DOI: http://doi.org/10.4038/sljastats.v20i2.8051

Sri Lankan Journal of Applied Statistics, Vol. (20-2)

# Joint Model for Study Programme Completion and Performance of Students in Distance Education in Sri Lanka

**Jayani Hapugoda[1*] and Roshini Sooriyarachchi[2]**

[1]Department of Organizational Studies, The Open University of Sri Lanka, Nawala.
[2]Department of Statistics, University of Colombo, Colombo.
[*]Corresponding Author: jchap@ou.ac.lk

Received: 7th May 2019 / Revised: 20th October 2019 / Published: 20th December 2019



## ABSTRACT

*The completion rates of study programmes by students under the open and distance learning (ODL) mode is important to analyze as the low completion rates or dropouts in ODL mode is an area which has taken much attention over the past years. The time to completion has a relationship with the number of pass grades in the first attempt at final examinations, as if the students cannot go through the exams by first attempt, it will take more time to complete the program. On the other hand, the number of pass grades in the first attempt indicates the performance of the students and the poorly performing students might dropout from programs. When the completion time and the number of pass grades in the first attempt are correlated, it is better to model the two variables together, creating a joint/bivariate model. Thus, this study investigates the impact of individual level factors of the students on the completion status of a study program and number of pass grades in the first attempt at final examinations through a joint/bivariate model of time and count. As the students are clustered/ grouped within the study centers they are attached to, the analysis was undertaken by adjusting for the cluster effect as well. The individual level student factors that are of interest in this study are gender, age, employment status, years of employment experience, civil status, medium of study and attached study center. The records of 1657 students who have registered for Bachelors of Management Studies study programme in 2010 and 2011 at The Open University of Sri Lanka were considered in this study. The findings revealed the significant factors that affect the two variables under consideration and it also suggests that the joint model performs better than models taking the two response variables separately.*

**Keywords:** Open and Distance Learning, Joint modeling, Study Programme Completion, Student performance

## 1. Introduction

While the ‘student success’ is considered as one of the factors that assess the quality of higher education, student dropout/ attrition is identified as a major threat in distance education institutes (Berge & Huang, 2004). Not only dropout rates, but time to dropout is also plays a significant role in this area of research (Ishitani & DesJardins, 2002). Not like in conventional universities, distance education institutes have flexible structure for completing the study programs, while providing multiple exit qualifications for the students. Thus, time taken to complete a particular study program is something important to consider in this context.

The time to complete/ dropout has a relationship with the number of pass grades in the first attempt at final examinations, as if the students cannot go through the exams by first attempt it will take more time to complete the program. On the other hand, the number of pass grades in the first attempt can be an indicator of student performance and the poorly performing students might dropout from programmes. Literature suggests that personal factors of students (such as gender, age, employment status etc.) can affect the dropout rates or the programme completion by the students (Berge & Huang, 2004; Xenos et al, 2002) and student performance (Beaumont-Walters & Soyibo ,2001); Richardson et al, 2012) in higher education. Factors such as age and employment status are important to consider in distance education context, as the study programs are open for students in any age level (except below 16 years of age) and allow part time students who are full time employed to enroll in these programmes. Thus, analyzing the impact of personal factors on time to dropout/ complete and performance of the students would be beneficial for the field of higher education and especially the field of distance education.

## 2. Data

The data obtained for this study represents records of the students who have registered for Bachelor of Management Studies (BMS) programme at the Open University of Sri Lanka (OUSL), which conducts the program in all of the regional/ study centers across the country. As the students are clustered in different regional/ study centers, the clustering effect is also present in the dataset.

These secondary data were retrieved from the OUSL student database for 1657 students who have registered for BMS programme in 2010 and 2011.

Out of the students, some have completed the programme by the end of 2017 (at the point of retrieving data), while some of them have dropped out and some of them are still ongoing. Out of the ongoing students, some are ongoing – active while some are ongoing – inactive. Ongoing inactive students represent the students who are inactive in the system (those who have not updated the studentship) for more than 5 years and hence for this study, they are assumed to be dropped out of the programme. Ongoing – active students are regarded as censored observations at the end of the study period in 2017.

## 3. Descriptive Data Analysis

With respect to gender, there were 737 (44.5%) females and 920 (55.5%) males in the dataset. There were 1271 (76.7%) students who have prior academic qualifications, 375 (22.6%) with prior professional qualifications, while 11 (11.7%) without any prior qualifications. According to the employment status of the students, 1524 (92%) are employed and 133 (8%) are not employed. Age of the students varies from 21 – 67, while the number of pass grades in first attempt ranges from 0 – 28.

Only 147 (8.9%) of the total number of students have completed the programme, 175 (10.4%) are ongoing – active, 280 (16.9%) are ongoing – inactive, while 1058 (63.9%) have been dropped out from the program.

The descriptions of variables used in this study are listed in Table 1.

Table 1: Description of Data

| Variable | Type | Description | Levels | Code |
|---|---|---|---|---|
| Completion Status | Categorical | Whether the students have completed the programme/ dropout/ ongoing | Completed | 1 |
| | | | Ongoing | 1* |
| | | | Dropout | 0 |
| Survival Time | Continuous | Time taken to complete the programme/ dropout from the programme | NA | NA |
| Count | Discrete | The number of pass grades in the first attempt at final examinations | NA | NA |
| Gender | Categorical | Gender of the student | Male | 0 |
| | | | Female | 1 |
| Medium | Categorical | Medium of study | English | 1 |
| | | | Sinhala | 2 |
| | | | Tamil | 3 |

| | | | | |
|---|---|---|---|---|
| Civil Status | Categorical | Civil status of students | Married | 1 |
| | | | Single | 2 |
| | | | Unknown | 3 |
| | | | Widowed | 4 |
| Center | Categorical | Regional/ study center that the students are attached to | Ambalangoda | 1 |
| | | | Anuradhapura | 2 |
| | | | Badulla | 3 |
| | | | Bandarawela | 4 |
| | | | Batticaloa | 5 |
| | | | Colombo | 6 |
| | | | Galle | 7 |
| | | | Gampaha | 8 |
| | | | Hatton | 9 |
| | | | Jaffna | 10 |
| | | | Kalutara | 11 |
| | | | Kandy | 12 |
| | | | Kuliyapitiya | 13 |
| | | | Kurunegala | 14 |
| | | | Matara | 15 |
| | | | Monaragala | 16 |
| | | | Polonnaruwa | 17 |
| | | | Puttalam | 18 |
| | | | Ratnapura | 19 |
| | | | Vavuniya | 20 |
| Age | Discrete | Age of the students | NA | NA |
| Prior Qualification | Categorical | Qualifications obtained prior to registration for the programme | Academic | 1 |
| | | | Null | 2 |
| | | | Professional | 3 |
| Employment Status | Categorical | Employment status of the students | Not Employed | 0 |
| | | | Employed | 1 |
| Years of Working Experience | Continuous | Number of years of working experience the students have once they are registering for the programme | NA | NA |

# 4. Methodology of Joint Modeling

## 4.1 Model Formulation

In this study, the time to dropout from a study program/ time taken to complete the study program is considered as survival time of students in the system (since registering for a study program till dropout/ completion). The other important variable in this study is the number of pass grades at the first attempt in the final examinations, which is a count variable. These two are regarded as response variables in the study.

Since the data are clustered within regional/ study centers, correlations exist between the response variables within clusters, as well as in between predictor variables and response variables. Due to this fact, associations between responses and predictors after adjusting for correlations within the clusters should be determined using the Generalized Cochran Mantel Haenszel test (Zhang & Boos, 1997).

According to the findings of the correlation analysis, there is a significant correlation exists between responses and between predictors and responses. Hence a joint model can be constructed using the two response variables. Here, when the data are clustered, the methodology to obtain a joint model for survival and count data by combining Discrete Time Hazard Model (DTHM) and Poisson Regression Model, has been developed by the first two authors of this paper (Hapugoda et al, 2017; Hapugoda&Sooriyarachchi, 2018). This joint modeling was carried out via Generalized Linear Mixed Models which perform maximum likelihood estimation based on a Laplace approximation of the marginal log likelihood (Capanu et al, 2013). For the modeling, PROC GLIMMIX procedure in SAS software version 9.4 was used (Schabenberger, 2005)

## 4.2 Model in Notation Form

According to (Hapugoda et al, 2017; Hapugoda & Sooriyarachchi, 2018), the responses of analysis are: $Y_{gij1}$ (binary – completed/ dropout) and $Y_{gij2}$ (count - number of first attempt pass grades). Here, the event (completing) is happening in time interval g for the $i^{th}$ student in the $j^{th}$ center given that event does not occur in (g-1) and $x_{ij}$$x_{ij}$are covariates at the center level j and student level i. Variables that impact $Y = (Y_1, Y_2)$ are the explanatory variables (student individual factors) denoted by $X_{ij}$(i = 1 (1)p; j=1(1)N).

To formulate a joint model, Generalized Linear Model (GLM) can be used to form marginal models for each response by considering mean $E(Y_k/X_i)$ and variance

$Var(Y_k/X_i)$ where k=1,2. The responses are linked by structuring a covariance matrix $Var(Y_k/X_i)$ to include potential correlations.

In GLM,

$$l_k(E(Y_{ik}/X_i)) = X_{ik}'\beta_k \tag{1}$$

where k=1,2 and i denotes each student record from each center and $l_k$ is the link function. Here, $l_1(u)$ is the logit link and $l_2(u)$ is the log link. $E_{ij}$ is the Expected count or offset where $E_{ij}$ = Total students$_{ij}$ * Rate, where Rate = Total completed students / Total students.

GLIMMIX in SAS is used to estimate two marginal models jointly.

This paper discusses the joint modelling of Binary and Poisson count outcomes. We use GLMM concepts to capture the correlation between the responses within the center. For simplicity however, single random variable u0j is used to capture the correlation. That is exchangeability is assumed. That is the correlation between the same outcome variable within the center and the correlation between different responses with the same individual are assumed to be the same. The ideal situation would be to have two random effects (Sunethra and Sooriyarachchi, 2019) for these two situations. However, as there are only two responses within an individual often convergence problems could occur in fitting the model. Apart from this, this situation can be efficiently modeled by having only one random effect. (Adikari and Sooriyarachchi, work in progress).

A structural formulation of the model is given as:

$$l_1(Y'_{gij1}) = \text{Logit}(\boldsymbol{\pi}_{\mathbf{gij}}) = \beta_{0j} + \sum_{g=1}^{n} \alpha_g T_{gij} + \boldsymbol{\beta}\mathbf{X}_{\mathbf{ij}} \tag{2}$$

where $\beta_{0j} = \beta_0 + u_{0j}$ and $u_{0j} \sim N(0,\sigma_u^2)$

and

$$l_2(Y_{ij2}') = \log(\boldsymbol{\mu}_{\mathbf{ij2}}) = \log(E_{ij}) = \beta_{0j} + \boldsymbol{\beta}\mathbf{X}_{\mathbf{ij}} \text{ where } \beta_{0j} = \beta_0 + u_{0j} \text{ and } u_{0j} \sim N(0, \sigma_u^2) \tag{3}$$

As mentioned earlier, for simplicity, we assume that both random effects are the same ($u_{0j}$) and have the same variance ($\sigma_u^2$).

The joint model variance covariance matrix of the two outcomes within the same centeris of the form

$$\begin{bmatrix} \sigma_u^2 & \rho_{12}\sigma_u^2 \\ \rho_{12}\sigma_u^2 & \sigma_u^2 \end{bmatrix}$$

PROC GLIMMIX in SAS will structure the variance matrix of Y = ($Y_1$, $Y_2$) as Var ($Y_i/X_i$) = $\mathbf{A_i^{1/2} R_i A_i^{1/2}}$where $\boldsymbol{R_i}$is a user specified 2x2 covariance structure and $\boldsymbol{A_i}$ is the diagonal matrix of the variances of ($Y_1$, $Y_2$).

The likelihood of the joint model is,

$$L = \int \prod_{i=1}^{n} \prod_{j=1}^{n_j} f(y_{gij1}|a) f(y_{ij2}|a) f(a)\, da$$

$$L = \int \prod_{i=1}^{n} \prod_{j=1}^{n_j} [P(y_{gij1}|a)^{y_{gij1}} (1 - P(y_{gij1}|a)^{(1-y_{gij1})}] P(y_{ij2}|a) f(a)\, da$$

$$\mathrm{L} = \int \prod_{i=1}^{n} \prod_{j=1}^{n_j} [(\frac{\exp(Z_{gij})}{1+\exp(Z_{gij})})^{y_{gij1}} (1 - (\frac{\exp(Z_{gij})}{1+\exp(Z_{gij})})^{(1-y_{gij1})}] \left(\frac{e^{-\mu_{ij}} (\mu_{ij})^{y_{ij2}}}{y_{ij2}!}\right) \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left(\frac{1}{2} a_i' \Sigma\, a_i\right) da \qquad (4)$$

## 5. Advanced Data Analysis - Joint Modeling

### 5.1 Steps of Modeling

The following steps were followed in data analysis.

**Step 1 -** The time to completion/ dropout, which is a continuous variable, should be categorized in appropriate time intervals. According to the percentiles of the data, it was categorized into three predetermined time intervals as 0 – 2 years, 2 to 4 years and more than 4 years. For each category of time to completion/ dropout, the completion status (whether completed/ dropout/ ongoing) and the number of first attempt pass grades should be recorded.

**Step 2 -** Data are restructured in to those time intervals.

This process can be explained using the following example.

Example:
Consider the following time to completion/ dropout data in 2 centers (C1, C2). The number of first attempt pass grades (count) was recorded for each student, with their respective time to completion/ dropout and the times were categorized in appropriate time categories coded as 1, 2 and 3.

Consider the following 4 observations [1-C1-3*(5), 2-C1-1*(6), 3-C2-2*(7), 4-C2-3(1)]. Here the count is mentioned within brackets after the time category and '*' denotes students who have completed the respective study programme. For example, the first data point should be read as; first student of C1 has completed the programme at the 3rd time category with a count of 5. The final data point should be read as the second student of C2 has not completed/ dropped out of the programme with a count of 1 in the third time category, though he/ she has been ongoing in the system up to the second time period. This dataset is shown in tabular format under table 2.

Table 2: Original Dataset

| Center | Student | Last Observed Time Category | Status/ Censoring (Completed – 1/ Dropout – 0) | Count |
|---|---|---|---|---|
| C1 | 1 | 3 | 1 | 5 |
| C1 | 2 | 1 | 1 | 6 |
| C2 | 3 | 2 | 1 | 7 |
| C2 | 4 | 3 | 0 | 1 |

Now, the data should be restructured according to the method of DTHM [12]. Here, all the data points from each student in each center should be listed for each time period up to the occurrence of the event (i.e. completion or dropout). The restructured data set is given in table 3.

Table 3: Restructured Dataset

| Time Category | Center | Student | Status/ Censoring (Completed – 1/ Dropout – 0) | Count |
|---|---|---|---|---|
| 1 | D1 | 1 | 1 | 5 |
| 2 | D1 | 1 | 1 | 5 |
| 3 | D1 | 1 | 1 | 5 |
| 1 | D1 | 2 | 1 | 6 |
| 1 | D2 | 3 | 1 | 7 |
| 2 | D2 | 3 | 1 | 7 |
| 1 | D2 | 4 | 1 | 1 |
| 2 | D2 | 4 | 1 | 1 |
| 3 | D2 | 4 | 0 | 1 |

As the next step, time indicators should be introduced to the data, in a way that each time category corresponds to a particular time indicator. For 3 time categories, 3 time indicators should be introduced. For each time indicator, Ti= 1 if time category = i and Ti = 0 if time category ≠ i. (Shown in Table 4). At each time category, the response variable has a code of 1 if the corresponding student has completed the programme in that time category, and 0 otherwise.

Table 4: Restructured dataset with time indicators

| Time Category | Center | Student | Status | Count | Time Indicator – T1 | Time Indicator – T2 | Time Indicator – T3 |
|---|---|---|---|---|---|---|---|
| 1 | C1 | 1 | 1 | 5 | 1 | 0 | 0 |
| 1 | C1 | 2 | 1 | 6 | 1 | 0 | 0 |
| 1 | C2 | 3 | 1 | 7 | 1 | 0 | 0 |
| 1 | C2 | 4 | 1 | 1 | 1 | 0 | 0 |
| 2 | C1 | 1 | 1 | 5 | 0 | 1 | 0 |
| 2 | C2 | 3 | 1 | 7 | 0 | 1 | 0 |
| 2 | C2 | 4 | 1 | 1 | 0 | 1 | 0 |
| 3 | C1 | 1 | 1 | 5 | 0 | 0 | 1 |
| 3 | C2 | 4 | 0 | 1 | 0 | 0 | 1 |

Now, a binary response variable (status) has been created, while the other response, count, is in discrete form and thus can usually be considered as a Poisson distributed variable.

Using the above mentioned approach, all 1657 data records were restructured.

## 5.2 Performance Comparison

Two separate models for two responses and a joint model were fitted to compare the joint model with the separate models. To model the status (completion/ dropout), a logistic regression model was used, while to model the count, a poisson regression model was used. The fit statistics were obtained for comparison of the models, as shown in table 5.

Table 5: Performance comparison of the models

| Fit Statistics | Joint model | Separate models | | |
|---|---|---|---|---|
| | | Completion model | Count model | Total |
| -2 Log Likelihood | 24,838.44 | 5,173.99 | 44,477.06 | 49,651.05 |
| AIC | 24,974.44 | 5,241.99 | 44,545.06 | 49,787.05 |
| AICC | 24,976.57 | 5,242.53 | 44,545.60 | 49,788.13 |
| BIC | 25,042.15 | 5,275.85 | 44,578.92 | 49,854.77 |

As per the model fit statistics in table 5, it can be seen that joint model has achieved lower AIC, AICC and BIC than the sum of those values obtained for the separate models. For example, the sum of AICs of the separate models were 49,787.05 (5,241.99 + 44,545.06) which is higher than the joint model's AIC 0f 24,974.44. In addition, -2 log L is decreased by 24,812.61 for 1 degree of freedom which is significant at 5% level of significance. These represent the superiority of the joint model than the two separate models.

### 5.3 Results

Upon confirming the superiority of the joint model, table 6 below presents the results of the joint model. Highlighted values are corresponding to the significant factors.

Interpretations should be carried out for the two marginal models separately; using odds ratio for the binary model and using expected values for the count model. These results are obtained by writing down the equations for the models according to equations (2) and (3) and manipulating the equations for the said parameter values.

According to the results from the binary marginal model, the odds of programme completion within 0 – 2 years of time is 0.23 times the odds of completion by taking more than 4 years of time, while the other factors held constant (i.e. higher dropouts in the first two years of the degree). The odds of completion is 2.1 times higher for a female than that of a male (i.e. the probability of females completing the programmes are higher than males), while it is 0.99 times higher for one year increase in age (i.e. the probability of completion decreases with the age). The centers can be ranked according to the odds of completion.

The list of centers ranked from highest to lowest on completion status is Ambalangoda, Badulla, Kurunegala, Kalutara, Galle, Anuradhapura, Polonnaruwa, Gampaha, Jaffna, Puttalam, Monaragala, Ratnapura, Matara, Vauniya, Hatton, Kandy, Batticaloa, Kuliyapitiya, Colombo and Bandarawela.

According to the results from the count marginal model (for student performance – number of pass grades in first attempt), the student performance is 0.51 higher when completing the programme within 0 – 2 years of time than in more than 4 years of time, while it is 1.09 higher in 2 - 4 years of time than in more than 4 years of time. Female students' performance is 1.41 higher than that of male students. The performance of English medium students is 0.94 higher than that of Tamil medium students, while Sinhala medium students perform 0.87 higher than Tamil medium students. When the age increases by one year, the performance increases by 0.98

and when the students have no prior qualification, the performance is 1.55 higher than having professional qualifications prior. For each year increase in working experience, the performance is increased by 0.99. The centers can be ranked according to the student performance in respective centers and the list of centers ranked from highest to lowest on student performance is Ambalangoda, Badulla, Polonnaruwa, Kurunegala, Kalutara, Galle, Gampaha, Anuradhapura, Ratnapura, Puttalam, Monaragala, Jaffna, Matara, Batticaloa, Vauniya, Kandy, Colombo, Hatton, Kuliyapitiya and Bandarawela.

Table 6: Results of the joint model

| Effect | Levels | Binary model | | Count model | |
|---|---|---|---|---|---|
| | | Estimate | Pr> \|t\| | Estimate | Pr> \|t\| |
| dist | | -7.0058 | 0.7928 | -0.5096 | 0.4018 |
| t1 | | -1.3953 | <.0001 | -0.6783 | <.0001 |
| t2 | | 0.1634 | 0.3614 | 0.08288 | 0.0002 |
| Gender | Female | 0.7436 | <.0001 | 0.3434 | <.0001 |
| Gender | Male | 0 | . | 0 | . |
| Center | Ambalangoda | 11.0121 | 0.5206 | 6.1219 | <.0001 |
| Center | Anuradhapura | 3.0172 | <.0001 | 2.2673 | <.0001 |
| Center | Badulla | 10.9378 | 0.4409 | 5.1361 | <.0001 |
| Center | Bandarawela | -1.9949 | 0.8947 | -2.9148 | 0.7681 |
| Center | Batticaloa | -0.6586 | 0.2802 | 0.5051 | 0.0041 |
| Center | Colombo | -1.7985 | 0.0013 | -1.3105 | <.0001 |
| Center | Galle | 3.0201 | 0.0015 | 3.3111 | <.0001 |
| Center | Gampaha | 1.7655 | 0.0149 | 2.3634 | <.0001 |
| Center | Hatton | -0.4819 | 0.9856 | -1.4558 | 0.9327 |
| Center | Jaffna | 1.4325 | 0.0161 | 1.3639 | <.0001 |
| Center | Kalutara | 3.6545 | 0.0007 | 3.4302 | <.0001 |
| Center | Kandy | -0.5377 | 0.3421 | -0.02976 | 0.8623 |
| Center | Kuliyapitiya | -0.8856 | 0.9735 | -1.9967 | 0.9077 |
| Center | Kurunegala | 4.1212 | <.0001 | 3.5784 | <.0001 |
| Center | Matara | 0.2427 | 0.6991 | 1.1227 | <.0001 |
| Center | Monaragala | 0.3509 | 0.7097 | 1.4381 | <.0001 |
| Center | Polonnaruwa | 2.8282 | 0.0025 | 3.8315 | <.0001 |
| Center | Puttalam | 0.4516 | 0.5599 | 1.6933 | <.0001 |
| Center | Ratnapura | 0.2581 | 0.8325 | 1.9565 | <.0001 |
| Center | Vavuniya | 0 | . | 0 | . |
| Medium | English | 0.2514 | 0.1836 | -0.06647 | 0.0224 |

| | | | | | |
|---|---|---|---|---|---|
| Medium | Sinhala | 0.1655 | 0.4137 | -0.1393 | <.0001 |
| Medium | Tamil | 0 | . | 0 | . |
| Civil Status | Divorced | 3.6464 | 0.8914 | -0.6938 | 0.2724 |
| Civil Status | Married | 4.0024 | 0.8807 | -0.2495 | 0.6672 |
| Civil Status | Single | 4.3788 | 0.8696 | -0.2854 | 0.6228 |
| Civil Status | Unknown | 4.3637 | 0.87 | -0.2757 | 0.6347 |
| Civil Status | Widowed | 0 | . | 0 | . |
| Age | | -0.01472 | 0.0641 | -0.02217 | <.0001 |
| Prior Qualification | Academic | 0.06468 | 0.5856 | 0.01456 | 0.4359 |
| Prior Qualification | Null | 0.01181 | 0.9829 | 0.4405 | <.0001 |
| Prior Qualification | Professional | 0 | . | 0 | . |
| Employment Status | Not Employed | -0.04377 | 0.8091 | -0.02575 | 0.3784 |
| Employment Status | Employed | 0 | . | 0 | . |
| Years of Working Experience | | -0.00321 | 0.6531 | -0.00498 | <.000 |

## 6. Conclusion

Literature reveals that student dropout in Open and Distance Learning (ODL) context is one problematic area of research, which is difficult to analyze. Few advanced statistical analyses on time to dropout/ completion and student performance has been done both in local and international level, though there are many qualitative studies and descriptive studies have been carried out. Thus, this study has analyzed time to dropout/ completion and number of pass grades in first attempt (which is an indicator of the student performance) in order to assess the individual student factors affecting them by utilizing a joint model. Accordingly, completion/ dropout time, gender, center and age are significantly affecting the completion status, while completion/ dropout time, gender, center, age, prior qualifications, medium of study and years of working experience are significantly affecting the student performance. The students are considered to be clustered in regional/ study centers and the findings reveal the ranked list of centers based on their students' completion status and performance. The findings of this analysis would facilitate the decision making process of distance education institutes, in terms of allocation of facilities to centers, improvement of teaching processes in centers etc.

## Acknowledgement

The authors profoundly thank The Open University of Sri Lanka for providing the data to carry out the study and University of Colombo for funding the project.

## References

1. Beaumont-Walters, Y. and Soyibo, K. (2001). An analysis of high school students' performance on five integrated science process skills. *Research in Science & Technological Education*, 19(2), 133-145.
2. Berge, Z. L. and Huang, Y. P. (2004). 13: 5 A Model for Sustainable Student Retention: A Holistic Perspective on the Student Dropout Problem with Special Attention to e-Learning. *DEOSNEWS [Online] Available: www. researchgate. net/profile/Zane_Berge/publication/237429805*.
3. Capanu, M., Gönen, M. and Begg, C. B. (2013). An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Statistics in medicine*, 32(26), 4550-4566.
4. Hapugoda, J. C. and Sooriyarachchi, M. R. (2018). Joint Modeling of Discrete Time Hazard Model with Poisson Regression Model: A Simulation Study. Accepted for publication at Jaffna University International Research Conference 2018.
5. Hapugoda, J. C., Sooriyarachchi, M. R., Kalupahana, R. S. and Satharasinghe, D. A. (2017). Joint Modeling of Mixed Responses: An Application to Poultry Data. *Proceedings of 5th Annual International Conference on Operations Research and Statistics (ORS) 2017, Singapore.*
6. Ishitani, T. T. and DesJardins, S. L. (2002). A longitudinal investigation of dropout from college in the United States. *Journal of college student retention: research, theory & Practice*, 4(2), 173-201.
7. Richardson, M., Abraham, C. and Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353.
8. Schabenberger, O. (2005). Introducing the GLIMMIX procedure for generalized linear mixed models. *SUGI 30 Proceedings, 196.*
9. Singer, J. D. and Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of educational statistics*, 18(2), 155-195.

10. Sunethra, A. A. and Sooriyarachchi, M. R. (2017). A bivariate model for analyzing recurrent multi-type automobile failures. In *Journal of Physics: Conference Series*, 890(1), 012142.
11. Xenos, M., Pierrakeas, C. and Pintelas, P. (2002). A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. *Computers & Education*, 39(4), 361-377.
12. Zhang, J. and Boos, D. D. (1997). Generalized Cochran-Mantel-Haenszel test statistics for correlated categorical data. *Communications in Statistics-Theory and Methods*; 26(8), 1813-1837.