

# Student Attrition Propabilities: A Case Study

Shafeeqah Al-Awadhi & Mohammed Ahmed

Kuwait University Dept. of Statistics & Operation Researches.

P.O.Box 5969, Safat 13060, Kuwait

[alawadi@kuc01.kuniv.edu.kw](mailto:alawadi@kuc01.kuniv.edu.kw); [wakeel@kuc01.kuniv.edu.kw](mailto:wakeel@kuc01.kuniv.edu.kw)

## ABSTRACT

Student retention is one major challenge facing different collages at Kuwait University and administrators. Previous research found that the students enrolled at Kuwait University have high probability of leaving study before graduation especially those belonging to science majors. Using the data on each faculty for student cohorts who first enrolled in the academic years of the 1993/1994 through 1999/2000, we examined the factors influencing student attrition. A multinomial logit choice model is developed to estimate the probability of continuation for a new student at each year.

## INTRODUCTION

There are many studies that advocate longitudinal, explanatory approaches as a key to understanding student departure. For example Tinto (1975,1982 and 1987) give theoretical model to analyzing the attrition question based on the concept of social and academic integration. This model asserts that a student comes to a particular institution with a range of background characteristic and goal commitments. Wetzel et al. (1999) proposed a comprehensive model of retention based on Tinto's goal commitment and institutional commitment combined with financial consideration for the case study of an urban public university. We believe that the background characteristics have great impact on the drop out decision. The purpose of this study is to control the acceptance of applicants to Kuwait University through sitting a model that depends on the characteristic background in order to reduce the attrition rate in the future.

## RETENTION MODEL

There are many factors of the characteristic background that affect the student retention rate at Kuwait University. The main factors and their definition is presented in Table 1. This work focuses on the attrition question by analyzing the marginal impacts that these factors have on attrition. The study involves all freshmen students at Kuwait University for the academic years 1993/1994 and they are about 4045 students. We are studying their cases at the end of the academic year 1998/1999.

**Table1: Variables names and definitions.**

Variable Name	Definition
Status	0 for continue or graduate 1 for drop out
Gender	0 male and 1 female
Percentage High School	Percentage of high school grade
Type of High School	Type of high school: 0 for general school 1 for courses school
Major of Study in High-School	0 for art 1 for science
Area-code	There are 5 areas in Kuwait: Capital (1), Hawalli (2), Farwaniyah (3), Ahmadi (4), and Jahra (5)
Collage Enrolled	There are 10 collages in Kuwait University: Collage of Commerce (1), Law (2), Art (3), Science (4), Medicine (5), engineer and Petroleum (6), Assistant of Medicine (7), Education (8), and Shareea & Islamic studies (9)

We set a model in order to control the acceptance of applicants by showing the attrition rate of each applicant based on given background characteristics about the applicant and the collage to be enrolled. The empirical model is a regression model involving a binary dependent variable. The dependent variable is the status of students: 0 for attrition and 1 for retention. A logistic specification will be used. The estimation method yields maximum likelihood estimates of the model's parameters. Now consider the logistic response regression functions

$$E(Y_i) = \exp(\hat{\alpha}' X_i) / (1 + \exp(\hat{\alpha}' X_i)).$$

And the equivalent simple logistic response extends to:

$$E(Y_i) = 1 / [1 + \exp(-\hat{\alpha}' X_i)]$$

Where  $Y_i$  is binary taking value of either 0 or 1. Let we denote  $E(Y)$  by  $\delta$ , since the mean response is the probability of retention. Because the logistic regression function is nonlinear in the parameter, the estimated coefficients cannot be interpreted as usual linear least squares interpretation. To measure this relation the logistic regression function can be linearized

easily by taking logarithm of the odd ratio:  $\delta' = \text{LOGe}(\delta_i / (1 - \delta_i)) = \hat{\alpha}' X_i$  which is called the logistic transformation of the probability  $\delta$ .

To study in more detail whether any significant results can be concluded from response rates by some factors when taking different subsets of the population, the data is separated by gender. (The performance of male and female students is significantly different). The model will be then re-estimated for both cases (male: 1565, female: 2485).

## EMPERICAL RESULTS

Results in Table 2, whose column 2 is organized to enable comparing the results for the entire student and columns 3 and 4 for the male and female students respectively. The output gives the value for each coefficient that provides the best fit to the observed data. Providing along with these values is a set of statistics used to conduct diagnostic tests of the model.

**Table 2: Logit estimates** (p-value of t-statistics test in parentheses)

<b>Variables</b>	<b>University</b>	<b>Male</b>	<b>Female</b>
Constant	5.247 (0.000)	4.778 (0.00)	5.638(0.000)
Sex (male)	0.554 (0.000)		
% High School	-0.083 (0.00)	-0.075 (0.00)	-0.085 (0.000)
General High School	-0.850 (0.00)	-1.012 (0.00)	-0.710(0.000)
Art Major	-0.870 (0.00)	-0.692 (0.00)	-0.964(0.000)
Capital Area	0.277 (0.063)	0.446(0.038 )	0.166 (0.423)
Hawalli Area	0.363 (0.017)	0.494 (0.03)	0.220 (0.282)
Farwaniah Area	0.285 (0.074)	0.215(0.414 )	0.271(0.188)
Ahmadi Area	0.513 (0.001)	0.675 (0.006)	0.394 (0.064)
Collage of Commerce	0.459 (0.012)	0.706 (0.011)	0.317(0.191)
Collage of Law	0.558 (0.016)	0.721 (0.031)	0.549(0.087)
Collage of Art	0.340 (0.063)	0.533 (0.082)	0.158 (0.485)
Collage of Science	2.352 (0.000)	2.798 (0.000)	2.116 (0.000)
Collage of Medicine	0.665 (0.169)	0.875 (0.138)	-0.066 (0.949)
Collage of Engineer and Petroleum	2.280 (0.000)	2.575 (0.000)	1.785 (0.000)
Collage of Assistant of Medicine	1.848 (0.000)	1.964 (0.010)	1.682 (0.000)
Collage of Education	0.709 (0.000)	0.978 (0.002)	0.461(0.044)
Observations	4045	1561	2484
X <sup>2</sup> (Goodness of Fit)	3990.8	1658.1(0.00)	2304.5(0.01)
p-value (Goodness of Fit)	(0.00)	)	0.182
Sensitivity (% of retention predicted correctly)	0.243 87.4%	0.254 72.8 %	93.7%
Specificity (% of attrition predicted correctly)	46.2%	65.1 %	25.9%

For judging the model validation, "goodness of fit" value is given ( $\chi^2$ -test = 3990.803, p-value = 0.000), indicating that the given model provides the best fit to the observed data. For checking the model validation, at the end of the column, Table 2 provide values of sensitivity and specificity, where the value of sensitivity measure the percentage of cases in which  $y=0$  (retention) is predicted correctly for the observed data while the specificity measures the percentage of cases in which  $y=1$  (attrition) is predicted correctly. These measures indicate that the model appears to predict quite well especially the retention probabilities.

## REFERENCES

- Tinto, V. (1975). Drop out from higher education: A theoretical synthesis of recent research. *Review of Education Research*, 45, 89:125.
- Tinto, V. (1982). Limit of theory and practice in student attrition., *Journal of Higher Education*, 43, 687:700.
- Tinto, V. (1978). *Leaving collage: rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.
- Wetzel, J. N., O' Toole, D. and Peterson, S., (1999). Factors affecting student retention probabilities: a case study., *Journal of Education and Finance*. 23 (1), 45:554.