11.

APPLICATION OF MULTIPLE LINEAR REGRESSION ANALYSIS ON ACADEMIC PERFORMANCE IN UNIVERSITIES

BY

NAREEBA CRISPUS

2018/KEP/0344/F

A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF BACHELOR OF EDUCATION WITH EDUCATION OF KABALE UNIVERSITY

FEBRUARY, 2022

12.

Declaration

This report is a result of my efforts and to the best of my knowledge, no part of it has been submitted for any award in any University.

Signature:……. Date: 22/02/22

NAREEBA CRlSPUS

2018/KEP/0344/F

13.

Approval

This report has been under my supervision and is submitted with my approval.

Signature:…….. Date: 27/02/22

DR AKUGIZIBWE EDWIN

Department of Mathematics,

Kabale University

Dedication

I dedicate this work to my parents Mr Ahimbisibwe Vicent and Mrs Tumwebaze Deziranta (MHSRIP) for the care and support they gave and still give me.

Acknowledgement

I wish to convey my sincere appreciation to the following people, whose assistance and guidance led to the accomplishment of this report:

First of all I wish to thank my parents who provided me everything I needed to accomplish my education.

My heartfelt thanks goes to my supervisor Dr Akugizibwe Edwin who guided me in every step of compiling research report

Finally great thanks goes to all my friends and course mates for the productive criticisms and discussions that made me to accomplish my research.

May God bless you all

14.

Table of Contents

15.

16.

Chapter One

Introduction

1.0 Introduction.

This chapter consisted of the background, statement of the problem, purpose of the study, objectives of the study, scope of the study, significance of the study and definition of key terms

1.1 Background of the Study

Often it is the case that mathematics provides a clear solution to a problem. However, when this is not the case, other methods must be applied so as to make decisions with some degree of certainty. Often, this is where statistics can be very useful. We can use statistical methods to classify, estimate, or predict the actual value of parameters with some degree of certainty that is satisfactory for that particular problem. Although these methods are known as statistical, the basis of their existence lies in mathematics (Gilbbert, 2002).

It is important to show and understand the theory behind these techniques before actually applying them. However, one cannot understand the theory involved without having some idea **of** which techniques are to be used. Only then can a proper procedure be chosen, broken down **and** then applied with all the restrictions and assumptions necessary to make the model both reliable and accurate.

With this in mind, the first step in any analysis should be to carefully define the problem at hand.

This may seem like a simple step, but it is critical and the overview will lead to a proper model. **In** any university, or school for that matter, there must be some way in placing students into classes that are suitable for the types of backgrounds and abilities they possess for example University grades students according to their faculties and departments. Most universities use a placement exam, entrance exam scores, high school grades or some combination of these to **place** students (Moore 2004). However, it is often the case that a student is misplaced. It is quite simple to see why this would happen. If a single exam score is used. It is only representative of **me** student's performance on that particular day. An ill student may place poorly. Also, most standardized tests given, including the Algebra and Calculus, are multiple choice. This brings in a "chance factor," that can place some students too high and force them to have poor or failing

grades in the class. Using only high school GPA in placing a student in a math class can cause a misplacement and a possible failure, or on the other end of the spectrum, it can cause a student to take a class that will not only bore them, but cause them to lose interest in the class altogether. This has been a topic of discussion among many Universities. (Sanders, R,1996) hence applicable to various Universities

In using a multiple regression, one must also consider the data. Questions arise as to what kind of data, and how it should be inputted, as well as which variables should be included in the regression. Understandably, variables that would be considered unethical to use will not be included. These include sex, race, social status, religion, as well as many others. The variables **that** will be included will be only those that reflect a student's ability based on past performance. Additionally, all data used in this analysis is totally confidentially. Students are listed by random **numbers** without any names or identification numbers (Antunes et al 2017)

The model chosen, least squares, can be solved in general by singular value decomposition. The **theory** behind this follows in the next section. However, this approach is only necessary if there **are** singular matrices involved. In the case of a nonsingular data matrix, the problem can be **solved** by a simpler derivation, resulting from a series of partial differential equations.

Once a model is refined, it can be applied to the problem. Data analysis will only give good results if the data is good to start with. Reducing sampling error is a huge factor in obtaining valid results. Even the best of models can be totally useless if the data is contaminated. Because **of** this, it is essential that all data is as random as possible and without bias. This can be a difficult task and one is often forced to work with restricted amounts of data, or with data that may be less than preferable. When this is the case, it is important to note any areas in which the data could cause problems (Bulmer,,1979).

Once the model and the data have been obtained, refined and reviewed, one may then apply the techniques to obtain results with some degree of confidence. It is at this stage that the results **must** be evaluated. The results can be tested statistically. Again, the mathematics behind the testing techniques are crucial. One must understand what the results actually represent and to

what degree they can be trusted. Statistical software can do the analysis and give results in the **form** of statistical tests (Mathworks,2002).

1.2 Problem Statement

Obviously, some students have the ability to do well but do not apply themselves, while others **with** less ability will study extremely hard and making a good grade. So, there is variability that **shows** up naturally. However, by looking at a large group of student's patterns, there should **appear** to be an estimate based on all of these different aspects of a student's past.

One could estimate a student's grade in a particular course, before they took that course, it would **be** wry helpful in placing them correctly. Therefore, the problem becomes quite clear. The goal is to find a model to estimate a student's grade in a particular course based on that student's background.

Once the problem is defined, a generalized model can be chosen and then altered if needed to fit the problem. It is at this point that assumptions must be made and restrictions applied to the **model** and the data, hopefully completing a model that will yield satisfactory results. If one wants to estimate a grade before a student in a class, multiple regression analysis can be used. This gave a linear estimate of a target variable (the estimated grade vs. the input variables (the student's background data). This kind of analysis is common in statistics and can be a powerful **tool** in making decisions when faced with some amount of uncertainty.

1.3 Purpose of the Study

The purpose of the study was to analyze the Application of multiple linear regression analysis on academic performance of students in Universities.

1.4 Objectives of the study

1. To review how Multiple linear regression analysis can be applied to assess the students' performance in Universities
2. **To** find out a model to estimate a student's grade in a particular course based on that student's background

1.5 Research Questions

- How is Multiple linear regression analysis applicable

17. What is a model to estimate a student's grade in a particular course based on that student's background?

1.6 Scope of the Study

The study was carried out on Application of Multiple linear regression analysis on academic performance in University. It was carried out in Universities for the period of one year that is

2021 6 2022

1.7 Significance of the study

The study will help other researchers who are interested in carrying out related to regression analvsis

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter involves the related literature review by different sources such as library, internet **and other** related books

2.1 Solving the Least Squares Problem

The least squares problem can be solved exactly or as close as possible depending on the data itself. Most of this chapter is based on the text "Applied Linear Algebra" by Noble and Daniel **(2001)** although other references are made.

The idea starts with a simple process of a relationship between variables and the error involved. **This** is easily seen in 2-diniension. If you start with a set of data points, any line can be drawn to **estimate** a relationship. In a linear regression one wants the line to be the best fit for the data. **Of course** any line could be drawn, but in order to get the best fit we must reduce the sum of the distances from the points to the line. We can do this by starting with a simple model,

.:.=x-e. \\here bis represented by a fixed, but unknown, part x, and also by a random fluctuation **atout the** fixed part, e. If there are multiple observations, we can index them. Thus, for i = 1, 2, - **m. the** band e will change as x remains fixed for all observations.

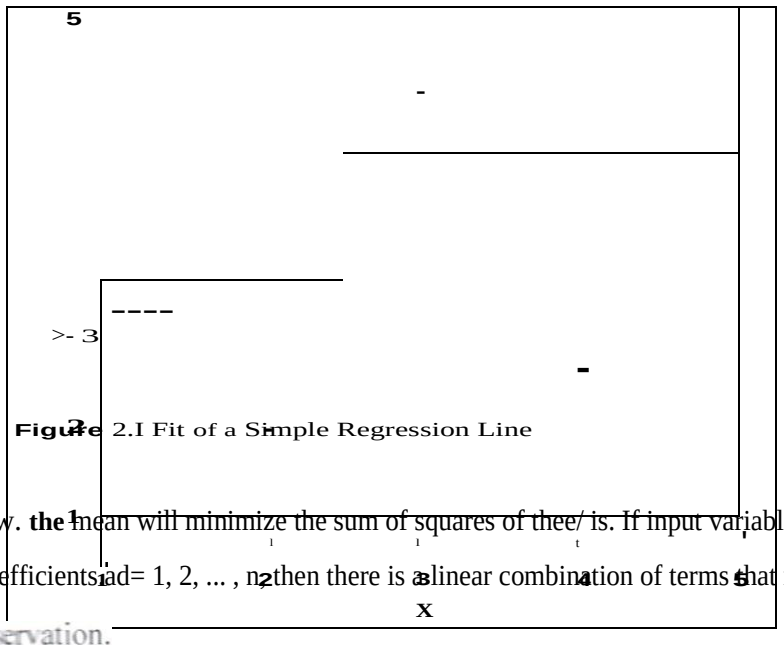


Figure 2.I Fit of a Simple Regression Line

Now. **the** mean will minimize the sum of squares of thee/ is. If input variables are represented ~coefficients;ad= 1, 2, ... , n, then there is a linear combination of terms that represent each bservation.

$$b = a_1X_1 + a_2X_2 + \cdots + a_jX_j + \varepsilon, j = 1,2,\ldots,n.$$

18.

There are n input variables and a fluctuation term for each observation. It can be rewritten as, $b = \text{"f.ai}\}Xj +$

$E_1, i = 1, 2, ..,m.$

Then changing to matrix notation

$$b = AX + \qquad (2.1)$$

border to minimize the sum of the squares of the ,,S,it is sufficient to minimize the L'-norm. **This norm** is given by,

$$\left(\text{"\'" } 2\right)^{1/2}$$

$$\text{III } 2i$$

We **can** solve for£, giving,

$$= b\text{-}Ax$$

So --tlai the vector,

$$x = min,\|b - AX\| = min,f\,a) \qquad (2.2)$$

- ~ mg the definition of the L'-norm gives,

$$f(x) = \sum\left(b_i - \sum a_{ij} X_j\right)^2.$$

~ **.::i .. ical** point is defined as a point where the derivative with respect to every variable vanishes. **So. there** is a system of partial derivatives that satisfy (for a critical point *x)* the following;

$$or/on«co =,2,- ya,M)\text{-}st9 = o,\kappa = 12....7$$

Scan be rearranged, as follows,

$$\sum a_{ik} b_i = \sum\sum a_{ik}\, a_{ij} X_j, \quad k = 1,2,\ldots..,n$$

19.

Definition 2.1. Let A be a p x g matrix.

The transpose A' of A is the q x p matrix obtained by interchanging the rows and columns of A, such that <A'>ij= <A>ji for l i<j and ij p.

Again changing to matrix notation gives:

$$A^T b = A^T Ax$$

..,.. erse of a Matrix.

-~.-\be a given matrix. Let I be the Identity Matrix .

.!. •.\m matrix L for which LA= 1 is called a left-inverse of A.

-) An~ matrix R for which AR= I is called the right-inverse of A

An matrix X for which XA = l and AX= I is called an inverse of A.

So the inverse of A is then denoted as A'

Mote that this is true for all matrices, but can be rearranged and solved such that

$$x = (A^T A)^{-1} A^T b$$

_ with the assumption that the inverse exists.

22Singular and non-singular matrices

A -~!1-singular matrix is a (necessarily square) matrix A that possesses an inverse X: For non-

-=- ..r .-\. there is an X with AX= XA = I. This inverse is denoted again by A

singular matrix is a square matrix that does not possess an inverse. ere are two

possibilities. Either,

A A is non-singular and thus (AT Ari exists, or

- !\TA is sim!Ular and therefore the inverse will not exist.

he special case that AT A is non-singular the solution can be found simply.

en a m::i.trix A, made up of input variables, first calculate A'. Next, find AT A and A Tb.

Find th

- ~ imerse of AT A and multiply it by AT b to get,

$$x = (A^T A)^{-1} A^T b$$

20.

-~mis.xis the optimum solution to the L 2 -norm Least Squares problem. The components **fx**are**the** coefficients in the regression equation, and thus, the best fit has been achieved.

Vassy **1965)**

---, ·,orks fine with direct calculations for the special case of a non-singular matrix.

eer. a problem arises if the matrix A'A is singular. Then the inverse (AT Ari will not exist. 1°":_~·s is the case, a different approach must be taken. Singular Value Decomposition (SVD) can be used to acquire the optimum solution for the general case.

The hermitian transpose A 8 of A is the q x p matrix found by taking the complex conjugates of e entries in A'.

Apxp matrix P for which p-1 = PH, so that ppH = pHp = T, is said to be unitary. An orthogonal **matrix** is a real unitary matrix P, so that p· 1 = pT and ppT= pTp = I.

-+ these definitions in mind, the theory behind Singular Value Decomposition can be given.

_ :1-rem 1.1: Let A be p x q.

eexists a p x q unitary matrix U (orthogonal if A is real), a q x q unitary matrix V (**gonal** if A is real), and a p x q matrix **I** with <>= 0 for i# j and <I>g = Oi ~0 with 01 02 where sis the min{p,q}, and the notation <I> represents an element in, such that the singular ~ decomposition

$$A = Uy''$$

is valid.

muse Singular Value Decomposition on the Least Squares problem, another theorem is **ed, AII** non-singu Jar matrices have an inverse. But, it is true that all matrices have a pseudo

erse as defined below

21.

Suppose that A= UEVis the singular value decomposition of the p xq matrix A of **rank** k, and that A' = V'U" is the pseudoinverse of A, where ' is the q x p matrix whose only nonzero entries are C'= 1/c,for 1 <i<k, then:

a) *x0*=A 11> minimizes I\Ax-b I\2 with respect to*x.*

b) Among al I minimizers X of || Ax-b || 2, xo = A-+b has the least 2-norm.

c' X minimizes 11 Ax-b I\z if and only if X= Xo + y, where vis an arbitrary linear combination of the final q-k columns of V and X = A'b. [6]

So. if the problem Ax b is considered, there is a solution to the least squares problem. The solution is found in the form $x = A'b$, and since b is known, the objective is to

calculate A',the pseudoinverse of A. From the definition,

$$A^{+} = V\sum^{+}U^{H}$$

Ce **that H** denotes the Hermitian matrix, which may include complex numbers. If all of the .----'" **oi-L'** are real then the notation can be changed to Ur, being simply the transpose of U. he the first step is to calculate V. Given A, it is easy to find AH and therefore AH A.

matrix is square, but not necessarily invertible. However, the eigenvalues, 1.1,"-2.... ***-e*** **found** from the characteristic equation. Then OI=vA., so all values of a can be found .

der **to** proceed in calculating V, the eigenvectors associated with each eigenvalue must be '" **These** are found by solving the system of equations for the AH A matrix and substituting

for each eigenvalue.

2~.

CHAPTER THREE

METHODOLOGY

3.1 forroduction

~ - .. **ch:1pter** presents the model development and the assumptions of Linear regression model in **it can be** applied to determine the performance of students Universities especially Calculas **algebra.** This study is based on data from the University of Tennessee but this methodology also =zed to Universities since all universities and teach almost the same course units as it was discussed **ncdusion** and recommendations. This methodology was based on the study (Masters Thesis) by (~ ,,,, K.More,2004) on A Multiple Linear Regression Analysis on Mathematics Placement at The - versity of Tennessee, Knoxville

31 Model development

of this chapter is based on material from the text "Applied Linear Regression Models" by eer. Kutner. Nachtsheim and Wasserman (1996). It is meant to explain some of the decisions – .. ,e to be made in doing an analysis of this kind.

rcer to run a statistical analysis on the data, there must first be a series of steps taken to eure that the data is both reasonable and valid. lt would make no sense to run a multiple **egression** on meaningless data or to have data that was statistically meaningful eliminated.

Se **of the** decisions can be made from statistical calculations.

raps. more mistakes in statistics occur from a lack of good judgment in one of two areas. :x. _ -:oosing the proper data to use, and second, deciding how to use that data. Bad sampling *ead* to errors in conclusions that no amount of analysis can fix. In fact, if not caught, it leads se conclusions and the work can prove to do more harm than good. Therefore, it is very imorant that the data be looked at carefully before being considered for analysis and equally u .. --; that the problem at hand be looked at thoroughly, so as to not leave out anything that _ ---~ ,he outcome of the work in a manner rendering it poor if not useless. (Hild, 1996)

26.

=~t ~ **is a** limited amount of data. Some would like to put every conceivable piece of

regression. but for two reasons this is not possible. One, some types of data are

-:L~ **unethical** to include; and two, only certain variables are available from the records

department.

this **leaves** is a list of variables that should be considered useful in evaluating a student's **neriomance based** on their past record. The high school grade point averages, Algebra **site. Algebra** score, Calculus composite, Calculus score, the class the student took and the **re they received** \\-ere all acquired from the records office. Also, the Mathematics Placement **scores were** matched with each student.

seems **simple** enough just to run a regression using these with the grade received as the target **e however,** one must look at each variable independently to assure that it makes sense in the analysis.

32**Assumptions**

_ - -;:_ it is warranted to approach each variable individually and make any assumptions **messary**. This does not mean that an assumption gives the correct or best assessment, but only es one "-hich tends to lead to a good amount of judgment.

kooks first at the Algebra vs. the Calculus composite scores, obviously, these are two e exams. They measure different subjects and in different ways. The Calculus · only has ~- ions. Math and English, while the Algebra has four sections including a science section. **-_ ... ;h** they are both multiple choice exams, one is penalized on the Calculus if they md are wrong, where this is not the case for the Algebra. So, these are different exams, **both** are accepted as entrance requirements into Universities especially in mathematics **ent**

.is an equivalency table that some universities use to convert one score to the other. For the - sof these exams, again they differ, but, can be converted. So, it is easy to see that an eminence **table** can be used to simplify the numbers. One might prefer not to do this, however, !'CC-"l.: ;:-:-:ictice and simplifies the number of variables in the regression. This is important

due to the fact that the more the data is broken down, the smaller the samples become for each class.

Small samples, especially those occurring in classes that have small enrolment to begin with, can lead to false conclusions. The Algebra was the exam taken by a majority of the students. So, Calculus scores were converted. This is the same table used by many universities to accept or deny student admissions.

The Mathematics Placement Exam should be a good indicator of a student's math ability. However, at the time these students entered college there were actually two exams. One was for Calculus and one was for Algebra. So, the regression had to be run in such a way that the exam was chosen as one or the other for some of the classes. Also, there were some problems with the exam itself. The Algebra exam.

The grade point averages of the students are a good indication of how well they did in high school. But, this represents an overall performance. There is no data on how many math classes the students took or the grades they received in them. Also, the quality of education cannot be looked at. Some high schools have higher standards and better teachers. So, although this might be an indication of if the student performed well overall,

it is not necessarily a good indicator of the student's mathematical abilities.

Finally, looking at the target variable, the grade made in the class, there has to be separation of classes. A grade of "B" in Calculus is not the same as a grade of "B" in Algebra. So, each class must be looked at individually and only those people taking that particular class were taken into account for that particular regression. This breaks down the data again, giving sometimes small data sets which may not be significant for some classes. However, for the larger classes, there is plenty of data and should yield good results.

As for the grade itself, at the university different teachers have different ways of assigning grades. Overall the system is uniform, but some instructors give a "B+" at 85% where others may not give one unless a student is very close to an "A" Having only the grades to go by, an assumption must be made as to what a grade represents. The grade assumption was made based on the average score for a student making a particular grade.

Keep in mind that these are assumptions. An average "B" for example is found by taking score of 80 as a "B" and a score of 85 as a "B+" and then finding the midpoint between them to be 82.5. The score of "F" is much more complicated. Someone could assign a grade of "F," a score of 30, being the average of O and 60. However, this would pull the regression way down. Obviously, there are some students that get 0% by not attending class at all. These would have to be considered outliers and thrown out of the regression. But, if all of the "F's" are thrown out the regression then it would only represent those that actually pass and would change the estimated grade dramatically.

Hence, a value for "F" had to be chosen; 50% seemed to be about an average "F," as much lower scores would tend to be outliers and a lot of students fail by only a few points. Again, this is a judgment call that can't really be termed as right or wrong, but it does at least take those failing into consideration without making the assumption that if a person fails they received a score of 30%.

From this chapter, I have learnt that how regression is calculates to determine students' academic performance using several statiscal software like JMP and ANOVA

CHAPTER FOUR

RESULTS

4.0 Introduction

In dealing with large amounts of data it becomes impossible to do all the calculations by hand. Luckily, we live in a day and age where computers can do the calculation for us very quickly. There is a wide range of possibilities that one may view in order to find the right package for the problem needed to be solved. Here, we will see that these are all very similar in approaching the least squares problem

Once a sofovare package is decided upon, whichever one it may be, one must understand the output. It is critical to assess the results to see if there is any real meaning to the regression. JMP, as do most packages, generate an ANOVA (analysis of variance) table along with other key output to allow the user to determine the validity of the analysis. This output is based on mathematical formulas that compare error Leitnaker M(1996).

The first step in the example is to let IMP test each variable to see which ones are significant. One can choose a forward or backward elimination procedure. JMP uses standard tests to check these. Some definitions are needed to understand the procedures

SSR= Sum of squares of the regression, which is associated with p-l degrees of freedom, representing the number of input variables Xl ... X.,. 1. SSR can be expressed in matrix notation

as SSR = 'A'b-(I/n)b'Ib, where J is an n x n matrix of l's. (4.1)

SSE=Sum of squares of the Error, which has n-p degrees of freedom since p parameters need to be estimated in the regression. SSE= b'b -TA'b. (4.2)

SST=The total sum of squares= SSR + SSE with n-1 degrees of freedom. (4.3) The mean squares are given by:

$$MSR = SSR/(p-I) \qquad (4.4)$$

$$MSE = SSE/(-p) \qquad (4.)$$

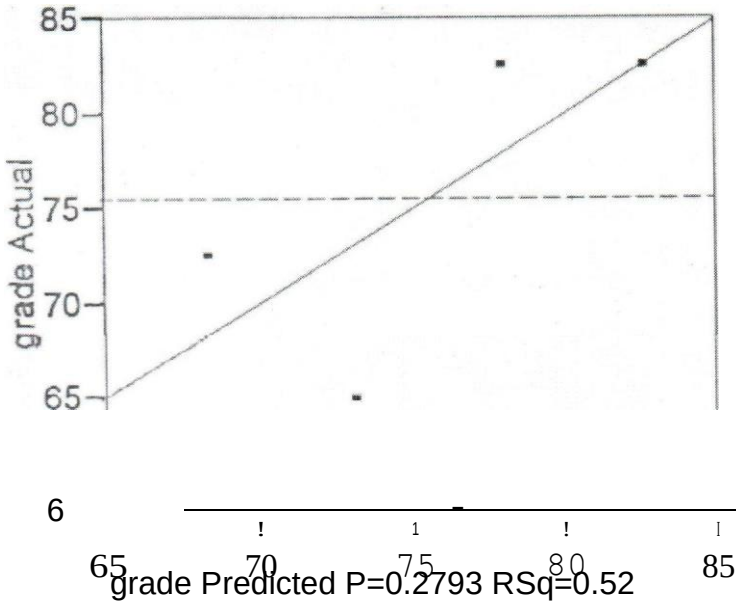
Finally, the F statistic is calculated as $F' = MSRMSE.$ (4.6)

Also, an overall view of the model is the R value. This is calculated from R 2 = SSR/SST, a ratio of the sum of squares of the regression to that of the total sum of squares. If this value is low (close to 0) then the model is not very good indicating a lot of error. If the value is high (close to l), then the model appears to be working well with little error attributed to the regression model.Neter Jet al ,(1996)

A simple look at the plot shows the grade predicted vs. the actual grade made and then a line is fit. Of course with only 4 data points it is easy to have a large amount of error. Looking at the A NOVA table, we get an F test of .2793. This tells us that there is definite correlation going on between the ACT score and the grade made in the class. One would prefer this to be lower, but it acceptable for the purpose of the example

The next things to look at are the students' T tests in the estimates table. The T test is done in the same manner and is directly related to the F test by F=(t)2 (Shavelson, 2013).

Response grade
\Vhole Model
Actual bv Predicted Plot



Summary of Fit

Rsquare Rsquare	0.519424
Adj	0.279137
Root Mean Square Error Mean	7.224092
of Response Observations (or	75.625
Sum Wgts)	4

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	112.81250	112.813	2.1617
Error	2	104.37500	52.187	Prob> F
C. Total	3	21718750		0.2793

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>l]
Intercept	41.583333	23.43349	1.77	0.2180
ACT	1.5833333	1.076904	1.47	0.2793

Figure **4.1 JM**P Data

Below, is the residual plot. Please see Figure 4.2. This shows how far away each actual grade is away from the predicted value. Note, that the grade off the most is about 8 points or less than one letter grade, while 2 points are about ½ of a letter grade off and the fourth is almost exact.

The Leverage plot (see Figure 4.3) shows the Algebra scores vs. grade residuals. This takes into account only the Algebra. In a multiple regression with many input variables there will be a leverage plot for each variable. The leverage plot shows just the correlation between that variable and the target variable residuals. This is an excellent way to quickly check the correlation of a single variable

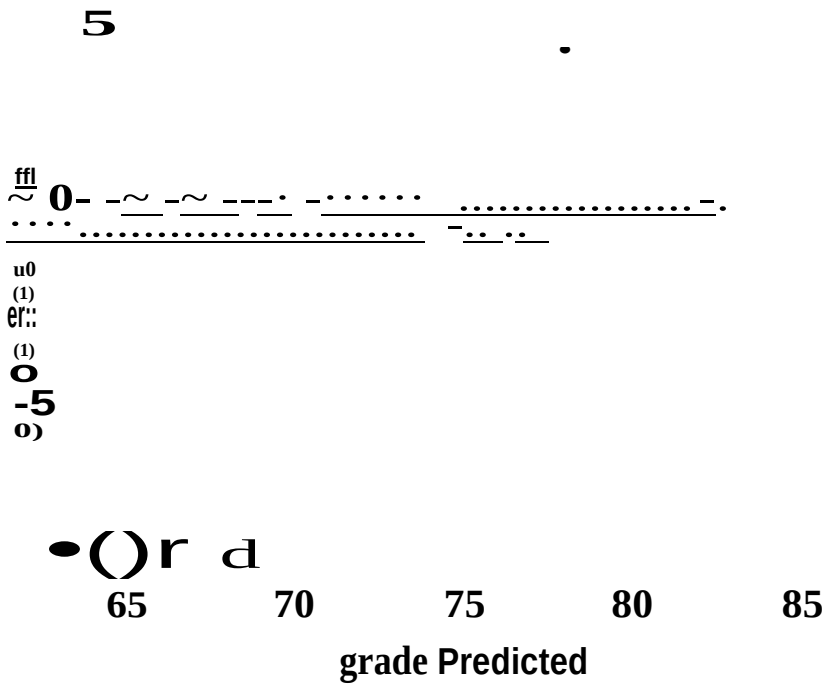


Figure 4.2 Residual Plot

31.

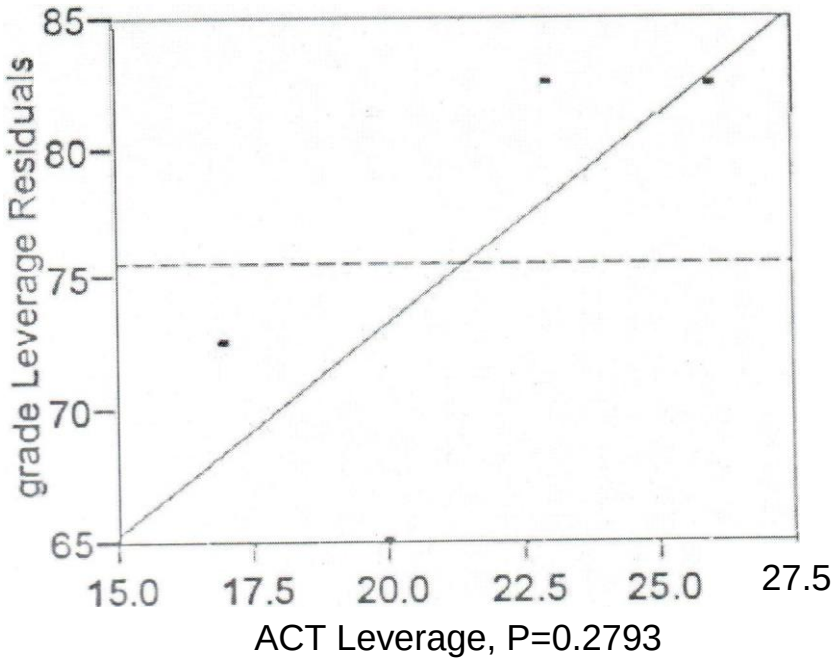


Figure **4.3** Leverage Plot

Finally, we look at the parameter estimates, which are calculated as shown in Chapter 2 by Singular Value Decomposition. They are:

Intercept	41.5833
ACT	1.5833

This leads us to our final regression equation to estimate the target variable (the score or grade). Score= 41.5833 + (l.5833)Algebra.

So, if a student made a 23 on the Algebra, the estimated grade for him to take this particular class would be as follows: 41.5833 + (L5833) 23 = 77.999. This would most likely be rounded off to an 78%. Then by using the regression equation we would estimate that the student would receive a score

of about 78% and therefore, it would be estimated that the student would probably get a "C+". Of course, if the student was placed into this class by the present system and was happy with a "C+" they may want to take that class. However, they might decide to take a lower class, perhaps 119, to improve their Algebra skills before moving ahead

CHAPTER FIVE

5.1 SUMMARY OF FINDINGS

After the data was split up, it was found that some classes did not contain enough information to run a regression with any confidence. Some of the data was deleted due to obvious input errors, such as scores that were out of the range of the test, etc. and then other data was left out due to missing data for that particular student. Once the data that was to be used was decided upon, it was split up into different courses. Then, for each course it was split again into 2 data sets per course. Those represent the students that took the Algebra exam and those that took the Calculus exam. Some courses only required one data set because the students that took the course almost all took the Calculus exam, leaving a handful that took the Algebra exam. For those courses, there is only one regression that was done. There was not enough data to do a regression, for example, for those who took Math 141 and took the Algebra placement exam.

The output and comments that follow will be helpful in understanding the validity of the analysis done. The confidence level for each regression was set at .8. This seemed like a reasonable value since representatives of the University mentioned this to me as a possible cut-off for success

The first pages listed are the actual JMP output pages so that all output can be viewed and analysed. Following the output, comments on the validity are made and a comparison to the multiple regression values vs. just analysis done on correlation between the math placement exam and the students score.

The classes that were analyzed were:

1) Math 115 Algebra placement;
2) t-.1ath 115 Calculus placement;
3) Math 119 Algebra placement;
4) Math 119 Calculus placement;
5) Math 1 23 Algebra placement;
6) Math 123 Calculus placement;
7) Math 125 Algebra placement;
8) Math 125 Calculus placement;
9) Math 130 Calculus placement; and
10) Math 141 Calculus placement.

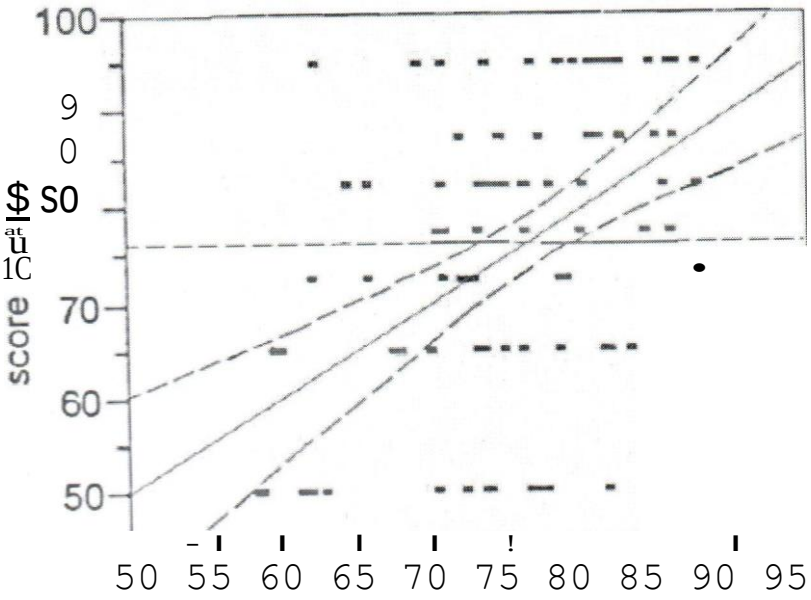
Notice there is not a regression for Algebra exam students that took 130 or 141 as there were not enough students taking that exam. Also, Math 110 was not analysed because it is a course that if analysis shows one to do poorly in the other classes, Math 110 is a course meant to prepare them for the others. Therefore, it being the least of the classes was not needed as a placement tool.

For the classes that were analysed, a list of the final regression equations is given at the end of the chapter. Please refer to Figures 5.1-5 .10.

For the classes that were analysed, a list of the final regression equations is given at the end of the chapter. Please refer to Figures 5.1-5.10.

32.

Response score
Whole Model
Actual by Predicted Plot



score Predicted P<.0001 RSq=0.21
RMSE=14.199

p

Summary of Fit

Rsquare	0.214752
R Square Adj	0.206821
Root Mean Square Error Mean	14.19925
of Re s.poose Observations (o	76.36139
Sum Wgts)	10
	1

Analysis of Variance

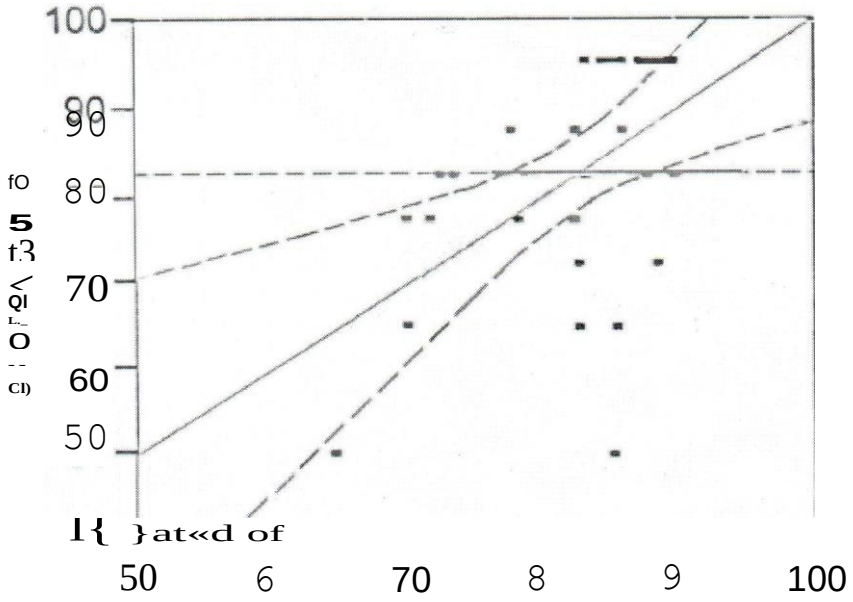
Source	OF	Sum of Squares	Mean Square	F Ratio
Modef	1	5458.803	5456.80	27.0749
Error	99	19960.257	201.62	Prob> f
C. Total	100	25419.059		<.0031

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> ft]
Intercept	23.5015€2	10.2565.8	2.29	0.0241
high sch::iol gpa	16.09538	3.093273	5.20	<.0001

33.

Response score
Whole Model
Actual b Predicted Plot



Summary of Fit

Rsquare	0.271192
Rsquare Adj	0.246001
Ro« Mean Square Error Mean or	11.51864
Response Observations **(or**	82.74194
Sum **Wgt.s}**	31

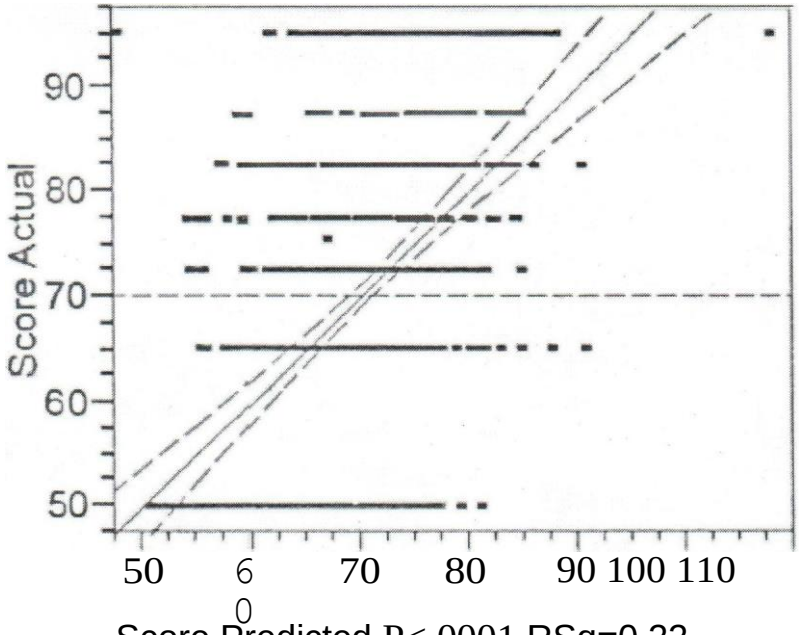
Analysis of Variance

Soorce	OF	Sum of Squares	Mean Square	F Ratio
Model	1	1431.7424	1431.74	10.7910
Error	29	3847.6931	132.68	Prob> F
C. Total	30	5279.4355		0.0027

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>/t]
Intercept	17.876914	19.85409	0.9	0.3753
high school gpa	17.977789	5.472743	0 3.28	0.0027

Response Score
Whole Model
Actual b Predicted Plot



Score Predicted P<.0001 RSq=0.22
RMSE.=13.363

Summary of Fit

Rsquare	0.224552
Rsquare Adj	0.221947
Root Mean Square Error Mean	13.36311
of Response Observations (o	70.27648
Sum Wgts)	8 7

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	46177.47	15392.5 176.6	6.1973
Error	693	159465.47		Prob> F
C. Total	696	205642.93		**<.0001**

Parameter Estimates

Term	Estimate	Std Eno	t Ratio	Prob> t]
intercept	18.256516	3.383759	5.40	**<.0001**
HS gpa	11.895833	0.943208	12.61	**<.0001**
Entrance math	0.1623701	0.075713	2.14	**0.0323**
Algebra exam	0.8190829	0.111703	7.33	**<.0001**

5.2 Conclusion

Unfortunately, I cannot say that this work will have a great impact. It should not change policy or be used for anything more than a tool to perhaps help a student decide in the case that they are in between choices. It would be nice to have better results, but in the real world, what does not work well, tells us what to try next. Therefore, I see value in what was done. In addition, a regression analysis could be better if done in a different matter. It could be that a linear regression is not the best way. Possibly, a logistic regression would be better(Shavelson,. (2013). "However, tried logistic regressions" and found no substantial difference. This is probably due to the high rate of variability in the data itself and this makes it inaccurate to be applied in Universities with ease in modulation and manipulation

However, the values found for a simple regression for just the math exam showed R squared values less than. 05 for some of the classes. That indicates that there is much information that can be used to classify students better.This is inadequate in Universities since students are graded according to subject combinations in Education faculty and their professional course units and this is also different since other courses can be assessed using one subject to obtain a specific performance in students courses at Universities especially Medical and Engineering

Courses.therefore that is why the researcher used data from Tennessee University since their data is readily published online to compaire and contract with University performance

5.3 Recommendations

My suggestion is to use the current system as a way to classify, however, in the case of person that is borderline, to use the regression equations to let them decide

I think the analysis is a good approach to starting a real system of class placement. I hope that it continues so that it can be refined. More data is a big key. Also, I think that unless something is done to change this system, so that not too many students are misplaced, we are looking at the same problems we already have. Looking at the fail out rate of the classes is enough to show me that we are failing as educators in the placement process. It is difficult to imagine a fail out rate of close to 50% for any class. I see that as a problem for the teachers, not just the students.

As a teacher, my primary concern is not grades or graduation, it is to teach and make sure a student learns. The whole point of an education is to learn.

The mathematics placement exam has been changed some since this analysis. I believe that if there will be any success in placing students properly, it has to be changed more to better evaluate mathematical ability.

The high rate of variability is a real issue. Only using the variables available leaves out a big part of usable data. Obviously, knowing the mathematics courses taken in high school and the grades that were made would be useful and other subjects one pursued. Another useful piece of information is the amount of time each student spent working in that class. This would bring in a factor not considered here.

Since all universities are supposed to produce almost same quality of graduates, the researcher took it that what happens in Tennessee University could also in Universities though per now its confidential therefore I would advise that data for various Universities on performance be published online so that to assist study purposes

References

Gilbert, M., (2002).An analysis on Scores vs. Grades

Leitnaker, M., Sanders, R., Hild, C., (1996). The Power of Statistical Thinking, Addison-Wesley, Mathworks, Learning Matlab 6.5, Mathworks (2002).

Noble, B., Daniels, J., Applied Linear Algebra, Prentice hall (1988) .. Sall, J., Lehman, A., Creighton, L., (200). JMP Start Statistics, Duxbury

Gilbert, A. G. (2002). *Teaching the Three R's Through Movement Experiences.* National Dance Education Organization, 4948 St. Elmo Avenue, Suite 301, Bethesda, MD 20814.

Moore, S. K. (2004). A Multiple Linear Regression Analysis on Mathematics Placement at The University of Tennessee, Knoxville.

Kinder, D. R., Sanders, L. M., & Sanders, L. M. (1996). *Divided by color: Racial politics and democratic ideals.* University of Chicago Press.

Antunes, T. P. C., de Oliveira, A. S. B., Crocetta, T. B., de Lima Antao, J. Y. F., de Almeida Barbosa, R. T., Guarnieri, R., ... & de Abreu, L. C. (2017). Computer classes and games in virtual reality environment to reduce loneliness among students of an elderly reference center:
Study protocol for a randomised cross-over design. *Medicine, 96(10).*

Bulmer, M. (1979). Concepts in the analysis of qualitative data. *The Sociological Review, 27(* 4), 651-677.

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics . .*Journal of statistical sofiware, 31,* 1-21.

Downs, T., Gates, K. E., & Masters, A. (200 I). Exact simplification of support vector solutions. *Journal of Machine Learning Research,* 2Dec), 293-297

Massy, W. F. (1965). Principal components regression in exploratory statistical research . .*Journal of the American Statistical Association, 60(309).* 234-256.

Shavelson, R. J. (2013). On an approach to testing and modeling competence. *Educational Psychologist, 48(2),* 73-86.

34. – - ---- –––––––

Shavelson, R. J. (2013). **On** an approach to testing and modeling competence. *Educational Psychologist, 48(2),* 73-86.

e