

Survival Analysis on Undergraduate Students' Persistence:

Case study

Mureithi
Department of Mathematics
Catholic University of Eastern
Africa Nairobi, Kenya.
*email:lawrence.mureithim@g
mail.com*

Dr. Karanjah Anthony N.
Department of Mathematics,
School of Science
Maasai Mara University
Narok, Kenya.
*email:karanjah@mmarau.ac.k
e*

Dr. F. Njui
School of Mathematics
University of Nairobi
Nairobi, Kenya.
emai:lnjui.francis@gmail.com

Abstract— A longitudinal study on 1,998 undergraduate students in the four-year program was conducted in one of the universities in Kenya. These were students who had been admitted in the years 2012, 2013 and 2014. Data collected was for 64 months, from January 1, 2012 to May 27, 2017. Survival analysis was used to model and analyze the persistence of these students at the university. It was found that majority of the students who withdrew from the university did so in the first 12 months of study. There was a significant difference in the persistence of the male and female students, with the female students having a higher survival rate than the male students. It was also found that the persistence of the Kenyan and Non-Kenyan students was significantly different, with the Kenyan students having a higher survival rate than the Non-Kenyan students.

Gender had a significant effect on the survival of these students with the risk of a male student withdrawing being 2 times higher than that of a female student.

Keywords- *Survival Analysis, Survival rate, survival function, hazard, Persistence*

1. INTRODUCTION

There has been a rapid increase of the number of universities in Kenya. This is due to increased access of university education. Records from the Commission of University Education (CUE) indicate that as of the year 2016, there were 33 registered public universities and 36 private ones with a total undergraduate student population of 475,750.

It has become evident that not every student who enrolls for university education graduates. There are those who fall on the way. The study of undergraduate student retention developed as a result of trying to explain the phenomenon of student attrition. *Bean, (1979)* defined student attrition as the cessation of individual student membership in an institution of higher education. *Swail (2004)* defined undergraduate student retention as an institution of higher education's ability to retain a student from admission until graduation.

A case study was carried out in one of the oldest universities in Kenya. Study on undergraduate student retention in the institution was found necessary. There are several models used in the analysis of such retention but in this paper, survival analysis has been used.

2. SURVIVAL ANALYSIS

Survival analysis is a collection of procedures for data analysis for which the outcome variable of interest is time until an event occurs. This statistical methodology accommodates the censoring of individuals in its analysis. In essence censoring occurs when some information on individual survival time is known, but the exact survival time is unknown, *Kleinbaum &*

Klein (2005). A longitudinal research design was used to study the persistence of the 1,998 undergraduate students in the four-year program who were admitted in the years 2012, 2013 and 2014. The period of investigation was 64 months beginning from January, 2012 to May 27, 2017. The survival time of each student was measured in months, from the time a student was admitted to the university until this student withdrew from the university without having graduated. The survival times of the students who had graduated within the 64 month period of investigation were censored at their times of graduation. The survival times of the students who quietly withdrew within this 64-month period were censored at the times they withdrew. The survival times of the students who had officially deferred their studies were censored at the times of withdrawal. Survival analysis was used to estimate the survival functions of these undergraduate students using the Kaplan-Meier product limit estimates of the survival function, *Kaplan & Meier (1958)*. The log-rank test, *Peto and Peto (1972)* was used to test the significance of the differences in the survival of these undergraduate students by gender, Nationality, method and faculty of study at 5% level of significance. The Cox proportional hazards regression model, *Cox (1972)* was fitted to the four-year program undergraduate students' data set to assess the relationships between the covariates of survival. The Breslow approximation, *Breslow (1975)* was then used to estimate partial maximum likelihood estimates of the regression coefficients of the Cox proportional hazards

regression model. This approximation works well and is implemented in most statistical package.

The main assumption of the Cox proportional hazards regression model is that the hazard ratio comparing any two specifications of predictors is constant over time. This assumption was evaluated on each covariate in the model using Schoenfeld residuals, *Schoenfeld (1982)*.

3. DATA

Secondary data on the gender, Nationality, method of study, and faculty of study were obtained from the office of the registrar. The data had been organized in MS EXCEL spreadsheet format. The categories of Nationality were Kenyan and Non-Kenyan. Students were categorized as full-time, part-time or School-based holiday students. Various disciplines of study were Science, Education, Arts and Social Sciences, Law and Business studies. Available data also contained the dates of admission for the 1,998 undergraduate students in the four-year program who were admitted in the years 2012, 2013 and 2014. There were nine graduation ceremonies within the 64-month period.

Primary data on the times of withdrawal or deferment were obtained through mobile phone interviews with the students who had not graduated from the university within the study period. Those students who could not be reached using any means were considered as drop-outs.

4. RESULTS

R statistical software package for data analysis and graphics was used to analyze the four-year program undergraduate student persistence data. Table 1 shows the Kaplan-Meier product limit estimates of the survival function for the entire cohort of the 1,998 students.

Times to withdrawal (months)	Number at risk	Number withdrew	Survival probability estimates	Standard error	95 % Confidence interval	
					LCL	UCL
5	1998	51	0.974	0.00353	0.968	0.981
8	1945	48	0.950	0.00486	0.941	0.960
12	1897	50	0.925	0.00588	0.914	0.937
13	1847	1	0.925	0.00590	0.913	0.937
17	1845	32	0.909	0.00644	0.896	0.922
20	1803	36	0.891	0.00699	0.877	0.904
24	1758	12	0.885	0.00716	0.871	0.899
29	1731	10	0.879	0.00730	0.865	0.894
32	1706	24	0.867	0.00762	0.852	0.882
36	1292	7	0.862	0.00778	0.847	0.878
41	736	3	0.859	0.00801	0.843	0.875

Table 1: Kaplan-Meier product limit survival estimates for the entire cohort of students.

A total 274 of students withdrew from the University without graduating in the 64-month study period. The highest proportion of students who withdrew from the university did so in the first 12 months of study, which is the first 3 trimesters of study. The Kaplan-Meier product limit survival estimate at the close of study is 86%. The survival curve for the 1,998 students is shown in figure 1.

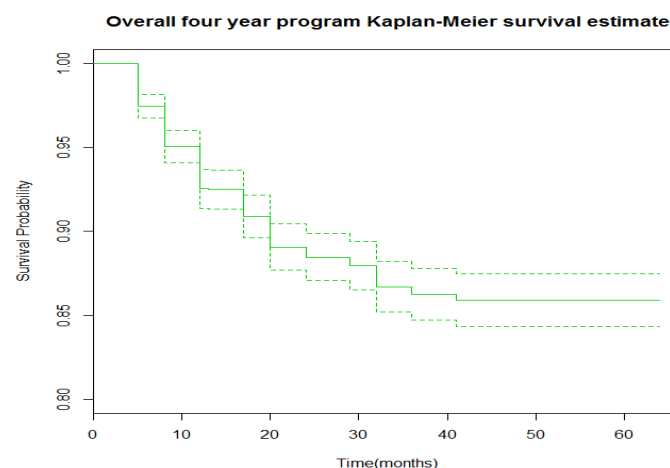


Figure1: The survival curve of the K-M product limit estimate of the overall survival function.

The survival curves for the female and male students are illustrated in figure 2.

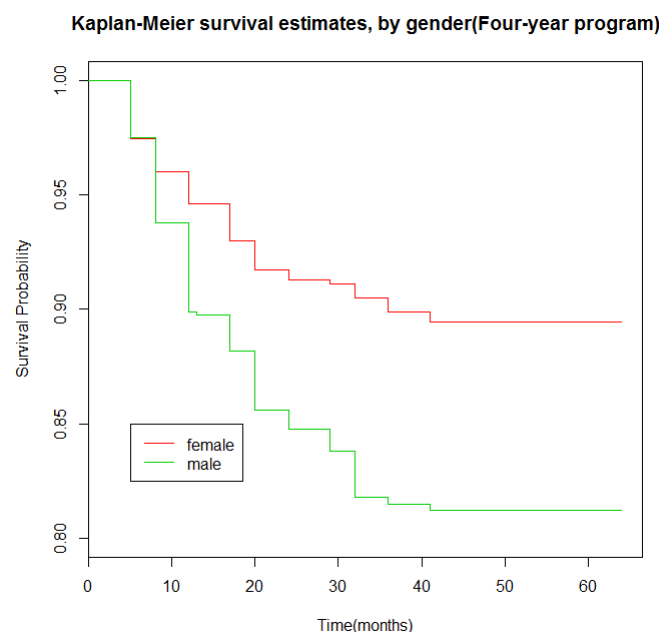


Figure 2: Survival curves for the male and female students.

As the figure shows, the female students have a higher survival rate than the male students and the difference in their survival is highly significant with a p-value of 0.000000106 at 5% level of significance.

The survival curves for the Kenyan and Non-Kenyan students are shown in figure 3 below. The Kenyan students have a

higher survival rate than the Non-Kenyan students. The difference in their survival is significant with a p-value of 0.0486 at 5% level of significance.

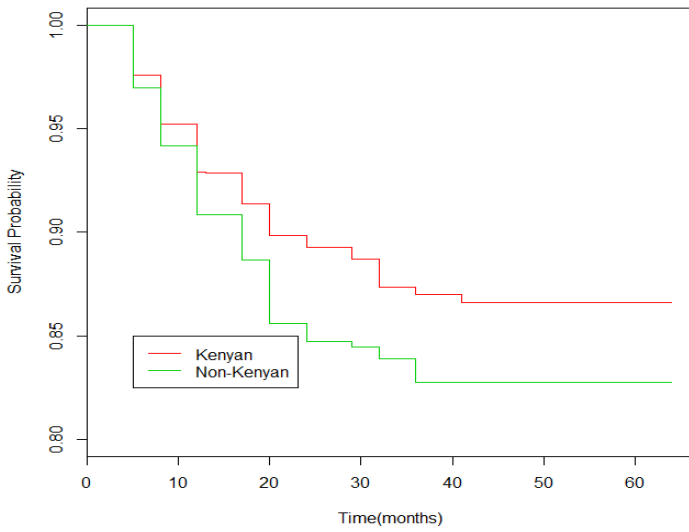


Figure 3: survival curves for the Kenyan and Non-Kenyan students.

There were no significant differences in the persistence of these students as far as the categorization of students and the various study disciplines of study were concerned at 5% level of significance with p-values of 0.127 and 0.0648 respectively.

Figures 4 and 5 illustrate these survival curves.

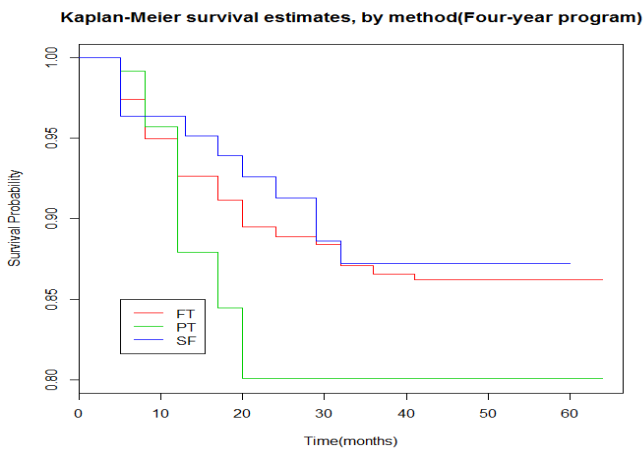


Figure 4: The survival curves of these undergraduate students by method of study.

The survival curves of these undergraduate students by faculty of study are illustrated in the following figure 5.

Kaplan-Meier survival estimates, by faculty(Four-year program)

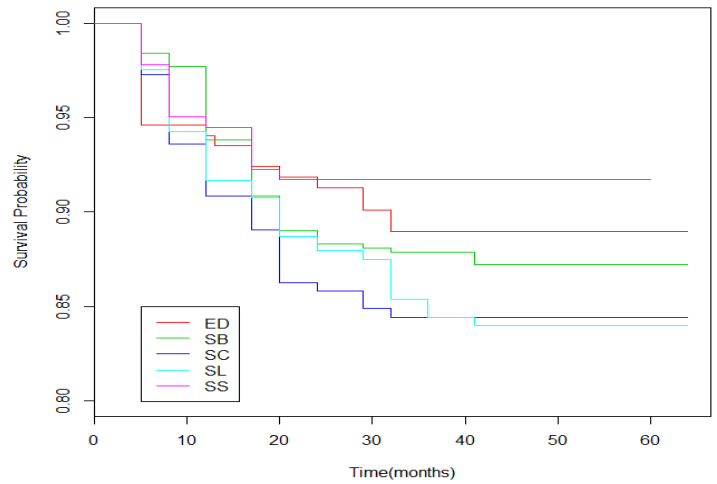


Figure 5: The survival curves of these undergraduate students by the faculty of study.

The results of fitting the Cox proportional hazards regression model to the four-year program undergraduate student data set suggest that only the predictor variable Gender (male) had a highly significant effect on the persistence of these undergraduate students at 5% level of significance, with a p-value of 0.00000039.

The results for evaluating the proportional hazards assumption for the covariate gender indicate that this assumption was not violated in the entire period of study, with a p-value of 0.10440 at 5% level of significance.

The estimated hazard ratio for this covariate is 1.90408, which means that the hazard rate for a male student withdrawing without graduating was almost 2 times higher than that for a female student in the study period.

5. CONCLUSION

The results of the survival analysis of the four-year program undergraduate student persistence data will indeed be valuable not only to the prospective or on-going students, but also to the university's management. This study through the use of survival analysis techniques has brought to light the fact that most students who withdrew from the university without graduating did so within the first 12 months of study, even though the estimated survival probability at this time was 92.5%. This study also revealed that the female students had a higher survival rate than the male students, and the difference in the survival of these two groups of undergraduate students was highly significant at 5% level of significance. It was also found that the Kenyan students had a higher survival rate than the Non-Kenyan students and the difference in the survival of these two groups of students was also significant at 5% level of significance. This study has shown that at 5% level of significance, the risk of a male student withdrawing was

nearly 2 times higher than that of a female student in the entire study period.

REFERENCES

1. Bean, J. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, 12(2), 155-187.
2. Breslow, N E. (1975). Analysis of Survival data under the Proportional Hazard Models. *International Statistical Review*. 43 (1); 45-57
3. Cox, D.R. 1972. Regression Models and life tales. *Journal of the Royal Statistical Society* 34 (Series B):187-202.
4. Demetriou, C. & Schmitz-Sciborski, A. 2011. Integration, motivation, strengths and optimism: Retention theories past, present and future. In R. Hayes (Ed.), *proceedings of the 7th National Symposium on Student Retention, 2011, Charleston*. (pp. 300-312). Norman, OK: The University of Oklahoma.
5. Kaplan, E. L., and P. Meier. 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, 55: 457-481.
6. Kleinbaum, D. G., & Klein, M. 2005. *Survival Analysis: A Self-Learning Text* (2nd Ed.). New York: Springer.
7. Schoenfeld, D., 1982. Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 7, pp. 145-153.
8. Swail, W. S. 2004. *The art of student retention: A handbook for practitioners and administrators*. Austin, TX: Educational Policy Institute.
9. The Commission for University Education (CUE), (2016). *State of University Education in Kenya*.
10. Peto, R., and Peto, J. (1972). Asymptotically efficient rank invariant procedures. *Journal of the Royal Statistical Society, Series A* 135: 185-207.
11. Kalbfleisch, J. and Prentice, R. 1980. *The statistical analysis of failure time data*. New York: Wiley.