MODELING OF ON TIME GRADUATION DURATION ON GRADUATE SCHOOL STUDENTS OF BOGOR AGRICULTURAL UNIVERSITY

Irmayani, I Made Sumertajaya,Indahwati

ABSTRACT : The cox proportional hazard model is a semiparametric distributed model because in the Cox model does not require information about the underlying distribution of survival time. In cox proportional hazard model, there is an assumption that must be fulfilled, that is proportional, if the assumption is not fulfilled hence model that formed is the nonproportional hazard. One method that can be used to address non-proportional hazard cases is the extended Cox Proportional hazard regression model. The extended Cox Proportional hazard model is a modification of the cox proportional hazard model if the proportional hazard assumption is not met. Estimation method used in the model is maximum partial likelihood estimation (MPLE). Data used in this research is the data of graduate students of Graduate School of Bogor Agricultural University in academic year 2011 to 2013 which amounted to 3443 data for 4 years and 10 independent variables of sex, marital status, age, working status, education funding source, accreditation, GPA in undergraduate, acceptance status at Bogor Agricultural University (IPB), University origin, and faculty. Based on the selection of the best cox proportional hazard model, there are four significant variables in the model: age, education funding source, University origin, and faculty. Then testing of proportional hazard assumptions is conducted and retrieved one of the variables which do not satisfy the assumption of proportional hazard i.e. faculty. Further, the faculty variable which does not satisfy the assumption interacts to the time function of g (t) = t. By using the calculation of the AIC (Akaike Information Criteria), it is proved that the Extended Cox Proportional hazard model is a suitable model for data on the right postgraduate students of Bogor Agricultural University (IPB). Furthermore, the results of the calculation by SAS 9.3 software, Extended Cox Proportional hazard model has significant influence of independent variables i.e. Marital status, education funding source, and university origin.

Keywords:Cox Proportional Hazard, Non Proportional Hazard, and Extended Cox Proportional Hazard

————————————◆◆————————————

1. INTRODUCTION

Bogor Agricultural University (IPB) is one of the public universities (PTN) in Indonesia with the vision of producing the best graduates of which it include a graduate school (SPs) of IPB that established in 1975. During the 41 years since it established, SPs-IPB is Experiencing rapid advance. Initially, SPs-IPB has only consisted of 7 courses and now it expanded to 67 study programs/majors of master degree courses and 43 study programs/majors of doctoral course (IPB 2014).

An indicator or benchmark of the quality of a university is the on time graduation duration of students. A student of master's degree program normally takes four semesters to complete his studies. In the fact, the student is not always able to complete their studies in normal time. It caused by several factors both external or internal.

The secondary data is obtained from the observation of students' on time graduation duration of the Graduate School of Bogor Agricultural University (SPs-IPB) is an example of survival data. Data about the length of time from initial observations to the occurrence of an event called the survival data. Generally, survival data are not complete it means the survival time can not be exactly known because of the limited research time and others. This causes the distribution of survival is not normal but positively skewed, so it is needed a method

————————————————————

- *Irmayani is currently pursuing masters degree program in applied statistics in Bogor Agricultural University, Indonesia,*
  E-mail: Irmayanistt@gmail,com
- *I Made Sumertajaya is Lecturer, Departement of Statistics, Bogor Agricultural University, Indonesia, E-mail: imsjaya @yahoo.com*
- *Indrawati is Lecturer, Departement of Statistics, Bogor Agricultural University, Indonesia, E-mail: indah.stk @gmail.com*

that can facilitate the unnormality of survival data i.e. survival analysis.

Survival analysis is a statistical procedure for analyzing the data that the independent variable is the duration of time until the occurrence of an event (Kleinbaum 2012). The observed event such as death due to a disease, accident, duration time of getting the job after a long time unemployed, duration of on time graduation of students, the durability of the students from dropping out of College, etc. One of the distinguishing points of survival analysis is the concept of censorship. The data is categorized as censored data if the data is not completely observed due to the loss of the research subject or another reason, because of ordinary linear regression analysis can not handle this problem..

One of the objectives of survival analysis is to find out the relationship between survival time and independent variables that are suspected to affect survival time. The relationship can be modeled with the Cox regression model. Cox regression models commonly called Cox proportional hazard regression (cox PH), because the cox regression must meet proportional hazard assumption that hazard function from the different individual hazard is proportional or the ratio of hazard function of two different individuals which is constant (Lee and Wang 2003). If the assumption is not met then the models that formed is the non-proportional hazard. In the case of on time graduation duration data on graduate students of Bogor Agricultural University (IPB-SPs) which is influenced by many factors, it tends to be nonproportional. It can be seen on the factor of Faculty performance which is not met the assumption because it is not constant in each year. Addressing the case of a non-proportional hazard is necessary to use the appropriate method. One method that can be used to resolve the case of nonproportional hazard (Kleinbaum

and Klein 2012) is by using the extension model of the cox proportional hazard regression that is Extended cox proportional hazard.

The Extended cox proportional hazard regression contains time-dependent variable or the multiplication of independent variables with the function of time.Several studies about Extended cox proportional hazard regression model and Stratified cox proportional hazards such as survival analysis of debtor by using extended cox and stratified cox (Kurniawan 2016) which states that the best model of the debtor's survival is the model of stratification cox . Furthermore, Ata and Sozer (2007) explained about the Cox regression model for addressing the nonproportional hazard to lung cancer survival data. In such research, the stratification Cox model without interaction and extended cox model deliver more suitable results based on the value of the Akaike's Information Criterion (AIC). In relation to data, Sabil (2016) conducted a study using data of IPB graduate students in 2008 until 2010 by modeling the durability of graduate students from dropping out of the College using survival analysis

Based on some previous finding, in this study, the researcher would like to study about Extended cox proportional hazard (PH) regression model. The data used are the data of regular student of IPB graduate school to find out the factors that affect the on time graduation duration of the student of IPB graduate school.

2. RESEARCH METHOD
2.1 Data

This research is to apply the method of survival analysis using secondary data. The data used is the data master student Graduate School of Bogor Agricultural University class of 2011 through 2013, were observed for 4 years where there are 63 regular classroom courses involved. This data was obtained from the academic part of the Graduate School of Bogor Agricultural University. Response variable in this study is the timing of an event that is passed on time (≤24 months). Some of the variables used in this study are as follows**:**

Table 1 Independent Variable

No	Variable	Categori
1	X_1 (Gender)	0=male 1=female
2	X_2 (Marrital Status)	1=married 2=single 3=widow/widower
3	X_3 (Age)	1= x<33 years old 2=33≤x< 49 years old 3= x ≥ 49 years old
4	X_4 (Employment Status)	0= unemployed 1= employed
5	X_5 (Source of fund)	1= own cost 2= scholarship
6	X_6 (Accreditation)	1= A accredited 2= B accredited 3= C accredited
7	X_7 (Undergraduate GPA)	0= GPA < 2.75 1= 2.75 ≤ GPA < 3 2= 3≤ GPA < 3.25 3= GPA ≥ 3.25
8	X_8 (Status of Acceptance)	1= ordinary 2= trial

| 9 | X_9 (university of origin) | 1= private
2= contry |
| 10 | X_{10} (faculty) | |

2.2 Methods of Analysis

Proceduresto analysis data were following:
1. Description of the data to obtain an overview of the data;
2. Analysis of survival data with semiparametric method, using Cox proportional hazard models. Cox proportional hazard models are used to see the effect of independent variables simultaneously or establish regression models. Generally cox proportional hazard models;

$$h(t|x) = h_0(t)\exp\left(\sum_{j=1}^{p} \beta_j x_{ij}\right)$$

where,
h(t|x)   = *hazard* fungsi of Cox PH models
t   = time until the occurence of an event
$h_0(t)$ = *baseline hazard function*
β_j = coefficien of Cox models variable j-th
x_{ij} = Independent variable j-th for indvidual   i-th

3. Testing the assumption of proportional hazards was using the Pearson correlation coefficient test based on the Schoenfeld residuals. Schoenfeld residuals are based on the contributions of individuals with partial derivatives of the log likelihood (Hosmer and Lemeshow 1999);
4. If there are independent variables that do not meet proportional hazard assumptions, then modeling using the extended cox proportional hazard regression is used. In general, the extended cox models;

$$h(t, x(t)) = h_0(t) \exp\left(\beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \delta_1 x_{i1} g(t)\right)$$

where
h(t,x(t))   = hazard function that adds time-dependent variables.
x g(t)   = independent variables that are interaction with the time function g(t)
$h_0(t)$ = baseline hazard function
δ   = coefficient X that are interaction with function of time

3.RESULT AND DISCUSSION
3.1 Descriptive Analysis

The data used is the data master Graduate School of Bogor Agricultural University (SPs IPB) of 2011 to 2013 as many as 3,600 observations and after cleaning process the total data used are 3443, by entering the time of observation for 48 months students who graduate on time as many as 453 and as many as 2990 data censored data. Details of total students per class are listed in Table 2**.**

Table 2 The number of students of GS-IPB

Level (Entry year)	Number of student	Students Graduating on time		duration of observation duration
		quantity	persentation	

2011	917	93	10.142	48 month
2012	1190	177	14.874	36 month
2013	1336	183	13.698	24 month

Based on table 2, the percentage of students graduating on time based on the duration of their observations shows an increase in each year. Percentage of students graduated on time in 2011 and 2012 increased by 4.732% from 10.142% to 14.874%. In 2013, however, slightly decreased by 1.176% to 13.698%. Overall data distribution based on other variables shown in Figure 1 are:
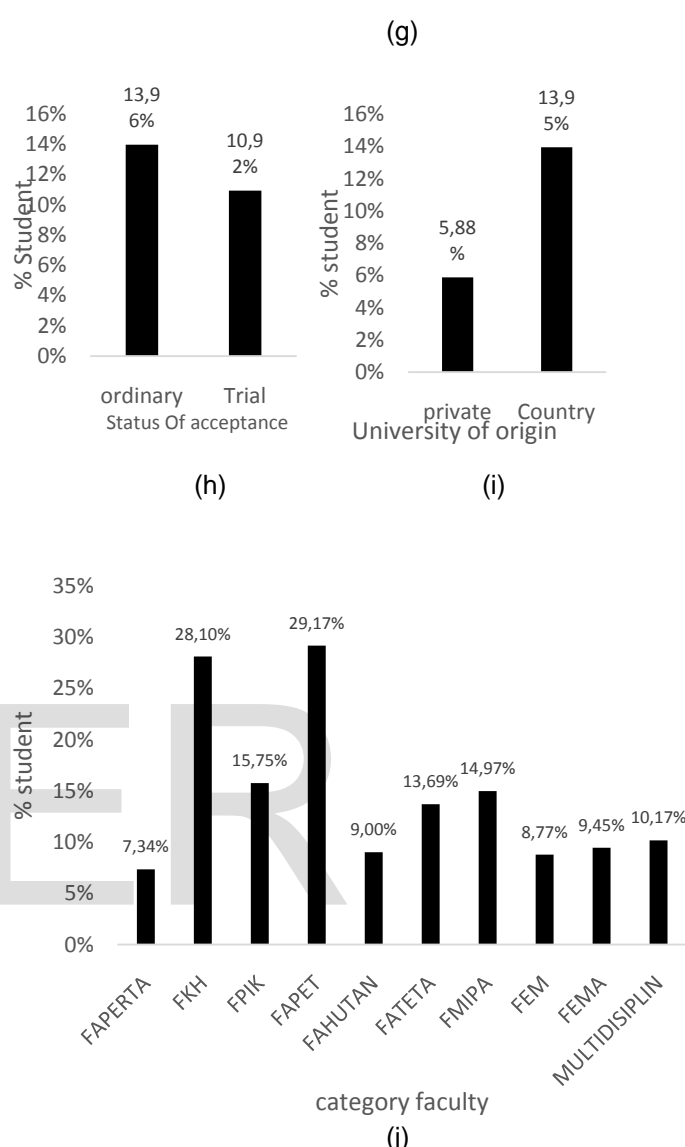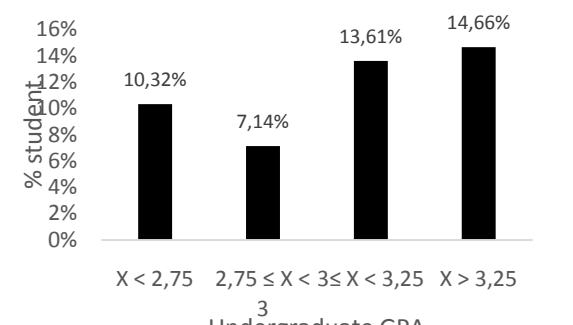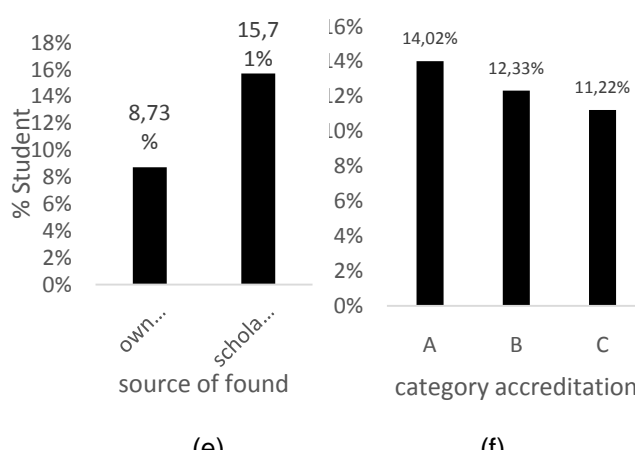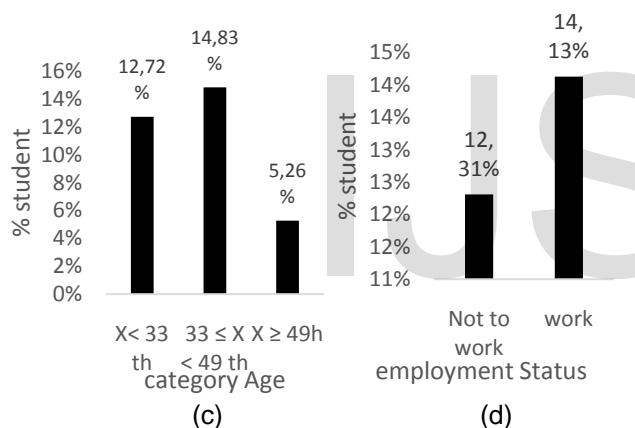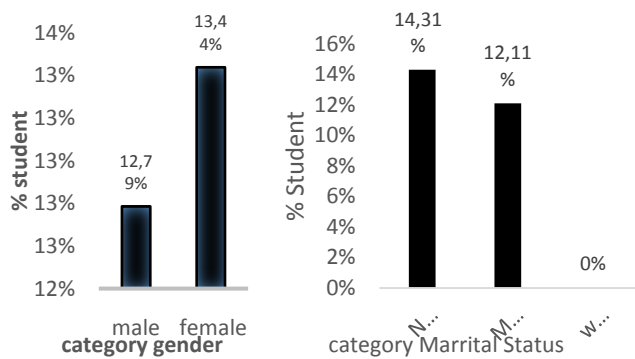
(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

Figure 1 Percentage Distribution of Data Student Graduate School of Bogor Agricultural University

Based on the illustration in figure 1, it is generally explained that the percentage of the graduation data distribution of the on time graduation for postgraduates based on sex shows that the percentage of female sex with men is not much different. The percentage of female sex graduated on time was 13.43% of the total students as many as 1957 people. The percentage for male gender is 12.78% from 1486 students. Based on the marital status, The student entered at Graduate School Bogor Agricultural University (SPs-IPB), for which single/unmarried status was 14.31% while the married status was 12.11%.

Based on the age status, students under 33 years old have a graduated percentage of 12.72%, 14.83% is in the range of 33 and 49 years old while the percentage of on time graduation duration for students aged over 49 years is 5.26%. Furthermore, the

percentage of on time graduation for unemployed students is 12.31% while those who have worked 14.31%, and for students who get scholarships as much as 15.71% while for students whose source of funds comes from self or private is only 8.73%. For students who have an origin GPA of under 2.75 is 10.32%, between 2.75-3 as much as 7.14%, GPA between 3-3.25 is 13.61% and above 3.25 has the percentage of on time graduation is 14.02%. Accreditation of the department of origin university of A is 14.02%, B accreditation of 12.33% and accreditation C only 11.22%. Furthermore, for the status of acceptance in IPB with ordinary status has a higher percentage than the status of acceptance of trial that is 13.96% and 10.93% respectively and students from public universities who graduated on time of 13.95%, and 5.88% for those coming from the private universities. Based on faculty variables, students from 2011 to 2013 came from the Faculty of husbandry (FAPET) with 29.17% percentage, while the smallest percentage is the faculty of agriculture (FAPERTA) of 7.34% .

3.2 Semiparametric model
3.2.1.Cox Proportional hazard

One of the purposes of Cox Proportional hazard model analysis is used to describe the influence of the independent variables on dependent variable that is simultaneously on time graduation in a given period. The analysis process by entering the independent variables into the model. Parameter test is done using the G test. Based on the result of SAS 9.3 software, the obtained value of Chi-squared test statistic is 141.437 and the P-value is <0.001. This indicates that there is at least one regression coefficient which is not equal to zero, meaning that there is at least one independent variable in the model that has a significant influence on the on time graduation of students of the Bogor Agricultural University. Partial parameter test is using Wald Test. Independent variables that have significant effect on the 5% of significance level on the on time graduation of graduate students of Bogor Agricultural University (IPB) are age, education funding source, undergraduate GPA, university origin and faculty.

3.2.2.Proportional Hazard Assumptions

The proportional hazard assumption is an assumption to note in the Cox Proportional hazard regression. The assumption is classified to be fulfilled if the failure rate (Hazard rate) of two observations is constant during the observation time. If the proportional hazard assumption is not met then it is necessary to test the proportional hazard assumption. In this study, the test of proportional hazard assumption used the pearson correlation coefficient test based on Schoenfeld residual. Based on Sehoenfeld residual test results by looking at the correlation value for each independent variables in Table 3, it can be concluded that there are independent variables that do not meet the proportional hazard assumption i.e the variable of faculty in IPB.

$$\lambda(t,x) = \exp\big(0.30860x_{3(2)} + 0.51683x_5 + 0.81061x_9$$
$$+ 1.42116x_{10(2)} + 0.79102x_{10(3)}$$
$$+ 1.46302x_{10(4)} + 0.67899x_{10(6)}$$
$$+ 0.84712x_{10(7)}\big)$$

Table 3 Pearson Correlation Coefficient Test

Variable Name	correlation	Value- P
Age(X_3)	-0.07248	0.1235
Source Education Fund (X_5)	0.00510	0.9138
University of origin (X_9)	-0.05309	0.2595
Faculty (X_{10})	0.28042	<.0001*

* real at α = 5%

3.3. Development Model for Addressing Non Proportional hazard
3.3.1. Extended Cox Proportional Hazard Model

The extended cox is one approach to examine the proportional assumptions of a covariate in addition to the graphical approach and goodness of fit test. In addition to examining the proportional assumptions of a model, extended cox can also simultaneously model the time-dependent variables and predict how much it will affect the durability. On the model of extended cox, variables which do not satisfy the assumption of proportional hazard interacted with a function of time. The function of time used is g (t) = (t) then interacted with the variable that did not meet the assumptions of the PH that is Faculty variable (X_{10}).

The interpretation of the cox coefficient of the extended cox model can be seen from the hazard level ratio. Hazard ratio (HR) is the ratio between the hazard value of one individual to another. Based on the age variable, students whose age is in the range of 33 to 49 years have a tendency to graduate on time of 1.359 times greater than students aged under 33 years. Students who get education funding source from the scholarship have the opportunity to graduate on time of 1.676 more than the students who finance their own college or in other words use the personal fund. Students who come from public universities have a tendency to graduate on time of 2.356 times greater than students from private universities.

The faculty variables are influenced by time thereby causing the change of probability of graduate on time. In the category variables of faculty, the faculty that has a significant effect on the student's on time graduation is FATETA, FMIPA, FEMA and MULTIDICIPLIONARY. The on time graduation rate of FATETA students is initially amounted to 1.282 than FAPERTA as a comparison, as influenced by the time, on time graduation of FATETA student tend to rise to 1.325 compared to FAPERTA. Likewise with the on time graduation duration of FMIPA students, initially of 0.003 compared to FAPERTA, because it is influenced by the time, the trend of on time graduation of FMIPA students rose to 1495 greater than FAPERTA. For the on time graduation duration of FEMA students of 0.000 than FAPERTA, increased to 2.736. For the Multidisciplinary Faculty initially of 0.000 compared to FAPERTA, as it was influenced by the time, the trend of on time graduation the multidisciplinary faculty students rose to 2.170 compared to FAPERTA.

3.4. Selection Of the best Model

Comparison of Cox Proportional Hazard and Extended Cox Proportional Hazard models was conducted to find out which model is appropriate to explain the factors that affect the on time graduation duration of graduate school students of the Bogor Agricultural University. In this study, the conformity values

of the models that can be used to determine the appropriate model are -2 log likelihood and AIC. The best modeling conformance criteria are shown in Table 4 below:

Table 4Value conformance criteria best model

	Model Cox Proportional Hazard	Model Extended Cox
AIC	7196,347	7159,298
-2 Log Likelihood	7150.347	7107.787

Based on table 4, the extended cox proportional hazard model has the -2log likelihood and AIC values smaller than the proportional Hazard cox model, i.e -2 log likelihood = 7107.787 and AIC = 7151.787. In the case of the on time graduation duration of graduate school students of the Bogor Agricultural University, the Extended cox proportional hazard model is better than the cox proportional hazard model.

4.CONCLUSIONS

Based on the results of this research, it can be concluded that:

1. The Cox proportional hazard model can be used to model the relationship between resistance time and the covariates that are suspected to affect the durability time. However, in some cases, if there are covariates whose value is time-dependent then the extended Cox model is able to provide a better description of the relationship because it has a smaller AIC value as in the case of on time graduation of graduate student of Bogor Agricultural University. Cox proportional hazard model produces an AIC value of 7196.347 while the extended Cox model has an AIC value of 7151.787 based on AIC criteria then the best model for the case is the extended Cox model with the time function used is g (t) = t.
2. Based on the modeling results using the extended cox proportional hazard model, the factors that affect the on time graduation duration of the graduate students of the Bogor Agricultural University are age, education funding source, university origin and faculty.

REFERENCES

[1] Hosmer D, Lemeshow S. *Applied Survival Analysis.* New York: Wiley, 1998.
[2] [IPB] Bogor Agricultural Institute. Catalog IPB Graduate School Academic Year 2014/2015. Bogor, IPB, 2014.
[3] Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text. 3rd ed. Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, editor. New York (US): Springer.doi: 10.1007 / 978-1-4419-6646-9, 2012
[4] Kurniawan I. 2016. Debtor Endurance Analysis Using Expansion of Cox and Cox Stratification Models. [Thesis]. Bogor: Graduate Program Bogor Agricultural University
[5] Sabil AL. Durability Modelling Graduate Student of the End Class Using Survival Analysis. [Thesis]. Bogor, IPB, 2016