A STUDY OF STUDENT ACADEMIC PERFORMANCE AT THE UNIVERSITY OF NATAL

by

Robert Naidoo

Submitted in fulfilment of the requirements
for the degree of Master of Science,
in the
Department of Mathematical Statistics,
University of Natal

Durban 1994

ABSTRACT

In this dissertation a study will be made of university performance in the Science Faculty of the University of Natal, Durban. In particular, we will develop models that can be used to predict the success rate of a student based on his or her matriculation results. These models will prove useful for selecting students to universities. They may also be used to assist sponsors, bursars and donors in allocating funds to deserving students. In addition, these models may be used to identify students who might experience difficulties in their studies at university.

PREFACE

The study described in this dissertation was carried out in the Department of Mathematical Statistics, University of Natal, Durban, during the period January 1993 to December 1994. It was completed under the excellent supervision of Dr M. Murray.

This study represents original work by the author except where use of the work of others has been duly acknowledged in the text and it has not been submitted in any form to another university.

ACKNOWLEDGEMENTS

A special thanks to my beloved mother and to my family for their tremendous support and encouragement.

The author would like to sincerely thank the following for their courteous co-operation in contributing to the success of this dissertation:

- Dr M. Murray for his expert assistance with the computer programs and the writing of this thesis.
- All authors of books and articles, whether mentioned by name or not, whose ideas this thesis contains.
- The financial support of the University of Natal through a Graduate Assistantship, the Foundation of Research and Development the German Academic Exchange Service (DAAD).
- Mrs J. Sylaides for her patience and assistance in typing this thesis.
- Professor L. Troskie for the use of departmental facilities.
- The administrative assistance rendered by Mrs B. Becker, Faculty Officer, Faculty of Science.

CONTENTS

CHAPTER ONE

INTRODUCTION

Due to an increasing number of student applications being made to universities in South Africa, the problem of selecting suitable students has become one of major concern to the university authorities. If it were possible to successfully identify those students who are most likely to succeed then one would be able to assist bursars, sponsors and donors in allocating funds to those "deserving" students. Furthermore, student advisers would be in a better position to offer guidance to students and academic support programmes could be adjusted so as to help these students with their course work.

In the United Kingdom students were, and still are, admitted to university solely on their school leaving results. Although interviews are also conducted by some universities, the most common entry requirement is that students obtain a sufficiently high number of "A"[1] level results for a specific combination of subjects. In the United States of America, university applicants are required to write Standardized Achievement Tests (SATS) in English and Mathematics. A weighted average score for these tests is then used to determine which students should be admitted, with the weights differing from one university to the next.

In South Africa, due to the apartheid policies of the past the problem of student selection has become far more complex. Because the average pupil to teacher ratio, in schools for 1992, for the various population groups has been as follows:

African $40,4:1$; Coloureds $23,7:1$; Asians $21,1:1$ and Whites $18,1:1$,

combined with the fact that a large percentage of teachers in the African schools are without proper professional qualifications (approximately 14% in

[1]School pupils, in the United Kingdom, wrote exams at two levels, namely "O"(lower level) and "A"(higher level).

primary and 8% in secondary schools) have prevented a significant proportion of the Department of Education and Training (DET)[2] students from being able to obtain a matriculation exemption[3] which is the minimum requirement necessary for entry into a South African university. This has resulted in a large number of DET students being deprived of a university education, where possibly with a little bit of extra academic support teaching they might in fact be able to obtain a university degree. This point of view has been strengthened by studies that have been conducted by Potter and Jamotte (1985) and Mitchell and Fridjohn (1987) who found that the DET matriculation performance has not been a reliable indicator of whether a student is likely to succeed at university or not. Mitchell and Fridjohn (1987) have also shown that some matriculation authorities tend to over-rate a student's ability to perform at university. Sochet (1986) found for most matriculation authorities that the success rate at university of students with a matriculation rating that was below a C aggregate was not significantly correlated with their matriculation record while at the top end of the scale a good ma-

[2]DET represents the matriculation examining body of the majority of the "African" students in South Africa.

[3]Matriculants are presently admitted to South African universities on the basis of a matriculation point count (MPC) in which points are awarded for each particular symbol obtained in their six matriculation subjects. The point scales applicable to the University of Natal are as follows

Symbol	H.G.	S.G.
A	8	5
B	7	4
C	6	3
D	5	2
E	4	1
F	1	0
G	0	0

where H.G. respresents Higher Grade and S.G. represents Standard Grade. A MPC score is then obtained for each student by adding together the points awarded for each of the six matriculation subjects. Admission to a particular faculty is then based on the student obtaining a score that lies above a certain cut-off point.

triculation result did tend for most matriculation boards to suggest that a student would be successful at university.

Turning our attention to an examination of the various selection policies that are currently being used at South African universities, a two-tier selection procedure has been employed at the university of the Witwatersrand for accepting disadvantaged students into the Science faculty. In particular students with a sufficiently high set of matriculation results are automatically accepted while those students who are below the minimum entrance requirement are subjected to a battery of tests before gaining access to a four-year (instead of the usual three-year) curriculum. Rutherford and Watson (1990) have shown that the above battery of tests, combined with their matriculation results, do have a predictive ability when it comes to identifying those students who will do well in the four-year curriculum. In fact, it has become evident that the Wits academic support programme (ASP) has had a positive impact on the achievements of these students (Agar, 1992)[4]. At the University of Natal, Durban, a similar academic support programme is provided for helping disadvantaged students. In particular, disadvantaged students wanting to do a Bachelor of Commerce degree are offered an Economics and Management Extended Curriculum (EMEC) programme. This programme essentially spreads the three year degree over four years with the students in their first year being provided with an Educational Development Programme

[4]Other findings of Agar (1992) were that

(i) Non-academic problems tended to influence ASP students' academic progress the most;

(ii) ASP students in residence tended to perform better than ASP students not in residence, provided that these students had few or no financial worries;

(iii) students in residence who had full financial support were more successful at university than those in residence who had limited financial support; and

(iv) ASP students whose home language was English performed better than those whose home language was one of the African languages.

which they take along with three credit bearing courses (instead of the usual five that the other first year students are required to take). Initial indication of the success of this programme has shown that (Flockemann et al. 1993)

(i) EMEC students with a high matriculation point count have tended to produce better results than EMEC students with a lower matriculation point count;

(ii) EMEC students who would not have gained entrance into the faculty based on the current faculty cut-off point of 32 points have produced better results than those students from the African Educational Departments that are not on the EMEC programme; and

(iii) as a group EMEC students have generally obtained pass rates, in the individual credit courses, that are comparable with those students from the Non-African Educational Departments.

In the Science faculty at the University of Natal, Durban, a special four-year curriculum is also offered to "disadvantaged" students. In particular an augmented first year curriculum is given where specific learning skills, language, communication and additional support is offered to students who are then only required to take two (instead of the usual four courses) in each semester. At this point in time we are unable to report on the "success" of this programme.

Turning our attention to an examination of some of the prediction models that have been developed for analysing student performance at universities in South Africa, Van Wyk and Crawford (1984) devised a method for predicting the probability of a student at the University of the Witswatersrand passing a single first year course that is based on the following matriculation system for each matriculation subject, namely

TABLE 1.1

Symbol	H.G.	S.G.
A	8	6
B	7	5
C	6	4
D	5	3
E	4	2
F	3	1

By doubling the number of points awarded for Mathematics and the best of Physical Science, Biology, Geography or Physiology and then adding the points obtained for each of the six matriculation subjects to obtain a total matriculation rating score which we will denote by x, Van Wyk and Crawford (1984) then proceeded to regress this score (for a given sample of students) against the marks obtained for a particular university subject thereby obtaining an estimate for the mean mark for that subject (which we will denote by μ) and for the variance for that subject (which we denote by σ^2). An estimate for the probability that a student with a matriculation rating score x will pass (assuming 48 to be the pass mark) the above university course can then be modelled using a function of the form

$$p(x) = (2\pi\sigma^2)^{-1/2} \int_{48}^{100} \exp\left\{-\frac{1}{2}\left[\frac{(y-\mu)}{\sigma}\right]^2\right\} dy$$

This type of modelling approach will be explored further in chapter three.

Fresen and Fresen (1987) chose to adopt a logistic modelling approach to the above prediction problem. In particular, by associating a pass mark in a particular course with a Bernoulli random variable, a logistic link function of the form

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x$$

was used to link the probability of a student passing this course with their matriculation rating score. This type of modelling approach will be explored further in chapter two.

Turning our attention now to the type of student performance problem that we will be attempting to analyse in this thesis, a study will be made of student performance in the Science faculty of the University of Natal (Durban). In particular the performance of students taking Mathematics 1 (DSMA1SX1), Physics 1 (DSPH1SC1) and then both DSMA1SX1 and DSPH1SC1 will be analysed as a function of the following variables[5]

$$\begin{aligned} X_1 &= \text{English} \\ X_2 &= \text{Afrikaans} \\ X_3 &= \text{Mathematics} \\ X_4 &= \text{Biology} \\ X_5 &= \text{Physics} \end{aligned}$$

Because these results are given in the form of a symbol, and not the actual mark, the following method of coding these symbols was used, namely

Matriculation Symbol (H.G)	Coded Value
A	0
B	1
C	2
D	3
E	4
F	5

Collecting data for the period 1989 to 1992 the following summary statistics were obtained for a student in our sample:

[5]Only Higher Grade (H.G) matriculation subjects were considered as we found that almost all of the the students who attempted Mathematics 1 and Physics 1 at university had attempted matriculation English, Afrikaans, Mathematics, Biology and Physics on the Higher Grade. Note that students in South Africa may attempt matriculation subjects on the Higher Grade, Standard Grade or Lower Grade, however Lower Grade students are usually automatically denied entrance to university.

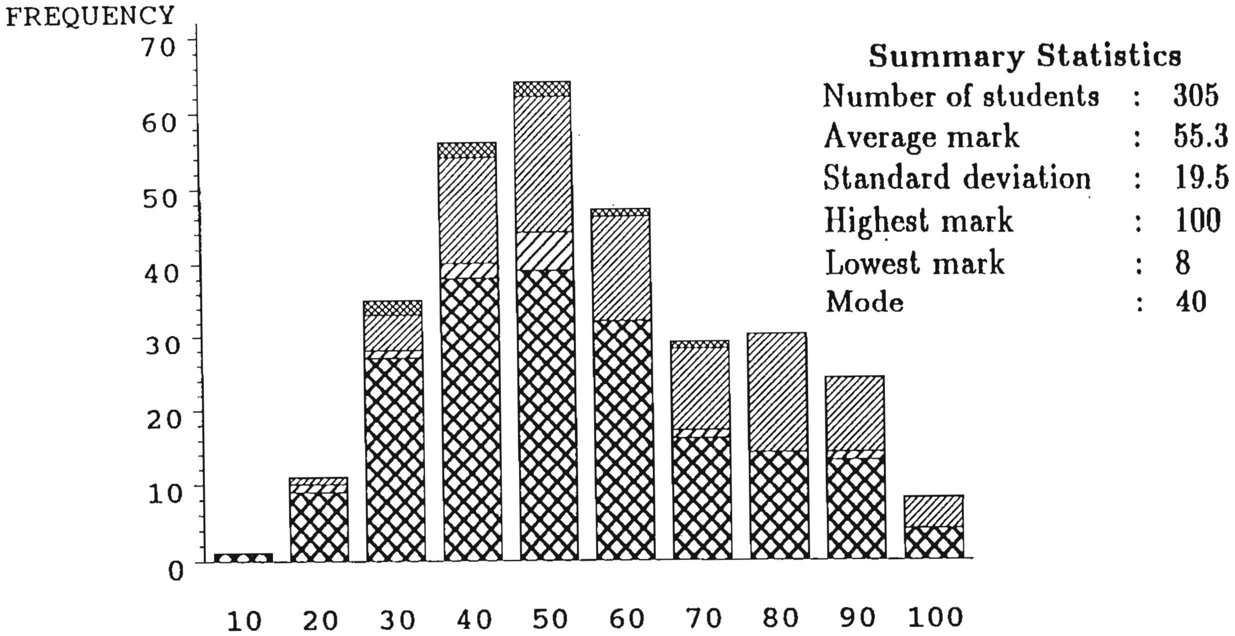


Figure 1.1: Graph representing the DSMA1SX1 result profile

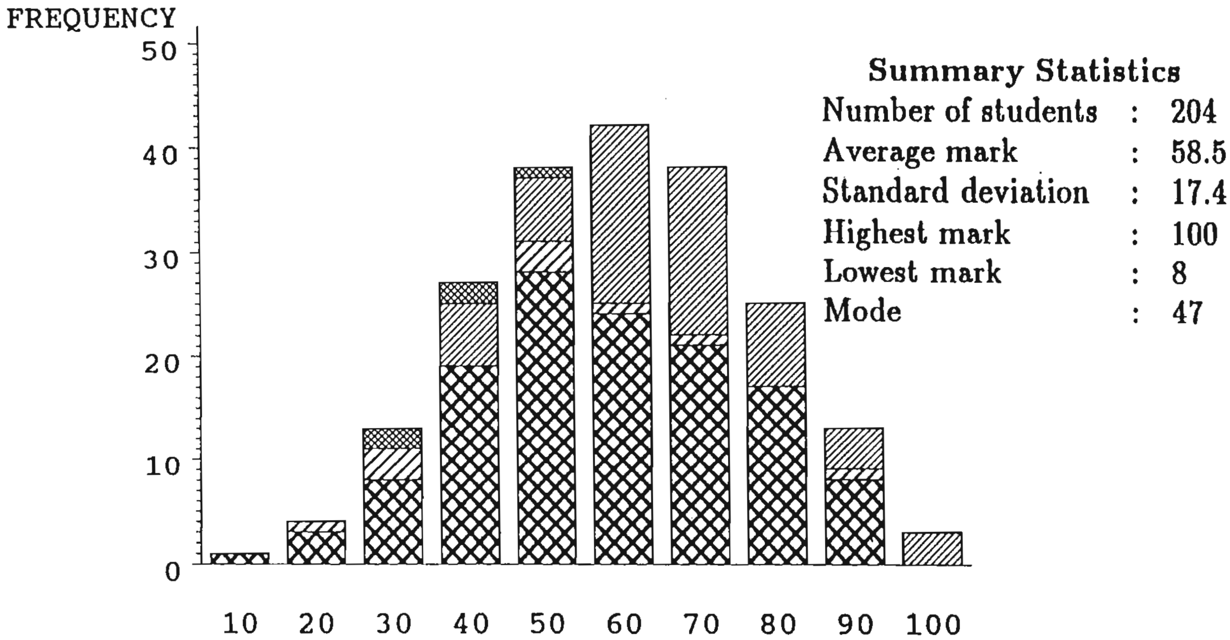
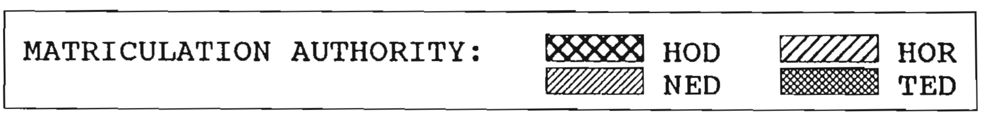


Figure 1.2: Graph representing the DSPH1SC1 result profile



HOD: House of Delegates
HOR: House of Representatives
NED: Natal Education Department
TED: Transvaal Education Department

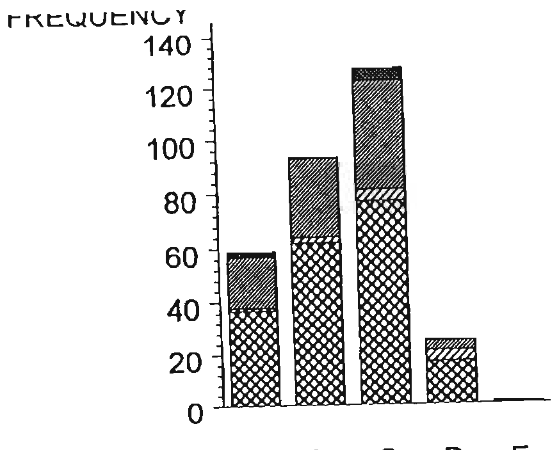


Figure 1.3

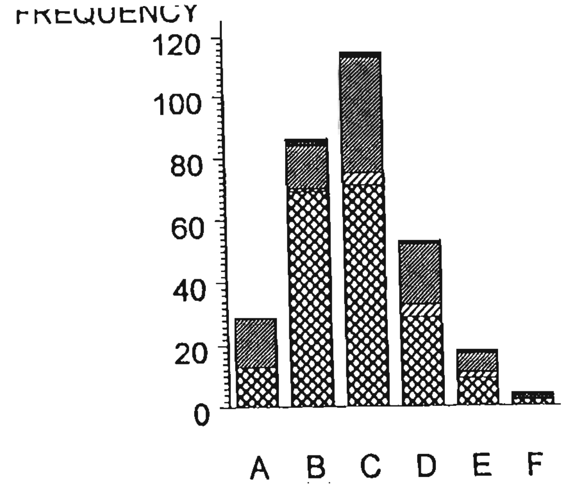


Figure 1.4

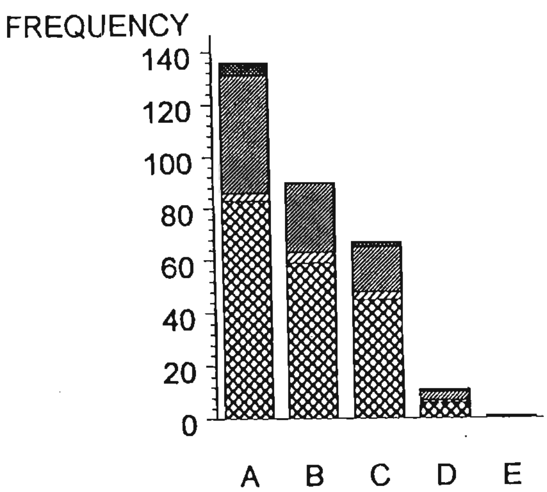


Figure 1.5

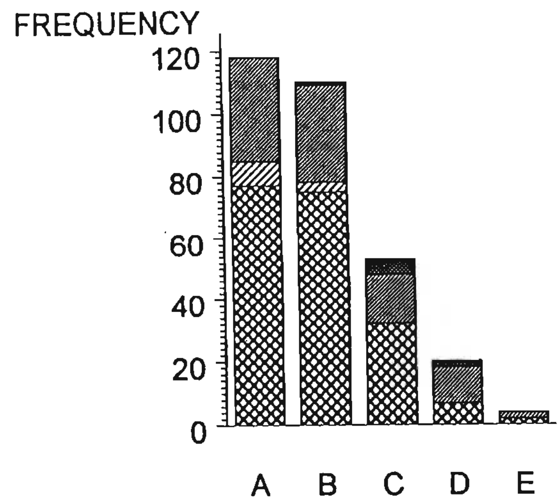


Figure 1.6

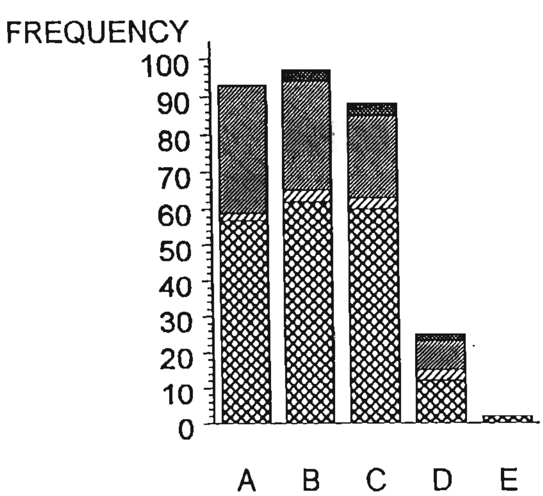
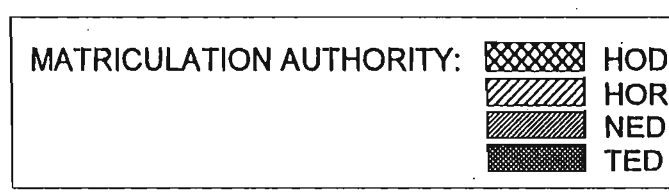


Figure 1.7



Figures 1.3 to 1.7 represent the matriculation result profiles for the matriculation subjects English, Afrikaans, Mathematics, Biology and Physics respectively for the students in our DSMA1SX1 data set.

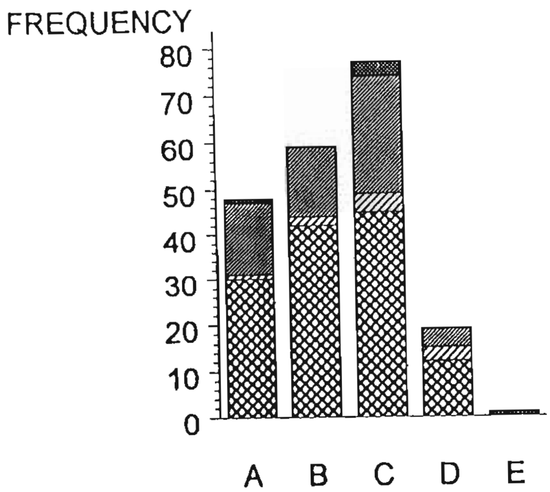


Figure 1.8

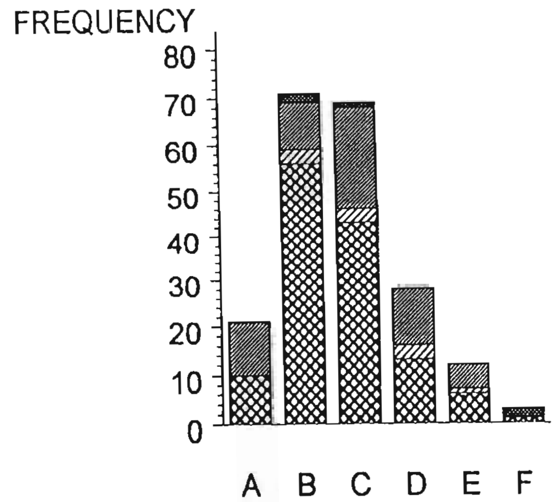


Figure 1.9

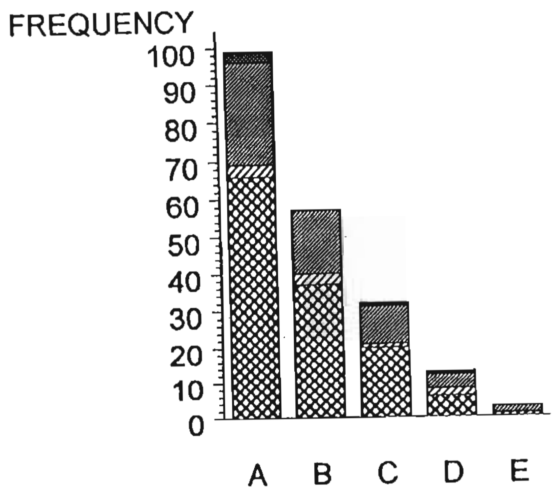


Figure 1.10

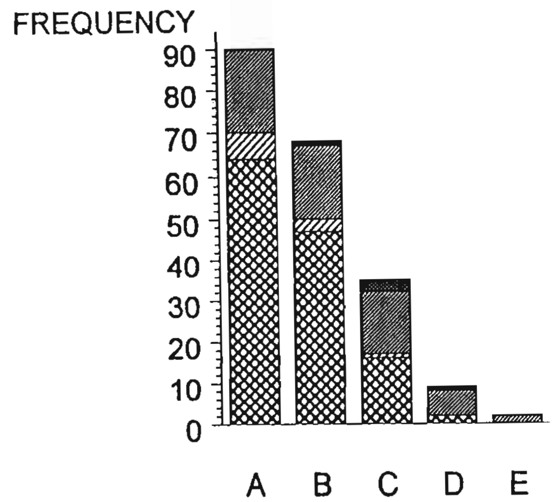


Figure 1.11

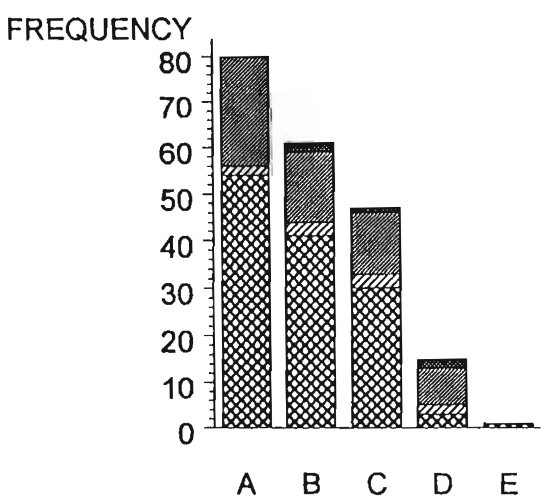
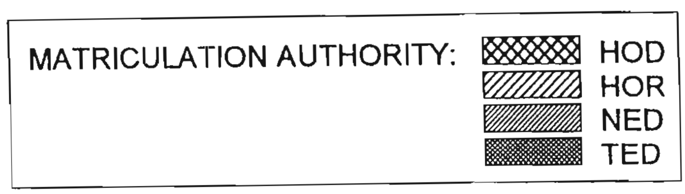


Figure 1.12



Figures 1.8 to 1.12 represent the matriculation result profiles for the matriculation subjects English, Afrikaans, Mathematics, Biology and Physics respectively for students in our DSPH1SC1 data set.

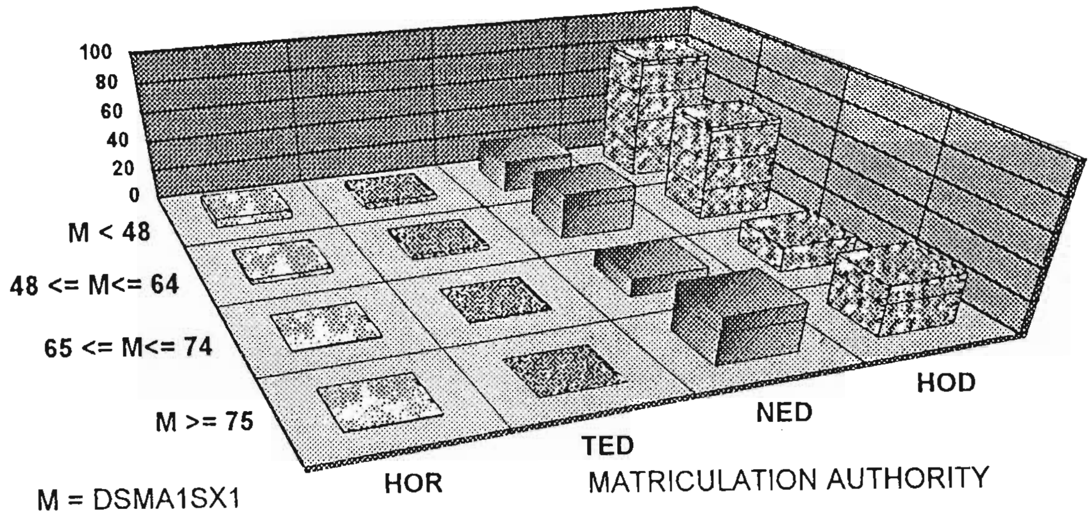


Figure 1.13: Graph illustrating the performance of students from the various matriculation authorities in the DSMA1SX1 course.

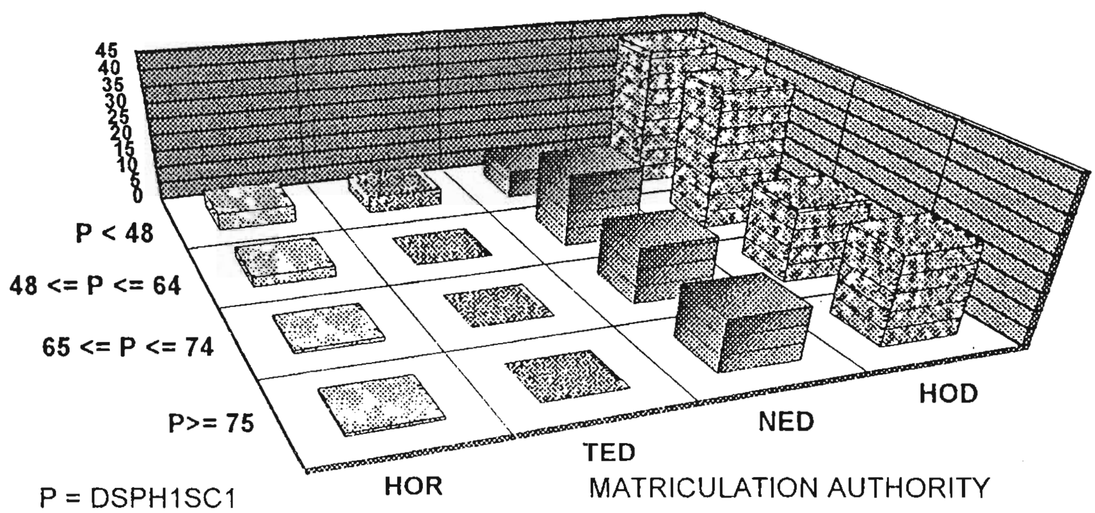


Figure 1.14: Graph illustrating the performance of students from the various matriculation authorities in the DSPH1SC1 course.

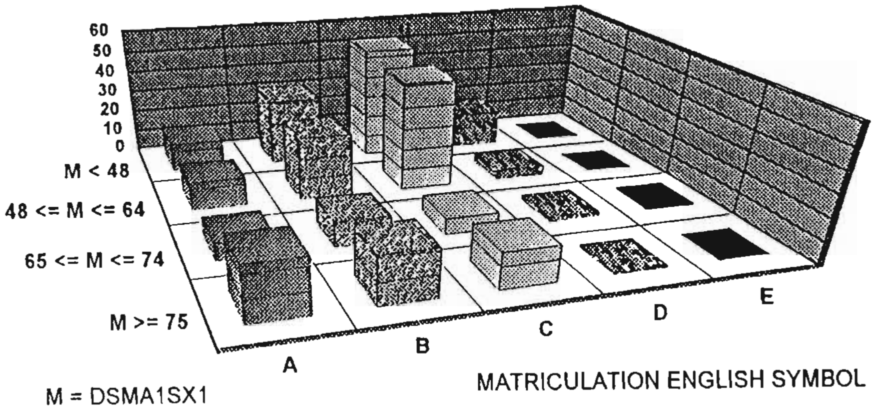


Figure 1.15: Graph illustrating the effect of the matriculation English symbol on the DSMA1SX1 course.

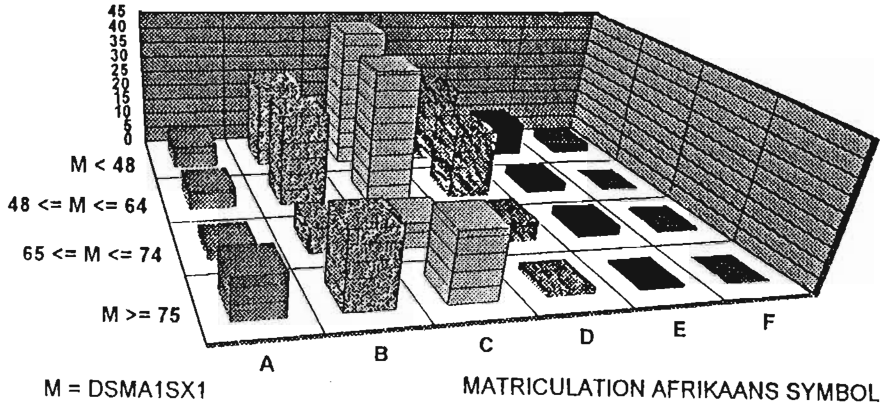


Figure 1.16: Graph illustrating the effect of the matriculation Afrikaans symbol on the DSMA1SX1 course.

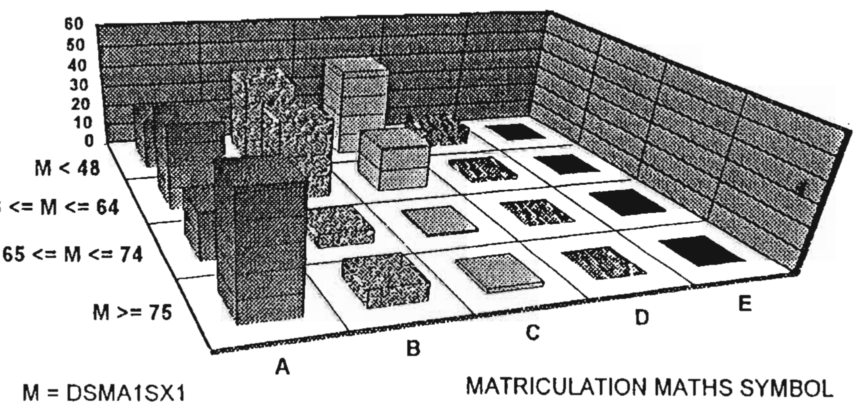


Figure 1.17: Graph illustrating the effect of the matriculation Mathematics symbol on the DSMA1SX1 course.

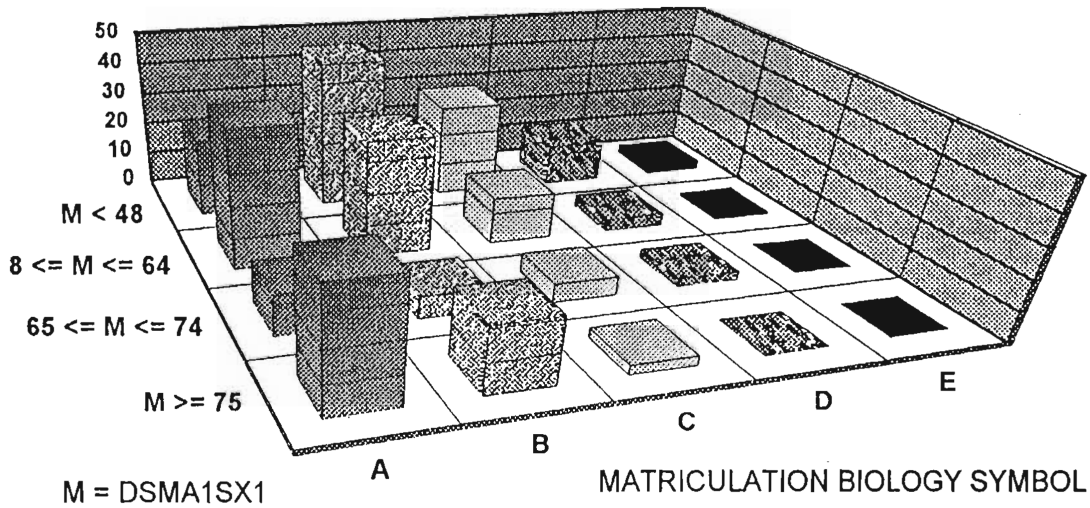


Figure 1.18: **Graph illustrating the effect of the matriculation Biology symbol on the DSMA1SX1 course.**

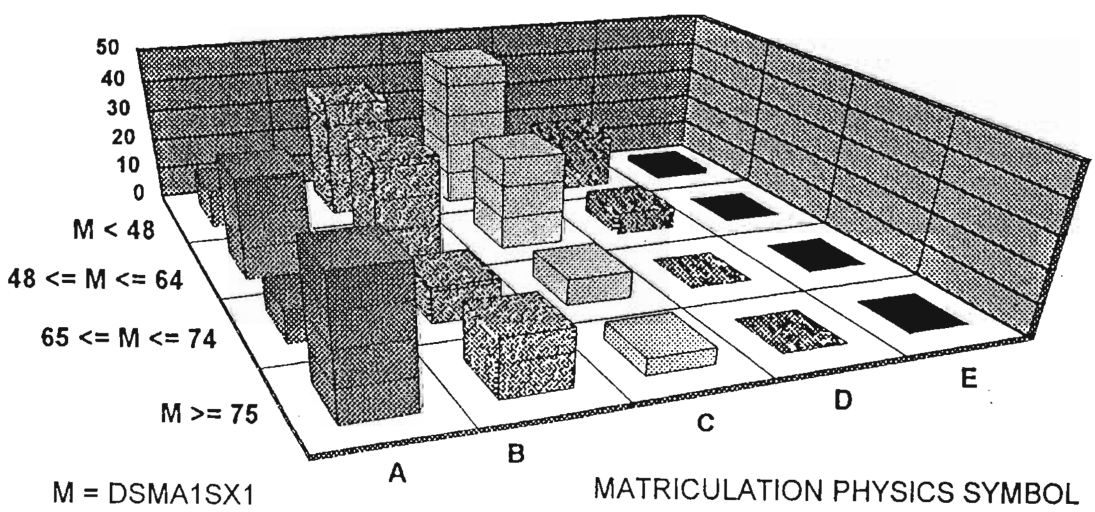


Figure 1.19: **Graph illustrating the effect of the matriculation Physics symbol on the DSMA1SX1 course.**

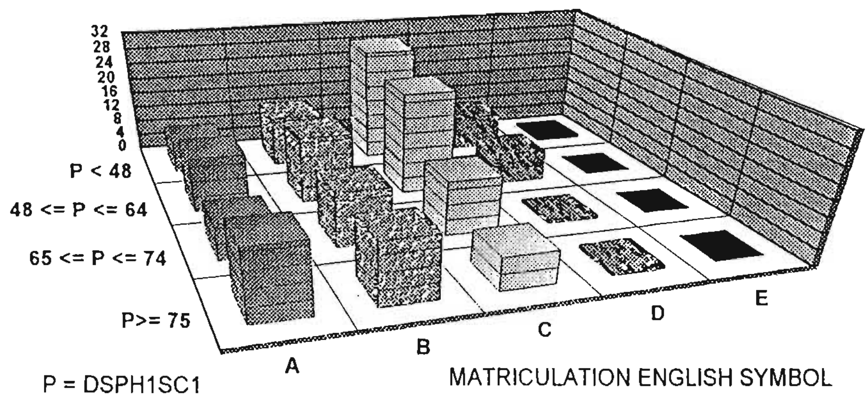


Figure 1.20: Graph illustrating the effect of the matriculation English symbol on the DSPH1SC1 course.

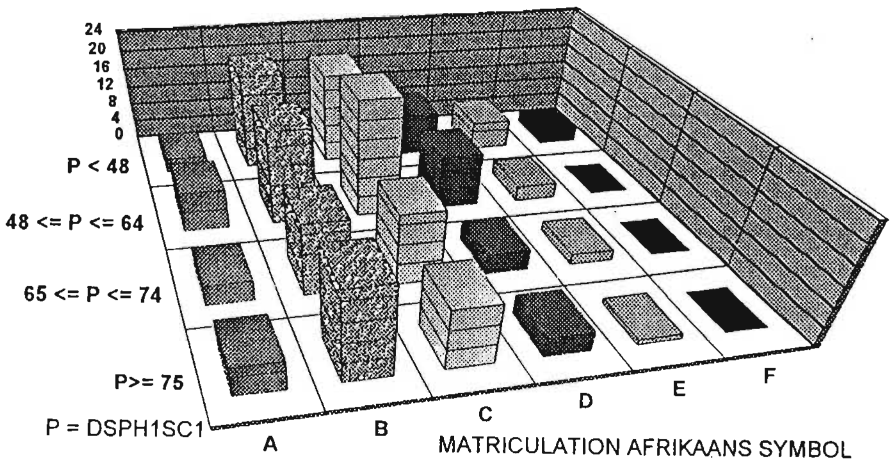


Figure 1.21: Graph illustrating the effect of the matriculation Afrikaans symbol on the DSPH1SC1 course.

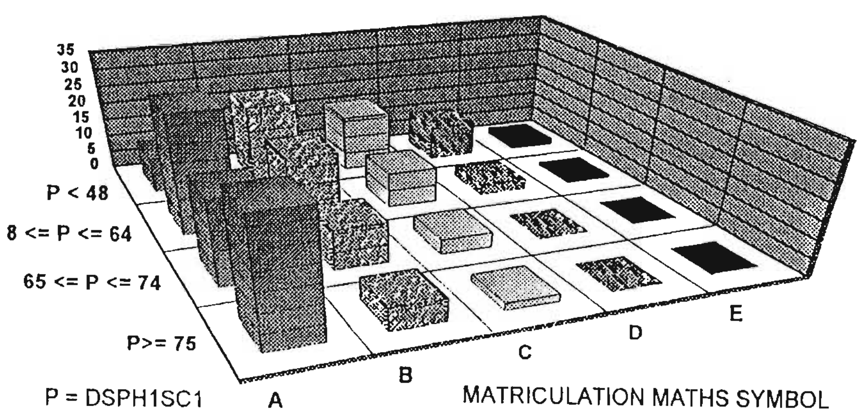


Figure 1.22: Graph illustrating the effect of the matriculation Mathematics symbol on the DSPH1SC1 course.

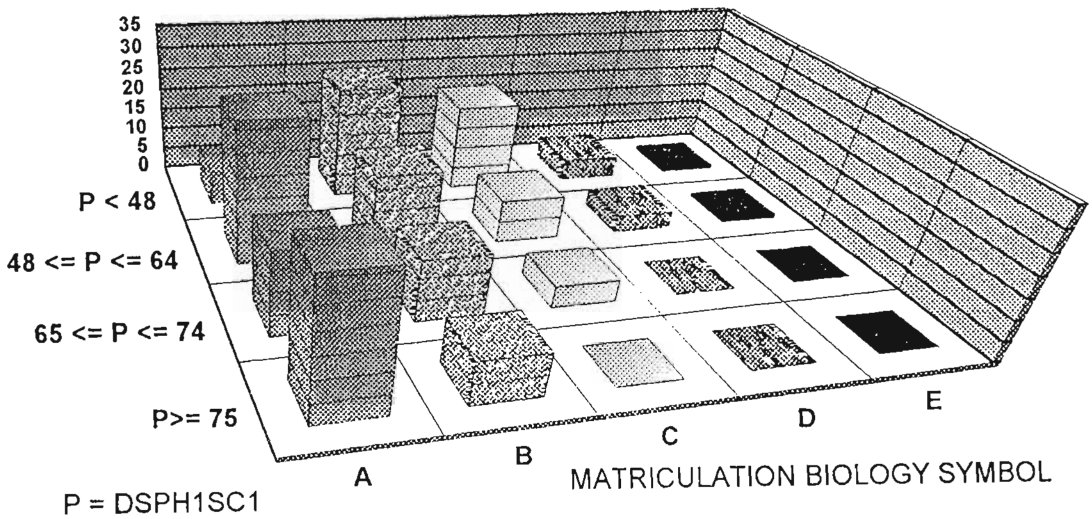


Figure 1.23: Graph illustrating the effect of the matriculation Biology symbol on the DSPH1SC1 course.

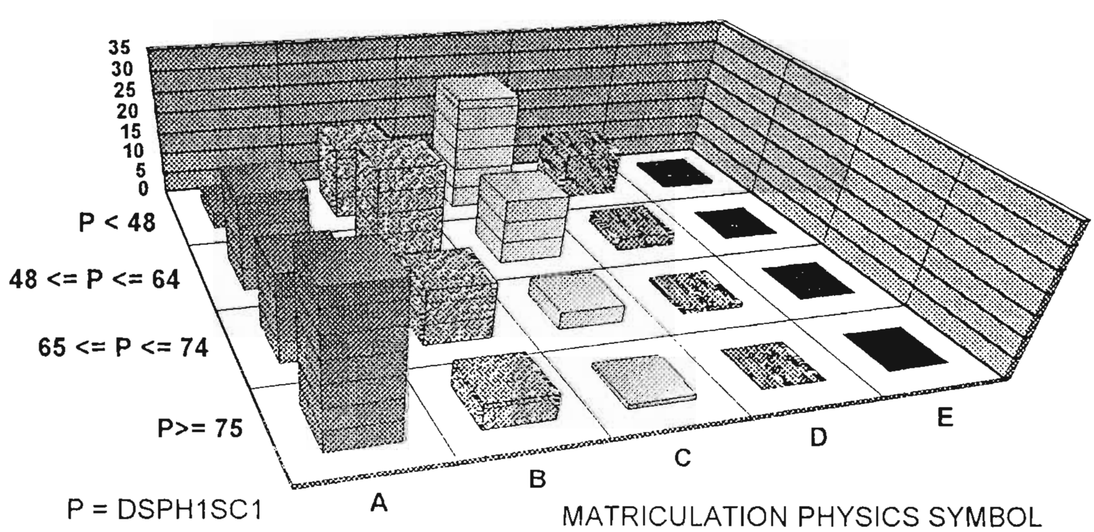


Figure 1.24: Graph illustrating the effect of the matriculation Physics symbol on the DSPH1SC1 course.

Turning our attention to an analysis of the effect that the type of matriculation authority[6] might have on the prediction process, since the majority of students in our sample were from the following matriculation authorities

- House of Delegates (HOD),
- Natal Education Department (NED),
- Transvaal Education Department (TED), and
- House of Representatives (HOR)

only the effect of these matriculation authorities will be considered in this dissertation. Due to the small number of students from the TED and HOR, these students will be grouped together in our analyses in the sense that the following two dummy variables will be used to model the effect of the matriculation authority, namely

$$\begin{aligned} D_1 &= 1 \quad \text{if the student's matriculation authority is the HOD,} \\ &= 0 \quad \text{otherwise and} \end{aligned}$$

$$\begin{aligned} D_2 &= 1 \quad \text{if the student's matriculation authority is the NED,} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

As a preliminary tool for identifying students who are likely to pass DSMA1SX1, DSPH1SC1 and then both DSMA1SX1 and DSPH1SC1, the following discriminant analysis procedure (Johnson and Wichern, 1988) was performed where a student with a matriculation profile $\mathbf{x}$ was allocated to the successful

[6]Figures 1.13 and 1.14 suggest a possible need for a matriculation authority variable due to the large proportion of students from the HOD failing. Also the large number of students with a matriculation A symbol and a university mark in the interval {48, 64} suggests an anomoly which could possibly be due to the matriculation authority.

group[7] if

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S_{\mathrm{p}}^{-1}\, \mathbf{x} > \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S_{\mathrm{p}}^{-1}\, (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)$$

where $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ represent the mean vectors that are to be associated with the successful group and unsuccessful group, respectively, and S_{p} the pooled sample variance for both groups.

Performing the discriminant analysis routine[8] on our DSMA1SX1 data set (given in Appendix C1) we obtained the following classification rule for identifying students who are likely to pass DSMA1SX1,[9] namely:
Allocate a student with a matriculation profile $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, d_1, d_2)$ to the successful group (pass) if

$$0.022\, x_1 + 0.035\, x_2 + 0.729\, x_3 + 0.434\, x_4 + 0.414\, x_5 - 0.044\, d_1 - 1.319\, d_2 \leq 1.415$$

and to the unsuccessful group (fail) if

$$0.022\, x_1 + 0.035\, x_2 + 0.729\, x_3 + 0.434\, x_4 + 0.414\, x_5 - 0.044\, d_1 - 1.319\, d_2 > 1.415$$

The small parameter estimates obtained for the English, Afrikaans and HOD authority variable suggests that these variables do not have an influence on the classification of a student to a particular group. Furthermore students writing the NED matriculation papers tend to perform much better than students coming from the other matriculation authorities.

Performing the discriminant analysis routine on the DSPH1SC1 data (given in Appendix C2) the following results were obtained, namely:

[7]For our analyses the successful group will represent students who pass and the unsuccessful group those who fail.

[8]A program for performing discriminant analysis is given in Appendix D1.

[9]A student passed if he or she obtained 48 marks or more as we assumed that the external examiner would pass those students with 48 marks or more.

Allocate a student to the successful (pass) group if

$$-0.132\,x_1 - 0.151\,x_2 + 0.573\,x_3 + 1.023\,x_4 + 1.147\,x_5 - 1.359\,d_1 - 3.80\,d_2 \leq 0.754$$

else allocate the student to the unsuccessful (fail) group. The parameter estimates obtained for English and Afrikaans when compared with the other variables seem to indicate that the matriculation English and Afrikaans marks do not drastically influence the classification of a student to a successful or unsuccessful group. Also the parameter estimate obtained for the NED matriculation authority variable seem to suggest that NED students have a much greater chance of passing than those students from any of the other matriculation authorities. Furthermore, students writing the HOD matriculation papers have a higher chance of passing DSPH1SC1 than students writing papers set by the House of Representatives or Transvaal Education Department.[10]

Finally, performing a discriminant analysis on our combined DSMA1SX1 and DSPH1SC1 data (given in Appendix C3) we obtained the following results, namely:
Allocate a student to the successful group (i.e., a student is termed successful if he passes both subjects; otherwise he is unsuccessful) if

$$0.198\,x_1 - 0.023\,x_2 + 0.836\,x_3 + 0.715\,x_4 + 0.618\,x_5 - 1.035\,d_1 - 2.912\,d_2 \leq 0.655$$

Once again the small parameter estimates obtained for English and Afrikaans seem to suggest that these variables have very little influence when identifying a successful or unsuccessful student. Furthermore, the parameter estimate obtained for the NED matriculation authority variable seem to suggest that NED students have a much better chance of passing both DSMA1SX1 and

[10]When trying to determine whether students from the HOR or TED are successful we should exercise caution as they represent only a small portion of the whole data set.

DSPH1SC1 when compared with students from the other matriculation authorities. Similarly it appears that students coming from the HOD have a greater chance of passing both DSMA1SX1 and DSPH1SC1 when compared with students from the TED and HOR matriculation authorities.

CHAPTER TWO

A GENERALIZED LINEAR MODELLING APPROACH

2.1 INTRODUCTION

The main purpose of this chapter will be to develop a technique that can be used for modelling binary responses. In particular we will examine a special class of models, called generalized linear models, that can be used to model our exam data problem. In section 2.2, a formal definition of the generalized linear model will be given with section 2.3 being devoted to the derivation of a maximum likelihood estimation technique for obtaining parameter estimates in the generalized linear model. The problem of subset selection in the generalized linear model will be examined in section 2.4 with section 2.5 being devoted to showing how a generalized linear model can be applied to our exam data by using first a logit link function and then a probit link function. Finally, in section 2.6 a discussion of our results will be given.

2.2 MODEL DESCRIPTION

Adopting the notation of Dobson (1990), consider a random variable y_i that comes from the following family of exponential distributions

$$f(y_i, \theta_i) = \exp\left[y_i b(\theta_i) + c(\theta_i) + d(y_i)\right] \tag{2.2.1}$$

where θ_i denotes a set of model parameters that are to be associated with the above distribution, and $b(\cdot)$, $c(\cdot)$ and $d(\cdot)$ represent known functions of θ_i and y_i respectively. Letting μ_i denote the expected value of y_i, the generalized linear model specification is then completed by introducing a monotonic link function that links a function of μ_i to a set of explanatory variables $(x_{i1}, \ldots, x_{ip})$ in the following way

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$$

Although a large number of choices for $g(\cdot)$ are possible one usually chooses $g(\mu_i)$ to be equal to $b(\theta_i)$[1] as this simplifies the parameter estimation problem in the above model structure. Introducing the notation

$$l(y_i, \theta_i) = y_i b(\theta_i) + c(\theta_i) + d(y_i) \tag{2.2.2}$$

to denote the log-likelihood function, and differentiating (2.2.2) with respect to θ_i yields the score statistic

$$u_i = \frac{dl}{d\theta_i} = y_i b'(\theta_i) + c'(\theta_i)$$

The vector $\mathbf{u} = (u_1, \ldots, u_n)'$ is known to have (for a regular density) a zero mean and variance-covariance matrix equal to the inverse of the Fisher Information matrix. Because of this one can obtain the result that

$$\mathcal{E}\{y_i\} b'(\theta_i) + c'(\theta_i) = 0$$

which implies that

$$\mu_i \triangleq \mathcal{E}\{y_i\} = -\frac{c'(\theta_i)}{b'(\theta_i)} \tag{2.2.3}$$

Furthermore, an expression for the variance of y_i can also be obtained by noting that

$$\mathcal{V}\{y_i b'(\theta_i) + c'(\theta_i)\} = [b'(\theta_i)]^2 \mathcal{V}\{y_i\}$$

[1]If $g(\mu_i) = b(\theta_i)$ then $g(\mu_i)$ is called the canonical link function.

and that

$$
\begin{aligned}
\mathcal{V}\left\{\frac{dl}{d\theta_i}\right\} &= \mathcal{E}\left\{\left(\frac{dl}{d\theta_i}\right)^2\right\} \\
&= -\mathcal{E}\left\{\frac{d^2 l}{d\theta_i^2}\right\} \\
&= \mathcal{E}\left\{-y_i b''(\theta_i) - c''(\theta_i)\right\} \\
&= -b''(\theta_i)\mathcal{E}\{y_i\} - c''(\theta_i) \\
&= \frac{b''(\theta_i)c'(\theta_i)}{b'(\theta_i)} - c''(\theta_i) \cdots \text{by (2.2.3)}
\end{aligned}
$$

On equating the above two expressions one can then obtain the result that

$$
\mathcal{V}\{y_i\} = \frac{[b''(\theta_i)c'(\theta_i) - b'(\theta_i)c''(\theta_i)]}{[b'(\theta_i)]^3} \tag{2.2.4}
$$

2.3 PARAMETER ESTIMATION

Given a set of observations $\{y_1, \ldots, y_n\}$ that have been drawn from the above family of distributions, an expression for the log-likelihood function can be given by

$$
l(\theta_1, \ldots, \theta_n; y_1, \ldots, y_n) = \sum_{i=1}^{n}[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \tag{2.3.1}
$$

Letting

$$
\eta_i \triangleq g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = \mathbf{x}_i'\boldsymbol{\beta} \quad i = 1, \ldots, n \tag{2.3.2}
$$

denote the link function that is to be associated with the i'th observation, maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$ can be obtained by solving the following system of equations (see Appendix A1)

$$u_j \triangleq \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\mathcal{V}\{y_i\}} \frac{\partial \mu_i}{\partial \eta_i} = 0 \quad j = 0, \ldots, p$$

Due to the non-linear form of the above set of equations, an iterative solution can be obtained by employing the following Newton-Raphson method (Gallant, 1987) that updates the previous estimate of $\boldsymbol{\beta}$, which we will denote by $\mathbf{b}^{(m-1)}$, as follows

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} - \mathbf{H}^{(m-1)}\,\mathbf{u}^{(m-1)} \qquad m = 1, 2, \ldots$$

where

$$\mathbf{H}^{(m-1)} = \left.\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\mathbf{b}^{(m-1)}}$$

denotes the value of the Hessian matrix that has been evaluated at the previous iteration point $\mathbf{b}^{(m-1)}$, and

$$\mathbf{u}^{(m-1)} = \left.\left(\frac{\partial l}{\partial \beta_0}, \ldots, \frac{\partial l}{\partial \beta_p}\right)'\right|_{\boldsymbol{\beta}=\mathbf{b}^{(m-1)}}$$

denotes the value of the score vector that has been evaluated at the previous iteration point $\mathbf{b}^{(m-1)}$. Alternatively, the Fisher method of scoring (Dobson, 1990) which replaces the Hessian matrix with its matrix of expected values can be used yielding the following iterative recursion formula

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + [\mathcal{I}^{(m-1)}]^{-1}\,\mathbf{u}^{(m-1)} \tag{2.3.3}$$

where

$$\mathcal{I}^{(m-1)} = -\mathcal{E}\left.\left\{\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right\}\right|_{\boldsymbol{\beta}=\mathbf{b}^{(m-1)}} = \mathcal{E}\left.\left\{\frac{\partial l}{\partial \boldsymbol{\beta}}\frac{\partial l}{\partial \boldsymbol{\beta}'}\right\}\right|_{\boldsymbol{\beta}=\mathbf{b}^{(m-1)}}$$

denotes the value of the Information matrix evaluated at the previous iteration point $\mathbf{b}^{(m-1)}$. The above iterative routine can however be rewritten in the form of an iteratively re-weighted least squares procedure, if on letting

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

and $\mathbf{W}^{(m-1)}$ denote a $n \times n$ diagonal matrix with i'th element

$$w_{ii}^{(m-1)} = \frac{1}{\mathcal{V}\{y_i\}} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \Bigg|_{\boldsymbol{\beta}=\mathbf{b}^{(m-1)}}$$

one notes that (see Appendix A2)

$$\mathcal{I}^{(m-1)} = \mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{X}$$

and that

$$\mathbf{u}^{(m-1)} = \mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{z}^{*(m-1)}$$

where $\mathbf{z}^{*(m-1)}$ denotes a n dimensional vector with i'th component

$$z_i^{*(m-1)} = (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \Bigg|_{\boldsymbol{\beta}=\mathbf{b}^{(m-1)}}$$

Then on pre-multiplying both sides of (2.3.3) by $\mathcal{I}^{(m-1)}$ one can obtain the result that

$$\mathcal{I}^{(m-1)}\,\mathbf{b}^{(m)} = \mathcal{I}^{(m-1)}\,\mathbf{b}^{(m-1)} + \mathbf{u}^{(m-1)}$$

and thus that

$$\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{X}\mathbf{b}^{(m)} = \mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{z}^{(m-1)}$$

where $\mathbf{z}^{(m-1)} = \mathbf{X}\mathbf{b}^{(m-1)} + \mathbf{z}^{*(m-1)}$.

The above result then implies that the m'th iteration of the Fisher method of scoring can be implemented by setting

$$\mathbf{b}^{(m)} = (\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(\mathbf{m}-1)}\mathbf{z}^{(m-1)} \quad m = 1, 2, \ldots \tag{2.3.4}$$

with the maximum likelihood estimate being obtained upon convergence of the above algorithm, that is

$$\hat{\boldsymbol{\beta}} = \lim_{m \to \infty} \mathbf{b}^{(m)}$$

Letting

$$\hat{\mathbf{W}} = \lim_{m \to \infty} \mathbf{W}^{(m-1)}$$

asymptotic distribution theory results then imply that

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \hat{\mathcal{I}}^{-1})$$

where

$$\hat{\mathcal{I}} = \mathbf{X}'\hat{\mathbf{W}}\mathbf{X}$$

2.4 SUBSET SELECTION

To determine which matriculation subjects and authorities are important for predicting university results we need to develop a suitable subset selection procedure. In particular, given a matriculation profile vector containing $p+1$ variables we need to determine which variables are to be excluded from the link function of the generalized linear model. In order to do this, let us partition our matrix of independent variables (and thus vector of link function parameters) as follows

$$\mathbf{X} = (\mathbf{X_1}\ \mathbf{X_2}) \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \tag{2.4.1}$$

where $\mathbf{X_1}$ denotes a $n \times (k+1)$ dimensional matrix of explanatory variables which we want to include in our model, and $\mathbf{X_2}$ a $n \times (p-k)$ dimensional matrix of variables that we want to exclude from our model. To determine whether the set of independent variables $\mathbf{X_2}$ can be left out of the generalized linear model we need then only derive a test procedure for testing

$$H_0: \ \boldsymbol{\beta}_2 = 0 \quad \text{vs} \quad H_1: \ \boldsymbol{\beta}_2 \neq 0$$

or alternatively

$$H_0: \ \mathbf{R}\boldsymbol{\beta} = 0 \quad \text{vs} \quad H_1: \ \mathbf{R}\boldsymbol{\beta} \neq 0$$

where one has appropriately set $\mathbf{R} = (\mathbf{I}, \mathbf{0})$, with $\mathbf{0}$ representing a $(p-k) \times (k+1)$ matrix of zeroes and $\mathbf{I}$ a $(p-k) \times (p-k)$ identity matrix. Such a test can be taken using either a likelihood ratio, Wald or a Lagrange multiplier test statistic (Buse, 1982; Lawless and Singhal, 1978). Letting $\hat{\boldsymbol{\beta}}$ denote the unrestricted estimate of $\boldsymbol{\beta}$ that has been obtained using the Fisher scoring routine and

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= \lim_{m\to\infty}[(\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{z}^{(m-1)} - (\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{X})^{-1}\mathbf{R}' \\ &\qquad \{\mathbf{R}(\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{X})^{-1}\mathbf{R}'\}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{(m-1)}\mathbf{z}^{(m-1)}]\end{aligned}$$

the estimate of $\boldsymbol{\beta}$ that has been obtained subject to the restrictions that $\mathbf{R}\boldsymbol{\beta} = 0$ (see Appendix A3), then the relevant test statistics can be given by:

$$D = 2\{l(\hat{\boldsymbol{\beta}}, \mathbf{y}) - l(\tilde{\boldsymbol{\beta}}, \mathbf{y})\} \tag{2.4.2}$$

for the likelihood ratio test statistic[2] which follows a chi-squared distribution with $v = p - k$ degrees of freedom when H_0 is true; by

$$\mathbf{WT} = (\mathbf{R}\hat{\boldsymbol{\beta}})'\{\mathbf{R}(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\}^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}) \tag{2.4.3}$$

for the Wald test statistic which follows asymptotically a chi-squared distribution with $v = p - k$ degrees of freedom when H_0 is true, and by

$$\mathbf{LM} \triangleq \mathbf{u}(\tilde{\boldsymbol{\beta}})'\mathcal{I}^{-1}\mathbf{u}(\tilde{\boldsymbol{\beta}}) \tag{2.4.4}$$

for the Lagrange multiplier test statistic[3] which also follows asymptotically a chi-squared distribution with $v = p - k$ degrees of freedom.

2.5 APPLICATION TO EXAM DATA

In order to model the success (or failure) of a student when taking a particular course at university and to associate this success (or failure) with the outcome of a Bernoulli random variable, we need to consider a student as

[2] Alternatively an approximation, derived in Appendix A4, for the likelihood ratio test statistic can be given by

$$D^* = 2[l(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) - l(\hat{\boldsymbol{\beta}}_1 + \mathcal{I}_{11}^{-1}\mathcal{I}_{12}\hat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}_2 = \mathbf{0})]$$

where

$$\mathcal{I} = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix}$$

denotes a partitioned form of the information matrix that is to be associated with the parameter vector $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1' \, \boldsymbol{\beta}_2')$. Asymptotically D^* then also follows a chi-squared distribution with $v = p - k$ degrees of freedom if H_0 is true.

[3] where

$$\mathbf{u}(\tilde{\boldsymbol{\beta}}) = \left(\frac{\partial l}{\partial \beta_0}, \ldots, \frac{\partial l}{\partial \beta_p}\right)'\Bigg|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}$$

and

$$\mathcal{I} = \mathcal{E}\left\{\frac{\partial l}{\partial \boldsymbol{\beta}}\frac{\partial l}{\partial \boldsymbol{\beta}'}\right\}\Bigg|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}$$

being successful if they obtain a mark that lies within a certain prespecified interval that is of interest. For example, if one is intending to model the probability that a student will pass DSMA1SX1, then the interval of interest will be given by that student's mark lying in the following set, namely

$$I_M = \{\text{DSMA1SX1 mark} \geq 48\}$$

Clearly the above definition can be extended to account for more than one university subject if for each subject an appropriate mark interval is constructed with the student being classified as being successful if all of his or her university marks lie in these intervals. For example, suppose that we are interested in a student obtaining a first class pass for DSMA1SX1 and a second class pass for DSPH1SC1. By letting

$$I_M = \{75 \leq \text{DSMA1SX1 mark} \leq 100\}$$

and

$$I_P = \{60 \leq \text{DSPH1SC1 mark} \leq 74\}$$

denote the appropriate mark intervals of interest, then a student can be classified as being successful if their DSMA1SX1 mark lies in the interval set I_M, and their DSPH1SC1 mark in the interval I_P. Introducing a Bernoulli random variable to model the occurrence of this event with y_i being set equal to 1 if the i'th student obtains the desired event and set equal to 0 if he or she does not, and letting

$$\pi_i = P(y_i = 1|\mathbf{x}_i')$$

denote the probability that this student will be successful, our approach will be to associate this probability value with the student's matriculation profile vector $\mathbf{x}_i'$ using initially a logit link function defined by[4]

$$g(\mu_i) = g(\pi_i) \triangleq \log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i'\boldsymbol{\beta} = \eta_i \tag{2.5.1}$$

and then a probit link function of the form

$$g(\pi_i) \triangleq \Phi^{-1}(\pi_i) = \mathbf{x}_i'\boldsymbol{\beta} = \eta_i \tag{2.5.2}$$

where Φ denotes the cumulative standard normal distribution function. Because

$$\begin{aligned} f(y_i, \pi_i) &= \pi_i^{y_i}(1-\pi_i)^{1-y_i} \\ &= \exp\left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log(1-\pi_i)\right] \end{aligned} \tag{2.5.3}$$

represents the probability density function of y_i, with a mean and a variance parameter given by π_i and $\pi_i(1-\pi_i)$ respectively, and on noting, for the logit link function, that

$$\begin{aligned} \frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \pi_i}{\partial \eta_i} &= \frac{\partial}{\partial \eta_i}\left(\frac{e^{\eta_i}}{1+e^{\eta_i}}\right) \\ &= \frac{(1+e^{\eta_i})e^{\eta_i} - e^{2\eta_i}}{(1+e^{\eta_i})^2} \\ &= \frac{e^{\mathbf{x}'_i\boldsymbol{\beta}}}{[1+e^{\mathbf{x}'_i\boldsymbol{\beta}}]^2} \\ &= \pi_i(1-\pi_i) \end{aligned}$$

[4]Manipulating (2.5.1) the following expression for π_i can be obtained, namely

$$\pi_i = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i'\boldsymbol{\beta}}}$$

where

$$\pi_i = \frac{e^{\mathbf{x'}_i\boldsymbol{\beta}}}{1 + e^{\mathbf{x'}_i\boldsymbol{\beta}}}$$

one can obtain the following substitutions for the Fisher scoring routine defined in section 2.3, namely

$$w_{ii}^{(m-1)} = \frac{1}{\mathrm{var}(y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \Bigg|_{\boldsymbol{\beta}=\mathbf{b}^{(m-1)}} = \pi_i^{(m-1)}(1 - \pi_i^{(m-1)})$$

and (for the i'th element of $\mathbf{z}^{(m-1)}$),

$$z_i^{(m-1)} = \mathbf{x}_i'\mathbf{b}^{(m-1)} + \frac{(y_i - \pi_i^{(m-1)})}{\pi_i^{(m-1)}(1 - \pi_i^{(m-1)})}$$

Having obtained upon suitable convergence of the Fisher's scoring routine a maximum likelihood estimate for $\boldsymbol{\beta}$, which we will denote by $\hat{\boldsymbol{\beta}}$, this estimate can then be substituted into the following formula

$$\hat{\pi}_i = \frac{e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}}}{1 + e^{\mathbf{x}_i'\hat{\boldsymbol{\beta}}}} \tag{2.5.4}$$

to obtain an estimate for the success probability of a student with a matriculation profile $\mathbf{x}_i'$. For large sample sizes a $(1-\alpha)100\%$ confidence interval for β_i can be defined by the expression

$$\hat{\beta}_i \pm \mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\hat{\mathcal{I}}_{ii}^{-1}}$$

where z represents a standardized normal random variable and on defining $\hat{\boldsymbol{\sigma}}$ to be a vector with i'th element

$$\hat{\sigma}_i = \sqrt{\hat{\mathcal{I}}_{ii}^{-1}}$$

a $(1-\alpha)100\%$ confidence interval for π_i can be given by (Hosmer et al. 1989)

$$\frac{\exp\left(\mathbf{x}_i'\left[\hat{\boldsymbol{\beta}} \pm z_{1-\frac{\alpha}{2}}\hat{\boldsymbol{\sigma}}\right]\right)}{1+\exp\left(\mathbf{x}_i'\left[\hat{\boldsymbol{\beta}} \pm z_{1-\frac{\alpha}{2}}\hat{\boldsymbol{\sigma}}\right]\right)} \tag{2.5.5}$$

Turning our attention to the Probit link function, one can obtain the result that

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \pi_i}{\partial \eta_i} = \frac{\partial}{\partial \eta_i}\Phi(\eta_i) = \phi(\eta_i) = \phi(\mathbf{x}_i'\boldsymbol{\beta})$$

where ϕ represents the standard normal density function, and thus the Fisher scoring routine substitutions takes the form

$$w_{ii}^{(m-1)} = \frac{1}{\operatorname{var}(y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\Bigg|_{\boldsymbol{\beta}=\mathbf{b}^{(m-1)}} = \frac{[\phi(\mathbf{x}_i'\mathbf{b}^{(m-1)})]^2}{\pi_i^{(m-1)}(1-\pi_i^{(m-1)})}$$

and

$$z_i^{(m-1)} = \mathbf{x}_i'\mathbf{b}^{(m-1)} + \frac{y_i - \pi_i^{(m-1)}}{\phi(\mathbf{x}_i'\mathbf{b}^{(m-1)})}$$

with an estimate for π_i, for a given set of independent variables $\mathbf{x}_i'$, being given by

$$\hat{\pi}_i = \Phi(\mathbf{x}_i'\hat{\boldsymbol{\beta}}) \tag{2.5.6}$$

and a $(1-\alpha)100\%$ confidence interval for π_i being given by the expression

$$\boldsymbol{\Phi}\left(\mathbf{x}_i'\left[\hat{\boldsymbol{\beta}} \pm z_{1-\frac{\alpha}{2}}\hat{\boldsymbol{\sigma}}\right]\right) \tag{2.5.7}$$

2.6 RESULTS

Turning our attention once again to the three different problems that we analysed earlier in chapter one, namely the problem of associating the success probability of a student passing DSMA1SX1, DSPH1SC1 and then both DSMA1SX1 and DSPH1SC1, with their matriculation profile record, a generalized linear model with a logit[5] link function was applied to our DSMA1SX1 data with the event of interest being that a student obtains a mark greater than or equal to 48 for that particular subject. Our analysis yielded the following results, namely

Table 2.6.1

Parameter	$\hat{\beta}$	SE*1	WT*2	PROB*3	LB*4	UB*5
β_0	1.8168972	0.6670501	7.4189739	0.006454	0.5094791	3.1243153
β_1	−0.001621	0.1860349	0.000076	0.993046	−0.36625	0.3630071
β_2	−0.069744	0.1590199	0.1923588	0.6609606	−0.381423	0.2419349
β_3	−0.650262	0.1793891	13.139672	0.0002891	−1.001864	−0.298659
β_4	−0.430301	0.1772003	5.8967842	0.0151686	−0.777613	−0.082988
β_5	−0.376458	0.1889741	3.9685174	0.0463586	−0.746847	−0.006069
β_6	0.0136196	0.5624001	0.0005865	0.9806795	−1.088685	1.1159239
β_7	1.1969564	0.614631	3.7925191	0.0514821	−0.00772	2.4016332

A backward elimination type procedure with a variable exclusion probability level of 0.1 was then applied to our model with the following results being achieved, namely

[5]The initial link function is given by

$$g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 d_{i1} + \beta_7 d_{i2}$$

where X_1, X_2, X_3, X_4 and X_5 are defined on page 5 and d_1 and d_2 on page 15.

*1standard error of the parameter estimate

*2the Wald Test Statistic

*3Probability that the Wald Test Statistic is greater than χ_1^2

*4lower bound for the 95% confidence interval of β

*5upper bound for the 95% confidence interval of β

Table 2.6.2

Parameter	$\hat{\beta}$	SE	WT	PROB	LB	UB
β_0	1.731933	0.2623238	43.589974	4.049E-11	1.2177782	2.2460877
β_3	−0.64638	0.1783844	13.129916	0.0002906	−0.996013	−0.296747
β_4	−0.449118	0.1687949	7.0795033	0.0077971	−0.779956	−0.11828
β_5	−0.396369	0.1819488	4.7457053	0.0293715	−0.752989	−0.039749
β_7	1.1877764	0.3349495	12.575082	0.0003909	0.5312755	1.8442773

From the above table an estimate of the probability that can be associated with a student passing DSMA1SX1, for a specified matriculation profile $\mathbf{x}_i' = (1\, x_{i3}\, x_{i4}\, x_{i5}\, d_{i2})$, can be given by

$$\begin{aligned}\hat{\pi}_i &= \frac{\exp(\hat{\beta}_0 + \hat{\beta}_3 x_{i3} + \hat{\beta}_4 x_{i4} + \hat{\beta}_5 x_{i5} + \hat{\beta}_7 d_{i2})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_3 x_{i3} + \hat{\beta}_4 x_{i4} + \hat{\beta}_5 x_{i5} + \hat{\beta}_7 d_{i2})} \\ &= \frac{1}{1 + \exp[-(\hat{\beta}_0 + \hat{\beta}_3 x_{i3} + \hat{\beta}_4 x_{i4} + \hat{\beta}_5 x_{i5} + \hat{\beta}_7 d_{i2})]}\end{aligned}$$

As would be expected, the negative parameter estimates obtained for Mathematics (X_3), Biology (X_4) and Physics (X_5) indicate that lower symbols for these matriculation subjects are associated with a lower probability of a student passing DSMA1SX1. Furthermore the magnitude of all the regression coefficients imply that a student's matriculation Mathematics symbol has the greatest influence on the probability of passing the DSMA1SX1 course. The significantly positive estimate obtained for the NED variable implies that students writing the Natal Education Department matriculation papers have a much greater probability of passing DSMA1SX1 when compared with students from any of the other matriculation authorities.[6]

[6] Analyzing student results from the Transvaal Education Department and the House of Representatives should be handled with caution as these students only represent a small percentage of the total number of students in our data set.

In Table 2.6.3 we have provided a brief summary of some of the estimated success probabilities, for passing DSMA1SX1, that can be associated with a certain type of matriculation profile result.[7]

Table 2.6.3

X_3	X_4	X_5	d_2	$\hat{\pi}$	LB($\hat{\pi}$)*6	UB($\hat{\pi}$)*7
0	0	0	0	0.8496595	0.7716723	0.9043125
1	0	0	0	0.7475434	0.5552151	0.8753748
0	1	0	0	0.7829285	0.6077399	0.8935767
0	0	1	0	0.7917595	0.6141498	0.9008173
0	0	0	1	0.9488122	0.8518334	0.9835423
1	0	0	1	0.906644	0.6798408	0.9779817
0	1	0	1	0.9220543	0.7249395	0.981514
0	0	1	1	0.925762	0.7302842	0.9828863

Turning our attention to the DSPH1SC1 data a generalized linear model implementation with a logit link function produced the following results, namely

Table 2.6.4

Parameter	$\hat{\beta}$	SE	WT	PROB	LB	UB
β_0	1.5416826	0.9428239	2.6737992	0.1020122	−0.306252	3.3896174
β_1	0.216685	0.2808714	0.5951725	0.4404259	−0.333823	0.7671929
β_2	0.1609419	0.2514019	0.4098277	0.522057	−0.331806	0.6536895
β_3	−0.537684	0.2577496	4.3516973	0.0369721	−1.042874	−0.032495
β_4	−1.1926	0.3289327	13.145444	0.0002882	−1.837308	−0.547892
β_5	−1.160835	0.3247551	12.77701	0.0003509	−1.797355	−0.524315
β_6	1.1429925	0.825125	1.9188776	0.1659804	−0.474252	2.7602375
β_7	4.8582069	1.1203755	18.802899	0.0000145	2.6622709	7.0541428

[7]From chapter one 0 represents an A symbol for a particular subject and 1 represents a B symbol. As an example the fifth row in Table 2.6.3 refers to an NED student with A's in matriculation Mathematics, Biology and Physics.

*6lower bound for the 95% confidence interval of the success probability

*7upper bound for the 95% confidence interval of the success probability

An application of the backward elimination routine, using a probability value of 0.1, yielded the following results, namely

Table 2.6.5

Parameter	$\hat{\beta}$	SE	WT	PROB	LB	UB
β_0	2.9552285	0.4270913	47.878523	4.535E-12	2.1181296	3.7923274
β_3	−0.542827	0.2478101	4.7982761	0.0284882	−1.028535	−0.057119
β_4	−1.009345	0.2971657	11.53671	0.0006824	−1.59179	−0.4269
β_5	−1.076835	0.2977927	13.075852	0.0002991	−1.660508	−0.493161
β_7	3.6106889	0.7464047	23.400836	1.3152E-6	2.1477356	5.0736422

Once again, we notice that lower matriculation symbols in Mathematics, Biology and Physics are associated with a lower probability of a student passing the DSPH1SC1 course. Also the large negative value associated with a student's matriculation Physics mark implies that this subject has the greatest influence on the probability of that student passing DSPH1SC1 when compared with his or her matriculation Biology and Mathematics marks. Furthermore, students writing the matriculation examination set by the Natal Education Department have a far greater probability of passing the DSPH1SC1 course when compared with students from the other matriculation departments. A few estimated probabilities, along with their associated matriculation profiles are summarised in Table 2.6.6.

Table 2.6.6

X_3	X_4	X_5	d_2	$\hat{\pi}$	LB($\hat{\pi}$)	UB($\hat{\pi}$)
0	0	0	0	0.95051	0.8926528	0.9779539
1	0	0	0	0.9177681	0.7483054	0.9766882
0	1	0	0	0.8749971	0.628629	0.9666064
0	0	1	0	0.8674265	0.6124498	0.9644002
0	0	0	1	0.9985944	0.9861547	0.9998589
1	0	0	1	0.9975837	0.9622152	0.9998506
0	1	0	1	0.9961529	0.9354794	0.9997838
0	0	1	1	0.9958854	0.9312056	0.999769

Turning our attention now to the final problem, namely that of determining the probability that a student will pass both DSMA1SX1 and DSPH1SC1, a generalized linear model implementation with a logit link function produced the following results, namely

Table 2.6.7

Parameter	$\hat{\beta}$	SE	WT	PROB	LB	UB
β_0	0.745716	0.9055002	0.6782189	0.4102007	−1.029064	2.5204964
β_1	−0.096568	0.248199	0.1513802	0.6972199	−0.583038	0.3899018
β_2	−0.041521	0.2406129	0.029778	0.862995	−0.513122	0.4300803
β_3	−0.827465	0.26855	9.493996	0.0020615	−1.353823	−0.301107
β_4	−0.782912	0.2867977	7.4520238	0.0063365	−1.345035	−0.220788
β_5	−0.584166	0.2685855	4.7304982	0.0296323	−1.110593	−0.057738
β_6	1.1815894	0.8104127	2.1257919	0.1448379	−0.406819	2.7699982
β_7	3.4470168	0.9688301	12.658771	0.0003738	1.5481098	5.3459238

Employing a backward elimination routine we then obtained the following results, namely

Table 2.6.8

Parameter	$\hat{\beta}$	SE	WT	PROB	LB	UB
β_0	1.696732	0.3091204	30.128062	4.044E-8	1.090856	2.3026079
β_3	−0.801685	0.2652619	9.1339401	0.0025091	−1.321599	−0.281772
β_4	−0.817072	0.2657783	9.451085	0.0021102	−1.337997	−0.296147
β_5	−0.655945	0.2606012	6.3355243	0.0118343	−1.166724	−0.145167
β_7	2.4002477	0.568151	17.847792	0.0000239	1.2866716	3.5138237

The only surprising result here, was that a student's matriculation Biology mark now seems to have the greatest influence on the probability of passing both DSMA1SX1 and DSPH1SC1 when compared with his or her other matriculation subjects. In Table 2.6.9 a few estimated probabilities for passing both DSMA1SX1 and DSPH1SC1 are given along with their associated matriculation profile results.

Table 2.6.9

X_3	X_4	X_5	d_2	$\hat{\pi}$	LB($\hat{\pi}$)	UB($\hat{\pi}$)
0	0	0	0	0.8451074	0.7485429	0.9090928
1	0	0	0	0.7099305	0.4425689	0.8829674
0	1	0	0	0.7067518	0.4385273	0.8814738
0	0	1	0	0.7390017	0.4810421	0.8963621
0	0	0	1	0.983649	0.9150975	0.9970306
1	0	0	1	0.964267	0.7419118	0.996068
0	1	0	1	0.9637331	0.7387594	0.9960113
0	0	1	1	0.9689626	0.770441	0.9965683

Repeating the above analysis but now with a probit link function we obtained (after employing the backward elimination procedure) the following results for the DSMA1SX1 data, namely[5]

Table 2.6.10

Parameter	$\hat{\beta}$	SE	WT	PROB	LB	UB
β_0	1.0436877	0.1495514	48.703466	2.977E-12	0.7505668	1.3368085
β_3	−0.381888	0.1057402	13.04343	0.0003044	−0.589139	−0.174637
β_4	−0.262207	0.099141	6.9949223	0.0081741	−0.456524	−0.067891
β_5	−0.243779	0.1080707	5.088337	0.0240873	−0.455597	−0.03196
β_7	0.6805545	0.1910991	12.682609	0.0003691	0.3060003	1.0551088

The probability associated with a student passing DSMA1SX1, given a particular matriculation profile $\mathbf{x}_i' = (1\, x_{i3}\, x_{i4}\, x_{i5}\, d_{i2})$, can then be estimated by

$$\hat{\pi}_i = \Phi(\hat{\beta}_0 + \hat{\beta}_3 x_{i3} + \hat{\beta}_4 x_{i4} + \hat{\beta}_5 x_{i5} + \hat{\beta}_7 d_{i2})$$

These results are very similar to those obtained when a logit link function was employed. In Table 2.6.11 we have provided a brief summary of some of the estimated success probabilities that can be associated with a given matriculation result profile, namely

[5]For these three problems the initial link function is defined by

$$g(\pi_i) = \Phi^{-1}(\pi_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 d_{i1} + \beta_7 d_{i2}$$

Table 2.6.11

X_3	X_4	X_5	d_2	$\hat{\pi}$	LB($\hat{\pi}$)	UB($\hat{\pi}$)
0	0	0	0	0.851685	0.7735433	0.9093574
1	0	0	0	0.7459501	0.5641218	0.877417
0	1	0	0	0.78274	0.6156375	0.8977648
0	0	1	0	0.7881182	0.6159914	0.9040277
0	0	0	1	0.9576679	0.8546454	0.9916197
1	0	0	1	0.9102594	0.6799032	0.986698
0	1	0	1	0.9281342	0.7257613	0.989938
0	0	1	1	0.9306252	0.7260699	0.9908615

An analysis of the DSPH1SC1 data (after employing the backward elimination procedure) yielded the following results, namely

Table 2.6.12

Parameter	$\hat{\beta}$	SE	WT	PROB	LB	UB
β_0	1.737115	0.2257253	59.223846	1.41E-14	1.2946934	2.1795366
β_3	−0.317817	0.1424659	4.9765785	0.0256927	−0.59705	−0.038583
β_4	−0.559217	0.1672549	11.179004	0.0008273	−0.887037	−0.231397
β_5	−0.650865	0.1689926	14.83359	0.0001174	−0.98209	−0.319639
β_7	2.0568496	0.4100291	25.163768	5.2662E-7	1.2531925	2.8605067

In Table 2.6.13 a few estimated probabilities for passing DSPH1SC1, that can be associated with a specific set of matriculation profiles, are provided.

Table 2.6.13

X_3	X_4	X_5	d_2	$\hat{\pi}$	LB($\hat{\pi}$)	UB($\hat{\pi}$)
0	0	0	0	0.958816	0.902287	0.9853541
1	0	0	0	0.922094	0.7572999	0.9838611
0	1	0	0	0.8805813	0.6582371	0.9743008
0	0	1	0	0.8613158	0.6227089	0.96855
0	0	0	1	0.9999259	0.9945811	0.9999998
1	0	0	1	0.9997547	0.9744617	0.9999997
0	1	0	1	0.9993912	0.9516281	0.9999992
0	0	1	1	0.9991642	0.9413018	0.9999988

Finally, turning our attention to the combined DSMA1SX1 and DSPH1SC1 data problem, a probit analysis (after the backward elimination procedure was employed) produced the following results, namely

Table 2.6.14

Parameter	$\hat{\beta}$	SE	WT	PROB	LB	UB
β_0	0.9830133	0.1723893	32.516086	1.1821E-8	0.6451303	1.3208963
β_3	−0.448145	0.1504812	8.8689487	0.0029006	−0.743088	−0.153202
β_4	−0.458362	0.1517864	9.1190964	0.0025295	−0.755864	−0.160861
β_5	−0.377692	0.1507272	6.2790372	0.0122174	−0.673118	−0.082267
β_7	1.2965121	0.3026867	18.347057	0.0000184	0.7032461	1.8897781

In Table 2.6.15 a few estimated success probabilities for passing both DSMA1SX1 and DSPH1SC1, that can be associated with a specific set of matriculation profiles, are summarised.

Table 2.6.15

X_3	X_4	X_5	d_2	$\hat{\pi}$	LB($\hat{\pi}$)	UB($\hat{\pi}$)
0	0	0	0	0.8371996	0.7405786	0.906732
1	0	0	0	0.7036296	0.460983	0.878535
0	1	0	0	0.7000871	0.4559139	0.8769828
0	0	1	0	0.7275171	0.4888361	0.8922586
0	0	0	1	0.9886821	0.9112313	0.9993379
1	0	0	1	0.9664781	0.7275063	0.9988839
0	1	0	1	0.965709	0.7232464	0.9988551
0	0	1	1	0.9714035	0.7502443	0.9991212

We notice that on employing the backward elimination procedure on both our probit and logit models the same variables were excluded based on the Wald test statistic procedure. Furthermore, the confidence intervals for the estimated probabilities obtained using both the probit and logit models were also of the same order. It can therefore be concluded that for our DSMA1SX1/DSPH1SC1 data problem both techniques lead to similar results.

CHAPTER THREE

A MONTE CARLO SIMULATION APPROACH

3.1 INTRODUCTION

The main purpose of this chapter will be to present an alternative method for relating the probability of success at university to a student's matriculation profile. In particular, letting $Y_1, Y_2, \ldots, Y_p$ denote a set of p university subject variables that are of interest, and

$$A = \{c_{11} \leq Y_1 \leq c_{12}\} \cup \{c_{21} \leq Y_2 \leq c_{22}\} \cup \ldots \cup \{c_{p1} \leq Y_p \leq c_{p2}\}$$

an event that we wish to model, a Monte Carlo simulation approach will be used to estimate the probability of a student being successful. For example, let Y_1 denote the DSMA1SX1 mark and Y_2 the DSPH1SC1 mark of a particular student, then an event of the form

$$A = \{Y_1 \geq 48\} \cup \{Y_2 \geq 48\}$$

will represent a pass mark for both DSMA1SX1 and DSPH1SC1 for that particular student. In section 3.2 we will begin by briefly discussing the assumptions that are necessary for our model, with section 3.3 being used to derive parameter estimates. In section 3.4, a seemingly unrelated regression modelling approach (Huang, 1970) will be developed with section 3.5 being devoted to a practical application of our model and a discussion of our results.

3.2 MODEL ASSUMPTIONS

Letting $\mathbf{y}_i' = (y_{i1}\, y_{i2} \ldots y_{ip})$ denote the set of marks obtained by student i for his or her p university subjects, and letting $\mathbf{x}_i' = (1\, x_{i1} \ldots x_{ik})$ denote the matriculation profile vector that is to be associated with this student, then the following assumptions will be made concerning our proposed model structure, namely that

i) the conditional density, $f(\mathbf{y}_i|\mathbf{x}_i)$ is Gaussian with a mean $\boldsymbol{\mu}_i' = \mathbf{x}_i'\mathbf{B}$ where $\mathbf{B}$ represents a matrix of regression coefficients of the form

$$\mathbf{B} = \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{kp} \end{pmatrix} \tag{3.2.1}$$

and a variance-covariance matrix

$$\boldsymbol{\Sigma}_p = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} \tag{3.2.2}$$

that is independent of the predictor variables, and that

ii) the observations $\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_n'$ are independent of each other.

Conditional on a matriculation profile record $\mathbf{x}_i$, a probability density function of $\mathbf{y}_i$ can therefore be given by the expression

$$f(\mathbf{y}_i|\mathbf{B}'\mathbf{x}_i, \boldsymbol{\Sigma}_p) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}_p|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}_i - \mathbf{B}'\mathbf{x}_i)'\boldsymbol{\Sigma}_p^{-1}(\mathbf{y}_i - \mathbf{B}'\mathbf{x}_i)\right] \tag{3.2.3}$$

Letting

$$A = \{c_{11} \leq Y_1 \leq c_{12}\} \cup \{c_{21} \leq Y_2 \leq c_{22}\} \cup \ldots \cup \{c_{p1} \leq Y_p \leq c_{p2}\}$$

denote the event of interest then providing parameter estimates are available for both $\mathbf{\beta}$ and $\mathbf{\Sigma}_p$, the probability of obtaining this event can then be given by

$$\begin{aligned} \pi(\mathbf{x}) &= \int P[\mathbf{y} \in A \,|\, \mathbf{x}] \\ &= \int_A f(\mathbf{y}|\mathbf{x})d\mathbf{y} \\ &= \int_A (2\pi)^{-\frac{p}{2}}|\mathbf{\Sigma}_p|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(\mathbf{y} - \mathbf{\beta}'\mathbf{x})'\mathbf{\Sigma}_p^{-1}(\mathbf{y} - \mathbf{\beta}'\mathbf{x})]d\mathbf{y} \qquad (3.2.4) \end{aligned}$$

A method for evaluating this integral will be discussed in the following section.

3.3 PARAMETER ESTIMATION

Given a sample of n student marks and matriculation profile results and adopting the notation

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1' \\ \mathbf{y}_2' \\ \vdots \\ \mathbf{y}_n' \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} \qquad (3.3.1)$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \qquad (3.3.2)$$

and

$$\mathbf{E} = \begin{pmatrix} \mathbf{e}_1' \\ \mathbf{e}_2' \\ \vdots \\ \mathbf{e}_n' \end{pmatrix} = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{np} \end{pmatrix} \tag{3.3.3}$$

where $\mathbf{E}$ represents a $n \times p$ matrix of random error terms, the model assumptions listed in the previous section imply that an appropriate statistical model (Graybill, 1976) for the n observations can be given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \tag{3.3.4}$$

where

$$\mathbf{E} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_p \otimes \mathbf{I}_n)$$

Employing the method of maximum likelihood the following parameter estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_p$ can be obtained, namely

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{3.3.5}$$

and correcting for the bias

$$\mathbf{S} = \frac{(\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB})}{n - (k+1)} \tag{3.3.6}$$

By the invariance principle, a maximum likelihood estimator for $\pi(\mathbf{x})$ can then be given by

$$\hat{\pi}(\mathbf{x}) = \int_A (2\pi)^{-\frac{p}{2}} |\mathbf{S}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{B}'\mathbf{x})'\mathbf{S}^{-1}(\mathbf{y} - \mathbf{B}'\mathbf{x})\right] d\mathbf{y} \tag{3.3.7}$$

We need therefore only to develop a method for evaluating the p dimensional integral expression given in (3.3.7). Fresen and Fresen (1993) have suggested that one of the following methods be used for evaluating this integral, namely

i) a numerical quadrature method as described by Schervish, (1984);

ii) the use of a tetrachoric series as described by Harris *et al*, (1980) and

iii) a Monte Carlo simulation method which will now be developed.

In particular, the Monte Carlo simulation algorithm, to estimate $\pi(\mathbf{x})$, proceeds as follows:

a) For any given vector $\mathbf{x}$ generate a large number, m say, of pseudo-random vectors from the multivariate Gaussian distribution, $N(\mathbf{B}'\mathbf{x}, \mathbf{S})$. Denote these p dimensional vectors by $\mathbf{y}_1^*, \mathbf{y}_2^*, \ldots, \mathbf{y}_m^*$.

b) Estimate $\pi(\mathbf{x})$ by the proportion of pseudo-random vectors (generated in step a) that lie in the product set defined by an event A, that is

$$\hat{\pi}(\mathbf{x}) \approx m^{-1} \sum_{i=1}^{m} I_A(\mathbf{y}_i^*)$$

where $I_A(\cdot)$ represents the indicator function of set A. The accuracy of this approximation can be improved by generating a larger number of pseudo-random vectors.

3.4 SUBSET SELECTION

In order to determine which independent variables to exclude from the model, we will need to write the above multivariate regression model in the form of a seemingly unrelated regression model and then employ a backward elimination technique. In particular letting $\mathbf{y}_{[i]}$ to denote the i'th column vector of the matrix $\mathbf{Y}$ and $\boldsymbol{\beta}_{[i]}$ the i'th column vector of $\mathbf{B}$, then the above multivariate regression model (given in 3.3.4) can be written in the following seemingly unrelated regression form

$$\begin{pmatrix} \mathbf{y}_{[1]} \\ \mathbf{y}_{[2]} \\ \vdots \\ \mathbf{y}_{[p]} \end{pmatrix} = \begin{pmatrix} \mathbf{X} & 0 & \cdots & 0 \\ 0 & \mathbf{X} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \mathbf{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{[1]} \\ \boldsymbol{\beta}_{[2]} \\ \vdots \\ \boldsymbol{\beta}_{[p]} \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_p \end{pmatrix} \tag{3.4.1}$$

where

$$\mathcal{E}\{\mathbf{e}_i\, \mathbf{e}_j'\} = \sigma_{ij}\mathbf{I}_n$$

Variable selection can then be implemented by employing a suitable backward elimination procedure to successively eliminate variables in the above model structure whose parameter coefficients are not significantly different from zero. Let us assume that at a particular stage in our backward elimination procedure we have a resulting seemingly unrelated regression model structure that takes the form

$$\underset{\tilde{\mathbf{y}}}{\begin{pmatrix} \mathbf{y}_{[1]} \\ \mathbf{y}_{[2]} \\ \vdots \\ \mathbf{y}_{[p]} \end{pmatrix}} = \underset{\tilde{\mathbf{X}}}{\begin{pmatrix} \mathbf{X}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{X}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \mathbf{X}_p \end{pmatrix}} \underset{\tilde{\boldsymbol{\beta}}}{\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_p \end{pmatrix}} + \underset{\tilde{\mathbf{e}}}{\begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_p \end{pmatrix}} \tag{3.4.2}$$

where $\mathbf{X}_i$ denotes the component matrix of $\mathbf{X}$ that remains and $\boldsymbol{\beta}_i$ the parameter coefficient vector that remains after elimination of those parameter coefficients in $\boldsymbol{\beta}_{[i]}$ that are not significantly different from zero. Parameter estimation can then proceed by first employing an ordinary least squares estimation technique on each of the p regression model structures

$$\mathbf{y}_{[i]} = \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{e}_i \qquad\qquad i = 1, \ldots, p$$

to obtain the following consistent estimates for σ_{ij}, namely

$$\hat{\sigma}_{ij} = \frac{\hat{\mathbf{e}}_i'\hat{\mathbf{e}}_j}{n} = (\mathbf{y}_{[i]} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_i)'(\mathbf{y}_{[j]} - \mathbf{X}_j\hat{\boldsymbol{\beta}}_j)/n \tag{3.4.3}$$

where

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_{[i]}$$

and then substituting these elements into $\boldsymbol{\Sigma}_p$ to yield the following feasible generalized least squares estimator for $\tilde{\boldsymbol{\beta}}$, namely

$$\hat{\tilde{\boldsymbol{\beta}}} \triangleq \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\Sigma}}_p) = (\tilde{\mathbf{X}}'(\hat{\boldsymbol{\Sigma}}_p^{-1} \otimes \mathbf{I}_n)\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'(\hat{\boldsymbol{\Sigma}}_p^{-1} \otimes \mathbf{I}_n)\tilde{\mathbf{y}} \tag{3.4.4}$$

where

$$\hat{\boldsymbol{\Sigma}}_p = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1} & \hat{\sigma}_{p2} & & \hat{\sigma}_{pp} \end{pmatrix}$$

Since the variance-covariance matrix of the feasible generalized least squares estimator can be approximated by the following expression

$$\mathcal{V}\{\hat{\tilde{\boldsymbol{\beta}}}\} = [\tilde{\mathbf{X}}'(\hat{\boldsymbol{\Sigma}}_p^{-1} \otimes \mathbf{I}_n)\tilde{\mathbf{X}}]^{-1} \tag{3.4.5}$$

standard errors for the j'th element of $\hat{\tilde{\boldsymbol{\beta}}}$ can be obtained by taking the square root of the j'th diagonal element of (3.4.5) and thus a backward elimination model selection technique can proceed by calculating t-values [1] for each component of $\tilde{\boldsymbol{\beta}}$ with variables being excluded from the seemingly unrelated regression model, using a 10% level of significance. Given a matriculation profile matrix of the form

$$\tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{x}_1' & 0 & \cdots & 0 \\ 0 & \mathbf{x}_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \mathbf{x}_p' \end{pmatrix}$$

---

[1] the t-values corresponding to each $\tilde{\beta}_j$ is given by $\frac{\hat{\tilde{\beta}}_j}{\sigma_{\hat{\tilde{\beta}}_j}}$ where $\sigma_{\hat{\tilde{\beta}}_j}$ represents the standard error of $\hat{\tilde{\beta}}_j$.

for a particular student where $\mathbf{x}_i'$ denotes the set of matriculation results that relate to the retained parameter coefficients $\boldsymbol{\beta}_i$ in (3.4.2), then the probability of success can be defined by

$$\hat{\pi}(\tilde{\mathbf{x}}) = \int_A (2\pi)^{-p/2} |\hat{\boldsymbol{\Sigma}}_p|^{-1/2} \exp\left[\frac{1}{2}(\mathbf{y} - \tilde{\mathbf{x}}\hat{\hat{\boldsymbol{\beta}}})' \hat{\boldsymbol{\Sigma}}_p^{-1} (\mathbf{y} - \tilde{\mathbf{x}}\hat{\hat{\boldsymbol{\beta}}})\right] d\mathbf{y} \qquad (3.4.6)$$

with the above integral being solved using the method of Monte Carlo[2] described earlier in Section 3.3. Furthermore 95% lower and upper bounds for the confidence intervals of $\tilde{\boldsymbol{\beta}}$ can be given by $\hat{\hat{\boldsymbol{\beta}}} - 1.96\,\hat{\boldsymbol{\sigma}}$ and $\hat{\hat{\boldsymbol{\beta}}} + 1.96\,\hat{\boldsymbol{\sigma}}$ respectively, where $\hat{\boldsymbol{\sigma}}$ represents the vector of standard errors of $\hat{\hat{\boldsymbol{\beta}}}$ (that are obtained by taking the square root of the diagonal elements of (3.4.5)), with the 95% confidence intervals for $\pi(\tilde{\mathbf{x}})$ being obtained by substituting the respective confidence intervals for $\tilde{\boldsymbol{\beta}}$ in (3.4.6).

3.5 RESULTS

We will now apply the results of the previous sections on our three problems, defined in chapter one, that is predicting the probability of a student passing firstly DSMA1SX1 , secondly DSPH1SC1 and finally both DSMA1SX1 and DSPH1SC1 simultaneously. Now in order to determine the probability of a student, with a given matriculation profile $\mathbf{x}$, passing[3] a single university

[2] We replace $N(\mathbf{B}'\mathbf{x}, \mathbf{S})$ in the Monte Carlo algorithm by $N(\tilde{\mathbf{x}}\hat{\hat{\boldsymbol{\beta}}}, \hat{\boldsymbol{\Sigma}}_p)$.

[3] 48 is defined to be the pass mark as in chapters 1 and 2.

course, (3.2.4) reduces to

$$\begin{aligned}
\hat{\pi}(\mathbf{x}) &= \int_{48}^{\infty} (2\pi s^2)^{-(\frac{1}{2})} \exp\left[-\tfrac{1}{2}\left(\frac{y-\mathbf{x}'\hat{\boldsymbol{\beta}}}{s}\right)^2\right] dy \\
&= 1-\Phi\left(\frac{48-\mathbf{x}'\hat{\boldsymbol{\beta}}}{s}\right) \\
&= \Phi\left(\frac{\mathbf{x}'\hat{\boldsymbol{\beta}}-48}{s}\right) \qquad (3.5.1)
\end{aligned}$$

where Φ denotes the standard cumulative distribution function and

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

and

$$s^2 = \frac{(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})}{n-(k+1)}$$

Employing the following variables in our predictor matrix, namely English (X_1), Afrikaans (X_2), Mathematics (X_3), Biology (X_4), Physics (X_5), NED matriculation authority (d_1) and HOD matriculation authority (d_2) and applying a backward elimination procedure on our DSMA1SX1 data set, the variables X_2 and d_2 were excluded from the model using a 10% level of significance and thus the following results were obtained (Table 3.5.1) for the remaining variables, namely

Table 3.5.1

Parameter	$\hat{\beta}$	SE*1	t*2	LB*3	UB*4	s^2 *5
β_0	69.238409	1.8365899	37.699439	65.638693	72.838125	225.51048
β_1	−2.643295	1.0766819	−2.455038	−4.753592	−0.532999	
β_3	−7.028165	1.1673725	−6.020499	−9.316215	−4.740115	
β_4	−2.228880	1.1134729	−2.001737	−4.411287	−0.046473	
β_5	−4.225204	1.2028948	−3.512530	−6.582878	−1.86753	
β_7	9.254687	1.9125960	4.838809	5.5059984	13.003375	

From Table 3.5.1, the estimated probability that can be associated with a student passing DSMA1SX1 for a specified matriculation profile $\mathbf{x}_i' = (1\, x_{i1}\, x_{i3}\, x_{i4}\, x_{i5}\, d_{i2})$, can be given by

$$\hat{\pi}_i = \Phi\left(\frac{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_3 x_{i3} + \hat{\beta}_4 x_{i4} + \hat{\beta}_5 x_{i5} + \hat{\beta}_7 d_{i2} - 48}{s}\right)$$

where Φ denotes the cumulative standard normal distribution function. The negative parameter coefficients of $\hat{\beta}_1$, $\hat{\beta}_3$, $\hat{\beta}_4$ and $\hat{\beta}_5$ indicate that higher marks (symbols) in English, Mathematics, Biology and Physics, at matriculation level, are associated with a higher probability of passing. Also, since the magnitude of $\hat{\beta}_3$ is the largest when compared with $\hat{\beta}_1$, $\hat{\beta}_4$ and $\hat{\beta}_5$, we infer that a matriculant's Mathematics mark has the most influence on the probability associated with a student passing DSMA1SX1. Furthermore, the large positive value of $\hat{\beta}_7$ indicates that Natal Education Department students have a higher probability of passing DSMA1SX1 compared to students from the other matriculation authorities (see also Table 3.5.2). In Table 3.5.2 we list a

*1 standard error of the parameter estimate
*2 t-values associated with the parameter estimates
*3 lower bound for the 95% confidence interval of the parameter estimates
*4 upper bound for the 95% confidence interval of the parameter estimates
*5 mean square error

few estimated probabilities associated with passing DSMA1SX1 for specific values of X_1, X_3, X_4, X_5 and d_2.

Table 3.5.2

X_1	X_3	X_4	X_5	d_2	$\hat{\pi}$	LB($\hat{\pi}$)*6	UB($\hat{\pi}$)*7
0	0	0	0	0	0.9213617	0.8799188	0.9509362
1	0	0	0	0	0.8921921	0.8045631	0.9472233
0	1	0	0	0	0.8279963	0.7102802	0.9096088
0	0	1	0	0	0.8972196	0.8107946	0.950621
0	0	0	1	0	0.871378	0.7692015	0.9369469
0	0	0	0	1	0.9788507	0.9383698	0.9941308
1	0	0	0	1	0.9681698	0.889653	0.9935118
0	1	0	0	1	0.9409216	0.8214368	0.986247
0	0	1	0	1	0.970092	0.8938891	0.9940789
0	0	0	1	1	0.9598726	0.8649583	0.9917022

Turning our attention to the DSPH1SC1 data, and implementing a backward elimination routine (using a probability value of 0.1) the variables X_1, X_2, X_3 and d_1 were excluded from the model with the following results being obtained for the remaining variables, namely

Table 3.5.3

Parameter	$\hat{\beta}$	SE	t	LB	UB	s^2
β_0	67.582542	1.3392143	50.464322	64.957682	70.207402	149.14551
β_4	−4.392492	1.1749480	−3.738457	−6.695390	−2.089594	
β_5	−9.162855	1.0722381	−8.545541	−11.264440	−7.061268	
β_7	13.104750	1.9554894	6.701519	9.2719902	16.937509	

Table 3.5.3 suggests that a matriculant's Physics mark influences the probability associated with a student passing DSPH1SC1 more than his or her Biology mark since $\hat{\beta}_5$ is more negative than $\hat{\beta}_4$. Furthermore the large positive value of $\hat{\beta}_7$ implies that students from the Natal Education Department

*6 lower bound for the 95% confidence interval of the success probability
*7 upper bound for the 95% confidence interval of the success probability

have a greater probability of passing DSPH1SC1 compared to students from the other matriculation authorities. In Table 3.5.4 we give a few estimated probabilities for passing DSPH1SC1 given specific values for X_4, X_5 and d_2.

Table 3.5.4

X_4	X_5	d_2	$\hat{\pi}$	LB($\hat{\pi}$)	UB($\hat{\pi}$)
0	0	0	0.9455858	0.9175151	0.9654995
1	0	0	0.8932153	0.7996326	0.9502529
0	1	0	0.8032251	0.679457	0.8925519
0	0	1	0.9962807	0.984134	0.9993254
1	0	1	0.9897446	0.9451485	0.9987942
0	1	1	0.9729638	0.889788	0.9956943

Finally turning our attention to the third problem that we have been considering in this thesis, namely that of determining the probability associated with a student passing both DSMA1SX1 and DSPH1SC1 and applying the seemingly unrelated regression model to this data set, the following results[4] were obtained, namely

Table 3.5.5

Parameter[*8]	$\hat{\hat{\beta}}$	SE	t	Prob
B01	69.647416	2.01102	34.63	0.0001
B11	−2.917197	1.05425	−2.77	0.0062
B31	−5.949945	1.25738	−4.73	0.0001
B41	−2.771966	1.62782	−1.70	0.0903
B51	−5.423292	1.57495	−3.44	0.0007
B71	11.317970	2.61120	4.33	0.0001
B02	67.594627	1.39130	48.58	0.0001
B42	−4.548053	1.25355	−3.63	0.0004
B52	−8.916156	1.16054	−7.68	0.0001
B72	13.103354	2.07635	6.31	0.0001

Letting $\mathbf{y}_{[1]}$ and $\mathbf{y}_{[2]}$ denote the DSMA1SX1 and DSPH1SC1 marks, respectively; the large positive values associated with the parameters B_{71} and B_{72}

[4]Note that a 10% level of significance was used. Also these results were obtained using the model procedure in SAS.

[*8] B_{i1} represents the regression coefficients associated with the DSMA1SX1 course and B_{i2} those regression coefficients associated with the DSPH1SC1 course.

imply that students writing matriculation papers set by the Natal Education Department perform better in DSMA1SX1 and DSPH1SC1 at university than students from the other matriculation authorities (see also Table 3.5.6). Based on the above model structure, with a covariance matrix of residuals that was given by

$$\hat{\Sigma}_p = \begin{pmatrix} 240.3949 & 118.5459 \\ 118.5459 & 154.7962 \end{pmatrix}$$

the Monte Carlo algorithm (outlined in section 3.3) was used to obtain the probability estimates given in Table 3.5.6.[5]

Table 3.5.6

$\tilde{x}$	$\hat{\pi}(\tilde{x})$	LB($\hat{\pi}$)	UB($\hat{\pi}$)
$\begin{pmatrix} 1&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&1 \end{pmatrix}$	0.9812	0.9274	0.996
$\begin{pmatrix} 1&0&1&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&1 \end{pmatrix}$	0.959	0.8436	0.9931
$\begin{pmatrix} 1&0&0&1&0&1&0&0&0&0 \\ 0&0&0&0&0&0&1&1&0&1 \end{pmatrix}$	0.9669	0.845	0.9958
$\begin{pmatrix} 1&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&1&1 \end{pmatrix}$	0.9448	0.7848	0.9916
$\begin{pmatrix} 1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0 \end{pmatrix}$	0.889	0.8312	0.9292
$\begin{pmatrix} 1&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0 \end{pmatrix}$	0.8163	0.7019	0.9012
$\begin{pmatrix} 1&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&1&0&0 \end{pmatrix}$	0.8231	0.6746	0.916
$\begin{pmatrix} 1&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&1&0 \end{pmatrix}$	0.7286	0.5734	0.8587

[5]A program for obtaining these probabilities using the Monte Carlo algorithm is given in Appendix D3.2 and is based on generating 100 000 2×1 dimensional vectors $\mathbf{y}^*$. As an example the first row in $\tilde{x} = \begin{pmatrix} 1&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&1 \end{pmatrix}$ refers to an NED student with A's in English, Mathematics, Biology and Physics.

CHAPTER FOUR

A BAYESIAN APPROACH

4.1 THE GIBBS SAMPLER

In this chapter we will consider a Bayesian approach for obtaining parameter estimates in the generalized linear model as opposed to the frequentist approaches that have been employed previously. In particular, consider the generalized linear model with logit link function discussed in chapter two. Suppose that we now specify a multivariate normal prior distribution for $\boldsymbol{\beta}$ with a mean vector, $\boldsymbol{\mu}_0$, and variance-covariance matrix, $\boldsymbol{\Sigma}_0$, then applying Bayes theorem (Zellner, 1970; Press, 1989) for a given set of observation $\mathbf{y} = \{y_1, \ldots, y_n\}$ we obtain the following expression for the posterior density of $\boldsymbol{\beta}$, namely[1]

$$P(\boldsymbol{\beta}|\mathbf{y}) = \frac{P(\boldsymbol{\beta})P(\mathbf{y}|\boldsymbol{\beta})}{\int \cdots \int P(\boldsymbol{\beta})P(\mathbf{y}|\boldsymbol{\beta})d\boldsymbol{\beta}}$$

$$= \frac{\exp\left[-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)'\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0) + \sum_{i=1}^{n} y_i\mathbf{x}_i'\boldsymbol{\beta} + \sum_{i=1}^{n}\log\left(1+e^{\mathbf{x}_i'\boldsymbol{\beta}}\right)^{-1}\right]}{\int \cdots \int \left\{\exp\left[-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)'\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0) + \sum_{i=1}^{n} y_i\mathbf{x}_i'\boldsymbol{\beta} + \sum_{i=1}^{n}\log\left(1+e^{\mathbf{x}_i'\boldsymbol{\beta}}\right)^{-1}\right]\right\} d\boldsymbol{\beta}} \tag{4.1.1}$$

[1]since the likelihood function of this model is given by

$$P(\mathbf{y}|\boldsymbol{\beta}) = \exp\left[\sum_{i=1}^{n} y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \sum_{i=1}^{n}\log(1-\pi_i)\right] \qquad y_i = 0, 1$$

where

$$\pi_i = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i'\boldsymbol{\beta}}}$$

Now in order to obtain an approximation of the above posterior density, several approaches can be adopted. One method, involves employing a Taylor series expansion of the log-posterior density about its posterior maximizer $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ (Knuiman and Speed, 1988) yielding

$$\log P(\boldsymbol{\beta}|\mathbf{y}) = \log P(\hat{\boldsymbol{\beta}}|\mathbf{y}) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\mathbf{u}(\hat{\boldsymbol{\beta}}|\mathbf{y}) - \frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}|\mathbf{y})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) + r(\boldsymbol{\beta}|\mathbf{y})$$

where $r(\boldsymbol{\beta}|\mathbf{y})$ represents a remainder term which we will assume is of negligible order,

$$\mathbf{u}(\hat{\boldsymbol{\beta}}|\mathbf{y}) = \left[\frac{\partial}{\partial\boldsymbol{\beta}}\log P(\boldsymbol{\beta}|\mathbf{y})\right]\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$$

denotes the score vector evaluated at the point $\hat{\boldsymbol{\beta}}$ and

$$\mathcal{I}(\hat{\boldsymbol{\beta}}|\mathbf{y}) = -\left[\frac{\partial^2 \log P(\boldsymbol{\beta}|\mathbf{y})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right]\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$$

the Information matrix evaluated at the point $\hat{\boldsymbol{\beta}}$, then because $\mathbf{u}(\hat{\boldsymbol{\beta}}|\mathbf{y}) = 0$, the posterior density of $\boldsymbol{\beta}$ can now be approximated by a multivariate normal density with mean vector $\hat{\boldsymbol{\beta}}$ and variance-covariance matrix $\boldsymbol{\Sigma} = \mathcal{I}^{-1}(\hat{\boldsymbol{\beta}}|\mathbf{y})$.

An alternative approach is to make use of Monte Carlo Markov Chain methods to implement what has become known as the Gibbs sampler (Geman and Geman, 1984) to generate a set of observations $\{\beta^{(j)}, j = 1 \cdots m\}$ from the posterior distribution[2] $P(\boldsymbol{\beta}|\mathbf{y})$. A sample based estimator for the posterior

[2]Functional forms for the $(p+1)$ univariate full conditional densities $P(\beta_i|\boldsymbol{\beta}_{j\neq i}\,\mathbf{y})$ can be easily obtained, at least up to a proportionality constant, from the given joint posterior density $P(\boldsymbol{\beta}|\mathbf{y})$ by regarding the joint density as a as a function of β_i for fixed values of the other parameters $\boldsymbol{\beta}_{j\neq i}$. In particular, rewriting the joint posterior density in the form

$$P(\beta_0,\ldots,\beta_p|\mathbf{y}) = P(\beta_0|\beta_1,\ldots,\beta_p,\mathbf{y}) \cdot P(\beta_1|\beta_2,\ldots,\beta_p,\mathbf{y}) \cdots P(\beta_p|\mathbf{y})$$

and fixing the values for $\beta_1,\ldots,\beta_p$ implies that

$$P(\beta_0,\ldots,\beta_p|\mathbf{y}) \propto P(\beta_0|\beta_1,\ldots,\beta_p,\mathbf{y})$$

and thus we may obtain an observation from $P(\beta_i|\boldsymbol{\beta}_{j\neq i}\,\mathbf{y})$ by employing a suitable rejection algorithm on the density function given in (4.1.1) (see Appendix B).

distribution associated with the i'th component of $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ can then be given by evaluating (Gelfand and Smith, 1990)

$$\hat{P}(\beta_i|\mathbf{y}) = \frac{1}{m}\sum_{j=1}^{m} P(\beta_i|\beta_0^{(j)}, \ldots, \beta_{i-1}^{(j)}, \beta_{i+1}^{(j)}, \ldots, \beta_p^{(j)}, \mathbf{y})$$

Alternatively a Bayesian estimator for β_i can be developed by evaluating

$$\hat{\beta}_i = \frac{1}{m}\sum_{j=1}^{m} \beta_i^{(j)}$$

Turning our attention to the actual implementation of the Gibbs sampler and letting $P(\beta_i|\beta_0, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_p, \mathbf{y})$ denote a full conditional density specification for β_i, the Gibbs sampling nature proceeds from an arbitrary set of starting values $(\beta_0^{(0)}, \beta_1^{(0)}, \ldots, \beta_p^{(0)})$ to draw a value $\beta_0^{(1)}$ from the conditional density

$$P(\beta_0|\beta_1^{(0)}, \beta_2^{(0)}, \ldots, \beta_p^{(0)}, \mathbf{y})$$

and then a value $\beta_1^{(1)}$ from the conditional density

$$P(\beta_1|\beta_0^{(1)}, \beta_2^{(0)}, \ldots, \beta_p^{(0)}, \mathbf{y})$$

Proceeding in this manner until a value $\beta_p^{(1)}$ has been generated from the conditional density

$$P(\beta_p|\beta_0^{(1)}, \beta_1^{(1)}, \ldots, \beta_{p-1}^{(1)}, \mathbf{y})$$

the algorithm then returns the first conditional density specification where a value $\beta_0^{(2)}$ is generated from

$$P(\beta_0|\beta_1^{(1)}, \beta_2^{(1)}, \ldots, \beta_p^{(1)}, \mathbf{y})$$

Continuing with this procedure, and letting

$$\boldsymbol{\beta}^{(t)} = (\beta_0^{(t)}, \beta_1^{(t)}, \ldots, \beta_p^{(t)})$$

denote the set of values that have been generated on completion of the t'th iteration of the Gibbs sampler, Geman and Geman (1984) have shown that under fairly general conditions the distribution of $\beta_i^{(t)}$ converge to $P(\beta_i|\mathbf{y})$ for t large enough, that is

$$\beta_i^{(t)} \xrightarrow{d} \beta_i \sim P(\beta_i|\mathbf{y})$$

Letting L denote the number of iterations that are necessary for the convergence of $\beta_i^{(t)}$ so as to represent a sample from $P(\beta_i|\mathbf{y})$, N independent replications of this entire process can then be implemented to obtain a collection of observations $\{\beta_{0j}^{(L)}, \ldots, \beta_{pj}^{(L)}, j = 1 \cdots N\}$ which can be used to estimate $P(\beta_i|\mathbf{y})$ as follows

$$\hat{P}(\beta_i|\mathbf{y}) = \frac{1}{N}\sum_{j=1}^{N} P(\beta_i|\boldsymbol{\beta}_{rj}^{(L)}\,\mathbf{y}\ r \neq i) \qquad r = 0, 1, \ldots, p$$

Furthermore an estimate of β_i can be obtained by averaging over the collection of N observations to yield

$$\hat{\beta}_i = \frac{1}{N}\sum_{j=1}^{N} \beta_{ij}^{(L)} \qquad i = 0, 1, \ldots, p$$

4.2 RESULTS

Applying the Gibbs Sampling technique with the method of Rejection Sampling outlined in Appendix B on our DSMA1SX1, DSPH1SC1 and our combined DSMA1SX1 and DSPH1SC1 data, we obtained the following results,[3] namely

[3] We used $L = 10$ and $N = 1000$. A program implementing the Gibbs Sampler is given in Appendix D4.

RESULTS OBTAINED USING THE DSMA1SX1 COURSE

Variable = β_0 (parameter associated with the intercept term)

Moments

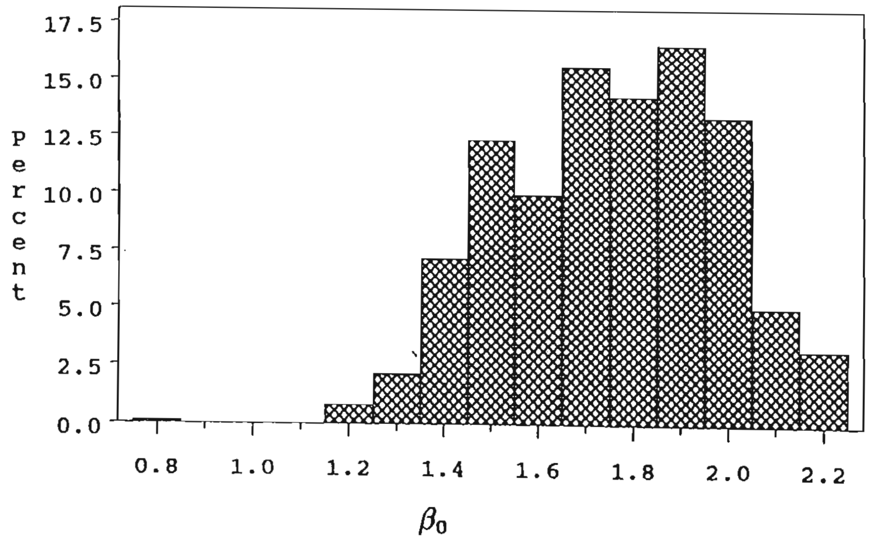
N	1000	Sum Wgts	1000
Mean	1.754678	Sum	1754.678
Std Dev	0.225494	Variance	0.050848
Skewness	-0.21164	Kurtosis	-0.41659
USS	3129.693	CSS	50.79688
CV	12.85104	Std Mean	0.007131
T:Mean=0	246.0717	Pr>|T|	0.0000
Sgn Rank	250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	2.243731	99%	2.206173
75% Q3	1.928956	95%	2.102733
50% Med	1.765661	90%	2.034419
25% Q1	1.584861	10%	1.448443
0% Min	0.767955	5%	1.388413
		1%	1.257246
Range	1.475776		
Q3-Q1	0.344095		
Mode	1.248647		

Extremes

Lowest	Obs	Highest	Obs
0.767955(	876)	2.218708(	617)
1.19778(	829)	2.230577(	160)
1.217521(	636)	2.235834(	124)
1.236478(	523)	2.238787(	28)
1.237835(	672)	2.243731(	64)



Variable = β_3 (parameter associated with the matriculation Mathematics symbol)

Moments

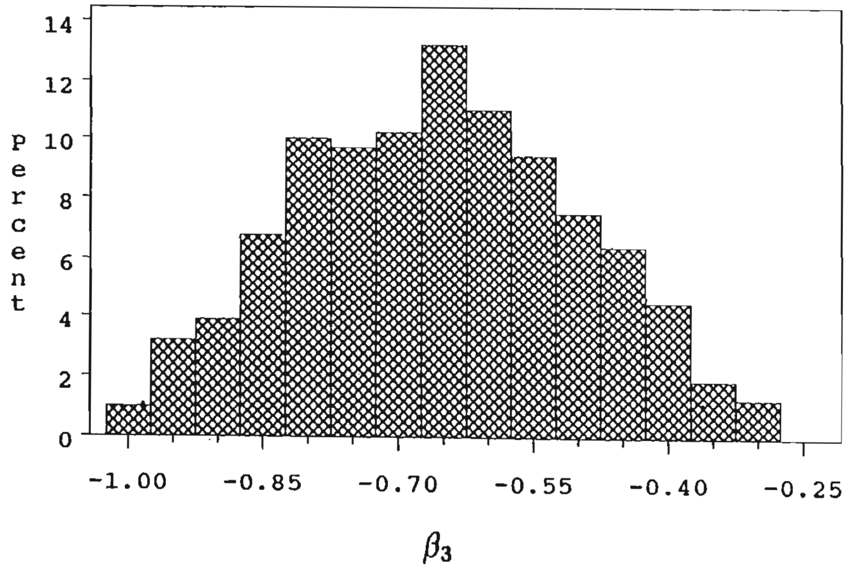
N	1000	Sum Wgts	1000
Mean	-0.65577	Sum	-655.769
Std Dev	0.155281	Variance	0.024112
Skewness	0.038356	Kurtosis	-0.60687
USS	454.121	CSS	24.08796
CV	-23.6792	Std Mean	0.00491
T:Mean=0	-133.547	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.29607	99%	-0.31431
75% Q3	-0.54225	95%	-0.39782
50% Med	-0.65568	90%	-0.44716
25% Q1	-0.77468	10%	-0.85978
0% Min	-0.99569	5%	-0.91408
		1%	-0.97593
Range	0.699618		
Q3-Q1	0.232424		
Mode	-0.96926		

Extremes

Lowest	Obs	Highest	Obs
-0.99569(	664)	-0.30546(	462)
-0.99259(	70)	-0.29986(	300)
-0.9905(	530)	-0.29971(	92)
-0.98967(	906)	-0.29814(	988)
-0.98803(	118)	-0.29607(	662)



Variable = β_4 (parameter associated with the matriculation Biology symbol)

Moments

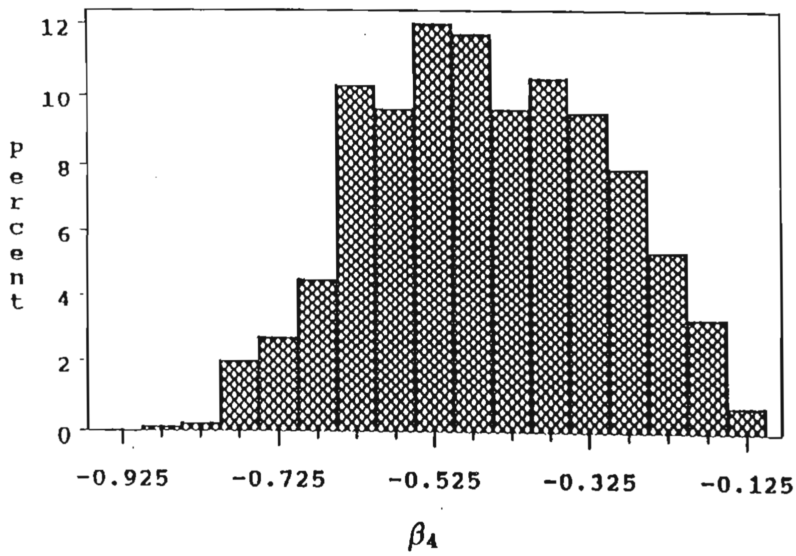
N	1000	Sum Wgts	1000
Mean	-0.45776	Sum	-457.759
Std Dev	0.149837	Variance	0.022451
Skewness	0.0038	Kurtosis	-0.70135
USS	231.972	CSS	22.42882
CV	-32.7328	Std Mean	0.004738
T:Mean=0	-96.6087	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.11683	99%	-0.15274
75% Q3	-0.34124	95%	-0.21251
50% Med	-0.46387	90%	-0.25354
25% Q1	-0.57312	10%	-0.64441
0% Min	-0.87909	5%	-0.69982
		1%	-0.77097
Range	0.762265		
Q3-Q1	0.231882		
Mode	-0.77888		

Extremes

Lowest	Obs	Highest	Obs
-0.87909(	563)	-0.13809(	647)
-0.82977(	135)	-0.1235(	927)
-0.82619(	391)	-0.12015(	232)
-0.78284(	393)	-0.12015(	453)
-0.7827(	986)	-0.11683(	815)



Variable = β_5 (parameter associated with the matriculation Physics symbol)

Moments

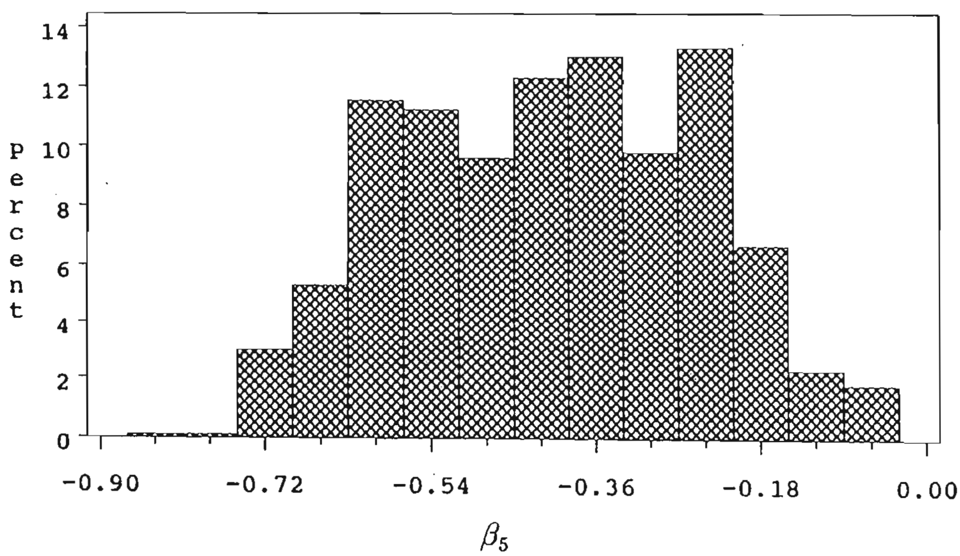
N	1000	Sum Wgts	1000
Mean	-0.40883	Sum	-408.828
Std Dev	0.16171	Variance	0.02615
Skewness	-0.0131	Kurtosis	-0.82018
USS	193.2645	CSS	26.12383
CV	-39.5544	Std Mean	0.005114
T:Mean=0	-79.9476	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.03422	99%	-0.06422
75% Q3	-0.27589	95%	-0.16043
50% Med	-0.40157	90%	-0.20035
25% Q1	-0.54452	10%	-0.62003
0% Min	-0.85399	5%	-0.67071
		1%	-0.7296
Range	0.819774		
Q3-Q1	0.268629		
Mode	-0.71671		

Extremes

Lowest	Obs	Highest	Obs
-0.85399(	540)	-0.04108(	845)
-0.75634(	88)	-0.03931(	623)
-0.74908(	590)	-0.03729(	402)
-0.74713(	543)	-0.03598(	611)
-0.7433(	410)	-0.03422(	274)



Variable = β_7 (parameter associated with the NED matriculation authority)

Moments

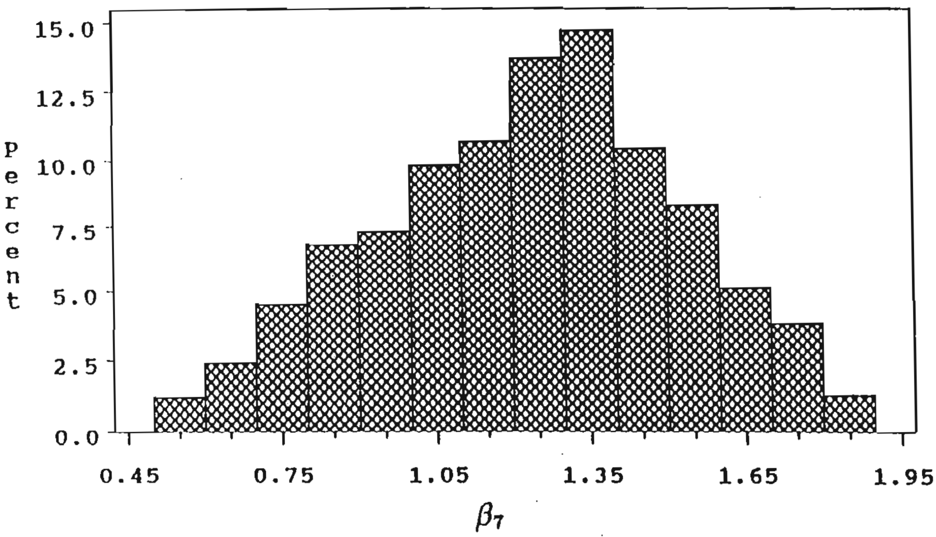
N	1000	Sum Wgts	1000
Mean	1.231384	Sum	1231.384
Std Dev	0.285401	Variance	0.081454
Skewness	-0.16053	Kurtosis	-0.53547
USS	1597.678	CSS	81.37211
CV	23.17724	Std Mean	0.009025
T:Mean=0	136.4389	Pr>|T|	0.0000
Sgn Rank	250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	1.850272	99%	1.814409
75% Q3	1.434285	95%	1.702133
50% Med	1.254945	90%	1.602976
25% Q1	1.023714	10%	0.832471
0% Min	0.523482	5%	0.748677
		1%	0.58753
Range	1.32679		
Q3-Q1	0.41057		
Mode	0.523482		

Extremes

Lowest	Obs	Highest	Obs
0.523482(	707)	1.834431(	230)
0.523482(	193)	1.834431(	451)
0.531307(	292)	1.850144(	582)
0.54079(	63)	1.850272(	265)
0.545394(	899)	1.850272(	486)



RESULTS OBTAINED USING THE DSPH1SC1 COURSE :

Variable = β_0 (parameter associated with the intercept term)

Moments

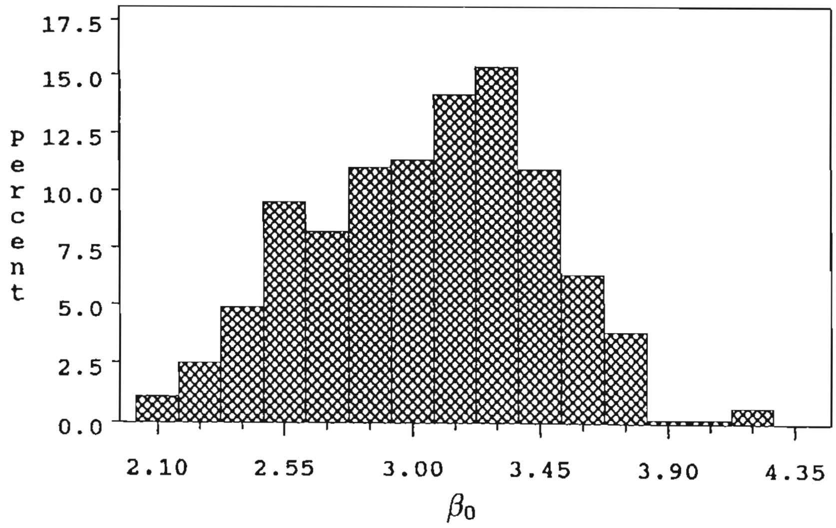
N	1000	Sum Wgts	1000
Mean	3.051592	Sum	3051.592
Std Dev	0.402061	Variance	0.161653
Skewness	-0.09305	Kurtosis	-0.47534
USS	9473.707	CSS	161.4915
CV	13.17545	Std Mean	0.012714
T:Mean=0	240.0128	Pr>|T|	0.0000
Sgn Rank	250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	4.230289	99%	3.851138
75% Q3	3.346965	95%	3.669392
50% Med	3.092712	90%	3.550887
25% Q1	2.753876	10%	2.486824
0% Min	2.095327	5%	2.389352
		1%	2.171735
Range	2.134962		
Q3-Q1	0.593089		
Mode	2.095327		

Extremes

Lowest	Obs	Highest	Obs
2.095327(	732)	4.159806(	360)
2.102426(	637)	4.165588(	965)
2.110648(	339)	4.19506(	10)
2.120716(	259)	4.197028(	627)
2.137425(	722)	4.230289(	123)



Variable = β_3 (parameter associated with the matriculation Mathematics symbol)

Moments

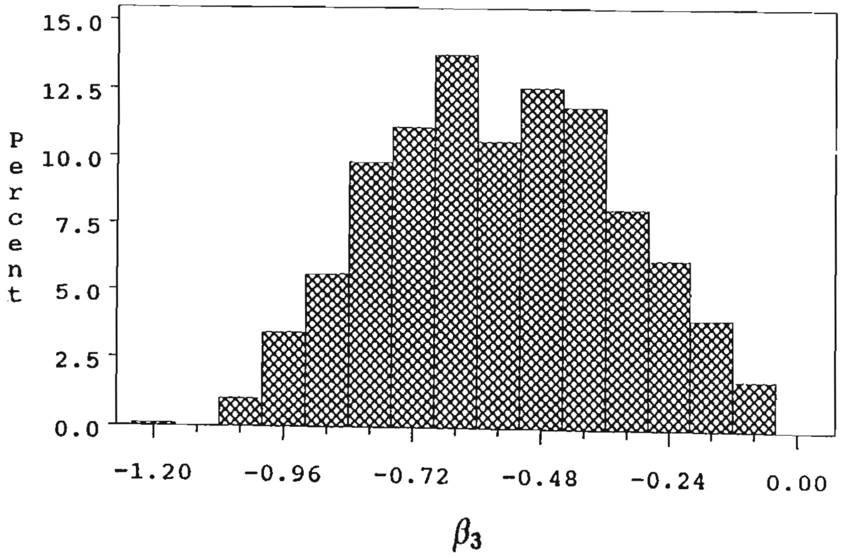
N	1000	Sum Wgts	1000
Mean	-0.55717	Sum	-557.173
Std Dev	0.216912	Variance	0.047051
Skewness	0.036099	Kurtosis	-0.68264
USS	357.4454	CSS	47.00365
CV	-38.9308	Std Mean	0.006859
T:Mean=0	-81.2282	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.0711	99%	-0.09464
75% Q3	-0.40233	95%	-0.19363
50% Med	-0.55774	90%	-0.25565
25% Q1	-0.72384	10%	-0.84167
0% Min	-1.19806	5%	-0.90428
		1%	-1.00352
Range	1.126953		
Q3-Q1	0.321512		
Mode	-1.19806		

Extremes

Lowest	Obs	Highest	Obs
-1.19806(	490)	-0.08462(	806)
-1.03416(	854)	-0.08445(	568)
-1.0232(	643)	-0.0804(	688)
-1.02306(	90)	-0.07627(	722)
-1.01971(	16)	-0.0711(	456)



Variable = β_4 (parameter associated with the matriculation Biology symbol)

Moments

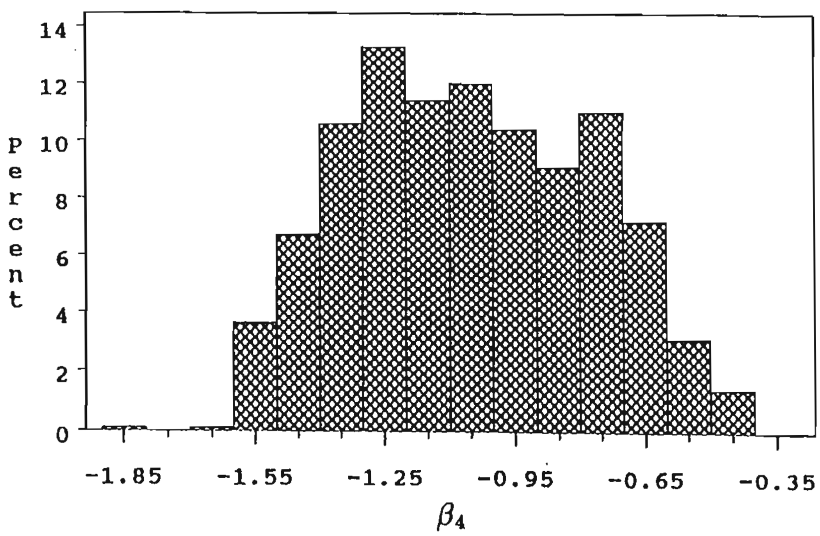
N	1000	Sum Wgts	1000
Mean	-1.05139	Sum	-1051.39
Std Dev	0.274389	Variance	0.075289
Skewness	0.090284	Kurtosis	-0.89214
USS	1180.631	CSS	75.21404
CV	-26.0978	Std Mean	0.008677
T:Mean=0	-121.17	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.42004	99%	-0.48433
75% Q3	-0.82354	95%	-0.60795
50% Med	-1.06017	90%	-0.68162
25% Q1	-1.28141	10%	-1.4102
0% Min	-1.85883	5%	-1.4782
		1%	-1.56341
Range	1.43879		
Q3-Q1	0.457873		
Mode	-1.85883		

Extremes

Lowest	Obs	Highest	Obs
-1.85883(	882)	-0.45543(	384)
-1.68065(	592)	-0.44474(	618)
-1.59517(	836)	-0.44364(	736)
-1.58283(	849)	-0.42574(	197)
-1.58014(	711)	-0.42004(	453)



Variable = β_5 (parameter associated with the matriculation Physics symbol)

Moments

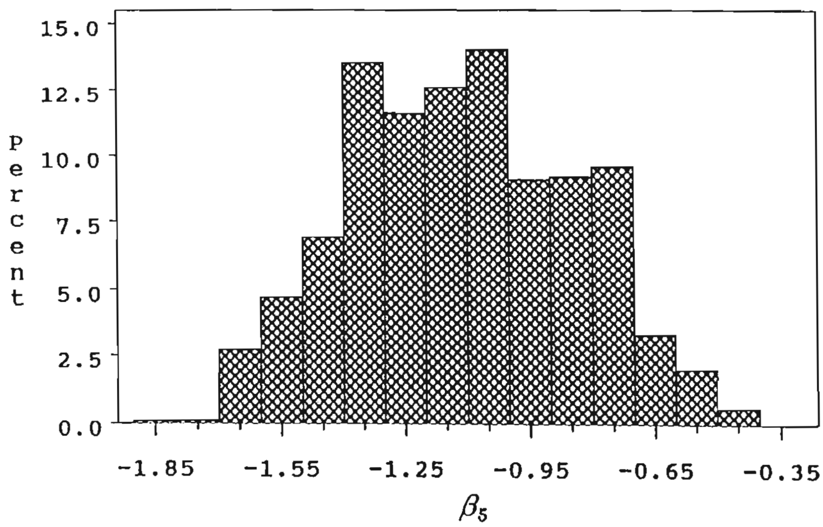
N	1000	Sum Wgts	1000
Mean	-1.11271	Sum	-1112.71
Std Dev	0.268217	Variance	0.071941
Skewness	0.072181	Kurtosis	-0.66816
USS	1309.987	CSS	71.86863
CV	-24.1049	Std Mean	0.008482
T:Mean=0	-131.188	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.48296	99%	-0.53376
75% Q3	-0.90125	95%	-0.68086
50% Med	-1.11965	90%	-0.75533
25% Q1	-1.32253	10%	-1.44423
0% Min	-1.85942	5%	-1.56109
		1%	-1.64707
Range	1.376459		
Q3-Q1	0.421284		
Mode	-1.19156		

Extremes

Lowest	Obs	Highest	Obs
-1.85942(	332)	-0.49356(	390)
-1.77563(	369)	-0.49348(	792)
-1.6775(	204)	-0.48817(	546)
-1.66499(	491)	-0.48427(	485)
-1.66395(	616)	-0.48296(	389)



Variable = β_7 (parameter associated with the NED matriculation authority)

Moments

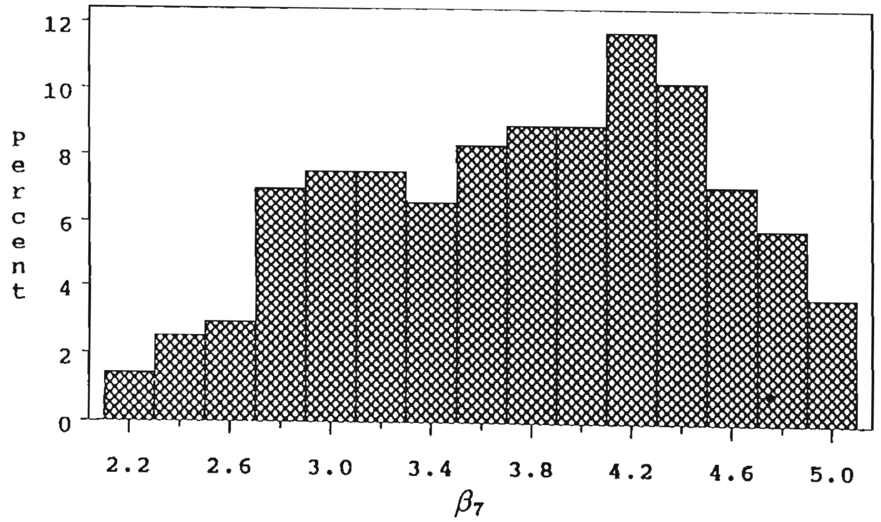
N	1000	Sum Wgts	1000
Mean	3.775604	Sum	3775.604
Std Dev	0.704626	Variance	0.496497
Skewness	-0.21376	Kurtosis	-0.87435
USS	14751.18	CSS	496.0008
CV	18.6626	Std Mean	0.022282
T:Mean=0	169.4447	Pr>|T|	0.0000
Sgn Rank	250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	5.091719	99%	5.038965
75% Q3	4.323396	95%	4.847246
50% Med	3.809289	90%	4.678922
25% Q1	3.204802	10%	2.795438
0% Min	2.14336	5%	2.597703
		1%	2.263675
Range	2.948359		
Q3-Q1	1.118594		
Mode	2.14336		

Extremes

Lowest	Obs	Highest	Obs
2.14336(	618)	5.074036(	975)
2.174244(	308)	5.082588(	27)
2.174814(	388)	5.08455(	392)
2.19366(	740)	5.085465(	302)
2.197419(	520)	5.091719(	646)



RESULTS OBTAINED FOR THE COMBINED DSMA1SX1 AND DSPH1SC1 DATA :

Variable = β_0 (parameter associated with the intercept term)

Moments

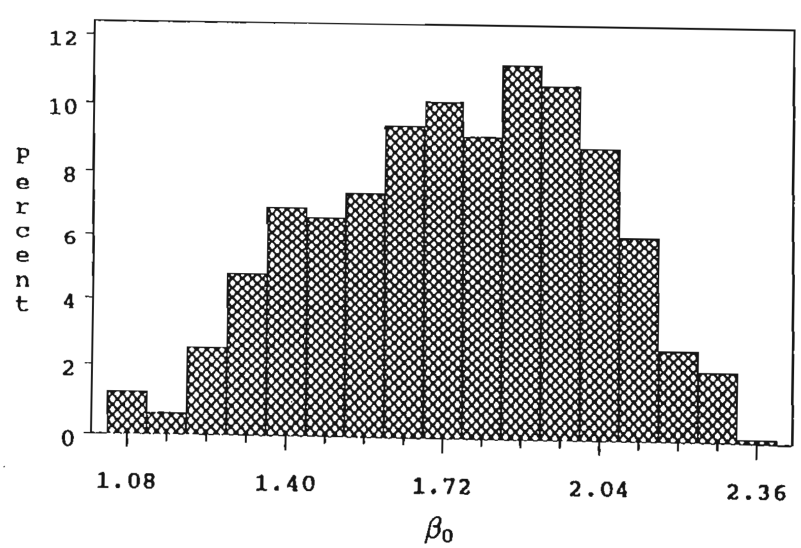
N	1000	Sum Wgts	1000
Mean	1.748324	Sum	1748.324
Std Dev	0.271513	Variance	0.07372
Skewness	-0.20394	Kurtosis	-0.67963
USS	3130.282	CSS	73.64584
CV	15.52993	Std Mean	0.008586
T:Mean=0	203.6247	Pr>|T|	0.0000
Sgn Rank	250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	2.365231	99%	2.281661
75% Q3	1.958854	95%	2.15688
50% Med	1.763682	90%	2.095101
25% Q1	1.547375	10%	1.369156
0% Min	1.078589	5%	1.292157
		1%	1.1155
Range	1.286642		
Q3-Q1	0.411479		
Mode	1.078589		

Extremes

Lowest	Obs	Highest	Obs
1.078589(	137)	2.298518(	95)
1.080434(	160)	2.299173(	178)
1.080603(	843)	2.309204(	992)
1.08488(	541)	2.310389(	282)
1.095208(	92)	2.365231(	655)



Variable = β_3 (parameter associated with the matriculation Mathematics symbol)

Moments

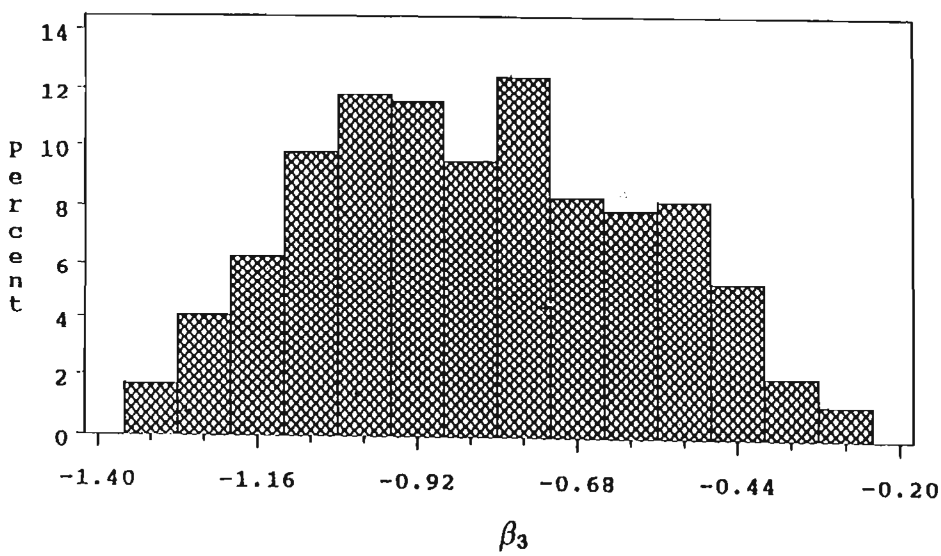
N	1000	Sum Wgts	1000
Mean	-0.82959	Sum	-829.588
Std Dev	0.24181	Variance	0.058472
Skewness	0.116416	Kurtosis	-0.80627
USS	746.6309	CSS	58.41382
CV	-29.1482	Std Mean	0.007647
T:Mean=0	-108.489	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.28491	99%	-0.31498
75% Q3	-0.64199	95%	-0.43362
50% Med	-0.8402	90%	-0.49678
25% Q1	-1.01903	10%	-1.13802
0% Min	-1.33205	5%	-1.21927
		1%	-1.30565
Range	1.047139		
Q3-Q1	0.377039		
Mode	-0.93924		

Extremes

Lowest	Obs	Highest	Obs
-1.33205(	849)	-0.29636(	672)
-1.32962(	324)	-0.2908(	264)
-1.32862(	545)	-0.28987(	179)
-1.32236(	608)	-0.28893(	446)
-1.32071(	575)	-0.28491(	313)



Variable = β_4 (parameter associated with the matriculation Biology symbol)

Moments

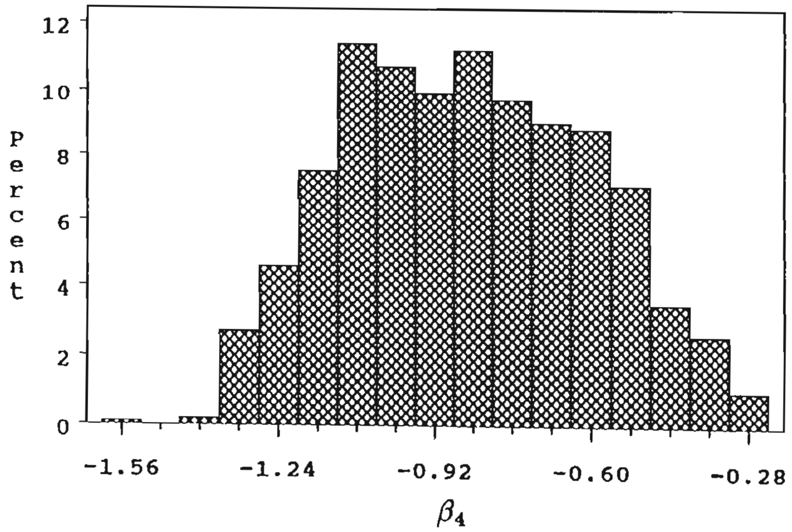
N	1000	Sum Wgts	1000
Mean	-0.85044	Sum	-850.44
Std Dev	0.247398	Variance	0.061206
Skewness	0.113575	Kurtosis	-0.75207
USS	784.3934	CSS	61.14477
CV	-29.0906	Std Mean	0.007823
T:Mean=0	-108.704	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.30056	99%	-0.3192
75% Q3	-0.65493	95%	-0.43123
50% Med	-0.86143	90%	-0.51624
25% Q1	-1.05005	10%	-1.17084
0% Min	-1.56938	5%	-1.23256
		1%	-1.32259
Range	1.268823		
Q3-Q1	0.395122		
Mode	-1.56938		

Extremes

Lowest	Obs	Highest	Obs
-1.56938(	918)	-0.30264(	339)
-1.43914(	692)	-0.30256(	238)
-1.37399(	670)	-0.30138(	93)
-1.34793(	102)	-0.30058(	823)
-1.3433(	972)	-0.30056(	997)



Variable = β_5 (parameter associated with the matriculation Physics symbol)

Moments

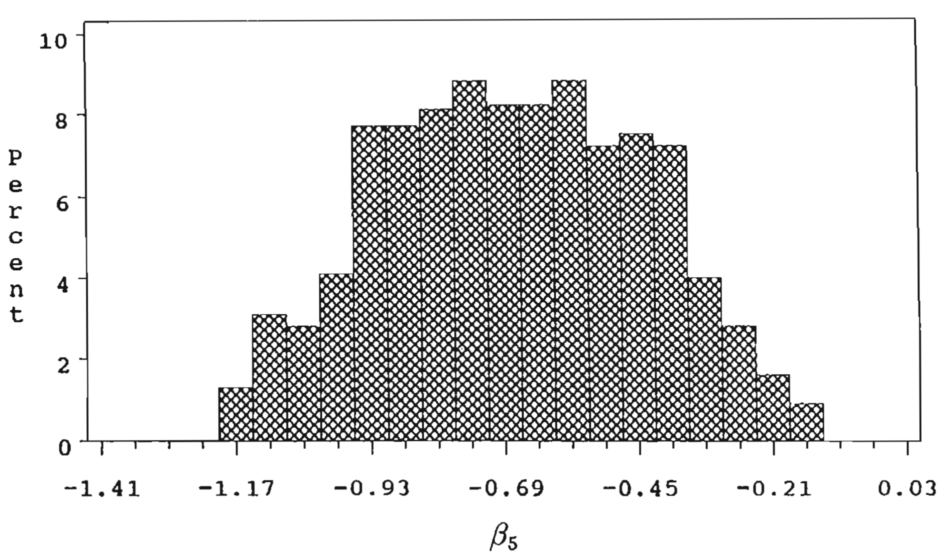
N	1000	Sum Wgts	1000
Mean	-0.67188	Sum	-671.875
Std Dev	0.234368	Variance	0.054928
Skewness	-0.00165	Kurtosis	-0.7576
USS	506.2895	CSS	54.8733
CV	-34.8826	Std Mean	0.007411
T:Mean=0	-90.6548	Pr>|T|	0.0000
Sgn Rank	-250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	-0.14266	99%	-0.18225
75% Q3	-0.4905	95%	-0.29101
50% Med	-0.66841	90%	-0.367
25% Q1	-0.85487	10%	-0.98011
0% Min	-1.17095	5%	-1.06815
		1%	-1.14812
Range	1.028289		
Q3-Q1	0.364368		
Mode	-1.17095		

Extremes

Lowest	Obs	Highest	Obs
-1.17095(	427)	-0.15128(	942)
-1.16777(	266)	-0.14631(	906)
-1.16657(	396)	-0.14441(	481)
-1.16428(	606)	-0.14353(	409)
-1.15984(	88)	-0.14266(	733)



Variable = β_7 (parameter associated with the NED matriculation authority)

Moments

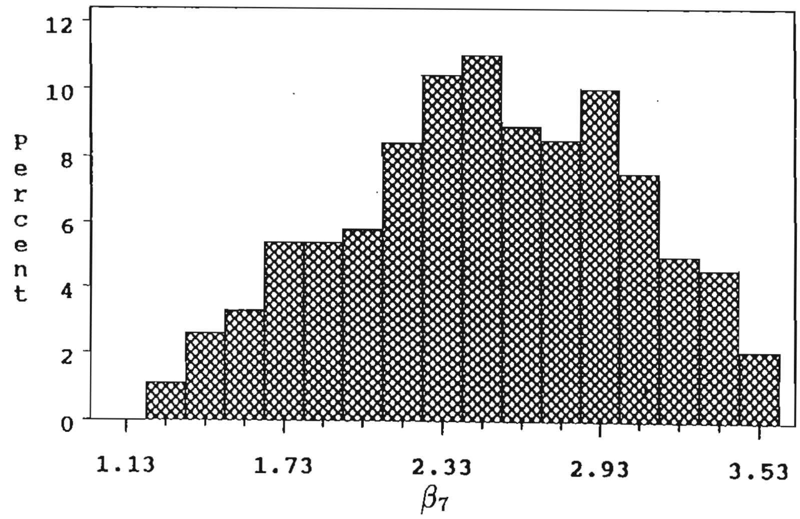
N	1000	Sum Wgts	1000
Mean	2.495623	Sum	2495.623
Std Dev	0.533757	Variance	0.284896
Skewness	-0.17959	Kurtosis	-0.70854
USS	6512.748	CSS	284.6115
CV	21.38772	Std Mean	0.016879
T:Mean=0	147.8549	Pr>|T|	0.0000
Sgn Rank	250250	Pr>=|S|	0.0000
Num ^= 0	1000		

Quantiles(Def=5)

100% Max	3.527636	99%	3.479743
75% Q3	2.922849	95%	3.352245
50% Med	2.507981	90%	3.187007
25% Q1	2.137842	10%	1.728642
0% Min	1.260725	5%	1.57857
		1%	1.340471
Range	2.266911		
Q3-Q1	0.785007		
Mode	1.260725		

Extremes

Lowest	Obs	Highest	Obs
1.260725(	543)	3.507784(	919)
1.269875(	561)	3.509482(	769)
1.271796(	279)	3.514332(	897)
1.280208(	353)	3.516201(	160)
1.282781(	651)	3.527636(	153)



CHAPTER FIVE

CONCLUSION

Predicting the probability of a student being "successful" at university is an extremely difficult task since it depends on many external factors that are difficult to quantify. For example a student may not study hard enough or attend lectures on a regular basis. Furthermore, a student entering university may attempt a set of subjects that he or she is not capable of passing. This will then in turn cause them to produce poor results even though they may have obtained excellent matriculation results. Thus when attempting to model the probability of a student being successful at university other variables such as socio-economic background of a student, the time spent studying by a student, the time spent travelling to and from university, the age of a student etc., need also to be considered.

However, from the models developed in chapters two, three and four it is evident that it is possible to predict the probability of a student being successful in a particular subject or set of subjects at university, based on the students matriculation result profiles. A significant conclusion that can be drawn from the results in chapters two and three is that students from the NED matriculation authority perform better, in the DSMA1SX1 and DSPH1SC1 courses, than the students from the other matriculation authorities. Unfortunately, students who wrote the Department of Education and Training (DET) matriculation examinations were excluded from our analyses due to the small number of DET students attempting both the DSMA1SX1 course and the DSPH1SC1 course. However, given a sample of DET students attempting the DSMA1SX1 course and the DSPH1SC1 course, it is then a fairly simple exercise to compare the probability of them being successful with students from the other matriculation authorities.

One of the weaknesses of the generalized linear model is that useful information is lost when introducing a Bernoulli random variable in order to obtain the desired event. However, this drawback is overcome by the fact that these models are easy to implement and also very useful in determining the probability of a student being successful as illustrated in chapter two. An important feature of the model described in chapter three is that the exact university results are used to determine the probability of a student being successful and thus no information is lost when implementing this model. Now, comparing the results obtained in chapter two with those obtained in chapter three, we see that in general the confidence intervals for the estimated probabilities in chapter three are smaller than those obtained in chapter two. Thus, for predicting the probability of a student being successful at university, I would prefer to use the model discussed in chapter three than those discussed in chapter two. In chapter four we adopted a Bayesian approach. This approach is different to the frequentist approaches discussed in chapters two and three. The advantage of the Bayesian approach, over those models developed in chapters two and three is that it can incorporate past experiences or beliefs of the parameters.

I strongly disapprove of the point system which is currently being used for determining which students should be admitted to university. A better method for selecting students would be to use the models described in chapters two, three and four in collaboration with the point system. It is also evident that these models can be used to determine which students should receive bursaries and in aiding student advisers in helping students with their course selection.

One of the shortcomings of the models that were developed in chapters two, three and four is that we have had to discard the data corresponding to those students who had not attempted each and every one of the following matriculation subjects, namely English, Afrikaans, Mathematics, Biology and Physics. In order to incorporate this data into our analysis Smyth et al. (1990) have developed an EM based technique which replaces any missing matriculation subject marks with an appropriately estimated value. The standard procedures given in chapters two, three and four can then be applied to this new augmented data set to determine the probability of a student being successful, no matter what matriculation subjects they have chosen.

APPENDIX A

APPENDIX A1

Letting the log-likelihood function, for the generalized linear model, be given by the expression

$$l(\boldsymbol{\theta}; y_1, y_2, \dots y_n) = \sum_{i=1}^{n} [y_i b(\theta_i) + c(\theta_i) + d(y_i)] = \sum_{i=1}^{n} l_i$$

then the *score* function associated with the parameter β_j is defined by

$$u_j = \frac{\partial l(\boldsymbol{\theta}; \mathrm{y})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j}$$

where $\boldsymbol{\theta}$ represents some function of $\boldsymbol{\beta}$. Noting that the partial derivative of l_i with respect to β_j is given by

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}$$

one can on differentiating l_i with respect to θ_i and substituting (from (2.2.3)) $c'(\theta_i) = -\mu_i b'(\theta_i)$ obtain the result that

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)[y_i - \mu_i]$$

Similarly differentiating μ_i with respect to θ_i and using (2.2.4) yields the result that

$$\frac{\partial \mu_i}{\partial \theta_i} = -\frac{c''(\theta_i)}{b'(\theta_i)} + \frac{c'(\theta_i) b''(\theta_i)}{[b'(\theta_i)]^2} = b'(\theta_i)\mathcal{V}\{y_i\}$$

and differentiating μ_i with respect to β_j, from (2.3.2), yields the result that

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = x_{ij} \frac{\partial \mu_i}{\partial \eta_i}$$

Thus an expression for the derivative of l_i with respect to β_j can be given by

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i}\frac{\partial \mu_i}{\partial \beta_j}\bigg/\frac{\partial \mu_i}{\partial \theta_i} = \frac{(y_i - \mu_i)}{\mathcal{V}\{y_i\}}x_{ij}\frac{\partial \mu_i}{\partial \eta_i}$$

with

$$u_j = \sum_{i=1}^{n}\frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n}\frac{y_i - \mu_i}{\mathcal{V}\{y_i\}}x_{ij}\frac{\partial \mu_i}{\partial \eta_i}$$

APPENDIX A2

Using the results obtained in Appendix A1, the jk'th element of the Information matrix can be given by

$$\begin{aligned}
\mathcal{I}_{jk} &= \mathcal{E}\left\{\frac{\partial l}{\partial \beta_j}\frac{\partial l}{\partial \beta_k}\right\} \\
&= \sum_{i=1}^{n}\mathcal{E}\left\{\frac{\partial l_i}{\partial \beta_j}\frac{\partial l_i}{\partial \beta_k}\right\} \\
&= \mathcal{E}\left\{\sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{(\mathcal{V}\{y_i\})^2}x_{ij}x_{ik}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\right\} \\
&= \sum_{i=1}^{n}\frac{x_{ij}x_{ik}}{\mathcal{V}\{y_i\}}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \\
&= \sum_{i=1}^{n}x_{ij}x_{ik}w_{ii}
\end{aligned}$$

and similarly the j'th component of $\mathbf{u}$ can be given by

$$u_j = \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n}\frac{(y_i - \mu_i)}{\mathcal{V}\{y_i\}}x_{ij}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2\frac{\partial \eta_i}{\partial \mu_i} = \sum_{i=1}^{n}(y_i - \mu_i)x_{ij}w_{ii}\frac{\partial \eta_i}{\partial \mu_i} = \sum_{i=1}^{n}x_{ij}w_{ii}z_i^*$$

where

$$w_{ii} = \frac{1}{\mathcal{V}\{y_i\}} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

and

$$z_i^* = (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)$$

Thus, writing in matrix notation we obtain the following results, namely

$$\mathcal{I} = \mathbf{X}'\mathbf{W}\mathbf{X}$$

and

$$\mathrm{u} = \mathbf{X}'\mathbf{W}\mathbf{z}^*$$

APPENDIX A3

Here we will attempt to derive an estimate for $\boldsymbol{\beta}$ that maximizes the log-likelihood function,

$$l(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i)$$

subject however to the restrictions $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, where $\mathbf{R}$ denotes a $(v \times p+1)$ matrix with j'th row $\mathbf{R}_j'$ and $\mathbf{r}$ a $(v \times 1)$ vector with j'th element r_j. In particular, and as an alternative to the more used Lagrange multiplier approach, we will introduce a penalty function (Nyquist, 1991) of the form

$$P(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \sum_{i=1}^{n} y_i b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i) - \frac{1}{2} \sum_{j=1}^{v} \lambda_j (r_j - \mathbf{R}_j'\boldsymbol{\beta})^2$$

and attempt to obtain an estimate for $\boldsymbol{\beta}$(that is dependent on the penalty function parameters $\boldsymbol{\lambda} = (\lambda_1 \ldots \lambda_v)$) that maximizes $P(\boldsymbol{\beta}, \boldsymbol{\lambda})$. Letting the penalty function parameters λ_i tend to infinity a restricted estimate for $\boldsymbol{\beta}$

can then be realized. Differentiating $P(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to β_j yields the following system of equations

$$
\begin{aligned}
q_j(\boldsymbol{\beta}, \boldsymbol{\lambda}) \triangleq \frac{\partial P(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} + \frac{\partial}{\partial \beta_j}\left[-\tfrac{1}{2}\sum_{k=1}^{v} \lambda_k (r_k - \mathbf{R}_k' \boldsymbol{\beta})^2\right] \\
&= \sum_{i=1}^{n} \frac{(y_i - \mu_i) x_{ij}}{\mathcal{V}\{y_i\}} \frac{\partial \mu_i}{\partial \eta_i} + \sum_{k=1}^{v} \lambda_k R_{kj} (r_k - \mathbf{R}_k' \boldsymbol{\beta}) \\
&= \sum_{i=1}^{n} (y_i - \mu_i) x_{ij} w_{ii} \frac{\partial \eta_i}{\partial \mu_i} + \sum_{k=1}^{v} \lambda_k R_{kj} (r_k - \mathbf{R}_k' \boldsymbol{\beta})
\end{aligned}
$$

for $j = 0, \ldots, p$, which must then be set equal to zero in order to obtain an estimate for $\boldsymbol{\beta}$. Note that the estimate obtained for $\boldsymbol{\beta}$ will depend on the values of the penalty function vector $\boldsymbol{\lambda}$. As such we will use the notation $\mathbf{b}(\boldsymbol{\lambda})$ to denote the set of parameter values that solve the above set of penalized likelihood function equations. Due to the nonlinear form that these equations take, Fisher's method of scoring as outlined in section 2.3 can be applied with

$$
\mathbf{b}(\boldsymbol{\lambda})^{(m)} = \mathbf{b}(\boldsymbol{\lambda})^{(m-1)} + [\mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\lambda})]^{-1} \; \mathbf{q}(\boldsymbol{\beta}, \boldsymbol{\lambda})
$$

with $\mathbf{S}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ and $\mathbf{q}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ being evaluated at $\boldsymbol{\beta} = \mathbf{b}(\boldsymbol{\lambda})^{(m-1)}$

where[1]

$$\mathbf{S}(\boldsymbol{\beta},\boldsymbol{\lambda}) = -\mathcal{E}\left\{\frac{\partial^2 P(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right\} = \mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{R}'\Lambda\mathbf{R}$$

and

$$\mathbf{q}(\boldsymbol{\beta},\boldsymbol{\lambda}) = \mathbf{X}'\mathbf{W}\mathbf{z}^* + \mathbf{R}'\Lambda\mathbf{r} - \mathbf{R}'\Lambda\mathbf{R}\boldsymbol{\beta}$$

with $\mathbf{z}^*$ denoting a n dimensional vector with i'th element

$$z_i^* = (y_i - \mu_i)\frac{\partial\eta_i}{\partial\mu_i}$$

and Λ a $v \times v$ dimensional diagonal matrix with i'th element λ_i. On making the appropriate substitutions we obtain the result

$$\mathbf{b}(\boldsymbol{\lambda})^{(m)} = \mathbf{b}(\boldsymbol{\lambda})^{(m-1)} + [\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X} + \mathbf{R}'\Lambda\mathbf{R}]^{-1}[\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}}^* + \mathbf{R}'\Lambda\mathbf{r} - \mathbf{R}'\Lambda\mathbf{R}\mathbf{b}(\boldsymbol{\lambda})^{(m-1)}]$$

[1] the jk'th component of of $\mathbf{S}(\boldsymbol{\beta},\boldsymbol{\lambda})$ is given by

$$\begin{aligned}
s_{jk}(\boldsymbol{\beta},\boldsymbol{\lambda}) &= -\mathcal{E}\left\{\frac{\partial^2 P(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial\beta_j\partial\beta_k}\right\} \\
&= -\mathcal{E}\left\{\frac{\partial^2 l}{\partial\beta_j\partial\beta_k} + \frac{\partial^2}{\partial\beta_j\partial\beta_k}\left(-\frac{1}{2}\sum_{i=1}^{v}\lambda_i(r_i - \mathbf{R}_j'\boldsymbol{\beta})^2\right)\right\} \\
&= -\mathcal{E}\left\{\frac{\partial^2 l}{\partial\beta_j\partial\beta_k}\right\} + \mathcal{E}\left\{\frac{\partial^2}{\partial\beta_j\partial\beta_k}\left(\frac{1}{2}\sum_{i=1}^{v}\lambda_i(r_i - \mathbf{R}_j'\boldsymbol{\beta})^2\right)\right\} \\
&= \mathcal{I}_{jk} + \mathcal{E}\left\{\sum_{i=1}^{v}\lambda_i R_{ij}R_{ik}\right\} \\
&= \mathcal{I}_{jk} + \sum_{i=1}^{v}\lambda_i R_{ij}R_{ik} \\
&= \sum_{i=1}^{n}x_{ij}x_{ik}w_{ii} + \sum_{i=1}^{v}\lambda_i R_{ij}R_{ik}
\end{aligned}$$

where $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{z}}^*$ denote the value of $\mathbf{W}$ and $\mathbf{z}^*$ evaluated at the previous iteration point $\mathbf{b}(\boldsymbol{\lambda})^{(m-1)}$. Premultiplying the above equation by $\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X} + \mathbf{R}'\Lambda\mathbf{R}$ we obtain

$$\begin{aligned}
&[\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X} + \mathbf{R}'\Lambda\mathbf{R}]\mathbf{b}(\boldsymbol{\lambda})^{(m)} \\
&= [\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X} + \mathbf{R}'\Lambda\mathbf{R}]\mathbf{b}(\boldsymbol{\lambda})^{(m-1)} + \mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}}^* + \mathbf{R}'\Lambda\mathbf{r} - \mathbf{R}'\Lambda\mathbf{R}\mathbf{b}(\boldsymbol{\lambda})^{(m-1)} \\
&= \mathbf{X}'\tilde{\mathbf{W}}\mathbf{X}\mathbf{b}(\boldsymbol{\lambda})^{(m-1)} + \mathbf{R}'\Lambda\mathbf{R}\mathbf{b}(\boldsymbol{\lambda})^{(m-1)} + \mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}}^* + \mathbf{R}'\Lambda\mathbf{r} - \mathbf{R}'\Lambda\mathbf{R}\mathbf{b}(\boldsymbol{\lambda})^{(m-1)} \\
&= \mathbf{X}'\tilde{\mathbf{W}}[\mathbf{X}\mathbf{b}(\boldsymbol{\lambda})^{(m-1)} + \tilde{\mathbf{z}}^*] + \mathbf{R}'\Lambda\mathbf{r}
\end{aligned}$$

Letting

$$\tilde{\mathbf{z}} = \mathbf{X}\mathbf{b}(\boldsymbol{\lambda})^{(m-1)} + \tilde{\mathbf{z}}^*$$

and noting that (Rao, 1965)

$$\begin{aligned}
&(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X} + \mathbf{R}'\Lambda\mathbf{R})^{-1} \\
&= [(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1} - (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}]
\end{aligned}$$

one can obtain the following iteratively weighted least squares formula for determining $\mathbf{b}(\boldsymbol{\lambda})^{(m)}$, namely

$$\begin{aligned}
\mathbf{b}(\boldsymbol{\lambda})^{(m)} &= [(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1} - (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}] \\
&\quad [\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} + \mathbf{R}'\Lambda\mathbf{r}] \\
&= (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} + (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\mathbf{r} \\
&\quad -(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} \\
&\quad -(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\mathbf{r}
\end{aligned}$$

$$
\begin{aligned}
&= (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} + (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1} \\
&\quad \{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}\mathbf{r} \\
&\quad -(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} \\
&\quad -(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\mathbf{r} \\
&= (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} + (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{r} \\
&\quad -(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} \\
&\quad -(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\mathbf{r} \\
&\quad +(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1}\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\mathbf{r} \\
&= (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} + (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\{\mathbf{I} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\Lambda\}^{-1} \\
&\quad \{\mathbf{r} - \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}}\} \\
&= (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} + (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\{\Lambda^{-1} + \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\}^{-1} \\
&\quad \{\mathbf{r} - \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}}\}
\end{aligned}
$$

Letting the penalty function parameters λ_i tend to infinity a restricted estimate for $\boldsymbol{\beta}$ can now be given by

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}} &\triangleq \lim_{m\to\infty}\left[\lim_{\lambda_1,\ldots,\lambda_v\to\infty}\mathbf{b}(\boldsymbol{\lambda})^{(m)}\right] \\
&= \lim_{m\to\infty}[(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}} + (\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\{\mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{R}'\}^{-1} \\
&\qquad \{\mathbf{r} - \mathbf{R}(\mathbf{X}'\tilde{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}'\tilde{\mathbf{W}}\tilde{\mathbf{z}}\}]
\end{aligned}
$$

APPENDIX A4

Since $\hat{\boldsymbol{\beta}}$ is asymptotically normally distributed with mean vector $\boldsymbol{\beta}$ and a variance-covariance matrix given by the inverse of the Information matrix, expanding $l(\boldsymbol{\beta})$ in a second order Taylor series around the point $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ yields

$$l(\boldsymbol{\beta}) \approx l(\hat{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathcal{I}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

which implies that for parameter estimation purposes

$$l(\boldsymbol{\beta}) \approx (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathcal{I}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

Now substituting our partitioned results (see footnote 2 in section 2.4) in the above expression we obtain the result

$$l(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \approx (\hat{\boldsymbol{\beta}}_1' - \boldsymbol{\beta}_1' \;\; \hat{\boldsymbol{\beta}}_2' - \boldsymbol{\beta}_2') \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \end{pmatrix}$$

Then substituting $\boldsymbol{\beta}_2 = \mathbf{0}$, in the above expression yields

$$\begin{aligned} l(\boldsymbol{\beta}_1, \mathbf{0}) \;\; &\approx \;\; [(\hat{\boldsymbol{\beta}}_1' - \boldsymbol{\beta}_1')\mathcal{I}_{11} + \hat{\boldsymbol{\beta}}_2'\mathcal{I}_{21} \;\; (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1')\mathcal{I}_{12} + \hat{\boldsymbol{\beta}}_2'\mathcal{I}_{22}] \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} \\ &\approx \;\; (\hat{\boldsymbol{\beta}}_1' - \boldsymbol{\beta}_1')\mathcal{I}_{11}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + \hat{\boldsymbol{\beta}}_2'\mathcal{I}_{21}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) + (\hat{\boldsymbol{\beta}}_1' - \boldsymbol{\beta}_1')\mathcal{I}_{12}\hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{\beta}}_2'\mathcal{I}_{22}\hat{\boldsymbol{\beta}}_2 \\ &\approx \;\; \hat{\boldsymbol{\beta}}_1'\mathcal{I}_{11}\hat{\boldsymbol{\beta}}_1 - 2\boldsymbol{\beta}_1'\mathcal{I}_{11}\hat{\boldsymbol{\beta}}_1 + \boldsymbol{\beta}_1'\mathcal{I}_{11}\boldsymbol{\beta}_1 + \hat{\boldsymbol{\beta}}_2'\mathcal{I}_{21}\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2'\mathcal{I}_{21}\boldsymbol{\beta}_1 \\ &\quad\;\; + \hat{\boldsymbol{\beta}}_1'\mathcal{I}_{12}\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_1'\mathcal{I}_{12}\hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{\beta}}_2'\mathcal{I}_{22}\hat{\boldsymbol{\beta}}_2 \end{aligned}$$

Finally maximizing the log-likelihood function, $l(\boldsymbol{\beta}_1, \mathbf{0})$, with respect to $\boldsymbol{\beta}_1$ yields

$$\frac{\partial l}{\partial \boldsymbol{\beta}_1} \approx -2\mathcal{I}_{11}\hat{\boldsymbol{\beta}}_1 + 2\mathcal{I}_{11}\tilde{\boldsymbol{\beta}}_1 - 2\mathcal{I}_{12}\hat{\boldsymbol{\beta}}_2 \stackrel{\text{set}}{=} 0$$

which implies that

$$\tilde{\boldsymbol{\beta}}_1 \approx \hat{\boldsymbol{\beta}}_1 + \mathcal{I}_{11}^{-1}\mathcal{I}_{12}\hat{\boldsymbol{\beta}}_2$$

where $\tilde{\boldsymbol{\beta}}_1$ represents as estimate for $\boldsymbol{\beta}$ obtained under the restrictions that $\boldsymbol{\beta}_2 = \mathbf{0}$. Thus, D^* is now obtained by substituting $\tilde{\boldsymbol{\beta}} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 \\ \mathbf{0} \end{pmatrix}$ in D, yielding

$$D^* = 2[l(\hat{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\beta}}_1 - \mathcal{I}_{11}^{-1}\mathcal{I}_{12}\hat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}_2 = \mathbf{0})]$$

APPENDIX B

REJECTION SAMPLING

Suppose that we want to obtain a sample of observations from a probability density function $f(x)$ whose form is known only up to a proportionality constant, that is $f(x) = Mg(x)$ for some unknown value of M. Employing a rejection sampling technique Gilks and Wild (1992) have developed a technique that is based on using the function $g(\cdot)$, an envelope function $g_u(\cdot)$ and optionally a squeezing function $g_l(\cdot)$, such that $g_l(\cdot) \leq g(\cdot) \leq g_u(\cdot)$ for all x. By independently sampling a value x^* from the density function $s(x) = \dfrac{g_u(x)}{\int g_u(x)dx}$ and a value u from a $U(0,1)$ distribution, the following test procedure can then be employed to generate an observation from the desired probability density function $f(\cdot)$, namely if

$$u \leq \frac{g_l(x^*)}{g_u(x^*)}$$

then accept x^* as an observation from $f(\cdot)$, otherwise evaluate $g(x^*)$ and accept x^* as an observation from $f(\cdot)$ if

$$u \leq \frac{g(x^*)}{g_u(x^*)}$$

The above procedure is then repeated until the desired number of observations have been sampled. Now, in general, finding a suitable envelope function $g_u(x)$ can be difficult and usually involves locating the supremum of $g(x)$. However, if the function $g(x)$ is log-concave[1] then the functions

[1]Any positive function g on an open convex set D in $\mathbb{R}^n$ is called log-concave if $\log g$ is a twice continuously differentiable real-valued function on D and its Hessian matrix, $H_{ij}(x) = \left[\dfrac{\partial^2 \log g(x_1 \ldots x_n)}{\partial x_i \partial x_j}\right]$, is negative semidefinite for every $x \in D$. If the Hessian matrix is negative definite, then the function g is said to be strictly log-concave.

$g_l(x)$ and $g_u(x)$ can easily be obtained by noting that any concave function is bounded above and below by its tangents and chords respectively.

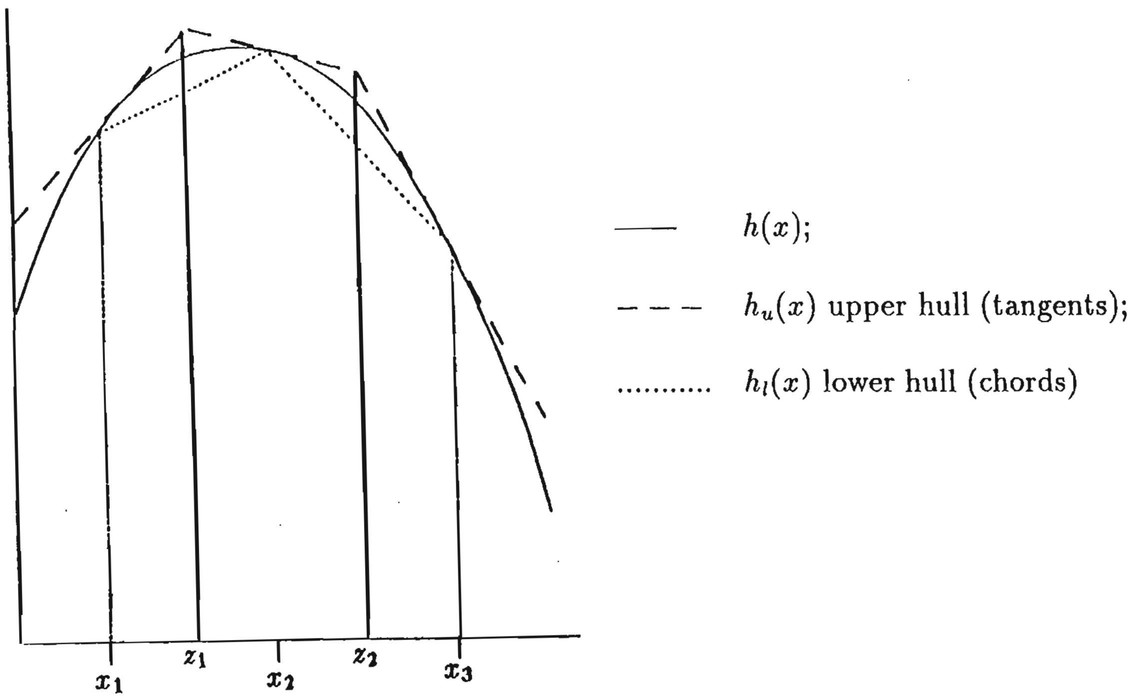


Figure B.1: Graph representing a concave function $h(x)$

In particular, letting $h(x) = \log g(x)$ and referring to the diagram given in Figure B.1, the envelope and squeezing functions for $g(x)$ can be obtained by setting

$$g_u(x) = \exp[h_u(x)]$$

and

$$g_l(x) = \exp[h_l(x)]$$

respectively where [2]

$$
\begin{aligned}
h_u(x) &= h(x_1) + (x - x_1)h'(x_1)\,, \text{ if } x < z_1 \\
&= h(x_2) + (x - x_2)h'(x_2)\,, \text{ if } z_1 \leq x \leq z_2 \qquad \text{(B1.1)} \\
&= h(x_3) + (x - x_3)h'(x_3)\,, \text{ if } x > z_3
\end{aligned}
$$

denotes the tangents[3] and

$$
\begin{aligned}
h_l(x) &= \frac{h(x_2) - h(x_1)}{x_2 - x_1} x + \frac{x_2h(x_1 - x_1h(x_2)}{x_2 - x_1}\,, \text{ if } x_1 \leq x \leq x_2 \\
&= \frac{h(x_3) - h(x_2)}{x_3 - x_2} x + \frac{x_3h(x_2) - x_2h(x_3)}{x_3 - x_2}\,, \text{ if } x_2 < x \leq x_3 \qquad \text{(B1.2)} \\
&= -\infty\,, \text{ otherwise}
\end{aligned}
$$

denotes the chords as shown in Figure B.1. Now in order to generate a value x^* from the exponentiated upper hull, we require an expression for the density

[2]Note that in order to reduce the number of points that are likely to be rejected, x_1 and x_3 should be chosen so that $h'(x_1) > 0$ and $h'(x_3) < 0$ where $h'(x_1)$ and $h'(x_3)$ represents the derivative of the function $h(x)$ evaluated at the points x_1 and x_3 respectively. For our applications we also choose

$$x_2 = \frac{[h(x_3) - h(x_1) - x_3h'(x_3) + x_1h'(x_1)]}{[h'(x_1) - h'(x_3)]}$$

which further reduces the rejection envelope.

[3]where the tangents intersect at the following points, namely

$$z_k = \frac{h(x_{k+1}) - h(x_k) - x_{k+1}h'(x_{k+1}) + x_kh'(x_k)}{h'(x_k) - h'(x_{k+1})} \;\; ; \; k = 1, 2$$

function that is associated with the exponentiated upper hull, namely[4]

$$
\begin{aligned}
s(x) &= \frac{\exp\left(h_u(x)\right)}{\int \exp\left(h_u(x)\right) dx} \\
&= \sum_{k=1}^{3} \frac{\exp[h(x_k) + h'(x_k)(x - x_k)] I_{(z_{k-1};z_k)}}{\int \exp\left(h_u(x)\right) dx}
\end{aligned}
\tag{B1.3}
$$

Letting V_1, V_2 and V_3 denote the areas under the exponentiated curves bounded by the tangents and the points $\{-\infty, z_1\}$, $\{z_1, z_2\}$ and $\{z_2, \infty\}$ respectively (refer to Figure B.1), then we obtain the result that

$$c_u = \int \exp\left[h_u(x)\right] dx = V_1 + V_2 + V_3$$

where [5]

$$
\begin{aligned}
V_j = \int_{z_{j-1}}^{z_j} \exp[h_u(x)] dx &= \int_{h_u(z_{j-1})}^{h_u(z_j)} \exp[h_u(x)] \left| \frac{dx}{dh_u(x)} \right| dh_u(x) \\
&= \int_{h_u(z_{j-1})}^{h_u(z_j)} \exp[h_u(x)] \frac{1}{h'(x_j)} dh_u(x) \\
&= \left. \frac{\exp[h_u(x)]}{h'(x_j)} \right|_{h_u(z_{j-1})}^{h_u(z_j)} \\
&= \frac{\exp[h_u(z_j)] - \exp[h_u(z_{j-1})]}{h'(x_j)} \qquad \text{for } j = 1, 2, 3
\end{aligned}
$$

and thus the following two stage method can be used to generate a value x^* from the piecewise exponential probability density function s(x), namely

[4]where $z_0 = -\infty$ and $z_3 = \infty$.

[5]note that

$$\frac{dh_u(x)}{dx} = h'(x_j)$$

i) generate a discrete random variable y^* based on the probability density function

$$P(y = k) = \frac{V_k}{V_1 + V_2 + V_3} \qquad\qquad k = 1, 2, 3$$

ii) if $y^* = k$, then generate a value x^* from the density $s_k(x)$ defined by

$$\begin{aligned} s_k(x) \triangleq \frac{\exp h_u(x)}{V_k} &= \frac{\exp[h(x_k) + h'(x_k)(x - x_k)]}{V_k} \\ &= \frac{h'(x_k)}{\exp\{z_k h'(x_k)\} - \exp\{z_{k-1} h'(x_k)\}} \exp\{x h'(x_k)\} \end{aligned}$$

for $x \in (z_{k-1}, z_k)$ and $k = 1, 2, 3$. In particular the random variable x^* can be generated by first generating a random variable u^* from a $U(0,1)$ distribution, and then setting [6]

$$x^* = \frac{\log\left[u^* \exp\{z_k h'(x_k)\} + (1 - u^*) \exp\{z_{k-1} h'(x_k)\}\right]}{h'(x_k)}$$

---

[6]Noting that the density function with respect to u^* is given by

$$f_{U^*}(u^*) = 1$$

and letting

$$x = \frac{\log[u^* \exp\{z_k h'(x_k)\} + (1 - u^*) \exp\{z_{k-1} h'(x_k)\}]}{h'(x_k)} \tag{1}$$

the density for x can then be given by

$$f_X(x) = f_{U^*}(u^*) \left| \frac{du^*}{dx} \right| = 1 \left| \frac{du^*}{dx} \right|$$

Now from (1) we obtain

$$u^* = \frac{\exp\{x h'(x_k)\} - \exp\{z_{k-1} h'(x_k)\}}{\exp\{z_k h'(x_k)\} - \exp\{z_{k-1} h'(x_k)\}}$$

which implies that

$$\frac{du^*}{dx} = \frac{h'(x_k) \exp\{x h'(x_k)\}}{\exp\{z_k h'(x_k)\} - \exp\{z_{k-1} h'(x_k)\}}$$

which corresponds to the expression given for $s_k(x)$

We can thus implement the Gibbs sampler, using the above method of Rejection Sampling, until the desired number of sample points from each conditional density have been obtained.

APPENDIX C

C1 DSMA1SX1 DATA SET

Z1 = DSMA1SX1 (MATHEMATICS1) MARK
Y1 = CODED UNIVERSTY MARK (IE. Y1 = 1 IF Z1 IS GREATER THAN
OR EQUAL TO 48; ELSE Y1 = 0)
X1 = MATRICULATION ENGLISH CODED VALUE
X2 = MATRICULATION AFRIKAANS CODED VALUE
X3 = MATRICULATION MATHEMATICS CODED VALUE
X4 = MATRICULATION BIOLOGY CODED VALUE
X5 = MATRICULATION PHYSICS CODED VALUE
D1 (HOD) AND D2 (NED) REPRESENT THE DUMMY VARIABLES THAT
ACCOUNT FOR THE TYPE OF MATRICULATION AUTHORITY
NB. IN CHAPTER 2 WE USE THE CODED MARK Y1 AND NOT THE ACTUAL MARK Z1.
AND IN CHAPTER 3 WE USE THE ACTUAL MARK Z1 AND NOT THE CODED MARK Y1.

OBS	Z1	X1	X2	X3	X4	X5	D1	D2	Y1
1	50	2	1	1	2	2	1	0	1
2	86	0	0	0	0	0	0	1	1
3	30	3	5	1	3	2	1	0	0
4	28	2	3	0	2	2	1	0	0
5	50	2	3	2	1	2	1	0	1
6	81	2	2	0	2	1	0	1	1
7	22	3	3	2	3	1	1	0	0
8	40	2	4	2	3	2	0	1	0
9	40	0	1	1	0	1	1	0	0
10	81	1	1	0	1	1	0	1	1
11	44	3	4	0	2	1	1	0	0
12	34	2	2	1	2	2	1	0	0
13	50	2	2	2	3	2	0	0	1
14	40	1	1	1	2	2	1	0	0
15	34	1	2	2	3	3	1	0	0
16	43	0	4	0	2	1	0	0	0
17	85	1	0	0	0	0	0	1	1
18	46	1	3	0	1	2	1	0	0
19	25	2	2	2	3	2	1	0	0
20	65	2	3	0	2	2	1	0	1
21	32	2	1	2	2	3	0	0	0
22	15	2	4	1	1	1	1	0	0
23	44	1	1	1	2	1	1	0	0
24	43	2	1	1	2	0	0	1	0
25	40	0	1	1	1	2	1	0	0
26	70	1	1	0	0	1	1	0	1
27	55	1	3	2	1	2	1	0	1
28	51	1	2	0	0	1	1	0	1
29	42	2	1	2	0	2	0	0	0
30	56	2	2	2	0	1	1	0	1
31	92	0	0	0	0	0	0	1	1
32	44	1	0	0	1	1	0	1	0
33	50	1	2	0	0	0	0	1	1
34	51	3	2	3	4	3	0	1	1
35	57	2	2	1	2	1	0	1	1
36	53	2	0	0	1	1	0	1	1
37	34	2	2	0	2	2	0	0	0
38	77	1	2	1	1	1	0	1	1
39	58	1	3	1	1	2	1	0	1
40	56	1	2	0	0	1	1	0	1
41	99	0	0	0	0	0	0	1	1
42	40	3	3	2	3	3	0	1	0
43	86	2	2	0	1	1	0	1	1
44	29	3	3	1	1	2	1	0	0
45	65	3	4	0	2	2	1	0	1
46	40	2	4	2	4	3	1	0	0
47	42	1	2	2	3	2	0	1	0
48	28	2	2	2	2	3	1	0	0
49	88	2	1	0	1	0	1	0	1
50	50	2	2	2	1	2	0	1	1
51	30	3	4	1	1	3	0	0	0
52	41	2	3	1	1	2	1	0	0
53	68	0	2	0	1	1	1	0	1
54	83	0	2	0	0	1	0	1	1
55	52	1	2	1	1	2	0	1	1
56	33	3	3	2	0	0	1	0	0
57	52	0	4	1	2	1	0	1	1
58	44	1	2	0	0	1	0	1	0
59	58	2	2	0	1	2	1	0	1
60	44	3	2	1	2	2	1	0	0
61	52	2	3	2	0	3	0	0	1
62	62	2	3	2	2	2	0	1	1
63	87	0	0	0	0	1	1	0	1
64	8	0	2	3	2	4	1	0	0
65	56	2	2	0	1	1	1	0	1
66	58	1	2	0	1	1	1	0	1
67	89	2	2	0	1	1	1	0	1
68	50	2	3	0	2	2	0	0	1
69	34	1	1	2	2	0	1	0	0
70	40	3	3	2	1	2	1	0	0
71	75	2	3	0	2	2	0	1	1
72	16	3	5	3	2	1	1	0	0
73	77	0	0	0	0	0	0	1	1
74	40	2	4	2	1	3	1	0	0
75	65	2	2	0	3	0	0	1	1
76	74	1	0	0	0	1	0	1	1
77	44	3	4	1	1	3	1	0	0
78	40	2	2	0	2	3	1	0	0
79	69	1	1	1	1	2	1	0	1
80	55	2	3	2	1	3	0	1	1
81	53	2	2	1	2	2	0	1	1
82	42	1	2	3	1	2	1	0	0
83	76	2	2	0	1	0	0	1	1
84	31	1	2	2	0	1	1	0	0
85	68	0	1	1	0	0	1	0	1
86	36	2	3	0	3	1	0	1	0
87	54	1	2	1	1	1	0	1	1
88	34	1	2	1	2	1	1	0	0
89	50	1	1	2	1	2	1	0	1
90	58	1	1	0	1	0	1	0	1
91	27	2	2	3	2	3	0	1	0
92	42	2	3	1	1	1	0	1	0
93	50	3	4	1	0	3	0	0	1
94	45	2	3	1	1	1	0	1	0
95	74	2	3	0	3	1	0	1	1
96	66	4	5	0	2	1	0	0	1
97	51	2	3	1	1	2	1	0	1
98	44	2	2	1	3	1	1	0	0
99	87	0	1	0	0	0	1	0	1
100	57	2	2	1	0	2	1	0	1
101	56	2	2	1	1	2	1	0	1
102	40	0	1	2	1	2	1	0	0

103	59	3	4	1	1	0	1	0	1
104	25	2	2	1	1	2	1	0	0
105	32	1	2	3	1	2	1	0	0
106	77	1	2	0	0	1	1	0	1
107	90	3	2	0	0	0	0	0	1
108	36	2	2	2	2	2	0	1	0
109	82	0	1	0	0	0	1	0	1
110	77	0	1	0	1	0	1	0	1
111	51	2	1	1	1	1	1	0	1
112	100	0	0	0	0	0	0	1	1
113	91	0	1	0	0	0	1	0	1
114	91	0	2	0	1	1	1	0	1
115	70	0	2	0	1	1	1	0	1
116	46	1	1	1	0	1	1	0	0
117	86	1	1	1	1	0	1	0	1
118	34	2	1	2	2	2	1	0	0
119	50	1	1	2	2	2	1	0	1
120	80	1	0	0	0	0	1	0	1
121	50	1	3	1	0	0	0	1	1
122	85	0	2	0	0	0	1	0	1
123	57	2	2	0	0	0	1	0	1
124	50	2	2	0	0	1	1	0	1
125	79	1	2	0	1	1	1	0	1
126	78	1	2	0	1	1	1	0	1
127	32	0	0	0	0	0	0	1	0
128	50	1	1	2	1	2	1	0	1
129	95	1	1	0	0	0	1	0	1
130	82	0	1	0	1	1	0	1	1
131	42	2	3	1	2	3	1	0	0
132	44	3	3	0	1	2	1	0	0
133	69	1	2	1	0	1	0	0	1
134	72	0	1	0	1	1	0	1	1
135	95	0	1	0	0	0	1	0	1
136	43	2	1	0	1	2	1	0	0
137	23	3	2	2	3	3	0	1	0
138	61	2	1	1	2	2	1	0	1
139	30	1	1	1	0	1	1	0	0
140	33	1	3	2	1	1	1	0	0
141	70	3	2	1	3	2	1	0	1
142	77	2	1	2	1	1	1	0	1
143	40	1	2	2	0	1	1	0	0
144	100	0	0	0	0	0	0	1	1
145	62	2	3	2	2	2	1	0	1
146	97	0	2	0	0	0	1	0	1
147	64	2	2	1	1	2	1	0	1
148	33	2	2	2	0	2	1	0	0
149	29	2	3	1	2	3	1	0	0
150	81	2	2	1	1	2	1	0	1
151	45	1	3	0	0	1	0	0	0
152	63	2	2	1	1	1	1	0	1
153	90	2	1	0	0	1	1	0	1
154	51	2	1	0	1	2	1	0	1
155	75	2	2	0	0	0	1	0	1
156	69	0	1	1	0	0	1	0	1
157	57	2	1	0	0	0	0	1	1
158	88	1	2	0	0	0	1	0	1
159	68	2	3	0	0	2	1	0	1
160	89	1	2	1	1	0	0	1	1
161	45	0	2	1	0	1	0	0	0
162	60	1	2	1	1	3	0	1	1
163	40	1	2	2	1	2	0	1	0
164	46	2	2	2	2	1	1	0	0
165	54	0	1	1	0	0	1	0	1
166	63	0	1	0	1	0	1	0	1
167	40	2	1	2	2	3	1	0	0
168	44	2	2	2	1	2	1	0	0
169	61	1	1	0	0	1	1	0	1
170	26	2	3	1	1	2	1	0	0
171	21	2	2	2	0	2	1	0	0
172	47	1	2	1	0	1	1	0	0
173	69	0	1	1	0	1	1	0	1
174	51	0	1	1	0	1	1	0	1
175	75	1	1	1	0	2	0	1	1
176	50	2	2	1	0	1	0	1	1
177	94	2	1	0	0	0	0	1	1
178	44	2	1	2	1	1	1	0	0
179	52	1	2	2	0	1	1	0	1
180	58	2	1	1	0	0	1	0	1
181	35	2	5	3	3	3	0	0	0
182	15	2	4	2	1	3	1	0	0
183	32	0	0	2	1	2	1	0	0
184	31	2	4	3	4	3	0	1	0
185	36	0	1	1	1	1	1	0	0
186	74	0	4	4	2	2	0	1	1
187	87	0	1	0	0	0	1	0	1
188	46	2	0	3	1	2	1	0	0
189	50	2	3	3	1	3	1	0	1
190	64	1	2	1	1	2	0	1	1
191	70	1	2	0	0	0	1	0	1
192	48	2	3	1	0	1	1	0	1
193	40	2	2	1	2	2	0	1	0
194	50	2	1	2	1	1	1	0	1
195	52	3	0	2	3	3	0	1	1
196	30	1	3	0	1	2	1	0	0
197	48	2	2	1	1	0	1	0	1
198	59	0	1	0	0	1	1	0	1
199	51	0	2	1	2	1	1	0	1
200	50	2	1	0	0	0	1	0	1
201	40	3	2	0	1	0	1	0	0
202	91	1	2	0	0	1	0	1	1
203	100	0	1	0	0	0	0	1	1
204	40	2	1	2	2	2	1	0	0
205	44	2	0	0	1	0	1	0	0
206	78	2	2	1	2	1	0	1	1
207	81	0	0	0	0	0	0	1	1
208	21	2	3	1	3	4	1	0	0
209	53	1	1	0	0	1	1	0	1
210	61	2	2	1	1	1	1	0	1
211	51	0	1	0	1	1	1	0	1
212	80	2	2	2	1	2	0	1	1
213	50	2	3	0	2	2	1	0	1
214	90	2	3	1	0	1	0	1	1
215	57	2	1	2	2	1	0	1	1
216	55	1	1	0	0	1	1	0	1
217	30	2	2	1	1	2	1	0	0
218	76	1	2	0	0	0	0	1	1
219	61	0	2	0	0	0	1	0	1
220	53	0	0	2	0	2	0	1	1
221	47	0	2	1	2	1	0	1	0
222	50	3	3	0	0	0	0	0	1
223	55	0	1	0	1	1	0	0	1
224	75	1	2	0	0	0	1	0	1
225	71	0	0	0	0	0	1	0	1
226	89	1	1	0	1	0	1	0	1
227	33	0	1	0	1	1	1	0	0
228	99	0	0	0	0	0	1	0	1
229	54	2	2	1	0	1	1	0	1
230	94	2	4	0	1	0	0	1	1
231	40	2	2	2	1	2	0	0	0
232	84	1	1	0	0	0	1	0	1
233	40	2	2	2	1	2	1	0	0
234	65	2	2	2	1	2	0	1	1
235	26	1	1	2	0	1	0	1	0
236	80	1	1	0	1	0	1	0	1
237	88	0	0	0	0	0	0	1	1
238	63	2	3	1	0	0	0	1	1
239	42	2	1	1	3	2	0	1	0
240	55	2	3	1	3	1	0	1	1
241	63	2	1	0	0	0	1	0	1
242	50	2	2	0	1	0	0	1	1
243	24	2	3	3	1	2	0	0	0
244	50	2	2	1	3	1	0	1	1
245	61	1	0	0	0	0	1	0	1
246	40	0	0	0	0	1	1	0	0
247	24	1	2	2	1	3	1	0	0
248	60	2	3	0	1	0	0	1	1
249	79	0	0	0	0	0	0	1	1
250	88	0	1	1	0	0	1	0	1
251	50	0	1	0	0	0	1	0	1
252	59	1	1	0	0	0	1	0	1
253	57	1	2	2	0	0	0	1	1
254	59	1	0	0	0	1	1	0	1
255	34	1	2	2	1	1	1	0	0
256	71	1	3	0	1	0	0	1	1

257	66	1	1	0	2	1	1	0	1
258	40	2	2	1	1	0	1	0	0
259	33	2	3	1	1	2	0	1	0
260	78	1	1	0	0	0	1	0	1
261	47	1	2	2	0	1	1	0	0
262	50	0	0	1	0	0	1	0	1
263	50	2	3	2	1	2	0	1	1
264	63	2	2	0	2	0	1	0	1
265	40	2	3	1	1	0	0	1	0
266	23	3	4	1	2	1	1	0	0
267	50	2	2	0	0	0	1	0	1
268	75	1	1	0	1	2	0	1	1
269	45	0	0	0	0	1	1	0	0
270	36	2	3	2	4	2	1	0	0
271	64	1	3	0	0	0	0	1	1
272	68	2	3	0	1	2	1	0	1
273	71	2	4	0	2	1	0	1	1
274	47	1	1	0	0	0	1	0	0
275	70	1	2	0	0	0	0	1	1
276	53	1	2	1	0	1	1	0	1
277	29	1	1	2	1	1	1	0	0
278	42	3	2	2	0	2	1	0	0
279	27	1	2	2	2	2	1	0	0
280	50	1	2	0	0	1	0	1	1
281	67	1	0	0	0	0	1	0	1
282	57	1	2	1	0	0	1	0	1
283	58	1	1	0	0	0	1	0	1
284	42	1	1	2	0	2	1	0	0
285	40	2	1	1	1	1	1	0	0
286	64	0	2	0	1	0	0	1	1
287	62	1	1	0	0	0	1	0	1
288	44	1	1	0	1	0	1	0	0
289	65	1	3	0	0	0	0	1	1
290	63	2	3	0	0	2	1	0	1
291	40	1	2	0	0	1	1	0	0
292	52	2	3	1	1	0	1	0	1
293	75	2	1	0	0	0	1	0	1
294	62	2	2	1	2	2	0	1	1
295	44	0	1	1	0	0	1	0	0
296	82	1	1	0	1	0	0	1	1
297	52	1	2	2	1	2	1	0	1
298	21	2	2	2	2	2	1	0	0
299	40	1	2	1	1	2	0	1	0
300	33	2	2	1	0	1	1	0	0
301	40	2	3	0	1	0	1	0	0
302	70	1	1	0	0	0	0	1	1
303	53	2	1	1	0	0	1	0	1
304	48	2	2	2	2	2	1	0	1
305	40	3	2	0	1	2	1	0	0

END OF DATA SET

C2 DSPH1SC1 DATA SET

Z2 = DSPH1SC1 (PHYSICS1) MARK

Y2 = CODED UNIVERSTY MARK (IE. Y2 = 1 IF Z2 IS GREATER THAN OR EQUAL TO 48; ELSE Y2 = 0)

X1 = MATRICULATION ENGLISH CODED VALUE

X2 = MATRICULATION AFRIKAANS CODED VALUE

X3 = MATRICULATION MATHEMATICS CODED VALUE

X4 = MATRICULATION BIOLOGY CODED VALUE

X5 = MATRICULATION PHYSICS CODED VALUE

D1 (HOD) AND D2 (NED) REPRESENT THE DUMMY VARIABLES THAT ACCOUNT FOR THE TYPE OF MATRICULATION AUTHORITY

NB. IN CHAPTER 2 WE USE THE CODED MARK Y2 AND NOT THE ACTUAL MARK Z2.
AND IN CHAPTER 3 WE USE THE ACTUAL MARK Z2 AND NOT THE CODED MARK Y2.

OBS	Z2	X1	X2	X3	X4	X5	D1	D2	Y2
1	81	0	0	0	0	0	0	1	1
2	8	3	5	1	3	2	1	0	0
3	22	2	3	0	2	2	1	0	0
4	55	2	2	0	2	1	0	1	1
5	54	1	1	0	1	1	0	1	1
6	47	3	4	0	2	1	1	0	0
7	32	2	2	1	2	2	1	0	0
8	25	2	1	2	2	3	0	0	0
9	23	2	4	1	1	1	1	0	0
10	40	1	1	1	2	1	1	0	0
11	65	2	1	1	2	0	0	1	1
12	31	0	1	1	1	2	1	0	0
13	50	2	1	2	0	2	0	0	1
14	20	2	1	3	2	3	0	0	0
15	33	2	2	0	2	2	0	0	0
16	83	0	0	0	0	0	0	1	1
17	41	3	3	2	3	3	0	1	0
18	69	2	2	0	1	1	0	1	1
19	32	3	3	1	1	2	1	0	0
20	46	1	2	2	3	2	0	1	0
21	29	3	4	1	1	3	0	0	0
22	50	0	2	0	1	1	1	0	1
23	50	0	2	0	0	1	0	1	1
24	40	1	2	1	1	2	0	1	0
25	61	3	3	2	0	0	1	0	1
26	47	2	2	0	1	2	1	0	0
27	38	3	2	1	2	2	1	0	0
28	58	2	2	1	1	2	1	0	1
29	60	0	0	0	0	1	1	0	1
30	18	0	2	3	2	4	1	0	0
31	51	2	2	0	1	1	1	0	1
32	71	2	2	0	1	1	1	0	1
33	47	1	1	2	2	0	1	0	0
34	51	3	3	2	1	2	1	0	1
35	60	2	3	0	2	2	0	1	1
36	60	0	0	0	0	0	0	1	1
37	28	2	4	2	1	3	1	0	0
38	55	1	1	1	1	2	1	0	1
39	63	0	1	1	0	0	1	0	1
40	57	1	2	1	1	1	0	1	1
41	47	1	2	1	2	1	1	0	0
42	72	1	1	0	1	0	1	0	1
43	60	2	2	3	2	3	0	1	1
44	45	4	5	0	2	1	0	0	0
45	71	2	3	1	1	2	1	0	1
46	82	0	1	0	0	0	1	0	1
47	73	3	4	1	1	0	1	0	1
48	83	1	2	0	0	1	1	0	1
49	85	3	2	0	0	0	0	0	1
50	85	0	1	0	0	0	1	0	1
51	86	0	1	0	1	0	1	0	1
52	100	0	0	0	0	0	0	1	1
53	83	0	1	0	0	0	1	0	1
54	40	1	1	1	0	1	1	0	0
55	76	1	1	1	1	0	1	0	1
56	45	2	1	2	2	2	1	0	0
57	66	1	3	1	0	0	0	1	1
58	77	0	2	0	0	0	1	0	1
59	54	2	2	0	0	0	1	0	1
60	65	2	2	0	0	1	1	0	1
61	67	1	2	0	1	1	1	0	1
62	75	1	2	0	1	1	1	0	1
63	58	0	0	0	0	0	0	1	1
64	44	1	1	2	1	2	1	0	0
65	77	1	1	0	0	0	1	0	1
66	65	0	1	0	1	1	0	1	1
67	50	3	3	0	1	2	1	0	1
68	72	1	2	1	0	1	0	0	1
69	38	1	1	3	2	2	1	0	0
70	56	2	2	3	2	2	0	1	1
71	87	0	1	0	0	0	1	0	1
72	45	1	1	1	0	1	1	0	0
73	78	2	1	2	1	1	1	0	1
74	70	1	2	2	0	1	1	0	1
75	98	0	0	0	0	0	0	1	1
76	62	2	3	2	2	2	1	0	1
77	88	0	2	0	0	0	1	0	1
78	53	2	2	2	0	2	1	0	1
79	50	1	3	0	0	1	0	0	1
80	77	2	1	0	0	1	1	0	1
81	59	2	1	0	1	2	1	0	1
82	79	2	2	0	0	0	1	0	1
83	82	0	1	1	0	0	1	0	1
84	69	2	1	0	0	0	0	1	1
85	94	1	2	0	0	0	1	0	1
86	84	1	2	1	1	0	0	1	1
87	47	0	2	1	0	1	0	0	0
88	67	1	2	1	1	3	0	1	1
89	41	2	2	2	2	1	1	0	0
90	56	0	1	1	0	0	1	0	1
91	66	0	1	0	1	0	1	0	1
92	41	2	2	2	1	2	1	0	0
93	63	1	1	0	0	1	1	0	1
94	47	2	3	1	1	2	1	0	0
95	63	1	2	1	0	1	1	0	1
96	70	0	1	1	0	1	1	0	1
97	43	0	1	1	0	1	1	0	0
98	56	2	2	1	0	1	0	1	1
99	53	1	2	2	0	1	1	0	1
100	33	2	2	3	3	2	1	0	0
101	63	2	1	1	0	0	1	0	1
102	41	2	5	3	3	3	0	0	0
103	31	2	4	2	1	3	1	0	0
104	51	2	4	3	4	3	0	1	1
105	40	0	1	1	1	1	1	0	0
106	63	0	4	4	2	2	0	1	1
107	71	0	1	0	0	0	1	0	1
108	42	2	0	3	1	2	1	0	0

109	56	1	2	0	0	0	1	0	1
110	46	2	3	1	0	1	1	0	0
111	44	2	2	1	2	2	0	1	0
112	32	2	1	4	1	2	0	0	0
113	61	2	1	2	1	1	1	0	1
114	55	3	0	2	3	3	0	1	1
115	40	2	2	1	1	0	1	0	0
116	53	0	1	0	0	1	1	0	1
117	57	2	1	0	0	0	1	0	1
118	55	3	2	0	1	0	1	0	1
119	96	0	1	0	0	0	0	1	1
120	42	2	1	2	2	2	1	0	0
121	44	2	0	0	1	0	1	0	0
122	63	2	2	1	2	1	0	1	1
123	65	0	0	0	0	0	0	1	1
124	48	1	1	0	0	1	1	0	1
125	42	2	2	1	1	1	1	0	0
126	62	0	1	0	1	1	1	0	1
127	36	2	3	0	2	2	1	0	0
128	86	2	3	1	0	1	0	1	1
129	65	2	1	2	2	1	0	1	1
130	55	1	1	0	0	1	1	0	1
131	52	0	2	0	0	0	1	0	1
132	66	0	0	2	0	2	0	1	1
133	72	0	2	1	2	1	0	1	1
134	60	3	3	0	0	0	0	0	1
135	44	0	1	0	1	1	0	0	0
136	57	1	2	0	0	0	1	0	1
137	69	0	0	0	0	0	1	0	1
138	69	1	1	0	1	0	1	0	1
139	45	0	1	0	1	1	1	0	0
140	80	0	0	0	0	0	1	0	1
141	93	2	4	0	1	0	0	1	1
142	77	1	1	0	0	0	1	0	1
143	68	2	2	2	1	2	0	1	1
144	90	1	1	0	1	0	1	0	1
145	59	0	0	0	0	0	0	1	1
146	54	2	1	1	3	2	0	1	1
147	61	2	3	1	3	1	0	1	1
148	76	2	1	0	0	0	1	0	1
149	70	2	2	0	1	0	0	1	1
150	25	2	3	3	1	2	0	0	0
151	62	2	2	1	3	1	0	1	1
152	58	1	0	0	0	0	1	0	1
153	57	0	0	0	0	1	1	0	1
154	62	2	3	0	1	0	0	1	1
155	87	0	0	0	0	0	0	1	1
156	87	0	1	1	0	0	1	0	1
157	66	0	1	0	0	0	1	0	1
158	43	2	3	2	2	2	0	1	0
159	67	1	1	0	0	0	1	0	1
160	77	1	2	2	0	0	0	1	1
161	53	1	0	0	0	1	1	0	1
162	50	1	2	2	1	1	1	0	1
163	81	1	3	0	1	0	0	1	1
164	53	2	2	1	1	0	1	0	1
165	59	1	1	0	0	0	1	0	1
166	47	0	0	1	0	0	1	0	0
167	60	2	2	0	2	0	1	0	1
168	72	2	3	1	1	0	0	1	1
169	50	3	4	1	2	1	1	0	1
170	66	2	2	0	0	0	1	0	1
171	50	2	3	0	1	2	1	0	1
172	43	2	3	4	2	3	0	1	0
173	68	2	4	0	2	1	0	1	1
174	66	1	1	0	0	0	1	0	1
175	56	3	3	2	2	2	0	1	1
176	67	1	2	1	0	1	1	0	1
177	38	2	2	3	1	2	1	0	0
178	29	1	1	2	1	1	1	0	0
179	36	3	2	2	0	2	1	0	0
180	75	1	2	0	0	1	0	1	1
181	67	1	0	0	0	0	1	0	1
182	75	1	2	1	0	0	1	0	1
183	70	1	1	0	0	0	1	0	1
184	47	1	1	2	0	2	1	0	0
185	40	2	1	1	1	1	1	0	0
186	86	0	2	0	1	0	0	1	1
187	79	1	1	0	0	0	1	0	1
188	50	1	1	0	1	0	1	0	1
189	47	3	4	3	4	3	0	1	0
190	78	1	3	0	0	0	0	1	1
191	75	2	3	0	0	2	1	0	1
192	68	1	2	0	0	1	1	0	1
193	92	2	1	0	0	0	1	0	1
194	69	2	2	1	2	2	0	1	1
195	52	0	1	1	0	0	1	0	1
196	77	1	1	0	1	0	0	1	1
197	41	1	2	3	1	3	1	0	0
198	57	1	2	1	1	2	0	1	1
199	58	2	2	1	0	1	1	0	1
200	68	2	3	0	1	0	1	0	1
201	73	1	1	0	0	0	0	1	1
202	41	1	1	2	1	3	0	1	0
203	56	2	1	1	0	0	1	0	1
204	32	3	2	0	1	2	1	0	0

END OF DATA SET

C3 DSMA1SX1 AND DSPH1SC1 DATA SET

Z1 = DSMA1SX1 (MATHEMATICS1) MARK

Z2 = DSPH1SC1 (PHYSICS1) MARK

Y = CODED UNIVERSTY MARK (IE. Y = 1 IF Z1 AND Z2 IS GREATER THAN OR EQUAL TO 48; ELSE Y = 0)

X1 = MATRICULATION ENGLISH CODED VALUE

X2 = MATRICULATION AFRIKAANS CODED VALUE

X3 = MATRICULATION MATHEMATICS CODED VALUE

X4 = MATRICULATION BIOLOGY CODED VALUE

X5 = MATRICULATION PHYSICS CODED VALUE

D1 (HOD) AND D2 (NED) REPRESENT THE DUMMY VARIABLES THAT ACCOUNT FOR THE TYPE OF MATRICULATION AUTHORITY

NB. IN CHAPTER 2 WE USE THE CODED MARK Y AND NOT THE ACTUAL MARKS Z1 AND Z2. AND IN CHAPTER 3 WE USE THE ACTUAL MARKS Z1 AND Z2 AND NOT THE CODED MARK Y.

OBS	Z1	Z2	X1	X2	X3	X4	X5	D1	D2	Y
1	86	81	0	0	0	0	0	0	1	1
2	30	8	3	5	1	3	2	1	0	0
3	28	22	2	3	0	2	2	1	0	0
4	81	55	2	2	0	2	1	0	1	1
5	81	54	1	1	0	1	1	0	1	1
6	44	47	3	4	0	2	1	1	0	0
7	34	32	2	2	1	2	2	1	0	0
8	32	25	2	1	2	2	3	0	0	0
9	15	23	2	4	1	1	1	1	0	0
10	44	40	1	1	1	2	1	1	0	0
11	43	65	2	1	1	2	0	0	1	0
12	40	31	0	1	1	1	2	1	0	0
13	42	50	2	1	2	0	2	0	0	0
14	34	33	2	2	0	2	2	0	0	0
15	99	83	0	0	0	0	0	0	1	1
16	40	41	3	3	2	3	3	0	1	0
17	86	69	2	2	0	1	1	0	1	1
18	29	32	3	3	1	1	2	1	0	0
19	42	46	1	2	2	3	2	0	1	0
20	30	29	3	4	1	1	3	0	0	0
21	68	50	0	2	0	1	1	1	0	1
22	83	50	0	2	0	0	1	0	1	1
23	52	40	1	2	1	1	2	0	1	0
24	33	61	3	3	2	0	0	1	0	0
25	58	47	2	2	0	1	2	1	0	0
26	44	38	3	2	1	2	2	1	0	0
27	87	60	0	0	0	0	1	1	0	1
28	8	18	0	2	3	2	4	1	0	0
29	56	51	2	2	0	1	1	1	0	1
30	89	71	2	2	0	1	1	1	0	1
31	34	47	1	1	2	2	0	1	0	0
32	40	51	3	3	2	1	2	1	0	0
33	75	60	2	3	0	2	2	0	1	1
34	77	60	0	0	0	0	0	0	1	1
35	40	28	2	4	2	1	3	1	0	0
36	69	55	1	1	1	1	2	1	0	1
37	68	63	0	1	1	0	0	1	0	1
38	54	57	1	2	1	1	1	0	1	1
39	34	47	1	2	1	2	1	1	0	0
40	58	72	1	1	0	1	0	1	0	1
41	27	60	2	2	3	2	3	0	1	0
42	66	45	4	5	0	2	1	0	0	0
43	51	71	2	3	1	1	2	1	0	1
44	87	82	0	1	0	0	0	1	0	1
45	59	73	3	4	1	1	0	1	0	1
46	77	83	1	2	0	0	1	1	0	1
47	90	85	3	2	0	0	0	0	0	1
48	82	85	0	1	0	0	0	1	0	1
49	77	86	0	1	0	1	0	1	0	1
50	100	100	0	0	0	0	0	0	1	1
51	91	83	0	1	0	0	0	1	0	1
52	46	40	1	1	1	0	1	1	0	0
53	86	76	1	1	1	1	0	1	0	1
54	34	45	2	1	2	2	2	1	0	0
55	50	66	1	3	1	0	0	0	1	1
56	85	77	0	2	0	0	0	1	0	1
57	57	54	2	2	0	0	0	1	0	1
58	50	65	2	2	0	0	1	1	0	1
59	79	67	1	2	0	1	1	1	0	1
60	78	75	1	2	0	1	1	1	0	1
61	32	58	0	0	0	0	0	0	1	0
62	50	44	1	1	2	1	2	1	0	0
63	95	77	1	1	0	0	0	1	0	1
64	82	65	0	1	0	1	1	0	1	1
65	44	50	3	3	0	1	2	1	0	0
66	69	72	1	2	1	0	1	0	0	1
67	95	87	0	1	0	0	0	1	0	1
68	30	45	1	1	1	0	1	1	0	0
69	77	78	2	1	2	1	1	1	0	1
70	40	70	1	2	2	0	1	1	0	0
71	100	98	0	0	0	0	0	0	1	1
72	62	62	2	3	2	2	2	1	0	1
73	97	88	0	2	0	0	0	1	0	1
74	33	53	2	2	2	0	2	1	0	0
75	45	50	1	3	0	0	1	0	0	0
76	90	77	2	1	0	0	1	1	0	1
77	51	59	2	1	0	1	2	1	0	1
78	75	79	2	2	0	0	0	1	0	1
79	69	82	0	1	1	0	0	1	0	1
80	57	69	2	1	0	0	0	0	1	1
81	88	94	1	2	0	0	0	1	0	1
82	89	84	1	2	1	1	0	0	1	1
83	45	47	0	2	1	0	1	0	0	0
84	60	67	1	2	1	1	3	0	1	1
85	46	41	2	2	2	2	1	1	0	0
86	54	56	0	1	1	0	0	1	0	1
87	63	66	0	1	0	1	0	1	0	1
88	44	41	2	2	2	1	2	1	0	0
89	61	63	1	1	0	0	1	1	0	1
90	26	47	2	3	1	1	2	1	0	0
91	47	63	1	2	1	0	1	1	0	0
92	69	70	0	1	1	0	1	1	0	1
93	51	43	0	1	1	0	1	1	0	0
94	50	56	2	2	1	0	1	0	1	1
95	52	53	1	2	2	0	1	1	0	1
96	58	63	2	1	1	0	0	1	0	1
97	35	41	2	5	3	3	3	0	0	0
98	15	31	2	4	2	1	3	1	0	0
99	31	51	2	4	3	4	3	0	1	0
100	36	40	0	1	1	1	1	1	0	0
101	74	63	0	4	4	2	2	0	1	1
102	87	71	0	1	0	0	0	1	0	1

103	46	42	2	0	3	1	2	1	0	0
104	70	56	1	2	0	0	0	1	0	1
105	48	46	2	3	1	0	1	1	0	0
106	40	44	2	2	1	2	2	0	1	0
107	50	61	2	1	2	1	1	1	0	1
108	52	55	3	0	2	3	3	0	1	1
109	48	40	2	2	1	1	0	1	0	0
110	59	53	0	1	0	0	1	1	0	1
111	50	57	2	1	0	0	0	1	0	1
112	40	55	3	2	0	1	0	1	0	0
113	100	96	0	1	0	0	0	0	1	1
114	40	42	2	1	2	2	2	1	0	0
115	44	44	2	0	0	1	0	1	0	0
116	78	63	2	2	1	2	1	0	1	1
117	81	65	0	0	0	0	0	0	1	1
118	53	48	1	1	0	0	1	1	0	1
119	61	42	2	2	1	1	1	1	0	0
120	51	62	0	1	0	1	1	1	0	1
121	50	36	2	3	0	2	2	1	0	0
122	90	86	2	3	1	0	1	0	1	1
123	57	65	2	1	2	2	1	0	1	1
124	55	55	1	1	0	0	1	1	0	1
125	61	52	0	2	0	0	0	1	0	1
126	53	66	0	0	2	0	2	0	1	1
127	47	72	0	2	1	2	1	0	1	0
128	50	60	3	3	0	0	0	0	0	1
129	55	44	0	1	0	1	1	0	0	0
130	75	57	1	2	0	0	0	1	0	1
131	71	69	0	0	0	0	0	1	0	1
132	89	69	1	1	0	1	0	1	0	1
133	33	45	0	1	0	1	1	1	0	0
134	99	80	0	0	0	0	0	1	0	1
135	94	93	2	4	0	1	0	0	1	1
136	84	77	1	1	0	0	0	1	0	1
137	65	68	2	2	2	1	2	0	1	1
138	80	90	1	1	0	1	0	1	0	1
139	88	59	0	0	0	0	0	0	1	1
140	42	54	2	1	1	3	2	0	1	0
141	55	61	2	3	1	3	1	0	1	1
142	63	76	2	1	0	0	0	1	0	1
143	50	70	2	2	0	1	0	0	1	1
144	24	25	2	3	3	1	2	0	0	0
145	50	62	2	2	1	3	1	0	1	1
146	61	58	1	0	0	0	0	1	0	1
147	40	57	0	0	0	0	1	1	0	0
148	60	62	2	3	0	1	0	0	1	1
149	79	87	0	0	0	0	0	0	1	1
150	88	87	0	1	1	0	0	1	0	1
151	50	66	0	1	0	0	0	1	0	1
152	59	67	1	1	0	0	0	1	0	1
153	57	77	1	2	2	0	0	0	1	1
154	59	53	1	0	0	0	1	1	0	1
155	34	50	1	2	2	1	1	1	0	0
156	71	81	1	3	0	1	0	0	1	1
157	40	53	2	2	1	1	0	1	0	0
158	78	59	1	1	0	0	0	1	0	1
159	50	47	0	0	1	0	0	1	0	0
160	63	60	2	2	0	2	0	1	0	1
161	40	72	2	3	1	1	0	0	1	0
162	23	50	3	4	1	2	1	1	0	0
163	50	66	2	2	0	0	0	1	0	1
164	68	50	2	3	0	1	2	1	0	1
165	71	68	2	4	0	2	1	0	1	1
166	47	66	1	1	0	0	0	1	0	0
167	53	67	1	2	1	0	1	1	0	1
168	29	29	1	1	2	1	1	1	0	0
169	42	36	3	2	2	0	2	1	0	0
170	50	75	1	2	0	0	1	0	1	1
171	67	67	1	0	0	0	0	1	0	1
172	57	75	1	2	1	0	0	1	0	1
173	58	70	1	1	0	0	0	1	0	1
174	42	47	1	1	2	0	2	1	0	0
175	40	40	2	1	1	1	1	1	0	0
176	64	86	0	2	0	1	0	0	1	1
177	62	79	1	1	0	0	0	1	0	1
178	44	50	1	1	0	1	0	1	0	0
179	65	78	1	3	0	0	0	0	1	1
180	63	75	2	3	0	0	2	1	0	1
181	40	68	1	2	0	0	1	1	0	0
182	75	92	2	1	0	0	0	1	0	1
183	62	69	2	2	1	2	2	0	1	1
184	44	52	0	1	1	0	0	1	0	0
185	82	77	1	1	0	1	0	0	1	1
186	40	57	1	2	1	1	2	0	1	0
187	33	58	2	2	1	0	1	1	0	0
188	40	68	2	3	0	1	0	1	0	0
189	70	73	1	1	0	0	0	0	1	1
190	53	56	2	1	1	0	0	1	0	1
191	40	32	3	2	0	1	2	1	0	0

END OF DATA SET

APPENDIX D

PROGRAMS

D1 DISCRIMINANT ANALYSIS

```
PROGRAM USED TO OBTAIN THE DISCRIMINANT ANALYSIS RESULTS ON
OUR DSMA1SX1 DATA.  RESULTS OUTINED IN CHAPTER ONE.  THIS
PROGRAM WAS ALSO APPLIED TO OUR DSPH1SC1 DATA AND OUR COMBINED
DSMA1SX1 AND DSPH1SC1 DATA.

libname in 'c:\robert';

data math1; /**CONTAINS ALL THOSE STUDENTS WHO FAILED DSMA1SX1**/
set in.mp2;

  if z1='0' then delete;
  if z1='.' then delete;
  if ma='4' then d1=1; else d1=0;
  if ma='6' then d2=1; else d2=0;
  if z1 ge 48 then y1=1; else y1=0;
  if y1=1 then delete;
run;

data math2; /**CONTAINS ALL THOSE STUDENTS WHO PASSED DSMA1SX1**/
set in.mp2;

  if z1='0' then delete;
  if z1='.' then delete;
  if ma='4' then d1=1; else d1=0;
  if ma='6' then d2=1; else d2=0;
  if z1 ge 48 then y1=1; else y1=0;
  if y1=0 then delete;
run;

proc iml;
reset noprint;
use math1;  /**UNSUCCESSFUL GROUP**/
read all var{x1 x2 x3 x4 x5 d1 d2};
x=x1||x2||x3||x4||x5||d1||d2;
n1=nrow(x);
v=ncol(x);
one=j(n1,1,1);
xbar1=1/n1 * one`*x;
xb1=one*xbar1;
s1=1/(n1-1) * (x-xb1)`*(x-xb1);
```

```
print xbar1 s1;
/**xbar1 = mean vector for the successful(pass) group**/
/**s1 = variance-covariance matrix for the successful group**/

use math2; /**SUCCESSFUL GROUP**/
read all var{x1 x2 x3 x4 x5 d1 d2};
x=x1||x2||x3||x4||x5||d1||d2;
n2=nrow(x);
one=j(n2,1,1);
xbar2=1/n2 * one'*x;
xb2=one*xbar2;
s2=1/(n2-1) * (x-xb2)'*(x-xb2);
print xbar2 s2;
/**xbar2= mean vector for the unsuccessful(fail) group**/
/**s2 = variance-covariance matrix for the unsuccessful group**/

spool=((n1-1)*s1 + (n2-1)*s2)/(n1+n2-2);
/**spool = pooled variance-covariance matrix**/

k=(xbar1-xbar2)*inv(spool);
c=(1/2)*(xbar1-xbar2)*inv(spool)*(xbar1+xbar2)';
print spool k c;
print n1 n2 ;
/*n1 - no. of failures        n2 - no. of successes   */

/**END OF PROGRAM**/
```

D2.1 LOGIT LINK FUNCTION

```
GENERALIZED LINEAR MODEL WITH LOGIT LINK FUNCTION

PROGRAM USED TO OBTAIN THE PARAMETER ESTIMATES, STANDARD ERROR
OF THESE ESTIMATES, WALD TEST STATISTIC VALUES, PROBAILITY VALUES
ASSOCIATED WITH THE WALD TEST STATISTIC AND 95 % LOWER AND UPPER
CONFIDENCE INTERVALS  FOR THE PARAMETER ESTIMATES FOR OUR DSMA1SX1
DATA (RESULTS GIVEN IN TABLE 2.6.2).  THIS PROGRAM WAS ALSO USED
TO OBTAIN THE ESTIMATED PROBABILITY (TOGETHER WITH ITS CONFIDENCE
INTERVALS) ASSOCIATED WITH A STUDENT PASSING DSMA1SX1 FOR A GIVEN
SET OF MATRICULATION PROFILE RESULTS (RESULTS GIVEN IN TABLE 2.6.3).
THIS PROGRAM WAS ALSO APPLIED ON OUR DSPH1SC1 DATA AND OUR COMBINED
DSMA1SX1 AND DSPH1SC1 DATA.**/

/**ORGANISING OUR DSMA1SX1 DATA SET **/

libname in 'c:\robert';
data math ;
set in.mp2;

if ma='4' then d1=1; else d1=0;
if ma='6' then d2=1; else d2=0;
if z1='0' then delete;
if z1='.' then delete;
if z1 ge 48 then y1=1; else y1=0;
run;


/** THE ACTUAL PROGRAM**/

proc iml worksize=100;
reset noprint;
use math;
read all var{x1 x2 x3 x4 x5 d1 d2 y1};
n=nrow(x3);
X=j(n,1,1)||x3||x4||x5||d2; /**MATRIX OF PREDICTOR VARIABLES**/

/** THE FISHER SCORING ROUTINE USING A CONVERGENCE
    CRITERION OF 0.0001**/

p=ncol(x);
b=j(p,1,0);
z=j(n,1,0);
w=j(n,n,0);
oldb=b+j(p,1,1);
```

```
do iter =1 to 60 while(max(abs(b-oldb))> 0.0001);
 oldb=b;
  do i=1 to n;
    yi=y1[i];
    xit=x[i, ];xi=xit`;
    pi=exp(xi`*b)/(1+exp(xi`*b));
    wii=pi*(1-pi);
    zi=xi`*b + (yi - pi)`*inv(pi*(1-pi));
    w[i,i]=wii;
    z[i]=zi;
  end;
  detw=det(x`*w*x);
  b=inv(x`*w*x)*x`*w*z;
end;

/** END OF THE FISHER SCORING ROUTINE (SEE EQUATION (2.3.10)**/

   covb=inv(x`*w*x);

/**LB = LOWER BOUND FOR THE CONFIDENCE INTERVAL OF THE PARAMETER ESTIMATE **/
LB=j(p,1,0);
  do j=1 to p;
    LCBj=b[j]-1.96*sqrt(covb[j,j]);
    LB[j]=LCBj;
  end;

/**UB = UPPER BOUND FOR THE CONFIDENCE INTERVAL OF THE PARAMETER ESTIMATE **/
UB=j(p,1,0);
  do j=1 to p;
    UCBj=b[j]+1.96*sqrt(covb[j,j]);
    UB[j]=UCBj;
  end;

/**SEB = STANDARD ERROR OT THE PARAMETER ESTIMATES **/
SEB=j(p,1,0);
 do j=1 to p;
    sebj=sqrt(covb[j,j]);
    seb[j]=sebj;
 end;

m=nrow(b);

/**WT = WALD TEST STATISTIC**/
WT=j(m,1,0); prob=j(m,1,0);
R=j(1,m,0);
```

```
 do i=1 to m;
  R[i]=1;
  WTi=(R*b)'*inv(R*inv(x'*w*x)*R')*(R*b);

  WT[i]=WTi;
  q=nrow(R);  probi=1-probchi(WTi,q);
  prob[i]=probi ;
  R[i]=0;
 end;
q=8 ;

/**PHAT = ESTIMATED PROBABILITY  **/
/**LBPHAT AND UBPHAT = LOWER AND UPPER BOUNDS FOR THE
   CONFIDENCE INTERVALS ASSOCIATED WITH PHAT**/

PHAT=j(q,1,0); LBPHAT=j(q,1,0);  UBPHAT=j(q,1,0);
x={1 0 0 0 0 , 1 1 0 0 0 , 1 0 1 0 0 , 1 0 0 1 0 , 1 0 0 0 1 ,
   1 1 0 0 1 , 1 0 1 0 1 , 1 0 0 1 1};
do i=1 to q;
  phati=exp(x[i, ]*b)*inv(1+exp(x[i, ]*b));
  phat[i]=phati;
  lbphati=exp(x[i, ]*LB)*inv(1+exp(x[i, ]*LB));
  ubphati=exp(x[i, ]*UB)*inv(1+exp(x[i, ]*UB));
  lbphat[i]=lbphati;
  ubphat[i]=ubphati;
end;
print b seb WT prob LB UB;
print X phat lbphat ubphat;

/**END OF PROGRAM**/
```

D2.2 PROBIT LINK FUNCTION

```
GENERALIZED LINEAR MODEL WITH PROBIT LINK FUNCTION

PROGRAM USED TO OBTAIN THE PARAMETER ESTIMATES, STANDARD ERROR
OF THESE ESTIMATES, WALD TEST STATISTIC VALUES, PROBAILITY VALUES
ASSOCIATED WITH THE WALD TEST STATISTIC AND 95 % LOWER AND UPPER
CONFIDENCE INTERVALS  FOR THE PARAMETER ESTIMATES FOR OUR DSMA1SX1
DATA (RESULTS GIVEN IN TABLE 2.6.10).  THIS PROGRAM WAS ALSO USED
TO OBTAIN THE ESTIMATED PROBABILITY (TOGETHER WITH ITS CONFIDENCE
INTERVALS) ASSOCIATED WITH A STUDENT PASSING DSMA1SX1 FOR A GIVEN
SET OF MATRICULATION PROFILE RESULTS (RESULTS GIVEN IN TABLE 2.6.11).
THIS PROGRAM WAS ALSO APPLIED ON OUR DSPH1SC1 DATA AND OUR COMBINED
DSMA1SX1 AND DSPH1SC1 DATA.**/

libname in 'c:\robert';
data math              ;
set in.mp2;

if ma='4' then d1=1; else d1=0;
if ma='6' then d2=1; else d2=0;

if z1='0' then delete;
if z1='.' then delete;
if z1 ge 48 then y1=1; else y1=0;
proc sort data=math;
by ma;
run;

proc iml worksize=100;
reset noprint;
use math;
read all var{x1 x2 x3 x4 x5 d1 d2 y1};
n=nrow(x3);
X=j(n,1,1)||x3||x4||x5||d2;

/**FISHERS SCORING ROUTINE**/
p=ncol(x);
b=j(p,1,0);
z=j(n,1,0);
w=j(n,n,0);
oldb=b+j(p,1,1);
 do iter =1 to 60 while(max(abs(b-oldb))> 0.0001);
  oldb=b;
   do i=1 to n;
     yi=y1[i];
```

```
     xit=x[i, ];xi=xit';
     pi=probnorm(xi'*b);
     vi=xi'*b;
     l=1/(sqrt(2)*gamma(0.5));
     phi=l*exp((-1/2)*vi*vi);
     wii=(phi*phi)/(pi*(1-pi));
     zi=xi'*b + (yi - pi)*inv(phi);
     w[i,i]=wii;
     z[i]=zi;
   end;
   detw=det(x'*w*x);
   b=inv(x'*w*x)*x'*w*z;
 end;
/**END OF FISHERS SCORING ROUTINE**/

   covb=inv(x'*w*x);
LB=j(p,1,0);
  do j=1 to p;
    LCBj=b[j]-1.96*sqrt(covb[j,j]);
    LB[j]=LCBj;
  end;

UB=j(p,1,0);
  do j=1 to p;
    UCBj=b[j]+1.96*sqrt(covb[j,j]);
    UB[j]=UCBj;
  end;
seb=j(p,1,0);
 do j=1 to p;
    sebj=sqrt(covb[j,j]);
    seb[j]=sebj;
 end;
m=nrow(b);
WT=j(m,1,0); prob=j(m,1,0);
R=j(1,m,0);
 do i=1 to m;
  R[i]=1;
  WTi=(R*b)'*inv(R*inv(x'*w*x)*R')*(R*b);

  WT[i]=WTi;
  q=nrow(R);  probi=1-probchi(WTi,q);
  prob[i]=probi ;
  R[i]=0;
 end;
```

```
q=8 ;
phat=j(q,1,0); lbphat=j(q,1,0);  ubphat=j(q,1,0);
x={1 0 0 0 0 , 1 1 0 0 0 , 1 0 1 0 0 , 1 0 0 1 0 , 1 0 0 0 1 ,
   1 1 0 0 1 , 1 0 1 0 1 , 1 0 0 1 1};
do i=1 to q;
  phati=probnorm(x[i, ]*b);
  phat[i]=phati;
  lbphati=probnorm(x[i, ]*LB);
  ubphati=probnorm(x[i, ]*UB);
  lbphat[i]=lbphati;
  ubphat[i]=ubphati;
end;
print b seb WT prob LB UB;
print X phat lbphat ubphat;

/**END OF PROGRAM**/
```

D3.1 A CUMULATIVE STANDARD NORMAL FUNCTION

PROGRAM USED TO OBTAIN THE REGRESSION PARAMETER ESTIMATES, STANDARD ERROR FOR THE PARAMETER ESTIMATES, T-VALUES, UPPER AND LOWER BOUNDS FOR THE CONFIDENCE INTERVALS OF THE PARAMETER ESTIMATES AND THE MEAN SQUARE ERROR (FOR THE PROBLEM DISCUSSED IN CHAPTER THREE) FOR OUR DSMA1SX1 COURSE (RESULTS GIVEN IN TABLE (3.5.1). THIS PROGRAM ALSO GIVES THE ESTIMATED PROBABILITY (TOGETHER WITH ITS CONFIDENCE INTERVAL) ASSOCIATED WITH A STUDENT PASSING THE DSMA1SX1 COURSE FOR A GIVEN MATRICULATION RESULT PROFILE (RESULTS GIVEN IN TABLE 3.5.2). THIS PROGAM WAS ALSO APPLIED ON OUR DSPH1SC1 COURSE.

```
libname perm 'c:\robert';
data math; /**DSMA1SX1 DATA**/
set perm.mp2;
if z1 = '0' then delete;
if z1 = '.' then delete;
if ma=4 then d1=1; else d1=0;
if ma=6 then d2=1; else d2=0;


proc iml worksize=100;
reset noprint;
 use math;
read all var{z1 x1 x2 x3 x4 x5 d1 d2};
n=nrow(z1);
x0=j(n,1,1);
x=x0 || x1 || x3 || x4 || x5 || d2;
p=ncol(x);
beta=inv(x'*x)*x'*z1;
sse=z1'*z1-beta'*x'*z1;
mse=sse/ (n-p);
covb=inv(x'*x)*mse;
seb=j(p,1,0);
do j=1 to p;
sebj=sqrt(covb[j,j]);
seb[j]=sebj;
end;

t=j(p,1,0);
do j=1to p;
tj=beta[j]/seb[j];
t[j]=tj;
end;

lb=beta-1.96*seb; ub=beta+1.96*seb;
```

```
x={1 0 0 0 0 0 , 1 1 0 0 0 0 ,1 0 1 0 0 0 ,1 0 0 1 0 0 ,1 0 0 0 1 0,
   1 0 0 0 0 1 , 1 1 0 0 0 1 ,1 0 1 0 0 1 ,1 0 0 1 0 1 ,1 0 0 0 1 1};

zhat=x*beta;  zhatlb=x*lb;    zhatub=x*ub;
sx=(zhat-48)/sqrt(mse);
sxlb=(zhatlb-48)/sqrt(mse);
sxub=(zhatub-48)/sqrt(mse);
phat=probnorm(sx);
phatlb=probnorm(sxlb);
phatub=probnorm(sxub);

print beta seb t lb ub mse;
print  X phat phatlb phatub;

/**END OF PROGRAM**/
```

D3.2 MONTE CARLO ALGORITHM

PROGRAM USED TO OBTAIN THE ESTIMATED PROBABILITY (TOGETHER WITH ITS CONFIDENCE INTERVALS) ASSOCIATED WITH A STUDENT PASSING BOTH DSMA1SX1 AND DSPH1SC1 FOR A GIVEN SET OF MATRICULATION PROFILES (RESULTS GIVEN IN TABLE 3.5.6). THIS PROBLEM IS DISCUSSED IN CHAPTER THREE.

```
libname in 'c:\robert';
data maph;
set in.mp2;

if ma='4' then d1=1; else d1=0;
if ma='6' then d2=1; else d2=0;

if z1='0' then delete;
if z1='.' then delete;
if z2='0' then delete;
if z2='.' then delete;
run;

proc iml worksize=100;
reset noprint;
use maph;
read all var{z1 z2 x1 x2 x3 x4 x5 d1 d2};
n=nrow(z1);
x0=j(n,1,1);
x= x0 || x1 || x2 || x3 || x4 || x5 || d1 || d2 ;
y=z1 || z2;
p=ncol(y);

/** ROUTINE TO GENERATE OBSERVATIONS FROM A MULTIVARIATE NORMAL DENSTIY**/
start multnor(ystar,ystarlb,ystarub,mu,mulb,muub,sigma,p);
seed = 456 ;
 z=j(p,1,0);
  sigma=(sigma` + sigma)/2;
  u=root(sigma);
  t=u`;
  do i=1 to p;
    z[i]=rannor(seed);
  end;
ystar=t*z + mu;
ystarlb=t*z + mulb;
ystarub=t*z + muub;
finish ;
/**END OF ROUTINE**/
```

```
x={1 0 0 0 0 1 0 0 0 0,0 0 0 0 0 0 1 0 0 1,
   1 0 1 0 0 1 0 0 0 0,0 0 0 0 0 0 1 0 0 1,
   1 0 0 1 0 1 0 0 0 0,0 0 0 0 0 0 1 1 0 1,
   1 0 0 0 1 0 0 0 0 0,0 0 0 0 0 0 1 0 1 1,
   1 0 0 0 0 0 0 0 0 0,0 0 0 0 0 0 1 0 0 0,
   1 0 1 0 0 0 0 0 0 0,0 0 0 0 0 0 1 0 0 0,
   1 0 0 1 0 0 0 0 0 0,0 0 0 0 0 0 1 1 0 0,
   1 0 0 0 1 0 0 0 0 0,0 0 0 0 0 0 1 0 1 0};

/*b1 and b2 obtained from the sur model*/


b1={69.647416,-2.917197,-5.949945, -2.771966,-5.423292,11.317970};
b2={67.594627,-4.548053,-8.916156,13.103354};

bet=b1//b2;
q=8;
prob=j(q,1,0); problb=j(q,1,0); probub=j(q,1,0);
do k=1 to q;
mu=(x[2*k-1,]//x[2*k,])*bet;

/* sigma2 obtained from the sur model*/

sigma={240.3949 118.5459, 118.5459 154.7962};

/*stder obtained from the sur model*/
stder={2.01102 , 1.05425 , 1.25738 , 1.62782 , 1.57495 , 2.61120 ,
       1.39130 , 1.25355 , 1.16054 , 2.07635};

lbeta=bet-1.96*stder;
ubeta=bet+1.96*stder;
mulb=(x[2*k-1,]//x[2*k,])*lbeta;
muub=(x[2*k-1,]//x[2*k,])*ubeta;

/* For a given mu and sigma , a set of m  p-dimensional N(mu,sigma)
   random vectors are generated. If the generated vector ystar
   lies in the region A={y[1]>=c1,...,y[p]>=cp} where the
   cutoff points cp are given in the vector c, then a count
   variable is updated by one. cf. Y.L.Tong- Monte Carlo
   integration method in 'THE MULTIVARITE NORMAL DISTRIBUTION'
   pp 185-189)  */

/** GENERATING 100000 OBSERVATIONS**/
```

```
m=100000;
c=j(p,1,48);

count=0; countlb=0;  countub=0;

  do r=1 to m;
     run multnor(ystar,ystarlb,ystarub,mu,mulb,muub,sigma,p);
     if all(ystar>=c) then count=count+1;
     if all(ystarlb>=c) then countlb=countlb+1;
     if all(ystarub>=c) then countub=countub+1;
  end;

prob[k] = count/m;
problb[k]=countlb/m;
probub[k]=countub/m;

end;      /* k loop */
print x;
print prob problb probub;

/**END OF PROGRAM**/
```

D4 GIBBS SAMPLING VIA REJECTION SAMPLING

THIS PROGRAM USES THE METHOD OF REJECTION SAMPLING WITH SQUEEZING TO INVOKE THE GIBBS SAMPLER THAT WAS IMPLEMENTED ON OUR DSMA1SX1 DATA. THE VARIABLES THAT WERE FOUND TO BE IMPORTANT IN CHAPTER TWO WERE USED HERE, NAMELY MATHEMATICS (X3), BIOLOGY (X4), PHYSICS X(5) AND THE NED (d2) VARIABLE (INCLUDING AN INTERCEPT TERM)

```
libname in 'c:\robert';
data math;
set in.mp2;
if z1= '.' then delete;
if z1= '0' then delete;
if z1 ge 48 then y=1;
else y=0;
if ma=4 then d1=1;
else d1 = 0;
if ma=6 then d2=1;
else d2 = 0;
keep y x1 x2 x3 x4 x5 d1 d2;

/* Non-adaptive Rejection sampling program with squeezing */

/* Assume density function f(x) from which we want to draw
   an observation is log-concave; i.e
            h(x)=ln g(x) propto ln f(x)
   is concave. This assumption then ensures that the support
   of x is in a finite interval [xlb,xub]                    */

proc iml worksize=1000;
 reset noprint;

/* This function routine computes h(x) for a fixed value of x */

start hx(s) global(beta,x,y,n,j,p);
    sum=0;
    beta[j]=s;
    betao= j(p,1,0)          ;
    sigmao=10000*i(p)          ;
      do i=1 to n;
        sum = sum + y[i]*x[i, ]*beta - log(1 + exp(x[i, ]*beta));
      end;
    val=-0.5*(beta-betao)'*inv(sigmao)*(beta-betao) + sum;
    return(val);
finish hx;
```

```
/* This function routine computes h'(x) for a fixed value of x */

start hpx(s) global(beta,x,y,n,j,p);
 sum1=0;
 IP=I(p);
 e=IP[ ,j];
 beta[j]=s;
 betao= j(p,1,0)     ;
 sigma0= 10000*i(p)        ;
 do i=1 to n;
   sum1=sum1+y[i]*x[i,]*e-inv(1+exp(x[i,]*beta))*exp(x[i,]*beta)*x[i,j];
 end;
    val=-e`*inv(sigma0)*(beta-betao) + sum1;
    return(val);
finish hpx;


/* MAIN PROGRAM */
use math;
read all var{y x1 x2 x3 x4 x5 d1 d2};
n=nrow(y);

x=j(n,1,1) || x3 || x4 || x5 || d2;
p=ncol(x);

/**we used the parameter estimates obtained in chapter 2
   for the starting values of beta                              **/

beta={1.73, -0.64 , -0.44 , -0.39 , 1.18};

m=10; G=1000;

do nobs=1 to G;

do count =1 to m;
/** Implement Gibbs sampler m times to obtain one observation**/

   do j=1 to 5;

/** Initialization step: Specify xstart and compute s1 and s3 **/
/** for xubd and xlbd we used the parameter estimates obtained
    in chapter 2 plus or minus 4 standard deviations            **/

     xlbd={0.68 , -1.35 , -1.12 , -1.12 , -0.15} ;
```

```
xubd={2.78 , 0.067 , 0.22 , 0.33 , 2.52};
xlb=xlbd[j] ; xub=xubd[j];
xstart=(xlb+xub)/2;
s=j(3,1,0); u=j(2,1,0); v=j(3,1,0);

h=hx(xstart); hp=hpx(xstart);

if hp=0 then do until (hp>0);
xstart=(xstart + xlb)/2;
hp=hpx(xstart);
end;

if hp > 0 then do;
  s[1]=xstart;
   do while (hp>=0);
     xstart=(xstart + xub)/2;
     hp=hpx(xstart);
  end;
s[3]=xstart;
end;

else if hp < 0 then do;
s[3]=xstart;
  do while (hp<=0);
    xstart=(xstart + xlb)/2;
    hp=hpx(xstart);
  end;
s[1]=xstart;
end;


/* Compute tangential point s2 and chord-intersection points u1
   and u2  */

hs1=hx(s[1]) ; hs3=hx(s[3]); hps1=hpx(s[1]); hps3=hpx(s[3]);

s[2]= (hs3-hs1-s[3]*hps3 +s[1]*hps1)/ (hps1-hps3);
hs2=hx(s[2]);
hps2=hpx(s[2]);

u[1]=(hs2-hs1-s[2]*hps2+s[1]*hps1)/(hps1-hps2);
u[2]=(hs3-hs2-s[3]*hps3+s[2]*hps2)/(hps2-hps3);

d1=hs1-s[1]*hps1; d2=u[1]*hps1; d3=xlb*hps1;
v[1]=exp(d1)*(exp(d2)-exp(d3))/hps1;
```

```
    d1=hs2-s[2]*hps2; d2=u[2]*hps2; d3=u[1]*hps2;
    v[2]=exp(d1)*(exp(d2)-exp(d3))/hps2;

    d1=hs3-s[3]*hps3; d2=xub*hps3; d3=u[2]*hps3;
    v[3]=exp(d1)*(exp(d2)-exp(d3))/hps3;

/* Repeat until an observation has been generated or a maximum
   number of iterations have been completed */

    output='none';
    max=1000;
   iter=0; seed=0;
   do while ((output='none') & (iter< max) );

/* generate discrete variable z=1,2 or 3 */

     uran=ranuni(seed);
     d1=v[1]/sum(v);
     d2=(v[1]+ v[2])/sum(v);
    if uran<=d1 then z=1;
    if uran>d1 & uran<=d2 then z=2;
    if uran>d2 then z=3;

/* generate sample value (obs) from envelope function */

      ustar=ranuni(seed);
      if z=1 then do;
      d1=u[1]*hps1; d2=xlb*hps1;
      d3=ustar*exp(d1)+(1-ustar)*exp(d2);
      obs=log(d3)/hps1;
    end;

    if z=2 then do;
       d1=u[2]*hps2;
       d2=u[1]*hps2;
       d3=ustar*exp(d1)+(1-ustar)*exp(d2);
       obs=log(d3)/hps2;
    end;

    if z=3 then do;
      d1=xub*hps3; d2=u[2]*hps3;
      d3=ustar*exp(d1)+(1-ustar)*exp(d2);
      obs=log(d3)/hps3;
    end;
```

```
/* Define envelope  function co-ordinates for
   acceptance-rejection step of generated observation s */

    infty=100000;
    if obs< u[1] then
    huobs=hs1+hps1*(obs-s[1]);

    if obs>=u[1] & obs<=u[2] then
    huobs=hs2+hps2*(obs-s[2]);

    if obs>u[2]  then
    huobs=hs3+hps3*(obs-s[3]);

/* Define  squeezing function co-ordinates for
   acceptance-rejection step of generated observation s */

    if obs< s[1] then
    hlobs=-infty;

    if obs>=s[1] & obs<=s[2] then
    hlobs=obs*(hs2-hs1)/(s[2]-s[1]) + (s[2]*hs1-s[1]*hs2)/(s[2]-s[1]);

    if obs>s[2] & obs<=s[3] then
    hlobs=obs*(hs3-hs2)/(s[3]-s[2]) + (s[3]*hs2-s[2]*hs3)/(s[3]-s[2]);

    if obs> s[3] then
    hlobs=-infty;

/* generate a uniform random variable w~U(0,1) and perform two-stage
   acceptance-rejection test */

   d1=hlobs-huobs;
   w=ranuni(seed);

   if w<=exp(d1) then do;   /* Accept s as an observation from f(x) */
     b=obs;
     output='yes';
   end;



   if w> exp(d1) then do;
     hobs=hx(obs);             /* Compute h(x) at the point obs */
     d2=hobs-huobs;
```

```
    if w<=exp(d2) then do;
      b=obs;
      output='yes';
    end;
  output='none';
  end;
  iter=iter+1;
 end;

 beta[j]=b;
 end;              /* j loop */


betap=beta`;

if count=1 then betall=betap;
  else betall=betall//betap;

 end;              /*count */

/* output last observation as coming from true density */


     if nobs=1 then   betamat=betap;
     else betamat=betamat//betap;
end;  /*nobs*/
print betamat;
/**betamat is a 1000 x 5 dimensional matrix**/

/**END OF PROGRAM**/
```

REFERENCES

Agar, D. 1992. Evaluating Academic Support Programmes — what have we learnt in the last six years? South African Journal of Education 12: (2) 93-100.

Buse, A. 1982. The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note. Journal of the American Statistical Association 36: (3) Part 1 153-157.

Dobson, A.J. 1990. An Introduction to Generalized Linear Models. London: Chapman & Hall.

Flockemann, J., Frame, J. & Kernick, G. 1993. Assessment of Academic Potential for Selection into Emec Programme. A working paper.

Fresen, J.L. & Fresen, J.N. 1987. Correlation between Matric Symbols and First Year University Results: Business Statistics at the University of the Western Cape as an example. South African Journal of Science 83: 492-494.

——— 1993. The Probability of Success at University: A Methodology. Draft Documents.

Gallant, A.R. 1987. Nonlinear Statistical Models. New York: Wiley.

Gelfand, A.E. & Smith, A.F.M. 1990. Sampling-Based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association 85: 398-409.

Geman, S. & Geman, D. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. Transactions on Pattern Analysis and Machine Intelligence. PAMI 6: (6) 721-744.

Gilks, W.R. & Wild, P. 1992. Adaptive Rejection Sampling for Gibbs Sampling. Applied Statistics 41: (2) 337-348.

Graybill, 1976. Theory and Application of the Linear Model. Massachusetts: Duxbury Press.

Harris, B. & Soms, A.P. 1980. The use of the Tetrachoric Series for Evaluating Multivariate Normal Probabilities. Journal of Multivariate Analysis 10: 252-267.

Hosmer, D.W. Jr & Lemeshow, S. 1989. Applied Logistic Regression. New York: Wiley.

Huang, D.S. 1970. Regression and Econometric Methods. New York: Wiley.

Johnson, R.A. and Wichern, D.W. 1988. Applied Multivariate Statistical Analysis. 2nd edition. London: Prentice-Hall International.

Knuiman, M.V. & Speed, T.P. 1988. Incorporating Prior Information into the Analysis of Contingency Tables. Biometrics 44: 1061-1071.

Lawless, J.F. & Singhal, K. 1978. Efficient Screening of Non-normal Regression Models. Biometrics 34: 318-327.

Mitchell, G. & Fridjohn, P. 1987. Matriculation Examinations and University Performance. South African Journal of Science 18: 555-559.

Nyquist, H. 1991. Restricted Estimation of Generalized Linear Models. Applied Statistics 40: (1) 133-141.

Potter, L.S. & Jamotte, A.N. 1985. African Matric Results: Dubious Indicators of Academic Merit. Indicator South Africa 3: (1) 10-13.

Press, S.J. 1989. Bayesian Statistics Principles, Models and Applications. New York: Wiley.

Rao, C.R. 1965. Linear Statistical Inference and its Applications. 2nd edition. New York: Wiley.

Rutherford, M. & Watson, P. 1990. Selection of Students for Science Courses. South African Journal of Education 10: (4) 353-359.

Schervish, M.J. 1984. Multivariate Normal Probabilities with Error Bounds. Applied Statistics 33: 81-87.

Smyth, G.K., Knuiman, M.W., Thornett, M.L. & Kiiveri, H. 1990. Using the EM Algorithm to Predict First Year University Performance. Australian Journal of Education 34: (2) 204-224.

Sochet, I. 1986. Manifest and Potential Performance in Advantaged and Disadvantaged Studies. Unpublished Ph.D. Thesis. Johannesburg: University of the Witwatersrand.

Tong, Y.L. 1990. The Multivariate Normal Distribution. London: Springer-Verlag.

Van Wyk, J.A. & Crawford, J.L. 1984. Correlation between Matric Symbols and Marks Obtained in a first-year Ancillary Physics Course at the University of the Witwatersrand. South African Journal of Science 80: 8-9.

Zellner, A. 1970. Introduction to Bayesian Inference in Econometrics. New York: Wiley.