

https://dspace.org/

Analysing first year students' performance in the Commerce Faculty at the University of the Witwatersrand

Yathavan, Vasuki

2009-09-16T07:10:58Z
http://hdl.handle.net/10539/7268

Downloaded from a DSpace repository. https://dspace.org/

ANALYSING FIRST YEAR STUDENTS' PERFORMANCE IN THE COMMERCE FACULTY AT THE UNIVERSITY OF THE WITWATERSRAND

Vasuki Yathavan

A research report submitted to Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfillment of the requirements for the degree of Master of Science

Johannesburg, 2008.

Declaration

I declare that this research report is my own, unaided work. It is being submitted for the degree of Masters in Mathematical Statistics in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

_____

(Vasuki Yathavan)

…….day of ……

ABSTRACT

With the increasing diversity of students attending University, there is a growing interest in the factors predicting academic performance. A large number of students who enter University do not continue beyond the first year of study. Academics seek explanations, whereas University administrators desire to manage their student enrolments by reducing failure rates. Decision on admissions to University and placement into University courses are usually based on the results of achievement (as in secondary school exams) and/or selection tests.

About half of the first year students in the Faculty of Commerce at the University of the Witwatersrand, do not continue to their second year. The drop out rate of first year students in this Faculty reported to range from roughly 24% to 32%. In this report an attempt is made to identify factors which affect the students' performance during the first year. The purpose of this report is to use a CHAID analysis to find the importance of some predictors and interactions between them as well as fitting a Multinomial Logistic Regression model to the same data.

This report presents the important predictors from the statistical analyses. The analyses were done on the first year students in the Faculty of Commerce, University of the Witwatersrand from 2003 to 2006. Previous Institution Type, Gender, Age, Matriculation Aggregate, First year performance and Matriculation courses (Accountancy, Biology, English, History, Mathematics and Physical Science) were used as predictor variables.

The CHAID analyses indicated that Matriculation Aggregate is the most important predictor, whereas Previous Institution Type, Age, Accountancy, English and Physical Science are also important predictors. Several of these variables interact with it. In the Multinomial Logistic Regression analysis, Age, Aggregate, Accountancy, English, Mathematics and Physical Science are the significant predictors. Most of these variables were significant as variables interacting with some of these variables. Age is the only single variable significant on its own in these models.

ACKNOWLEDGEMENTS

I thank my supervisor Mr. Peter Fridjhon for their untiring support and guidance throughout this research. This work would not have been possible without his commitment.

I also thank to Prof. J. Galpin and Mr. P. Fridjhon, who organized the internship work in Strategic Planning Division.

My appreciation and gratitude is to Strategic Planning Division for this opportunity and financial assistance. Thanks also the staffs of IRU and SPD for your continual support, practical advice and friendship.

My appreciation goes to the University of the Witwatersrand's Postgraduate Merit Award Scholarship for providing me with the much needed financial support.

I would like to thank the MIU for providing me the data. Ms. D. Harshila and
Ms. V. Pooja provided the data all the studies.

Prof. G. Fernandez has been especially helpful in extending the programme of CHAID macro for my data.

I would like to thank my husband Yathavan and daughter Nethra for the support. I thank my parents and my sister with all my heart. Thank you to my friends and staff at School of Statistics and Actuarial Science.

Most importantly, I thank God for giving me the strength and desire to conduct this study.

TABLE OF CONTENTS

Declaration ii

Abstract iii

Acknowledgements iv

Table of contents v

List of figures viii

List of tables ix

CHAPTER 1 : INTRODUCTION 1

1.1 Literature review 3

1.2 Purpose of study 6

1.3 Research questions 7

1.4 Methodology 7

CHAPTER 2 : LITERATURE REVIEW AND METHODOLOGY 11

2.1 Scales of measure 11

2.2 Statistical methods 12

2.2.1 Chi-Square Automatic Interaction Detection (CHAID) 13

2.2.1.1 Basic tree-building algorithm 14

2.2.1.2 Mathematical description of CHAID 16

2.2.1.2.1 Outline of the technique 17

2.2.1.2.2 Case of a dichotomous dependent variable 19

2.2.1.2.3 Some properties of the technique 21

2.2.1.2.4 Convergence of the procedure 27

2.2.1.3 Validating tree results 30

2.2.1.4 Statistical distributions 31

Table of contents continued…..

2.2.2	Binary Logistic Regression	31
	2.2.2.1 The model	32
	2.2.2.2 Tests of significance	34
2.2.3	Multinomial Logistic Regression	34
	2.2.3.1 Fitting the Multinomial Logistic regression Model	35
	2.2.3.2 Interpreting the fit and odds ratio	39
2.3	Student performance	40
CHAPTER 3 : METHODOLOGY		44
3.1	Data source	45
3.2	Data cleaning	46
3.3	Data analysis	50
CHAPTER 4 : RESULTS		53
4.1	CHAID	53
4.2	Multinomial Logistic Regression	65
CHAPTER 5 : CONCLUSION AND DISCUSSION		75
5.1	Discussion on the results of the CHAID analyses	76
5.2	Discussion on the multinomial logistic regression analyses	80
5.3	General discussions on the models	81
5.4	Important categories of the significant variables	82

REFERENCES 87

Table of contents continued…..

APPENDICES

APPENDIX 1 : DISTRIBUTION OF THE VARIABLES OF TRAINING AND VALIDATION DATA 96

APPENDIX 2: STUDENTS' PERFORMANCE BY AGE AND AGGREGATE 102

LIST OF FIGURES

Figure	Caption	Page
2.1	The CHAID algorithm Perreault & Barksdale (1980)	17
2.2	Four category example whose middle two categories have identical means	23
4.1	CHAID output for students' performance on the training data	55
4.2	Donut chart showing the training data classification summary display generated by using the SAS macro CHAID	61
4.3	Donut chart showing the validation data classification summary display generated by using the SAS macro CHAID	62
4.4	Validation tree	65

LIST OF TABLES

Table	Caption	Page
2.1	Summary of data for one predictor	18
2.2	Dichotomous dependent variable	19
2.3	A sub table for two rows	22
3.1	Variables and associated response categories	48
4.1	Description of terminal nodes	60
4.2	Observed and predicted percentages	63
4.3	Maximum likelihood analysis of variance for the single variables	66
4.4	Maximum likelihood analysis of variance for the two factor interaction	67
4.5	Maximum likelihood analysis of three factor interactions	68
4.6	Maximum likelihood analysis of four factor interactions	68
4.7	Maximum likelihood analysis of variance for five factor interactions	69
4.8	Maximum likelihood analysis of variance for six factor	

List of tables continued……

	interaction	69
4.9	Maximum likelihood analysis of variance	71
4.10	Maximum likelihood analysis of variance	72
4.11	Estimated Coefficients, Estimated standard errors, Wald statistics and two tailed p-values for the full multivariable model fit	72
4.12	Predicted Frequencies	74