the censored right type I. The analysis showed that the best survival model in explaining the survival study is a student survival model by Log Logistic distribution. Covariates that influenced survival of UT student studies of Non Basic EducationProgram were: chosen study programs, age, marital status, employment status, educational background, number of credits taken, Performance Index acquired in the 1st half, GPA, and the courses taken per semester.

BIODATA and CONTACT ADDRESSES of AUTHORS

Dewi Juliah RATNANINGSIH is an Associate Professor of Faculty of Mathematics and Natural Sciences at Universitas Terbuka (UT). She is currently the student of doctoral program at Bogor Agricultural University in statistics field. Her academic interest areas are statistics, applied statistics in education, studies of services, customer satisfaction, tracer study, developing research instrument, open and distance learning and needs assessment for distance education. She has three articles that have been published in international indexes and scopus indexed (Asian Association of Open Universities Journal, Open Praxis, and Journal of Applied Statistics). She was awarded First Prize in Excellent Paper Winner of the 25th International Council for Open and Distance Education (ICDE) World Conference and ICDE Prizes for Innovation and Best Practice (2013). She has published five books on statistics.

Dewi Juliah RATNANINGSIH
Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Terbuka (UT)
Address: Universitas Terbuka, Jalan Cabe Raya, Pondok Cabe, Pamulang, Tangerang Selatan, 15418, Banten, Indonesia
Phone: +62-21-7490941 ext.1808, +62 818884292,
E-mail: djuli@ecampus.ut.ac.id, dewijuliahr@gmail.com

Asep SAEFUDDIN is Professor of Departemen of Statistics at Bogor Agricultural University. He obtained a master's and doctorate degrees in Biostatistics and Populations Medicine/ Quantitative Epidemiology from the University of Guelph. He served as Vice Rector for Planning, Development and Cooperation, Bogor Agricultural University in 2003-2008. Year 2013-2017 as Rector of Trilogi University. Currently, he served as Rector at University of Al Azhar Indonesia as well as President of Pusat Bahasa Mandiri University of Al Azhar Indonesia. In addition, he is also active in several national organizations, including Vice President of Dewan Pertimbangan Forum Rektor Indonesia, Vice President of ISSOSS (Islamic Country Society of Statistical Science) (2009-2014), and Chairman of the Advisory Committee of Yayasan Inovasi Teknologi (2007-present). His research interest include: geoinformatics, spatial analysis, and statistical modelling. He has published any journal and conference papers and also coordinated some researches project in statistics.

Asep SAEFUDDIN
Department of Statistics, Faculty of Mathematics and Natural Sciences
Bogor Agricultural University
Jalan Meranti, Kampus IPB Dramaga, Bogor, Indonesia
Phone: +62-251-8624535
E-mail: asaefuddin@gmail.com

Anang KURNIA is an Assosiate Professor of Departemen of Statistics, Faculty of Mathematics and Natural Sciences at Bogor Agricultural University. His research interests include: small area estimation, generalized linear (mixed) model, robust statistics, and high-dimensional data analysis. He is head of Department of Statistics in Bogor Agricultural University and former head of Indonesian Statistics Higher Education Association. He has published any journal and conference papers and also coordinated some researches project in statistics.

Anang KURNIA
Department of Statistics, Faculty of Mathematics and Natural Sciences
Bogor Agricultural University
Jalan Meranti, Kampus IPB Dramaga, Bogor, Indonesia
Phone: +62-251-8624535
E-mail: anangk@apps.ipb.ac.id

Wayan MANGKU is Professor of Departemen of Mathematics, Faculty of Mathematics and Natural Sciences at Bogor Agricultural University. He completed his Master's degree in Mathematical Modeling at School of Mathematics and Computer Science, Curtin University of Technology, Western Australia. He received his Ph.D degree in Mathematics at University of Amsterdam, The Netherlands. He is the head of Mathematical Modeling Division at Department of Mathematics, Bogor Agricultural University. His research interest include: stochastic process and mathematical statistics. He has published two books and more than 90 articles in scientific journals.

Wayan MANGKU
Department of Mathematics, Faculty of Mathematics and Natural Sciences
Bogor Agricultural University
Jalan Meranti, Kampus IPB Dramaga, Bogor, Indonesia
Phone: +62-251-8625276
E-mail: wayanma@apps.ipb.ac.id, wayan.mangku@gmail.com

REFERENCES

Bean, J.P. (1982). Student attrition, intentions, and confidence: Interaction effect in the a path model. Research in Higher Education, 17, 291-320.

Cain, K.C., Harlow, S.D., Little, R.J., Nan, B., Yosef, M., Taffe, J.R. & Elliot, M.R. (2011). Bias due to left trucation and left censoring in longitudinal studies of developmental and disease processes. American Journal of Epidemiology,173(9), 1078-1084.

Collet, D. (1997). Modelling survival data in medical research. Second Edition. London: Chapman & Hall.

Cox, D.R. and Oakes, D. (1984). Analysis of survival data. London: Chapman & Hall.

Dobson, A.J. (2002). An introduction to generalized linear models. Second Edition. London: Chapman & Hall.

Gijbels, I. (2010). Censored data. Wiley Interdisciplinary Reviews: Computational Statistics, 2(2), 178-188.

Islam, S. (2010). Readiness for self learning of Universitas Terbuka and high school students in open and distance learning higher education system in Indonesia. Journal of Open and Distance Education, 11(1), 1-14.

Jones, B.S. and Branton, R.P. (2005). Beyond logit and probit: Cox duration models of single, repeating, and competing events for state policy adoption. State Politics and Policy Quarterly, 5(2), 420-443.

Kadarko, W. (2000). Understanding students' learning styles and strategies. Journal of Open Distance Education, 3(2), 1-15.

Klapproth, F. and Schaltz, P. (2014). Who is retained in school, and when? Survival analysis of predictors of grade retention in Luxembourgish secondary school. European Journal of Psychology of Education A Journal of Education and Development, 29(3), 119-136.

Law, A.M. and Kelton, W.D. (2000). Simulation modeling and analysis (3rd ed). New York: McGraw-Hill.

Lee, E.T. (1992). Statistical methods for survival data analysis. New York: John Wiley & Sons Inc.

Lee, M.C. (2014). Business bankruptcy prediction based on survival analysis approach. International Journal of Computer Science & Information Tecnology, 6(2), 103-119.

McCullagh, P. and Nelder, J.A. (1983). Generalized linear models. Second Edition. London: Chapman & Hall.

McCormick, N.J. and Lucas, M.S. (2014). Student retention and success: Faculty initiatives at Middle Tennessee State University. Journal of Student Success and Retension,1(1), 1-12.

Okello, J.O. and Abdou, D.K. (2014). Parametric model and future event prediction base on right cencored data. American Journal of Mathematics and Statistics, 4(5), 205-213.

Rahayu, D.P. (2009). An analysis of characteristics of OU's non active students with cluster Encamble approach. Unpublished thesis, Postgraduate Program. Bogor: Bogor Agricultural University.

Ratnaningsih, D.J., Saefuddin, A. & Wijayanto, H. (2008). An analysis of drop-out students' survival in distance higher education. Journal of Open Distance Education, 9(2), 101–110.

Soeleiman, N. (1991). Continuity of registration and its relation to the examination results. Research Report. Jakarta: Indonesia Open University.

Schuemer, R. (1993). Some psychological aspects of distance education. Hagen. Germany: Institute for Research into Distance Education. (ED 357 266).

Thomas, L., Herbert, J. & Teras, M. (2014). A sense of belonging to enhance participation, success and retention in online programs. The International Journal of the First Year in Higher Education, 5(2), 69- 80.

Universitas Terbuka (2014). Universitas Terbuka Catalogue 2014 (3rd ed). Ministry of Education and Culture, South Tangerang: Universitas Terbuka.