# AN EVALUATION OF PRE-ENTRY PERFORMANCE PREDICTORS FOR BACHELOR OF COMMERCE STUDENTS AT THE UNIVERSITY OF NAIROBI

By

JOHN O. ONUONG'A

A Management Science Research Project Submitted in
Partial Fulfilment of the Requirements for
the Award of the M.B.A. Degree
Faculty of Commerce,
University of Nairobi

June 1988

i

This Research Project is my original work and has not been presented for a degree in any other University

JOHN O. ONUONG'A

This Research Project has been submitted for Examination with my Approval as University Supervisor

MR. GITURO WAINAINA

To my parents, Luka Onuong'a and Sabina Gesare

TABLE OF CONTENTS

# LIST OF TABLES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to extend my profound gratitude to the many individuals who directly or indirectly contributed to the successful completion of this research project.

I wish to extend special thanks to my supervisor Mr. Gituro Wainaina who not only aroused my interest in this area but also provided much of the needed suggestions, comments and general guidance in the production of this report.

I must also thank the Dean, Faculty of Commerce, Professor J.H. Kimura for allowing me to have access to the Faculty records which were the source of virtually all the necessary data.

I'm very grateful to the MBA II class, 1987/88 academic year and the entire academic staff for the useful suggestions made during the presentation of the proposal which helped to refine the study.

To my entire family, especially my brother Capt. P.J. Migisi, I must say thanks a lot for the moral and material support, the patience and appreciation they have accorded me during my studies.

Lastly I must sincerely thank Mrs. Joyce Muthama having kindly accepted to type this work within the set deadline. I must however take full responsibility for any errors or shortcomings in this research project.

ABSTRACT

This research project reports on the results of an investigation into the relationship between certain Bachelor of Commerce student attributes and their performance in the final year of study at the University of Nairobi.

The total sample used was 300 students divided into three sub-samples of 100 students each. Each sample of 100 students were selected from a period of four years. So the total period for the study was twelve years from 1973/74 to 1984/85 academic years inclusive.

The scores obtained by the students in their final year of study in the Bachelor Commerce programme was used as a measure of performance and as the criterion variable. The attributes against which their performance was measured (the predictor variables) were the scores attained in the subjects taken at Ordinary (secondary school) level and two demographic variables, age, and sex.

The results show that the attributes that had the strongest association with performance at the University are scores in Mathematics, age, Commerce-related subjects, Science subjects and O-level aggregate score. Those that seemed to lack significant association with University level performance were,

sex, scores in English Language and Literature, General Sciences, other languages, and the Liberal Arts subjects.

Many of these predictor variables were also found to be strongly correlated.

The main statistical tools that were used in analysing the data were multiple, stepwise regression analyses and correlation analysis.

Though there were limitations in the study, the findings from this research still have strong implications as concerns the admission criteria into the Faculty of Commerce programme.

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Studies pertaining to the prediction of academic performance can be put into various uses. Three most important ones are:[1]

- to show the extent to which a criterion behaviour pattern can be predicted;
- provide information for theory-building about possible determinants of the criterion behaviour pattern;
- provide evidence regarding the predictive validity of the test or tests that are correlated with the criterion behaviour.

Whereas a study could be used to provide all the above information, not all predictive studies will of necessity serve all the three purposes. Therefore it is the specific purpose to which a study is intended to be used that will determine which of the above sets of information will be relevant.

The current study, which is predictive in nature is mainly an attempt to identify possible determinants

---

1. Gatumu, H.J.N.: A study of predictive validity of O-level, A-level and an aptitude test in relation to the performance at the University of Nairobi. (Unpublished M.Ed. Thesis, U.o.N. 1976).

of academic performance in the Faculty of Commerce, University of Nairobi (U.o.N.). The ultimate aim is to establish which subjects should be emphasized on in admitting students into the Faculty.

Currently, the country is undergoing a change in the education system from 7-6-3 to 8-4-4, referring to the number of years spent in Primary, Secondary/ High schools and University levels respectively. Such an event has implications on the admission criteria into the various public universities in the country.

The Faculty of Commerce is currently among the most competitive faculties in the U.o.N. Over the years the Faculty has always received more applicants than could be admitted. The selection therefore has gone beyond consideration of the minimum entry requirements.

For any selection criteria to be effective, it is imperative that it bears a relationship with the students' performance at the University. In order to be effective, the admission criteria should not be viewed as a selection criteria per se but rather, as a measure which would provide supplementary information to identify those candidates unlikely for whatever reason, to complete successfully a college course (Astin 1971). Though there are various methods of deriving a relationship between admission criteria and performance, empirical analyses are the most widely used since they are testable as opposed to such simple methods as "educated" or intuitive guesses. The most widely applied empirical methods in determining pre-entry student attributes that are relevant to academic performance include correlation factor, discriminand and multiple regression analyses. All these have been widely applied in studies at various universities and other educational institutions. Some of these include studies by Dockweiler and Willis (1984), Dunn and Hall (1984), Ingram and Petersen (1987), Astin (1971), Delaney, Keys, Norton and Simon (1979), Gatumu (1976) and Frakes (1977).

In the current study only two of the above mentioned statistical tools will be utilized, namely, multiple regression analysis and correlation analysis.

The interest in this study was stimulated by two main factors. First and foremost by the results (findings) of past investigations in the Faculty's examination performances. The findings that particularly aroused the interest of the researcher are those by a committee that was set up in 1979 by the Senate to investigate into the causes of high failure rates in some of the Faculty's examinations.[2]

---

2. Kohler, Dr. D.: Ad Hoc Committee's report on the high failure rates in some of the Faculty's examinations - 1980.

The second factor was the change in the Kenyan education system. Because the admission criteria under the 7-6-3 system had been found by the above committee to be in a way ineffective the researcher thought that such pitfalls could be prevented if some empirical method was used as a basis of setting an admission criteria.

In the past, in some years, there have been abnormally high failure rates in some of the Faculty's examinations. Sometimes, a large number of students have had either to be discontinued, required to repeat, sit supplementary examinations or have their marks adjusted. At one time the failure rates were so high that the Senate expressed its concern by setting up the above-mentioned committee to look into the causes of such failures and suggest possible remedies.

One of the major findings of the committee was that the Advanced (A) level total score did not seem to have any significant correlation with performance at the University level, whereas it is the same results that are given most weight in selecting students into the Faculty of Commerce. The issue that arises then is whether the reasons for the high failure rates lies with the admission (selection) criteria or some other factors. From the Dr. Kohler Report (1980), there are indications that the selection criteria is partly responsible for the failure rates in the Faculty.

Now that the Faculty (and the University as a whole) has to come up with entry requirements for the 8-4-4 system, there is a need to try and find out which Ordinary (O) level student attributes might affect student performance in the Faculty. This would show which subjects should be focused on in setting entry requirements.

The researcher, being a graduate of the Faculty has a conviction that other than 0-level Mathematics which forms part of the current entry requirements, other factors may also be important in predicting performance in the Faculty, hence an investigative study would reveal any such factors that could be of relevance in the 8-4-4 programme.

It was also felt that there was need to undertake this study because the only similar study done in this area was geared towards a specific report to be presented to the Senate, thus the investigation did not in any way support the findings with any theoretical background. Further, the sample taken of 104 students from the 1976 intake of students might not have been representative due to the problem of stationality.

## 1.2 Statement of the Problem

Each student aspiring to join (or is already admitted into) the University would like to know his/

her chances of successfully completing the programme. Astin (1971), has aptly summarized the student's problem in one paragraph:

> "Nearly every student planning to go to college is concerned about how well he will do once he gets there. This concern relates not only to the grades that he is likely to receive, but also to his chances of staying in college through graduation."[3]

Apart from the students, the University as a whole and the Faculty in particular would like to ensure that the admitted students will successfully complete the courses offered. One way to achieve this aim is to set minimum entry requirements that if met, will ensure that the selected students have a high probability of passing. At the same time, such entry requirements should be able to identify those students who are unlikely to successfully complete the course.

However, arriving at such a selection criteria is not practically easy. This is evident from the findings of a previous investigation into the performances in the Faculty which concluded that the current criteria is not adequate (effective). This problem is compounded by the present change in the education system which will require new entry

---

3. Astin, A.W.: <u>Predicting Academic performance in Colleges</u>. (The Free press, N.Y., 1973), pg. 3.

requirements. The question that arises is whether it is possible to decide ex-ante, which particular pre-entry attributes are likely to be predictive factors of students' performance in the Faculty of Commerce under the 8-4-4 programme. This study is an attempt to answer this question by using the approach suggested by Harnett and Murphy (1985) who state that:

> "...given sufficient historical information on the enrollees' characteristics and their eventual graduate school record, a mathematical relationship can be determined among the variables. The relation, allowing for uncertainty, is a regression equation that can be used to make predictions for other enrollees"[4]

1.3 Objectives of the Study

The study is basically aimed at identifying the variables that could be used as a basis of a predictive model for the performance of students in the Faculty of Commerce before they are admitted. Hence the objectives of this study are threefold:

(1) to identify those pre-entry variables that have a significant relationship with eventual performance and hence could be used to predict student performance in the Faculty;

(2) identify the nature of the prediction model, if any, that could possibly be

---

4. Harnett, D.L. and Murphy, J.L.: Statistical Analysis for Business and Economics (3rd Ed. Addison Wesley Publishing Co., 1985) pg. 633.

used to predict student performance in the Faculty;

(3) based on first and second objectives above, suggest possible areas of emphasis in formulating a selection criteria under the 8-4-4 programme.

1.4 Importance of the study

The findings of this study will be useful to a number of individuals and/or institutions. These include:

the Faculty of Commerce in deciding on an effective admission (selection) criteria and offering advice to students already admitted (especially those identified as weak);

any other university intending to introduce a Faculty of Commerce, in setting up the minimum entry requirements;

for careers masters at secondary schools in offering guidance and advice to students intending to join the Faculty of Commerce;

as a basis for further research.

1.5 Research Design

1.5.1 The Population

The population for the study comprises of all the students who were admitted into the Faculty of Commerce in the academic years 1974/75 to 1984/85, a period of ten years.

### 1.5.2 The Sample

To decide on the appropriate sample size, the following formula was used:

$$n = \frac{z^2 s^2}{E^2}$$

where: n = sample size
z = reliability factor
s = sample standard deviation
E = maximum tolerable error.

Ordinarily, the correct formula to use in determining the sample size given a finite population (as in the present case) is:

$$n = \frac{Nz^2 s^2}{E^2(N-1)+z^2 s^2}$$

where N = population size

But because the populations were different in all the years, the formula cannot be used.

In this study, the reliability factor was set at 95% and so z = 1.96. Since the population standard deviation is unknown it is necessary to use a sample standard deviation. To get this statistic, a pilot study was carried out by selecting fifty students from the 1979/80 and fifty from the 1980/81 graduating classes and getting the respective Grade Point Average

(GPA) means and standard deviations. The results obtained were:

$\bar{x}_{79/80} = 56.49$ $\bar{x}_{80/81} = 57.14$

$s_{79/80} = 5.28$ $s_{80/81} = 4.36$

$E(s) = (5.28 + 4.36)/2 = 4.82$

Since $n \geq 30$ these sample parameters could be taken to be unbiased estimators of the population parameters (Kane D.J. 1968: pg. 196, and Daniel and Terrell 1983: pg. 128).

The maximum tolerable error was set at 0.98 which is normally arrived at by dividing the reliability factor by two (Daniel and Terrell 1983: pg. 172). Applying the above statistics, the appropriate sample size was determined as:

$$n = \frac{(1.96)^2(4.82)^2}{0.98^2} = 92.9 \text{ rounded up to } 100.$$

This sample was then divided equally over a period of four years, twenty five from each year.

Still, using only one sample might produce results that are unique to only that sample. To overcome this problem of stationality, it is advisable to test more samples and identify those variables that appear consistently significant. Accordingly, two different samples of 100 students each were used to develop the model. Therefore in total, eight years were used. To cross-validate the model developed from these initial two samples, a further 100 students were used.

### 1.5.3 Data Collection

In the current study, it would have been meaningless to include the students' A-level results as variables in light of the fact that in 1990 the A-level will be phased out. However it is necessary to include A-level Economics as a surrogate for O-level Economics because currently the subject is not offered in O-level but under the 8-4-4 it will be offered in the four years of secondary level education.

For each of the students in the sample, the scores were obtained for all the subjects taken at O-level and the score in A-level Economics. The students' final year Grade Point Average (G.P.A.) were used, as a measure of performance. All the above scores were obtained from the Faculty of Commerce records office.

### 1.5.4 Data Analysis

The data was analysed using regression and correlation analysis. In this respect, the Statgraphics computer package was used on the Compaq micro-computer.

1.6 <u>Scope of the Study</u>

This project consists of five chapters. The first chapter - introduction gives a brief background on the subject matter of the study, the statement of the problem, objectives of the study, importance of the study and the research design.

The second chapter is on the literature review that is pertinent to the understanding of the study.

Chapter three is a description of the statistical tool that was used to analyse the data.

The fourth chapter is devoted entirely to data analysis and findings from the study.

The fifth and final chapter is the conclusion which is a discussion of the findings, their implications, limitations of the study and lastly, suggests areas for further research from the current study.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Background to the Faculty of Commerce

The Faculty of Commerce was started at the University of Nairobi (then the Royal College of East Africa) in the academic year 1956/57 during which time it was referred to as the Faculty of Professional Studies. The actual Faculty of Commerce as it is known and structured today was established in 1964 with fifty students. At this time, it had four departments, namely Accounting, Business Administration, Law and Home Economics. In 1970, the Department of Law became a fully-fledged Faculty while Home Economics was absorbed into the Faculty of Science. Over the last two and half decades, the Faculty population has steadily grown both in terms of academic staff and under and post-graduate students.

Currently the Faculty offers three options to students enrolled in the undergraduate programme. These options are Accounting, Marketing and Insurance. However, plans are underway to introduce other options such as Management Science, Banking, Finance and Cooperative Management. The Bachelor of Commerce undergraduate programme is undertaken over a period of three years of full time study. During the first year, common courses are taught to all the new students. These are mainly introductory courses aimed at acquainting the fresh undergraduates with the entire programme and enabling them to identify potential areas of specialization. In their second year, the students choose specified areas of specialization which will constitute their final degree option. Further introductory and intermediate core courses relevant to specific specialization areas are offered to the students. During the final year of study, a common and integrative course, Business Policy and Decisions is offered to all the students in addition to the more advanced core courses along with a number of diversified but relevant elective courses (see appendix A ).

The minimum entry requirements for admission into the Faculty of Commerce as per the 1986/7 University Calendar are a credit or better (grades 1-6) in Ordinary ('O') level Mathematics, five or more passes at O-levels and passes (grades A-E) in at least any two principal subjects at Advanced ('A') level. However in actual practice, the entry requirements are more stringent than these and depend mainly on the A-level results. Currently the minimum number of points one can get in order to be admitted are thirteen. In A-level results, points are awarded to the various grades as follows: 6 points for grade A, 5 for grade B, ......, 2 for grade E and 1 point for a subsidiary pass. Candidates are then ranked according to their total score in all the subjects taken.

The actual selection into the University is performed by a selection committee composed of the Deans of faculties and chaired by the Vice-Chancellor. Each faculty has a specified maximum number of students it can admit. In the first round of selection, those candidates whose first choice was Commerce are selected in order of total point score. However, because large numbers of students have equal point scores, several places may be left unfilled. For example, if Commerce has 180 places and 120 students have scores of 13 or better and 100 have scores of 12, then only the top 120 students will be selected. In the second and subsequent rounds, second, third and sometimes fourth choices of the candidates are considered to see whether they qualify for admission to other faculties.

When considering the subjects that have been taken by the students at A-level, the Faculty of Commerce does not distinguish between any of the subjects. Therefore, for entry into Commerce, a grade B in Christian Religious Education carries the same weight as a grade B in Mathematics. However, for a short period, during the 1973/74 academic year, the Faculty of Commerce used a different weighting scheme for A-level subjects:

| | |
|---|---|
| Mathematics | x 3 |
| Economics | x 2 |
| Other approved subjects | x 1 |
| Art | x 0 |

The scheme was however abandoned because it was felt that the Faculty "should try to test for basic intelligence rather than A-level in a particular area." (Dr. Kohler, 1980).

Once a student has been admitted into the Faculty, he/she is required to pass all the courses taken. A pass implies a student must score 40 or more in a subject. However, depending on how many courses one fails in the first sitting in any academic year, he/she could be required to sit supplementary examinations, be discontinued or asked to repeat the year. (For details see appendix B ). The marks attained in the first year of study do not count for the final degree Grade Point Average (G.P.A.). This G.P.A. is composed of the student's performance in the second and third years where the former gets a weight of 1/3 while 2/3 go to the latter.

## 2.2 Factors that predict academic performace

Many authors have undertaken studies aimed at identifying the variables or factors that could be used to predict student performance either at university level or for some particular examination, professional or academic. When undertaking such studies, the inherent assumption by the researchers is not that the factors identified actually have a cause-effect relationship with subsequent performance but rather because:

> "...in using such variables as... it is not because they are thought or believed to be related to the eventual degree performance, but rather, because all these examinations are supposed to test what the final examination will test, that is, recall, intelligence, expression etc."[5]

However, apart from the examination-related variables that the above authors have referred to, there could be other qualitative and quantitative (but not examination-related) factors that could predict academic performance. These include such factors as sex, age, communicating ability, integrity and ability to deal with people. For computation purposes however, the variables that have mainly been used for the prediction of academic performance are those though qualitative can be assigned quantitative values.

The need to identify and evaluate academic performance predictors has been emphasized by Williams (1969) who has said that:

> "...the future of the Accounting profession will be determined not only by the opportunities which confront it, but by the intellectual capability, knowledge and skills of its members.
>
> It is important therefore for the Accounting profession to evaluate <u>periodically</u> the characteristics of its

---

5. Ibid.: pg. 635.

recruits to determine
if those who seek to enter
into its ranks are equipped
with the attributes
necessary to deal
effectively with the
increasingly complex
problems of the
profession."6

The author suggests that such periodic evaluations will assist those responsible for the profession's recruiting activities in determining the effectiveness of their recruiting programmes. "In addition, educators may find such an analysis useful in counselling students."

The above proposition by Williams is applicable not only in the Accounting profession (which is a branch of Commerce) but virtually in all professions. The author however raises the issue of a periodic evaluation as opposed to a once-a-lifetime evaluation because the latter could be overtaken by events. Or in other words, predictor variables might be changing with time for one reason or another.

Baldwin and Howe (1982), carried out a study in an attempt to relate secondary school level study of Accounting and subsequent performance in the introductory level college course in Accounting. Thus they sought to answer the question - does high

---

6. Williams, D.Z.: A Profile of C.P.A. candidates (The Accounting Review - Jan. 1969) pg. 153.

school exposure to Accounting benefit the university student? They used two samples - one of students not previously exposed to Accounting in high school and another group that had prior Accounting knowledge. They then compared the performance of the two groups at college level. Their primary conclusion was that college students who had a bookkeeping course in high school perform no better in university Accounting courses than those who had not been previously exposed to such a course. The results of their research even implied that prior study of bookkeeping may be dysfunctional to the student.

The authors then suspected that the student with previous bookkeeping knowledge may be over-confident at the beginning of the college course as they expect the course to be easy. This leads them to feel they can "sail" through the course. However, with time, new and more rigorous topics are introduced but these students make no extra effort as is required by the course and thus perform poorly, much as the other students may have done at the beginning of the course. The result is that their relative performance falls considerably. On the other hand, those students with no prior Accounting knowledge performed relatively more poorly early in the course and they either drop the course or begin to work more diligently. Thus when the more rigorous material comes later in the course, this group of students is usually working at a high level of proficiency. Little wonder then, in the authors' words, "that they outperform the oldmasters from the high school bookkeeping course" (pg. 625).

The authors thus recommend that "forceful counselling is therefore necessary especially to those with prior accounting knowledge."

There is a clear parallel between the above study and the current one. In the 8-4-4 system, Accounting, Commerce and Economics are among the subjects that will be offered during the four years of secondary level education. There is a possibility that this group of subjects could be significantly correlated with university performance. If this be the case, utmost care will have to be taken in considering these group of subjects as part of the selection criteria. And whether these subjects are included in the admission criteria or not, it might become necessary to counsel students who have taken these subjects prior to university education against assuming, and wrongly so, that they have an advantage over the "non-starters".

Astin, A.W. (1971) from his study on "Predicting academic performance in colleges" concluded that there are three major predictors of academic success in colleges. These are:[7]

---

7. Op. cit.

academic record in high school

tests of academic ability (aptitude tests)

sex.

The tests of academic ability are administered for the prime purpose of gauging the suitability of each candidate for college education. It is imperative therefore that they are administered before the student(s) are admitted to colleges in the United States of America, where the above study was carried out. In Kenya, such pre-university entrance examinations do not exist. This does not however, imply that they have been found to be irrelevant. On the contrary, a study that was carried out by Gatumu (1976)[8] did conclude that such aptitude tests have a reasonably high correlation with performance in certain faculties in the University of Nairobi. Unfortunately, Commerce was not among the faculties selected for his study and so, the relevance of aptitude tests in the Faculty of Commerce cannot be ruled out.

In his overall conclusion, Astin hypothesised that the students who would be expected to perform better are "those who have been honoured for excellent performance at high school; those who attended private schools and those who rated themselves high on academic ability and the drive to achieve." Likewise, the students most likely to

---

8. Op. cit.

perform worse than expected are those that turned in papers and assignments late, came late to class, made wisecracks in class and those who went to movies and discos frequently.

Gatumu H.J.N. (1976), having used samples from different faculties for his study, concluded that different variables appeared to be more significant predictors in one faculty than in another. Intuitively, this finding is to be expected because the course content in each faculty will usually be structured on the assumption that students will have taken particular subjects (gained particular knowledge) or attained above a certain score at secondary school level. This therefore suggests that the various faculties should set their own minimum entry requirements distinctly.

Dockweiler and Willis (1984)[9], using correlation analysis, concluded that there were three major factors that were strong predictors of success or failure in the Faculty of Accounting, University of Missouri. These three factors were, the students' overall GPA score prior to entering the Accounting programme; the grades in the first and second introductory Accounting courses taken prior to entry into the Faculty. In the same study, the authors

---

9. Dockweiler, R.C. and Willis, C.G.: On the use of Entry Requirements for Undergraduate Accounting Programs" (<u>The Accounting Review,</u> July 1984, Vol. LIV, No. 3).

found out that College Aptitude Test scores also had positive correlation with performance but this was substantially lower than the correlation between the first and second grades in introductory Accounting courses and eventual performance.

Because correlation analysis ignores the joint contribution of predictor variables, the authors went further to use stepwise multiple regression to check for the contribution of the multiple variables taken together. From an original set of eleven predictor variables, the conclusion was that high school GPA and the 2 introductory Accounting courses were strong predictors of performance in the third and final year in college.

By way of validating their results the researchers carried out a discriminant analysis using these three variables. They were found to be highly significant. Age was also found to be a significant predictor but the authors caution, and correctly so, that it would be illegal and unfair to use age as an entry requirement. Therefore they dropped this variable and used the three other variables to predict and achieved an 81% prediction success. They therefore concluded that the best selection criteria would be a three-factor model using the three significant predictor variables.

In evaluating the findings of their study, Dockweiler and Willis concluded that with the 81%

prediction success, the study was a great success since "any set of admission standards can be expected to result in some screening errors",[10] that is to say, some students who would have been successful will be screened out whereas others who are admitted will subsequently prove to be unsuccessful - which they referred to as Type I and Type II screening errors respectively.

The finding that introductory Accounting courses taken prior to university admission were significant predictors of performance is in contrast to the conclusions by Baldwin and Howe (1982)[11] who found the opposite to be true. Indeed what these two studies serve to highlight is that research findings cannot be overgeneralized. Perhaps if Baldwin and Howe had used correlation or regression analysis then it would be possible to know if prior Accounting knowledge bears a negative or positive relation to performance.

The overall conclusion and contribution of the Dockweiler and Willis study was that whereas there was nothing wrong with considering one most important variable when screening enrollees, the process could be improved by considering additional variables which depict high correlations with the eventual performance.

---

10. Ibid. pg. 502.

11. Op. cit.

In attempting to relate various predictor variables and candidate performance, Dunn and Hall (1984)[12] arrived at similar conclusions as those by Dockweiler and Willis. The former authors sought to find out the relationship between Certified Public Accountants (CPA) Examination candidate attribute and candidate performance. The statistical tool used was regression analysis from which findings indicated that Accounting GPA at college was the single most important attribute in predicting a candidate's eventual score. Other variables that appeared to be important were hours of self-study, aptitude tests and type of school attended. These variables appeared to have consistently significant associations with examination performance in the series of regressions that were performed. However work experience, age and non-Accounting subject scores did not have any significant relationship with performance.

Dunn and Hall seem to be the only authors who did a complete analysis because, instead of performing one regression, they opted for a series of regressions to determine the variables that appeared to be consistently significant. They even went further to test the regression assumptions. These two approaches will be used in the current study.

---

12. Dunn, W.M. and Hall, T.W.: An Emphirical analysis of the relationship between CPA Examination candidate attributes and candidate performance (The Accounting Review, Vol. LIV, No. 4 Oct. 1984).

Ingram and Petersen (1987)[13] attempted to evaluate American Institute of Certified Public Accountants (AICPA) to test scores for predicting the performance of Accounting majors. The predictor variables were AICPA aptitude tests, GPA and first level AICPA total score, whereas the criterion variable was the score in upper division Accounting courses at the University of Iowa. By far the most important predictor was the students' GPA for the first two years of college. None of the other variables provided a significant improvement in predictive power. To cross-validate their model, the author calculated the average correlation coefficient between the actual and predicted values from fifty validation groups and this was found to be on average 0.880 with a standard deviation of 0.021, which is reasonably close to the correlation coefficient of the original sample used to develop the model (0.912).

It should, however, be noted that the most important variable in the above study were scores attained after admission into college, hence such variables are not of much use if one is interested in evaluating a student before actual admission.

---

13. Ingram, R.W. and Petersen, R.J.: An Evaluation of AICPA tests for predicting the performance of Accounting majors (<u>The Accounting Review,</u> Vol.LXII, No.1 Jan. 1987).

Frakes, A.H. (1977), another Accounting educator, using two samples from two different universities (Washington State University and University of Washington), attempted to "unearth" the relationship between "Introductory Accounting objectives and Intermediate Accounting performance". The most important conclusion that is pertinent to the current study was that:

> "...prediction equations developed for use in admission decisions for accounting study beyond the introductory level should not be limited to Accounting achievement test results. General academic ability and demographic variables also should be considered and, in some areas, may be more important predictor variables than achievement test scores."[14]

This conclusion is strongly supported by findings by Dockweiler and Willis (1984) and Dr. Kohler (1980) both of who found age and sex to be significant predictors of performance but cautioned, that it would not only be illegal but also unfair to consider such demographic factors for admission purposes. On the average, Dr. Kohler found that younger students tend to do better than older ones whereas males tend to perform better than females, on the average.

---

14. Frakes, A.H.: Introductory Accounting objectives and Intermediate Accounting performance. (<u>The Accounting Review,</u> Vol. LII, No. 1, Jan. 1977) pg. 209.

Delaney, P.R. et al (1979)[15] tried to relate scores attained in an admission test with eventual performance in Intermediate Accounting. The researchers came up with a multiple correlation coefficient of 0.57 and their conclusion was that there was need for additional explanatory variables. They further suggest that these additional variables might well be demographic variables, similar to those previously suggested by Frakes (1977). These two studies therefore serve as pointers to the importance of demographics in the prediction of academic performance despite the inherent practical problems of using such variables as a basis of student admission. Two demographic variables, sex and age have, however, been included in the current study to see if they have an effect on performance.

In a report tabled by an Ad Hoc Committee chaired by Dr. Kohler, D.[16] that was set up to investigate the high failure rates in some examinations in the Faculty of Commerce, several conclusions were made. Some of the conclusions were quite contrary to some of the above literature already reviewed. In order to facilitate the investigation, the committee used a sample of 104 students from the 1976 intake of students.

---

15. Delaney, P.R. et al.: An Admission test for Intermediate Accounting (The Accounting Review, Vol. LIV, No. 1, Jan. 1979).

16. Op. cit.

Their GPA in the final year was used as a measure of performance since "it was indicative of the performance of both of good and bad students." Several other items of information concerning the same students were gathered and correlations derived between this information and the students' final year GPA. The most important findings were that on average, males do better than females; younger students tend to do better than older ones, O-level Mathematics was a significant determinant of success or failure; O-level science had a strong correlation; and there was a very strong correlation between A-level Mathematics and final year GPA (it was actually the best indicator of the ability to do well in the Bachelor of Commerce programme) and lastly, A-level Commerce (Accounting and Economics) was highly correlated with final performance.

Perhaps the finding that had the strongest implications was that total A-level score, which is the criteria currently used in selecting students into the Faculty was not correlated with final performance. The committee further concluded that of all the variables, the following had the strongest correlation with final year GPA (in order of importance):

A-level Mathematics

A-level Commerce

O-level Mathematics.

Perhaps the reason that A-level total score was not significantly correlated with final performance is the fact that a student might take A-level subjects and score highly but the same subjects are not in any way relevant to the Bachelor of Commerce programme. Thus it becomes important to identify the specific subjects that are correlated with university level performance as a basis for setting up the admission criteria.

Several other authors have carried out similar studies as those reviewed and each of these has come up with a list of predictor variables, the majority of which have already been mentioned. They include Somercet (1968) and the National Foundation for Education Research (1973). In summary, the author who has tried to categorize the various possible predictor variables into some form of uniformity is Astin (1971) who has categorized the variables into four groups:

background characteristics

type of school attended (private, mission, or government)

race

religion

education level of parents

number of family members

family income

students' age

high school achievements

average grade in high school

specific achievements (e.g. won a contest or a scholarship, was a member of a certain society)

future plans

probable field of study in college

probable career choice

highest degree sought

expectations about getting married during or after graduation

concern about finances

interests and personal characteristics

interests in various job activities (e.g. teaching, research, service etc.)

dating patterns during high school

daily activities or behaviour (playing music, disco, studying, drinking, smoking, church activities, etc.).

Even though the above variables were found by Astin to have some predictive ability, generally the additional accuracy in prediction contributed by each of them was insignificant compared to the three major ones earlier mentioned (academic record in high school, aptitude tests and sex) of which academic record in high school was the best single indicator of students' performance at college level. This latter conclusion is similar to that by Gatumu (1976) and Dockweiler and Willis (1984).

It should be noted that even though most of the variables mentioned above could be expected to be significant predictors in the Faculty of Commerce, it is not a foregone conclusion that a variable found to have strong predictive ability in America will have a similar kind of relationship in the Kenyan situation. For example, the conclusion that students who went to private schools would be expected to perform better may not be the case in Kenya.

As is evident from the above studies, it is possible to come up with an almost limitless list of predictor variables that could explain a student's academic performance. However for practical purposes it may not be necessary to include all of the possible variables in the model. In order to reduce the complexities of dealing with large volumes of data which may in any case be adding little explanatory power to the final results, the number of variables may have to be reduced to save computational effort and time. Indeed as Kendall, M.G. points out:

> "...it may be necessary to avoid
> variables that are expensive to
> observe or involve a lot of delay
> in measurement, provided that
> nothing is lost in the purpose
> of the study."17

---

17. Kendall, Sir M.G.: Multivariate Analysis (Griffin Publishing Co., London, 1975) pg. 5.

Daniel and Terrell (1983) further also justify the non-inclusion of certain variables by saying:

> "...Usually we base the decision on both statistical and non-statistical considerations. We may have to omit some variables because the results of statistical analyses cast doubt on their usefulness as predictor, or explanatory, variables. Some writers suggest that under some circumstances variables that fair poorly when subjected to statistical evaluation should remain in the model, either because measurements on them are easily obtained or because the logic of their preserve is so strong."[18]

There are also theoretical reasons that underlie the preference for as few variables as possible. For example, one may wish to reduce the dimensions or number of variables even at the expense of sacrificing some information or transform the data so as to get rid of "nuisance" parameters. An example is the case where some variables might not be quantifiable though they might be important. If such parameters cannot be transformed into quantitatives, then they cannot be included.

But, if there is no a priori reason to believe that a particular variable is insignificant, then it must be included in the model so that there are as many variables as possible, after which statistical methods can be used to eliminate some, if necessary.

---

18. Daniel, W.W. and Terrell, J.C.: Business Statistics: Basic Concepts and Methodology (Third Ed., Houghton Mifflin Co. 1983) pg. 363.

# CHAPTER 3

# THE MODEL

## 3.1 Introduction

The statistical tool that will be used in this study is multiple regression analysis and the related correlation analysis. The general predictive equation derived from the analysis is of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_n X_n + E$$

where $Y^1$ = response variable

$\beta_0$ = a constant

$\beta_i$ = the partial regression coefficient for variable i.

$X_i$ = score on predictor variable i

E = Error term.

The reason why multiple regression and correlation analysis were selected in preference to other statistical tools is that from the literature review on similar studies, it is apparent that these tools are the most widely used. So there is reason to believe that the tools could be appropriate in the current study since it is similar to those reviewed. As Dunn and Hall suggest, in such studies linear models are used

> "...because there is no a priori reason to expect a nonlinear relationship..."[19]

---

19. Op. cit.

In any case a simple plot of the data should show the kind of relationship that exists between the variables.

The assumption underlying the model is that given measurements on a set of predictor variables $X_1, X_2, X_3, \ldots, X_n$ and one criterion variable Y, for a group of individuals or occurrences, the multiple regression problem then is to try and construct a linear function of the form:

$$Y^1 = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

such that the sum of squared errors (deviations),

$$e^2 = \sum(Y-Y^1)^2 = \sum(Y-a-b_1x_1 - b_2x_2 - \ldots - b_nX_n)^2$$

is minimized for the specific data. In effect the problem reduces to that of determining the values $a, b_1, b_2, \ldots b_n$ that will minimize the magnitude of $e^2$. In order to minimize $e^2$, we get the partial coefficients ($b_i$) by using the method of least squares. For one predictor variable, the equation is of the form:

$$Y^1 = a + bx.$$

Therefore minimizing $e^2 = (Y_i - Y_2^1)^2$ is the same as minimizing:

$$\sum[Y - (a + bx)]^2$$

In this function, there are two unknowns, a and b. To solve for them, we make use of the first order conditions.

Let the function to be minimized be F. Then:

$$\frac{\partial F}{\partial a} = \sum_{i=1}^{n} 2(Y-a-bx)(-1)$$

$$\frac{\partial F}{\partial b} = \sum_{i=1}^{n} 2(Y-a-bx)(-x)$$

From the above partial derivatives, the following two normal equations can be obtained, which can then be solved simultaneously to obtain the two unknowns:

$$\sum Y = na + b \sum x \quad \ldots\ldots \text{(i)}$$

$$\sum XY = a \sum x + b \sum x^2 \quad \ldots\ldots \text{(ii)}$$

Note: In the case of two predictor variables, there will be three unknowns and therefore three normal equations. In general, if there are n predictor variables, there will be n+1 unknowns and the same number of normal equations resulting from the partial derivatives.

## 3.2 Assumptions of the model

For the results from the regression model to be valid, a number of assumptions are usually made. These assumptions are necessary mainly for purposes of interpreting the various measures of the goodness of fit of the resulting predictive equation. For most of these assumptions, there are tests that can be performed to find out if the assumptions have been violated or not. These assumptions are that:

1. the random variable e (error term) is assumed to be statistically independent of each of the predictor variables. This implies that at each observation i, the covariance between a predictor variable and the corresponding error term is zero;

2. the error term $e_i$ for all possible sets of given values $x_1$, $x_2$....., $x_n$ are assumed to be nornally distributed;

3. the expected value (the mean) of the errors is zero for all possible sets of given values $x_1$, $x_2$, $x_3$ ...$x_n$, that is $E(e_i) = 0$. This implies that, for a given $x_i$, the differences between $Y_i$ and $Y_1^1$ are on the average zero, though some of the differences may be positive, others negative;

4. the variance of the errors is finite and is constant for all possible sets of given values $x_1$, $x_2$, $x_3$,.....,$x_n$ (homoscedasticity). This means that the dispersion or variability of points in the population from the regression line must be constant;

5. any two error terms are statistically independent of any other, that is, their covariance is zero. What the assumption implies is that the error term at one point in the population cannot be systematically related to the error term of any other point in the population. In other words, knowledge about the size or sign of one or more

errors does not help in predicting the size or sign of any other error;

6. none of the predictor variables is an exact linear combination of another predictor variable(s). This assumption requires that no perfectly linear relationship exists (multicollinearity);

7. the number of observations (m) must exceed the number of coefficients (n+1) being estimated, where n is the number of predictor variables. With (n+1) coefficients, the number of degrees of freedom is m - (n + 1). Therefore the condition specifies that there must be at least one degree of freedom.

The assumptions allow us to know the characteristics (properties) of the estimators obtained by the least squares rules. Using the assumptions, it can be determined that each estimator is a random variable with a normal probability distribution. So the assumptions give us three important properties of a, $b_i$ and $Y^1$[20]:

- they are correct on the average (unbiased)
- relatively reliable on a single trial (efficiency)
- they are more and more accurate on the average as the sample size increases (consistency).

---

20. Harnett, D.L. and Murphy, J.L.: <u>Statistical Analysis for Business and Economics</u> (3rd Ed. Addison Wesley Publishing Co., 1985) pg. 574.

The robustness of the regression model lies in the fact that various tests are available to test both the significance of the estimators and violations of some of the assumptions.

3.3 Predicting using the Regression model

If the resulting equation has a significant fit and the significant predictor variables have also been identified, then it can be used in determining point forecasts. But while performing the estimations, it must be borne in mind that the model is derived from a sample, hence gives only results from one set of data. So no matter how significant the fit is, it is not true that the predictions will be accurate for other sets of data. However, to overcome this problem of stationality, the model can be improved by using a series of samples in order to find out which variables appear consistently significant.

Another solution out of the problem is to develop a model, then test its appropriateness by using a different sample (cross-validation). To perform the test, all that is needed is to calculate a correlation coefficient between the actual and predicted values for the validation group. Gatumu (1976) and Ingram and Petersen (1987: pg. 220) have adopted this method. If the ordinary product moment correlation coefficient is reasonably close to that of the original sample, the researcher could be confident that in subsequent samples too, the predictive efficiency will be more or less of this degree (Gatumu, 1976: pg. 36).

Once the model has been validated it can then be used for prediction but still with a caveat. This is because as Neter and Wasserman (1985 put it:

> "...it is important to remember that the validity of the regression application depends upon whether basic conditions in the period ahead will be similar to those in existence during the period upon which the regression is based"[21]

But generally, the equation will be quite accurate near the point of means or averages of the variables used. Given that no substantial changes occur in the conditions surrounding the time, place and observation procedures under which the data was collected, the predictive equation will usually be valid.

In using the model for prediction, regression equations are mostly used to get a point estimate, that is, "the best guess numerical score" on the criterion variable for an individual with a given combination of predictor scores. However, it is possible in certain cases to utilize regression

---

21. Neter, J. Wasserman, W. and Kutner, M.H.: Applied Linear Statistical methods: Regression, Analysis of Variance and Experimental Designs (2nd Ed., Richard D. Irwin Inc., 1985) pg. 83.

models in a more sophisticated manner. One of these, suggested by Gatumu (1976) is for example, to get answers to questions such as what is the approximate probability that an individual with a particular combination of predictor scores will get a criterion score above a specified value? (pg.39). So it can be used to answer the question, what are the chances that a particular individual will succeed or fail? In the present study, we shall look into the question: what are the chances that a student who obtains certain results at O-level will succeed in Commerce?

# CHAPTER 4

# DATA ANALYSIS AND FINDINGS

## 4.1 Introduction

For the model development, the score attained in the final year of study at the University was used as the criterion variable.

The predictor variables are the subjects that will be offered in the four years of secondary level education under the 8-4-4 system. But some of the subjects are not currently being offered. So, where possible, a surrogate was used. Also because the number of subjects are so many, it was found necessary to group the predictor variables into categories and develop an index of measurement. Even if all the predictor variables could be analysed separately without grouping them, the approach of categorizing them is more appropriate because the prime objective of the study is to come up with a group of subjects that could possibly be used in formulating a selection criteria. After the identification of this group, the individual subjects in the group can then be tested as separate predictor variables.

Since there is no a priori basis of grouping the subjects, a useful beginning point is the categorization adopted in the 8-4-4 syllabus for the subjects to be taught in secondary school level. Presented below are the categories as per the syllabus

8-4-4 SYLLABUS

Group 1 (Compulsory core subjects)

English

Kiswahili

History/Government

Geography

Mathematics

Biology or Biological Sciences

Physical Science or Physics and Chemistry

Group 2

Christian Religious Education

Islamic Religious Education

Social Education and Ethics

Group 3

Home Science

Agriculture

Woodwork and Metal work

Building Construction

Power Mechanics

Electricity

Art and Design

Drawing and Design

<u>Group 4</u>

French

German

Music

Accounting

Commerce

Economics

Typing and Office Practice

<u>CATEGORIZATION USED IN THE STUDY</u>

<u>Group 1</u> (Main language)

English

<u>Group 2</u> (Quantitative)

Mathematics

<u>Group 3</u> (Arts)

History

Geography

Christian Religious Education

Islamic Religious Education

<u>Group 4</u> (Sciences)

Biology/Biological Sciences

Physical Sciences/Physics and Chemistry

<u>Group 5</u> (Other Sciences)

Agriculture

Home Science/Domestic Science

Health Science

General Science

<u>Group 6</u> (Other languages)

French

German

Music

<u>Group 7</u> (Commerce)

Principles of Accounts

Commerce

Economics

All the data used was from the students' 0-level results. In O-level, the results, per subject are awarded in terms of numerics and designated as below:

1-2 Distinction

3-6 Credit

7-8 Pass

9 Fail

Therefore the lower the score, the better the performance. In situations where A-level results were used as a surrogate (only for Economics) it was necessary to transform the A-level results to O-level 'equivalents' so as to get a meaningful score in the category in which that particular subject falls. For example, A-level Economics has been grouped with O-level Commerce subjects. To get a score for the whole category, all the subjects need to be on the same point score scale. In A-level, the results are given in the form of grades A to E with the following points:

| A | B | C | D | E | Subsidiary (0) | F |
|---|---|---|---|---|---|---|
| 6 | 5 | 4 | 3 | 2 | 1 | 0 |

Therefore, to convert the A-level scores into O-level score equivalents, the following scheme was used:

| A-Level | O-Level equivalent |
|---|---|
| A | 1-2 (1.5) |
| B | 3 |
| C | 4 |
| D | 5 |
| E | 6 |
| O | 7-8 (7.5) |
| F | 9 |

In getting the index for each category of subjects, the total score was calculated for the subjects taken and then an average was calculated by dividing the total score by the number of subjects taken in that category, thus giving a representative index for the category. If a student does not take any subject in a particular category, a score of 9 was awarded for that category.

The following variable names were used to represent each of the categories used:

| variable name | category |
|---|---|
| sex | sex |
| Age | Age |
| Eng | English |
| Art | Arts subjects |
| Mat | Mathematics |
| Sci | Sciences |
| Gen | Other Sciences |
| Lan | Other languages |
| Com | Commerce |
| Gce | O-level aggregate score |
| GPA | Grade Point Average in the final year at University |

Since there were two different samples used, the results of each were analysed separately.

## 4.2 1974/75-1976/77 results

### 4.2.1 Correlation analysis

Before regression analysis could be done a correlation analysis was done to check for any multicollinearity among the predictor variables. The results are given in the correlation matrix as shown in table 4.1 below.

## TABLE 4.1 : CORRELATION MATRIX 1974/77

| | GPA7477 | sex 7477 | Age7477 | Eng7477 | Mat7477 | Art7477 | Sci7477 | Gen7477 | Lan7477 | Com7477 | Gce7477 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GPA7477 | 1.00000 | | | | | | | | | | |
| Sex7477 | .10562 | 1.00000 | | | | | | | | | |
| Age7477 | -.10086 | .35935* | 1.00000 | | | | | | | | |
| Eng7477 | -.07400 | .01261 | .06271 | 1.00000 | | | | | | | |
| Mat7477 | -.54576* | .54576* | -.10100 | .08223 | 1.00000 | | | | | | |
| Art7477 | -.13146 | .05652 | -.14434 | .00218 | .14658 | 1.00000 | | | | | |
| Sci7477 | -.17743 | .03944 | -.11057 | .35109* | .14714 | .2116* | 1.00000 | | | | |
| Gen7477 | -.05757 | .10076 | .13281 | -.14747 | .13661 | -.05383 | -.17276 | 1.00000 | | | |
| Lan7477 | .07837 | -.11577 | -.01141 | -.02965 | .13644 | -.04949 | -.09435 | .00490 | 1.00000 | | |
| Com7477 | -.20400* | -.63489* | -.23037* | .05632 | .22445* | .27236* | .04032 | -.10874 | -.04282 | 1.00000 | |
| Gce7477 | -.32339* | .01253 | -.00380 | .39968* | .47850* | .51498* | .45425* | .18029* | .21378* | .10564 | 1.00000 |

*Significant at $\alpha = 0.05$

From the matrix above, it is evident that many of the variables are strongly correlated. At $\alpha = 0.05$, a correlation of 0.19 for the 100 observations is significant. Accordingly, the variable representing O-level aggregate score (GCE) is found to be highly correlated with virtually all the other predictor variables particularly, English, Mathematics, Arts subjects, Sciences, Other languages and General Sciences. But it should also be noted that O-level aggregate is significantly correlated with GPA. However, this strong correlation with the other predictor variables is to be expected because afterall the variable is an aggregate of the best six scores in O-level. Because this multicollinearity effect makes it difficult to interpret the regression coefficients it is necessary to remove O-level aggregate from the model.

In evaluating the correlation between the predictor variables and the criterion variable, only four have a significant correlation with GPA. These are, Mathematics (-0.55), GCE (-0.32), Commerce (-0.20) and Science(-0.18).

Age and sex are highly correlated at 0.40 though this has no meaningful interpretation except maybe that the majority of the students who are admitted at more advanced ages are likely to be males. Sex is further highly correlated with Commerce (-0.63).

These two variables affecting sex tend to justify the removal of sex from the prediction model. Age is also significantly correlated with Commerce (0.23) though this is not too serious. Some of the correlations, though significant are meaningless, for example English and Science. However, the correlation between Mathematics and Commerce (0.22) can be explained in that both variables are quantitative.

### 4.2.2 Regression Analysis

Because of the strong correlation between O-level aggregate and virtually all the other variables, an ordinary least squares regression was run using the GPA as the criterion variable and the other nine (excluding O-level aggregate) as the predictor variables. Table 4.2 next page shows the results.

The results show a low coefficient of determination which is 0.35 (adjusted = 0.29). Further, the effects of multicollinearity are quite evident. In the correlation matrix (see table 4.1) English is negatively correlated with GPA whereas the regression coefficient for this variable implies a positive relationship, thus giving confusing signals making it difficult to interpret the meaning of this coefficient. This is the same case for the General Sciences. Otherwise, all the other variables have the same relationship as that ble 4.2: MULTIPLE REGRESSION RESULTS

MODEL FITTING RESULTS

| ARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>~T~ |
|---|---|---|---|---|
| CONSTANT | 73.459128 | 7.197612 | 10.2060 | .000 |
| sex7477 | 1.558919 | 1.680064 | .9279 | .355 |
| age7477 | -0.541991 | 0.327295 | -1.6560 | .100 |
| eng7477 | 0.065609 | 0.29949 | .2191 | .827 |
| mat7477 | -1.83597 | 0.319867 | -5.7398 | .000 |
| art7477 | -0.216581 | 0.44476 | -.4870 | .627 |
| sci7477 | -0.239373 | 0.22144 | -1.0810 | .282 |
| gen7477 | 0.003733 | 0.150428 | .0248 | .980 |
| lan7477 | 0.262256 | 0.154963 | 1.6924 | .093 |
| com7477 | -0.058189 | 0.274245 | -.2122 | .832 |

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(> |
|---|---|---|---|---|---|
| MODEL | 913.68957 | 9 | 101.52106 | 5.49195 | .000 |
| ERROR | 1663.6875 | 90 | 18.4854 | | |
| TOTAL (CORR.) | 2577.3771 | 99 | | | |

R-SQUARED = 0.354504

R-SQUARED (ADJ. FOR D.F.) = 0.289954

STND. ERROR OF EST. = 4.29947

implied from the correlation matrix.

Table 4.2 shows that at 95% confidence level only one predictor variable, Mathematics, and the constant term are significant, using the t-statistic values. However, this does not imply that all the other variables are insignificant. Due to the problem of multicollinearity, a further regression analysis was done after removing the following variables:

| variable | Reason for removal |
| --- | --- |
| sex | multicollinearity with age and Commerce and relatively large standard error |
| English | insignificant correlation with GPA |
| Other languages | insignificant correlation with GPA |
| General sciences | insignificant correlation with GPA |
| O-level aggregate | multicollinearity |
| Arts | Multicollinearity with Commerce and Science. |

After the removal of the above variables the results of the regression are shown in table 4.3.

From table 4.3 the coefficient of determination

TABLE 4.3: MULTIPLE REGRESSION RESULTS, 1974/77

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 74.978999 | 6.85765 | 10.9336 | .0000 |
| age7477 | -0.453114 | 0.303107 | -1.4949 | .1381 |
| mat7477 | -1.748262 | 0.305607 | -5.7206 | .0000 |
| sci7477 | -0.262336 | 0.197537 | -1.3280 | .1872 |
| com7477 | -0.258913 | 0.196233 | -1.3194 | .1901 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

RESIDUALS PLACED IN VARIABLE: RESIDUALS

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 850.85180 | 4 | 212.71295 | 11.70428 | .00000 |
| ERROR | 1726.5253 | 95 | 18.1740 | | |
| TOTAL (CORR.) | 2577.3771 | 99 | | | |

R-SQUARED = 0.330123
R-SQUARED (ADJ. FOR D.F.) = 0.301918
STND. ERROR OF EST. = 4.26309

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = -6.39488E-16
SAMPLE VARIANCE = 17.4396
SAMPLE STANDARD DEVIATION = 4.17608
COEFF. OF SKEWNESS = 0.428472 STANDARDIZED VALUE = 1.74923
COEFF. OF KURTOSIS = 3.72319 STANDARDIZED VALUE = 1.47621
DURBIN-WATSON STATISTIC = 1.58185

has only dropped from 0.35 (0.29 adjusted) to 0.33 (0.30 adjusted) which is an insignificant (negligible) decline. Moreover, of those variables in the model all of them are significant at $\alpha = 0.20$ level

So on the basis of the above analysis for this sample, the most significant predictor variables are, Mathematics, age, Science and Commerce subjects in that order.

However, the above process of removing certain variables and retaining others on the basis or correlation has some weakness. Therefore stepwise regression analysis was performed using the forward inclusion method. (See Appendix C for details of the stepwise regression procedure).

Using stepwise regression analysis the critical F-value at $\alpha = 0.05$ level, is 3.94 (1 and 98 df). After specifying this value, the results are as shown in table 4.4(A) next page.

As from the earlier analyses, it is also clear that only Mathematics has an F-value higher than the specified critical value. In fact it is significant at $\alpha = 0.01$. Thus at $\alpha = 0.05$ level, it is the only significant predictor and it explains 0.30 (0.29 adjusted) of the variables.

TABLE4.4(A) - STEPWISE REGRESSION, 1974/77

STEPWISE REGRESSION

SELECTION: FORWARD CONTROL: AUTOMATIC
F-TO-ENTER = 3.94 MAX STEPS = 50 F-TO-REMOVE = 3.94

STEP 1

R-SQUARED = 0.297852
R-SQUARED (ADJ.) = 0.290687 MSE = 18.4663 WITH 98 D.F.

| VARIABLES CURRENTLY IN MODEL | | | VARIABLES CURRENTLY NOT IN MODEL | | |
|---|---|---|---|---|---|
| VARIABLE | COEFF. | F-REMOVE | VARIABLE | PARTIAL CORR. | F-ENTER |
| 7. mat7477 | -1.90783 | 41.5716 | 1. sex7477 | .0606 | .3571 |
| | | | 2. com7477 | -.0998 | .9760 |
| | | | 3. lan7477 | .1841 | 3.4033 |
| | | | 4. gen7477 | .0205 | .0406 |
| | | | 5. sci7477 | -.1172 | 1.3506 |
| | | | 6. art7477 | -.0621 | .3754 |
| | | | 8. eng7477 | -.0349 | .1181 |
| | | | 9. age7477 | -.1072 | 1.1271 |

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 63.347096 | 1.250847 | 50.6434 | .0000 |
| mat7477 | -1.907833 | 0.295898 | -6.4476 | .0000 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 767.67564 | 1 | 767.67564 | 41.57162 | .00000 |
| ERROR | 1809.7015 | 98 | 18.4663 | | |
| TOTAL (CORR.) | 2577.3771 | 99 | | | |

R-SQUARED = 0.297852
R-SQUARED (ADJ. FOR D.F.) = 0.290687
STND. ERROR OF EST. = 4.29725

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = -2.34479E-15
SAMPLE VARIANCE = 18.2798
SAMPLE STANDARD DEVIATION = 4.27549
COEFF. OF SKEWNESS = 0.674506 STANDARDIZED VALUE = 2.75366
COEFF. OF KURTOSIS = 4.58963 STANDARDIZED VALUE = 3.24481
DURBIN-WATSON STATISTIC = 1.47133

However, to find out which of the other variables could be important predictors, the critical F-value was reduced to 2.76 ($\alpha = 0.10$ and one more variable was included in the model; Other languages and the $r^2$ increased to 0.32 (0.31 adj.). However this additional variable tends to have a "brushing effect" because the number of students who took this category subjects was very small compared to the total sample. To remove this illusionary effect, a stepwise regression was performed at F = 2.76 without this variable and only Mathematics was included in the model. (See Table 4.4(B)).

A further stepwise regression analysis was performed at an F-value of 0.459 (significant at $\alpha$= 0.50), to find out which other variables could be important though not very significant. Table 4.4(C) shows the results of the regression and it can be seen that omission of Other languages, in addition to Mathematics three other variables are included in the model that is

TABLE 4.4(B) STEPWISE REGRESSION, 1974/77

STEPWISE REGRESSION

SELECTION: FORWARD  CONTROL: AUTOMATIC
F-TO-ENTER = 2.76  MAX STEPS = 50  F-TO-REMOVE = 2.76
STEP 2

R-SQUARED = 0.321651
R-SQUARED (ADJ.) = 0.307665  MSE = 18.0243 WITH 97 D.F.

| VARIABLES CURRENTLY IN MODEL | | | VARIABLES CURRENTLY NOT IN MODEL | | |
|---|---|---|---|---|---|
| VARIABLE | COEFF. | F-REMOVE | VARIABLE | PARTIAL CORR. | F-ENTER |
| 4. mat7477 | -1.98211 | 45.1162 | 1. sex7477 | .0814 | .6410 |
| 8. lan7477 | .27416 | 3.4033 | 2. age7477 | -.1064 | 1.0986 |
| | | | 3. eng7477 | -.0278 | .0740 |
| | | | 5. art7477 | -.0500 | .2407 |
| | | | 6. sci7477 | -.0980 | .9313 |
| | | | 7. gen7477 | .0234 | .0528 |
| | | | 9. com7477 | -.0875 | .7415 |

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 61.892809 | 1.465816 | 42.2241 | .0000 |
| mat7477 | -1.98211 | 0.295095 | -6.7169 | .0000 |
| lan7477 | 0.274164 | 0.148615 | 1.8448 | .0681 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

RESIDUALS PLACED IN VARIABLE: RESIDUALS

STEPWISE REGRESSION

SELECTION: FORWARD  CONTROL: AUTOMATIC
F-TO-ENTER = 2.76  MAX STEPS = 50  F-TO-REMOVE = 2.76
STEP 1

R-SQUARED = 0.297852
R-SQUARED (ADJ.) = 0.290687  MSE = 18.4663 WITH 98 D.F.

| VARIABLES CURRENTLY IN MODEL | | | VARIABLES CURRENTLY NOT IN MODEL | | |
|---|---|---|---|---|---|
| VARIABLE | COEFF. | F-REMOVE | VARIABLE | PARTIAL CORR. | F-ENTER |
| 4. mat7477 | -1.90783 | 41.5716 | 1. sex7477 | .0606 | .3571 |
| | | | 2. age7477 | -.1072 | 1.1271 |
| | | | 3. eng7477 | -.0349 | .1181 |
| | | | 5. art7477 | -.0621 | .3754 |
| | | | 6. sci7477 | -.1172 | 1.3506 |
| | | | 7. gen7477 | .0205 | .0406 |
| | | | 8. com7477 | -.0998 | .9760 |

TABLE 4.4(C): STEPWISE REGRESSION, 1974/77

STEPWISE REGRESSION

SELECTION: FORWARD                                    CONTROL: AUTOMATIC
F-TO-ENTER = 0.459    MAX STEPS = 50                  F-TO-REMOVE = 0.459

STEP 4

R-SQUARED = 0.330123
R-SQUARED (ADJ.) = 0.301918                           MSE = 18.174 WITH 95 D.F.

| VARIABLES CURRENTLY IN MODEL | | | VARIABLES CURRENTLY NOT IN MODEL | | |
|---|---|---|---|---|---|
| VARIABLE | COEFF. | F-REMOVE | VARIABLE | PARTIAL CORR. | F-ENTER |
| 2. age7477 | -.45311 | 2.2347 | 1. sex7477 | .0531 | .2663 |
| 4. mat7477 | -1.74826 | 32.7254 | 3. eng7477 | .0286 | .0771 |
| 6. sci7477 | -.26234 | 1.7637 | 5. art7477 | -.0274 | .0707 |
| 8. com7477 | -.25891 | 1.7409 | 7. gen7477 | -.0055 | .0028 |

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 74.978999 | 6.85765 | 10.9336 | .0000 |
| age7477 | -0.453114 | 0.303107 | -1.4949 | .1383 |
| mat7477 | -1.748262 | 0.305607 | -5.7206 | .0000 |
| sci7477 | -0.262336 | 0.197537 | -1.3280 | .1873 |
| com7477 | -0.258913 | 0.196233 | -1.3194 | .1902 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 850.85180 | 4 | 212.71295 | 11.70428 | .00000 |
| ERROR | 1726.5253 | 95 | 18.1740 | | |
| TOTAL (CORR.) | 2577.3771 | 99 | | | |

R-SQUARED = 0.330123
R-SQUARED (ADJ. FOR D.F.) = 0.301918
STND. ERROR OF EST. = 4.26309

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = -6.39488E-16
SAMPLE VARIANCE = 17.4396
SAMPLE STANDARD DEVIATION = 4.17608
COEFF. OF SKEWNESS = 0.428472   STANDARDIZED VALUE = 1.74923
COEFF. OF KURTOSIS = 3.72319   STANDARDIZED VALUE = 1.47621
DURBIN-WATSON STATISTIC = 1.58185

Age, Sciences, and Commerce. It should be noted that these are the same variables identified using correlation analysis.

### 4.2.3 Conclusion

Based on the results above;

there is evidence of strong correlations between the various subjects taken at O-level;

demographic variable age, appears to be strongly related to performance

Mathematics is the best single predictor of performance.

## 4.3 1977/78-1979/80 Results

### 4.3.1 Correlation analysis

Using the same variables as previously defined, correlation analysis was done for the 1977/80 sample. The correlation matrix shown in table 4.5 next page presents the results.

As in the previous sample, a number of predictor variables are highly correlated with student performance. These are: Mathematics (-0.54), Commerse (-0.27), O-level aggregate (-0.33), age (-0.26) and Science (-0.215).

It should be noted that the correlation between

TABLE 4.5 CORRELATION MATRIX 1977/80

| | GPA7780 | Sex7780 | Age7780 | Eng7780 | Mat7780 | Art7780 | Sci7780 | Gen7780 | Lan7780 | Com7780 | Gce7780 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GPA7780 | 1.00000 | | | | | | | | | | |
| Sex7780 | .13026 | 1.00000 | | | | | | | | | |
| Age7780 | -.25643* | .26214* | 1.00000 | | | | | | | | |
| Eng7780 | -.06792 | .16350 | -.13678 | 1.00000 | | | | | | | |
| Mat7780 | -.54360* | .00175 | .18013 | -.01609 | 1.00000 | | | | | | |
| Art7780 | -.14332 | -.03652 | -.07823 | -.26531* | .14703 | 1.00000 | | | | | |
| Gen7780 | -.04500 | -.07059 | .08411 | .08411 | .10249 | .04695 | -.12061 | 1.00000 | | | |
| Lan7780 | -.10921 | .04662 | .09062 | .02670 | -.09814 | .03568 | .03216 | -.17799 | 1.00000 | | |
| Com7780 | -.26632* | -.33395* | -.01061 | -.04914 | .26500* | -.07334 | -.02259 | .00490 | -.02557 | 1.00000 | |
| Gce7780 | -.33445* | .08468 | .01102 | .53940* | .41147* | .64226* | .42557* | .19465* | .08090 | -.01174 | 1.00000 |

*significant at 0.05

the subjects at O-level and GPA are negative. This indicates that the lower (the better) the score at O-level, the higher the performance at university level.

From the correlation matrix, there is evidence of multicollinearity. Sex is significantly correlated with age. Sex is also significantly correlated with O-level Commerce (-0.33) which indicates that males tend to perform better in the subject. Arts and English are highly positively correlated. Science and Mathematics are also significantly correlated (0.26) and this can be expected since the Sciences are related to Mathematics in nature. Mathematics is highly correlated with Commerce (0.26) because the two are generally related to some extent. Science and Arts are siginificantly correlated (0.32). O-level aggregate score is significantly correlated with quite a number of other variables; namely English (0.54), Mathematics (0.41), Arts (0.64), Sciences (0.426) and General Sciences (0.19). However the O-level aggregate score has a very high correlation with GPA (-0.33) which therefore implies that despite the collinearity with other predictor variables, it is an important predictor variable. Infact after Mathematics, it is the next most significant predictor variable. So whereas it might have to be removed from the model for the sake of removing the multicollinearity effect, it should be considered as an important variable for gauging the general knowledge of a student.

From the above, it is clear that the correlations between the variables are almost a replica of the 1974/77 correlations. This reflects on the uniformity of samples though taken from different years.

### 4.3.2 Regression Analysis

All the predictor variables with the exception of O-level aggregate were used to run a regression analysis. The results of the regression analysis are shown in table 4.6(A)(p63). Mathematics, Sex and Age are significant. The coefficient of determination ($r^2$) is low at 0.41 (0.35 - adjusted).

As earlier pointed out, there is evidence of multicollinearity and therefore it was necessary to remove certain variables so as to get a meaningful interpretation of the beta coefficients. The following variables were removed:

| variable name | reasons |
|---|---|
| O-level aggregate | multicollinearity with six other predictor variables. |
| Arts | multicollinearity with English and Science |
| General Sciences | insignificant correlation with the criterion variable |
| Other languages | insignificant correlation with the criterion variable. |

TABLE 4.6(A) MULTIPLE REGRESSION RESULTS, 1977/80

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>¯T¯) |
|---|---|---|---|---|
| CONSTANT | 86.16752 | 7.299498 | 11.8046 | .0000 |
| sex7780 | 2.009411 | 1.00808 | 1.9933 | .0490 |
| age7780 | -0.91411 | 0.351517 | -2.6005 | .0107 |
| eng7780 | -0.371484 | 0.268161 | -1.3853 | .1691 |
| mat7780 | -1.512453 | 0.290623 | -5.2042 | .0000 |
| art7780 | -0.138975 | 0.305075 | -.4555 | .6497 |
| sci7780 | -0.105577 | 0.20769 | -.5083 | .6123 |
| gen7780 | 0.062475 | 0.159068 | .3928 | .6953 |
| lan7780 | -0.237958 | 0.149357 | -1.5932 | .1143 |
| com7780 | -0.16857 | 0.16503 | -1.0215 | .3095 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

RESIDUALS PLACED IN VARIABLE: RESIDUALS

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 946.59061 | 9 | 105.17673 | 6.83549 | .00000 |
| ERROR | 1384.8169 | 90 | 15.3869 | | |
| TOTAL (CORR.) | 2331.4075 | 99 | | | |

R-SQUARED = 0.406017
R-SQUARED (ADJ. FOR D.F.) = 0.346618
STND. ERROR OF EST. = 3.92261

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = 1.87583E-14
SAMPLE VARIANCE = 13.988
SAMPLE STANDARD DEVIATION = 3.74006
COEFF. OF SKEWNESS = 0.231781 STANDARDIZED VALUE = 0.946243
COEFF. OF KURTOSIS = 2.93671 STANDARDIZED VALUE = -0.129189
DURBIN-WATSON STATISTIC = 2.50156

Therefore a regression analysis was done using the six remaining predictor variables. See table 4.6(B) on the next page.

From the results, the $r^2$ has been reduced from 0.41 to 0.38 (0.34, adj.), which is not a significant drop. However, the standard errors for each variable vis-a-vis the Beta coefficients reveals that science has a standard error that is larger than even the coefficient itself thus resulting in a very low t-value which makes the variable insignificant.

Therefore to further improve the model, Science was omitted from the predictor variables and a second regression equation was run. The results of this regression are shown in table 4. 6(C). The $r^2$ remains at 0.38 (0.35 adj.) and so nothing significant was lost by dropping Science.

Also, after removing Science, the coefficient and standard error of Commerce have changed so much as to make it insignificant. It was therefore removed from the regression model. The results after removing this variable are presented in table 4.6(D) . The results show that the coefficient of multiple determination dropped slightly 0.37 (0.35 adj.) from 0.38 (0.35 adj.). Further, the standard error of the estimate changed by a very small margin.

Stepwise regression analysis was also used to identify the variables that are statistically significant

TABLE 4.6(B) multiple regression results, 1977/80

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 85.446181 | 7.249146 | 11.7871 | .0000 |
| sex7780 | 1.951175 | 0.991826 | 1.9673 | .0520 |
| age7780 | -0.942025 | 0.342283 | -2.7522 | .0070 |
| eng7780 | -0.392629 | 0.257842 | -1.5227 | .1310 |
| mat7780 | -1.453266 | 0.284458 | -5.1089 | .0000 |
| sci7780 | -0.163569 | 0.198015 | -.8260 | .4108 |
| com7780 | -0.170692 | 0.163573 | -1.0435 | .2992 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

RESIDUALS PLACED IN VARIABLE: RESIDUALS

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 895.96711 | 6 | 149.32785 | 9.67472 | .00000 |
| ERROR | 1435.4404 | 93 | 15.4348 | | |
| TOTAL (CORR.) | 2331.4075 | 99 | | | |

R-SQUARED = 0.384303
R-SQUARED (ADJ. FOR D.F.) = 0.344581
STND. ERROR OF EST. = 3.92872

## TABLE 4.6(C) MULTIPLE REGRESSION RESULTS 1977/80

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 84.849079 | 7.200819 | 11.7833 | .0000 |
| sex7780 | 2.031542 | 0.985374 | 2.0617 | .0419 |
| age7780 | -0.93923 | 0.341687 | -2.7488 | .0071 |
| eng7780 | -0.432724 | 0.252804 | -1.7117 | .0901 |
| mat7780 | -1.520772 | 0.272006 | -5.5910 | .0000 |
| com7780 | -0.154256 | 0.162084 | -.9517 | .3436 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

RESIDUALS PLACED IN VARIABLE: RESIDUALS

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 885.43518 | 5 | 177.08704 | 11.51210 | .00000 |
| ERROR | 1445.9723 | 94 | 15.3827 | | |
| TOTAL (CORR.) | 2331.4075 | 99 | | | |

R-SQUARED = 0.379786
R-SQUARED (ADJ. FOR D.F.) = 0.346796
STND. ERROR OF EST. = 3.92208

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = 1.82609E-14
SAMPLE VARIANCE = 14.6058
SAMPLE STANDARD DEVIATION = 3.82175
COEFF. OF SKEWNESS = 0.0728705 STANDARDIZED VALUE = 0.297492
COEFF. OF KURTOSIS = 2.8623 STANDARDIZED VALUE = -0.281069
DURBIN-WATSON STATISTIC = 2.45632

## TABLE 4.6(D): MULTIPLE REGRESSION RESULTS, 1977/80

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 84.425864 | 7.183509 | 11.7527 | .0000 |
| sex7780 | 2.350268 | 0.926257 | 2.5374 | .0127 |
| age7780 | -0.951198 | 0.341287 | -2.7871 | .0064 |
| eng7780 | -0.436997 | 0.252639 | -1.7297 | .0868 |
| mat7780 | -1.591033 | 0.261665 | -6.0804 | .0000 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

RESIDUALS PLACED IN VARIABLE: RESIDUALS

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 871.50243 | 4 | 217.87561 | 14.17776 | .00000 |
| ERROR | 1459.9051 | 95 | 15.3674 | | |
| TOTAL (CORR.) | 2331.4075 | 99 | | | |

R-SQUARED = 0.37381
R-SQUARED (ADJ. FOR D.F.) = 0.347444
STND. ERROR OF EST. = 3.92013

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = 1.64846E-14
SAMPLE VARIANCE = 14.7465
SAMPLE STANDARD DEVIATION = 3.84012
COEFF. OF SKEWNESS = 0.0647599 STANDARDIZED VALUE = 0.264381
COEFF. OF KURTOSIS = 2.59834 STANDARDIZED VALUE = -0.819888
DURBIN-WATSON STATISTIC = 2.51621

at the critical value of F = 3.94 ($\alpha = 0.05$). From the initial list of nine predictor variables, only Mathematics was significant (at $\alpha = 0.05$) level)- see table 4. (A). This variable alone has an $r^2$ 0.30 (0.29 adj.) which is 72.5% of the $r^2$ for all the variables.

Two further stepwise regressions analysis were performed at F-levels 2.76 ($\alpha = 0.10$) and 0.459 ($\alpha = 0.50$) respectively. At F = 2.76, three more predictor variables were included in the model. These are Sex, Age and English (see table 4.7(B)). The multiple coefficient of determination is 0.37 (0.35 adjusted).

At F = 0.459 Science and Commerce were included in the model thus making a total of six predictor variable (see table 4.7(C)). Both of these variables only increased the $r^2$ to 0.380 (0.340 adjusted) from 0.37 (0.35 adjusted) and did not change the standard error of the estimate much. So these two variables can be safely discarded.

### 4.3.3 Conclusion

From the foregoing correlation and regression analysis, the following conclusions can be made:

inter-relationships exist between many of the predictor variables;

Mathematics is the best predictor of performance at university level;

other important predictors are age, sex, English, Commerce and Sciences.

TABLE 4.7(A): STEPWISE REGRESSION, 1977/80

STEPWISE REGRESSION

SELECTION: FORWARD CONTROL: AUTOMATIC
F-TO-ENTER = 3.94 MAX STEPS = 50 F-TO-REMOVE = 3.94

STEP 1

R-SQUARED = 0.295496
R-SQUARED (ADJ.) = 0.288307 MSE = 16.7601 WITH 98 D.F.

| VARIABLES CURRENTLY IN MODEL | | | VARIABLES CURRENTLY NOT IN MODEL | | |
|---|---|---|---|---|---|
| VARIABLE | COEFF. | F-REMOVE | VARIABLE | PARTIAL CORR. | F-ENTER |
| 4. mat7780 | -1.72103 | 41.1049 | 1. sex7780 | .1563 | 2.4299 |
| | | | 2. age7780 | -.1920 | 3.7123 |
| | | | 3. eng7780 | -.0914 | .8163 |
| | | | 5. art7780 | -.0764 | .5689 |
| | | | 6. sci7780 | -.0913 | .8154 |
| | | | 7. gen7780 | .0128 | .0160 |
| | | | 8. com7780 | -.1511 | 2.2653 |
| | | | 9. lan7780 | -.1946 | 3.8186 |

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 64.727714 | 1.096656 | 59.0228 | .0000 |
| mat7780 | -1.721033 | 0.268437 | -6.4113 | .0000 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 688.92079 | 1 | 688.92079 | 41.10489 | .00000 |
| ERROR | 1642.4867 | 98 | 16.7601 | | |
| TOTAL (CORR.) | 2331.4075 | 99 | | | |

R-SQUARED = 0.295496
R-SQUARED (ADJ. FOR D.F.) = 0.288307
STND. ERROR OF EST. = 4.09391

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = 1.42109E-14
SAMPLE VARIANCE = 16.5908
SAMPLE STANDARD DEVIATION = 4.07318
COEFF. OF SKEWNESS = 0.102179 STANDARDIZED VALUE = 0.417144
COEFF. OF KURTOSIS = 2.89338 STANDARDIZED VALUE = -0.217627
DURBIN-WATSON STATISTIC = 2.49604

TABLE 4.7(B): STEPWISE REGRESSION, 1977/80

STEPWISE REGRESSION

SELECTION: FORWARD — CONTROL: AUTOMATIC

F-TO-ENTER = 2.76 — MAX STEPS = 50 — F-TO-REMOVE = 2.76

STEP 4

R-SQUARED = 0.37381

R-SQUARED (ADJ.) = 0.347444 — MSE = 15.3674 WITH 95 D.F.

| VARIABLES CURRENTLY IN MODEL | | | VARIABLES CURRENTLY NOT IN MODEL | | |
|---|---|---|---|---|---|
| VARIABLE | COEFF. | F-REMOVE | VARIABLE | PARTIAL CORR. | F-ENTER |
| 1. sex7780 | 2.35027 | 6.4383 | 5. art7780 | -.0560 | .2959 |
| 2. age7780 | -.95120 | 7.7679 | 6. sci7780 | -.0724 | .4956 |
| 3. eng7780 | -.43700 | 2.9920 | 7. gen7780 | .0898 | .7649 |
| 4. mat7780 | -1.59103 | 36.9714 | 8. com7780 | -.0977 | .9057 |

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 84.425864 | 7.183509 | 11.7527 | .0000 |
| sex7780 | 2.350268 | 0.926257 | 2.5374 | .0128 |
| age7780 | -0.951198 | 0.341287 | -2.7871 | .0064 |
| eng7780 | -0.436997 | 0.252639 | -1.7297 | .0869 |
| mat7780 | -1.591033 | 0.261665 | -6.0804 | .0000 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 871.50243 | 4 | 217.87561 | 14.17776 | .00000 |
| ERROR | 1459.9051 | 95 | 15.3674 | | |
| TOTAL (CORR.) | 2331.4075 | 99 | | | |

R-SQUARED = 0.37381
R-SQUARED (ADJ. FOR D.F.) = 0.347444
STND. ERROR OF EST. = 3.92013

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = 1.64846E-14
SAMPLE VARIANCE = 14.7465
SAMPLE STANDARD DEVIATION = 3.84012
COEFF. OF SKEWNESS = 0.0647599 STANDARDIZED VALUE = 0.264381
COEFF. OF KURTOSIS = 2.59834 STANDARDIZED VALUE = -0.819888
DURBIN-WATSON STATISTIC = 2.51621

## TABLE 4.7(C): STEPWISE RGRESSION1977/80

STEPWISE REGRESSION

```
SELECTION:    FORWARD                                          CONTROL:       AUTOMATIC
F-TO-ENTER =  0.459          MAX STEPS =  50                   F-TO-REMOVE = 0.459
                                   STEP 6
R-SQUARED = 0.384303
R-SQUARED (ADJ.) = 0.344581                    MSE = 15.4348 WITH 93 D.F.
 VARIABLES CURRENTLY IN MODEL               VARIABLES CURRENTLY NOT IN MODEL
VARIABLE               COEFF.  F-REMOVE VARIABLE            PARTIAL CORR.   F-ENTER
  1. sex7780          1.95118    3.8701  5. art7780                 -.0527     .2561
  2. age7780          -.94203    7.5745  7. gen7780                  .0699     .4522
  3. eng7780          -.39263    2.3188
  4. mat7780         -1.45327   26.1008
  6. sci7780          -.16357     .6823
  8. com7780          -.17069    1.0889
```

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 85.446181 | 7.249146 | 11.7871 | .0000 |
| sex7780 | 1.951175 | 0.991826 | 1.9673 | .0521 |
| age7780 | -0.942025 | 0.342283 | -2.7522 | .0071 |
| eng7780 | -0.392629 | 0.257842 | -1.5227 | .1312 |
| mat7780 | -1.453266 | 0.284458 | -5.1089 | .0000 |
| sci7780 | -0.163569 | 0.198015 | -.8260 | .4109 |
| com7780 | -0.170692 | 0.163573 | -1.0435 | .2994 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 895.96711 | 6 | 149.32785 | 9.67472 | .00000 |
| ERROR | 1435.4404 | 93 | 15.4348 | | |
| TOTAL (CORR.) | 2331.4075 | 99 | | | |

```
R-SQUARED = 0.384303
R-SQUARED (ADJ. FOR D.F.) = 0.344581
STND. ERROR OF EST. = 3.92872

NUMBER OF RESIDUALS = 100
SAMPLE AVERAGE = 1.75504E-14
SAMPLE VARIANCE = 14.4994
SAMPLE STANDARD DEVIATION = 3.80781
COEFF. OF SKEWNESS = 0.132733   STANDARDIZED VALUE = 0.541879
COEFF. OF KURTOSIS = 2.98818   STANDARDIZED VALUE = -0.0241276
DURBIN-WATSON STATISTIC = 2.46889
```

## 4.4 The General (Optimal) model

### 4.4.1 Introduction

From the analysis of the two samples the variables that appear to be significant predictors of performance are (in order of significance):

| 1974/77 | 1977/80 |
|---|---|
| Mathematics | Mathematics |
| Age | Age |
| Science | Sex |
| Commerce | English |
| | Commerce |

The regression equations for the two samples are;

1974/77:

$$GPA = 74.98 - 0.45\ AGE - 1.75\ MAT - 0.26\ SCI - 0.26\ COM$$

1977/80:

$$GPA = 85.45 + 1.95\ SEX - 0.94\ AGE - 0.39\ ENG - 1.45\ MAT - 0.165\ SCI - 0.17\ COM.$$

The two equations are different and thus makes it difficult to decide which is the more accurate equation. On the basis of $r^2$ alone the two equations are not quite different as in 1974/77 the $r^2$ is 0.33 whereas in the 1977/80 case it is 0.38. To come up with a predictive equation it is therefore necessary to "compromise" between the two sample equations so as to have properties of both. To achieve this objective, the two samples were combined to form one sample of 200 observations in order to derive a 'general' equation. Only the variables identified as significant in either of the two samples were included in the correlation and regression analyses.

### 4.4.2 Correlation analysis

The results of the correlation analysis are presented in table 4.8 below.

TABLE 4.8

CORRELATION MATRIX , FOR OPTIMAL MODEL

| | GPA | Age | Sex | Eng | Mat | Com | S |
|---|---|---|---|---|---|---|---|
| GPA | 1.00000 | | | | | | |
| Age | -.16588* | 1.00000 | | | | | |
| Sex | .08216 | .30371* | 1.00000 | | | | |
| Eng | -.07266 | -.02971 | .09666 | 1.00000 | | | |
| Mat | -.54175* | .09494 | -.03531 | .03215 | 1.00000 | | |
| Com | -.25560* | -.11153 | -.43356* | -.00128 | .25909* | 1.00000 | |
| Sci | -.18806* | -.06437 | -.00071 | .26222* | .20291* | .00247 | 1 |

*significant at 0.05

The correlation matrix shows that some of the variables are inter-related. The following variables are inter-related, Sex and Commerce (0.43), English and Science (0.26), Commerce and Mathematics (0.26) and Science and Mathematics (0.20). It is difficult to explain why English and Science are correlated. Because of the strong correlation between Sex and Commerce, and its low correlation with GPA (0.08), Sex are removed from the regression analysis. Note that the correlation between Commerce and GPA is (-0.26). English was also removed due to its low correlation with GPA (-0.07).

4.4.3 <u>Regression analysis</u>

Before the omission of Sex and English from the model, all the six variables were included in the regression analysis and the results are shown in table 4.9(A) next page. The t-values show that the two variables, Sex and English indeed appear insignificant except at $\alpha = 0.30$ and $\alpha = 0.46$ respectively. Because of this low significance, the two variables were therefore omitted from further analysis and a new regression analysis was run. The results are shown in table 4. 9 (B). From the results, the coefficient of multiple determination is the same as shown in table 4.9(A). So by dropping the two variables, there is nothing lost. All the variables in the model are significant at $\alpha = 0.10$ level.

Further, a stepwise regression analysis done was at the following levels; $\alpha = 0.05$, $\alpha = 0.10$ and $\alpha = 0.25$. At the $\alpha = 0.05$ level, ($F = 6.76$), only Mathematics was entered into the model. This is consistent with the results of earlier findings on the two different samples. The results of this analysis at $\alpha = 0.05$ are shown in table 4.10 (A). The inclusion of Mathematics emphasizes that this variable

**TABLE $.)(A): MULTIPLE REGRESSION RESULTS**

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 78.469191 | 4.990511 | 15.7237 | .0000 |
| sex | 0.803776 | 0.834539 | .9631 | .3366 |
| age | -0.610912 | 0.234416 | -2.6061 | .0099 |
| eng | -0.145324 | 0.196856 | -.7382 | .4612 |
| mat | -1.615721 | 0.214409 | -7.5357 | .0000 |
| sci | -0.214111 | 0.14733 | -1.4533 | .1477 |
| com | -0.248954 | 0.137401 | -1.8119 | .0715 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 1768.6604 | 6 | 294.7767 | 16.5583 | .0000 |
| ERROR | 3435.8554 | 193 | 17.8024 | | |
| TOTAL (CORR.) | 5204.5158 | 199 | | | |

R-SQUARED = 0.339832

R-SQUARED (ADJ. FOR D.F.) = 0.319309

STND. ERROR OF EST. = 4.21928

NUMBER OF RESIDUALS = 200

SAMPLE AVERAGE = -1.27542E-14

SAMPLE VARIANCE = 17.2656

SAMPLE STANDARD DEVIATION = 4.15519

COEFF. OF SKEWNESS = 0.213364 STANDARDIZED VALUE = 1.23185

COEFF. OF KURTOSIS = 3.30049 STANDARDIZED VALUE = 0.867454

DURBIN-WATSON STATISTIC = 1.84976

TABLE4.9(B): MULTIPLE REGRESSION RESULTS, OPTIMAL MODEL

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 77.48639 | 4.900672 | 15.8114 | .0000 |
| age | -0.546303 | 0.224659 | -2.4317 | .0159 |
| sci | -0.241409 | 0.142008 | -1.7000 | .0907 |
| com | -0.305014 | 0.124339 | -2.4531 | .0150 |
| mat | -1.601736 | 0.213674 | -7.4961 | .0000 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

RESIDUALS PLACED IN VARIABLE: RESIDUALS

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 1745.1543 | 4 | 436.2886 | 24.5931 | .0000 |
| ERROR | 3459.3615 | 195 | 17.7403 | | |
| TOTAL (CORR.) | 5204.5158 | 199 | | | |

R-SQUARED = 0.335315
R-SQUARED (ADJ. FOR D.F.) = 0.321681
STND. ERROR OF EST. = 4.21193
NUMBER OF RESIDUALS = 200
SAMPLE AVERAGE = -1.15108E-14
SAMPLE VARIANCE = 17.3837
SAMPLE STANDARD DEVIATION = 4.16938
COEFF. OF SKEWNESS = 0.249053 STANDARDIZED VALUE = 1.43791
COEFF. OF KURTOSIS = 3.38031 STANDARDIZED VALUE = 1.09787
DURBIN-WATSON STATISTIC = 1.85719

is indeed a significant predictor variable. At a level of $\alpha = 0.10$, ($F = 3.89$) two more variables, Age and Commerce were included in the regression model. These additional variables increased the value of $r^2$ from 0.29 (0.29, adjusted) to 0.33 (0.32 adjusted), see table 4.10 (B) below. Lastly, at a level of $\alpha = 0.25$ ($F = 2.73$), Science was included into the model increasing the value of $r^2$ from 0.33 (0.32 adjusted) to 0.34 (0.32 adjusted) see table 4.11(C).

4.4.4 <u>Conclusion</u>

From the foregoing, the four most significant predictor variables are, Mathematics, Commerce, Science and Age. The predictive equation that therefore can "best" explain performance in the Bachelor of Commerce programme is:

$$GPA = 77.49 - 1.6\ MAT - 0.55\ AGE - 0.245\ SCI - 0.31\ COM$$

From the coefficients of the academic scores it would appear as if the appropriate weighting method should be

Mat : 1.6

Sci : 0.245

Com : 0.31

Rounding these figures gives a weighting of 1.5:0.3:0.3 which is the same as:

Mat 5

Sci 1

Com 1

## TABLE 4.10(A): STEPWISE REGRESSION, OPTIMAL MODEL

```
                              STEPWISE REGRESSION
SELECTION:   FORWARD                                        CONTROL:      AUTOMATIC
F-TO-ENTER =  6.76          MAX STEPS =  50                 F-TO-REMOVE = 6.76
                                   STEP 1
R-SQUARED = 0.293489
R-SQUARED (ADJ.) = 0.28992                    MSE = 18.571 WITH 198 D.F.
 VARIABLES CURRENTLY IN MODEL             VARIABLES CURRENTLY NOT IN MODEL
VARIABLE                 COEFF.  F-REMOVE VARIABLE         PARTIAL CORR.   F-ENTER
 4. mat                -1.85245   82.2503  1. age                -.1368    3.7556
                                           2. sex                 .0750    1.1155
                                           3. eng                -.0658     .8556
                                           5. com                -.1419    4.0507
                                           6. sci                -.0949    1.7915
```

```
--------------------------------------------------------------------------------
                              MODEL FITTING RESULTS
--------------------------------------------------------------------------------
VARIABLE                      COEFFICIENT  STND. ERROR    T-VALUE  PROB(>|T|)
--------------------------------------------------------------------------------
CONSTANT                        64.176519     0.849084    75.5833       .0000
mat                             -1.852453     0.204258    -9.0692       .0000
--------------------------------------------------------------------------------
0 CASES WITH MISSING VALUES WERE EXCLUDED.
--------------------------------------------------------------------------------
                 ANALYSIS OF VARIANCE FOR THE FULL REGRESSION
--------------------------------------------------------------------------------
SOURCE             SUM OF SQUARES    DF  MEAN SQUARE   F-RATIO    PROB(>F)
MODEL                   1527.4663     1    1527.4663   82.2503       .0000
ERROR                   3677.0495   198      18.5710
--------------------------------------------------------------------------------
TOTAL (CORR.)           5204.5158   199

R-SQUARED = 0.293489
R-SQUARED (ADJ. FOR D.F.) = 0.28992
STND. ERROR OF EST. = 4.3094
NUMBER OF RESIDUALS = 200
SAMPLE AVERAGE = -8.81073E-15
SAMPLE VARIANCE = 18.4776
SAMPLE STANDARD DEVIATION = 4.29856
COEFF. OF SKEWNESS = 0.333771   STANDARDIZED VALUE = 1.92703
COEFF. OF KURTOSIS = 3.45752   STANDARDIZED VALUE = 1.32075
DURBIN-WATSON STATISTIC = 1.83234
```

## TABLE 4.10(B): STEPWISE REGRESSION, OPTIMAL MODEL

SELECTION: FORWARD CONTROL: AUTOMATIC
F-TO-ENTER = 3.89 MAX STEPS = 50 F-TO-REMOVE = 3.89

STEP 3

R-SQUARED = 0.325465
R-SQUARED (ADJ.) = 0.31514 MSE = 17.9114 WITH 196 D.F.

| VARIABLES CURRENTLY IN MODEL | | | VARIABLES CURRENTLY NOT IN MODEL | | |
|---|---|---|---|---|---|
| VARIABLE | COEFF. | F-REMOVE | VARIABLE | PARTIAL CORR. | F-ENTER |
| 1. age | -.51026 | 5.1552 | 2. sex | .0618 | .7464 |
| 4. mat | -1.68166 | 64.4698 | 3. eng | -.0744 | 1.0860 |
| 5. com | -.29106 | 5.4509 | 6. sci | -.1208 | 2.8899 |

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 75.751064 | 4.81623 | 15.7283 | .0000 |
| age | -0.510258 | 0.224732 | -2.2705 | .0243 |
| mat | -1.68166 | 0.20944 | -8.0293 | .0000 |
| com | -0.291057 | 0.124664 | -2.3347 | .0206 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 1693.8870 | 3 | 564.6290 | 31.5235 | .0000 |
| ERROR | 3510.6288 | 196 | 17.9114 | | |
| TOTAL (CORR.) | 5204.5158 | 199 | | | |

R-SQUARED = 0.325465
R-SQUARED (ADJ. FOR D.F.) = 0.31514
STND. ERROR OF EST. = 4.23218

NUMBER OF RESIDUALS = 200
SAMPLE AVERAGE = -1.05871E-14
SAMPLE VARIANCE = 17.6414
SAMPLE STANDARD DEVIATION = 4.20016
COEFF. OF SKEWNESS = 0.27676 STANDARDIZED VALUE = 1.59787
COEFF. OF KURTOSIS = 3.32587 STANDARDIZED VALUE = 0.940697
DURBIN-WATSON STATISTIC = 1.82059

**TABLE 4.10(C): STEPWISE REGRESSION, OPTIMAL MODEL**

```
SELECTION:   FORWARD                                   CONTROL:      AUTOMATIC
F-TO-ENTER =  2.73          MAX STEPS =  50            F-TO-REMOVE = 2.73
                                  STEP 4
R-SQUARED = 0.335315
R-SQUARED (ADJ.) = 0.321681                  MSE = 17.7403 WITH 195 D.F.
 VARIABLES CURRENTLY IN MODEL               VARIABLES CURRENTLY NOT IN MODEL
VARIABLE                COEFF.  F-REMOVE VARIABLE        PARTIAL CORR.   F-ENTER
  1. age               -.54630    5.9131  2. sex                  .0632     .7772
  4. mat              -1.60174   56.1923  3. eng                 -.0450     .3929
  5. sci               -.24141    2.8899
  6. com               -.30501    6.0176
```

MODEL FITTING RESULTS

| VARIABLE | COEFFICIENT | STND. ERROR | T-VALUE | PROB(>\|T\|) |
|---|---|---|---|---|
| CONSTANT | 77.48639 | 4.900672 | 15.8114 | .0000 |
| age | -0.546303 | 0.224659 | -2.4317 | .0159 |
| mat | -1.601736 | 0.213674 | -7.4961 | .0000 |
| sci | -0.241409 | 0.142008 | -1.7000 | .0907 |
| com. | -0.305014 | 0.124339 | -2.4531 | .0150 |

0 CASES WITH MISSING VALUES WERE EXCLUDED.

ANALYSIS OF VARIANCE FOR THE FULL REGRESSION

| SOURCE | SUM OF SQUARES | DF | MEAN SQUARE | F-RATIO | PROB(>F) |
|---|---|---|---|---|---|
| MODEL | 1745.1543 | 4 | 436.2886 | 24.5931 | .0000 |
| ERROR | 3459.3615 | 195 | 17.7403 | | |
| TOTAL (CORR.) | 5204.5158 | 199 | | | |

```
R-SQUARED = 0.335315
R-SQUARED (ADJ. FOR D.F.) = 0.321681
STND. ERROR OF EST. = 4.21193
NUMBER OF RESIDUALS = 200
SAMPLE AVERAGE = -1.19016E-14
SAMPLE VARIANCE = 17.3837
SAMPLE STANDARD DEVIATION = 4.16938
COEFF. OF SKEWNESS = 0.249053   STANDARDIZED VALUE = 1.43791
COEFF. OF KURTOSIS = 3.38031   STANDARDIZED VALUE = 1.09787
DURBIN-WATSON STATISTIC = 1.85719
```

## 4.5 Model Validation

### 4.5.1 Introduction

Before the model could be used for validation purposes tests for any serious violations of the regression assumption was done. The problem of multicollinearity has already been dealt with.

The test for heteroscedasticity (presence of serial correlation among the residuals) was also done. To perform this test the Durbin-Watson statistic is used. The calculated Durbin-Watson statistic of 1.857 (see table 4-10(C)is greater than the critical Durbin-Watson statistic (1.1810) and so we fail to reject the null hypothesis and therefore conclude that the error terms are independent where:

$H_o$: the error terms are independent

$H_a$: the error terms are dependent.

The t-test for the significance of the correlation coefficient r was also done. The t-value was calculated as shown below:

$$t = \frac{r - 0}{\sqrt{\frac{(1 - r^2)}{(n - 2)}}}$$

Statement of hypotheses

$H_o$: $r = 0$

$H_a$: $r \neq o$

Calculations

$$t = \frac{0.5672 - 0}{\sqrt{\dfrac{1 - 0.3217}{200 - 2}}}$$

$$= \frac{0.5672}{0.0614}$$

$$= 9.2378$$

The critical value of t at $\alpha = 0.05$ is 1.645. Since the calculated value is greater than the critical value, we reject the null hypothesis and conclude that there is multiple correlation. The analysis of variance results were also used to test statement of the hypothesis

$H_o : B_1 = B_2 = B_3 = B_4 = 0$

$H_a :$ At least one $B_i \neq 0$

The calculated F-ratio is 24.5931 (see table 4.10(C). From the ANOVA table in table 4.10(C) the probability that the critical F-value is greater than the calculated F-value is zero so we reject the null hypothesis and conclude there is linear regression.

The model therefore does not violate the regression assumptions, and so it can be applied to the validation sample to test its accuracy.

4.5.2 Validation of the Model

A sample of size 100, from the 1981/82 to 1983/84 academic years was used for the validation of the model. The scores of the predictor variables were entered into a computer program (see appendix D) for the purpose of predicting the GPA. The results are shown on Appendix E.

To test the aptness of the model, a 95% confidence interval was constructed about each of the actual GPA scores and it turned out that 88 out of the total 100 predictions were within the 95% confidence interval, which is a good prediction.

Therefore in applying this model, on the average a very small percentage of the predictions will be expected to be significantly different from the actual GPA scores. Of course we have to make exceptions for the outliers.

# CHAPTER 5

# CONCLUSION AND RECOMMENDATIONS

## 5.1 Conclusions

From the results of the study, several conclusions can be drawn. One of this is that there are several factors that bear a significant correlation with the students' performance at University level. However, not all of these factors relate to academic achievements before entry into the University. For example age has been found to have a strong negative correlation with performance, with the younger students performing relatively much better than the older students. Sex also has a significant correlation with performance with males performing relatively better than the females.

Among the pre-entry academic scores, the following scores did not seem to have any significant association with performance: English, Other languages, General Sciences, Liberal arts (History, Geography and Religious Education). The rest of the academic score variables however had significant association with performance in the Bachelor of Commerce programme (Science, Mathematics, Commerce and O-level aggregate score).

The results of this study strongly indicate that the most important factor for predicting the performance of Bachelor of Commerce students is O-level Mathematics score. This therefore implies that the selection criteria should emphasize on "O"-level Mathematics. However, the results also indicate that other factors had strong correlation with the student's performance. These factors are Age, O-level Commerce (Economics, Commerce and Accounts), O-level Science and O-level aggregate score.

Despite the relatively good prediction power of the model, the explanatory power ($r^2$) is however quite low. This is possibly due to the omission of important variables such as performance at A-level, performance in the first two years at University, degree options taken and hours of self-study, variables which might considerably increase the predictive ability of the model but which were irrelevant in the current study because the focus was on pre-entry performance predictors.

## 5.2 Recommendations of the Study

Several recommendations follow from the foregoing conclusions. Although the current admission criteria takes into consideration the most important factor, the process could be further improved by including more requirements into the admission criteria. Though age has a significant correlation with performance, its inclusion in the selection process would be discriminatory. The other variables should however be included in the selection criteria.

The '0'-level aggregate score can be incorporated to determine the cut-off points so that only those who attain the cut-off points or above can get admitted.

Lastly, the 'O'-level Commerce score can also be incorporated into the admission criteria by setting the minimum grade to be attained in these subjects. However, students will need to be counselled against assuming that since they did Commerce at secondary school they can relax. This has been found to be detrimental to the students' performance (see the Baldwin and Howe study, 1982).

In incorporating the above variables for selection purposes the following weightings are suggested: Mathematics (x5) Commerce (x1) Science (x1) and the rest (x0).

5.3 Limitations of the Study

The generalization of the results in this study should be done after careful consideration of the limitations below.

Although this study utilized robust statistical methods and also included as many variables as possible, the results only indicate associations between the variables and not causal relationships.

The fact that several of the variables were found to be statistically insignificant might have several interpretations. There might, in fact, be no association or there might be an association but the statistical tools and tests used in the study lacked sufficient power to detect the association or the nature of the data made the identification of the actual relationships impossible.

Further, in a study such as this one, it is not always possible to come up with all the possible variables that might explain academic performance. Certain data cannot be quantified, for example, the motivation to study. Also certain important data though quantifiable are difficult to collect such as hours spent by the students on self-study. Lastly, the data analysed was for the students who had gone through the 7-6-3 system but since the system is changing, certain aspects might change and might thus affect the prediction model developed, for example, the grading system.

Therefore in attempting to generalize the findings, one should take into consideration the above limitations.

## 5.4 Suggestions for further research

In this study, regression and correlation analyses were used. Other tools such as factor and discriminant analysis could also be used in such a study.

Since the population of study was from students who had been through the 7-6-3 education system, the same study could be carried out with the 8-4-4 education programme when it becomes operational.

Future research could also be done to examine the relationship between aptitude tests and performance in the Faculty since this variable is considered important in many universities world-wide especially in the Commerce Faculties.

And lastly, such factors as performance in the first two years at University and options taken could be incorporated with pre-entry performance predictors in a further study.

# APPENDIX A: COURSES OFFERED IN THE FACULTY

## FIRST YEAR

**Required full courses:**
D 100 Introduction to Economics
D 101 Business Law I
D 102 Fundamentals of Accounting
D 103 Quantitative Methods I
**Required Half Courses:**
D 104 Business Studies
D 105 Behavioural Science I

## SECOND YEAR

**(Accounting Option)**
Required full courses
D 200 Economic Theory
D 201 Intermediate Accounting

**Required half courses:**
D 202 Behavioural Science II
D 203 Managerial Accounting
D 204 Organization Theory
D 205 Finance I
D 206 Computing Science I
D 207 Business Statistics I

**Business Administration Option:**
**Required full course.:**
D 200 Economic Theory

**Required half courses:**
D 202 Behavioural Science II
D 203 Managerial Accounting
D 204 Organization Theory
D 205 Finance I
D 206 Computing Science I
D 207 Business Statistics I
Elective Course. The equivalent of one full course from the list of Approved Courses.

**Insurance Option:**
**Required full course:**
D 200 Economic Theory

**Required half courses:**
D 202 Behavioural Science II
D 203 Managerial Accounting
D 204 Organization Theory
D 205 Finance I
D 206 Computing Science I
D 207 Business Statistics I
D 209 Elements of Risk and Insurance
D 20C Introduction to Insurance Law

## APPROVED ELECTIVES

D 201 Intermediate Accounting (non-accounting option only).
D 208 Marketing I
D 209 Elements of Risk and Insurance (non insurance option only)
D 20c Introduction to Insurance Law (non-insurance option only)

## THIRD YEAR

**Accounting Option:**
**Required full courses:**
D 300 Advanced Accounting
D 301 Auditing
D 302 Business Policy and Decisions

**Required Half Courses:**
D 303 Taxation
D 304 Cost Accounting

**Elective courses:**
Equivalent of two full courses

**Business Administration Option:**
**Required full courses:**
D 302 Business Policy and Decisions.

**Elective courses:**
The equivalent of five full courses.

**Insurance Option**
**Required full courses:**
D 302 Business Policy and Decisions
D 319 Liability Insurance
D 320 Assurance of the Person
D 208 Marketing I (Second Year Course).

**Required half courses:**
D 321 Property Insurance
D 322 Elements of Actuarial Science

Source: University of Nairobi Calendar, 1986/87 Academic year pp. 376-378.

APPENDIX A (contd.)

**Elective Courses:**
**The equivalent of one full course.**

**APPROVED ELECTIVES**

D 303 Taxation
D 304 Cost Accounting
D 305 Marketing II
D 306 Banking Practice and Law
D 307 Quantitative Methods II
D 308 Business Statistics
D 309 Computing Science II
D 310 Systems Analysis
D 311 Company Law.
D 312 Finance II
D 313 International Marketing
D 314 Labour Relations and Law
D 315 Management of Co-operatives
D 316 Insurance Practice and Law
D 317 Personnel Administration
D 318 Accounting Theory
D 319 Liability Insurance
D 320 Assurance of the Person
D 321 Property Insurance
D 322 Elements of Actuarial Science
D 323 Marine Insurance
D 324 Life Insurance
D 325 Pensions Scheme
D 326 Motor Vehicle Insurance
D 327 Consequential Loss Insurance
D 328 Risk Theory
D 329 Introductory Econometrics
D 32B Aviation Insurance
D 32C Demographic Statistics
C 204 Comparative Economic Systems
C 205 Agricultural Economics
C 300 Economic Development
C 304 Economics of Industry and Labour
C 305 Money, Banking and Finance
C 306 International Economics

# APPENDIX B: EXAMINATION REGULATIONS

**Examination Regulations**

3. Final examinations in all courses are University examinations.

4. Final examinations in all courses will form 70 per cent of the basis on which the degree is awarded, the other 30 per cent being coursework, including term assignments, tests and papers.

5. The grade obtained in all courses will be classified and published.

6. A candidate who fails in the equivalent of no more than two full courses in the University Examinations prescribed for any year may, on the recommendation of the Board of Exxaminers to the University Senate, be admitted to Supplementary Examinations within a period of four months after the end of the academic year. A candidate who passes his required Supplementary Examinations is deemed to have passed the University Examinations for the year.

7. A candidate who satisfies the Board of Examiners on *either* his University Examinations *or* his Supplementary Examinations may, on the recommendation of the Board of Examiners to the University Senate, by admitted to the following year and, in the case of candidates in the final year, be considered as a candidate for the award of the degree.

8. A candidate, in any year of the programme, who fails to satisfy the Board of Examiners on the equivalent of more than two full courses at the University Examinations may, on the recommendation of the Board of Examiners to the University Senate, normally be required.

   (a) On not more than one occasion, to repeat the year internally and resit the University Examinations at their next scheduled resitting provided that the candidate has not previously repeated the year internally.
   (b) To be discontinued from the University.

9. A candidate who fails to satisfy the Board of Examiners at the University Supplementary Examinations may, on the recommendation of the Board of Examiners to the University Senate, normally be required:

   (a) On not more than one occasion, to repeat the year internally and resit the University Examinations at their next scheduled resitting provided that the candidate has not previously repeated the year internally.

   (b) To be discontinued from the University.

10. A final year candidate who has failed in not more than one full course in the University Examinations or in not more than one elective course in Supplementary Examinations, and whose overall performance in the other papers is 50% or above may, on the recommendation of the Board of Examiners to the University Senate, be awarded a Pass Degree.

11. A candidate who passes the University Examinations prescribed for the end

Source: University of Nairobi Calendar 1986/87 pp. 379.

APPENDIX B (Contd.)

of the third year of study and who, in other respects qualifies for the award of the degree, shall be placed in one of the three classes to be described as First, Second (Upper Division or Lower Division) and pass. Honours shall be awarded to a candidate whose name is placed in the First Class or Second Class (Upper or Lower Division).

11. Candidates who write Supplementary Examinations in the third year shall not be eligible for the award of Honours.

12. A candidate who qualifies for the award of the degree only after repeating the entire University Examinations for either the second or third year of study shall not be eligible for the award of Honours in terms of Regulations

13. The Bachelor of Commerce degree is classified on the basis of the percentage grades obtained by the candidates in all courses taken in Commerce II and Commerce III of the programme. In the determination of the degree classification, the percentage grades obtained in the second year courses (Commerce II) will be given a weight of 0.5 and the percentage grades obtained in the third year courses (Commerce III) will be given a weight of 1.0.

14. Ther Bachelor of Commerce degree certificate is inscribed as either in "Accounting Option", "Business Administration Option" or "Insurance Option" depending upon the candidate's study programme.

# APPENDIX C: STEPWISE REGRESSION PROCEDURE

## Stepwise Multiple Linear Regression

When the independent variables that should be included in the regression model are not known, stepwise multiple linear regression is the most efficient way to determine the variables and regression coefficients. Because of the number of computations, a computer must be used to perform the analysis. We cover only the logic of the method here. By definition,

> *Stepwise* multiple linear regression is a technique that places only the independent variables into the regression equation that remove a significant portion of the variation in $Y$, the dependent variable. The criterion used is the amount of variability (sum of squares) removed by each variable, as measured by the $F$ test, *independent* of the order in which the variable enters solution.

Initially the candidate variable must have the maximum correlation between itself and $Y$ to be selected. Correlation, a quantitative measure of the relation between variables, is studied in Sec. 28.2. The steps below track the progress of a computerized stepwise program. For convenience, variables are called $Y$ (response), $X_1$ (first entering variable), $X_2$ (second entry), etc. The value of the regression coefficients $a, b_1, \ldots$ changes with each new model.

1. Enter into solution the $X$ variable having the highest correlation with $Y$, because it will explain the most variability in $Y$. The model is now

$$\hat{Y} = a + b_1X_1$$

2. Select as $X_2$ the $X$ variable having the highest correlation with $Y$ after the effect of $X_1$ is accounted for by the regression model. Use the $F$ test to determine if more variability could have been explained if $X_2$ had entered *prior* to $X_1$. If it had, $X_1$ is removed and the model is

$$\hat{Y} = a + b_2X_2$$

   If more is removed with the order $X_1$ than $X_2$, the model is

$$\hat{Y} = a + b_1X_1 + b_2X_2$$

3. Select as $X_3$ the $X$ variable having the largest correlation with $Y$ after $X_1$ and $X_2$ are accounted for. Again use the $F$ test to see if some other entry order of $X_1$ and $X_2$ removes more variability. The orders checked are $X_1$ after $X_2$ and $X_3$, and $X_2$ after $X_1$ and $X_3$. Remove $X_1$ and/or $X_2$ from the model if the test does not justify their presence.
4. Continue this routine until either no more $X$ variables remove a significant amount of variability or no new $X$ variables remain to be considered.

See the Draper and Smith text for complete details and computer examples of stepwise regression.

Source: Blank, L.: Statistical Procedures for Engineering Management and Science - McGraw Hill Book Company 1980. pp. 502-503.

## APPENDIX D: COMPUTER PROGRAMME TO VALIDATE THE MODEL

```
1 REM          PROGRAM TO VALIDATE PERFORMANCE PRDICTION MODEL
2 REM          ==========================================
3 LET W=O             :REM w=no. of predictions within 95% confidence interval
5 LPRINT     "---------------------------PROJECT--------------------------"
6 LPRINT "---------------------------------------------------------------"
7 LPRINT "No.";"OBSERVED","LOWER LIM.","PREDICTED","UPPER LIM","RESIDUAL"
8 LPRINT "-------------------------------------------------------"
9 FOR I=1 TO 100
10 READ  A,M,S,C,G
11 REM                    a=age
12 REM                    m=mathematics score
13 REM                    s=science score
14 REM                    c=commerce score
15 REM                    g=actual g.p.a.
16 REM                    p=predicted g.p.a.
20 LET P=77.48639 -(.546303*A) -(1.601736*M) -(.241409*S) -(.305014*C)
22 LET R=G-P           :REM r=residual
25 LET L=G-(1.96*4.21193)        :REM l=lower confidence limit
26 LET U=G+(1.96*4.21193)        :REM p=upper confidence limit
27 IF P>=L THEN 29
28 GOTO 290
29 IF P<=U THEN 31
30 GOTO 290
31 LET W=W + 1
35 DATA 23,2,1,1.5,65.2,20,1,5,3.5,73.6,19,6,7,2,60.8,21,2,5,4,59.1
40 DATA 19,3,4,5,59.6,20,2,3,0,55,20,1,3,3,56.2,22,6,5,1.5,53.3,20,1,0,1.5,65.3
50 DATA 23,3,3,3,62.9,20,2,2,3,65.5,22,5,5.5,3,59.3,22,1,3,3,70.1
60 DATA 19,1,5,0,50.5,23,6,6,3,58.8,21,6,5,0,60,19,4,3,0,59.7,19,3,4.5,0,55.4
70 DATA 22,3,6,3,59.4,21,3,5,3,61.6,19,3,5,3,64.8,19,4,6,3,64.4,19,1,6.7,5,65.4
80 DATA 22,5,6,0,58.1,20,6,5.5,0,54.7
90 DATA 20,4,6,0,61.5,21,3,0,3,62.7,20,3,6,3,61.8,21,4,5,0,58.7,21,6,2,2,59.7
100 DATA 20,1,6.5,3,63.4,22,5,3,0,63.0,20,5,5,3.5,57.1,19,3,5,3,71.1
110 DATA 21,3,3,3.3,54.4,21,3,4,0,62.9,20,6,5.5,0,54.7,21,5,7,4,59.8
120 DATA 20,6,4,3,54.5,19,4,5.5,4,56.8,25,6,5,1.5,59.4,22,2,3,3,64.2
130 DATA 20,1,5.5,3,59.1,23,3,4,5,56.7,20,3,0,3,63.6,20,3,6,2.5,69.7
140 DATA 20,3,4.5,0,60.5,21,3,3,3,64.5,21,4,6.5,0,63.5,20,1,2,0,58.7
150 DATA 23,3,2,3.5,54.8,24,3,8,3,53,23,5,4,4,60.5,22,3,5,3,51.7
160 DATA 22,4,3.5,1.5,52.4,21,6,7.5,3,55.8,22,4,4.5,0,55.8,21,3,4,4,63.1
170 DATA 21,6,4.5,4,63.9,23,3,6.3,7.5,57.6,21,6,3,6,58.9,25,4,3,5.5,56.6
180 DATA 20,2,4.3,0,62.7,23,2,3,1,57.9,21,4,8,1.5,59.6,22,3,5,3.7,49.1
190 DATA 21,4,2,0,68.8,21,3,6,3,60.7,21,3,3,3,65.1,21,6,1,2.5,56.8
200 DATA 25,6,3.5,4,65,21,5,5.5,1.5,58.2,20,3,6,5.5,55,23,4,7,6,55.2
210 DATA 21,6,6,7.5,59.1
220 DATA 20,6,3.5,0,54.2,22,3,4,3,56.7,20,2,4.5,2,59.4,23,4,5,4,52.6
230 DATA 24,5,6.3,5,47.4,21,1,3.5,0,52,20,1,3,0,60.1,20,6,4.7,0,49.3
240 DATA 23,6,4.5,3,51.1,22,5,5,4,54.5,21,2,2,3,58.5,21,3,2.3,1.5,77.4
250 DATA 22,4,3.3,3,62.9,21,6,4,3,60.9,22,1,5,2.7,55.8,22,5,6,0,53.6
260 DATA 21,3,5,5,63.2,22,6,7.3,2,57.9,22,2,4,3,57,21,3,4.3,0,51.2
270 DATA 23,6,5,3,58.9,25,5,3,4,56,22,1,1.3,1.5,70.6,22,1,3.5,3,53.8
280 DATA 22,5,3.5,4,54.2
290 LPRINT I; G,L,P,U,R
300 NEXT I
301 LPRINT "---------------------------------------------------------------"
303 LPRINT "---------------------------------------------------------------"
304 LPRINT
305 LPRINT "Number of predictions within 95% confidence interval  :";W
310 LPRINT "--------------------------------------------------------------"
320 STOP
```

## APPENDIX E: RESULTS FROM COMPUTER PROGRAMME FOR VALIDATION OF THE MODEL

PROJECT

| No. | OBSERVED | LOWER LIM. | PREDICTED | UPPER LIM | RESIDUAL |
|---|---|---|---|---|---|
| 1 | 65.2 | 56.94462 | 61.01901 | 73.45538 | 4.180985 |
| 2 | 73.6 | 65.34462 | 62.684 | 81.85538 | 10.916 |
| 3 | 60.8 | 52.54462 | 55.19633 | 69.05538 | 5.603676 |
| 4 | 59.1 | 50.84462 | 60.38345 | 67.35538 | -1.283451 |
| 5 | 59.6 | 51.34462 | 59.81072 | 67.85538 | -.2107163 |
| 6 | 55 | 46.74462 | 62.63264 | 63.25539 | -7.632633 |
| 7 | 56.2 | 47.94462 | 63.31933 | 64.45538 | -7.119324 |
| 8 | 53.3 | 45.04462 | 54.19274 | 61.55538 | -.8927422 |
| 9 | 65.3 | 57.04462 | 64.50108 | 73.55539 | .7989273 |
| 10 | 62.9 | 54.64462 | 58.47694 | 71.15538 | 4.423062 |
| 11 | 65.5 | 57.24462 | 61.959 | 73.75539 | 3.541 |
| 12 | 59.3 | 51.04462 | 55.21626 | 67.55538 | 4.083744 |
| 13 | 70.1 | 61.84462 | 62.22672 | 78.35538 | 7.87328 |
| 14 | 50.5 | 42.24462 | 64.29785 | 58.75539 | -13.79784 |
| 15 | 58.8 | 50.54462 | 52.9475 | 67.05538 | 5.852497 |
| 16 | 60 | 51.74462 | 55.19656 | 68.25539 | 4.80344 |
| 17 | 59.7 | 51.44462 | 59.97546 | 67.95538 | -.2754555 |
| 18 | 55.4 | 47.14462 | 61.21508 | 63.65539 | -5.815079 |
| 19 | 59.4 | 51.14462 | 58.29903 | 67.65538 | 1.100979 |
| 20 | 61.6 | 53.34462 | 59.08673 | 69.85538 | 2.513271 |
| 21 | 64.8 | 56.54462 | 60.17934 | 73.05539 | 4.620671 |
| 22 | 64.4 | 56.14462 | 58.33619 | 72.65538 | 6.063816 |
| 23 | 65.4 | 57.14462 | 62.36238 | 73.65538 | 3.037621 |
| 24 | 58.1 | 49.84462 | 56.0106 | 66.35538 | 2.089405 |
| 25 | 54.7 | 46.44462 | 55.62217 | 62.95539 | -.9221649 |
| 26 | 61.5 | 53.24462 | 58.70494 | 69.75539 | 2.795067 |
| 27 | 62.7 | 54.44462 | 60.29378 | 70.95538 | 2.406227 |
| 28 | 61.8 | 53.54462 | 59.39163 | 70.05538 | 2.408371 |
| 29 | 58.7 | 50.44462 | 58.40003 | 66.95538 | .2999687 |
| 30 | 59.7 | 51.44462 | 55.31076 | 67.95538 | 4.389241 |
| 31 | 63.4 | 55.14462 | 62.47439 | 71.65538 | .9256096 |
| 32 | 63 | 54.74462 | 56.73482 | 71.25539 | 6.265179 |
| 33 | 57.1 | 48.84462 | 56.27706 | 65.35538 | .8229408 |
| 34 | 71.1 | 62.84462 | 60.17934 | 79.35538 | 10.92067 |
| 35 | 54.4 | 46.14462 | 59.47805 | 62.65539 | -5.078041 |
| 36 | 62.9 | 54.64462 | 60.24318 | 71.15538 | 2.656822 |
| 37 | 54.7 | 46.44462 | 55.62217 | 62.95539 | -.9221649 |
| 38 | 59.8 | 51.54462 | 55.09543 | 68.05538 | 4.704571 |
| 39 | 54.5 | 46.24462 | 55.06924 | 62.75539 | -.5692368 |
| 40 | 56.8 | 48.54462 | 58.15188 | 65.05538 | -1.351879 |
| 41 | 59.4 | 51.14462 | 52.55384 | 67.65538 | 6.846169 |
| 42 | 64.2 | 55.94462 | 60.62499 | 72.45538 | 3.575012 |
| 43 | 59.1 | 50.84462 | 62.7158 | 67.35538 | -3.615803 |
| 44 | 56.7 | 48.44462 | 57.62551 | 64.95538 | -.9255028 |
| 45 | 63.6 | 55.34462 | 60.84009 | 71.85538 | 2.759915 |
| 46 | 69.7 | 61.44462 | 59.54414 | 77.95538 | 10.15586 |
| 47 | 60.5 | 52.24462 | 60.66879 | 68.75539 | -.1687851 |
| 48 | 64.5 | 56.24462 | 59.56955 | 72.75539 | 4.930455 |
| 49 | 63.5 | 55.24462 | 58.03792 | 71.75539 | 5.462082 |
| 50 | 58.7 | 50.44462 | 64.47578 | 66.95538 | -5.775776 |

APPENDIX E (contd.)

| | | | | | |
|---|---|---|---|---|---|
| 51 | 54.8 | 46.54462 | 58.56584 | 63.05538 | -3.765843 |
| 52 | 53 | 44.74462 | 56.7236 | 61.25539 | -3.723602 |
| 53 | 60.5 | 52.24462 | 54.72705 | 68.75539 | 5.772953 |
| 54 | 51.7 | 43.44462 | 58.54043 | 59.95539 | -6.840431 |
| 55 | 52.4 | 44.14462 | 57.75833 | 60.65539 | -5.35833 |
| 56 | 55.8 | 47.54462 | 53.678 | 64.05538 | 2.122002 |
| 57 | 55.8 | 47.54462 | 57.97444 | 64.05538 | -2.174442 |
| 58 | 63.1 | 54.84462 | 59.02313 | 71.35538 | 4.076874 |
| 59 | 63.9 | 55.64462 | 54.09721 | 72.15538 | 9.802792 |
| 60 | 57.6 | 49.34462 | 56.30773 | 65.85538 | 1.292271 |
| 61 | 58.9 | 50.64462 | 53.8493 | 67.15538 | 5.050709 |
| 62 | 56.6 | 48.34462 | 55.02007 | 64.85538 | 1.579933 |
| 63 | 62.7 | 54.44462 | 62.3188 | 70.95538 | .3811989 |
| 64 | 57.9 | 49.64462 | 60.6887 | 66.15538 | -2.7887 |
| 65 | 59.6 | 51.34462 | 57.21829 | 67.85538 | 2.381714 |
| 66 | 49.1 | 40.84462 | 58.32693 | 57.35538 | -9.226925 |
| 67 | 68.8 | 60.54462 | 59.12426 | 77.05539 | 9.675743 |
| 68 | 60.7 | 52.44462 | 58.84532 | 68.95538 | 1.854683 |
| 69 | 65.1 | 56.84462 | 59.56955 | 73.35538 | 5.530453 |
| 70 | 56.8 | 48.54462 | 55.39966 | 65.05538 | 1.400337 |
| 71 | 65 | 56.74462 | 52.15341 | 73.25539 | 12.84659 |
| 72 | 58.2 | 49.94462 | 56.22007 | 66.45538 | 1.979931 |
| 73 | 55 | 46.74462 | 58.6291 | 63.25539 | -3.629093 |
| 74 | 55.2 | 46.94462 | 54.99452 | 63.45539 | .2054787 |
| 75 | 59.1 | 50.84462 | 52.66755 | 67.35538 | 6.432453 |
| 76 | 54.2 | 45.94462 | 56.10499 | 62.45539 | -1.904984 |
| 77 | 56.7 | 48.44462 | 58.78184 | 64.95538 | -2.081841 |
| 78 | 59.4 | 51.14462 | 61.66049 | 67.65538 | -2.260491 |
| 79 | 52.6 | 44.34462 | 56.08737 | 60.85538 | -3.487373 |
| 80 | 47.4 | 39.14462 | 53.3205 | 55.65539 | -5.920494 |
| 81 | 52 | 43.74462 | 63.56735 | 60.25539 | -11.56735 |
| 82 | 60.1 | 51.84462 | 64.23437 | 68.35538 | -4.134369 |
| 83 | 49.3 | 41.04462 | 55.81529 | 57.55538 | -6.515293 |
| 84 | 51.1 | 42.84462 | 53.30962 | 59.35538 | -2.209618 |
| 85 | 54.5 | 46.24462 | 55.03195 | 62.75539 | -.5319481 |
| 86 | 58.5 | 50.24462 | 61.41269 | 66.75539 | -2.912689 |
| 87 | 77.4 | 69.14463 | 60.19606 | 85.65538 | 17.20395 |
| 88 | 62.9 | 54.64462 | 57.34909 | 71.15538 | 5.550911 |
| 89 | 60.9 | 52.64462 | 54.52293 | 69.15538 | 6.377075 |
| 90 | 55.8 | 47.54462 | 61.83541 | 64.05538 | -6.035404 |
| 91 | 53.6 | 45.34462 | 56.0106 | 61.85538 | -2.410595 |
| 92 | 63.2 | 54.94462 | 58.4767 | 71.45538 | 4.723301 |
| 93 | 57.9 | 49.64462 | 53.485 | 66.15538 | 4.415005 |
| 94 | 57 | 48.74462 | 60.38358 | 65.25539 | -3.383576 |
| 95 | 51.2 | 42.94462 | 60.17076 | 59.45539 | -8.970756 |
| 96 | 58.9 | 50.64462 | 53.18891 | 67.15538 | 5.71109 |
| 97 | 56 | 47.74462 | 53.87586 | 64.25539 | 2.124146 |
| 98 | 70.6 | 62.34462 | 63.09464 | 78.85538 | 7.505364 |
| 99 | 53.8 | 45.54462 | 62.10602 | 62.05538 | -8.306015 |
| 100 | 54.2 | 45.94462 | 55.39406 | 62.45539 | -1.194061 |

Number of predictions within 95% confidence interval : 88

# BIBLIOGRAPHY


Astin, A.W.: Predicting Academic Performance in Colleges (The Free Press, N.Y., 1971).

Baldwin, B.A. and Howe, K.R.: "Secondary level study of Accounting and Subsequent Performance in the First College Course" (The Accounting Review - Vol. LVII No. 3, July 1982).

Blank, L.: Statistical Procedures for Engineering Management and Science (McGraw Hill Book Company, 1980).

Blommers, P. and Lindquist, F.: Elementary Statistical Methods in Psychology and Education (University of London Press Ltd., 1965).

Canavos, G.C.: Applied Probability and Statistical Methods (Little, Brown and Company, Boston, 1984).

Chirchir, M.K.; Asiko, N.; Ojode, LA.; Omosa, F.Y.B.; An Analysis of Student Performance in the University of Nairobi: Case Study of the Faculty of Commerce. (Unpublished Term Paper, 1988)

Churchill, G.A., Jr.: Marketing Research - Methodological Foundations, 3rd Ed. (The Dryden Press, 1983).

Daniel, W.W. and Terrell, J.C.: Business Statistics: Basic Concepts and Methodology, 3rd Ed. (Houghton Miffin Co., Boston, 1983).

Dean's Office: Faculty of Commerce Handbook; 1982.

Delaney, P.R., Keys, D.E., Norton, C.L. and Simon, J.R.: An Admission Test for Intermediate Accounting (The Accounting Review, Vol. LIV, No. 1, Jan. 1979.

Dockweiler, R.C. and Willis, C.G.: On the use of Entry Requirements for Undergraduate Accounting Programs (The Accounting Review, July 1984 Vol. LIV, No.3).

Dunn, W.M. and Hall, T.W.: An Empirical Analysis of the Relationship between CPA Exam candidate attributes and candidate performance (The Accounting Review, Vol. LIV No. 4, Oct. 1984).

Frakes, A.H.: Introductory Accounting Objectives and Intermediate Accounting Performance (The Accounting Review, Vol. LII, No.1 Jan. 1977).

Gatumu, H.J.N.: A study of Predicting Validity of O-level, A-level and an aptitude test in relation to the Performance at the University of Nairobi. (Unpublished M.A. Thesis, U.o.N., 1976).

Geer, Van de: Introduction to Multivariate Analysis for the Social Sciences (W.H. Freeman and Co., 1971).

Hamburg, M.: Statistical Analysis for Decision-Making 3rd Ed. (Harcourt Brace Jovanovich, Inc., 1983).

Harnett, D.L.: Statistical Methods, 3rd Ed. (Addison-Wesley Publishing Co., Inc., 1982).

Harnett, D.L. and Murphy, J.L.: Statistical Analysis for Business and Economics - 3rd Ed. (Addison - Wesley Publishing Co., 1985.

Ingram, R.W. and Petersen, R.J.: An Evaluation of AICPA Tests for Predicting the Performance of Accounting Majors (The Accounting Review, Vol. LXII, No. 1 Jan. 1987).

Kane, E.J.: Economic Statistics and Econometrics: An Introduction to Quantitative Economics (Harper and Row, Publishers, N.Y., 1960).

Kendall, M.G. (Sir): Multivariate Analysis (Griffin Publishing Company, London, 1975).

Kimura, D. J.H.: An Evaluation of Performance Predictors for Accounting Option Students. (Unpublished Research Proposal, 1987).

Kohler, Dr. D.: Ad hoc Committee's Report on the High Failure Rates in Some of the Faculty's (Commerce) Examinations (1980).

Koutsoyiannis, A.: Theory of Econometrics: An Introductory Exposition of Econometric Methods, 2nd Ed. (Macmillan Publishers Ltd., London, 1977).

Mbeche, I.M.: An Investigation into Factors Contributing to Teaching Effectiveness Among Student Teachers. Unpublished M.A. Thesis, U.o.N. 1979).

National Foundation for Education Research (NFER): The Prediction of Academic Success. (N.F.E.R. Publishing Company Ltd., 1973).

Neter, J.: Wasserman, W.; and Kutner, M.H.: Applied Linear Statistical Methods: Regression, Analysis of Variance and Experimental Designs, 3rd Ed. (Richard D. Irwin, Inc., Homewood, Illinois, 1985).

Nie, N.H., Hull, C.H., Jenkins, I.G. Steinbrenner, K., and Bent, D.H.: Statistical Package for the Social Sciences (SPSS) (McGraw Hill Book Co., 1975).

Somercet, H.C.A.: Predicting Success in School Certificate - A Uganda Case Study (East African Publishing House, 1968).

Tatsuoka, M.M.: Multivariate Analysis - Techniques for Educational and Psychological Research (John Wiley and Sons, Inc., 1971).

University of Nairobi: 1986-7 University of Nairobi Calendar.

Williams, D.Z.: A Profile of CPA candidates (The Accounting Review, Jan. 1979).