STATISTICAL MODEL ON STUDENT PERFORMANCE IN UTHM
BY USING NON-PARAMETRIC, SEMI-PARAMETRIC AND
PARAMETRIC SURVIVAL ANALYSIS

CHE AZMEEZA BINTI CHE HAMAT

A thesis submitted in
fullfilment of the requirement for the award of the
Degree of Master of Science

Faculty of Applied Sciences and Technology
Universiti Tun Hussein Onn Malaysia

JUNE 2019
For my beloved parents Rahana and Che Hamat.
ACKNOWLEDGEMENT

Thankful to God for the mercy, blessings and strength that were bestowed upon me in the quest to complete my thesis. This project bears the beautiful imprints of many peoples.

A token of gratitude to my beloved parents, Che Hamat Bin Che Harun and Rahana Binti Saamah for their love, support and inspiration during my study. Not to forget are my siblings, Azmiral, Azreen, Aznil and Ruslina; thank you very much for the support and motivation showered on me during this study.

My sincere gratitude and special thanks to my supervisor Professor Dr. Azme Bin Khamis for his insights, invaluable guidance, continuous assistance and encouragement throughout this research journey. I am especially indebted to him as a true mentor, the ultimate supervisor and motivator. I may not be able to make it this far without his perseverance, faith and patience. I am also grateful to my co-supervisor, Dr. Mohd Asrul Affendi Bin Abdullah for his guidance, opinion and motivation as well. I really appreciate the guidance from the initial to the final levels that enables me to develop thorough understanding of this research.

I am also sincerely thankful to the Research, Innovation, Commercialization & Consultancy Office (ORICC) Universiti Tun Hussein Onn Malaysia for providing the support and sponsors to do this project.

I am also indebted to many other people who have helped me directly or indirectly in this study. This thesis may not be successfully completed without the memorable involvement of my many lecturers, friends and others. I sincerely thank you.
Student performance defined as students who are capable to success during their studies. This study explored the use of survival analysis to investigate the performance of Bachelor’s degree students in Universiti Tun Hussein Onn Malaysia (UTHM). The data was collected from the Academic Management Office (AMO), UTHM. The main objective of this study is to estimate the survival rates of students with different entrance qualifications. The study also aim to identify the covariates that dominate the student performance, investigate the performance of Cox model based on the violation of the Proportional Hazard (PH) assumption, compare the model performance by using the survival and Accelerated Failure Time (AFT) models and estimate the time ratio (TR) of student performance in accordance to the selected best model. The survival analysis considered the survival approach such as the Kaplan-Meier (KM) method in the non-parametric method, Cox model in semi-parametric model and survival and AFT models in parametric model. The results revealed that students with STPM-entrance qualification had the highest survival rate compared to Diploma and Matriculation holders. The Cox model in the semi-parametric model identified the GPA, entrance qualification and course as the significant covariates to be included in the study. Faculty covariate was excluded since the $p$-value insignificant at 90% significance level. The result provided by the Cox model violated the PH assumptions. Then, the performance of the Cox model is less accurate. The invalidation performance of Cox model prompted the need to conduct other parametric survival and AFT models to produce more precise results. As a conclusion, the Log-normal AFT model is the best alternative model to estimate student performance in UTHM and other similar higher educational institution.
ABSTRAK

TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>TITLE</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF SYMBOLS AND ABBREVIATIONS</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF PUBLICATIONS</td>
<td>xvi</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Research background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Problem statement</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Research objectives</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Research scope</td>
<td>5</td>
</tr>
<tr>
<td>1.5 Research significance</td>
<td>6</td>
</tr>
<tr>
<td>1.6 Thesis organization</td>
<td>6</td>
</tr>
<tr>
<td>CHAPTER 2 LITERATURE REVIEW</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Students performance</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Statistical techniques applied to study on student performance</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Survival analysis for student performance</td>
<td>10</td>
</tr>
<tr>
<td>2.4 Survival analysis in other sector</td>
<td>12</td>
</tr>
<tr>
<td>2.5 Research gap</td>
<td>14</td>
</tr>
</tbody>
</table>
## METHODOLOGY

### 3.1 Data collection

### 3.2 Data description

### 3.3 Kaplan-Meier (KM) method

#### 3.3.1 Log-rank test

#### 3.3.2 Wilcoxon test

#### 3.3.3 Tarone-Ware test

### 3.4 Cox Model

#### 3.4.1 Hazard ratio

#### 3.4.2 Cox-Snell residual

#### 3.4.3 Schoenfeld residual

#### 3.4.4 Martingale residual

### 3.5 Parametric Model

#### 3.5.1 Parametric survival model

##### 3.5.1.1 Exponential model

##### 3.5.1.2 Weibull model

##### 3.5.1.3 Log-logistic model

##### 3.5.1.4 Log-normal model

#### 3.5.2 Parametric AFT Model

##### 3.5.2.1 The general AFT model

##### 3.5.2.2 Log-linear form AFT model

##### 3.5.2.3 Exponential AFT model

##### 3.5.2.4 Weibull AFT model

##### 3.5.2.5 Log-logistic AFT model

##### 3.5.2.6 Log-normal AFT model

##### 3.5.2.7 Time ratio (TR)

### 3.6 Model Selection

#### 3.6.1 Log-likelihood value

#### 3.6.2 Akaike Information Critera (AIC)

#### 3.6.3 Bayesian Information Critera (BIC)

#### 3.6.4 Corrected Akaike Information Criteria

### 3.7 Research Framework
### CHAPTER 4  RESULT AND DISCUSSION  48

1. Descriptive analysis  48
2. Nonparametric survival analysis  49
   - 2.1 KM analysis  49
   - 2.2 Comparison of survival result  54
3. Cox model analysis  55
   - 3.1 Covariates selection by Cox model  55
   - 3.2 Hazard ratio for Cox model  56
   - 3.3 Cox model residual analysis  57
4. Parametric model  63
   - 4.1 Parametric survival model  63
   - 4.2 Parametric AFT model  66
   - 4.3 Comparative result of the parametric survival and parametric AFT model  66
   - 4.4 The analysis of covariates of the best model  67

### CHAPTER 5  CONCLUSION  69

1. Main findings  69
2. Research contribution  71
3. Limitation and recommendation  72

REFERENCES  73

APPENDICES  86

VITA
**LIST OF TABLES**

2.1 The summary of literature review ........................................... 15
3.1 Description of variables and codes for data 2011/2012 ...... 21
3.2 Data of Bachelor’s degree student ................................. 22
3.3 Student-semester data set ................................................ 22
4.1 Number of students based on entrance qualification for each semester .................................................. 49
4.2 Probability of survival and failure for diploma-qualification ................................................................. 51
4.3 Probability of survival and failure for matriculation-qualification ......................................................... 52
4.4 Probability of survival and failure for STPM-qualification ................................................... 52
4.5 χ² test and p-value for three entrances qualification ................................................................. 54
4.6 AIC and p-value for covariates selection in Cox PH model ................................. 55
4.7 Analysis of the covariates that contribute to student performance ................................................................. 57
4.8 Scaled Schoenfeld residual of covariates in the Cox model ......................................................... 60
4.9 Log-likelihood, AIC, AIC_C and BIC values of parametric survival models ................................................................. 65
4.10 Log-likelihood, AIC, AIC_C and BIC values of parametric AFT models ......................................................... 66
4.11 Comparison results of best parametric and parametric AFT model ................................................................. 67
4.12 Coefficient estimates, TR and p-value for log-normal AFT model ................................................................. 67
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Research framework on student performance in UTHM</td>
<td>47</td>
</tr>
<tr>
<td>4.1</td>
<td>Survival probability plot by using KM curve</td>
<td>50</td>
</tr>
<tr>
<td>4.2</td>
<td>Survival and failure probability plot for diploma, matriculation and STPM</td>
<td>53</td>
</tr>
<tr>
<td>4.3</td>
<td>Cumulative hazards plot of Cox-Snell residuals</td>
<td>58</td>
</tr>
<tr>
<td>4.4(a)</td>
<td>Scaled Schoenfeld residuals plot for GPA</td>
<td>59</td>
</tr>
<tr>
<td>4.4(b)</td>
<td>Scaled Schoenfeld residuals plot for entrance qualification</td>
<td>59</td>
</tr>
<tr>
<td>4.4(c)</td>
<td>Scaled Schoenfeld residuals plot for course</td>
<td>60</td>
</tr>
<tr>
<td>4.5(a)</td>
<td>Plot of covariates GPA against Martingale residual</td>
<td>61</td>
</tr>
<tr>
<td>4.5(b)</td>
<td>Plot of covariates qualifications against Martingale residual</td>
<td>62</td>
</tr>
<tr>
<td>4.5(c)</td>
<td>Plot of covariates courses against Martingale residual</td>
<td>62</td>
</tr>
<tr>
<td>4.6(a)</td>
<td>Survival probability plot for exponential fitted with KM plot</td>
<td>63</td>
</tr>
<tr>
<td>4.6(b)</td>
<td>Survival probability plot for Weibull fitted with KM plot</td>
<td>64</td>
</tr>
<tr>
<td>4.6(c)</td>
<td>Survival probability plot for log-normal fitted with KM plot</td>
<td>64</td>
</tr>
<tr>
<td>4.6(d)</td>
<td>Survival probability plot for log-logistic fitted with KM plot</td>
<td>65</td>
</tr>
</tbody>
</table>
# LIST OF SYMBOLS AND ABBREVIATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_i$</td>
<td>Coefficients estimate of the covariates</td>
</tr>
<tr>
<td>$e_{ij}$</td>
<td>Expression of proportion groups at risk to fail by number of failure</td>
</tr>
<tr>
<td>$q - 1$</td>
<td>Degrees of freedom</td>
</tr>
<tr>
<td>$i$</td>
<td>Groups or covariates in the study, $i = 1, 2, \ldots, n$</td>
</tr>
<tr>
<td>$j$</td>
<td>Failure time, $j = 1, 2, \ldots, p$</td>
</tr>
<tr>
<td>$k$</td>
<td>Number of constant parameter of each parametric model</td>
</tr>
<tr>
<td>$m_i$</td>
<td>Number of failure that happened at time $t_i$</td>
</tr>
<tr>
<td>$m_{ij}$</td>
<td>Number of failure that happened by groups, $i$ at $j^{th}$ failure time</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of observations in the data set</td>
</tr>
<tr>
<td>$n_i$</td>
<td>Number of observations at risk at the beginning of time period $t_i$</td>
</tr>
<tr>
<td>$n_{ij}$</td>
<td>Number of observations at risk by groups, $i$ at $j^{th}$ failure time</td>
</tr>
<tr>
<td>$q$</td>
<td>Number of covariates in the model</td>
</tr>
<tr>
<td>$t$</td>
<td>Time of study</td>
</tr>
<tr>
<td>$t_i$</td>
<td>A time when an event happened by group in a study</td>
</tr>
<tr>
<td>$t_j$</td>
<td>A time when failure happened at $j^{th}$ failure time</td>
</tr>
<tr>
<td>$T_i$</td>
<td>Log-linear model for the random variable associated with the lifetime of the $i^{th}$ observation in a survival study</td>
</tr>
<tr>
<td>$x_i$</td>
<td>Covariate values in the study by groups</td>
</tr>
<tr>
<td>$X_{ij}$</td>
<td>Covariate values in the study by groups at $j^{th}$ failure time</td>
</tr>
<tr>
<td>$Y_i$</td>
<td>Process determining if the $i^{th}$ subject at risk to fail at time, $t$</td>
</tr>
<tr>
<td>$f(t)$</td>
<td>Probability density function</td>
</tr>
<tr>
<td>$F(t)$</td>
<td>Cumulative distribution functions</td>
</tr>
<tr>
<td>$h(t)$</td>
<td>Hazard function</td>
</tr>
<tr>
<td>$h_0(t)$</td>
<td>Baseline hazard function</td>
</tr>
</tbody>
</table>
$H_0$ - Cumulative baseline hazard function
$\hat{H}_0$ - Estimate values of cumulative baseline hazard function
$S(t)$ - Survival function
$S_0(t)$ - Baseline survival function
$w(t_j)$ - Weight at a time when failure happened at $j^{th}$ failure time
$\lambda$ - Scaling parameter
$\gamma$ - Shape parameter for parametric model
$\theta$ - Location parameter
$\mu$ - Mean log of distribution, intercept parameter for parametric AFT model
$\sigma$ - Shape parameter for parametric AFT model
$\phi$ - Cumulative distribution function of the normal distribution
$\beta$ - Residual value
$\hat{\beta}$ - Estimated residual value
$\delta_i$ - Define status for the event
$\epsilon_l$ - Random variable, assumed to have a particular probability distribution
$\eta_i$ - Linear component of the model
AMO - Academic Management Office
AFT - Accelerated Failure Time
AIC - Akaike Information Criterion
AIC_C - Corrected Akaike Information Criterion
BIC - Bayesian Information Criterion
CeDS - Centre for Diploma Studies (CeDs)
CGS - Centre of Graduate Studies (CGS)
FKAAS - Faculty of Civil and Environmental Engineering
FKEE - Faculty of Electrical and Electronic Engineering
FKMP - Faculty of Mechanical and Manufacturing Engineering
FPTP - Faculty of Technology Management and Business
FPTV - Faculty of Technical and Vocational Education
FSKTM - Faculty of Computer Science and Information Technology
GPA - Grade Point Average
HR - Hazard ratio
KM - Kaplan Meier
PH - Proportional Hazard
PPA - Pejabat Pengurusan Akademik
SPSS - Statistical Package for the Social Sciences
STPM - Malaysian Higher School Certificate
TR - Time ratio
UTHM - Universiti Tun Hussein Onn Malaysia
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Call the variables in R</td>
<td>86</td>
</tr>
<tr>
<td>B</td>
<td>Kaplan Meier analysis by using R</td>
<td>87</td>
</tr>
<tr>
<td>C</td>
<td>Non-parametric statistical test</td>
<td>89</td>
</tr>
<tr>
<td>D</td>
<td>Covariates selection for Cox model</td>
<td>91</td>
</tr>
<tr>
<td>E</td>
<td>Fit Cox model and estimate hazard ratio</td>
<td>94</td>
</tr>
<tr>
<td>F</td>
<td>Testing the Cox Proportional Hazards assumptions</td>
<td>95</td>
</tr>
<tr>
<td>G</td>
<td>Parametric model and Parametric AFT model</td>
<td>97</td>
</tr>
<tr>
<td>H</td>
<td>Estimate time ratio for the best model</td>
<td>101</td>
</tr>
</tbody>
</table>
LIST OF PUBLICATIONS


CHAPTER 1

INTRODUCTION

This chapter introduces the research background that mainly focuses on student performance. Overviews of the survival analysis application along with the significant models that are relevant to the present study are discussed to provide insight of the main critical topic of the study. The problem statement, research objective and research significance are also presented in this chapter.

1.1 Research background

Student performance is an important issue in education especially in higher learning institutions. Kyoshaba (2009) defined student performance as students who are capable to succeed during their studies. Student performance is the outcome of education when students achieved their educational goals. Performance is characterised by performance in a case method course which may assess along various dimensions including class participation, student work on papers and exams, and group activities such as projects and presentations (Steinmayr et al., 2014). Mat et al. (2014) characterised student performance by exam oriented performance and represented by grade point average (GPA) once the students complete the study.

Performance of students can be measured by estimating the covariates of characteristics that affect student performance. In general, covariates are characteristics, factors or explanatory variables that exist within the study. Investigations of covariates on the performance of students have become a topic of growing interest in higher educational circle. Recent studies were carried out to
explore the covariates that affected university student performance. These covariates may range from student education background (Boghikian-Whitby & Mortagy, 2008), previous school prior to admission into a programme (Kyoshaba, 2009), GPA (Bruinsma & Jansen, 2009) and various entry qualifications obtained by the students for admission into a higher academic programme (Amasuomo, 2014). The other covariates that contribute to student performance are geographical background (Hovdhaugen, 2011), gender (Zuilkowski & Jukes, 2014) and study habits (Shahzadi & Ahmad, 2011).

Study on student performance can be conducted using various statistical techniques including regression analysis (Hijazi & Naqvi, 2006), categorical regression analysis (Cilan & Can, 2014), data mining techniques of the hierarchical linear modelling (Miller et al., 2015), and meta-analysis approach (Chang, Jian & Chou, 2016). These studies were carried out to identify the most significant factors or covariates affecting the student performance and identify any risk of failure among students during their study. Survival techniques have applied in education to focus on student failure rate in higher education study program (Paura & Arhipova, 2014). The study found that survival analysis technique showed excellent estimation of student performance by applying time together with discrete or continuous independent variables. Survival analysis also has binary outcome variables that can be used to detect the status of students during the study.

In survival analysis, Kaplan-Meier (KM) method is a simple method (Goel, Khanna & Kishore, 2010) as it has no ability to test or investigate the covariates effect on student performance. KM method is limited to analyse one covariate in each study. However, existence of many covariates in a study may affect the outcome of the study (Clark et al., 2003). The survival analysis proposes the use of the Cox model to estimate the hazard ratio (HR) of each covariate. The estimation in Cox model is biased and the results derived from the model may be erroneous when the proportional hazard (PH) assumption in the Cox model is violated (Persson & Khamis, 2005). The biased estimation is likely because a PH assumption in Cox model is the HR for each covariate and it is always assumed as constant overtime.

Violation of PH assumption causes the unreliable and biased results from the Cox model. Therefore, Accelerated Failure Time (AFT) models such as exponential, Weibull, log-logistic and log-normal are better choices to apply in the study. As AFT models consider a statistical distribution for survival time and do not rely on PH
assumptions, they are deemed as suitable alternatives to Cox model (Zare et al., 2015). In order to overcome the weaknesses in Cox model, the parametric AFT models is applied in the study to estimate the effect of covariates on accelerating or decelerating the student survival or known as time ratio (TR).

The aims of this study are to propose the best survival model to model and formulate a good conclusion to the covariates that exist in the study of student performance. Besides, this study also estimates the HR and TR of covariates and determines the effect of the covariates to student performance. The novelty of this research focuses on suggesting various analytical techniques on student performance by using survival techniques such as KM method, Cox model and parametric AFT model.

1.2 Problem statement

Study on student performance comes in various techniques such as regression analysis, categorical regression analysis, data mining techniques and meta-analysis approach. The techniques that are applied in previous studies of student performance are limited to only one categorical variable and numeric variable for covariate (Cilan & Can, 2014). The missing data are ignored in the calculation of student performance (Miller et al., 2015). Limitations from previous studies of student performance prompted the study to use the survival techniques. Survival technique is the simplest method to find the survival rate or student performance without ignorant the missing data.

The benefit of the non-parametric via KM method in survival techniques lies in its flexibility to include information about students who are lost at any point in time which can be used to estimate the survival rates (Leonavicius, 2009). Various studies have been explored on the estimate of student performance according to the demographic characteristic (Zuilkowski & Jukes, 2014) and geographical background (Hovdhaugen, 2011). However, within the context of the effect of covariates proposed in the study of student performance in UTHM, very little studies in the literature applied have been dedicated to study the effect of entrance qualification as an important covariate and estimate survival rate in the research based on the covariate.

Estimate survival rate in the study of student performance using KM method was not able to estimate the multiple covariates used in the study at a time of analysis.
By using semi-parametric model via Cox model, the effect of covariates on the performance of student can be investigated. Thus, many covariates can be utilised to seek information within the sets of data pertaining the student at risk to fail or calculation of students HR. Previous study has been considered type of school, gender, age and GPA covariates together in the study of student performance (Bruinsma & Jansen, 2009). Gender, grades and course covariates have employed by Min et al. (2011) while Box-Steffensmeir et al. (2015) have considered faculty and gender covariates in the study of student performance. No studies in the literature have combined the faculty, course and student grades covariates in the study of student performance. Incorporation of these variables prompts the question on which covariates dominate the performance of students and can be utilised for further analysis.

Estimate of student at risk to fail or a HR of students through the existing covariates demonstrates the expectations about the future developments of student performance. A related question is how the Cox model can provide an accurate result for this estimated based for the covariates. By using semi-parametric model via PH assumption through the residual analysis, the performance of Cox model in analysing student performance can be investigated. When the PH assumption in Cox model is violated, the result provided by Cox model is considered as less accurate (Baghestani et al., 2015).

The parametric models have many distributions that can be utilised in the survival analysis. Very few studies in the literature have applied the parametric survival model and parametric AFT model in accessing student performance. Hence, it is questionable which of the models gives a good fitted data and can be used for further analysis. The best model will be selected to further study by using model selection method to select the parametric and the parametric AFT models.

Estimate of students’ ability to survive or TR of students through significant covariates will increase the validation of the study. Henceforth, the study will need to ascertain which the best model can estimate the TR of each significant covariate to conclude the overall study. The TR on student performance can be estimated by using the best model. Therefore, the aim of this study is to propose the best alternative model in evaluating the student performance.
1.3 Research objective

This study aims to achieve the following objectives:

1. To estimate the survival rates of students with different entrance qualifications using KM method.
2. To identify the covariates that dominate the student performance via Cox model.
3. To investigate the performance of Cox model based on the violation of the PH assumptions.
4. To compare the model performance by using the parametric survival and parametric AFT models.
5. To estimate the time ratio (TR) of student performance in accordance to the best model selected in Objective 4.

1.4 Research scope

This research involves Bachelor’s degree students in UTHM. Total numbers of Bachelor’s degree students in UTHM of 2011/2012 intake are 2606 based on standard entrance qualifications that are diploma, matriculation and STPM. All 2606 Bachelor’s degree students for 2011/2012 intake are collected from the Academic Management Office (AMO) in UTHM and analysed by using Statistical Package for the Social Sciences (SPSS) and R statistical software. In 2011, UTHM only have six faculties for Bachelor’s degree students namely the Faculty of Technical and Vocational Education (FPTV), Faculty of Electrical and Electronic Engineering (FKEE), Faculty of Civil and Environmental Engineering (FKAAS), Faculty of Mechanical and Manufacturing Engineering (FKMP), Faculty of Computer Science and Information Technology (FSKTM) and Faculty of Technology Management and Business (FPTP). All Bachelor’s Degree students from these six faculties for 2011/2012 intake are involved in the study of student performance. Centre for Diploma Studies (CeDS) and Centre of Graduate Studies (CGS) are excluded since this study focused only on Bachelor’s Degree students.

This study focuses on the potential of student failure in each semester. The event of interest in this study is student failure with GPA of less than 2.00 for the current semester of study. Students are considered as censored if they achieved any GPA value
that is greater or equal to 2.00. The dependent variable is the time of the event of interest occurs. It is defined as 1 for the event status with GPA less than 2.00 and 0 if it is censored with GPA greater and equal to 2.00. The independent variables are the entrance qualification, faculty, course and the GPA of students.

This study considered several survival approaches including non-parametric, semi-parametric and parametric, for example KM method, Cox model, exponential model, Weibull model, log-logistic model, log-normal model, exponential AFT model, Weibull AFT model, log-logistic AFT model and log-normal AFT model.

1.5 Research significance

This research aims to estimate the survival rate of students based on different entrance qualification. The Cox model applied design to identify the covariates that dominate the student performance. This study will investigate the performance of the Cox model based on PH assumption, whether the PH assumption in the Cox model is accepted or violated. The parametric model involved is intended to overcome the limitations encountered by the Cox model. The parametric AFT model is the extension method of the parametric survival model that is capable to estimate the TR for significant covariates in the study. Finally, this study plans to compare the student performance by using parametric survival and parametric AFT model in a way to propose an alternative survival model to evaluate the student performance.

1.6 Thesis organization

Chapter 2 which provides a literature review on student performance and survival analysis. Chapter 3 presents the data set and its description where the research methodologies that explain well before performed the survival analysis. Chapter 4 summarises and discusses the overall results. Chapter 5 discusses and concludes the main finding of the study, highlights the research contribution and limitations, and provides recommendations for future study on the similar topic.
CHAPTER 2

LITERATURE REVIEW

This chapter discusses the previous studies on student performance and the methods that have been used in the survival analysis. It is also provides the reader with the survival techniques that have been applied to the higher education and other sectors. The first part of this chapter gives a brief description of the past and current studies conducted on student performance followed by a brief review of various statistical methods. Finally, the research gap on the application of statistical techniques in education is discussed in this chapter.

2.1 Student performance

Hijazi & Naqvi (2006) carried out a case study on factors affecting student performance and found that study hour was not a significant factor as far as student performance is concerned. Student performance may depend on intelligence level, intellect, memory or learning method of the students. Meanwhile, Rasul & Bukhsh (2011) examined a relationship among psychological, physical, socio-economic and educational factors on the performance of student in examination. The study found that most of the psychological, physical, socio-economic and educational factors affected the performance of students in examination at university level.

Higher education institutions have always greatly emphasised on the factors governing the performance of students. Lesik (2007) performed a study to identify a casual impact of education programs on student performance. He stated that most of the existing research has been non-experimental and focused exclusively on student background and demographic characteristics to study student success. Belanger & Liu
also agreed that household characterised with high income impacted student performance. Cilan & Can (2014) conducted a study to measure the factors affecting student academic performance and found that demographic factors, which are age and student grades, are the factors that significantly affected students’ academic success. These studies offered a framework that student background, demographic characteristics, and student grades have causal impact on student success.

Shahzadi & Ahmad (2011) found that academic interaction with teachers and peers had significant effect on the student academic performance. Home environment and study habits were also important factors in the student academic performance at university level. Students must improve their study habits especially allocating proper study time every day and revise the lecture notes on daily basis. The research concluded that better study habits prompted higher academic interaction and ultimately led to higher academic performance of students. Chang et al. (2016) conducted a study to identify the effects of student engagement on academic achievement and revealed strong relationship existing between student behavioural, student engagement and academic achievement.

The academic achievements determined the factors that contributed to student performance in the university. Tianqi & Ganges (1995) conducted a study to examine the performance of students. The study included student gender, ethnicity, graduation date and GPA as the determinants and found that students’ gender and ethnicity affected on student performance. The second, third and fourth semesters are always that most critical periods for the students to leave their study. In 1999, Murtaugh, Burns & Schuster (1999) conducted a study on undergraduate students by considering the students’ college residency, ethnicity, age and GPA covariates to examine the performance of the students. Age, high GPA and non-residents college students were shown to have higher performance rates than the resident and international students. This analysis helped the university to guide and improve student performance in the university.

Leonavicius (2009) also conducted a study focusing on student performance and student retention. This study used gender and entrance grade of each student as the covariates. The study found that gender is not a significant factor for study on student performance and student retention. However, Bruinsma & Jansen (2009) revealed that 61% of the students in their study passed their first year examination within 24 months, with female students passing it earlier than male students and those...
with higher GPA passed their examination earlier than students with lower GPA. Min et al. (2011) focused on the survival analysis of the loss rate of undergraduate students. This study included grades, genders, course of each student, and found the existence of statistical differences between levels of grades received by different courses. Additionally, female students tend to leave engineering course study earlier than their male counterparts.

Paura & Arhipova (2014) analysed the student performance rate in higher education study program by using genders, grades and financial source covariates. The results of this study showed that the main reasons for student non-perform were student low grades during the study and low motivation to continue the study. In 2015, a research on student performance was conducted at the university level by using the year of study (Kartal & Ozan, 2015) and faculty (Box-Steffensmeier et al., 2015) covariates. The study showed that performance of students neither depended on the year of study nor the faculty that the students belonged to.

### 2.2 Statistical techniques applied to study on student performance

Various methods have been explored regarding the study on student performance such as regression analysis (Hijazi & Naqvi, 2006), mean score (Rasul & Bukhsh, 2011), confirmatory factor analysis (Shahzadi & Ahmad, 2011), categorical regression analysis (Cilan & Can, 2014) and meta-analysis approach (Chang et al., 2016).

Regression analysis was used to explore the factors that affected student performance. Hijazi & Naqvi (2006) applied regression analysis for the study on student performance, dealing with student attitude towards study, attendance, home environment and previous academic achievements as the covariates. The study concluded that the student attitude towards study was positively related to student performance.

Rasul & Bukhsh (2011) applied the mean score to determine the performance of students in examination considering the examination is an integral part of the education system. Examination is useful as it measures a student progress and acts as a process to test the abilities or achievement of the student in any area of academic programs. Examination system has many factors that directly and indirectly affect student performance. Shahzadi & Ahmad (2011) applied the confirmatory factor analysis and developed structural equation models to access the student academic
REFERENCES


