NON-PARAMETRIC SURVIVAL MODELLING OF TIME TO EMPLOYMENT AMONGST 09/10 COHORT OF MATHEMATICS GRADUATES

NOR LIYANA BINTI SABARI

UNIVERSITI TEKNOLOGI MALAYSIA

NON-PARAMETRIC SURVIVAL MODELLING OF TIME TO EMPLOYMENT AMONGST 09/10 COHORT OF MATHEMATICS GRADUATES

NOR LIYANA BINTI SABARI

The thesis is submitted in fulfillment of the requirements for the award of the degree of Master of Science (Mathematics)

Faculty of Science

Universiti teknologi Malaysia

JANUARY 2014

To my beloved family

Sabari bin Md. Yassin, Zainab bt Atan,

Apai, Ajim, and Adib, infinitely supportive

AKNOWLEDGEMENT

Alhamdulillah, all praises to Allah for the strengths and His blessing in completing this research. Special appreciation goes to my supervisor, Dr. Zarina binti Mohd Khalid, for her supervision and constant support. The co-operation is much indeed appreciated.

Sincere thanks to all my friends' especially Firman, Miza, Amirah and Azira for their kindness and moral support during my study. Thanks for the friendship and memories.

Last but not least, my deepest gratitude goes to my beloved parents, Mr. Sabari bin Md. Yassin and Mrs. Zainab bt Atan for their endless love, prayers and encouragement. To those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

ABSTRACT

The length of waiting time to employment may be considered as the determinant of the employability of a graduate. Hence, this project aims to study the behaviour of waiting time to employment of mathematics graduates. In particular, this study analyzes the differentials of the factors levels as well as identifies the factors that may influence the waiting time to employment by applying the survival analysis techniques. This study involves 40 mathematics graduates from Department of Mathematical Sciences who graduated in October 2012 at Universiti Teknologi Malaysia. In this project, Kaplan-Meier estimator is used to estimate the waiting time to employment of the graduates while the logrank test is used to compare the difference between two or more groups of factors influencing the employability. Moreover, the other method which is Cox proportional hazard regression model is used to examine the relationship between the variables (covariates) with the hazard. There is evidence that residential, English communication skill and confidence level were the factors that influence the employability of the mathematics graduates.

ABSTRAK

Tempoh masa menunggu sehingga mendapat kerja boleh dianggap sebagai penentu kebolehpasaran seseorang graduan. Oleh itu, projek ini bertujuan untuk mengkaji tabiat masa menunggu sehingga mendapat kerja graduan matematik. Secara amnya, kajian ini menganalisis perbezaan paras faktor dan juga mengenal pasti faktor-faktor yang mungkin mempengaruhi masa menunggu sehingga mendapat kerja dengan menggunakan teknik-teknik analisis survival. Kajian ini melibatkan 40 graduan matematik daripada Jabatan Sains Matematik yang telah tamat pengajian pada Oktober 2012 di Universiti Teknologi Malaysia. Dalam projek ini, penganggar Kaplan-Meier digunakan untuk menganggar masa menunggu sehingga mendapat kerja para graduan manakala ujian logrank digunakan untuk membandingkan perbezaan antara dua atau lebih kumpulan faktor yang mempengaruhi pekerjaan itu. Selain itu, kaedah lain iaitu model regresi Cox risiko proporsional digunakan untuk mengkaji hubungan di antara pembolehubah (kovariat) dengan risiko bahaya. Terdapat bukti bahawa kediaman, kemahiran komunikasi Bahasa Inggeris dan tahap keyakinan merupakan faktor yang mempengaruhi kebolehpasaran graduan matematik.

TABLE OF CONTENTS

REFERENCES

APPENDIXES

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF SYMBOLS

T	-	Random variable for survival time of a person or subject
t	-	Time
$S(t)$	-	Survival function at time *t*
$h(t)$	-	Hazard function at time *t*
$f(t)$	-	Probability density function at time *t*
$F(t)$	-	Cumulative distribution function at time *t*
n	-	Number of sample
d	-	Number of failure
$H(t)$	-	Cumulative hazard function at time *t*
O_i	-	Total number of events occur in group *i*
E_i	-	Total number of expected events in group *i*
$h_0(t)$	-	Baseline hazard at time *t*
β	-	Regression coefficient
$SE(\beta)$	-	Standard error for the regression coefficient

CHAPTER 1

INTRODUCTION

1.1 Introduction

Statistical methods for survival data analysis is one branch of mathematics that have continued to flourish in the last two decades. Its historical use in cancer and reliability research to business, criminology, epidemiology, and social behavioral sciences has widened the application of these methods (Lee and Wang, 2003). Moreover, biomedical researchers, consulting statisticians, economists and epidemiologists also use survival time study most extensively nowadays. This study deals with statistical methods for analyzing survival data developed from laboratory study of animals, clinical and epidemiologic studies of humans and other appropriate applications (Lee and Wang, 2003).

Based on Kleinbaum (1996), survival analysis is a collection of statistical procedures for data analysis for which the time until an event occurs become the outcome variable of interest. This means that the subjects are tracked until the event occurs (failure) or we lose the subjects from the sample (censored observations). The interest of the survival analysis is in observing how long the subjects stay in the sample and their risk of failure is analyzed.

1.2 Background of the Study

Nowadays, issues on unemployment are becoming increasingly serious in Malaysia. Based on the statistics of graduates in the labour force Malaysia 2011 analyzed by the Department of Statistics, Malaysia (2011), the number of unemployed degree holders is between 2,700 to 8,200 persons from 1982 until 1997. However, the number has shown a significant increase between 12,000 to 33,800 persons since 1998 until 2010. To make things worse, many students have graduated every year from institutions of higher learning either in public or private and these numbers will keep on adding.

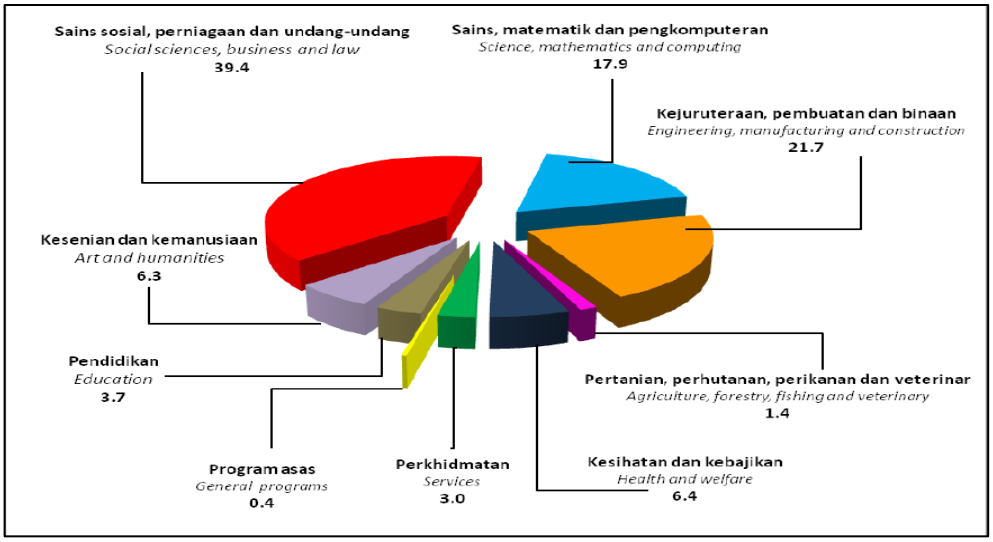

Figure 1.1 Percentage distributions of unemployed graduates by selected field of study, Malaysia, 2010. (Department of Statistics, Malaysia, 2011)

In term of field of study, majority of unemployed graduates were in social science, business and law (39.4%), followed by engineering, manufacturing and construction (21.7%) while for science, mathematics and computing fields unemployment percentage is 17.9 percent in 2010 (Department of Statistics, Malaysia, 2011).

Consequently, there exist a lot of competitions among graduates in obtaining jobs and it is more difficult for the graduates who lack interpersonal skills and low qualifications to get employment. Employers on the other hand tend to hire graduates with the best quality. Therefore, authorities of higher learning institutions, either in private or government sectors should play their part in producing the best quality of graduates that have high employability attributes. Besides that, future graduates also need to take this issue seriously and understand the salient factors that influence unemployment among graduates. They need to strive for success and prepare themselves with the skills needed for employment.

1.3 Statement of the Problem

As stated in statistics of graduates in labor force Malaysia 2011, science, mathematics and computing fields was included in high percentage of unemployment of graduates (17.9%) from nine fields of study (Department of Statistics, Malaysia, 2011), thus this work will be highly useful for the Department of Mathematical Sciences in order to know the employability of their graduates towards employment after graduation.

First issue that been taken to consideration in this study is the time taken by the mathematics graduates to get employment after their graduation. However, in many situations we want to do more than just estimate the survival times for a single group. Since the factors (variables) of interest involve in this study are categorical variables, then it has two or more levels that been classified within the factors. Hence we want to check whether or not the difference between the factor levels for each factor gives significance impact on the time to employment. Besides that, significant factors that influence the employability of the graduates also need to be identified throughout this study.

1.4 Objectives of the Study

The following are the objectives of the study:

1. To estimate the time-to-employment experiences of mathematics graduates using Kaplan-Meier approach.
2. To compare survival curves for different factor levels for each factor using logrank test.
3. To identify significant factors influencing the employability of mathematics graduates using Cox Proportional Hazard Regression model.

1.5 Scope of the Study

Collected data for this study involves data from graduates of the Department of Mathematical Science of Faculty of Science of UTM who had undergone a 3-years mathematics programme; Bachelor in Science (Mathematics) that is SSE and Bachelor in Science (Industrial Mathematics) that is SSM. The respondents involved in this study graduated in October 2012 and must be registered for three years mathematic programs, SSE and SSM in 2009/2010 session. 109 respondents have been followed up from October 2012 to September 2013 in order to participate in the study. This study only focuses on non-parametric methods which are Kaplan-Meier method and logrank test in analyzing the survival curves. Furthermore, the usage of Cox proportional hazard regression model, a semi-parametric model has also been included in this study to model the survival data. Besides that, we only considered the right-censored data in this study and the censoring was assumed to be uninformative.

1.6 Significance of the Study

The results from this study can be used by the Department of Mathematical Sciences of UTM in order to see the implementation of both programs in producing a employable graduates. Moreover, this study also can help future graduates in preparing themselves with the skills needed for employment and strive for success.

REFERENCES

Akram, M. Aman Ullah, M., and Taj, R. (2007). Survival Analysis of Cancer Patients Using Parametric and Non-Parametric Approaches. *Pakistan Vet. J.* 27(4): 194-198.

Clack, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003a). Survival Analysis Part I: Basic Concepts and First Analyses. *British Journal of Cancer*. 89: 232-238.

Clack, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003b). Survival Analysis Part II: Multivariate Data Analysis – An Introduction to Concepts and Methods. *British Journal of Cancer*. 89: 431-436.

Clack, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003c). Survival Analysis Part III: Multivariate Data Analysis – Choosing a Model and Assessing Its Adequacy and Fit. *British Journal of Cancer*. 89: 605-611.

Cox, D. R., and Oakes, D. (1984). *Analysis of Survival Data.* New York: Chapman and Hall.

Department of Statistic Malaysia. *Statistics of Graduates in the Labour Force Malaysia 2011*. November 2011.

Flynn, R. (2012). Approaches for Data Analysis: Survival Analysis. *Journal of Clinical Nursing*. 21: 2789-2797.

Goel, M. K., Khanna, P., and Kishore, J. (2010). Understanding Survival Analysis: Kaplan-Meier Estimate. *International Journal of Ayurveda Research*. 1(4): 274-278.

Hoon, T. S. (2008). Using Kaplan Meier and Cox Regression in Survival Analysis: An Example. *ESTEEM*. 4(2): 3-14.

Kleinbaum, D. G. (1996). *Survival Analysis: A Self Learning Text.* New York: Springer-Verlag.

Kolb, J. and Werwatz, A. (2001). The Duration of Marginal Employment in West Germany: A Survival Analysis Based on Spell Data. *Quarterly Journal of Economic Research.* 70(1): 95-101.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* (2nd. Ed.) New Jersey: John Wiley & Sons.

Lee, E. T. and Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis.* (3rd. ed.) New York: John Wiley & Sons.

Machin, D., Cheung, Y. B. and Parmar, M. K. B. (2006). *Survival Analysis: A Practical Approach.* England: John Wiley & Sons.

Mackes, M. (2005). Becoming an Employer of Choice. *ASHRAE Journal.* 47(11): 2-3.

Norusis, M. J. (2004). *SPSS 13.0 Advanced Statistical Procedures Companion.* New Jersey: Prentice Hall.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., Hiller, W., Fisher, E. R., Wickerham, D. L., Bryant, J., and Wolmark, N. (2004). A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *The New England Journal of Medicine.* 351(27): 2817-2825.

Persson, I. (2002). *Essays on the Assumption of Proportional Hazards in Cox Regression.* Sweden: Uppsala University, Tryck & Medier.

Singh, N. S., Singh, N. S. and Narendra, R. K. (2011). Survival Analysis of Duration of Waiting Time to Conception. *Electronic Journal of Applied Statistical Analysis.* 4(2): 144-145-154.

Smith, T., Smith, B. and Ryan, M. A. K. (2003). Survival Analysis Using Cox Proportional Hazards Modeling for Single and Multiple Event Time Data. *Statistics and Data Analysis. SUGI28.* Paper 254-28. Retrieved April 24th, 2013 from http://medicine.yale.edu/labs/ma/www/BIS643/254-28.pdf.

Somers, M. J. (1996). Modelling Employee Withdrawal Behaviour over Time: A Study of Turnover Using Survival Analysis. *Journal of Occupational and Organizational Psychology*. 69: 315-326.

Stolberg, H. O., Norman, G., and Trop, I. (2005). Survival Analysis. *American Journal of Roentgenology*. 185: 19-22.

Taylor, M. P. (1999). Survival of the Fittest? An Analysis of Self-Employment Duration in Britain. *The Economic Journal*. 109 (March): 140-155.