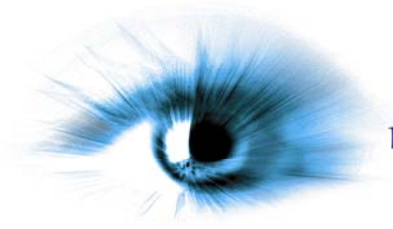


LOTHIANS EQUAL ACCESS PROGRAMME FOR SCHOOLS

LEAPS

Tracking LEAPS students through their years in Higher Education

A SEF Funded Report for LEAPS by Paula L McClements

South East Forum
widening access to
increase participation

Abstract

There is growing concern that a reduction in the number of students successfully completing their time at university is linked to an increased intake of students from non traditional backgrounds. Lothians Equal Access Programme for Schools (LEAPS) works with 46 secondary schools and six partner Higher Education Institutions (HEIs) to facilitate increased participation and success in higher education. This report assesses the influence of academic, economic, social and cultural factors on the performance of LEAPS students at six HEIs at the end of first year through to fourth year. Simple logistic regression is used to estimate the probability that a LEAPS student will successfully complete an academic year at university. The analysis indicates that completion is influenced, as expected, by qualifications but also by subject studied, ethnic background, gender and destination institution. There is also strong evidence to suggest that completing the LEAPS summer school improves the probability of success. The main conclusion is that LEAPS students perform at least as well as the general student population.

I would like to acknowledge the support of the admissions staff in the LEAPS partner institutions and the University of Stirling, the LEAPS team and SEF.

Contents

List of Tables

List of Graphs

Chapter One Introduction

Widening access to and participation in higher education for under represented groups is a key priority for the Scottish Executive. One of the goals of its lifelong learning strategy, Life Through Learning; Learning Through Life (2003), is "a Scotland where people have the chance to learn, irrespective of their background or current personal circumstances". The Scottish Higher Education Funding Council Report, Learning for All (2005), states that educational participation and achievement (in Scotland) is highly skewed particularly by socio-economic background, geography and gender. Furthermore the report states that the absence of confidence, aspirations, a sense of value of learning, the drive to learn, a determination to work at it and a lack of awareness of the options are the most significant barriers to learning for disadvantaged groups.

There are many wider access initiatives throughout Scotland and the UK in place to address these inequalities. These aim to ensure that students who have not reached their full potential as a result of economic, social or cultural factors receive a fair chance to enter higher education.

Lothians Equal Access Programme for Schools (LEAPS) is one such initiative.
Its main remit is to encourage able pupils from state schools in the region to consider higher education as an option when leaving school. LEAPS works within Edinburgh and the Lothians with 46 secondary schools, four local councils and six Higher Education Institutions (HEIs).

1.1 Study Aims

Recent research by Sinclair and McClements (2003) monitored and tracked a group of students through their first year at university matching their LEAPS eligibility details and student characteristics with their first year outcome. This current research will extend further to include additional first year data and data from second, third and fourth years. The scope of the statistical modelling will also be extended.

Monitoring and tracking of wider access students is a necessary step in determining the success of any initiative or policy. It is frequently neglected. This research addresses this issue.

1.2 Study Objectives

- To track LEAPS students through their years at LEAPS partner HEIs and the University of Stirling; from entering in October 2001 through to their academic status at the end of the 2004/05 session.
- To evaluate the effect of the LEAPS summer school throughout these years.
- To ascertain which student characteristics are significant for those who have progressed and for those who have not progressed.
- To develop a statistical model for each HEI taking into account their student background characteristics.
- To provide useful information to all wider access initiatives and educational institutions throughout the UK.
- To develop a set of data libraries and 'desk notes' for additional or future use.

This study will enhance the evaluation of LEAPS as a wider access initiative and as a means of evaluating the extent to which LEAPS summer school plays a positive role in student retention rates. It will also act as an evaluation tool for government policy. In a time when organisations are increasingly being held accountable to prove their impact, this research will be of interest to other HEIs and wider access initiatives.

Although this study sets the framework for any long term tracking study of wider access students, it will, in principle, be equally applicable to tracking any student population.

1.3 Background to LEAPS

LEAPS emerged from the University of Edinburgh's 'University Special Entrance' scheme. In 1996 a partnership of the four local councils in Lothian, four HEIs and Careers Scotland were brought together to move the project forward to meet the

common goal of widening access to higher education in Edinburgh and the Lothians. This partnership initiated LEAPS (Lothians Equal Access Programme for Schools).

LEAPS has undergone many changes since conception and was re-launched in the autumn of 2001. In September 2006 the partners are:

City of Edinburgh Council
East Lothian Council
Midlothian Council
West Lothian Council
Edinburgh College of Art
Heriot-Watt University
Napier University
Queen Margaret University College
Scottish Agricultural College
The University of Edinburgh
Careers Scotland.

LEAPS aims to promote social inclusion and equality of opportunity by facilitating increased participation and success in higher education of young people (in Edinburgh and the Lothians) whose ability to choose higher education as a post school option and/or to demonstrate their potential may have been inhibited by economic, social or cultural factors.

1.3.1 LEAPS Objectives

LEAPS Objectives are to**:**

- Provide (a) *young people* and (b) *their parents* with advice, information and encouragement to consider higher education, accessed directly or through further education, as an attractive and attainable option.
- Provide impartial information and advice about courses and routes to higher education.

- Raise awareness of widening participation issues and the need to challenge traditional assumptions about admissions criteria within (a) *higher education institutions* and (b) *schools.*
- Enhance prospects of young people fulfilling academic potential by promoting positive attitudes to learning and acquisition of learning skills to ensure effective transition to and success in higher education.
- Monitor and evaluate student progression into, through and beyond higher education.

LEAPS achieves these objectives through a comprehensive programme of activities involving student tutoring, student shadowing, workshops, one to one interviews, an annual summer school and a pre application enquiry service. For further details of the LEAPS activities access www.leapsonline.org.

1.3.2 School Groupings

The service LEAPS provides to each of the 46 state secondary schools in Edinburgh and the Lothians is based upon categorisation that depends on the Higher Education progression rate of each individual school, as well as advice from local councils and Careers Scotland. Every two years schools are allocated into one of three groups, based upon their perceived need for involvement with the LEAPS programme:

- Group 1 schools are eligible for all elements of the LEAPS schools programme.
- Group 2 schools are eligible for a slightly reduced programme of events.
- Group 3 schools receive the least support based upon the fact that progression onto higher education is an established and popular route for school leavers.

1.3.3 LEAPS Eligibility

LEAPS works with guidance staff in school in order to identify students who may be entitled to participate in LEAPS activities. In Group 1 schools this is simple - all students match this entitlement (providing they are judged capable or achieving three

or more Higher qualifications over two diets of exams). Students attending schools in Groups 2 or 3 must match one or more of the criteria listed under point 2 below.

The premise of social, economic and cultural disadvantage is the main eligibility criterion. Consequently, the criteria are intended to identify school students who are at risk of academic underachievement or have limited aspirations due to:

1. Negative peer or community influence, for example, attending schools serving communities with little or no tradition of sending young people into higher education and high levels of social deprivation.

2. Other social or economic factors that might inhibit a school student from choosing or progressing onto higher education:

    - Low income family
    - First generation in family to apply to higher education
    - Subject to negative peer and/or community influence
    - Long-term parental unemployment
    - Family in receipt of benefits
    - Overcrowded substandard accommodation
    - Large or complex family
    - Family break up/single parent families
    - Long-term mental/physical illness or disability of a family member
    - Other adverse circumstances (e.g. alcoholism/domestic violence).

Essentially, a student identified as being 'LEAPS eligible' will have an academic ability appropriate to Higher Education but whose ability to fulfil their potential at school may have been affected by one or more of the above social or economic circumstances.

Chapter Two Literature Review

There seems no general consensus on the best term used to describe wider access students. Terminology in use includes describing these students as from under represented groups, low participation neighbourhoods, non traditional and less advantaged back grounds. LEAPS eligible students fall into all these categories and for clarity the term non traditional will be used to broadly describe the wider access student background. Equally studies that have examined how well students perform at university have measured student progression, completion, retention, drop out, wastage, withdrawal, success and failure. Again for clarity this study will look at the successful completion of an academic year in terms of attendance and meeting the required academic standard.

The numbers of students entering higher education has been increasing for more than the last ten years. The numbers of students entering higher education from non traditional backgrounds has also been increasing but not at the same rate as those from more traditional backgrounds. This has most recently been confirmed by Ianelli and Paterson (2005) who found that this growth in participation has “benefited all social classes without reducing social inequalities”.

The gap between the traditional and non traditional students has been attributed to a variety of reasons, not least by Forsyth and Furlong (2000) who found that poorer school performance was a primary cause. Tinklin (2000) suggested that not only were students from non traditional backgrounds leaving school with lower qualifications they were also less likely to apply to higher education and less likely to start the degree courses that they applied for. Many school leavers whose background is not that traditionally associated with going to university feel that higher education is not for them.

Laing and Robinson (2003) define traditional students as normally coming from a family background with previous experiences of higher education. Non traditional are those currently under represented, for example, through lower social class and

whose family members have little experience of higher education. It is against this background that initiatives like LEAPS target able students.

The Select Committee Education and Employment Sixth Report (2001) stated that widening access to higher education is not only a matter of getting into university but "also a matter of staying in and emerging in good standing". While there is a great deal of literature about the "getting in", there is much less about the "staying in". And, the effort has been concentrated on examining the characteristics of those students who dropped out rather than the characteristics of those who successfully stayed in.

Forsyth and Furlong (2003) found that students from non traditional backgrounds are more likely to drop out of courses or follow more complicated paths – changing and repeating courses. Yorke (1999) also recognised that completion is a result of interconnected factors, course choice, fitting in and financial security to name but a few. While many students have considerable pressures such as money worries and a lack of family support, it is fair to say that these pressures may be more acute for students from non traditional backgrounds.

One of the government performance indicators for HEIs is the numbers of students failing to complete. The tendency to attribute lower levels of successful completion with an increase in student numbers from non traditional backgrounds has particularly important implications as it could result in higher education institutions being unwilling to risk admitting wider access students for fear of being branded a poor performer. It is difficult to establish with any degree of confidence whether wider access students do perform differently to those students from more traditional backgrounds without reliable tracking information.

The tracking of wider access students through HEIs is not straightforward. It is made more difficult as there is no standard mechanism for monitoring progress. Musselbrook (2003) recognises that the tracking of particular cohorts of students, particularly those from non traditional backgrounds, has not been a key activity for many institutions and points out the need for clear definitions and consistent terminology. Tracking more often than not is piecemeal in its approach.

Overall figures on student completion and drop out rates can be obtained from the higher education funding councils but different sectors define and measure completion and non completion in different ways. Completion rates vary significantly according to the type and size of institution, the student cohort, the subject studied and the qualification gained. These differences across universities in the recording of student characteristics and courses offered can lead to the higher education institution being negatively represented and, as a result, are not publicised.

Universities UK (2005) identified "barriers to evaluation" and include the:

- difficulty of tracking and impact measuring
- lack of skills and knowledge to implement suitable systems
- reluctance to collect information about the background of students
- limited funding
- concerns about data protection.

There is an absence of hard evidence of tracking wider access students.

For the general student population exhaustive studies have shown that qualifications continue to be the best indicator for academic success. Behkradnia and Thompson (2002) showed that there is a strong association between non progression and academic points. Many LEAPS students do not possess the academic qualifications required to do their chosen course but not necessarily as a result of lack of academic ability. While qualifications are important it can be shown that they are not the only significant factor in predicting successful student progression.

One of the most influential and widely cited models of student non completion is that of Tinto (1975, 1987). He suggests that the students' social and academic integration into the HEI is a major determinant of completion. Factors affecting the integration include the students' family background, personal characteristics in addition to qualifications and extend to include the students' interaction with faculty and staff. Johnes (1990) found that parental social class and school type are significant factors when predicting non completion probabilities. This has been confirmed by Woodley et

al (1992) who also revealed that no one variable could be used as a predictor of success. Instead it was likely to be a combination of factors.

Many of the recent studies examine in particular students failing to complete first year. There is a need for studies to address subsequent years of study. Although the greatest attrition has been shown to occur after year one (56% in the study by Smith and Naylor (2000)), students do continue to drop out after that. Woodley et al (1992) found in their study of non completion rates in eight Scottish Universities that in three of the eight, a larger proportion of students left in the second year or later than in first year.

This study is unlikely to produce a model that will predict with absolute certainty which wider access students will be successful in any academic year as there are factors beyond the student characteristics captured in this study which may have a strong influence. However, Johnston (2000) views a lack of absolute certainty as an advantage as a student with all or most of the characteristics associated with failure may not necessarily fail but may just encounter more problems achieving success.

Chapter Three Statistical Modelling

This chapter sets the framework for the statistical modelling. For those more interested in the results than the statistical methods the chapter can be skipped.

3.1 Linear Regression

The most common example of statistical modelling is linear regression. Linear regression models are of the form

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

where y is a continuous random (dependent) variable and $x_1, x_2, ...x_k$ are regressor or independent variables. It is possible for y to take on any value as *k* ranges between $-\infty$ and $+\infty$.

Simple linear regression provides information relating to how much a dependent variable *y* will change as a result of changes in the independent variables $x_1, x_2, ...x_k$ so we can predict the value of y given the values of x.

3.2 Logistic Regression

Logistic regression analysis is one of the most popular regression techniques available for modelling dichotomous dependent variables. It has the same objectives as linear regression. There are two general applications for logistic regression – prediction and explanation. Prediction involves using a sample to create a regression equation to make credible forecasts concerning an outcome for a particular set of circumstances. Explanation involves attempting to understand an outcome by exploring the relationships between multiple variables and generalising the understanding to a new population.

The simplest logistic regression model is where y is a binary variable with possible values 0 and 1. There are many practical situations where the dependent variable takes only one of two values. For example, life/death, diseased/not diseased and, in

the case of this study, the value 1 represents the successful completion of an academic year and 0 is failure to complete an academic year.
Logistic regression models are used to predict the proportion or probability of an outcome associated with a particular set of circumstances. In linear regression the value of the dependent variable is unrestricted. A proportion or probability must lie between 0 and 1 so a transformation is required to limit the range to between 0 and 1 (Krzanowski, 1998).

The odds of an event happening is the probability that the event will happen divided by the probability that the event will not happen. The odds is a way of expressing how often something happens relative to something else happening.

The odds of an event can be written

$$odds = \frac{p}{1-p} = \frac{probability\_of\_event}{probability\_of\_no\_event}$$

and rewritten

$$probability\_of\_event = \frac{odds}{1+odds}$$

In logistic regression the logarithm of the odds of an event is used as the dependent variable. This is known as the logit of p.

$$\log it(p) = \ln\left(\frac{p}{1-p}\right)$$

This important transformation has many of the desirable properties of a linear regression model.

The general logistic regression model in terms of independent variables $x_1, x_2, ...x_k$ can be written

$$\ln\left(\frac{p}{1-p}\right) = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$$

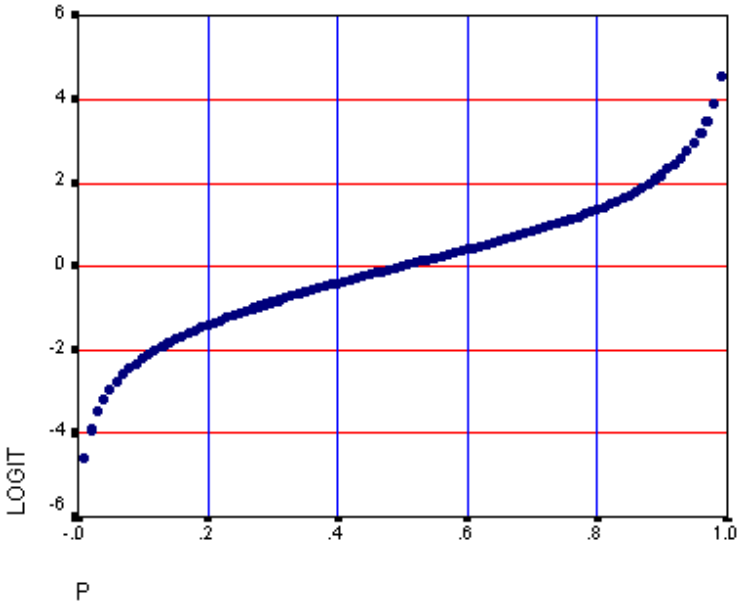
where p is the proportion and $\beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$ is the linear function of the independent variables.

This can be rewritten

$$p = \frac{e^{\beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k}}{(1 + e^{\beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k})}$$

The logit model assumes a non linear relationship between the probability of an event occurring and the independent variables.

Plotting p against logit (p) gives an s shaped curve which approaches 0 and 1 asymptotically.



This S shaped curve is characteristic of all logistic regression models ensuring that the probabilities are always between 0 and 1. The coefficients can then be interpreted as the amount of increase (or decrease) in the log-odds of the dependent variable y for each unit increase in x.

3.2.1 Dummy Variables

If some of the independent variables are discrete, such as sex and ethnicity it is quite common for them to be defined as factors. The numbers used to identify the different levels of factors have no numerical relationship and are called dummy variables. For example, suppose that one of the independent variables is sex. The dummy variables can have the value zero for female and one for male. Two or more levels of a factor can be distinguished by each dummy variable taking the value one for the level under consideration and zero value for the rest of the levels. Continuous variables can be grouped in this way. The number of dummy variables is one minus the number of factor levels.

A full explanation of logistic regression analysis can be found in Hosmer and Lemeshow (2000).

3.3 Model Building Strategy

It is widely accepted and cited by Krzanowski (1998) that the simpler or more parsimonious the statistical model the better. When, as in this study, there are many independent variables there is a compromise between including as many variables as possible to best explain the data and ending up with a simple workable and cost efficient model.

Although logistic regression finds a best fit equation as in linear regression, the principles on which it does it are different. Instead of using least squares it uses maximum likelihood which seeks to maximise the probability of getting the observed values of the dependent variable from the observed values of the independent variables. Maximum likelihood estimates produce results through an iterative procedure. When two successive approximations yield no significant improvement in the probability then the algorithm has converged.

As a result the goodness of fit and overall significance statistics used in logistic regression are different from those used in linear regression. Choosing the most appropriate summary statistic upon which to base the decisions requires careful consideration.

3.4 Developing the Model

There are three main phases in developing the statistical model. These involve identifying the initial model variables, evaluating and refitting the model, and interpreting the model.

3.4.1 Identifying the Initial Model Variables

The steps followed in developing this model follow those recommended by Hosmer and Lemeshow (2000). The selection process begins with univariate analysis of the variables and a chi-square test. The Pearson chi-square is used to test the hypothesis that there is no association between the variables. It calculates a double sided probability value (p-value) for the relationship between two dichotomous variables. It is this chi-square value which first signals which variable should or should not be included in the model.

Some variables are kept in the model as it is believed by experts in the field that they are both relevant and pertinent to the result regardless of statistical significance. In addition variables with a p-value < 0.25 are selected. This p-value, whilst capturing important variables, does lead to the inclusion of variables which only exhibit a questionable level of significance; therefore it is important to rigorously reassess variables before deciding on the final model.

3.5 Evaluating and Refitting the Model

Armed with the initial model this next phase involves some fine tuning. The process by which the variables are next tested for significance for inclusion or elimination from the model involves different criteria.

3.5.1 Log Likelihood

The log likelihood is one of the criterion for establishing parameters. Garson (2006) describes likelihood as the probability that the observed values of the dependent variable may be predicted from the observed values of the independent variables. The log likelihood, when multiplied by -2 has approximately a chi-square distribution.

3.5.2 Likelihood Ratio Test or Model Chi-Square for Overall Model

Statistical packages present the –2LL (intercept only) and -2LL (model with all explanatory variables). The difference between these two values forms the basis of the likelihood ratio test. This is used to assess the *overall* logistic model but does not give the relative importance of the variables. Larger values of –2LL indicate a worse prediction of the dependent variable and can be used as a measure of how poorly the model fits.

3.5.3 Likelihood Ratio Test for Individual Model Parameters

The likelihood ratio test can also be used to indicate whether particular independent variables are more important than others. When the models in question are nested, that is all the variables in the smaller model are in the larger model, the likelihood ratio test can be used to determine the statistical significance of the contribution of an independent variable to the explanation of the dependent variable. Firstly compare the difference in -2 times the log of the likelihood (-2LL) for the overall model with and without the variables in question. Secondly if the difference in the associated chi-square values is below the critical value then it can be concluded that the variable can be dropped as it makes no difference to the prediction. Menard (1995) declares this the best and most accurate test.

3.5.4 Wald Statistic

The Wald Statistic is also commonly used to test the statistical significance of the coefficients. It is the coefficient divided by its standard error and squared, similar to a z test. In general the Wald Statistic and the likelihood test give approximately the same value when the sample size is large. Menard (1995) warns of the use and pitfalls of the Wald Statistic. For very high coefficient values the Wald Statistic can

give wrong results (an increased type II error). Bannerjee and Wellner (2005) also recognised the Wald Statistic as being less powerful than the likelihood ratio test, so, the preferred measure will remain the difference in -2LL values.

3.6 Interactions

Once a model has been built that contains the essential variables the need for interactions should be considered. When the effect of one independent variable depends upon the level of another then it can be said that an interaction exists between these two variables. Consider, again, the general logistic model in terms of independent variables x_1 and x_2. If x_1 is ethnicity and x_2 is sex, then the interaction x_1x_2 shows whether the effect of ethnicity is different for males or females.

$$\ln\left(\frac{p}{1-p}\right) = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

The likelihood ratio test can be used to assess the significance of interactions. Variables and interactions that are statistically significant are likely to have predictive value.

3.7 Assessing the Model Fit

There are many goodness of fit indices. In some significance means fit and in others significance means lack of fit.

3.7.1 Hosmer and Lemeshow Goodness of Fit Test

The Hosmer-Lemeshow Goodness of Fit Test is performed by creating 10 ordered groups of subjects based on predicted probabilities and then comparing the number actually in the each group (observed) to the number predicted by the logistic regression model (expected). Thus, the test statistic is a chi-square statistic with a desirable outcome of non-significance, i.e. a large p-value, indicating that the model prediction does not significantly differ from the observed.

3.7.2 Area Under the ROC (Receiver Operating Characteristic) Curve

The ROC curve is a graphical representation of the trade off between false negative rates and false positive rates. It provides a measure of the model's ability to discriminate between cases with and without the outcome of interest. The area under the ROC curve ranges from zero to one, one being a perfect ability and 0.5 being worthless. This discriminative ability or accuracy can be judged by the general rule from Hosmer and Lemeshow (2000).

Area under ROC	Discrimination Ability
If ROC=0.5	None, flip a coin instead
If 0.7≤ROC<0.8	Acceptable Discrimination
If 0.8≤ROC<0.9	Excellent Discrimination
If ROC≥0.9	Outstanding Discrimination

3.7.3 R-Square

The strength of association can also be measured using a variety of R-squares. These measures are intended to mimic R-square used in assessing model fit in linear regression but do not represent the proportion of variance explained. In logistic regression low R-square values are the norm and their routine publishing is not recommended by Hosmer and Lemeshow (2000). Their use is only recommended in addition to other fit statistics though they can be useful to evaluate competing models in the model building stage.

3.7.4 Akaike's Information Criterion (AIC)

The AIC is useful when comparing the goodness of fit of different models. It takes into account the number of parameters in the model and penalises a model that has too many parameters. In general the preferred model is usually the one with the smaller AIC.

3.8 Parameter Estimates

The parameter estimate is the maximum likelihood estimate of the parameter values which makes the observed data the most probable. Parameter estimates are not straightforward to interpret but they are used to calculate probabilities.

3.9 Interpreting Coefficients Using the Odds Ratios

Parameter estimates can be converted easily into an odds ratio by using the exponential function.

Odds Ratio = $e^{ßi}$ where $ßi$ is the maximum likelihood parameter estimate

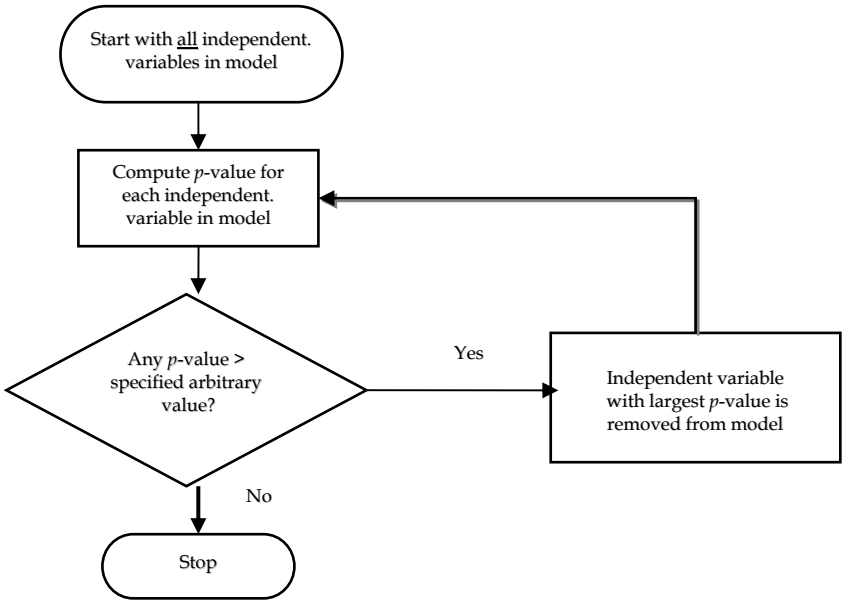
The odds ratio compares whether the probability for a certain event is the same for two groups. The odds ratio provides no additional information but is easier to interpret and as a result is a widely used measure of effect. If an odds ratio is less than 1 then the outcome is less likely to occur; greater than 1 then the outcome is more likely to occur. If the confidence interval on the odds ratio includes 1 then that variable is not a useful predictor.

3.10 Automatic Selection Procedures

These include backwards elimination, forward selection, stepwise regression and best subsets selection. They add or remove variables one at a time until all variables meet the required level of significance.

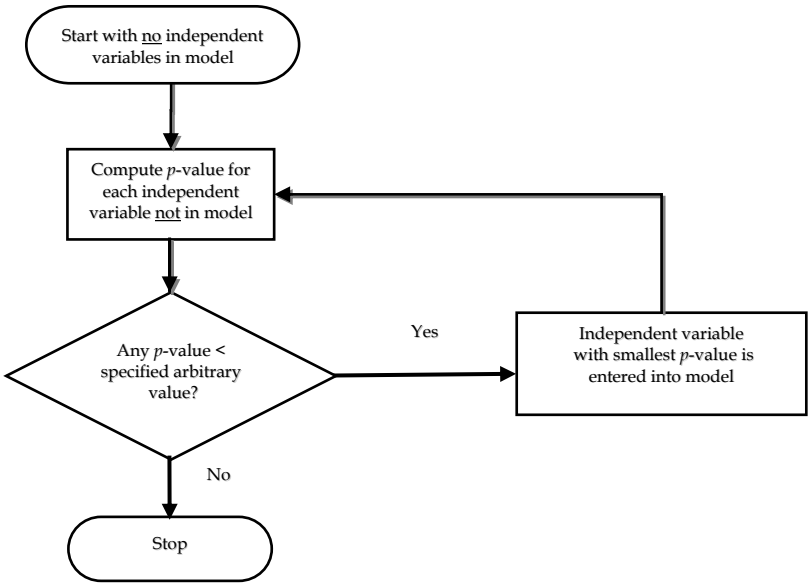
3.10.1 Backward Elimination

Backward elimination starts with a model that includes all the independent variables. The variable that is least significant, the one with the largest p-value, is removed and the model is refitted. Each subsequent step removes the least significant variable in the model until all remaining variables have individual p-values smaller than an arbitrary small value, typically 0.05. Once a variable is removed from the model it stays out. See flowchart overleaf.



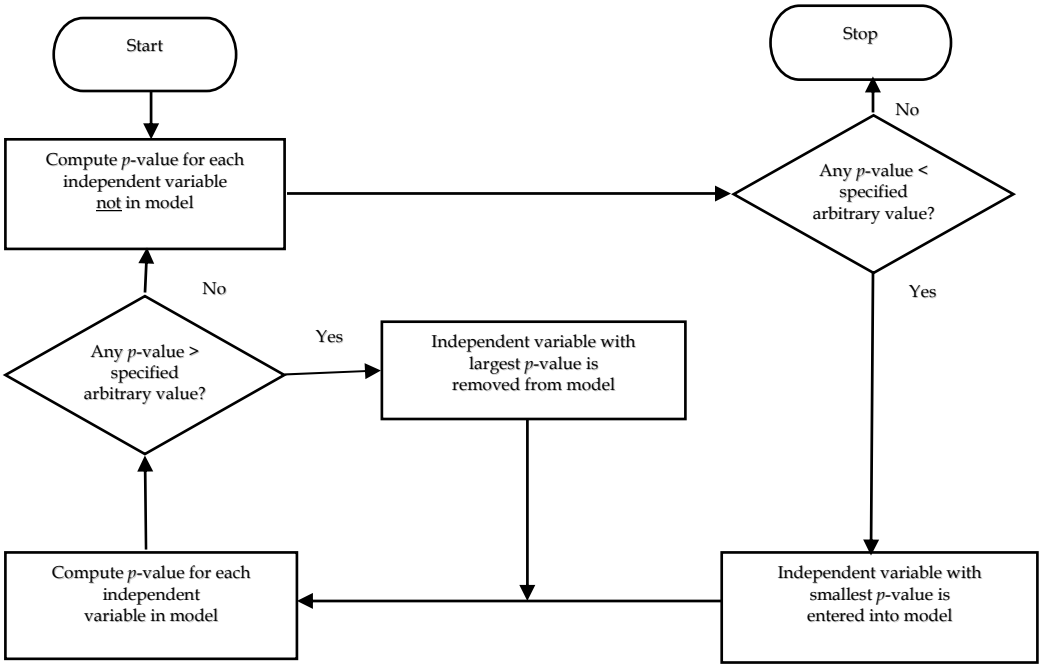
3.10.2 Forward Selection

Forward selection starts with no independent variables. The variable which has the smallest p-value when it is the only variable in the model is the first to be included in the model. Each subsequent step adds the variable that has the smallest p-value. Variables are added one at a time as long as their p-values are small enough, typically less than 0.05. Once a variable is included in the model it remains there. See flowchart below.



3.10.3 Stepwise Regression

Stepwise regression is similar to forward selection except that variables are removed from the model if they become non significant as other variables are added in. Unlike backward elimination and forward selection, variables already in or out of the model do not necessarily stay there. See flowchart below.



3.10.4 Best Subsets Selection

An alternative to stepwise selection of variables is best subset selection. The procedure uses the branch and bound algorithm of Furnival and Wilson (1974) to find a specified number of best models containing one, two, three variables and so on, up to the single model containing all of the explanatory variables. This technique compares models of the same size.

A common view about automatic selection procedures is cited by Dallal (2001) when he states that backwards elimination has an advantage over forward selection and stepwise regression because it is possible for a set of variables to have considerable predictive capability even though any subset of them does not. Forward selection and stepwise regression will fail to identify these variables because the variables do not predict well individually. As a result they will never enter the final model and their joint

behaviour will be unnoticed. Backwards elimination starts with everything in the model, so their joint predictive capability will be seen. Allison (1999) is simply ‘not a fan of such methods’.

3.11 Quasi Complete Separation

Complete separation, described by Allison (1999), occurs when there is some linear combination of independent variables that perfectly predicts the dependent variable. Quasi complete separation is the same situation except for a single or only a few values of the predictor for which both the values of the dependent variable occur.

If the data are partially or completely separated it may not be possible to obtain maximum likelihood estimates. The analysis may fail to converge and other symptoms include large coefficients and large standard errors. This can occur if there are more parameters in the model than can be estimated because the data are sparse.

Chapter Four Univariate Statistics

4.1 The Data

LEAPS undertakes a series of one to one interviews throughout the 46 secondary schools in Edinburgh and the Lothians. Every student that is LEAPS eligible and deemed able by school staff is interviewed. The interviews take approximately half an hour. The interviewers are all education professionals who have undergone additional training. Many of them are very experienced in LEAPS interviewing. Throughout the course of the interview a seven page form is completed covering student details, academic qualifications, other qualifications and work experience, course choices and other relevant data relating to their eligibility. The student retains a copy of the completed application form which forms the basis of the personal statement required by the Universities and Colleges Admission Service (UCAS).

One of the strengths of this study is the quality and reliability of the data.

The data set for this study is based on the anonymised LEAPS student records for all students who attended one of the LEAPS partner institutions or the University of Stirling. The data contain information on 1455 students who entered university at the start of the academic year 2001/2002 to 2004/2005 up to their academic status at the end of the 2004/2005 academic year.

Table 4.1.1 Student Numbers by Academic Year

	Academic Year				
	2001/2002	2002/2003	2003/2004	2004/2005	Total
First Year	305	337	384	429	1455
Second Year		250	274	322	846
Third Year			226	254	480
Fourth Year				199	199

Table 4.4.1 shows that 305 LEAPS students were tracked in their first year from entry in 2001/2001. Of those, 250 returned and were tracked through their second year in 2002/2003 as well as 337 LEAPS students who entered first year in 2002/2003.

4.2 Methodology

An Excel spreadsheet was created from the records held in the LEAPS office. It included the student's name and a coded set of background characteristics. The HEIs added the student qualifications and the end of year outcome. The student's name was then deleted. As each subsequent year of data was sent, the previous year's data was updated with the latest end of year outcome. Although the LEAPS database contained some information relating to the student qualifications it was decided to use the university information as it would have taken account of any changes resulting from appeals. The previous tracking study (Sinclair and McClements (2003)) had shown there to be some discrepancies between the qualifications the student thought they had and the qualifications the HEI was accepting them with.

The first, and one of the most time consuming steps, was the checking of data. The Excel file sent to the HEIs contained only enough student details for the HEIs to be able to identify them. Once the HEIs had added the qualification data and progression codes the files were expanded to include all other data held by LEAPS. Careful checks were required to ensure that the data were correct for all students.

4.3 Results

This first section of results reports some univariate statistics for the whole group and by destination. (Note: Due to rounding the percentages in the tables may not always total 100%).

4.3.1 Overall numbers

Table 4.3.1 details the overall student numbers in the study and the destination HEI they attended.

Table 4.3.1 Overall Numbers

Destination	Frequency	Percent
Edinburgh College of Art (ECA)	15	1%
The University of Edinburgh	415	28.5%
Heriot-Watt University	367	25.2%
Napier University	418	28.7%
Queen Margaret University College (QMUC)	112	7.7%
University of Stirling	128	8.8%
Total	1455	100%

Just over 82% of the students in the study attended The University of Edinburgh, Heriot-Watt University and Napier University.

4.3.2 Summer School

Table 4.3.2 indicates that around one quarter of all the students had an involvement with the LEAPS summer school with 20% successfully completing it. No students that applied to Edinburgh College of Art attended summer school. Heriot-Watt University had the highest proportion of its intake as summer school students at 23% and (ECA apart) QMUC the lowest at 17%.

Table 4.3.2 Destination by Summer School

	Successfully completed Summer School	Withdrew from Summer School	Did not attend Summer School	Total
ECA	0	0	15	15
	0%	*0%*	*100%*	
Edinburgh	84	15	316	415
	20%	*4%*	*76%*	
Heriot-Watt	84	20	263	367
	23%	*5%*	*72%*	
Napier	77	29	312	418
	18%	*7%*	*75%*	
QMUC	19	9	84	112
	17%	*8%*	*75%*	
Stirling	26	8	94	128
	20%	*6%*	*73%*	
Total	290	81	1084	1455
	20%	*6%*	*74%*	

Additionally:

- 63% of summer school students are female
- Almost 50% of summer school students are from LEAPS group1 schools
- 14% of summer school students are from a non white family background
- 82% of summer school students are first generation in their family to attend university.

Summer school students are:

- more likely to study Combined Studies; Languages, English, History, Geography, Divinity and Journalism; Medicine, Dentistry and Veterinary Science; Social Studies, Law and Accountancy; Business and Administrative Studies, Hospitality and Tourism
- less likely to study Creative Arts; Mathematical, Computer and Information Science; and Subjects Allied to Medicine.

4.3.3 Gender

Table 4.3.3 shows that ECA and QMUC have a very strong female bias to their intake. Heriot-Watt University is the only institution with a male bias. The LEAPS student mix is representative of the whole student population in terms of gender with an overall 58/42 female/male split.

Table 4.3.3 Destination by Gender

	Female	Male	Total
ECA	12	3	15
	80%	*20%*	
Edinburgh	272	143	415
	66%	*34%*	
Heriot-Watt	149	218	367
	41%	*59%*	
Napier	221	197	418
	53%	*47%*	
QMUC	102	10	112
	91%	*9%*	
Stirling	88	40	128
	69%	*31%*	
Total	844	611	1455
	58%	*42%*	

Additionally:

- Females are more likely to study Subjects Allied to Medicine; Education; and Medicine, Dentistry and Veterinary Science
- Males are more likely to study Mathematical, Computer and Information Science; Engineering and Technology; and Architecture, Building and Planning.

4.3.4 Ethnicity

Student ethnicity can be quantified by the student assigning themselves to an ethnic group. The groups and associated two digit codes are shown in Appendix (i).Table 4.3.4 establishes that the student group are largely white (overall 92%). It is difficult to compare this with published figures for the student body as a whole as published figures contain a significant percentage of 'unknowns' or 'information declined'. In this study there are less than 1% unknowns or information declined.

Table 4.3.4 Destination by Ethnicity

	white	non white	unknown	Total
ECA	13	0	2	15
	87%	*0%*	*13%*	
Edinburgh	387	25	3	415
	93%	*6%*	*1%*	
Heriot-Watt	325	41	1	367
	89%	*11%*	*<1%*	
Napier	378	31	9	418
	91%	*7%*	*2%*	
QMUC	107	3	2	112
	95%	*3%*	*2%*	
Stirling	126	2	0	128
	98%	*2%*	*0%*	
Total	1336	102	17	1455
	92%	*7%*	*1%*	

Additionally:

- Heriot-Watt University has the greatest number (11%) of students from a non white family background
- More than half (52%) of non white students attended summer school
- Less than a quarter of white students attended summer school

- 12% of students from the City of Edinburgh are non white
- 4% of students from Midlothian and West Lothian are non white
- East Lothian has the highest percentage of white students (98%).

4.3.5 LEAPS School Group

Table 4.3.5 reveals that ECA has the highest proportion (60%) of LEAPS group 1 school students in its intake. Heriot-Watt University has the largest number (187 students).

Table 4.3.5 Destination by LEAPS School Group

	Group 1	Group 2	Group 3	Total
ECA	9	3	3	15
	60%	*20%*	*20%*	
Edinburgh	168	74	173	415
	40%	*18%*	*42%*	
Heriot-Watt	187	59	121	367
	51%	*16%*	*33%*	
Napier	165	73	180	418
	39%	*17%*	*43%*	
QMUC	41	21	50	112
	37%	*19%*	*45%*	
Stirling	46	32	50	128
	36%	*25%*	*39%*	
Total	616	262	577	1455
	42%	*18%*	*40%*	

Additionally

- There are even proportions of gender and age group across the LEAPS school groupings
- Group 3 schools have the highest percentage of non white students (10%)
- The highest percentage of LEAPS group 3 school students come from Midlothian (62%)
- The highest percentage of LEAPS group 2 school students come from East Lothian (35%)
- The highest percentage of LEAPS group 1 school students come from the City of Edinburgh (52%).

4.3.6 Destination by Council

It can be seen from Table 4.3.6 that all the partner council areas are (fairly) equally represented in the study based on school roll numbers. In addition there were three students from Fife.

Table 4.3.6 Destination by Council

	City of Edinburgh	East Lothian	Midlothian	West Lothian	Fife	Total
ECA	6	2	2	5	0	15
	40%	*13%*	*13%*	*33%*	*0%*	
Edinburgh	200	58	70	85	2	415
	48%	*14%*	*17%*	*20%*	*<1%*	
Heriot-Watt	150	43	49	125	0	367
	41%	*12%*	*13%*	*34%*	*0%*	
Napier	192	51	88	86	1	418
	46%	*12%*	*21%*	*21%*	*<1%*	
QMUC	54	19	14	25	0	112
	48%	*17%*	*13%*	*22%*	*0%*	
Stirling	47	21	12	48	0	128
	37%	*16%*	*9%*	*38%*	*0%*	
Total	649	194	235	374	3	1455
	45%	*13%*	*16%*	*26%*	*<1%*	

Additionally:

- Three quarters of students attending Heriot-Watt University and the University of Stirling are from the City of Edinburgh or West Lothian
- QMUC takes the highest percentage of students from East Lothian (17%)
- 52% of summer school students are from the City of Edinburgh
- West Lothian Council has the lowest percentage of attendees at summer school (14%).

4.3.7 Age Group

Students are divided into two age groups, shown in Table 4.3.7. Students who were 18 or over by the 1st of October in their first year are assigned to group 1 and those under 18 are assigned to group 2. The overall student mix is a ratio of around 60:40 group 1 to group 2.

Table 4.3.7 Destination by Age Group

	Age Group 1	Age Group 2	Total
ECA	9	5	14
	64%	*36%*	
Edinburgh	246	164	410
	60%	*40%*	
Heriot-Watt	231	134	365
	63%	*37%*	
Napier	256	156	412
	62%	*38%*	
QMUC	64	45	109
	59%	*41%*	
Stirling	76	52	128
	59%	*41%*	
Total	882	556	1438
	61%	*39%*	

4.3.8 First Generation

Students are asked by their school if they are the first in their family to go to university. If this is the case then they automatically become LEAPS eligible and if they attend a LEAPS group 2 or group 3 school then they are eligible to take part in the LEAPS activities selected for that school and to attend summer school. A student is still classed as first generation if a sibling is (or has been) attending university. The number of first generation students by destination is presented in Table 4.3.8 and it shows that Napier University has the highest percentage of first generation students (86%).

Table 4.3.8 Destination by First Generation

	First Generation		
	No	Yes	Total
ECA	4	11	15
	27%	*73%*	
Edinburgh	125	284	409
	31%	*69%*	
Heriot-Watt	72	288	360
	20%	*80%*	
Napier	58	355	413
	14%	*86%*	
QMUC	27	82	109
	25%	*75%*	
Stirling	43	81	124
	35%	*65%*	
Total	329	1101	1430
	23%	*77%*	

Additionally:

- Almost 60% of females are first generation
- 70% of LEAPS group 1 school students are first generation
- 81% of LEAPS group 2 and group 3 school students are first generation
- More than three quarters (77%) of white students are first generation, compared with 69% of non white students
- There are more first generation students from Midlothian – just over 83%.

4.3.9 Subject Grouping

The subject groups were coded in accordance with the Higher Education Funding Council for England (HEFCE) guidelines (see Appendix (ii)).

The top five subject groupings were

Code	Subject Groupings	Percentage of Students
H	Social Studies, Law and Accountancy	19%
I	Business and Administrative Studies, Hospitality and Tourism	14%
C	Biological and Physical Sciences	12%
E	Mathematical, Computer and Information Science	12%
F	Engineering and Technology	9%

Some subjects are only offered at individual universities, for example, Medicine and Education are only offered at the University of Edinburgh. This compares with Business and Administrative Studies which is offered by all except Edinburgh College of Art.

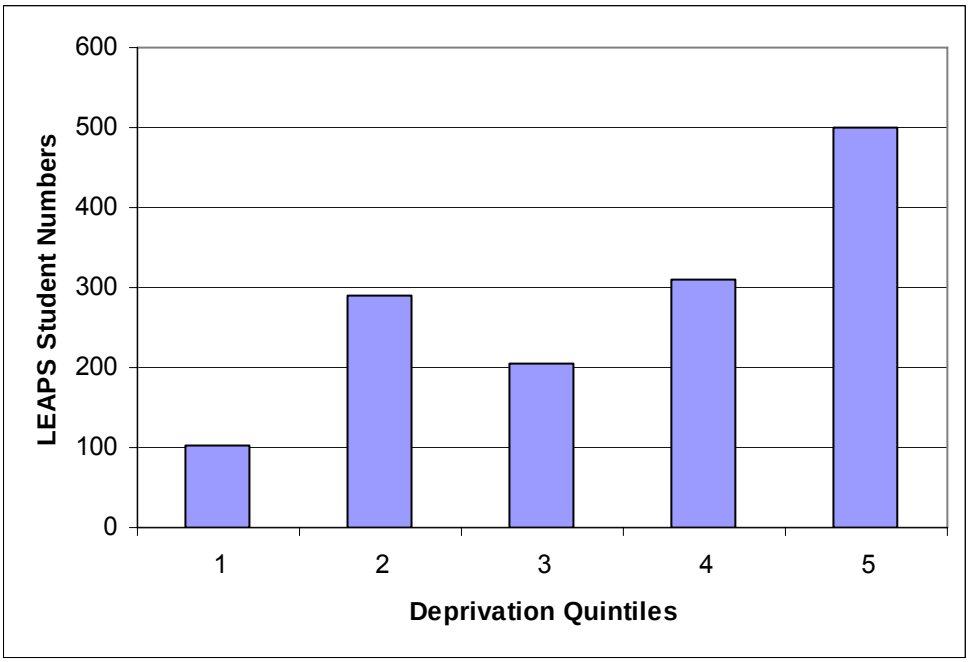
Additionally:

- Almost 40% of those students studying Biological and Physical Science do so at the University of Edinburgh.
- 53% of those studying Subjects Allied to Medicine do so at Napier University and a further 41% at QMUC.
- Heriot-Watt University takes almost 50% of the students studying Mathematical, Computer & Information Science and Engineering and Technology while Napier University takes around one third of the students in these two groups
- The University of Edinburgh take nearly 50% of the students studying Languages, English, History, Geography, Divinity and Journalism while the University of Stirling take around a quarter.

4.3.10 Post Code Analysis

The Scottish Index of Multiple Deprivation (SIMD) was used to analyse the post codes. The SIMD has been designed by the Scottish Executive to rank the data zones used for the production of Scottish Neighbourhood Statistics in order of deprivation. There are 6,505 data zones and the SIMD is based on 31 indicators in the seven individual divisions of Current Income, Employment, Housing, Health, Education, Skills and Training, and Geographic Access to Services and Telecommunications. More information can be found at the SIMD website http://www.scotland.gov.uk/stats. There is always a population associated with a post code and this has always been one of the criticisms of any post code analysis. However the SIMD uses the *whole* post code for categorisation therefore minimising the population associated with any one post code.

Graph 4.3.10 Post Codes by SIMD Quintiles

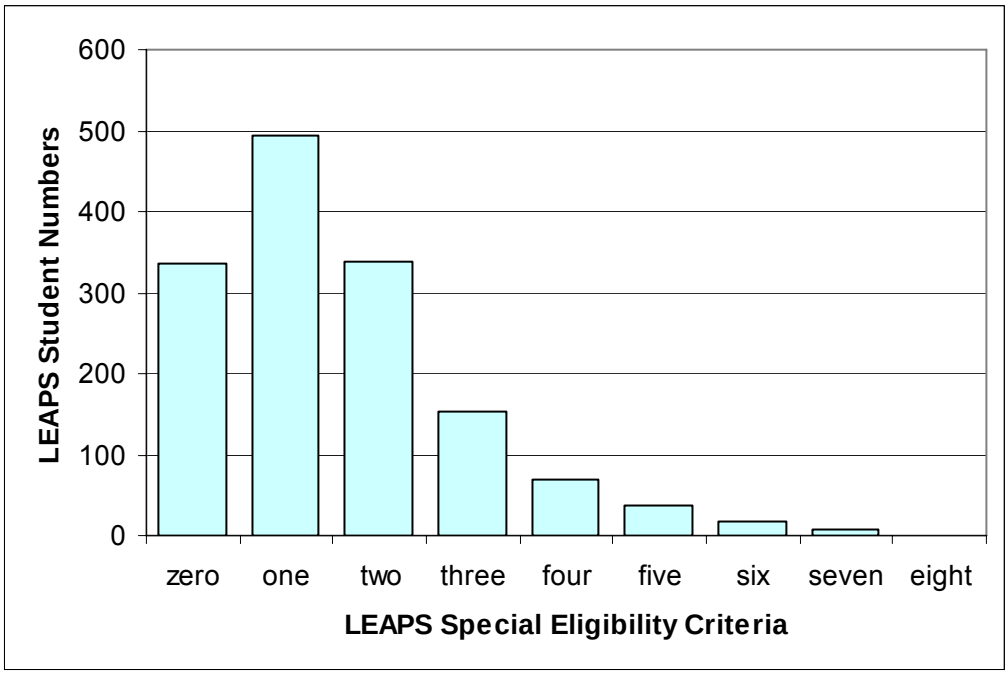


Deprivation quintile 1 is the most deprived 20% of the population through to deprivation quintile 5 which represents the least deprived 20% of the population. The ideal situation would be equal numbers of students from all deprivation quintiles. The graph indicates that there are almost five times as many LEAPS students from the least deprived area as the most. This figure is disappointing. It seems in line with the Scottish Funding Council (2004) figure for ancient universities but only one of the universities in this study is an ancient university. What is encouraging is the number of students from the second most deprived areas (21%). This suggests that the LEAPS programme is succeeding for students from deprived areas but that there is still a requirement for further help for those students from the *most* deprived areas.

4.3.11 Special Eligibility Criteria

The special eligibility criteria are discussed with the student during the pre-application interview and the category (or categories) that most closely resemble the situation of the student is recorded. Students can highlight any number of categories that apply to them. 625 (43%) have indicated that there are two or more personal reasons for them being LEAPS eligible.

Graph 4.3.11.1 Students by Special Eligibility Criteria

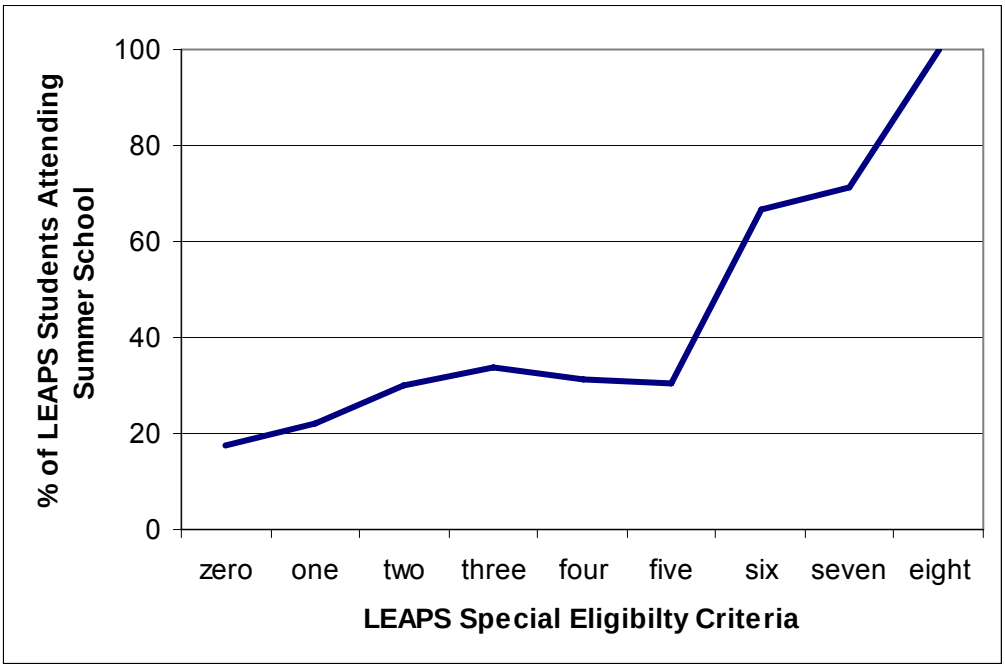


The relatively high number (335) of students shown with zero special eligibility criteria will have qualified for the LEAPS programme by being either first generation and/or attending a LEAPS group 1 school. Less than 10% of the students qualify with four or more special eligibility criteria.

Additionally:

- Around one fifth (20% to 23%) of students studying at Heriot-Watt University, Napier University, QMUC and the University of Stirling have three of more eligibility criteria
- 26% of students with three or more eligibility criteria are in the lowest quartile of qualifications after fifth year while 13% of these students are in the highest quartile of qualifications after fifth year (see Section 4.4.1)
- 24% of students with three or more eligibility criteria are in the lowest quartile of qualifications achieved in both fifth and sixth year while 14% of these students are in the highest quartile of qualifications achieved in both fifth and sixth year (see Section 4.4.2).

Graph 4.3.11.2 Summer School Attendance by Special Eligibility Criteria

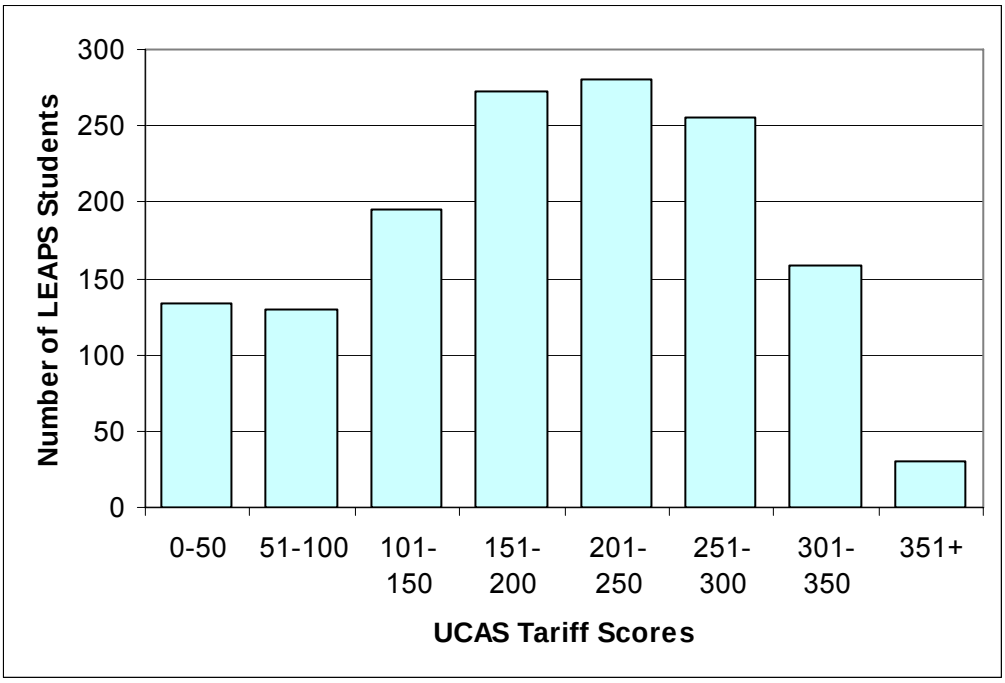


The more eligibility criteria that a student meets, the more likely they are to attend summer school.

4.4 Entrance Qualifications

Secondary school qualifications form the basis for an offer of a place at university. The qualifications achieved by each student in both fifth year (S5) and sixth year (S6) are known and include Intermediate 2 and Higher Grades in S5 and Intermediate 2, Higher and Advanced Higher Grades in S6. Students entering S5 are normally aged 16+. The qualification profile of the students was examined after S5 and after S6. In order to make any comparisons the qualifications have been converted to the UCAS Tariff scheme. The UCAS Tariff scheme is a points system designed to provide a basis for comparison between applicants with different types of qualifications. The points associated Scottish Qualifications are shown in Appendix (iii).

Graph 4.4.1 S5 Qualifications



Graph 4.4.1 shows the distribution of S5 scores. At the end of S5, 50% of students do not have the minimum entry requirement of 200 UCAS points as recommended by Napier University.

For ease of analysis the S5 qualifications are divided into four groups using dummy variables. These groups correspond to the quartiles with 25% of the student body in each group. Group 1 is associated with the lowest qualified students and group 4 is associated with the highest qualified students.

The S5 UCAS points score for each quartile are:

Qualification Group 1 (lowest qualified)	Up to 126 points
Qualification Group 2	127 – 197 points
Qualification Group 3	198 – 263 points
Qualification Group 4 (most highly qualified)	264+

Table 4.4.1 Destination by Qualifications after S5

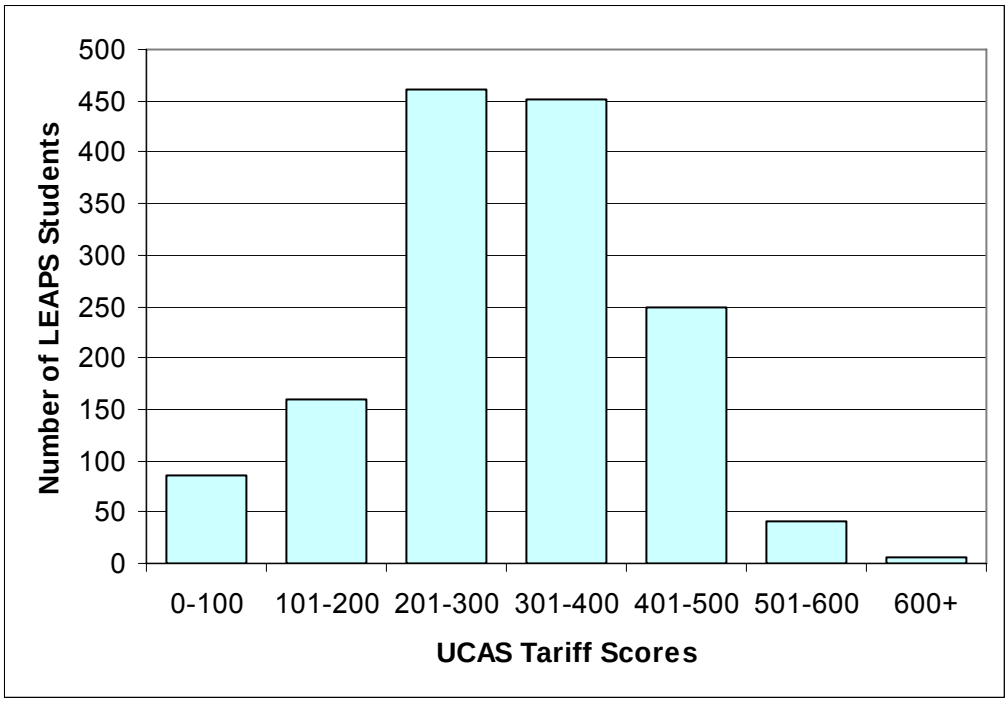
	S5 Qual Grp 1	S5 Qual Grp 2	S5 Qual Grp 3	S5 Qual Grp 4	Total
ECA	1	3	6	5	15
	7%	*20%*	*40%*	*33%*	
Edinburgh	29	45	90	251	415
	7%	*11%*	*22%*	*61%*	
Heriot-Watt	45	94	131	97	367
	12%	*26%*	*36%*	*26%*	
Napier	206	123	60	29	418
	48%	*31%*	*14%*	*7%*	
QMUC	29	42	28	13	112
	26%	*38%*	*25%*	*12%*	
Stirling	52	43	27	6	128
	41%	*34%*	*21%*	*5%*	
Total	362	350	342	401	1455
	25%	*24%*	*34%*	*28%*	

The University of Edinburgh takes the most students (61%) from the highest S5 qualification group. Almost half the Napier University intake is from S5 qualification group 1.

Additionally:

- 62% of students who attended summer school are in qualification groups 1 and 2
- Males and females are fairly evenly distributed among the qualification groups
- Three quarters of the students attending the University of Stirling are in qualification groups 1 and 2
- There are proportionally more non white students (30%) than white students (24%) in the lowest S5 qualification group
- There are proportionally more white students (28%) than non white (24%) students in the top S5 qualification group.

Graph 4.4.2 Combined Qualifications



A combined total score for the qualifications at the end of S6 is carefully coded to avoid double counting for the same or similar qualifications. Graph 4.4.2 shows the distribution of these combined scores. At the end of S6, 17% of students still do not have the minimum recommended 200 UCAS points.

Again, for ease of analysis the combined qualifications are divided into four groups corresponding to the quartiles. Once again group 1 is the lowest qualified group and group 4 the highest qualified group.

The combined qualification UCAS points score for each quartile are:

Qualification Group 1 (lowest qualified)	Up to 226 points
Qualification Group 2	227 – 305 points
Qualification Group 3	306 – 383 points
Qualification Group 4 (most highly qualified)	384+

Table 4.4.2 Destination by Combined Qualifications after S6

	Qual Grp 1	Qual Grp 2	Qual Grp 3	Qual Grp 4	Total
ECA	4	6	5	0	15
	27%	*40%*	*33%*	*0%*	
Edinburgh	17	60	116	222	415
	4%	*14%*	*28%*	*53%*	
Heriot-Watt	38	104	118	107	367
	10%	*28%*	*32%*	*29%*	
Napier	201	129	59	29	418
	48%	*31%*	*14%*	*7%*	
QMUC	27	32	36	17	112
	24%	*29%*	*32%*	*15%*	
Stirling	48	52	15	13	128
	38%	*41%*	*12%*	*10%*	
Total	335	383	349	388	1455
	23%	*26%*	*24%*	*27%*	

More than half (53%) of the students entering the University of Edinburgh are from Qualification Group 4 while just under half (48%) of the students entering Napier University are from Qualification Group 1.

Additionally:

- Again, 62% of students who attended summer school are in qualification groups 1 and 2
- Nearly 80% of the students attending Napier University and the University of Stirling are in qualification groups 1 and 2
- Females are evenly distributed among all combined qualification groups (around a quarter in each). Males vary more with 21% in group1, 29% in group 2, 23% in group 3 and 28% in group 4
- There is almost double the number of non first generation students in group 4 than in any other combined qualification group
- There are proportionally more non white students (31%) than white students (22%) in the lowest combined qualification group
- There are proportionally more white students (27%) than non white (19%) students in the top combined qualification group.

4.5 End of Year Outcomes

Staff from the HEIs were asked to allocate one of the following nine 'end of year' outcome categories to each of their students:

1. Normal progression to same programme/group of programmes
2. Normal progression to different programme/group of programmes
3. Repeating some or part of a year
4. Temporary Withdrawal
5. Award Achieved
6. Admission/transfer to year1 of a new programme within Institution
7. Transfer to another HE Institution
8. Permanent withdrawal
9. Other

What is a successful outcome? Categories 1, 2, 3 and 5 are considered a successful outcome. Category 7 could be a successful outcome for the student if they are continuing their studies at another institution rather than restarting them but this level of information is not known. It is arguable, though, whether category 7 is ever a successful outcome for the institution from which the student is transferring.

It should be noted here that many of the institutions code the end of year outcomes differently and the converting of one set of codes to another does leave areas of grey.

Table 4.5.1 Successful Outcome by Destination

	First Year	Second Year	Third Year	Fourth Year
ECA	*93%*	*100%*	*n/a*	*n/a*
Edinburgh	*89%*	*91%*	*93%*	*100%*
Heriot-Watt	*84%*	*93%*	*95%*	*100%*
Napier	*87%*	*93%*	*97%*	*96%*
QMUC	*86%*	*97%*	*89%*	*100%*
Stirling	*92%*	*96%*	*96%*	*100%*
Overall	*87%*	*93%*	*94%*	*99%*
	1455 students	846 students	480 students	199 students

The figures in Table 4.5.1 represent the LEAPS students who successfully complete a year, that is, they pass the minimum academic and attendance requirements to start the subsequent year. This is not always the same as those who turn up for any

subsequent years. There are a small number of students who pass a particular year but do not return.

Table 4.5.2 End of Year Outcome by Destination for Students Completing First Year

	End of Year Outcome Category									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	% success
ECA	0	14	0	0	0	0	0	1	0	
	0%	*93%*	*0%*	*0%*	*0%*	*0%*	*0%*	*7%*	*0%*	*93%*
Edinburgh	325	22	21	4	0	2	0	37	1	
	79%	*5%*	*5%*	*1%*	*0%*	*<1%*		*9%*	*<1%*	*89%*
Heriot-Watt	263	23	19	13	0	1	3	44	1	
	72%	*6%*	*5%*	*4%*	*0%*	*<1%*	*1%*	*12%*	*<1%*	*83%*
Napier	316	5	24	0	18	4	0	44	7	
	76%	*1%*	*6%*	*0%*	*4%*	*1%*	*0%*	*11%*	*2%*	*87%*
QMUC	85	3	6	2	1	1	1	12	1	
	76%	*3%*	*5%*	*2%*	*1%*	*1%*	*1%*	*11%*	*1%*	*85%*
Stirling	85	30	2	0	0	0	1	10	0	
	66%	*23%*	*2%*	*0%*	*0%*	*0%*	*1%*	*8%*	*0%*	*92%*
Total	1074	97	72	19	19	8	5	148	10	
	74%	*7%*	*5%*	*1%*	*1%*	*<1%*	*<1%*	*10%*	*<1%*	*87%*

Table 4.5.2 shows a more detailed breakdown of those successfully completing first year. Napier University has the largest number of students receiving an award after first year. These are students that have received recognition for the exams they have passed but have no desire to continue with their degree. There is a framework for qualifications of HEIs in Scotland to ensure that student academic achievement is recognised. After first year students can be awarded a Certificate in Higher Education, after second year a Diploma, after third year a Degree and after fourth year an Honours Degree.

4.6 A Comparison of Summer School and Non Summer School Students After First Year

Attendance at summer school does not guarantee successful completion of first year at University but it does appear that, generally, there is a greater percentage of success from summer school students. Table 4.6 presents the differences in first year percentage success.

Table 4.6.1 A Comparison of Summer School and Non Summer School Students After First Year

		First Year Success	
		Summer School	Non Summer School
	Overall	89%	86%
Gender	Female	93%	88%
	Male	83%	84%
Destinations	Edinburgh	93%	88%
	Heriot-Watt	86%	84%
	Napier	90%	86%
	QMUC	89%	85%
	Stirling	91%	91%
S5 Qualifications	QS5 1	85%	83%
	QS5 2	91%	86%
	QS5 3	92%	85%
	QS5 4	94%	91%
Combined Qualifications	QCOM 1	86%	81%
	QCOM 2	87%	88%
	QCOM 3	93%	84%
	QCOM 4	97%	91%
School Group	1	88%	88%
	2	91%	90%
	3	90%	83%
Ethnicity	White	88%	86%
	Non White	96%	87%
Council	City of Edinburgh	90%	87%
	East Lothian	89%	86%
	Midlothian	89%	82%
	West Lothian	87%	87%
Subject Areas	Med, Den, Vet A	100%	95%
	Allied to Med B	100%	88%
	Bio, Phys Sciences C	87%	81%
	Maths, Comp E	76%	77%
	Eng, Tech F	89%	82%
	Social, Law, Acc H	87%	88%
	Bus, Hosp, Tour I	88%	89%
	Lang, Eng, Hist, Geog J	97%	91%
	Creative Arts K	100%	88%
	Education L	94%	95%

In particular:

- Female students are more successful than male students and they have an increased chance if they have completed summer school
- All institutions benefit from an increased success rate from summer school students except the University of Stirling which remains the same
- Students from LEAPS group 3 schools seem to benefit the most from the experience of summer school
- Both first generation and non white students who attended summer school show increased levels of first year success
- Students from Midlothian and the City of Edinburgh, in particular, show increased levels of success over non summer school students from the same council areas.

Some of the numbers studying several subject areas are quite small and it would be unwise to read too much into them but of the subject areas studied by larger numbers of students (C,F, H, I, J) the success levels are similar or increased for summer school students.

Chapter Five Modelling the First Year Data

The aim of this analysis is to increase understanding of the characteristics associated with the probability that LEAPS students will successfully complete first year at university; and how these characteristics combine to determine a probable outcome.

A number of different models are presented in this chapter:

- Firstly, a model based on all the first year data is developed and aims to give LEAPS useful information relating to the impact of the LEAPS summer school and school group classification. Furthermore the information is used to predict how well selected students perform by destination.
- Secondly, models by gender are generated to see if the predictors of success are the same for females and males.
- Finally automated selection procedures are exercised.

A variety of characteristics or variables are examined and the extent to which, for example, qualifications, school and family background influence the success probability. The main focus is on first year success probabilities for two reasons. Firstly, the greatest number of those who do not succeed (drop outs) occur by the end of first year. Secondly, there is a greater amount of data available for first year students, both in this study and in the wider context. The overall success rate for first year LEAPS students is 87%.

5.1 The Modelling Approach

The statistical approach combines the data from the six HEIs over the four years of data with the aim of fitting a suitable statistical model to the data. The use of the binary logistic regression model gives the probability of successfully completing first year at university. The binary dependent variable is defined as taking the value 1 if the student successfully completes a particular year and 0 if the student fails to complete. The independent variables included in the model and their associated coding are listed in Appendix (iv). The analysis allows the identification of the effect of any one of the independent variables on the outcome of success or not. The use of

several different forms of the models is investigated before selecting a final model which best fits the data, both in terms of fit and in terms of predicted success rate.

5.2 SAS

The statistical software used is SAS. SAS offers four procedures capable of performing logistic regression, PROC LOGISTIC, PROC CATMOD, PROC GENMOD and PROBIT. PROC LOGISTIC is the most widely used and is used here. It has the advantage over the other procedures in that it contains additional output options like odds ratios, goodness of fit indices and ROC curves. The SAS code used is contained in Appendix (v).

5.3 Model Building Strategy

The model building process follows the strategy outlined in Chapter Three. Prior to developing the LEAPS model a number of different reference models were created to aid variable selection and assessment.

5.3.1 A Full Model

This model, which includes all the independent variables, is for reference and can be used, with the Likelihood ratio test, to confirm that the removal of certain variables does not affect the model's predictive power.

5.3.2 A Model for Each of the Variables

These models proved useful to confirm the significance of each of the variables and also to highlight any variable(s) that exhibited quasi complete separation problems.

5.4 Developing the LEAPS Model

5.4.1 Identifying Initial Variables

A chi square test checking for association between successful completion of first year and each of the variables revealed the following results:

Variable	**Chi-Square p-value**
Combined Qualifications (QCOM)	0.0172
S5 Qualification (QS5)	0.0023
FIRST_GENERATION	0.9525 *
ETHNICITY	0.0998
SUBJECT_GROUPING_1	0.0022
DESTINATION	0.1014
SUMMER_SCHOOL	0.1084
SEX	0.0007
AGE_GROUP	0.3294 *
Q_CODE	0.7790 *
COUNCIL	0.3373 *
SCHOOL_GROUP	0.2318
ELIGIBILITY	0.0401

The variables marked with a * are dropped as they do not meet the requirement of chi-square<0.25.

5.4.2 Evaluating and Refitting

The wisdom of the elimination of the variables deemed insignificant can be confirmed by running a model with all the variables and comparing, using the likelihood ratio, with a model minus the four * variables.

Model	*Likelihood Ratio*	**Degrees of freedom (df)**
Model with all 13 variables	93.0450	47
Model with 9 variables (- four *)	94.1779	37
Absolute Difference	1.1329	10

Chi-square statistic (tabular) for 10 df is 18.307 which is greater than 1.1329 therefore the variables in question can be dropped from the model as it will make no difference to the prediction.

Continuing this process with a view to parsimony the next least significant variable from the 9 variable model is removed - ELIGIBILITY

Model	*Likelihood Ratio*	**df**
Model with 9 variables	94.1779	37
Model with 8 variables (- ELIGIBILITY)	82.0135	29
Difference	12.1644	8

Chi-square statistic for 8 df = 15.507, > 12.1644, so ELIGIBILITY variable is dropped.

The next least significant variables at this stage are SEX and QCOM. These are kept in the model through choice.

The next least significant variable from the 8 variable model is considered for removal – SCHOOL_GROUP

Model	*Likelihood Ratio*	**df**
Model with 8 variables	82.0135	29
Model with 7 variables (- SCHOOL_GROUP)	77.6065	27
Difference	4.407	2

Chi-square statistic for 2 df = 5.991, > 4.407, so SCHOOL_GROUP is dropped.

The next least significant variable from the 8 variable model was considered for removal – ETHNICITY

Model	*Likelihood Ratio*	**df**
Model with 7 variables	77.6065	27
Model with 6 variables (- ETHNICITY)	73.5591	26
Difference	4.0474	1

Chi-square statistic for 1 df = 3.841, < 4.0474, so ETHNICITY kept in model.

5.4.3 Interaction Testing

This model is tested for all interactions. No interactions are significant at the 5% level. QCOM*SUMMER_SCHOOL is significant at the 10% level with a p value of 0.0847. The same check is made using the likelihood ratio to see if the interaction aids the prediction.

Model	*Likelihood Ratio*	**df**
Model with 7 variables	77.6065	27
Model with 7 variables and 1 interaction	86.2685	33
Absolute Difference	8.662	6

Chi-square statistic for 6 df = 12.592, > 8.662, so the interaction is dropped.

5.4.4 The Final LEAPS Model

QCOM and SEX are not significant but are retained out of interest. As qualifications remain the main selector for places at University, and because there are four qualification groups QCOM is retained. SEX is kept in the model as it is now widely recognised that females have been outperforming males at school for at least the last decade. It is advantageous to see whether there is any evidence of this enhancement in relation to LEAPS students at university. All other variables in the model are highly significant so the final model contains seven independent variables relating to qualifications both combined score at end of sixth year and score at end of fifth year, ethnicity, gender, summer school attendance (and completion), destination and subject grouping. This discerning approach to variable selection improves the predictive power as well as providing a better understanding of the underlying concepts of the data.

From the SAS output

Effect	DF	Wald Chi-Square	Pr > ChiSq
QCOM	3	5.3046	0.1508
QS5	3	10.2980	0.0162
DESTINATION	5	17.2612	0.0040
SS	2	8.1374	0.0171
SUBJECT_GROUPING_1	12	22.2950	0.0343
ETHNICITY	1	3.4601	0.0629
SEX	1	1.5280	0.2164

5.4.5 Assessing the Fit of the Final Model

Before using the results of the final model with any confidence the goodness of fit of the model and its predictive power should be assessed through objective measures.

5.4.5.1.1 Hosmer and Lemeshow Goodness of Fit Test

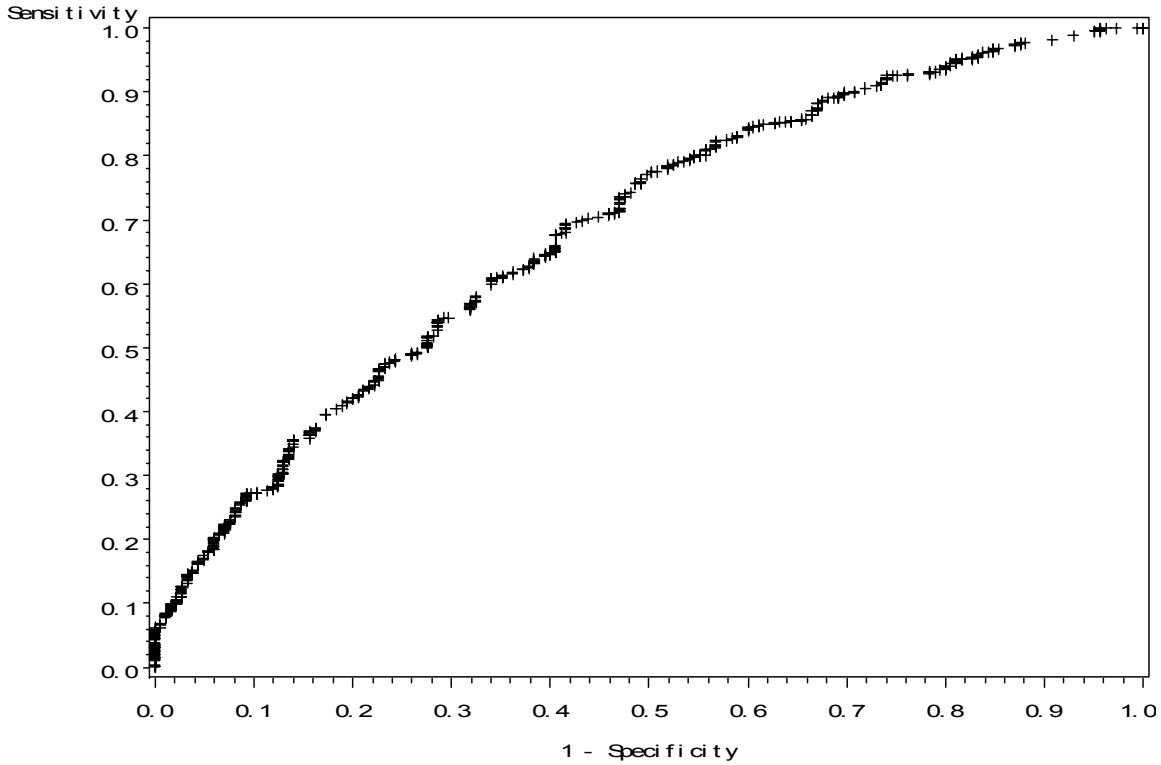
		Progression = 1		Progression = 0	
Group	Total	Observed	Expected	Observed	Expected
1	142	97	97.93	45	44.07
2	140	106	109.39	34	30.61
3	143	123	118.24	20	24.76
4	143	128	122.58	15	20.42
5	143	126	125.46	17	17.54
6	144	127	129.16	17	14.84
7	143	130	130.54	13	12.46
8	148	133	137.48	15	10.52
9	141	132	133.41	9	7.59
10	132	130	127.78	2	4.22

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
7.2195	8	**0.5131**

This chi-square statistic has the desirable outcome of non significance.

5.4.5.2 Area Under the ROC Curve



Sensitivity, on the x axis, is the true positive rate and 1-Specificity, on the y axis, is the false positive rate. The area under the curve is a measure of test accuracy. Recall that the area under the ROC curve ranges from 0 to 1. Anything over 0.5 implies a diagnostic accuracy greater than chance. The closer the curve follows the left hand border and top border of the diagram, the more accurate the test. The closer the curve is to the 45° diagonal, the less accurate the test. This curve with its area of 0.69 just reaches the level of acceptable discrimination as outlined by Hosmer and Lemeshow (2000). See Chapter Three.

5.4.6 Other Measures

The max-rescaled R-square value is 0.0983 and AIC is 1084.538. The AIC is slightly higher than the model including all the variables and has resulted in the less common situation of a higher value for the simpler and more desirable model.

5.4.7 Interpreting the Coefficients

The regression parameter estimates corresponding to the significant independent variables and the independent variables chosen to be kept in the model are presented in Table 5.4.7.1.

Table 5.4.7.1 SAS Output – Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	2.1408	1.0752	3.9641	0.0465
QCOM	2	1	0.2417	0.2508	0.9286	0.3352
QCOM	3	1	0.3337	0.3038	1.2067	0.2720
QCOM	4	1	0.7957	0.3580	4.9405	0.0262
QS5	2	1	0.4903	0.2534	3.7448	0.0530
QS5	3	1	0.3473	0.2851	1.4839	0.2232
QS5	4	1	1.0210	0.3493	8.5418	0.0035
DESTINATION	HW	1	0.3124	0.2558	1.4918	0.2219
DESTINATION	Napier	1	0.9477	0.2973	10.1614	0.0014
DESTINATION	QMUC	1	0.2012	0.3961	0.2581	0.6114
DESTINATION	Stirling	1	1.3228	0.4147	10.1727	0.0014
DESTINATION	ECA	1	1.0103	1.1451	0.7784	0.3776
SS	WSS	1	-0.7027	0.3892	3.2592	0.0710
SS	no	1	-0.6719	0.2385	7.9345	0.0049
SUBJECT_GROUPING_1	B	1	-0.4554	1.1076	0.1691	0.6809
SUBJECT_GROUPING_1	C	1	-1.1796	1.0567	1.2460	0.2643
SUBJECT_GROUPING_1	D	1	10.5231	703.3	0.0002	0.9881
SUBJECT_GROUPING_1	E	1	-1.5579	1.0639	2.1443	0.1431
SUBJECT_GROUPING_1	F	1	-1.1319	1.0767	1.1052	0.2931
SUBJECT_GROUPING_1	G	1	-0.7301	1.1511	0.4023	0.5259
SUBJECT_GROUPING_1	H	1	-0.8943	1.0552	0.7183	0.3967
SUBJECT_GROUPING_1	I	1	-0.6458	1.0707	0.3638	0.5464
SUBJECT_GROUPING_1	J	1	-0.4753	1.0935	0.1889	0.6638
SUBJECT_GROUPING_1	K	1	-0.7826	1.1182	0.4898	0.4840
SUBJECT_GROUPING_1	L	1	0.7645	1.2006	0.4054	0.5243
SUBJECT_GROUPING_1	M	1	-0.4291	1.3012	0.1087	0.7416
ETHNICITY	2	1	0.7273	0.3910	3.4601	0.0629
SEX	M	1	-0.2353	0.1904	1.5280	0.2164

These are not as easy to interpret as the odds ratios but they are used to give predictions relating to students with particular characteristics.

5.4.8 Interpreting the Odds Ratios

The odds ratio estimates are easier to interpret than the parameter estimates. The odds ratio is a measure of association which reflects how much more (or less) likely it is for the outcome to occur with or without the variables at different levels. The table below shows the SAS output for the odds ratio estimates.

Table 5.4.8.1 SAS Output - Odds Ratio Estimates

Effect			Point Estimate	95% Wald Confidence Limits	
QCOM	2 vs 1		1.273	0.779	2.082
QCOM	3 vs 1		1.396	0.770	2.532
QCOM	4 vs 1		2.216	1.099	4.470
QS5	2 vs 1		1.633	0.994	2.683
QS5	3 vs 1		1.415	0.809	2.474
QS5	4 vs 1		2.776	1.400	5.505
DESTINATION	HW	vs Edinburgh	1.367	0.828	2.256
DESTINATION	Napier	vs Edinburgh	2.580	1.441	4.620
DESTINATION	QMUC	vs Edinburgh	1.223	0.563	2.658
DESTINATION	Stirling	vs Edinburgh	3.754	1.665	8.463
DESTINATION	ZECA	vs Edinburgh	2.746	0.291	25.914
SS	WSS vs SS		0.495	0.231	1.062
SS	no vs SS		0.511	0.320	0.815
SUBJECT_GROUPING_1	B vs A		0.634	0.072	5.559
SUBJECT_GROUPING_1	C vs A		0.307	0.039	2.439
SUBJECT_GROUPING_1	D vs A		>999.999	<0.001	>999.999
SUBJECT_GROUPING_1	E vs A		0.211	0.026	1.694
SUBJECT_GROUPING_1	F vs A		0.322	0.039	2.660
SUBJECT_GROUPING_1	G vs A		0.482	0.050	4.600
SUBJECT_GROUPING_1	H vs A		0.409	0.052	3.235
SUBJECT_GROUPING_1	I vs A		0.524	0.064	4.275
SUBJECT_GROUPING_1	J vs A		0.622	0.073	5.301
SUBJECT_GROUPING_1	K vs A		0.457	0.051	4.092
SUBJECT_GROUPING_1	L vs A		2.148	0.204	22.590
SUBJECT_GROUPING_1	M vs A		0.651	0.051	8.341
ETHNICITY	2 vs 1		2.070	0.962	4.453
SEX	M vs F		0.790	0.544	1.148

If the confidence limits include the value of 1, a change in the dependent variable is not associated in the change in odds of the dependent variable. The variable in question is not considered a useful predictor.

5.4.8.1 Key Points from the Odds Ratios

All else being equal:

- Students in QCOM group 4 (the most highly qualified) are more than twice as likely to successfully complete first year than those in QCOM 1
- Students in QS5 group 4 (the most highly qualified) nearly three times more likely to successfully complete first year than those in QS5 1
- LEAPS students studying at Napier University are 2.7 times as likely to successfully complete first year as LEAPS students at the University of Edinburgh
- LEAPS students at the University of Stirling are around four times more likely to successfully complete first year than LEAPS students at the University of Edinburgh

- Students attending (and completing) summer school are almost twice as likely to successfully complete first year as those who did not attend summer school
- Non white students are twice as likely to successfully complete first year than white students.

None of the SUBJECT_GROUPING_1 variables individually seem particularly strong predictors. This is more than likely a result of the small numbers associated with particular subject areas. Subject grouping E (Mathematics, Computer and Information Science) is the strongest predictor. Students studying this subject grouping seem to do less well than any other group. This is and has been the case for the general student population.

5.4.9 Calculating the Probabilities and Making Predictions

Based on the univariate statistics it is possible to identify a 'typical' and an 'atypical' LEAPS student for each of the destinations for the variables in the model. Typical can be defined as the type of student most common to the institution. Atypical can be defined as the type of student less common to the institution. Using the characteristics of these students and the associated parameter estimates the probability of success at the end of first year can be calculated.

University of Edinburgh

The typical LEAPS University of Edinburgh student is in the top qualification group both after S5 *(QS5 4*) and S6 *(QCOM 4)*, *white*, studying *Social Studies, Law and Accountancy*, *female* and did not attend *summer school*.

The regression equation is:

$$\ln\left(\frac{p}{1-p}\right) = 2.1408(\text{intercept}) + 0.7957(\text{QCOM}) + 1.0210(\text{QS5}) + 0(\text{DESTINATION}) - 0.6719(\text{SUMMER\_SCHOOL}) - 0.0843(\text{SUBJECT\_GROUPING\_1}) + 0(\text{ETHNICITY}) + 0(\text{SEX})$$

$$= 3.2013$$

probability of success = $\frac{e^{3.2013}}{(1+e^{3.2013})}$ = 0.96

A 96% chance of successfully completing first year.

The atypical LEAPS University of Edinburgh student is in the bottom qualification group both after S5 *(QS5 1*) and S6 *(QCOM 1), non white*, studying *Mathematical, Computer and Information Science, male* and completed *summer school*.

The regression equation is:

$$\ln\left(\frac{p}{1-p}\right) = 2.1408(\text{intercept}) + 0(\text{QCOM}) + 0(\text{QS5}) + 0(\text{DESTINATION}) - 0(\text{SUMMER\_SCHOOL}) - 1.5579(\text{SUBJECT\_GROUPING\_1}) + 0.7273(\text{ETHNICITY}) - 0.2353(\text{SEX})$$

$$= 1.0749$$

probability of success = $\frac{e^{1.0749}}{(1+e^{1.0749})}$ = 0.75

A 75% chance of successfully completing first year.

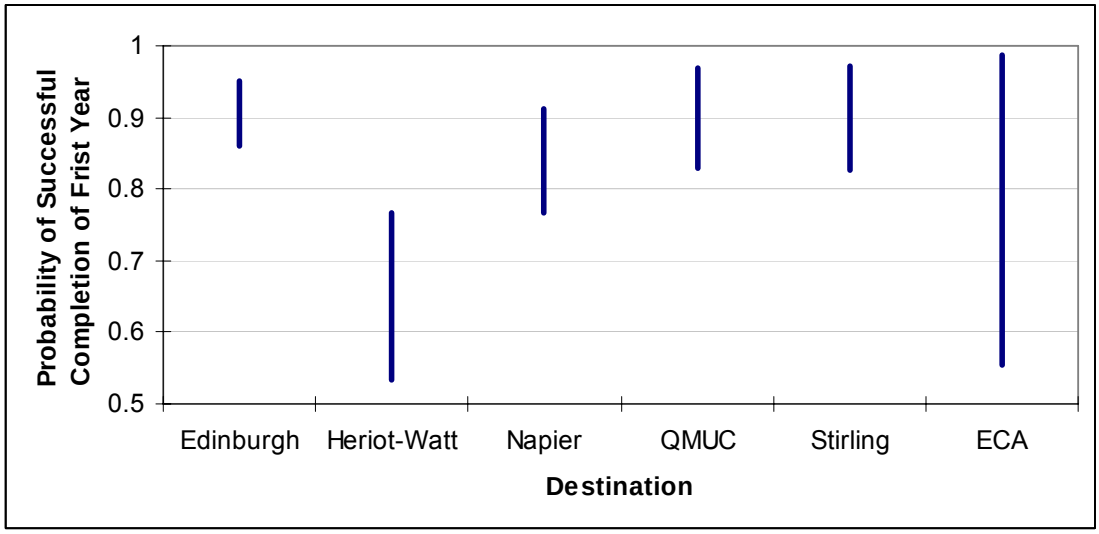
These calculations represent the first two entries in Table 5.4.9. Further probabilities were calculated in the same way for the two student types for each of the institutions. The parameter estimates can be used to calculate probabilities of success for a student with any combination of the predictor variables. The figures in brackets for the typical students are the 95% confidence limits. The combinations shown are purely to give an indication of the probabilities.

Table 5.4.9.1 Probabilities of Successful Completion of First Year for Typical and Atypical Students

	Student Characteristics						
Destination	**Combined Qualifications**	**S5 Qualifications**	**Summer School**	**Subject Grouping**	**Ethnicity**	**Sex**	**Probability of Success**
Edinburgh (typical)	top	top	no	Social Studies, Law, Accounts	White	Female	0.916 (0.861-0.951)
Edinburgh (atypical)	bottom	bottom	yes	Maths, Comp, Info Science	Non white	Male	0.746
Heriot-Watt (typical)	second top	second top	no	Maths, Comp, Info Science	White	Male	0.661 (0.534-0.768)
Heriot-Watt (atypical)	bottom	bottom	yes	Architecture, Building	Non white	Female	0.921
Napier (typical)	bottom	bottom	no	Business, Admin, Hosp, Tour	White	Female	0.854 (0.767-0.913)
Napier (atypical)	top	top	yes	Lang, Eng, Hist, Geog, Div, Jour	Non white	Male	0.988
QMUC (typical)	top	second bottom	no	Subjects Allied to Medicine	White	Female	0.924 (0.829-0.968)
QMUC (atypical)	top	top	yes	Social Studies, Law, Accounts	Non white	Male	0.972
Stirling (typical)	second bottom	bottom	no	Lang, Eng, Hist, Geog, Div, Jour	White	Female	0.928 (0.826-0.972)
Stirling (atypical)	top	top	yes	Creative Arts	Non white	Male	0.975
ECA (typical)	second bottom	second top	n/a	n/a	White	Female	0.908 (0.554-0.987)
ECA (atypical)	bottom	bottom	n/a	n/a	Non white	Male	0.933

Edinburgh College of Art students would not normally attend the LEAPS summer school as there is no art course offered. Also, all the students study the same subject grouping (creative arts) so any prediction is based only on the student qualifications, school group, ethnicity and gender.

Graph 5.4.9.1 A Comparison of the Confidence Limits for Typical Students



Graph 5.4.9.1 shows that the University of Edinburgh has the narrowest range of probabilities and ECA the widest. The wide confidence limit in this case is as a result of the small sample size. There are marked areas of overlap between Napier University, QMUC and the University of Stirling.

5.4.10 Probabilities of Successful Completion of First Year for Typical Students by Qualifications and Destination

The typical Napier University student is poorly qualified relative to students that attend the other HEIs in this study. It is of interest to establish how successful the typical Napier University student would be if they attended, for example, Heriot-Watt University. All other things being equal the probability of the typically qualified Napier student succeeding at Heriot-Watt is 0.757 and a typically qualified University of Edinburgh student has the probability of 0.976 of succeeding at the University of Stirling. Table 5.4.10 illustrates how the typically a qualified student from each of the institutions would perform in another. These students did not attend summer school but if they had then the probabilities of success would be even higher.

Table 5.4.10.1 Probabilities of Successful Completion of First Year for Typical Students by Qualifications and Destination

Typical Qualification	Destination: Edinburgh	Heriot-Watt	Napier	QMUC	Stirling
Edinburgh	0.916	0.937	0.966	0.930	0.976
Heriot-Watt	0.588	0.661	0.787	0.634	0.843
Napier	0.695	0.757	0.854	0.736	0.895
QMUC	0.859	0.893	0.940	0.924	0.958
Stirling	0.693	0.756	0.854	0.734	0.928

ECA has not been included in this table as the institution specialises in one subject area only.

5.5 Models Generated through Automatic Selection Procedures

Forward selection, backward elimination and stepwise regression are used to produce models for comparison with the final LEAPS one. Each procedure produced

5.5.1 Backward Elimination Using the LEAPS Data

Initially all the variables are included in the model and then each of the variables are removed in turn. The final model contains five predictor variables QS5, DESTINATION, SUMMER_SCHOOL, ETHNICITY and SUBJECT_GROUPING_1. The Hosmer and Lemeshow statistic (0.1935) implies a fairly good fit and the area under ROC curve figure of 0.681 implies acceptable discrimination ability.

- The best performing students would be in the highest qualification group at the end of S5, non white, having attended summer school and be studying Education at the University of Stirling.
- The worst performing students are in the lowest qualification group after S5, withdrew from summer school and are studying Mathematics, Computer and Information Science at QMUC.

Neither of these scenarios is possible with the courses available in the different HEIs.

5.5.2 Forward Selection Using the LEAPS Data

The forward selection procedure produces a model with six predictor variables – QS5, DESTINATION, SUMMER_SCHOOL, ETHNICITY, SUBJECT_GROUPING_1 and SEX. The Hosmer and Lemeshow Statistic implies a good fit (0.6323) and the area under the ROC curve figure of 0.683 implies acceptable discrimination ability. With this model the conclusions are the same as for the backward elimination model except that females are more successful than males.

5.5.3 Stepwise Regression Using the LEAPS Data

The stepwise regression yields a model with three predictor variables – QS5, ETHNICITY and SUBJECT_GROUPING_1. The Hosmer and Lemeshow statistic implies poor fit with a significant value of 0.0375. The area under the ROC curve figure of 0.646 implies acceptable discrimination ability.

5.5.4 Best Subsets

The best subsets procedure cannot be carried out for variables that have more than one level.

5.6 Modelling by Gender

It is a worthwhile exercise to ascertain whether there are differences in the predictors for males and females. Indeed there are.

5.6.1 LEAPS Males

The strongest predictors for the probability of males successfully completing first year are QCOM and DESTINATION. A model containing these variables is indicative of a good fit with a Hosmer and Lemeshow goodness of fit statistic of 0.8360. The area under the ROC curve is 0.65 which implies a discriminatory ability of just about acceptable.

From the odds ratios it can be inferred that

- Males from the highest combined qualification group are more than five times more likely to successfully complete first year than those from the lowest combined qualification group

- Males attending the University of Stirling or Napier University are the most successful

The SAS output for this model is contained in Appendix (vi).

5.6.2 LEAPS Females

The strongest predictors for the probability of females successfully completing first year are SUMMER_SCHOOL, ELIGIBILITY and QS5. A model containing these variables is indicative of a good fit with a Hosmer and Lemeshow goodness of fit statistic of 0.9943. The area under the ROC curve is 0.661 which implies a discriminatory ability of just about acceptable.

From the odds ratios it appears that

- Females who completed LEAPS summer school are nearly three times more likely to successfully complete first year than those who did not attend summer school
- The more eligibility criteria that apply to females the less likely they are to succeed.

The SAS output for this model is contained in Appendix (vii).

Chapter Six Modelling Subsequent Years

The percentage success through subsequent years is very high, 93% for second year, 94% for third year and 99% for fourth year. The sample numbers are lower (minus those who failed to complete the previous year) and the predictors are likely to be weaker. Nevertheless it is a worthwhile exercise in case a variable proves highly significant and results in a strong predictor. The same strategy and steps as outlined in chapters 3 and 5 are followed.

6.1 Second Year

A model can be run using the second year data. SEX and SCHOOL_GROUP are the best predictors showing enough significance to be kept in a model. For ease of comparison with the first year data where DESTINATION was significant, it is kept in this model.

- The best performing students are female, studying at QMUC or ECA and from a LEAPS group 3 school. The associated probability of success =>0.96
- The worst performing students are male, studying at Edinburgh and from a LEAPS group 2 school. The associated probability of success for this combination is 0.83.

The Hosmer and Lemeshow statistic (0.3926) implies a good fit but the area under ROC curve figure of 0.632 implies slightly less than acceptable discrimination ability.

In a model with SEX as the only predictor females are 1.6 times more likely to successfully complete second year than males.

6.2 Third Year

The model run using the third year data shows QS5, DESTINATION and SEX as the strongest predictors and includes for interest QCOM.

- The best performing students are female, in the highest qualification group by the end of S6 (QCOM 4) but were in the second lowest qualification group in S5 (QS5 2) and studying at Napier. The associated probability of success is 0.998

- The worst performing students are male, in the lowest qualification group both at the end of S5 (QS5 1) and at the end of S6 (QCOM 1) and studying at QMUC. The associated probability of success is 0.58

The Hosmer and Lemeshow statistic (0.8374) implies a good fit and the area under the ROC curve figure of 0.739 is within the acceptable discrimination band.

6.3 Fourth Year

99% of students in this year successfully complete it. It is not possible to run a model using the fourth year data as ten of the 13 independent variables exhibited quasi-complete separation. Reliable Maximum Likelihood estimates cannot be obtained as the data did not converge. The statistical package SAS prints a warning message when this occurs but does create a model. The interpretation of the model is at best limited and at worst misleading.

There are other more powerful statistical techniques such as discriminant analysis or artificial neural networks that can be used in such a situation.

Chapter Seven Conclusions, Recommendations and Critical Appraisal

7.1 Conclusions

This research set out to identify whether LEAPS students perform differently from traditional students at University. The evidence suggests that LEAPS students perform at least as well as more traditional students. 87% of LEAPS students successfully complete first year and, as expected, the greatest attrition is at the end of first year. The levels of success for LEAPS students are higher than the corresponding Higher Education Statistics Agency (HESA) figures for young students from low participation neighbourhoods (for Scotland) which ranged from 81% to 84% for the academic years in the study. LEAPS students fare better and LEAPS must take the credit. The extensive programme of events available, particularly to group 1 schools and the LEAPS summer school, equip the students for university both in making the right course choices and in the confidence to succeed.

The research also set out to try and identify the characteristics associated with successful completion of an academic year and to create a statistical model to predict the probabilities of success based on different mixes of the student characteristics. A number of characteristics have been identified as having predictive power. The probability of success in first year is influenced significantly by school qualifications but also the subject chosen, completion of the LEAPS summer school, destination, gender and ethnic background.

Qualifications, as expected, prove one of the strongest predictors but interestingly in this study S5 qualifications have more predictive power than the combined qualifications at the end of S6.

The subject studied undoubtedly has an influence on success but in this study it is a better indicator of lack of success as the strongest indicator comes from those studying Mathematics, Computer and Information Science where more fail than in any other subject.

LEAPS firmly believes that summer school is a very worthwhile exercise. It takes a great deal of time and effort from a number of parties to ensure its smooth running. Encouragingly students who complete summer school are more than twice as likely to successfully complete first year at university as non summer school students.

The variables that emerge from this study as good predictors largely agree with previous studies.

The institution in which LEAPS students are most successful is the University of Stirling. There is also evidence to suggest that LEAPS students do well at Napier University, despite being the least qualified group. This may be due to 'fitting in', first discussed by Tinto (1974). All students need to feel comfortable in their environment and it would appear that Napier University offers an environment more conducive to LEAPS student success.

Females are performing better at university than males. To many this will not be a surprise as females have been out performing males in school and university for the last decade. Woodfield and Saunders (1998) found that women work consistently harder than men throughout their degrees. Consider here Heriot-Watt University – the only university to have a male bias and the greatest number of students studying the highest drop out area of Mathematics, Computer and Information Science. Yet they still achieve an overall first year success level of 83%. They must be doing something right!

First generation is not significant. Previous research (Sinclair and McClements (2003)) has shown that students who are not first generation are more likely to succeed than those that are. In this study there appears no difference in the success of students whose parents have attended university and those that have not.

Although not a strong predictor there is evidence to suggest that the younger students perform better. It may be that the two age groups here are too close and valid age comparisons should be between students studying straight from school and

another level of maturity. Certainly though being under 18 at the start of their university career does not seem to disadvantage LEAPS students.

Table 4.5.1 shows that students who do complete first year then stand an even higher chance of successfully completing their degree. Johnes (1990) also found first year results to be a much stronger predictor of performance.

7.2 Recommendations

The following recommendations are personal to the author rather than LEAPS policy and many apply not just to LEAPS students but to the student body as a whole.

There is a need for HEIs to adopt the same terminology and aim for cross board agreement in how student progression should be coded.

The HEIs should have systems in place to track different cohorts of students and continuously monitor them throughout their time at university. A versatile system could track all students who do not follow the normal route to achieving their degree. The true causes of success (or failure) of wider access students cannot be ascertained without the required information.

LEAPS summer school works. It is already oversubscribed and LEAPS staff have to make difficult decisions regarding which students can and cannot be accommodated. A predictive model would make an extremely useful tool to help the LEAPS staff allocate summer school places. Furthermore the possibility of increasing the size of summer school should be investigated.

This study has shown that qualifications are a strong predictor of success, particularly for males, so any relaxing of entry requirements by the HEIs should be accompanied by appropriate measures such as focussed student support.

Students that are relatively poorly qualified can perform well. HEIs need to continue to look beyond qualifications as a sole measure of capability. LEAPS does not

adhere to the idea of a blanket fashion relaxing of entrance requirements other than for those schools and/or individuals whose potential has been identified.

There is a requisite to stop the negativity associated with dropping out. Some students leave university because it is simply not for them regardless of qualifications, social or cultural factors. It is still better to have tried and failed than not to have tried at all. Can there be an acceptable drop out rate?

LEAPS has an abundance of information about LEAPS eligible students. Respecting data protection this information can still be used to identify high risk students and help achieve a balance between the costs of support strategies and the costs of the students who do not proceed. Measures should be directed at first year students.

This study will provide a bench mark against which results for future years cohorts can be evaluated. The study should be repeated as often as is fundable but certainly every four to five years. This would validate the results.

7.3 Critical Appraisal

It is widely accepted that the factors relating to how well a student performs can be placed in three broad categories. They relate to academic qualifications, socio economic and personal factors, and involvement and integration with the institution. The LEAPS data adequately cover the first two but it is much harder to assess the third.

There will always be difficulty in explaining why students do or do not complete their time at university because there are so many variables which play a part. The information that LEAPS has on its database relating to the sample of students combined with the information supplied from the HEIs provides a valuable source of information to investigate. One of the strengths of this study is the quality of the data.

The sample of LEAPS students is representative of the student body as a whole in terms of age, gender and spread of qualifications but is, as expected, biased in terms of LEAPS eligibility criteria and summer school attendance. There is additional

information available about summer school students relating to parental profession but this was not included as only 20% of the students attended summer school. A broad indication regarding the profession of parents can be inferred from whether or not the student is first generation though this proved not to be a significant factor.

The school qualifications of the students are split into two groups, those attained in S5 and those attained after both S5 and S6. There is scope here, with the information available, to analyse the qualifications in a number of different ways. It would be interesting to investigate whether there are any school subjects and grades that could be used as an indicator of university success. Or to compare the students who have a similar UCAS point score but through a narrow range of high qualifications like Advanced Highers versus the students who have a broader range of lower qualifications like Intermediate 2 and Highers.

It is not possible to achieve a clear picture of the effect of having a number of the LEAPS special eligibility characteristics as (thankfully) there are few students that have multiple eligibility.

There are alternative statistical techniques that can be used to assess the LEAPS data. Logistic regression is the most widely used and it is popular because the assumptions associated with it are less restrictive. Logistic regression does not assume a linear relationship between the dependent and independent variables, nor does the dependent variable need to be normally distributed. However the outcome in logistic regression must be discrete and logistic regression does encounter limitations with small sample sizes. As mentioned there are small numbers of students associated with the different levels of eligibility criteria and with particular subject areas, for example agriculture and architecture. This can lead to problems with overfitting. Models with overfitting do not replicate well and some of the relationships that appear statistically significant are actually just noise. Too many cells with no responses gives high standard errors and quasi complete separation problems can result in important predictors being removed from the model, both of these problems occur in this study with the subject grouping variable. If the

distributional assumptions can be met then discriminant function analysis may be a more powerful statistical tool. Neural networks are also used in this type of situation.

The Hosmer and Lemeshow goodness of fit statistics for almost all of the models created indicate that there is no significant difference between the observed and the predicted results, a good model fit in other words. The poorest fitting model is the one created using the automated procedure stepwise regression. The discriminatory ability of the models determined using the ROC curve is less favourable. The best discriminatory ability resulted from the model using the third year data where success rates are very high. Generally the discriminatory ability for the models can be classified as just acceptable. So, although the model fit statistics are good, the models do not have much predictive power. This is understandable as it has already been mentioned that there are many characteristics that affect a student's performance at university which are not included in this study.

There are no variables in this study relating to personality or 'fitting in'. The degree to which a student may fit in or become involved in a university, faculty or their course is unknown. Perhaps this information may only be deduced from an exit interview with those students who do not complete. Studies by Tinto (1975) and Forsyth and Furlong (2003) have shown the importance of fitting in and Johnes (1990) found evidence to support the view that successful completion is associated with the degree of student involvement in university life. The difficulty in measuring a variable that indicates an ability to fit in or become involved means it would be challenging to include it in a statistical model.

There is a quandary regarding what to compare this study with. As already mentioned there is little information relating to the tracking of wider access students or students from non traditional backgrounds. HESA produces figures relating to 'non-continuation following year of entry' for students from 'low participation neighbourhoods' but it is not clear how the students which pass but fail to return are categorised. Comparisons with this data should be circumspect. The lack of clear definitions relating both to particular student cohorts and measures of progression make it difficult to state clear cut conclusions.

This research provides a much needed source on how wider access students perform at university. The model is limited to wider access students attending university straight from school – it does not include students applying from further education or mature students. However it does include students from the 46 secondary schools in Edinburgh and the Lothians attending one of the LEAPS partner institutions or the University of Stirling.

References

Allison, P.D. (1999) *Logistic Regression using the SAS system: Theory and Application*. SAS Institute. USA

Arulampalam, W., Naylor, R. and Smith, J. (2003) Factors affecting the probability of first-year medical student dropout in the UK: a logistic analysis for the intake of cohorts of 1980-1992. Warwick Economic Research Papers.

Bannerjee, M. and Wellner, J.A. (2205) Score Statistics for Current Status Data: Comparisons with Likelihood Ratio and Wald Statistics. *International Journal of Biostatistics*, Vol 1, (2005).

Bekhradnia, B. and Thompson, J. (2002) Who does best at University? *The Guardian*, 15th October 2002.

Dallal, G.E (2001) *The Little Handbook of Statistical Practice* http://www.tufts.edu/~gdallal/LHSP.HTM *accessed 29th June 2006.*

Forsyth, A and Furlong, A. (2000) *Socio-economic disadvantage and access to higher education*. Joseph Rowntree Foundation. The Policy Press.

Forsyth, A and Furlong, A. (2003) *Socio-economic disadvantage and experience in higher education*. Joseph Rowntree Foundation. The Policy Press.

Furnival, G.M. and Wilson, R.W. (1974) Regressions by Leaps and Bounds *Technometrics,* Vol. 16, No. 4, pp 499-511.

Garson, D.G. (2006) http://www2.chass.ncsu.edu/garson/pa765/logistic.htm *accessed 29th June 2006.*

Higher Education Funding Council for England (1999) Performance Indicators in Higher Education in the UK. *Report 99/66*. Higher Education Funding Council for England, Bristol http://www.hefce.ac.uk/pubs/hefce/1999/99_66/annexes.htm#a accessed 18th August 2006

Higher Education Statistics Agency, Scottish Funding Council (2002 (onwards)). Performance Indicators. Higher Education Statistics Agency. Cheltenham. http://www.hesa.ac.uk/pi/0203/continuation.htm *accessed 4th August 2006*

Hosmer, D. W. and Lemeshow, S. (2000) *Applied Logistic Regression*. John Wiley and Sons.

House of Commons Select Committee on Education and Employment (2001), *Higher Education: Student Retention Sixth Report.* London: Stationary Office

Iannelli, C. and Paterson, L. (2005) Does Education Promote Social Mobility? Centre for Educational Sociology, University of Edinburgh.

Johnes, J (1990). Determinants of Student Wastage in Higher Education. *Studies in Higher Education*, 15, 87-99.

Johnston, V. (2000) Identifying Students at Risk of Non-progression: The development of a diagnostic test. British Educational Research Association.

Krzanowski, W.J. (1998) *An Introduction to Statistical Modelling*. Arnold.

Laing, C and Robinson, A. (2003) The Withdrawal of Non-traditional Students: developing an explanatory model. *Journal of Further and Higher Education*, Vol 27, No. 2, 2003.

Menard, S. (1995) *Applied Logistic Regression Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-106. Thousand Oaks, CA:Sage.

Musselbrook, K (2003) Monitoring and Tracking Students through Higher Education and Beyond: A review of Management Information Systems in Higher Education Institutions (HEIs). SESWARF

Scottish Executive (2003) *Life Through Learning; Learning Through Life* www.scotland.gov.uk/library5/lifelong/llsm-00.asp *accessed 29th June 2006.*

Scottish Executive (2003) *A framework for higher education in Scotland: Higher Education Review* www.scotland.gov.uk/library5/education *accessed 29th June 2006.*

Scottish Higher Education Funding Council (2005) Learning for All. The Report of the SFEFC/SHEFC Widening Participation Review Group. SHEFC, Edinburgh.

Sinclair, H. and McClements, P. (2001) An evaluation of the progression of LEAPS students through their first year of higher education. Research Paper. SESWARF. www.snap.ac.uk/seswarf/leapsreport.pdf *accessed 31st March 2006*

Smith, J.P. and Naylor, R.A. (2000) Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society*, Series A, Statistics in Society, Volume 164, Part 2, 2001.

Tinklin, T. (2000) The influence of Social Background on Application and Entry to Higher Education in Scotland: a multilevel analysis. *Higher Education Quarterly*, Volume 54, No. 4, pp 343-385.

Tinto, V. (1975) Dropout from higher education: a theoretical synthesis of recent research. Rev. Educ. Res., 45, 89-125.

Tinto, V. (1987) *Leaving College: Rethinking the Causes of Student Attrition*. Chicago: University of Chicago Press

Universities UK (2005) From the margins to the Mainstream. Embedding widening participation in Education. Universities UK, London.

Woodfield, R. and Saunders, P. (1998) The Hidden Factor Fuelling Female Success. University of Sussex Media Release http://www.sussex.ac.uk/press_office/media/media37.html *accessed 4th August 2006*

Woodley, A. Thompson, M and Cowan, J. (1992) Factors Affecting Non-Completion Rates in Scottish Universities. Student Research Centre, Institute of Educational Technology, Open University.

Yorke, M. (1999*) Leaving early: undergraduate non-completion in higher education*. London: Falmer.

Appendix (i) Ethnic Groups

Code	Description
11, 12, 13	White – British
19	White – Other
21	Black or Black British – Caribbean
22	Black or Black British – African
29	Other Black background
31	Asian or Asian British – Indian
32	Asian or Asian British – Pakistani
33	Asian or Asian British – Bangladeshi
34	Chinese or other ethnic background
39	Other Asian background
41	Mixed – White and Black Caribbean
42	Mixed – White and Black African
43	Mixed – White and Asian
49	Other Mixed background
80	Other ethnic background
90	Not known
98	Information refused

Appendix (ii) HEFCE Subject Groupings

Code	Subject Groupings
A	Medicine, Dentistry and Veterinary Science
B	Subjects allied to Medicine
C	Biological Sciences and Physical Sciences including Sports Science
D	Agriculture and related subjects
E	Mathematical, Computer & Information Science
F	Engineering and Technology
G	Architecture, Building and Planning
H	Social studies, Law and Accountancy
I	Business and Administrative Studies, Hospitality and Tourism
J	Languages, English, History, Geography, Divinity and Journalism
K	Creative Arts
L	Education
M	Combined Studies

Appendix (iii) UCAS Points Tariff and Examples of UCAS Higher Grade Scoring

Score	Advanced Higher	Higher	Intermediate 2	Standard Grade Credit
120	A			
100	B			
80	C			
72		A		
60		B		
48		C		
42			A	
38				Band 1
35			B	
28			C	Band 2

Grades	Score
C	48
B	60
A	72
CC	96
BB	120
AA	144
CCC	144
BBB	180
CCCC	192
AAA	216
BBBB	240
CCCCC	240
AAAA	288
BBBBB	300
AAAAA	360

Appendix (iv) Independent Variables Included in the Modelling

QCOM - Combined Qualifications achieved by the end of S6 (adjusted for double counting) and split into four groups corresponding to the quartiles.

QS5 – S5 qualifications split into four groups corresponding to the quartiles.

FIRST_GENERATION – whether the student is first generation from their family to enter higher education.

ETHNICITY – ethnic background of student.

SUBJECT_GROUPING_1 - the subject area of the degree chosen by the student in first year.

DESTINATION – the destination higher education institution.

SUMMER_SCHOOL – whether or not the student attended and successfully completed LEAPS summer school.

SEX – gender of student.

AGE_GROUP – age of student on the 1st October of the academic year in question, split into two groups of over 18 and under 18.

Q_CODE – Scottish Index of Multiple Deprivation quintile for the student post code.

COUNCIL – Council area relating to the location of students school.

SCHOOL_GROUP - LEAPS School grouping.

ELIGIBILTY – number of eligibility criteria associated with each student.

Appendix (v) SAS Code for Logistic Regression

```
libname restotal 'J:\results';
/*Sorting of datasets required before merging*/
proc sort data=restotal.bigtotal;
by PID;
proc sort data=restotal.dblqualsall;
by PID;
run;
/*dataset containing student information merged with data set
containing qualification information*/
data restotal.thebigone;
merge restotal.bigtotal restotal.dblqualsall;
by PID;
run;
data logregprog;
set restotal.thebigone (KEEP = Destination SS PID Sex DOB Age_Grp
Q_code SRN Council SG _st_Gen Ethnicity
       eth2 Prog_after_yr1 Prog_after_yr2 Prog_after_yr3
Prog_after_yr4 s5fintot s6fintot
     s5s6com Eligfin subject_grouping_1 subject_grouping_2
subject_grouping_3 subject_grouping_4);
run;
data logregprog;
set logregprog;
/*ECA recoded so that Edinburgh becomes reference category*/
if destination = 'ECA' then Destination = "ZECA";
/* subject area recoding
if subject_grouping_1 = 'G' then subject_grouping_1 = 'K';
if subject_grouping_1 = 'D' then subject_grouping_1 = 'C';
run;*/
/*check that variables are read in correctly*/
/*proc contents data = logregprog position;
run;*/
/*setting into qualification groups*/
data logregprog;
set logregprog;
/*s5 and s6 combined scores taking into account double counting,
groups approximately equal to quartiles*/
qcom=0;
if s5s6com<226 then qcom=1;
else if 227<s5s6com<=305 then qcom=2;
else if 306<s5s6com<=383 then qcom=3;
else qcom=4;
/*s5 only scores, groups approximately equal to quartiles*/
qs5=0;
if s5fintot<126 then qs5=1;
else if 127<s5fintot<=197 then qs5=2;
else if 198<s5fintot<=263 then qs5=3;
else qs5=4;
run;
proc freq data=logregprog;
tables (Destination ss sex age_grp council _st_gen eth2)*qcom;
tables (Destination ss sex age_grp council _st_gen eth2)*qs5;
run;
```

```
/* setting up variable names to be used in proc logistic*/
data logregprog;
set logregprog;
/*destinations*/
Edinburgh=0; HeriotWatt=0; Napier=0; QMUC=0; Stirling=0;ECA=0;
if Destination ='Edinburgh' then Edinburgh=1;
if Destination ='HW' then HeriotWatt=1;
if Destination ='Napier' then Napier=1;
if Destination ='QMUC' then QMUC=1;
if Destination ='Stirling' then Stirling=1;
if Destination = 'ECA' then ECA=1;
/*summer school*/
SSno=0;SSwd=0;SSyes=0;
if SS ='no' then SSno=1;
if SS =' ' then SSno=1;
if SS ='WSS' then SSwd=1;
if SS ='SS' then SSyes=1;
/*gender*/
male=0; female=0;
if sex='M' then male=1;
if sex='F' then female=1;
/*age group*/
under18=0; over18=0;
if age_grp='1' then over18=1;
if age_grp='2' then under18=1;
/*council*/
Edin=0; West=0; Mid=0; East=0; Fife=0;
if council='City of Edinburgh' then Edin=1;
if council='West Lothian' then West=1;
if council='Midlothian' then Mid=1;
if council='East Lothian' then East=1;
if council='Fif' then Fife=1;
/*school group*/
SG1=0; SG2=0; SG3=0;
if SG = '1' then SG1=1;
if SG = '2' then SG2=1;
if SG = '3' then SG3=1;
/*first generation*/
FGyes=0; FGno=0; FGdk=0;
if _st_Gen ='yes' then FGyes=1;
if _st_Gen ='no' then FGno=1;
if _st_Gen =' ' then FGdk=1;
/*No of eligibility criteria*/
el0=0;el1=0;el2=0;el3=0;el4=0;
if Eligfin='0' then el0=1;
if Eligfin='1' then el1=1;
if Eligfin='2' then el2=1;
if Eligfin='3' then el3=1;
if Eligfin='4' then el4=1;
if Eligfin='5' then el4=1;
if Eligfin='6' then el4=1;
/*qualification grouping*/
qcom1=0;qcom2=0;qcom3=0;qcom4=0;
if qcom='1' then qcom1=1;
if qcom='2' then qcom2=1;
```

```
if qcom='3' then qcom3=1;
if qcom='4' then qcom4=1;
qs51=0;qs52=0;qs53=0;qs54=0;
if qs5='1' then qs51=1;
if qs5='2' then qs52=1;
if qs5='3' then qs53=1;
if qs5='4' then qs54=1;
/*ethnicity*/
white=0;nonwhite=0;dontknow=0;
if eth2='1' then white=1;
if eth2='2' then nonwhite=1;
if eth2='3' then delete;
/*subject codes*/
Medicine=0; Allied=0;Science=0;Maths=0;Engineering=0;
Architecture=0;Law=0;Business=0;Humanities=0;
CreativeArt=0;Education=0;
Combined=0;
if subject_grouping_1='A' then Medicine=1;
if subject_grouping_1='B' then Allied=1;
if subject_grouping_1='C' then Science=1;
if subject_grouping_1='E' then Maths=1;
if subject_grouping_1='F' then Engineering=1;
if subject_grouping_1='H' then Law=1;
if subject_grouping_1='I' then Business=1;
if subject_grouping_1='J' then Humanities=1;
if subject_grouping_1='K' then CreativeArt=1;
if subject_grouping_1='L' then Education=1;
if subject_grouping_1='M' then Combined=1;
/*Q code - SIMD deprivation quintiles*/
Qcode1=0;Qcode2=0;Qcode3=0;Qcode4=0;Qcode5=0;
if Q_code='1' then Qcode1=1;
if Q_code='2' then Qcode2=1;
if Q_code='3' then Qcode3=1;
if Q_code='4' then Qcode4=1;
if Q_code='5' then Qcode5=1;
/*separate the group into two groups - successfully completed yr1 and
those that didn't. Success is a 1 and failure is a 0, unless use
desc*/
if prog_after_yr1 in (1,2,3,5) then Progression=1;
if prog_after_yr1 in (4,6,7,8,9) then Progression=0;
run;
proc freq data = logregprog;
table progression/nocol nopercent;
run;
proc freq data = logregprog;
table qcom*progression/nocol nopercent;
run;
proc freq data = logregprog;
table destination*progression/nocol nopercent;
run;
proc freq data = logregprog;
table (Destination ss sex age_grp council _st_gen eth2 qcom
qs5)*eligfin;
run;
proc freq data = logregprog;
```

```
table qs5*subject_grouping_1/chisq;
table qcom*subject_grouping_1/chisq;
run;
proc freq data = logregprog;
table Destination*subject_grouping_1;
run;
proc freq data = logregprog;
table eth2*qs5;
table eth2*qcom;
run;
proc freq data = logregprog;
tables qcom*(ss sex sg Age_Grp Q_code council subject_grouping_1
_st_Gen eth2)/nocol nopercent;
tables qcom*Destination;
tables qs5*Destination;
tables ss*(q_code);
run;
proc freq data = logregprog;
tables qs5*(ss sex sg Age_Grp Q_code council subject_grouping_1
_st_Gen eth2)/nocol nopercent;
run;
/*Which variables are significant - a first indication?*/
proc freq data = logregprog;
table progression*qcom/chisq;
table progression*qs5/chisq;
table progression*_st_Gen/chisq;
table progression*eth2/chisq;
table progression*subject_grouping_1/chisq;
table progression*(Destination SS Sex Age_Grp Q_code Council SG
Eligfin)/chisq;
run;
/*set up logistic regression for single factors at a time only to see
if any variable
is significant on its own. Also highlights separation problems*/
proc logistic data=logregprog descending;
class qcom/param=ref ref=first;
model progression=qcom/nodummyprint rsq ctable lackfit pprob=.5
outroc=roc;;
run;
proc logistic data=logregprog descending;
class qs5/param=ref ref=first;
model progression=qs5;
run;
proc logistic data=logregprog descending;
class destination/param=ref ref=first;
model progression=destination;
run;
proc logistic data=logregprog descending;
class ss/param=ref ref=last;
model progression=ss;
run;
proc logistic data=logregprog descending;
class sex/param=ref ref=first;
model progression=sex;
run;
```

```
proc logistic data=logregprog descending;
class age_grp/param=ref ref=first;
model progression=age_grp;
run;
proc logistic data=logregprog descending;
class council/param=ref ref=first;
model progression=council;
run;
proc logistic data=logregprog descending;
class SG/param=ref ref=first;
model progression=SG;
run;
proc logistic data=logregprog descending;
class _st_gen/param=ref ref=first;
model progression=_st_gen;
run;
proc logistic data=logregprog descending;
class eligfin/param=ref ref=first;
model progression=eligfin;
run;
proc logistic data=logregprog descending;
class eth2/param=ref ref=first;
model progression=eth2;
run;
proc logistic data=logregprog descending;
class subject_grouping_1/param=ref ref=first;
model progression=subject_grouping_1;
run;
proc logistic data=logregprog descending;
class q_code/param=ref ref=first;
model progression=q_code;
run;
/*all together for information only*/
proc logistic data=logregprog descending;
class qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code/param=ref ref=first;
model progression=qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code/nodummyprint rsq
lackfit outroc=roc;
run;
proc gplot data=roc;
plot _sensit_*_1mspec_;
run;
quit;
/* This is the first model attempt-all minus first gen, age, q code
and council*/
proc logistic data=logregprog descending;
class qcom qs5 destination SS sex SG
eligfin eth2 subject_grouping_1/param=ref ref=first;
model progression=qcom qs5 destination SS sex SG
eligfin eth2 subject_grouping_1/nodummyprint rsq lackfit outroc=roc;
run;
proc gplot data=roc;
plot _sensit_*_1mspec_;
run;
```

```
/*After much trial and error removing eligfin*/
proc logistic data=logregprog descending;
class qcom qs5 Destination SS SG sex eth2
subject_grouping_1/param=ref ref=first;
model progression=qcom qs5 Destination SS SG sex
eth2 subject_grouping_1/nodummyprint rsq lackfit outroc=roc;
run;
/*After much trial and error, removing school group*/
proc logistic data=logregprog descending;
class qcom qs5 Destination SS Sex eth2 subject_grouping_1/param=ref
ref=first;
model progression=qcom qs5 Destination SS Sex eth2
subject_grouping_1/nodummyprint rsq lackfit outroc=roc;
run;
/*Keep school group but remove eth2*/
proc logistic data=logregprog descending;
class qcom qs5 Destination SS Sex SG subject_grouping_1/param=ref
ref=first;
model progression=qcom qs5 Destination SS Sex SG
subject_grouping_1/nodummyprint rsq lackfit outroc=roc aggregate
scale=none;
run;
proc gplot data=roc;
plot _sensit_*_1mspec_;
run;
quit;

/*Replace eth2 and test for interactions*/
proc logistic data=logregprog descending;
class qcom qs5 Destination SS Sex SG eth2
subject_grouping_1/param=ref ref=first;
model progression=qcom qs5 Destination SS Sex SG eth2
subject_grouping_1
qcom*qs5 qcom*destination qcom*ss qcom*sex qcom*ss qcom*eth2
qcom*subject_grouping_1
qs5*destination qs5*SS qs5*Sex qs5*SG qs5*eth2 qs5*subject_grouping_1
destination*ss destination*Sex destination*SG destination*eth2
destination*subject_grouping_1 SS*sex SS*SG SS*eth2
SS*subject_grouping_1
eth2*subject_grouping_1 Sex*SG Sex*eth2 sex*subject_grouping_1/rsq;
run;
quit;
/*FINAL MODEL! tested with and without interaction (qcom*SS)*/
proc logistic data=logregprog descending;
class qcom qs5 Destination SS subject_grouping_1 eth2 SG
sex/param=ref ref=first;
model progression=qcom qs5 Destination SS subject_grouping_1 eth2 SG
Sex/nodummyprint rsq ctable lackfit pprob=.5 outroc=roc;
output out=new p =Predicted_prob_of_response;
run;
proc gplot data=roc;
title 'ROC curve for final model';
plot _sensit_*_1mspec_;
run;
quit;
```

```
/*automatic selection procedures*/
proc logistic data=logregprog descending;
class qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code/param=ref ref=first;
model progression=qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code
/*qcom|qs5|destination|SS|sex|age_grp|
council|SG|_st_gen|eligfin|eth2|subject_grouping_1|q_code @2*//rsq
ctable lackfit pprob=.5 selection=backward;
run;
proc gplot data=roc;
plot _sensit_*_1mspec_;
run;
quit;
proc logistic data=logregprog descending;
class qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code/param=ref ref=first;
model progression=qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code
/*qcom|qs5|destination|SS|sex|age_grp|
council|SG|_st_gen|eligfin|eth2|subject_grouping_1|q_code @2*//rsq
ctable lackfit pprob=.5 selection=forward;
run;
proc gplot data=roc;
plot _sensit_*_1mspec_;
run;
quit;
proc logistic data=logregprog descending;
class qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code/param=ref ref=first;
model progression=qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code
/*qcom|qs5|destination|SS|sex|age_grp|
council|SG|_st_gen|eligfin|eth2|subject_grouping_1|q_code @2*//rsq
ctable lackfit pprob=.5 selection=stepwise;
run;
proc gplot data=roc;
plot _sensit_*_1mspec_;
run;
proc logistic data=logregprog descending;
class qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code/param=ref ref=first;
model progression=qcom qs5 destination SS sex age_grp council SG
_st_gen eligfin eth2 subject_grouping_1 q_code/rsq selection=score
best=2;
run;
quit;
```

Appendix (vi) SAS Output for the Males Model

Effect	DF	Wald Chi-Square	Pr > ChiSq
qcom	3	16.8557	0.0008
Destination	5	11.0697	0.0500

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.4657	0.4011	1.3486	0.2455
qcom	2	1	0.6841	0.3300	4.2961	0.0382
qcom	3	1	0.9991	0.3696	7.3084	0.0069
qcom	4	1	1.6954	0.4181	16.4411	<.0001
Destination	HW	1	-0.0508	0.3092	0.0270	0.8694
Destination	Napier	1	0.7424	0.3807	3.8034	0.0511
Destination	QMUC	1	-0.3219	0.7671	0.1761	0.6748
Destination	Stirling	1	1.5319	0.6836	5.0219	0.0250
Destination	ZECA	1	-0.3064	1.2835	0.0570	0.8113

Odds Ratio Estimates

Effect			Point Estimate	95% Wald Confidence Limits	
qcom	2 vs 1		1.982	1.038	3.785
qcom	3 vs 1		2.716	1.316	5.604
qcom	4 vs 1		5.449	2.401	12.365
Destination	HW	vs Edinburgh	0.950	0.519	1.742
Destination	Napier	vs Edinburgh	2.101	0.996	4.431
Destination	QMUC	vs Edinburgh	0.725	0.161	3.260
Destination	Stirling	vs Edinburgh	4.627	1.212	17.666
Destination	ZECA	vs Edinburgh	0.736	0.059	9.109

Association of Predicted Probabilities and Observed Responses

Percent Concordant	60.2	Somers' D	0.300
Percent Discordant	30.2	Gamma	0.332
Percent Tied	9.5	Tau-a	0.086
Pairs	50082	c	0.650

Partition for the Hosmer and Lemeshow Test

Group	Total	Progression = 1 Observed	Progression = 1 Expected	Progression = 0 Observed	Progression = 0 Expected
1	90	63	64.45	27	25.55
2	26	20	19.74	6	6.26
3	91	70	70.07	21	20.93
4	69	59	55.50	10	13.50
5	37	27	30.16	10	6.84
6	57	50	49.53	7	7.47
7	74	65	65.84	9	8.16
8	81	74	72.63	7	8.37
9	68	63	63.07	5	4.93

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
3.4208	7	0.8435

Appendix (vii) SAS Output for the Females Model

Effect	DF	Wald Chi-Square	Pr > ChiSq
SS	2	7.9873	0.0184
Eligfin	5	15.5304	0.0083
qs5	3	5.7037	0.1269

Analysis of Maximum Likelihood Estimates

Parameter		DF	Standard Estimate	Wald Error	Chi-Square	Pr > ChiSq
Intercept		1	3.1750	0.4738	44.8986	<.0001
SS	WSS	1	-0.5608	0.5877	0.9104	0.3400
SS	no	1	-0.9864	0.3557	7.6913	0.0055
Eligfin	1	1	-0.7563	0.3510	4.6428	0.0312
Eligfin	2	1	-0.3971	0.4016	0.9779	0.3227
Eligfin	3	1	-0.6476	0.4543	2.0317	0.1540
Eligfin	4	1	-1.6685	0.4874	11.7189	0.0006
Eligfin	5	1	-1.4853	0.6025	6.0769	0.0137
qs5	2	1	0.4145	0.3133	1.7503	0.1858
qs5	3	1	0.3241	0.3146	1.0615	0.3029
qs5	4	1	0.7814	0.3320	5.5383	0.0186

Odds Ratio Estimates

Effect		Point Estimate	95% Wald Confidence Limits	
SS	WSS vs SS	0.571	0.180	1.806
SS	no vs SS	0.373	0.186	0.749
Eligfin	1 vs 0	0.469	0.236	0.934
Eligfin	2 vs 0	0.672	0.306	1.477
Eligfin	3 vs 0	0.523	0.215	1.275
Eligfin	4 vs 0	0.189	0.073	0.490
Eligfin	5 vs 0	0.226	0.070	0.738
qs5	2 vs 1	1.514	0.819	2.797
qs5	3 vs 1	1.383	0.746	2.562
qs5	4 vs 1	2.184	1.140	4.188

Association of Predicted Probabilities and Observed Responses

Percent Concordant	63.7	Somers' D	0.322
Percent Discordant	31.4	Gamma	0.339
Percent Tied	4.9	Tau-a	0.062
Pairs	64328	c	0.661

Partition for the Hosmer and Lemeshow Test

Group	Progression = 1 Total	Observed	Expected	Progression = 0 Observed	Expected
1	89	66	67.23	23	21.77
2	93	79	78.92	14	14.08
3	89	79	77.13	10	11.87
4	91	82	81.58	9	9.42
5	85	78	76.63	7	8.37
6	87	79	80.02	8	6.98
7	85	78	79.05	7	5.95
8	73	68	68.79	5	4.21
9	82	78	78.14	4	3.86
10	45	44	43.51	1	1.49

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
1.3976	8	0.9943