ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 8 , August 2020

Parametric Accelerated Failure Time Model in the Analysis of Undergraduates Students Survival

AzmeKhamis, CheAzmeezaCheHamat, MohdAsrul Effendi Abdullah, NorzihaChe Him ,Shuhaida Ismail

Department of Statistics and Mathematic, Faculty of Applied Science and Technology,
UniversitiTun Hussein Onn Malaysia

ABSTRACT: Survival analysis methods can be applied to a wide range of data. This study focused on the data of undergraduate students in UniversitiTun Hussein Onn Malaysia (UTHM). The application of the parametric survival model can be extrapolated for the purpose of the study were to identify the best parametric model, to propose the best model for predicting the chance of students to survive and to find out the covariates affect the survival time among UTHM students. Accelerated Failure Time (AFT) model such as Exponential, Weibull and Lognormal model for right censored data have been used as an improved method of parametric survival model. This AFT model provided less value of AIC, AICc and BIC. Accelerated Failure Time models also used to estimate the effects of covariates on prolonged survival time or associated with a decrease in survival time. The Weibull AFT model found to be the best model as obtained the least value of AIC, AICc and BIC. Hence, this Weibull AFT model proposed for applied in predicting the chance of students survives. The result of the parametric AFT model easier to interpret as the covariates effects was directly expressed in terms of time ratio (TR). The time ratio obtained from the Weibull AFT model was less than one and it associated the results students obtained for each Semester shorten a survival of students.

KEYWORDS*:* Survival Analysis, Accelerated Failure Time model, AIC, AICc, BIC, Time Ratio

I. INTRODUCTION

The random variable of most interest in survival analysis was time to event data. Survival analysis methods can be applied to a wide range of data, not just biomedical survival data. Another time to event data can be included such as length of time students pursue the Degree, the length of a contract, the duration of a policy and time to finishing a master thesis [1].The survival model based on parametric model was effectively used to achieve the objective. The best distribution selected bythe lowest value of AIC, AICc and BIC. Then, the best model was chosen to propose as the model to predict of the survivability of each Degree student in UTHM.

The parametric model focus on general study of survival time. This study focused only status and semester enrolled by students. Therefore, using this classical parametric model to further the study can be misleading. The other extended parametric model must be proposed for the modeling of students' achievement. This new model proposes to incorporate all of the covariates factor of the Degree students since this covariates had an impact on the survival time. The covariates exist in this study were GPA and Semester enrolled by each students for every semester from Semester 1 until Semester 8.

II. SIGNIFICANCE OF THE SYSTEM

This study embarks on the following objectives which were to identify the best parametric model, to propose the best model for predicting the chance of students to survive and to find out the benefit of covariates affect the survival time among UTHM students.The analysis then aims to identify which the best-fitted survival models. In the survival analysis, the three models yielded different results. The AIC, AICc and BIC values for each model was compared to identify the best model that will be propose for prediction the chance for students to survive during the Degree.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 8 , August 2020

III. LITERATURE SURVEY

The two methods that often used to incorporate the effect of covariates on lifetimes of students were based on the Accelerated Failure Time model and the Proportional hazards (PH) models [2]. The Accelerated Failure Time model was an attractive alternative to the popular Cox proportional regression model. The accelerated failure time model was a linear regression model in which the response variable was the logarithm or a known monotone transformation of a failure time [3]. The proportional hazards model was appropriate when there was a permanent difference between the groups in the longer term in the context of the follow-up period [4],[5]. The accelerated failure time model was more appropriate when the group differences were seen over a shorter timeframe [6].

There were some studied the effect of education programs to promote students retention and performance. The application of the AFT model used for the analysis by others in many contexts including studies of sociology [7], bladder cancer [8] and cardiovascular disease [9]. However, this research was done to be the first application of AFT model to study UTHM undergraduates' students' performance.

Parametric Accelerated Failure Time model was one such model and most commonly used were Exponential, Weibull and Lognormal models. The parametric approach offers more in the way of predictions and the AFT formulation allows the derivation of a time ratio, which more arguably and more interpretable [10]. Furthermore, AFT model was the technique used to incorporate all of the covariates in the analysis that will increase the robustness of the model and provided less error in the prediction [11].

IV. METHODOLOGY

The undergraduates' students' data were taken from UniversitiTun Hussein Onn Malaysia (UTHM). The data consist of students who have been pursuing a Degree. Students who were lost to follow up due to other competing causes or with the incomplete data were excluded from the study. The variable time in the semester is the survival time, which was measured from the students was diagnosed with GPA more than 2.00 up to time of students end the study. Therefore, time was considered as the variables of interest. In the analysis, other variables that were considered were faculty, course and the level of study. Statistical package of R package was used to perform the data analysis.

A. Accelerate Failure Time (Aft) Model with Covariates

The Accelerate Failure Time model (AFT) was the technique that used to incorporate covariate into the survival model. Thus, the AFT model gave a lot of important information. This AFT model differs with the classical parametric model (Weibull model, Exponential and Lognormal model) since the AFT model can help to predict the survival time of a specific student rather than the general population.

The AFT model assumes that,

$$S_j(t) = S_i(\emptyset_{ij} t) \text{for all } t. \quad (1)$$

where $\emptyset_{ij}$ is a constant that is specified to the pair of individuals (i, j). This model says that what makes one individual different from another is the rate at which time.

Let (L_i, R_i) be the interval in which $T_i$is being observed, such that L_i<T_i<R_i. If the event does not occur till the end of study then the patient is said to be right censored, in this case, it assumed that $T_i$can occur in the interval (L_i, ∞), where L_i was the tie period from the beginning of the study until the last semester. Let $X = (X_{i1}, X_{i2}, \dots X_{ip})'$ be the values of p covariates for the i^{th} patient. Then the log linear form of Accelerated Failure Time model describes a linear relationship between logarithm of survival time and covariates, as given by[12]

$$\log(T_i) = Y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \sigma\varepsilon_i$$
$$(T_i) = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \sigma\varepsilon_i} \quad (2)$$

where β_0 and σ were intercept and scale parameters respectively and ε_i assumed to have some distribution (Exponential, Weibull and Lognormal). This transformation leads to the Exponential, Weibull or Lognormal AFT models for T_i[13].

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 8 , August 2020

B.Maximum Likelihood estimation

Suppose that $Y_i = log(T_i)$, then the density function of Y_i given by $\left(\frac{f(w)}{\sigma}\right)$) , where $\left(w = \frac{(y - X'\beta)}{\sigma}\right)$. Here assume that $X_{i1} = 1$ for all i. Then the likelihood function for the set of observed intervals $\{(L_i, R_i], X_i, i = 1,2 \dots n\}$ can be written as:

$$L(\beta,\sigma) = \prod_{i=1}^{n}\left[\left(\frac{f(W_{i1})}{\sigma}\right)\right]^{\delta_i}\left[S(w_{i1}) - S(w_{R1})^{1-\delta_i}\right] \quad (3)$$

where:

$$w_{i1} = \frac{(\log L_i - X_i'\beta)}{\sigma}, w_{R1} = \frac{(\log R_i - X_i'\beta}{\sigma}$$

$$S(w) = \int_W^{\infty} f(s)\,ds \quad (4)$$

And the δ_i defined as:

$$\delta_i = \begin{cases} 1 & if\ L_i = R_i \\ 0 & otherwise \end{cases}$$

The function $f(w)$ and $S(w)$ denote the probability density and survival functions of the error variable w in model 2, respectively. The estimates can be obtained by maximizing the likelihood equation (3) with respect to $\beta$and σ. A complete description of the procedure is given [14] [15].

Results obtained from AFT models can be summarized in the exponentiated form as time ratio, $TR = exp(\beta)$) unlike Cox model hazard ratio. Thus $TR > 1$, indicates prolonged survival time and $TR < 1$, associated with a decrease in survival time.

C. Model comparison

In order to select best fit model, the Goodness of fit test were used. The AIC provides a practical and versatile way to identify a parsimonious model from a set ofcompeting models, by adding a penalty term proportional to the number of parameters in the model. This penalty term guards against over fitting. The AIC wasdefined as:

$$AIC = -2 * Log - likelihood + 2(p + k) \quad (5)$$

where, p was the number of covariates in the model, $k = 1$ for exponential and $k = 2$ for Weibull, and Lognormal models. The model with smaller AIC value was termed as the best model. The AIC penalizes the number of parameters less strongly than the BIC [16]. BIC is defined as:

$$BIC = -2 * Log - likelihood + p.log(n) \quad (6)$$

where, p represents the number of covariates in the model and n represents the number data points. The main advantage of the BIC approximation is that it includes the BIC penalty for the number of parameters being estimated. The model with smallest BIC values chose as the best model. The corrected AICc also assess the model fit.

The lowest value of AICc indicated the best fitted model. The AICc formula defined as:

$$AIC_c = -2 * \log(likelihood) + 2(k) + 2(k)(k+1)/\,(n-k-1) \quad (7)$$

where$k$ is the number of estimated parameters in the model and n is the dataset. This value can be used to compare various models for the best fitting model. The model had the smallest value, preferred as the best model.

V. EXPERIMENTAL RESULTS

Parametric survival models possess some advantages such as utilization of full likelihood to estimate the parameters, and providing estimates in terms of survival. In theories, classical parametric models overcome the deficiencies of Kaplan Meier and the Cox methods[17]. These classical parametric models allowed easy computational of AIC, AICc and BIC regardless of any factor. The result of selected the best classical parametric model was presented in Table 1.

Table 1: Log-likelihood, AIC, AICc and BIC for classical parametric models

Model	Log-Likelihood	AIC	AICc	BIC
Exponential	-50594.3000	101190.6000	101190.60002	101198.3735
Weibull	-21975.3500	43954.7000	43954.7007	43970.2469
Lognormal	-25479.9000	50963.88	50963.88	509579.4

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 8 , August 2020

From the result obtained, this indicates that Weibull model was the best model to fit the data. The classical parametric model able to provide smooth and analysis of survivor. However, this classical model hardly used in the research because of the simple answer and the major problem with classical parametric models was simply lack of fit because of limited flexibility with higher values obtained through AIC, AICc and BIC.

Parametric Accelerated Failure Time models developed through extended of the classical parametric model to provide the real step forward. This parametric AFT model retains the theoretical strengths of the classical parametric model and overcome the limited flexibility of the classical parametric model through incorporating all of the covariates in the model.

Table 2: Log-likelihood, AIC, AICc and BIC for parametric models with covariates

Model	Log-Likelihood	AIC	AICc	BIC
Exponential AFT	-50298.9	100621.0000	100621.8178	100644.4903
Weibull AFT	-14569.0	29164.0000	29164.0207	29184.6903
Lognormal AFT	-18807.8	37641.6000	37641.6207	37662.2903

The AIC, AICc and BIC values of Weibull AFT model in Table 2 found to be smaller than the classical parametric model, hence it suggests that the Weibull AFT model could be the best model for study on student achievement. A large values of AIC, AICc and BIC obtained by Exponential AFT. It showed that this model was the least efficient. From this analysis, it found that the Weibull AFT model provides the best fit to the data.

A. Comparative Result

Table 3: The comparison AIC, AICc and BIC values obtained by the best classical parametric model and the best parametric AFT model.From this result, it summarized that, the classical parametric model estimates the general students' survival time. From the classical model, most questions about the general Degree students can be answered. But, the Weibull Accelerated Failure Time model gave a lot of information from it focused on every covariates included in this model. The covariates included in the model was semester students pursued the Degree which Semester 1 until Semester 8 and GPA. The covariates existed in the analysis using Parametric Accelerated Failure Time model increase the efficiency of the model through lower values of AIC, AICc and BIC obtained. Hence it suggested that the Weibull AFT model could be the best model for the further study.

Table 3: The values of AIC, AICc and BIC for Weibull and Weibull AFT models

Model	Weibull model	Weibull AFT model
Log-Likelihood	-21975.35	-14569.00
AIC	43954.700	29164.00
AICc	43954.701	29164.02
BIC	43970.247	29184.69

The time ratio helps the management to determine the benefits of the covariates exist whether increasing or increasing the survival time. Hence, time ratio was estimate to find out the trend of covariates affect the survival time among UTHM students. For the sake of simplicity and ease of interpretation, the exponentiated regression coefficients ($\exp(\beta_i)$) called time ratio (TR) was recommended to report. In the AFT model, the TR is recommended to report, whereas a hazard ratio (HR) was reported in proportional hazard models. TR > 1 for covariates implies that prolongs the survival time and TR < 1 for covariates decrease the survival time.Table 4 showed the coefficients and Time Ratio (TR) for the Weibull AFT models. The covariates which were Semester 1 until Semester 8 and GPA were found to be statistical significant factors for the survival of undergraduate students in UTHM.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 8 , August 2020

Table 4: The parameter estimated for Accelerated Failure Time (AFT) model

Parameters	B	Std. Error	Time Ratio (TR)	p-value
Intercept	2.434	0.00422	11.3997	0.0000
Sem_1	-0.373	0.00284	0.6887	0.0000
Sem_2	-0.369	0.00285	0.6916	0.0000
Sem_3	-0.370	0.00285	0.6908	0.0000
Sem_4	-0.373	0.00285	0.6888	0.0000
Sem_5	-0.381	0.00287	0.6829	0.0000
Sem_6	-0.373	0.00287	0.6886	0.0000
Sem_7	-0.351	0.00306	0.7041	0.0000
Sem_8	-0.339	0.00012	0.7122	0.0000
GPA	-0.046	0.00109	0.9550	0.0000
Log (scale)	-2.677	0.00583		
Scale	0.068877			
Shape	14.54075			

For all of the Semester variables, the TR < 1 indicated that when these covariates 1 unit change shorten survival times compared to the baseline survival. It can be seen that 1 unit changes in Semester 1 shortens survival times by 0.6887 which is 68.87%, 1 unit changes in Semester 2 shortens survival times by 69.16%. The interpretation of TR for other covariates can also be made in similar fashion referred to the Table 4. From the result obtained, on semester 8, 1 unit changes shortens survival times by 71.22%. The 1 unit changes shorter survival time with the highest percentage and it means that on semester 8, students had a better result compared to the other semester.

B. Discussion

Published studied in the field of survival analysis were often interested in the Cox proportional hazard model instead of parametric survival models. However, in the analysis of students' achievement, it found that many previous studied focused on the data mining and lack of studies using survival model. As mentioned earlier, the main objective of this study was to obtain a survival model for undergraduate students in UTHM. In particular, two methods were applied to the analysis, namely classical parametric survival model and Parametric Accelerated Failure Time model.

The advantages of using parametric survival model were the simplicity and completeness. For parametric survival models, times were assumed to follow some distributions whose probability density function can be expressed in terms of unknown parameters. Once a probability density function was specified for survival time, the corresponding survival functions can be determined. Since there were several parametric survival models, the model that gave a good fitted data will choose to further the study. However, parametric survival model cannot obtain an estimated coefficient of the categorical variables effect [18].

Finally, parametric accelerated failure time model was presented as an alternative and have many situations where the AFT model provides the best description of data than Cox PH model [19]. The findings suggest that the semester students enrolled from semester 1 until semester 8 and GPA obtained were statistically significant (p<0.0000). The analysis conducted using classical parametric models such as Exponential model, Weibull model and Lognormal model. These models were conducted from the survival times of all students, regardless of any factor. These classical parametric models focused only on the time variables which were semester enrolled by students and status variables which were either students GPA > 2.00 or GPA < 2.00.

Among the result obtained, the classical Weibull model was identified to provide the best fit and among the parametric AFT model, the Weibull AFT model found to provide the best fit. In order to identify the best model to be selected, the values of AIC, AICc and BIC obtained by the classical Weibull model and the Weibull AFT model was compared. The result can clearly see in Table 3.The Weibull AFT model provides the model with sufficient flexibility to adequate fit real data with values of AIC = 29164.00, AICc = 29164.020 and BIC = 29184.69 lower than the classical Weibull model with AIC = 43954.7, AICc = 43954.701 and BIC = 43970.247. Model comparison using the goodness of fit showed the improvement of model fit when switch from classical parametric model to the parametric AFT model. On the basis of log-likelihood, AIC, AICc and BIC, the Weibull AFT model was the best fitted model.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 8 , August 2020

VI. CONCLUSION AND FUTURE WORK

The estimate parameter of the Weibull Accelerated Failure Time (AFT) model was by applying the maximum likelihood estimation [20]. Since the prediction using the classical Weibull model can be misleading, the final choice of the model to predict students' performance depends on building model with covariates which Weibull Accelerated Failure Time model and assesses the goodness of fit through Akaike Information Criterion, Akaike Information Criterion Corrected and Bayesian Information Criterion [21].

The AFT model used to incorporated covariates into the survival model. AFT model was a failure time model which used for the analysis of time to event data. The model works to measure the effect of covariate to accelerate or to decelerate survival time. From this study, the time ratio obtained from each semester using Weibull AFT model shorten a survival time of each students. Because of that, an initiative should be taken to improve students' performance during the Degree.

The Parametric AFT models would be the best model of survival time was recognized to see the effect of each semester in survival time of the Degree students. On the basis of log-likelihood and Goodness of fit test, the conclusion was the data adequately fitted by the Weibull AFT model as AIC, AICc and BIC of this model less than the Weibull model. The Weibull AFT model suggests that the presence of covariates which Semester 1 until Semester 8 and GPA showed considerable association with survival time in the models. The result ofthe AFT models were easier and more reliable to interpret not only for students, but also as a reference for the management on the appropriate explanation of survival data.

Acknowledgement

The authors would like to thank to University Tun Hussein Onn Malaysia, Academic Students' Department (PPA) for providing us with the secondary data of the student achievement for every semester and valuable source of information. We thank to the editor, associate editor, and the reviewers for their constructive comments that improved the quality of this article. This research was sponsored, in part, by Postgraduate Research Grant (GPPS) U604.

REFERENCES

[1] T. R.Fleming, & D. P. Harrington, "*Counting process of survival analysis"*. A John Wiley& Sons Inc Publication: Canada, 2008.
[2]D. R. Cox, & d. Oakes, "*Analysis of survival data"*, 2nd Ed. Chapman & Hall: London, 1984.
[3]E. T. Lee,& J. W. Wang, "*Statistical Methods for survival Data Analysis".* New Jersey: John Willey, 2003.
[4]M. A. Maggard, J. E. Thompson, & Y., Koc."Why do breast cancer mortality rates vary across states?"*American Journal Surgery.* 2003, 69:59–62.
[5]A. R. Rajaeefard, M. R.Baneshi, A. R.Talei,&D. Mehrabani,"Survival Models in BreastCancer Patients". *Iranian Red Crescent Medical Journal*. 2009, 11: 295-300.
[6]P. H. Zahl,& S.Tretli, "Long-term survival of breast cancer in Norway by age and clinical stage". *Statistics in Medicine*. 1997, 16:1435–1449.
[7]K. Yamaguchi, "Some accelerated failure-time regression models derived from diffusion process models". *Sociology Methodology Journal*. 1994, 24: 267- 300.
[8]D. Y. Lin, L. J. Wei,&Z. Ying, Z. "Accelerated failure time models for counting processes". *Biometrika*, 1998, 85(3): 605-618.
[9]A. Menotti,P. E. PudduM. Lanni, D. Kromhou, H. Blackburn, & A. Nissinen, 'Twenty-five-year coronary mortality trends in the seven countries stud using the accelerated failure time model". *Eur. J. Epidemiology*, 2003, 18, 113-122.
[10]R. Kay, & N. Kinnersley, "On The Use of Accelerated Failure Time Model as an Alternative to the Proportional Hazards Model in the Treatment of Time to Event Data: A Case Study in Influenza". *Drug Inf. Journal*. 2002, 36: 571-579.
[11]M. Schmid, & T.Hothorn, "Flexible boosting of accelerated failure time models". *BMC Bioinformatics.* 2008,9: 269.
[12]J. D. Kalbfleish,& R. L. Prentice, "*The Statistical Analysis of Failure Time Data"*. 1st Ed. Wiley: New York. 1980.
[13]D. Collet, "*Modeling Survival Data in Medical Research"*. 2nd Ed. Chapman & Hall: London, 2003.
[14]J. P. Klein,& M. L. Moeschberger. "*Survival Analysis, Technique for Censored and Truncated Data"*. 1st Ed. Springer: New York, 1997.
[15]J. Sun, "*The Statistical Analysis of Interval Censored Failure Time Data"*, 1st Ed, Springer: New York, 2006.
[16]G. E. Schwarz,"Estimating Dimension of a Model". *Ann Stat Journal*. 1978, 6: 461-464.
[17]G. Reibnegger, "Modeling Time in Medical Education Research: The Potential of New Flexible Parametric Methods of Survival Analysis". *Scientific Research*, 2012, 3: 916-922.
[18]M. J. Bradburn, T. G. Clark, S. B. Love,& D. G. Altman, "Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods". *British Journal of Cancer*. 2003, 89: 431 – 436.
[19]J. Orbe, E. Ferreira,&V. Nunez, "Comparing proportional hazards and accelerated failure time models for survival analysis". *Statistics in Medicine*. 2002, 21: 3493-3510.
[20] M. Youngkyoung, Z. Guili, A. L.Russeli, J. A. Timothy, & W. O. Matthew, "Nonparametric survival analysis of the loss rate of undergraduate engineering students". *Journal of Engineering Educations.* 2011, 100(22): 349–373.
[21]M. K. William,*"Research Method knowledge Base: Inferential Statistics"*.Web Centre of Social Research Methods,