

Abstract

The study explores the application of statistical methods to determine factors influencing learners to commit offences at school. Focus is on the time taken to commit a second offence and the number of repeat offences committed by learners after the first offence. The survival time from the first offence to a second offence is analysed by using Kaplan-Meier estimate of the survival function, the tests for equality of survivor functions, the Cox proportional hazards model, and the parametric hazards models. The number of repeat offences is analysed by fitting the Poisson regression model and Negative Binomial regression model. As many learners are expected not to offend again, the Zero Inflated Poisson (ZIP) model is also fitted to determine which factors influence learners not to commit a repeat offence. Factors that are related to the school environment were used in the study of learner offence. The categorical variables are learner grade (class), gender, home location, parental involvement, repeating a grade, suspected substance abuse, and hostel residence. The learner performance in the three selected subjects were used as continuous variables. The results of the analysis should also assist the schools in assessing the effectiveness of other forms of punishment since corporal punishment was abolished.

Contents

List of Tables	v
List of Figures	vii
1. Introduction	1
1.0 Introduction	1
1.1 Background of study	2
1.2 Aims and objectives	3
1.3 Research data descriptions	4
1.4 Data Source	5
2. Literature Review	6
2.1 Preliminary literature review	6
2.2 Theoretical review.	9
2.2.1 Survival analysis	9
2.2.2 Poisson regression model	15
2.2.3 Negative binomial regression model	16
2.2.4 Zero-Inflated models	17
3. Methodology	20
3.1 Introduction	20
3.2 Data.	20
3.3 Descriptive Statistics	21
3.4 Model building	22
3.4.1 Survival analysis methods	22
3.4.2 Modelling the number of repeat offences	24

4. Analysis	26
4.0 Introduction	26
4.1 Descriptive Statistics	26
4.2 Nonparametric Analysis	34
4.2.1 Applying Kaplan-Meier Estimate of the Survivor Function	34
4.2.2 Comparisons between groups according to Treatment versus Control	39
4.2.3 Grouping of subjects based on other Categorical Variables	43
4.2.4 Tests of Hypothesis	48
4.3 Semi-Parametric Analysis	53
4.3.1 Applying the Cox Proportional Hazards Model	53
4.3.2 Examining the Categorical Variables	53
4.3.3 Examining the Continuous Variables	57
4.3.4 Examining the Interaction effect of Variables	58
4.3.5 Examining the Cox Proportional Hazards Model with all the Covariates	
Without Interaction	59
4.3.6 Examining the Cox Proportional Hazards Model with all the Covariates and	
Interactions	61
4.4 Applying Parametric Hazards Models to Learner Offence Data	62
4.4.1 Regressing survival time on the Parental Involvement group	62
4.4.2 Regressing survival time on Suspected Substance Abuse membership	63
4.4.3 Regressing survival time on Repeaters	64
4.4.4 Regressing survival time on the variable (Home location)	64
4.4.5 Regressing survival time on Hostel residence membership	65
4.4.6 Regressing survival time on Gender	66

4.4.7 Regressing survival time on (Grade) membership	67
4.4.8 Regressing survival time on Learner Performance	67
4.5 Fitting Parametric Models with all the Covariates Included	69
4.6 Modelling the Number of Repeat Offences	70
4.6.1 Summary Statistics	70
4.6.2 Examining the Number of Repeat Offences in relation to Parental Involvement . .	71
4.6.3 Examining the Number of Repeat Offences in relation to Suspected Substance Abuse	72
4.6.4 Examining the Number of Repeat Offences in relation to Repeating a Grade	73
4.6.5 Examining the Number of Repeat Offences in relation to Gender	74
4.6.6 Examining the Number of Repeat Offences in relation to Home Location	75
4.6.7 Examining the Number of Repeat Offences in relation to Hostel Residence	76
4.6.8 Examining the Number of Repeat Offences in relation to Grade	77
4.6.9 Exploring Models for Count Data	78
4.7 Summary	81
5. Summary and Conclusions	84
5.0 Introduction	84
5.1 Summary	84
5.2 Conclusion	86
5.3 Recommendations	87
References	89
Appendices	92

List of Tables

Table 3.1 Description of variables in dataset	21
Table 4.1 Summary statistics for continuous variables	26
Table 4.2 Summary of time (in weeks) to second offence	27
Table 4.3 Frequency of parent involvement	28
Table 4.4 Frequency of suspected substance abuse	29
Table 4.5 Subset of learner offence data for 5 weeks	34
Table 4.6 Subset of learner offence data with conditional and unconditional probabilities	35
Table 4.7 Survival probability for parent involvement	40
Table 4.8 Survival probability for suspected substance abuse	41
Table 4.9 Parent Involvement: Tests for equality of survivor functions	48
Table 4.10 Suspected substance abuse: Tests for equality of survivor functions	49
Table 4.11 Gender: Tests for equality of survivor functions	49
Table 4.12 Home location: Tests for equality of survivor functions	50
Table 4.13 Hostel residence: Tests for equality of survivor functions	51
Table 4.14 Repeating: Tests for equality of survivor functions	51
Table 4.15 Grade: Tests for equality of survivor functions	52
Table 4.16 Parental Involvement Likelihood Ratio (LR) test	54
Table 4.17 Suspected substance abuse LR test	54
Table 4.18 Repeating LR test	55

Table 4.19 Home location LR test	55
Table 4.20 Hostel residence LR test	56
Table 4.21 Gender LR test	56
Table 4.22 Grade LR test	57
Table 4.23 Learner performance LR test	58
Table 4.24 LR test for Interaction effect	59
Table 4.25 LR test for combined effects without interaction	60
Table 4.26 LR test for combined effects including interaction	61
Table 4.27 Parametric regression: Parent involvement LR test	62
Table 4.28 Parametric regression: Suspected substance abuse LR test	63
Table 4.29 Parametric regression: Repeating LR test	64
Table 4.30 Parametric regression: Home location LR test	65
Table 4.31 Parametric regression: Hostel residence LR test	65
Table 4.32 Parametric regression: Gender LR test	66
Table 4.33 Parametric regression: Grade LR test	67
Table 4.34 Parametric regression: Learner performance LR test	68
Table 4.35 Parametric regression: Combined effect LR test	69
Table 4.36 Summary Statistics for repeat offences	70
Table 4.37 Comparison of Count Models (Substance abuse and Maths in the Zero part)	78
Table 4.38 Comparison of Count Models (Maths in the Zero part)	80

List of Figures

Figure 4.1 Bar graph for learner grade	29
Figure 4.2 Histogram showing frequency of survival times (in weeks)	30
Figure 4.3 Survival times by (a) Parental involvement (b) Substance abuse (c) Gender	31
Figure 4.4 Survival time across Parental Involvement within Learner Grade	32
Figure 4.5 Survival time across Suspected Substance Abuse within Learner Grade	33
Figure 4.6 Kaplan –Meier survival estimate of learner offence	36
Figure 4.7 Nelson-Aalen estimate of the hazard	37
Figure 4.8 Nelson-Aalen estimate of the cumulative hazard	38
Figure 4.9 Kaplan-Meier survival estimates: Parent involved vs Parent not involved	39
Figure 4.10 Survival estimates: Suspected substance abuse vs not suspected	41
Figure 4.11 Survival estimates for Gender	43
Figure 4.12 Survival curves: repeaters vs non-repeaters	44
Figure 4.13 Survival curves: Local home location vs non-local home location	45
Figure 4.14 Survival curves: Hostel resident vs non-hostel resident	46
Figure 4.15 Survival curves for learner grades	47
Figure 4.16 Repeat offences and Parental involvement	71
Figure 4.17 Repeat Offences and Suspected substance abuse	72
Figure 4.18 Repeat Offences and Grade Repeating	73
Figure 4.19 Repeat Offences and Gender	74
Figure 4.20 Repeat Offences and Home location	75
Figure 4.21 Repeat Offences and Hostel residence	76
Figure 4.22 Repeat Offences and Grade	77