A Mathematical Meta Model of Academic Attainment Based on Class Attendance and Socio-Economic Changes

by

Nazmin Naher
Roll No. 1451559

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Mathematics

Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
September 2016

Dedication

To

My parents

Md. Nurul Islam & Shahanaz Begum

Declaration

This is to certify that the thesis work entitled "**A mathematical meta model of academic attainment based on class attendance and socio-economic changes**" has been carried out by **Nazmin Naher** in the Department of **Mathematics**, Khulna University of Engineering & Technology, Khulna, Bangladesh. The above thesis work or any part work of this work has not been submitted anywhere for the award of any degree or diploma.

Signature of Supervisor

Signature of Student

Approval

This is to certify that the thesis work submitted by **Nazmin Naher** entitled "**A mathematical meta model of academic attainment based on class attendance and socio-economic changes**" has been approved by the board of examiners for the partial fulfillment of the requirements for the degree of **Master of Science** in the Department of **Mathematics**, Khulna University of Engineering & Technology, Khulna, Bangladesh in September 2016.

BOARD OF EXAMINERS

1. Chairman (**Supervisor**)

……………………………………

Prof. Dr. A. R. M. Jalal Uddin Jamali
Department of Mathematics
Khulna University of Engineering & Technology

2. Member

…………………………………..

Head
Prof. Dr. M. M. Touhid Hossain
Department of Mathematics
Khulna University of Engineering & Technology

3. Member

……………………………………

Prof. Dr. Md. Bazlar Rahman
Department of Mathematics
Khulna University of Engineering & Technology

4. Member

……………………………………

Prof. Dr. Mohammad Arif Hossain
Department of Mathematics
Khulna University of Engineering & Technology

5. Member (**External**)

……………………………………

Prof. Dr. Subrata Majumdar
UGC Professor
Department of Mathematics
University of Rajshahi.

Acknowledgements

First and foremost, praises and thanks to Allah, the Almighty for His blessing upon me for the successful completion of this paper. Except His special kindness, this thesis paper would not be completed.

I would like to express my sincere gratitude and intense appreciation to my thesis supervisor, Dr. A. R. M. Jalal Uddin Jamali, Professor, Department of Mathematics, Khulna University of Engineering & Technology (KUET), Khulna, for his continuous supervision, guidance, support, valuable suggestions, constant encouragement and regular advice to prepare this thesis work. His continued support led me to the right way. I shall remain grateful to him.

I express my gratitude to my estimable teacher, Professor, Dr. M. M. Touhid Hossain the honourable Head, Department of Mathematics, Khulna University of Engineering & Technology for providing me all kinds of departmental help. I also want to express my all thanks, gratefulness and appreciations to my all teacher Department of Mathematics, Khulna University of Engineering & Technology, Khulna for their necessary advice, cordial co-operation and authentic information during the period of study. Especially, I am really grateful to Professor Dr. Mohammad Arif Hossain for his kind and important suggestions to prepare my thesis.

I am thankful to all my family members especially to my parents Md. Nurul Islam and Shahanaz Begum and two younger brothers Sohel Chowdhury and Sohanur Rahman.

Finally, I would like to thanks all of my friends especially Rakib Hossain, Joyanta Kumar Kundu, Md. Mahmudul Hasan, Ram Proshad, Borna, Chumky, Chhanda, Deepa and Elder brother Md. Sah Alam for their cordial encouragement and help.

Nazmin Naher
September, 2016

Abstract

The performance of students depend on many factors such as, attendance, motivation, level of engagement etc. which may be considered as the students attributes towards learning and some other attributes of the teachers on the process. No research is available relating the academic attainment and class attendance of our universities, especially at KUET where a 60% mandatory attendance is imposed to appear in the final examination. Though there exist many factors on student academic attainment, namely Class Test Marks and Final Grade, here we are interested and selected one important factor namely class attendance. There are many departments in this university (KUET) and a lot of subjects are taught. But Mathematics is common to all. Hence Mathematics is chosen for this study. This study is done among the students of first year and second year in several engineering departments regarding mathematics courses during the period 2000 -2013. It is of no doubt that for the existence of many proxy variables (Teacher's attribute, student's attribute, subjects, socio-economic environment etc.), it is very difficult to assess the impact of class attendance on academic attainment. Moreover, due to impose of mandatory percentage on class attendance, it becomes much more difficult to find out the impact of attendance on Class Test Marks as well as on Final Grade.

In this study we have considered only existing old data (student attendance and academic performance), where the effect of proxy variables are ignored. Moreover, for better comparison as well as for finding some test statistics, all the data are normalized. We have rigorously studied about the correlation between class Attendance and Class Test Marks for each course. We have also investigated thoroughly in each course for the existence of correlation between class Attendance and Final Grade. In spite of 60% mandatory attendance, from the experimental results, it revealed that attendance has a great effect on academic attainments. But the effect is varying department to department as well as semester to semester. Though there exist correlations between Attendance and Class Test Marks but there exist relatively much strong correlation between Attendance and Final Grade in perceptive of all departments. It reveals that class attendance grow some intuitive knowledge to the students which affect on their final Grade. According to the existence of correlations, some Meta regression models are proposed regarding both Class Test Marks and Final grade depending on attendance. Though for the lacking of continuous data, we could not find out socio-economic effect on academic attainment but it is revealed from the experimental study that introducing any new system effect on Attendance as well as academic performance.

Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Educators, parents, and politicians are continuously searching for that magic solution which will reform our public education system and establish a flawless system of education for our youth providing them with a quality education. In European Union, it is commonly assumed that university students are benefitted from attending lectures. This assumption, however, needs to be tested, as developments in information technology are increasingly calling for a reassessment of the traditional approach towards university education. Outside this assumption based on physical attendance in lectures and classes, a number of alternative weightless educational models, based on distance learning, are being introduced [Stanca (2006), Rocca (2003)]. In the last decade, a number of studies have examined the relationship between students' attendance (or absenteeism) and academic performance, with a general finding that attendance does matter for academic achievement [Devadoss and Foltz (1996), Marburger (2001), Kirby and McElroy (2003), Vanblerkon (1992), Sander et al. (2000)].

Now a days, regular school attendance is an important factor in school success [Rothman (2001)]. Research has shown a direct correlation between good attendance and student achievement. Poor attendance has been linked to poor academic achievement [Ziegler (1972)]. “Students who are absent from school receive fewer hours of instruction; they often leave education early and are more likely to become long term unemployed, homeless, caught in the poverty trap, dependent on welfare and involved in the justice system” [U.S. Department of Education (1996)]. Jones (2006) has studied the impact of student attendance, socio-economic status and mobility on student achievement. Literature suggests that a relationship exists among attendance, socio-economic status, and mobility and student achievement.

Studies have shown that learning and academic performance should be considered from a more holistic approach and the four main factors which are considered critical to learning are demography, active learning, students' attendance, and involvement in extracurricular activities [Ali et al. (2009)]. The consistent result of the holistic approach is that it enhances learning and is likely to increase academic performance. It also features attendance as being a contributing factor for such enhanced learning. One could argue that attendance increased as a result of more interesting class sessions or that the holistic approach requires active participation from the students hence attendance is crucial to the success of the learning and teaching style. In a meta-analysis reviewing the relationship of class attendance in college with grades and student characteristics, it was shown that attendance has strong correlations with both class grades and Grade Point Average (GPA) [Crede et al. (2010)]. Even with such strong evidence regarding the two variables, that meta-analysis also showed that mandatory attendance polices appear to have a small positive impact on average grades. According to Patel's study (based on his nine year teaching experience using Kelly's Personal Construct Theory (PCT)) the sustained high levels of student attendance at lectures and seminars improved student significant progress and satisfied cohort [Cohall (2009)].

Some studies [Hancock (1994), Shimoff and Catania (2001)] have showed that there is a positive correlation between attendance and academic performance. In addition to some studies showing that attendance and academic performance are directly correlated, some studies show a relatively consistent relationship between attendance and grades, regardless of the course subject or level of student [Ali et al. (2009)]. However, in some instances, the degree of change may be negligible [Crede et al. (2010)]. These latter reports mention other confounding factors in the learning process, such as student motivation and levels of engagement, which may have a greater contribution to academic performance than attendance. Therefore, it is still a burning question whether attendance has significant impact on academic performance or not. More researches are going on regarding this aspect here and abroad. It is also worthwhile to mention here that there are many factors exist in this aspect; one of them is socio-economic changes.

After **Chapter I** in which the introduction of the research works is presented, the literature review is discussed in **Chapter II**. **Chapter III** presents the overview of correlation and regression analysis and Test of Hypothesis**.** In **Chapter IV**, extensive investigations are carried out to find out the correlation of class attendance on Class Test as well as on Final Grade. In **Chapter V** extensive experiments are performed to establish regression models of academic performance on class attendance. Finally concluding remarks and brief discussion about the research works are given in **Chapter VI.** The list of the references and appendix are presented at the end of the thesis as well.

CHAPTER II

LITERATURE REVIEW

Student attendance is an important issue in today's' higher education. Many universities have compulsory attendance policies, while others refrain from making it as such. Despite the different policies, there seems to be a consensus among the professors about the positive effect of attendance in academic performance. Not attending in classes is seen as one of the reasons for academic failure. The recent developments in information and technology require a re-evaluation of the traditional method of study and the belief that undergraduate students are benefitted from class attendance should be tested. Moreover the presence of the new study methods based on distance learning requires a further analysis and discussion on the physical course attendance. In the last decade, a number of studies have investigated the relation between class attendance and academic performance reaching to the conclusion that there exists a positive correlation between these two [Durden and Ellis (1995), Devadoss and Foltz (1996), Marburger (2001), Kirby and McElroy (2003)].

It is mentioned earlier that research has shown a direct correlation between good attendance and student achievement and poor attendance has been linked to poor academic achievement [Ziegler (1972)]. Moreover studies have shown that a more holistic approach should be considered for assessment of the learning and academic performance and the four main factors which are considered critical to learning are demography, active learning, students' attendance, and involvement in extracurricular activities [Ali et al. (2009)]. When top-performing medical students were questioned about the main factors for their success, some of the main factors highlighted were "attitude, beliefs and motivation" and "effort and perseverance" (The University of the West Indies, 2009). Attendance was not mentioned or attributed to their success (The University of the West Indies, 2009). Moreover Crede et al.

(2010) also showed that mandatory attendance polices appear to have a small positive impact on average grades.

A large number of researches have been performed to investigate the impact of class attendance on academic attainment. A brief literature reviews are presented here regarding this matter. Romer (1993) provided the analysis of the relationship between lecture attendance and examination performance. Using attendance records in six sessions of his large (n = 195) Intermediate Macroeconomics course, he found that attendance had a positive and significant impact on academic performance. On the basis of these findings, Romer recommended experimenting with mandatory attendance policies to enhance student performance. Following on Romer’s (1993) seminal paper, several studies have attempted to measure the impact of attendance on learning outcomes. Durden and Ellis (1995) used students’ self-reported number of absences to explore the relationship between absenteeism and academic achievement in several sections (n = 346) of an undergraduate course.

Mitchell (1993), in his dissertation compared between the achievement and attendance of fifth grade African American male and female students attending same-gender classes and coeducational classes in Polytechnic Institute and State University. Poor attendance has been linked to poor academic achievement. Applegate (2003), in his Ph.D. dissertation, tried to establish a relationship among attendance, socio-economic status and mobility and the achievement of the students.

Sexton (2003) considered a case study of the effect of year round education on attendance, academic performance, and behavior patterns in his Ph. D dissertation. For the case study he considered the statistical data of Blacksburg University, Virginia. On the other hand Gamble (2004) and Jones (2006), in their doctorial dissertations, studied about the relationship among student population stability, academic achievement and gain-score test results. Some studies [Zamudio (2004), Hancock (1994)] showed a relatively consistent relationship between attendance and grades, regardless of the course subject or level of student. They have shown that attendance has strong correlations with both class grades and Grade Point Average (GPA). However, in some instances, the degree of change may be negligible.

Ali et al. (2009), investigated to find the influencing factors on performance. An empirical investigation was undertaken, using the simple correlation analytical technique, specifically the Pearson product movement correlation coefficient. After the data collection and analysis, they found that the result of the survey indicated that the Attendance of students at Simad University was highly affected by the following factors: demographic, active learning, students' attendance and involvement in extracurricular activities. On the other hand this study also examined the relationship between attendance and academic performance. A strong positive relationship between student's attendance and academic performance in Simad University has been found in this study. Similarly, Suleiman et al. (2012), in their study, revealed a strong positive relationship between Class Attendance and Cumulative GPA for Academic success in Industrial Engineering Classes.

However, Crede et al. (2010) showed that the degree of academic performance with respect to class attendance was negligible. They mentioned other confounding factors in the learning process, such as student motivation and levels of engagement, which may have a greater contribution to academic performance than attendance. The study showed that learning and academic performance should be considered from a more holistic approach including "attitude, beliefs and motivation" and "effort and perseverance".

On the other hand Cohall (2009) showed that, for the presence of mandatory attendance polices, attendances have a small positive impact on average grades. Moreover, Damian et al. (2012) tried to determine the significance of attendance in the improvement of academic performance of a first year course in medical program (University West Indies). They imposed that students must have an attendance rate of 80% of all timetabled sessions to sit final course exams which improved students' performance. Results showed that there was significant increase in attendance during Semester 2 of the academic year 2009-2010. This significant improvement in attendance was not reciprocated with an improvement in academic performance in course assessments when the two semesters were compared. The findings suggest that other factors are more critical to academic success. Some of these factors may be well indicated in the holistic approach which is regarded as the best approach to the learning process.

Sirin (2005) reviewed the literature on socioeconomic status (SES) and academic achievement in journal articles published between 1990 and 2000. The sample included 101,157 students, 6,871 schools, and 128 school districts gathered from 74 independent samples. The results showed a medium to strong SES–achievement relation. The author conducted a replica of White's (1982) meta-analysis to see whether the SES–achievement correlation had changed since White's initial review was published. Socioeconomic status (SES) is probably the most widely used contextual variable in education research. Increasingly, researchers examine educational processes, including academic achievement, in relation to socioeconomic background [Bornstein and Bradley (2003), Brooks-Gunn and Duncan (1997), Coleman (1988), McLoyd (1998)]. White (1982) carried out the first meta-analytic study that reviewed the literature on this subject by focusing on studies published before 1980 examining the relation between SES and academic achievement and showed that the relation varies significantly with a number of factors such as the types of SES and academic achievement measures. Since the publication of White's meta-analysis, a large number of new empirical studies have explored the same relation. On the other hand Authors [Lamdin (1996), Sutton and Soderstrom (1999)] show that these new results are inconsistent. They range from a strong relation to no significant correlation at all [Ripple and Luthar (2000), Seyfried (1998)]. Apart from a few narrative reviews that are mostly exclusive to a particular field [Entwisle and Astone, (1994), Haveman and Wolfe (1994), McLoyd (1998), Wang et al. (1993)], there has been no systematic review of these empirical research findings. The present meta-analysis is an attempt to provide such a review by examining studies published between 1990 and 2000.

More recently, the most comprehensive study to date is reported in Stanca (2006). He uses a large panel data set collected from an Introductory Microeconomics course (n = 766) in a

Italian university. The data combine administrative and survey sources. However, a limit of the data is that attendance to classes and tutorials is self reported by students. Applying three different econometric approaches (OLS-proxy regression, instrumental variables and panel estimators) to address the endogeneity of attendance rate variable, he bases his conclusions on panel data estimates indicating that attendance has an important independent effect on learning. Although most studies find positive effects of attendance on performance, the extent to which we can rely on the evidence presented in the cited studies is not always clear. Most of the studies leave unresolved the two main problems usually affecting the attendance rate variable.

Credé et al. (2010) considered a meta-analysis of the relationship between class attendances in college. They observed that the attendance has strong relationships with both class grades (k = 69, N = 21,195, r = .44) and GPA (k = 33, N = 9,243, r = .41). They also observed that mandatory attendance policies appear to have a small positive impact on average grades (k = 3, N = 1,421, d = .21).

Many college instructors exhort their students to attend class as frequently as possible, arguing that high levels of class attendance are likely to increase learning and improve student grades. Such arguments may hold intuitive appeal and are supported by findings linking class attendance to both learning [Jenne (1973)] and better grades [Moore et al. (2003)] but both students and some educational researchers appear to be somewhat skeptical of the importance of class attendance. This skepticism is reflected in high class absenteeism rates ranging from 18.5% [Marburger (2001)] and 25% [Friedman et al. (2001)] to 40% [Romer (1993)] and even as high as 59% and 70% (in two separate biology classes) [Moore et al. (2003)] and in explicit arguments against the importance of attendance in general and mandatory attendance policies in particular [Hyde and Flournoy (1986), St. Clair (1999)].

Indeed, a recent meta-analytic review of the training literature [Arthur et al. (2003)] showed lecture-based instruction to be effective for increasing cognitive, interpersonal, and even psychomotor skills and behaviors. Students who deny themselves the benefit of attending lectures (and the full range of activities involved in lecture attendance) and who rely only on other contact with class material are unlikely to retain relevant material as well as those attending class and subsequently perform less well on class tests and exams.

In the article [Guleker and Keci (2014)], a summary of the latest studies on attendance and academic performance will be given along with a deeper analysis of this relation in Albanian context. Data are collected from two courses in the civil engineering department of a private university taught by the same lecturer during 2009-2012. The results of the study are discussed in the light of the attendance policy enforced in today's Albanian higher education institutions.

CHAPTER III

OVERVIEW OF CORRELATION AND REGRESSION, AND TEST OF HYPOTHESIS

3.1 Introduction

There are many situations in which the objective in studying the joint behavior of two set of variables is to see whether they are related, rather than to use one to predict the value of the other. The most commonly used techniques for investigating the relationship between two quantitative variables are correlation and linear regression. Correlation quantifies the strength of the linear relationship between a pair of variables, whereas regression expresses the relationship in the form of an equation. For example, in patients attending an accident and emergency unit (A&E), we could use correlation and regression to determine whether there is a relationship between age and urea level, and whether the level of urea can be predicted for a given age.

The word correlation is used in everyday life to denote some form of association. We might say that we have noticed a correlation between foggy days and attacks of wheeziness. However, in statistical terms we use correlation to denote association between two quantitative variables. We also assume that the association is linear, that one variable increases or decreases by a fixed amount for a unit increase or decrease in the other. The other technique that is often used in these circumstances is regression, which involves estimating the best straight line to summarize the association.

3.2 Correlation

Correlation means association - more precisely it is a measure of the extent to which two variables are related. If an increase in one variable tends to be associated with an increase in the other then this is known as a positive correlation. If an increase in one variable tends to be associated with a decrease in the other then this is known as a negative correlation. When there is no relationship between two variables this is known as a zero correlation. Correlation is a widely used statistical technique. Correlation coefficients are the index of the measurement of the relationship among the sets of variables.

A correlation can be expressed visually. This is done by plotting a scatter diagram - that is one can plot the figures for one variable against the figures for the other on a graph. On a scatter diagram for a linear correlation, the closer the points lay to a straight line, the stronger the linear relationship between two variables (see Figure 3.1, 3.2). To quantify the strength of the relationship, we can calculate the correlation coefficient.

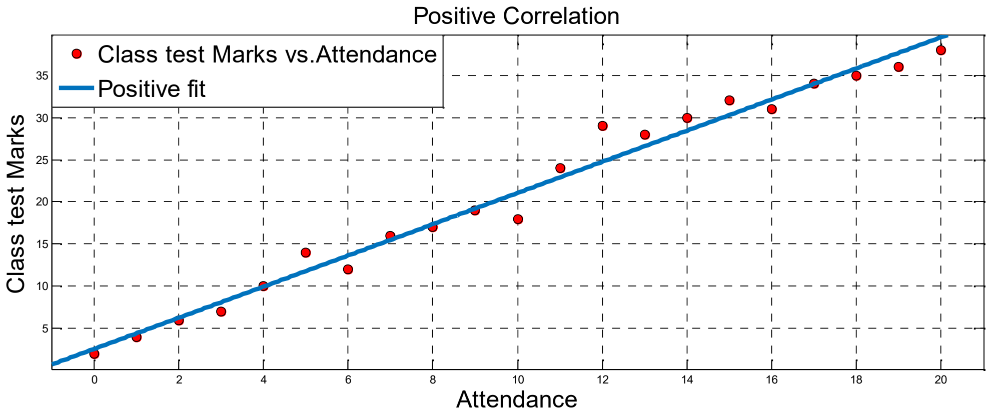

Figure 3.1 Linear fitting (positive correlation).

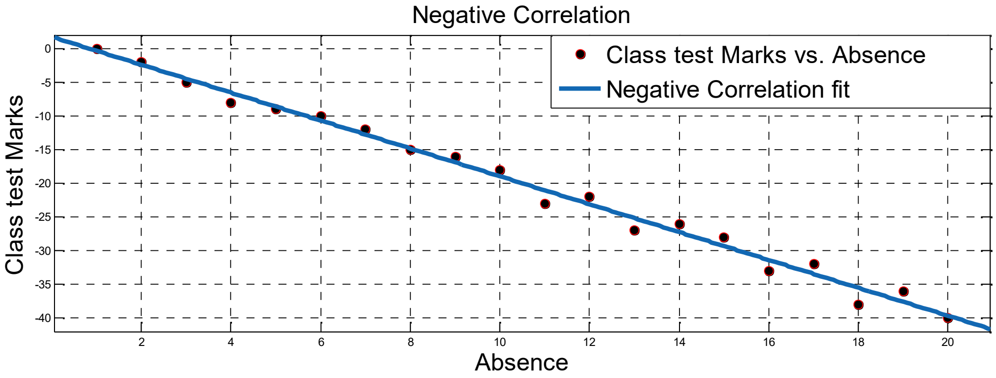

Figure 3.2 Linear fitting (negative correlation).

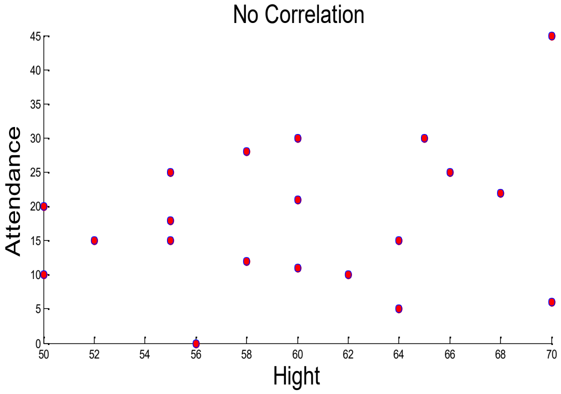

Figure 3.3 Scatter diagram of non-correlated data.

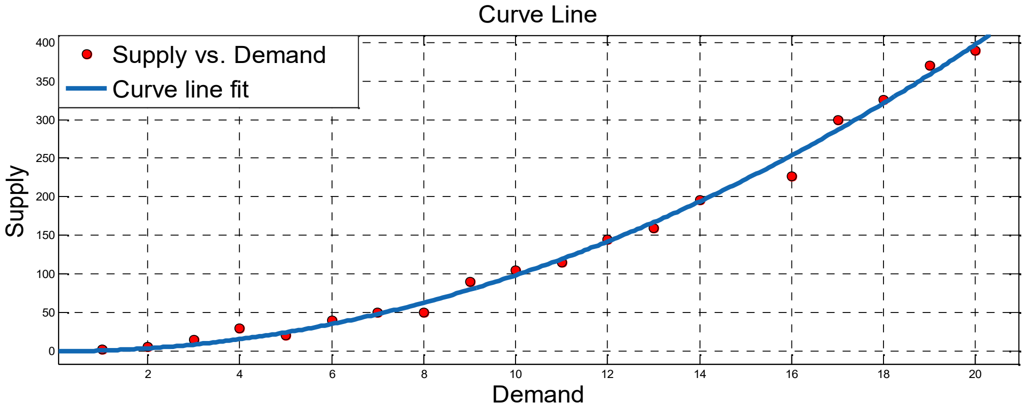

Figure 3.4 Non-linear fitting of correlated data.

Non Linear Correlation

When the amount of change in one variable is not in a constant ratio to the change in the other variable, we say that the correlation is non linear. Non linear correlation is also known as curvilinear correlation (Figure 3.4).

3.2.1 Classification of Methods

The relationship between more than one variable is considered as correlation. Correlation is considered as a number which can be used to describe the relationship between two variables. The number which quantifies the strength of the relationship is called the coefficient of correlation. There are several methods available to calculate the correlation coefficient.

Among them we will display two formulas namely (i) product-moment (Pearson product-moment) method and (ii) Rank (Spearman) method based on simple correlation between two variables.

In algebraic notation, if we have two variables *x* and *y*, and the data take the form of *n* pairs (i.e. [x_1, y_1], [x_2, y_2], [x_3, y_3], ..., [x_n, y_n]), then

(i) **Pearson product moment (Blanched formula) method:** The most important algebraic method of measuring correlation is Karl Pearson's Coefficient of correlation or Pearsonian's coefficient of Correlation. It has widely used application in Statistics. It is denoted by *r*. The mathematical formula of linear correlation coefficient of Pearson product moment (Blanched formula) method:

$$r = \frac{n\sum x \;\; -(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}} \tag{3.1}$$

which can be rewritten as

$$r = \frac{S_x}{\sqrt{(S_x \;)(S_y \;)}} \tag{3.2}$$

where *x* and *y* are *n* paired observations,

$$S_x \;\; = \sum(x_i - \bar{x})(y_i - \bar{y})$$

$$S_x \;\; = \sum(x_i - \bar{x})^2$$

$$S_y \;\; = \sum(y_i - \bar{y})^2$$

Here $\bar{x}$ is the mean of the *x* values, and $\bar{y}$ is the mean of the *y* values.

Requirements for Pearson's correlation coefficient:

- Scale of measurement should be interval or ratio.
- Variables should be approximately normally distributed.
- The association should be linear.
- There should be no outliers in the data.

(ii) **Rank (Spearman) method:** Spearman's rank correlation coefficient allows us to identify easily the strength of correlation within a data set of two variables, and whether the correlation is positive or negative. The Spearman coefficient is denoted with the Greek letter rho (ρ). Instead of using precise values of the variables, or when such precision is unavailable, the data may be ranked from 1 to *n* in order of size, importance, etc. If *x* and *y* are ranked in such a manner, the coefficient of rank correlation, or Spearman's formula for rank correlation (as it is often called), is given by the formula of linear correlation coefficient of Rank (Spearman) method:

$$R = 1 - \frac{6\sum D^2}{n(n^2 - 1)} \tag{3.3}$$

Where D denotes the differences between the ranks of corresponding values of *x* and *y*, and where *n* is the number of pairs of values (*x*, *y*) in the data.

Interpretation of Coefficient of correlation

Coefficient of correlation denoted by r is the degree of correlation between two variables. The value of *r* always lies between -1 and +1.

- When r is 1, we say there is a perfect positive correlation. A value of the correlation coefficient close to +1 indicates a strong positive linear relationship (i.e. one variable increases with the other; Figure 3.1).
- When r is a value between –1 and 0, we say that there is a negative correlation.
- When r is 0, we say there is no correlation. A correlation of zero means there is no relationship between the two variables. A value close to 0 indicates no linear relationship (Figure 3.3).
- When r is a value between 0 and 1, we say there is a positive correlation.
- When r is –1, we say there is perfect negative correlation. A value close to -1 indicates a strong negative linear relationship (i.e. one variable decreases as the other increases; Figure 3.2).

However, there could be a nonlinear relationship between the variables (Figure 3.4). As we noted, sample correlation coefficients range from -1 to +1. In practice, meaningful correlations (i.e., correlations that are clinically or practically important) can be as small as 0.4 (or -0.4) for positive (or negative) associations. There are also statistical tests to determine whether an observed correlation is statistically significant or not (i.e., statistically significantly different from zero). Procedures to test whether an observed sample correlation is suggestive of a statistically significant correlation can be found in Kleinbaum et al. (1988).

Properties of the Coefficient of correlation

© It has a well defined formula.

© It is a number and is independent of the unit of measurement.

© It lies between –1 and 1.

© Coefficient of correlation between *x* and *y* will be same as that between *y* and *x*.

3.2.2 Classification of Correlation

Correlation is described or classified in several different ways. Three of the most important are:

(i) Positive, negative and zero correlation

(ii) Simple, Partial and multiple correlation

(iii) Linear and non-linear correlation

We have already discussed about case (i). Now we will briefly discuss about the two cases (ii) and (iii). According to the number of variables there are three types of correlation coefficients. They are (i) Simple correlation (ii) Multiple correlation and (iii) Partial correlation. When only two variables are studied it is a problem of simple correlation. When three or more variables are studied it is a problem of either multiple or partial correlation. In multiple correlation three or more variables are studied simultaneously. For example, when we study the relationship between the yield of rice per acre and both the amount of rainfall and the amount of fertilizers used, it is a problem of multiple correlations. Similarly the relationship of plastic hardness, temperature and pressure is multivariate. In partial correlation we recognize more than two variables, but is considered that only two variables to be influencing each other, the effect of other influencing variable being kept constant. For example, in the rice problem, if we limit our correlation analysis of yield and rainfall with the assumption that the amount of fertilizer used remained same, it becomes a problem of partial correlation.

A brief discussion on them is given below :

(i) Simple correlation:

If there are only two variables, then the measure of the relation is called simple correlation. In order to compute simple correlation, we must have two variables, with values of one variable (*x*) paired in some logical way with values of the second variable (*y*). Such an organization of data is referred to as a bivariate (two-variable) distribution. Two variables may be positively correlated, be negatively correlated, or have no relationship to each other (zero correlation) [see Figure 3.1, 3.2 and 3.3].

In the case of a positive correlation between two variables, high measurements on one variable tend to be associated with high measurements on the other and low measurements on one with low measurements on the other. With negative correlation, high scores of one variable are associated with low scores of the other. The two variables thus tend to vary together but in opposite directions. A zero correlation means that there is no relationship between the two variables. High and low scores on the two variables are not associated in any predictable manner.

A simple correlation coefficient is a measure of the relationship between two variables. It describes the tendency of two variables to vary together (co-vary); that is , it describes the tendency of high or low values of one variable to be regularly associated with either high or low values of the other variable. The method of finding simple linear correlation between is discussed in section 3.2.1.

(ii) Multiple Correlations

The degree of relationship existing between three or more variables is called multiple correlation. When one variable is related to a number of other variables, the correlation is not simple. It is multiple if there is one variable on one side and a set of variables on the other side. To allow for generalizations to large numbers of variables, it is convenient to adopt a notation involving subscripts.

We shall let x_1, x_2, x_3, . . . x_k denote the variables under consideration. Then the partial correlation between the factor *i* and *j* are given by ρ_{ij}. Where each ρ_{ij} be the simple pair-wise correlation between factors x_i and x_j are computed as Eq. (3.1).

Then the measure the multicollinearity among the factors can be defined by the following measure of average pair-wise correlations

$$\circ^{2}\ \mathrm{X}\frac{\overset{k\quad i\mathrm{Z}1\quad \circ^{2}_{ij}}{i\mathrm{X}2\quad j\mathrm{X}1}}{k(k\ \mathrm{Z}1)/2} \qquad (3.4)$$

Note that this definition is frequently used in literature.

(iii) Partial Correlation

It is often important to measure the correlation between a dependent variable and one particular independent variable when all other variables involved are kept constant; that is, when the effects of all other variables are removed (often indicated by the phrase “other things being equal”). This can be obtained by defining a *coefficient of partial correlation*, except that we must consider the explained and unexplained variations that arise both with and without the particular independent variable.

If we denote by $\mathrm{r}_{12,3}$ the coefficient of partial correlation between x_1 and x_2 keeping x_3 constant, we find

$$r_{1\ ,3} = \frac{r_1 - r_1\ r_2}{\sqrt{(1-r_1^2\ )(1-r_2^2\ )}} \qquad (3.5)$$

Linear and Non-linear (curvilinear) Correlation. The distinction between linear and non-linear correlation is based upon the constancy of the ratio of change between the variables. If the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable then the correlation is said to be linear. It is clear that the ratio of change between the two variables is the same. If such variables are plotted on a graph paper all the plotted points would fall around a straight line (see Figure 3.1 and 3.2). Correlation would be called non-linear or curvilinear if the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable. For example, if we double the amount of rainfall, the production of rice or wheat, etc., would not necessarily be doubled. It may be pointed out that in most practical cases we find a non-linear relationship between the variables. However, since techniques of analysis for measuring non-linear correlation are far more complicated than those for linear correlation, we generally make an assumption that the relationship between the variables is of the linear type.

3.2.3 Limitations of Correlations

a) Correlation is not and cannot be taken to imply causation. Even if there is a very strong association between two variables we cannot assume that one causes the other. For example suppose we found a positive correlation between watching violence on T.V. and violent behavior in adolescence. It could be that the cause of both these is a third (extraneous) variable - say for example, growing up in a violent home - and that both the watching of T.V. and the violent behavior are the outcome of this.

b) Correlation does not allow us to go beyond the data that is given. For example suppose it was found that there was an association between time spent on homework (1/2 hour to 3 hours) and number of (G.C.S.E.) passes (1 to 6). It would not be legitimate to infer from this that spending 6 hours on homework would be likely to generate 12 (G.C.S.E.) passes.

3.2.4 Some uses of Correlations

Prediction

- If there is a relationship between two variables, we can make predictions about one from another.

Validity

- To test the validity of correlation between a new measure with an established measure.

Reliability

- Using different measures as well as samples for the consistence of the correlation.

Theory verification

- Performed further experiments on the population to verify the correlation.

3.3 Regression

People use regression on an intuitive level every day. In business, a well-dressed man is thought to be financially successful. A mother knows that more sugar in her children's diet results in higher energy levels. The ease of waking up in the morning often depends on how late you went to bed the night before. Quantitative regression adds precision by developing a mathematical formula that can be used for predictive purposes.

Regression is a statistical measure that attempts to determine the strength of the relationship between one dependent variable (usually denoted by *y*) and a series of other changing variables (known as independent variables). The two basic types of regression are simple regression and multiple regressions. Simple regression uses one independent variable to explain and/or predict the outcome of *y*, while multiple regression uses two or more independent variables to predict the outcome. On the other hand the regression may be Linear or non-linear. The general form of Simple Linear Regression is:

$$y = a + bx + \varepsilon \qquad (3.6)$$

Where:

y= the variable that we are trying to predict

x= the variable that we are using to predict y

a= the intercept

b= the slope denote regression coefficient

ε = the regression residual.

Similarly the Multiple Linear Regression can be express as

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + ... + b_tx_t + \varepsilon \qquad (3.7)$$

In multiple regression the separate variables are differentiated by using subscripted numbers.

Simple regression is used to examine the relationship between one dependent and one independent variable. After performing an analysis, the regression statistics can be used to predict the dependent variable when the independent variable is known. Regression goes beyond correlation by adding prediction capabilities. The regression line (known as the *least squares line*) is a plot of the expected value of the dependent variable for all values of the independent variable. Technically, it is the line that "minimizes the squared residuals". The regression line is the one that best fits the data on a scatter plot.

Using the regression equation, the dependent variable may be predicted from the independent variable. The slope of the regression line (b) is defined as the rise divided by the run. The y intercept (a) is the point on the y axis where the regression line would intercept the y axis. The slope and y intercept are incorporated into the regression equation. The intercept is usually called the constant, and the slope is referred to as the coefficient. Since the regression model is usually not a perfect predictor, there is also an error term in the equation. In the regression equation, y is always the dependent variable and x is always the independent variable. Here are three equivalent ways to mathematically describe a linear regression model.

- *y* = intercept + (slope x) + error
- *y* = constant + (coefficient x) + error
- *y* = *a* + *bx* + *ε* (3.8)

For $(x_i y_j)$: $i = 1,2,\cdots,n$ n pairs of observation, the parameter a and b can be estimated by using Least Square method. Mathematically

As we have seen, the least-squares regression line of *y* on *x* is

$$y = a + b \tag{3.9}$$

where *a* and *b* are obtained from the normal equations

$$\sum y = a \quad + b \sum x \tag{3.10}$$

$$\sum x \quad = a \sum x + b \sum x^2 \tag{3.11}$$

which yield

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum x \ )}{N \sum x^2 - (\sum x)^2} \tag{3.12}$$

$$b = \frac{n \sum x \ - (\sum x)(\sum y)}{N \sum x^2 - (\sum x)^2} \tag{3.13}$$

Note that if the slope is zero, it has no prediction ability because for every value of the independent variable, the prediction for the dependent variable would be the same. Knowing the value of the independent variable would not improve our ability to predict the dependent variable. Thus, if the slope is not significantly different than zero, don't use the model to make predictions.

Standard Error

The standard error of the estimate for regression measures the amount of variability in the points around the regression line. It is the standard deviation of the data points as they are distributed around the regression line. The standard error of the estimate can be used to develop confidence intervals around a prediction. A standard error is the standard deviation of the sampling distribution of a statistic. Standard error is a statistical term that measures the accuracy with which a sample represents a population. In statistics, a sample mean deviates from the actual mean of a population; this deviation is the standard error. Mathematically, if we let y_e represent the value of y for given values of *x* as estimated from equation (3.9), a measure of the scatter about the regression line of *y* on *x* is supplied by the quantity

$$s_{y.x} = \sqrt{\frac{\sum(y-y_e)^2}{n}} \qquad (3.14)$$

which is called the standard error of estimate of *y* on *x*.

3.4 Coefficient of determination

Before define coefficient of determination we would like to define explained and unexplained variation.

Explained and Unexplained Variation

The total variation of y is defined as $\sum(y-\bar{y})^2$: that is, the sum of the squares of the deviations of the values of y from the mean $\bar{y}$. As shown in this can be written

$$\sum(y-\bar{y})^2 = \sum(y-y_e)^2 + \sum(y_e-\bar{y})^2 \qquad (3.15)$$

The first term on the right of equation (3.15) is called the *unexplained variation*, while the second term is called the *explained variation*—so called because the deviations $y_e - \bar{y}$ have a definite pattern, while the deviations $y - y_e$ behave in a random or unpredictable manner. Similar results hold for the variable x.

The ratio of the explained variation to the total variation is called the *coefficient of determination.* If there is zero explained variation (i.e., the total variation is all unexplained), this ratio is 0. If there is zero unexplained variation (i.e., the total variation is all explained), the ratio is 1. In other cases the ratio lies between 0 and 1. Since the ratio is always nonnegative, we denote it by r^2. The quantity r, called the coefficient of correlation (or briefly coefficient correlation), is given by

$$\mathrm{R}^2 = \pm \frac{\mathrm{e} \qquad \mathrm{v}}{\mathrm{t} \qquad \mathrm{v}} = \pm \frac{\Sigma(y_e \ -\bar{y})^2}{\Sigma(y-\bar{y})^2} \tag{3.16}$$

and varies between -1 and +1. The + and – signs are used for positive linear correlation and negative linear correlation, respectively. Note that r is a dimensionless quantity; that is, it does not depend on the units employed.

For linear regression, the coefficient of determination (r-squared) is the square of the correlation coefficient. Its value may vary from zero to one. It has the advantage over the correlation coefficient in that it may be interpreted directly as the proportion of variance in the dependent variable that can be accounted for by the regression equation. For example, an r-squared value of 0.49 means that 49% of the variance in the dependent variable can be explained by the regression equation. The other 51% is unexplained. The **coefficient of determination** (R^2) is a measure of the proportion of variance of a predicted outcome.

In regression, the R^2 **coefficient of determination** is a statistical measure of how well the regression line approximates the real data points. The value of **coefficient of determination** (R^2) lies between 0 and 1. R^2 is a statistic that will give some information about the goodness of fit of a model. An R^2 of 1 indicates that the regression line perfectly fits the data. There are several definitions of R^2 that are only sometimes equivalent. One class of such cases includes that of simple linear regression where r^2 is used instead of R^2. When an intercept is included, then r^2 is simply the square of the sample correlation coefficient (i.e., r) between the outcomes and their predicted values. If additional regressors are included, R^2 is the square of the coefficient of multiple correlation. In both such cases, the coefficient of determination ranges from 0 to 1.

Important cases where the computational definition of R^2 can yield negative values, depending on the definition used, arise where the predictions that are being compared to the corresponding outcomes have not been derived from a model-fitting procedure using those data, and where linear regression is conducted without including an intercept. Additionally, negative values of R^2 may occur when fitting non-linear functions to data. In cases where negative values arise, the mean of the data provides a better fit to the outcomes than do the fitted function values, according to this particular criterion.

3.5 Assumptions and limitations of Correlation and Regression

The use of correlation and regression depends on some underlying assumptions. The observations are assumed to be independent. For correlation both variables should be random variables, but for regression only the response variable y must be random. In carrying out hypothesis tests or calculating confidence intervals for the regression parameters, the response variable should have a Normal distribution and the variability of y should be the same for each value of the predictor variable. The same assumptions are needed in testing the null hypothesis that the correlation is 0, but in order to interpret confidence intervals for the correlation coefficient both variables must be normally distributed. Both correlation and regression assume that the relationship between the two variables is linear.

A scatter diagram of the data provides an initial check of the assumptions for regression. The assumptions can be assessed in more detail by looking at plots of the residuals. Commonly, the residuals are plotted against the fitted values. If the relationship is linear and the variability constant, then the residuals should be evenly scattered around 0 along the range of fitted values.

3.6 Hypothesis test of correlation

A hypothesis is an assumption to be tested. The statistical testing of hypothesis is the most important technique in statistical inference. Hypothesis tests are widely used in business and industry for making decisions. It is noted that probability and sampling theory plays an ever increasing role in constructing the criteria on which business decisions are made. Very often in practice we are called upon to make decisions about population on the basis of sample information. For example, we may wish to decide on the basis of sample data whether a new medicine is really effective in curing a disease, whether one training procedure is better than another, etc. Such decisions are called statistical decisions.

In attempting to reach decisions, it is useful to make assumptions or guesses about the populations involved. Such assumptions, which may or may not be true, are called *statistical hypothesis* and in general are statements about the probability distributions of the population. The hypothesis is made about the value of some parameter but the only facts available to estimate the true parameter are those provided by a sample. If the sample statistic differs from the hypothesis made about the population parameter, a decision must be made as to whether or not this difference is significant. If it is, the hypothesis is rejected. If not, it must be accepted. Hence the term "tests of hypothesis".

Now if θ be the parameter of the population and $\hat{\theta}$ is the estimate of θ in the random sample drawn from the population, then the difference between θ and $\hat{\theta}$ should be small. In fact there will be some difference between θ and $\hat{\theta}$ because $\hat{\theta}$ is based on sample observations and is different for different samples. Such a difference is known as difference due to sampling fluctuations. If the difference between θ and $\hat{\theta}$ is large, then the probability that it is exclusively due to sampling fluctuations is small. Difference which is caused because of sampling fluctuations is called insignificant difference and the difference due to some other reasons is known as significant difference. A significant difference arises due to the fact that either the sampling procedure is not purely random or sample is not from the given population.

Procedure of Hypothesis Testing

The general procedure followed in testing hypothesis comprises the following steps:

(1) *Set up a hypothesis*. The first step in hypothesis testing is to establish the hypothesis to be tested. Since statistical hypotheses are usually assumptions about the value of some unknown parameter, the hypothesis specifies a numerical value or range of values for the

parameter. The conventional approach to hypothesis testing is not to construct single hypothesis about the population parameter, but rather to set up two different hypotheses. These hypotheses are normally referred to as (i) null hypothesis denoted by H_0, and (ii) alternative hypothesis denoted by H_1. The null hypothesis asserts that there is no true difference in the sample statistic and population parameter under consideration (hence the word "null" which means invalid, void or amounting to nothing) and that the difference found is accidental arising out of fluctuations of sampling.

A hypothesis which states that there is no difference between assumed and actual value of the parameter is the null hypothesis and the hypothesis that is different from the null hypothesis is the alternative hypothesis. If the sample information leads us to reject H_0, then we will accept the alternative hypothesis H_1. Thus, the two hypotheses are constructed so that if one is true, the other is false and *vice-versa*. The rejection of the null hypothesis indicates that the differences have statistical significance and the acceptance of the null hypothesis indicates that the differences are due to chance. As against the null hypothesis, the alternative hypothesis specifies those values that the researcher believes to hold true. The alternative hypothesis may embrace the whole range of values rather than single point.

(2) *Set up a suitable significance level*. Having set up a hypothesis, the next step is to select a suitable level of significance. The confidence with which an experimenter rejects or retains null hypothesis depends on the significance level a opted. The level of significance, usually denoted by "α", is generally specified before any samples are drawn, so that results obtained will not influence our choice. Though any level of significance can be adopted, in practice we either take 5 per cent or 1 per cent level of significance. When we take 5 per cent level of significance then there are about 5 chances out of 100 that we would reject the null hypothesis when it should be accepted, i.e., we are about 95% confident that we have made the right decision. When we test a hypothesis at a 1 per cent level of significance, there is only one chance out of 100 that we would reject the null hypothesis when it should be accepted, i.e., we are about 99% confident that we have made the right decision. When the null

hypothesis is rejected at $\alpha = 0.5$, the test result is said to be “significant”. When the null hypothesis is rejected at $\alpha = 0.01$, the test result is said to be “highly significant”.

(3) *Determination of a suitable test statistic*. The third step is to determine a suitable test statistic and its distribution. Many of the test statistics that we shall encounter will be of the following form :

$$\text{Test statistic} = \frac{\text{Sa} \quad \text{s} \quad -\text{H} \quad \text{p} \quad \text{pa}}{\text{S} \quad \text{e} \quad \text{o tl} \quad \text{sa} \quad \text{sl}}$$

(4) *Determine the critical region*. It is important to specify, before the sample it taken, which values of the test statistic will lead to a rejection of H_0 and which lead to acceptance of H_0. The former is called the critical region. The value of α, the level of significance, indicates the importance that one attaches to the consequences associated with incorrectly rejecting H_0. It can be shown that when the level of significance is α, the optimal critical region for a two-sided test consists of that $\alpha/2$ per cent of the area in the right-hand tail of the distribution plus that $\alpha/2$ per cent in the left hand tail. Thus establishing a critical region is similar to determining a 100 $(1 - \alpha)$% confidence interval. In general one uses a level of significance of $\alpha = 0.05$, indicating that one is willing to accept a 5 per cent chance of being wrong to reject H_0.

(5) *Doing computations*. The fifth step in testing hypothesis is the performance of various computations from a random sample of size *n*, necessary for the test statistic obtained in step (3). Then we need to see whether sample result falls in the critical region or in the acceptance regions.

(6) *Making decisions*. Finally, we may draw statistical conclusions and the management may take decisions. A statistical decision or conclusion comprises either accepting the null hypothesis or rejecting it. The decision will depend on whether the computed value of the test criterion falls in the region of rejection or the region of acceptance. If the hypothesis is being tested at 5 per cent level of significance and the observed set of results has a probability less than 5 per cent, we reject the null hypothesis and the difference between the sample statistic

and the hypothetical population parameter is considered to be significant. On the other hand, if the testing statistic falls in the region of non-rejection, the null hypothesis is accepted and the difference between the sample statistic and the hypothetical population parameter is not regarded as significant, i.e., it can be explained by chance variations.

Type I and Type II Errors

When a statistical hypothesis is tested, there are four possible result :

(1) The hypothesis is true but our test rejects it.

(2) The hypothesis is false but our test accepts it.

(3) The hypothesis is true and our test accepts it.

(4) The hypothesis is false and our test rejects it.

Obviously, the first two possibilities lead to errors. If we reject a hypothesis when it should be accepted (possibility No. 1) we say that a *Type I error* has been made. On the other hand, if we accept a hypothesis when it should be rejected (possibility No. 2) we say that a *Type II error* has been made. In either case a wrong decision or error in judgment has occurred.

TWO KINDS OF ERROR IN

HYPOTHESIS TESTING

Condition

Decision	H_0 :True	H_0 :False
Accept H_0	Correct Decision	Type II Error
Reject H_0	Type I Error	Correct Decision

There are two types of statistical hypotheses.

- **Null hypothesis**. The null hypothesis, denoted by H_0, is usually the hypothesis that sample observations result purely from chance.
- **Alternative hypothesis**. The alternative hypothesis, denoted by H_1 or H_a, is the hypothesis that sample observations are influenced by some non-random cause.

When the null hypothesis states that there is no difference between the two population means (i.e., d = 0), the null and alternative hypothesis are often stated in the following form.

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

CHAPTER IV

CORRELATION ANALYSIS AMONG ATTENDANCE AND ACADEMIC ATTAINMENTS

4.1 Introduction

As it has been seen from the earlier researches, the performance of students depend on many factors such as, attendance, motivation, level of engagement etc. which may be considered as the students attributes towards learning and some other attributes of the teachers on the process. No research is available relating the academic attainment and class attendance of our universities. It should be noted here that KUET has introduced a 60% mandatory attendance to appear in the final examination. It is of no doubt that in existence of proxy variables (Teacher’s attribute, student’s attribute, subjects, socio-economic environment etc.), the investigation of the impact of class attendance on academic attainment is very difficult to evaluate. For the limitations of resources and financial constraints, in this study the proxy variables will be ignored to calculate the impact of attendance on academic performance.

4.2 Experimental Study

Before investigate the existence of correlation between class Attendance and the student academic attainment, it is worthwhile to mention here that at least 60% attendance is imposed to a student of KUET to appear in the final examination. Therefore to find the rigorous effect of attendance on academic performance is difficult and hard working. Anyway, as KUET is a technical and engineering university, all of the students must take at least two basic mathematics courses. Therefore we considered mathematics courses for the study. To ignore teacher’s attributes, all the courses considered here was carried by the same teacher but in different departments as well as in different time epoch. Namely we considered random selection of mathematics courses as well as departments in the duration of 2000 – 2013 subject to the availability of data. For the investigations we considered following departments for the reason of availability of data.

Name of Departments	Courses
CE	Both 1st and 2nd semester/term of 1st year and both 1st and 2nd semester/term of 2nd year
EEE	Both 1st and 2nd semester/term of 1st years and both 1st and 2nd semester/term of 2nd years
ME	both 1st and 2nd semester/term of 2nd years
CSE	Both 1st and 2nd semester/term of 1st year and both 1st and 2nd semester/term of 2nd year
ECE	1st semester/term of 1st year and both 1st and 2nd semester/term of 2nd year
IEM	2nd semester/term of 1st year
LE	2nd semester/term of 1st year

In this study we performed extensive experiments to analyze the existing of correlation between student's class attendance and academic performance. According to the rules and regulation of the university, the student's final Grade is calculated by considering student attendance, performance in Class Tests and in final examinations. Therefore we will investigate for the existence of correlation between Attendance and Class Test marks and correlation between Attendance and final Grade separately. It is known that number of classes needed of each course depends on credit assigned to the course. So in this study, total numbers of classes of each course are normalized to 30 classes. Class Test marks are also normalized to 30 marks and final Grade points are calculated out of 4. Also note that to find the coefficient of correlation among attendance, Class Test marks and final Grade, we considered here the product-moment formula which is presented in Chapter III.

For the study of correlation between class attendance and student academic attainment, we will investigate to find (a) the existing of correlation between class attendance and Class Test marks and (b) the existing of correlation between class attendance and final Grade. We will use $R_{A,C}$ to denote simple linear sample correlation coefficient between Attendance (A) and Class Test marks (C) whereas $R_{A,G}$ means simple linear sample correlation coefficient (A)

Attendance and Final Grade (G). Moreover **t_{Cal}** will mean Student's t-statistics value corresponding to sample correlation value (coefficient of correlation value), on the other hand **t_{tab}** will denote standard t value for ν (=*N*-2) degree of freedom with α level of significance where *N* indicates number of pairs of data. If it is not mentioned otherwise, the value of α be 0.005 for one tailed test.

Again to test the significance of correlation between Attendance and Class Test marks to calculate $R_{A,C}$ value, we considered following test hypothesis with α = 0.005 (in one tailed) significance level.

$H_0 : \rho = 0$

$H_1 : \rho > 0$

Now we have to find out **t_{Cal}** value. We will find out each **t_{Cal}** value corresponding to each (*N*, $R_{A,C}$) values by using following formula

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \qquad (4.1)$$

Here *r* means coefficient of sample correlation which indicates $R_{A,C}$ and *N* means number of pair of data which indicates number of students. It is also to note that the standard t value denoted as $t_{(\nu,\alpha)}$ are available in any standard Statistics book as well as in web portal. It is known that the comments about the Null hypothesis are as follows:

S.L	Comparison	Decision about H_0
1.	if $t_{Cal} < t_{(\nu,\alpha)}$	accept
2.	if $t_{Cal} = t_{(\nu,\alpha)}$	No conclusion
3.	if $t_{Cal} > t_{(\nu,\alpha)}$	reject

4.2.1 Correlation Analysis between Class Test Marks and Class Attendance

We will find the coefficient of correlation between Class Attendance and Class Test Marks. To investigate the correlation between Class Attendance and Class Test Marks, at first, we have considered Civil Engineering (CE) Department. We have obtained the simple coefficient of correlation between Class Attendance and Class Test marks of each courses by using product- moment formula. The experimental results are displayed in the Table 4.1. The value of coefficient of correlation ($R_{A,C}$) of each course is calculated and presented in the column 5 of the Table 4.1. It is noted that in the Tables (4.1 – 4.7), 1st column indicates conducting year of the course. Code of each course is given in the second column. It is observed that, in most of the cases, the value of $R_{A,C}$ are near 0.5. In some few cases, the value of $R_{A,C}$ are less than 0.3 too.

Table 4.1 Correlation between Attendance and Class Test Marks and their test of significance in CE department

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,C}$	$\mathbf{t_{Cal}}$	$\mathbf{t_{(\nu,\alpha)}}$	$\mathbf{H_0}$
2001	1101	1st yr, 1st sem.	47	0.57589	4.72548	2.68	REJECT
2002	1101	1st yr, 1st sem.	52	0.64913	6.03419	2.68	REJECT
2001	1201	1st yr, 2nd sem.	45	0.76884	7.88446	2.68	REJECT
2002	1201	1st yr, 2nd sem.	49	0.43335	3.29652	2.68	REJECT
2004	1201	1st yr, 2nd sem.	104	0.36182	3.91978	2.62	REJECT
2011	1201	1st yr, 2nd sem.	117	0.16056	1.7444	2.62	ACCEPT*
2003	1201	1st yr, 2nd sem.	95	0.25377	2.5301	2.64	ACCEPT*
2009	2101	2nd yr, 1st term	115	0.51071	6.3145	2.62	REJECT
2012	2101	2nd yr, 1st term	116	0.23062	2.5305	2.62	ACCEPT*
2013	2101	2nd yr, 1st term	114	0.30802	3.4263	2.62	REJECT
2010	2201	2nd yr, 2nd term	110	0.32613	3.5852	2.62	REJECT

Now to test the significance of correlation between Attendance and Class Test marks regarding calculated sample $R_{A,C}$ value, we have to find *t* value for each $R_{A,C}$ value. The calculated *t* values denoted as t_{Cal} for each $R_{A,C}$ are presented in the column 6 of the Table 4.1. Also standard $t_{(\nu,\alpha)}$ values, with ν d.f. and 0.5% level of significance i.e. 99.5% confidence level, are given in column 7 of the table.

The comments about the test of hypothesis are mentioned in column 8 of the table. It is observed that, except three cases, all the Null hypotheses are rejected with 99.5% confidence level. That is the $R_{A,C}$ values are significant with 0.5% level of significance. We would also mention here that though, in three cases, Null hypotheses are accepted with 0.5% level of significance but with 5% level of significance the Null hypotheses of all of those are rejected. Therefore it may conclude that with 5% level of significance all calculated $R_{A,C}$ values are significant.

Now we have calculated the average value of coefficient of correlation between Class Attendance and Class Test Marks regarding CE department which is 0.416258 with average 88 students. Now for correlation coefficient 0.416258 with N=88, the corresponding t_{Cal} value is 4.245515 and standard $t_{(\nu,\alpha)}$ values, with 86 d.f. and 0.005 level of significance is about 2.63. As t_{Cal} (=4.245515)> $t_{(\nu,\alpha)}$(= 2.63) therefore the average value of coefficient of correlation between Class Attendance and Class Test marks regarding CE department is significance.

Table 4.2 Correlation between Attendance and Class Test Marks and their test of significance in EEE department

Year	**Course Code (Math)**	**Semester/ term**	**No. of Student (*N*)**	$R_{A,C}$	$\mathbf{t_{Cal}}$	$\mathbf{t_{(\nu,\alpha)}}$	$\mathbf{H_0}$
2001	1103	1st yr, 1st sem.	63	0.55183	5.16801	2.66	REJECT
2000	1203	1st yr, 2nd sem.	60	0.41576	3.48152	2.66	REJECT
2005	2103	2nd yr, 1st sem.	114	0.20707	2.23994	2.62	ACCEPT*
2012	2103	2nd yr, 1st sem.	118	0.30727	3.47760	2.62	REJECT
2001	2103	2nd yr, 1st sem.	61	0.68645	7.25092	2.66	REJECT
2000	2203	2nd yr, 2nd sem.	51	0.35494	2.65762	2.68	ACCEPT*

Now we consider Electrical and Electronic Engineering (EEE) Department to investigate the correlation between Class Attendance and Class Test Marks. The experimental results are displayed in the Table 4.2. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4.2. It is observed that in all the cases, except two, the value of $R_{A,C}$ are near 0.5. In only one case, the value of $R_{A,C}$ are less than 0.3.

Again to test the significance of correlation between Attendance and Class Test marks regarding calculated sample $R_{A,C}$ value, we have to calculate *t* for each $R_{A,C}$ value. The calculated t values for each $R_{A,C}$ are presented in the column 6 of the Table 4.2. Also standard $t_{(v,\alpha)}$ values, with v d.f. and 0.5% level of significance i.e. 99.5% confidence level, are given in column 7 of the table.

The comments about the test of hypothesis are mentioned in column 8 of the table. It is noticed that, except two cases, all the Null hypotheses are rejected with 99.5% confidence level. That is the corresponding $R_{A,C}$ values are significant with 0.5% level of significance. Moreover in that two cases, Null hypotheses are rejected with 5% level of significance rather than with 0.5% level of significance. Therefore it may conclude that with 5% level of significance all calculated $R_{A,C}$ values are significant.

Again we calculate the average value of coefficient of correlation between Class Attendance and Class Test marks regarding EEE department which is 0.420553 with average 78 students. Now for the correlation coefficient 0.420553 with N=78, the corresponding t_{Cal} value is 4.041033 and standard $t_{(v,\alpha)}$ values, with 76 d.f. and 0.005 level of significance is about 2.64. As t_{Cal} (=4.041033)> $t_{(v,\alpha)}$(= 2.64) therefore the average value of coefficient of correlation between Class Attendance and Class Test marks regarding EEE department is significance.

Table 4.3 Correlation between Attendance and Class Test Marks and their test of significance in ME department

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,C}$	$\mathbf{t_{Cal}}$	$\mathbf{t_{(v,\alpha)}}$	$\mathbf{H_0}$
2010	2105	2nd yr, 1st sem.	115	0.38536	4.43928	2.62	REJECT

Now we consider Mechanical Engineering (ME) Department to investigate the correlation between Class Attendance and Class Test Marks. The experimental results are displayed in the Table 4.3. It is noted that there is only one set of information is available. The value of coefficient of correlation of the course is calculated and presented in the column 5 of the Table 4.3. It is observed that the value of $R_{A,C}$ is near about 0.4.

To test the significance of correlation between Attendance and Class Test marks regarding calculated sample $R_{A,C}$ value, we have to calculate *t* for the $R_{A,C}$ value. The calculated *t* values for the $R_{A,C}$ is presented in the column 6 of the Table 4.3. Also standard $t_{(\nu,\alpha)}$ values, with ν d.f. and 0.5% level of significance i.e. 99.5% confidence level, are given in column 7 of the table.

The inference about the test of hypothesis is mentioned in column 8 of the table. It is noticed that the Null hypothesis is rejected with 99.5% confidence level. Therefore it may conclude that with 0.005 level of significance calculated $R_{A,C}$ value is significant.

Table 4.4 Correlation between Attendance and Class Test Marks and their test of significance in CSE department

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,C}$	$\mathbf{t_{Cal}}$	$\mathbf{t}_{(\nu,\alpha)}$	$\mathbf{H_0}$
2005	2107	2nd yr, 1st sem.	52	0.438312	3.448217	2.68	REJECT
2003	2107	2nd yr, 1st sem.	57	0.378429	3.031997	2.66	REJECT
2012	2207	2nd yr, 2nd sem.	58	0.256601	1.986745	2.66	ACCEPT*
2002	2207	2nd yr, 2nd sem.	45	0.256851	1.742752	2.68	ACCEPT*
2003	2207	2nd yr, 2nd sem.	58	0.616889	5.865408	2.66	REJECT
2010	2207	2nd yr, 2nd sem.	59	0.424192	3.536524	2.66	REJECT

Let us consider Computer Science and Engineering (CSE) Department to investigate the correlation between Class Attendance and Class Test Marks. The experimental results are shown in the Table 4.4. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4.4. It is observed that, in most of the cases, the value of $R_{A,C}$ are near 0.4. In some few cases, the value of $R_{A,C}$ are less than 0.3.

Now to test the significance of correlation between Attendance and Class Test marks regarding calculated sample $R_{A,C}$ value, we have to calculate *t* for each $R_{A,C}$ value. The calculated *t* values for each $R_{A,C}$ are presented in the column 6 of the Table 4.4. Also standard $t_{(\nu,\alpha)}$ values, with ν d.f. and 0.005 level of significance, are given in column 7 of the table.

The comments about the test of hypothesis are mentioned in column 8 of the table. It is noticed that, except two cases, all the Null hypotheses are rejected with 99.5% confidence level. Moreover for all the test statistics, the Null hypotheses are rejected 5% level of significance. Therefore it may conclude that with 0.05 level of significance all calculated $R_{A,C}$ values are significant.

Now we calculate the average value of coefficient of correlation between Class Attendance and Class Test marks regarding CSE department which is 0.395212 with average 55 students. Now for correlation coefficient 0.395212 with N=55, the corresponding t_{Cal} value is 3.132181 and standard $t_{(\nu,\alpha)}$ values, with 53 d.f. and 0.005 level of significance is about 2.67. As t_{Cal} (=3.132181)> $t_{(\nu,\alpha)}$(= 2.67) therefore the average value of coefficient of correlation between Class Attendance and Class Test marks regarding CSE department is significance.

Now to investigate the correlation between Attendance and Class Test marks we consider Electronics and Communication Engineering (ECE) Department. The experimental results are displayed in the Table 4.5. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4.5. It is observed that, in most of the cases, the value of $R_{A,C}$ are near 0.4. In two cases, the value of $R_{A,C}$ are near 0.2.

To test the significance of correlation between Attendance and Class Test marks regarding calculated sample $R_{A,C}$ value, we have calculated *t* values for each $R_{A,C}$ which are presented in the column 6 of the Table 4. 5. Also standard $t_{(\nu,\alpha)}$ values, with ν d.f. and 0.005 level of significance i.e. 99.5% confidence level, are given in column 7 of the table.

The inferences about the test of hypothesis are mentioned in column 8 of the table as well. It is noticed that, for two $\mathbf{t_{Cal}}$ values, the Null hypotheses are rejected with 99.5% confidence level. But, for all the $\mathbf{t_{Cal}}$ values, the Null hypotheses are rejected with 95% confidence level. Therefore it may conclude that with 0.05 level of significance all calculated $R_{A,C}$ values are significant.

Table 4.5 Correlation between Attendance and Class Test Marks and their test of significance in ECE department

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,C}$	$\mathbf{t_{Cal}}$	$\mathbf{t_{(\nu,\alpha)}}$	$\mathbf{H_0}$
2002	1109	1st yr, 1st sem.	31	0.53833	3.43997	2.75	REJECT
2003	2109	2nd yr, 1st sem.	31	0.14741	0.80258	2.75	ACCEPT*
2009	2109	2nd yr, 1st sem.	30	0.4202	2.45029	2.75	ACCEPT*
2012	2209	2nd yr, 2nd sem.	53	0.22696	1.66428	2.68	REJECT
2009	2209	2nd yr, 2nd sem.	29	0.3349	1.84683	2.76	ACCEPT*

Now we calculate the average value of coefficient of correlation between Class Attendance and Class Test Marks regarding ECE department which is 0.33356 with average 35 students. Now for correlation coefficient 0.33356 with N =33, the corresponding t_{Cal} value is 2.032564 and standard $t_{(\nu,\alpha)}$ values, with 31 d.f. and 0.005 level of significance is about 1.697. As t_{Cal} (=2.032564)> $t_{(\nu,\alpha)}$(= 1.697) therefore the average value of coefficient of correlation between Class attendance and Class Test marks regarding ECE department is significance.

Now we consider Industrial Engineering and Management (IEM) Department to investigate the correlation between Attendance and Class Test marks. The experimental results are shown in the Table 4.6. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4.6. It is observed that, there is only one information available and the value of $R_{A,C}$ is approximately 0.5.

Table 4.6 Correlation between Attendance and Class Test Marks and their test of significance in IEM department

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,C}$	$\mathbf{t_{Cal}}$	$\mathbf{t}_{(\nu,\alpha)}$	$\mathbf{H_0}$
2011	1211	1st yr, 2nd sem.	58	0.50295	4.35457	2.66	REJECT

It is also noticed that the calculated *t* values for the $R_{A,C}$ value is 4.35457 which is greater than $t_{(\nu,\alpha)}$ value (= 2.66) for 57 d.f with 99.5% confidence level.

The comments about the test of hypothesis are mentioned in column 8 of the table. It is noticed that except one case all the Null hypotheses are rejected with 99.5% confidence level. That is the corresponding $R_{A,C}$ values are significance with 0.005 level of significance. Therefore it may conclude that with 0.005 level of significance calculated $R_{A,C}$ value is significant.

Table 4.7 Correlation between Attendance and Class Test Marks and their test of significance in LE department

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,C}$	$\mathbf{t_{Cal}}$	$\mathbf{t}_{(\nu,\alpha)}$	$\mathbf{H_0}$
2013	2119	2nd yr, 1st sem.	43	0.60143	4.82021	3.551	REJECT

Now we consider Leather Engineering (LE) Department to investigate the correlation between Attendance and Class Test marks. The experimental results are displayed in the Table 4.7. It is observed that, there exist only one set of information and the value of $R_{A,C}$ is near about 0.6. It is also noticed that the calculated $\mathbf{t_{Cal}}$ values is significantly greater than tabulated $\mathbf{t}_{(\nu,\alpha)}$

with 41 d.f. and 0.005 level of significance. In consequence we reject the Null Hypothesis with 41 d.f. and 0.005 level of significance. Therefore there exist highly correlation between Attendance and Class Test marks regarding LE department.

4.2.2 Correlation Analysis between Class Attendance and Final Grade

Now we will perform extensive experiments to study the existence of correlation between class Attendance and Final Grade. We will find the simple coefficient of correlation between Class Attendance and Final Grade of each courses considered by using product- moment formula as well.

Table 4.8 Correlation between Attendance and Final Grade in CE department and their test of significance

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,G}$	$\mathbf{t_{Cal}}$	$\mathbf{t}_{(\nu,\alpha)}$	$\mathbf{H_0}$
2001	1101	1st yr, 1st sem.	47	0.651826	5.765775	2.68	REJECT
2002	1101	1st yr, 1st sem.	52	0.694241	6.820521	2.68	REJECT
2001	1201	1st yr, 2nd sem.	45	0.874814	11.84114	2.68	REJECT
2002	1201	1st yr, 2nd sem.	49	0.363451	2.674599	2.68	ACCEPT*
2004	1201	1st yr, 2nd sem.	104	0.346499	3.730578	2.62	REJECT
2011	1201	1st yr, 2nd sem.	117	0.438759	5.236079	2.62	REJECT
2003	1201	1st yr, 2nd sem.	95	0.351952	3.6261084	2.64	REJECT
2002	2101	2nd yr, 1st term	44	0.88629	12.40212	2.68	REJECT
2004	2101	2nd yr, 1st term	95	0.402555	4.240895	2.64	REJECT
2012	2101	2nd yr, 1st term	115	0.585991	7.721266	2.62	REJECT
2013	2101	2nd yr, 1st term	116	0.535068	6.702844	2.62	REJECT
2009	2101	2nd yr, 1st term	114	0.673547	9.686753	2.62	REJECT
2010	2201	2nd yr, 2nd term	110	0.632979	8.497001	2.62	REJECT

For this study we first consider Civil Engineering (CE) Department. The experimental results are displayed in the Table 4.8. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4. 8. Note that in the Tables (4. 8 – 4. 13), 1st column indicates conducting year of the courses and course code is given in the second column of the tables. It is observed that, in most of the cases, the value of $R_{A,G}$ are greater than or near equal to 0.6. In some few cases the value of $R_{A,G}$ are less than 0.5 but always greater than 0.3.

To test the significance of correlation between Attendance and Final Grade regarding $R_{A,G}$ value, we have to calculate *t* for each $R_{A,G}$ value. The calculated *t* values for each $R_{A,G}$ are reputed in the column 6 of the Table 4.8. Also standard $t_{(\nu,\alpha)}$ values, with ν d.f. and 0.005 level of significance i.e. 99.5% confidence level, are given in column 7 of the table.

The inferences about the test of hypothesis are mentioned in column 8 of the table as well. It is noticed that, except one case, all the Null hypotheses are rejected with 99.5% confidence level. That is the $R_{A,G}$ values are significance with 0.005 level of significance. Now we would again mention here that though, in one case, Null hypotheses is not rejected with 0.005 level of significance but with 0.05 level of significance, the Null hypotheses of all of those are rejected. Therefore it may conclude that with 5% level of significance almost all $R_{A,G}$ values are highly significant.

In our study we were supposed to study the effect of socio-economic changes. But due to lack of data from ordinary sources it cannot be done. So we may require questionnaire survey to get the data for further analysis. But due to some other constraints it has not been done. Fortunately within the period of study, the institute has undergone in different changes: BIT is converted to KUET at 2000 and the course credit system has been also introduced at 2009. With these changes, if the change in the class Attendance or in the Results is assessed, we may consider them as the impact due to Socio-economic changes. For the purpose, the time series analysis will be effective.

In order to study the effect of time series analysis regarding the correlation between class attendance and Final Grade, we plot $R_{A,G}$ with respect to years which is depicted on the Figure 4.1. In the figure 'rho' indicates $R_{A,G}$ values. It is observed from the figure that the data are clustered into two groups. One group ranges from 2000 – 2005 and another from 2009 – 2013. It is worthwhile to mention here that we have no data during 2005 to 2008 (as conducting teacher was not available at KUET in that time). Anyway, it is remarkable that the rho value is highest near 2000-2001. The trend of rho decreases gradually up to 2005. In next cluster, the value of rho is increased at 2009 and again gradually decreases. Perhaps the reason of this kind of nature of rho values is as follows: In 2000-2001, BIT is converted to KUET and course-credit system was introduced in this period. Again in 2009-2010, term system was introduced from semester based system. Therefore, in any initial stage of a new system, student performance is highly correlated with class attendance.

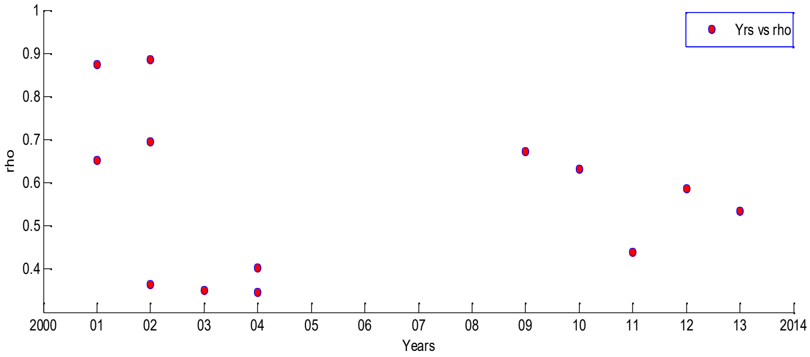

Figure 4.1 Scatter diagram of correlation coefficient with respect to time (CE Department)

Now we have calculated the average value of coefficient of correlation between Class attendance and Final Grade regarding CE department. The average correlation is 0.572152 and corresponding t_{Cal} value is 5.88 on average 73 students. We observed that the standard $t_{(\nu,\alpha)}$ values, with 71 d.f. and 0.005 level of significance is about 3.416. Since t_{Cal} (=3.857539) > $t_{(\nu,\alpha)}$(= 3.416) therefore the average value of coefficient of correlation between Class attendance and Final Grade regarding CE department is highly significance.

It is worthwhile to remark that the correlation between Attendance and Final Grade are more significance than correlation between Attendance and Class Test marks in the case of CE department. In other way we may say that class attendance is highly correlated with the final Grade on comparison to Class Test marks. The tentative reason may as follows: In general, the schedule of Class Test is declared about four/five days before the examination but the syllabus is very short. On the other hand, though the schedule of final examination is predefined but students get only a few days for preparation and the length of syllabus is relatively large. Therefore the student who does not attend the class frequently, he could not understand all of the course material within these few days. As a result the students who have poor class attendance, his Final Grade also become relatively poor. Therefore if students do not follow the classes i.e. do not attend in the class soundly, they cannot perform better in final examinations.

Table 4.9 Correlation between Attendance and Final Grade in EEE department and their test of significance

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,G}$	$\mathbf{t_{Cal}}$	$\mathbf{t_{(v,\alpha)}}$	$\mathbf{H_0}$
2001	1103	1st yr, 1st sem.	63	0.677395	7.19208	2.66	REJECT
2000	1203	1st yr, 2nd sem.	60	0.57119	5.2997	2.66	REJECT
2005	2103	2nd yr, 1st sem.	114	0.33321	3.74004	2.62	REJECT
2012	2103	2nd yr, 1st sem.	118	0.57633	7.59560	2.62	REJECT
2001	2103	2nd yr, 1st sem.	61	0.78954	9.88198	2.66	REJECT
2000	2203	2nd yr, 2nd sem.	51	0.21633	1.55106	2.68	ACCEPT*

Now we consider Electrical and Electronic Engineering (EEE) Department to investigate the correlation between Attendance and final Grade. The experimental results are displayed in the Table 4.9. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4.9. It is observed that, in most of the cases, the value of $R_{A,G}$ are greater than 0.5. In only one case, the value of $R_{A,G}$ are near 0.3 and at least greater than 0.2.

To test the significance of correlation between Attendance and Final Grade regarding $R_{A,G}$ value, we have calculated *t*, for each $R_{A,G}$ value and incorporated in the column 6 of the Table

4.9. Also standard $t_{(\nu,\alpha)}$ values, with ν (= *n*-2) d.f. and 0.005 level of significance, are displayed in column 7 of the table.

The inferences about the test of hypothesis are mentioned in column 8 of the table as well. It is noticed that except one case all the Null hypotheses are rejected with 99.5% confidence level. That is the corresponding $R_{A,G}$ values are significance with 0.005 level of significance. Note that though, in one case, the Null hypothesis is not rejected with 0.005 level of significance but with 5% level of significance the Null hypothesis of this case is also rejected and obviously all are rejected with 5% level of significance. Therefore it may conclude that with 5% level of significance all $R_{A,G}$ values are highly significant.

Now we consider Mechanical Engineering (ME) Department to investigate the correlation between Attendance and Final Grade. The experimental results are displayed in the Table 4.10. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4.10. It is observed that there exist only two set of information and in both cases the value of $R_{A,G}$ is greater than 0.5. Again we have calculated t_{Cal} values and compared with tabulated $\mathbf{t}_{(\nu,\alpha)}$ for 0.005 level of significance. We observed that for both cases H_0 are rejected for 5% level of significance. Therefore there exists significant correlation between Class Attendance and Final Grade with 5% level of significance.

Table 4.10 Correlation between Attendance and Final Grade in ME department and their test of significance

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,G}$	$\mathbf{t_{Cal}}$	$\mathbf{t}_{(\nu,\alpha)}$	$\mathbf{H_0}$
2010	2105	2nd yr, 1st sem.	115	0.538843	6.79954	2.62	REJECT
2012	2105	2nd yr, 1st sem.	112	0.573806	7.34822	2.62	REJECT

Now to investigate the correlation between Attendance and Final Grade for Computer Science and Engineering (CSE) Department. The experimental results are displayed in the Table 4.11. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4.11. It is observed that, in most of the cases, the value of $R_{A,G}$ are greater than or near equal to 0.4. In one case the value of $R_{A,G}$ are less than 0.2.

Now to test the significance of correlation between Attendance and Final Grade regarding $R_{A,G}$ value, we have calculated t_{Cal} for each $R_{A,G}$ and are reputed in the column 6 of the Table 4.11. Also standard $t_{(\nu,\alpha)}$ values, with ν d.f. and 0.5% level of significance i.e. 99.5% confidence level, are given in column 7 of the table.

Table 4.11 Correlation between Attendance and Final Grade in CSE department and their test of significance

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,G}$	$\mathbf{t_{Cal}}$	$\mathbf{t_{(\nu,\alpha)}}$	$\mathbf{H_0}$
2000	1207	1st yr, 2nd sem.	60	0.10781	0.82583	2.66	ACCEPT*
2005	2107	2nd yr, 1st sem.	52	0.68087	6.57357	2.68	REJECT
2003	2107	2nd yr, 1st sem.	57	0.43946	3.62824	2.66	REJECT
2012	2207	2nd yr, 2nd sem.	58	0.56608	5.13873	2.66	REJECT
2011	2207	2nd yr, 2nd sem.	59	0.62167	5.99207	2.66	REJECT
2002	2207	2nd yr, 2nd sem.	45	0.38360	2.72379	2.68	REJECT
2003	2207	2nd yr, 2nd sem.	58	0.55999	5.05808	2.66	REJECT
2009	2207	2nd yr, 2nd sem.	53	0.69051	6.81741	2.68	REJECT
2010	2207	2nd yr, 2nd sem.	59	0.60107	5.67816	2.66	REJECT

The inferences about the test of hypothesis are mentioned in column 8 of the table as well. It is noticed that except one case all the Null hypotheses are rejected with 99.5% confidence level. That is the $R_{A,G}$ values are significance with 0.5% level of significance. But in one case where t_{Cal} is less than 1.0 and therefore Null hypothesis is not rejected at 70% confidence too. Therefore except one case it may conclude that with 0.5% level of significance there exists strong evidence of correlation between Attendance and Final Grade in perspective to CSE department.

Now we consider Electronics and Communication Engineering (ECE) Department to investigate the correlation between Attendance and Final Grade. The experimental results are displayed in the Table 4.12. The value of coefficient of correlation of each course is calculated and presented in the column 5 of the Table 4. 12. It is observed that, in most of the cases, the value of $R_{A,G}$ are greater than or near equal to 0.4. In two cases the value of $R_{A,G}$ are less than 0.2.

Table 4.12 Correlation between Attendance and Final Grade in ECE department and their test of significance

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,G}$	$\mathbf{t}_{\mathbf{Cal}}$	$\mathbf{t}_{(\nu,.05)}$	$\mathbf{H_0}$
2002	1109	1st yr, 1st sem.	31	0.41273	2.4402	1.697	REJECT
2003	1109	1st yr, 1st sem.	27	0.34632	1.8458	1.703	REJECT
2003	2109	2nd yr, 1st sem.	31	0.07498	0.40494	1.697	ACCEPT*
2009	2109	2nd yr, 1st sem.	30	0.75673	6.12514	1.699	REJECT
2012	2209	2nd yr, 2nd sem.	53	0.55243	4.73291	1.671	REJECT
2009	2209	2nd yr, 2nd sem.	29	0.12373	0.64792	1.7	ACCEPT*

To test the significance of correlation between Attendance and Final Grade regarding $R_{A,G}$ value, we have calculated t_{Cal} for each $R_{A,G}$ and the values are presented in the column 6 of the Table 4.12. Also standard $t_{(\nu,\alpha)}$ values, with ν d.f. and 5% (rather than 0.5%) level of significance, are displayed in column 7 of the table.

The inferences about the test of hypothesis are mentioned in column 8 of the table as well. It is noticed that except two cases all the Null hypotheses are rejected with 95% confidence level. That is the $R_{A,G}$ values are significance with 5% level of significance. But for the remaining two cases, Null hypotheses are not rejected with 30% level of significance too. Therefore it may conclude that except two the $R_{A,G}$ values are significant with 5% level of significance.

Now we consider IEM and LE Departments to investigate the correlation between Attendance and Final Grade. The experimental results are displayed in the Table 4.13. It is observed that there is only one set of information is available for IEM and one set of information is available for LE department. We notice the $R_{A,C}$ value for IEM is greater than 0.3 and the $R_{A,C}$ value for LE is greater than 0.6. It is also observed that for both cases the Null hypothesis is rejected with 1% level of significance. Therefore there exist significance correlation between class Attendance and Final Grade with 99% confidence level.

Table 4.13 Correlation between Attendance and Final Grade in IEM and LE department and their test of significance

Conducted Year	**Course Code (Math)**	**Semester/term**	**No. of Student (*N*)**	$R_{A,G}$	$\mathbf{t_{Cal}}$	$\mathbf{t}_{(v,.01)}$	$\mathbf{H_0}$
2011	1211	1st yr, 2nd sem.	58	0.33441	2.65537	2.390	REJECT
2013	2119	2nd yr, 1st sem.	43	0.66872	5.75898	2.423	REJECT

CHAPTER V

REGRESSION ANALYSIS OF ATTENDANCE AND ACADEMIC ATTAINMENTS

5.l Introduction

In Chapter IV we have investigated extensively for the existence of the correlation between attendance and Class Test as well as correlation between Attendance and Final Grade. We observe that attendance has significance impact on both Class Test Marks as well as Final Grade regarding Mathematics courses. Here we will perform extensive experiments to establish a simple meta regression model of academic performance on attendance.

5.2 Experimental study

Before carry on experiments about regression analysis among attendance and academic performance, it is noted here that we will consider all data and conditions used in Chapter IV in perceptive CE department. It is worthwhile to mention here that among the all departments; we have able to collect comparatively more data about CE department. Therefore we will investigate regression between attendance and academic attainments only for CE department.

In this section, we will perform several experiments to examine the existence of the simple meta regression model of Class Test Mark on Class Attendance and also examine the existence of the simple meta regression model of Final Grade on Class attendance by ignoring proxy variables. Note that total numbers of classes of each course are normalized to 30 classes. Class test marks are also normalized to 30 marks and Final Grade are calculated in out of 4. Also note that, to find the regression model, we use Least Square Curve fitting tools of MatLab package. It is also noted that in the figures CT means Class Test Marks.

5.2.1 Regression analysis between Class Attendance and Class Test (CT) Marks

Firstly, we will investigate regression between class attendance and class test marks. For this experimental study we have considered data of Civil Engineering Department. For the regression model, we have considered three type of regression model given bellow:

- Linear model: $f(x) = p1 \cdot x + p2$ (5.1)
- Polynomial model $f(x) = p1 \cdot x^2 + p2 \cdot x + p3$ (5.2)
- Power/Exponential model $f(x) = a \cdot x^b$ (5.3)

At first we consider the data of 1st year 1st semester, CE: 2001. The data are plotted and then fitted by using MatLab fitting tools. The plotted data as well as fitting curves are displayed in Figure 5.1(a), 5.1(b) and 5.1(c).

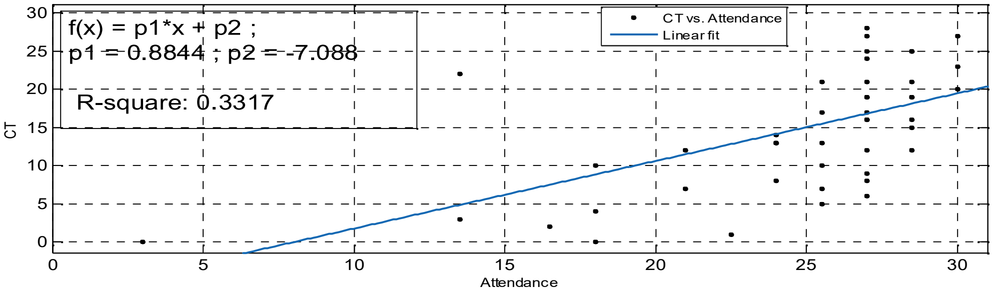

Figure 5.1(a) Linear fitting of Attendance vs. CT (1st year 1st semester, CE: 2001)

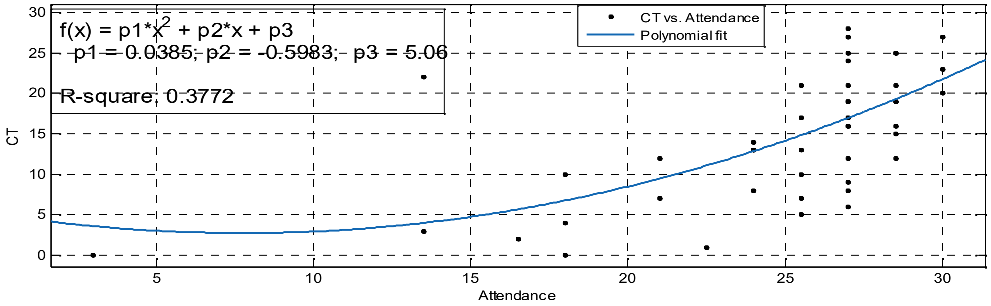

Figure 5.1(b) Polynomial fitting of Attendance vs. CT (1st year 1st semester, CE: 2001)

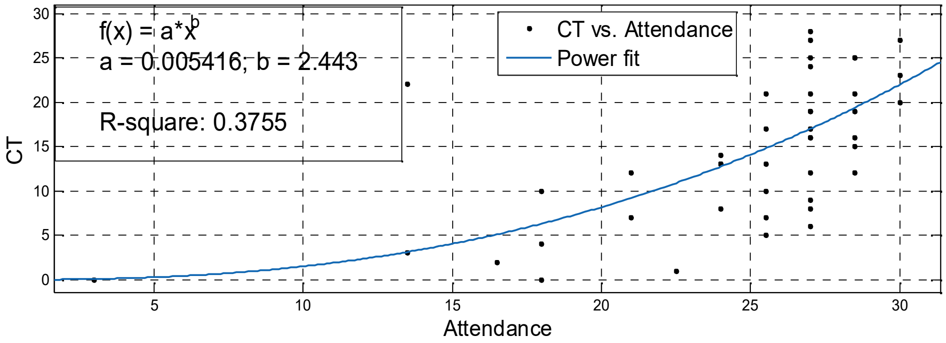

Figure 5.1(c) Exponential fitting of Attendance vs. CT (1st year 1st semester, CE: 2001)

It is observed that most of the data regarding attendance are lies between 20 – 30 because 60% attendance is imposed. Now Figure 5.1(a) represents the linear fitting whereas Figure 5.1(b) represents polynomial fitting and Figure 5.1(c) represents exponential fitting for the same data. The values of coefficient of determination (R^2) of (a) linear fitting is 0.3317, (b) polynomial fitting is 0.3772 and (c) exponential fitting is 0.3755. In this data set, the R^2 values of nonlinear fitting are a bit better than that of linear model.

Again we consider another data set namely 1st year 2nd semester, CE: 2001. The data are plotted and then fitted by using Mat Lab fitting tools. The plotted data as well as fitting curves are displayed in Figure 5.2(a), 5.2(b) and 5.2(c). Here Figure 5.2(a) represents the linear fitting whereas Figure 5.2(b) represents polynomial fitting and Figure 5.2(c) represents exponential fitting for the same data. Here it is noticed that though much amount of data are greater than 20 regarding attendance but there exist more significant data less than 15. Consequence the values of coefficient of determination (R^2) value are significantly large. We notice that R^2 values of all the fitting curves are almost identical and approximately 0.6.

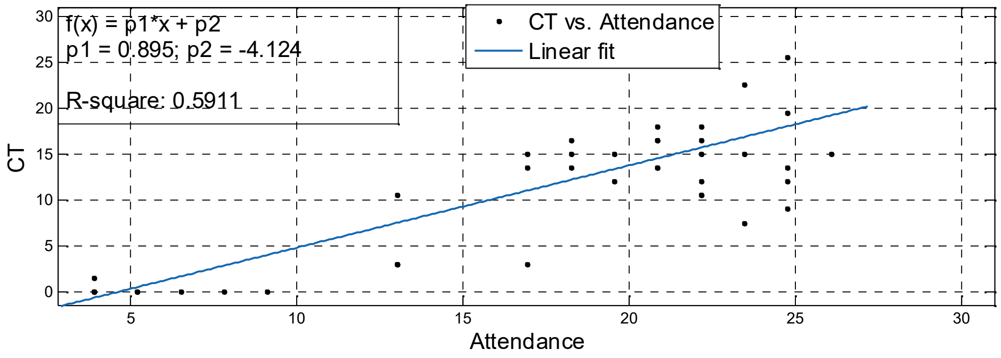

Figure 5.2(a) Linear fitting of Attendance vs. CT (1st year 2nd semester, CE: 2001)

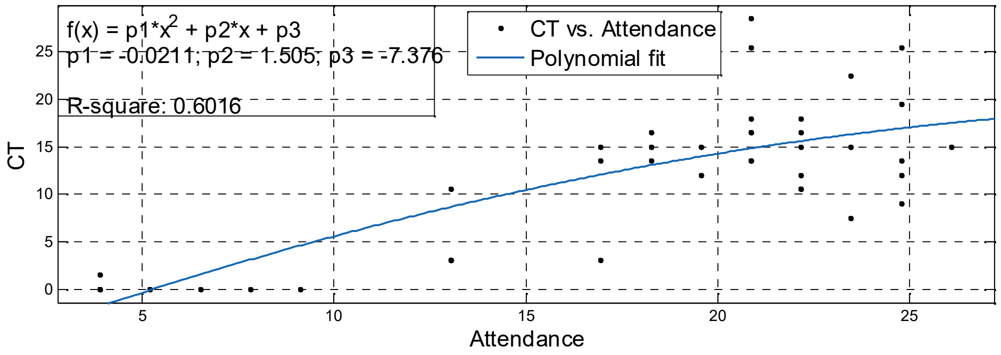

Figure 5.2(b) Polynomial fitting of Attendance vs. CT (1st year 2nd semester, CE: 2001)

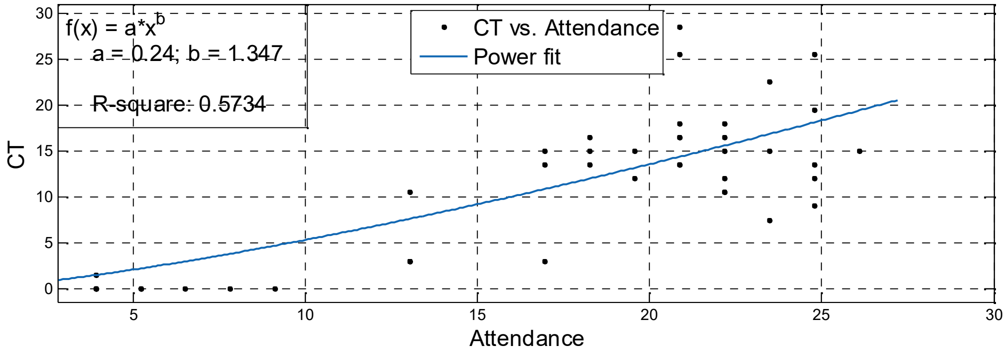

Figure 5.2(c) Exponential fitting of Attendance vs. CT (1st year 2nd semester, CE: 2001)

Again we consider another data set namely 2^st^ year 1^st^ semester, CE: 2009. The data are plotted and then fitted by using Mat Lab fitting tools. The plotted data as well as fitting curves are displayed in Figure 5.3(a), 5.3(b) and 5.3(c). Again it is noticed that most of the data regarding attendance are lies between 20 – 30 because 60% attendance is imposed.

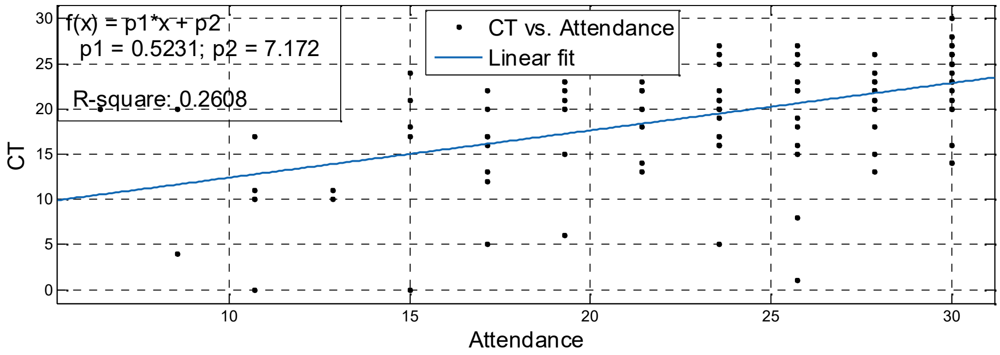

Figure 5.3(a) Linear fitting of Attendance vs. CT (2nd year 1st semester, CE: 2009)

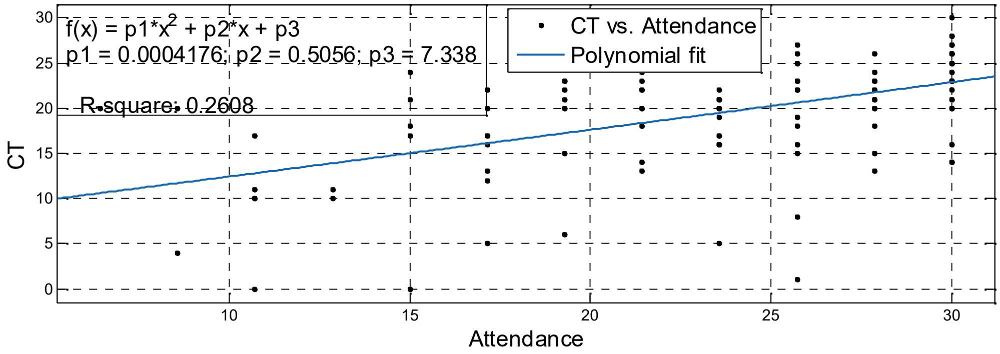

Figure 5.3(b) Polynomial fitting of Attendance vs. CT (2nd year 1st semester, CE: 2009)

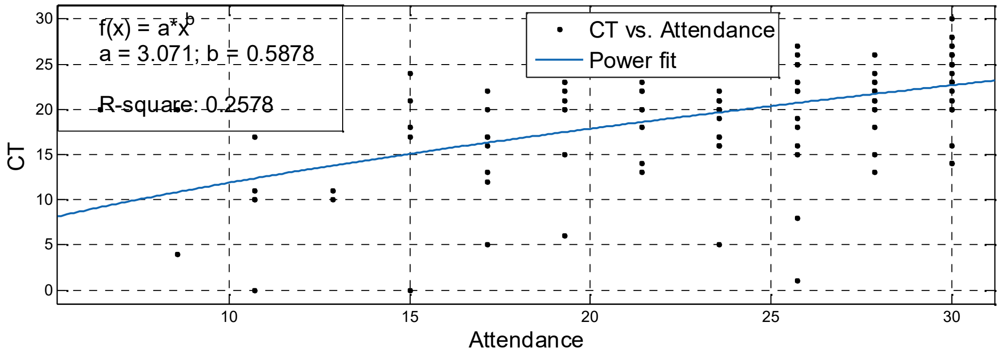

Figure 5.3(c) Exponential fitting of Attendance vs. CT (2nd year 1st semester, CE: 2009)

Table 5.1: Parameter values of fitted model for Attendance vs. CT and R^2 values in CE department

Conducted Year	Course Math-	Model	R^2	p1 (or a)	p2 (or b)	p3
2001	1101	Linear	0.33	0.8844	-7.088	
		Polynomial	0.3772	0.0385	-0.5983	5.06
		Power	0.3755	0.00542	2.443	
2002	1101	Linear	0.42	1.015	-2.399	
		Polynomial	0.4227	-0.0059	1.204	-3.445
2001	1201	Linear	0.59	0.895	-4.124	
		Polynomial	0.6016	-0.0211	1.505	-7.376
		Power	0.5734	0.24	1.347	
2002	1201	Linear	0.19	1.531	-23.8	
		Polynomial	0.215	0.156	-5.888	62.91
		Power	0.2049	0.0003949	3.249	
2004	1201	Linear	0.13	0.7343	3.032	
		Polynomial	0.1838	-0.1078	5.579	-49.26
		Power	0.1338	1.629	0.8005	
2011	1201	Linear	0.03	0.2891	12.45	
		Polynomial	0.027	0.01313	-0.3452	19.87
		Power	0.0248	6.414	0.3495	
2003	1201	Linear	0.06	0.3279	11.9	
		Polynomial	0.0657	0.00857	-0.05742	16.1
		Power	0.0616	6.091	0.3713	
2002	2101	Linear	0.4	2	1.9	
		Polynomial	0.3	0.59	1.204	3.445
		Power	0.59	0.895	4.124	
2004	2101	Linear	0.4	1.9	2	
		Polynomial	0.3	0.9	1.2	
		Power	0.029	0.995	4.12	
2012	2101	Linear	0.05	0.3553	11.49	
		Polynomial	0.0637	0.03622	-1.361	31.24
		Power	0.0505	5.393	0.4137	
2013	2101	Linear	0.09	0.3019	18.05	
		Polynomial	0.1078	0.01914	-0.5492	26.93
		Power	0.0843	11.52	0.2494	
2009	2101	Linear	0.26	0.5231	7.172	
		Polynomial	0.2608	0.00041	0.5056	7.338
		Power	0.2578	3.071	0.5878	
2010	2201	Linear	0.11	0.8074	-5.341	
		Polynomial	0.1393	0.1015	-4.158	53.84
		Power	0.1117	0.09267	1.569	
For Linear Model		Linear	0.42	0.889569	1.941692	

Figure 5.3(a) represents the linear fitting whereas Figure 5.3(b) represents polynomial fitting and Figure 5.3 (c) represents exponential fitting for the same data. The values of coefficient of determination (R^2) of (a) linear fitting is 0.2608, (b) polynomial fitting is 0.2608 and (c) Exponential fitting is 0.2578. In this data set, the R^2 values of nonlinear fitting as well as linear fitting are almost identical. Similarly we have tested several model equations with all data regarding CE and try to fit each of the above mentioned models. The characteristics of the fitted models are summarized in Table 5.1.

It is observed in the Table 5.1 that on an average the linear model is the best among the three models considered here. It is also noticed that though the values of regression coefficient are vary from data to data but on an average in case of linear regression model the value of R^2 is 0.42, the value of regression coefficient $p1$ = 00.89 and $p2$ = 1.94. So angle of the slope of the linear regression line is about 40^0 and the line cut the ordinate at about 2.0 unit ahead.

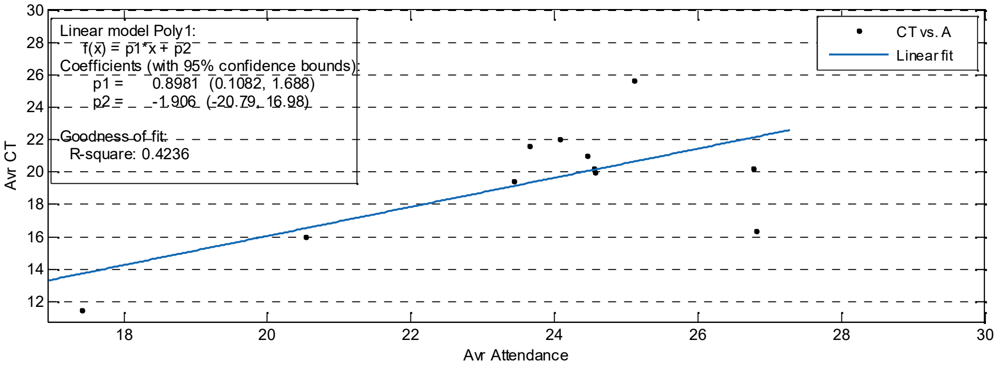

Figure 5.4(a) Linear fitting of average Attendance vs. average CT of CE

Again we have performed regression analysis on average Attendance vs. average CT marks for CE. The experimental results are depicted in Figure 5.4(a), 5.4(b) and 5.4(c). Here Figure 5.4(a) represents Linear fitting whereas Figure 5.4(b) represents polynomial fitting and Figure 5.4(c) represents exponential fitting for the same data. We observed that the R^2 values of linear fitting as well as Power fitting are almost identical and is 0.42.Whereas the R^2 values of Polynomial fitting is 0.69. But the graphical view (Figure 5.4(b)) imposes that it would be not a valid model. Since according to this Polynomial model, when attendance becomes larger than 24 the CT marks goes down. Again the graphical view as well as the estimated value of the coefficients of Power methods (Figure 5.4(c)) reveals that the model is almost linear. Now according to the estimated Linear model corresponding to the average Attendance vs. average CT Marks of CE department $p1$ = 0.9 and $p2$= 1.9 which is agree with the previously obtained average value of $p1$ and $p2$ of all linear models regarding CE departments.

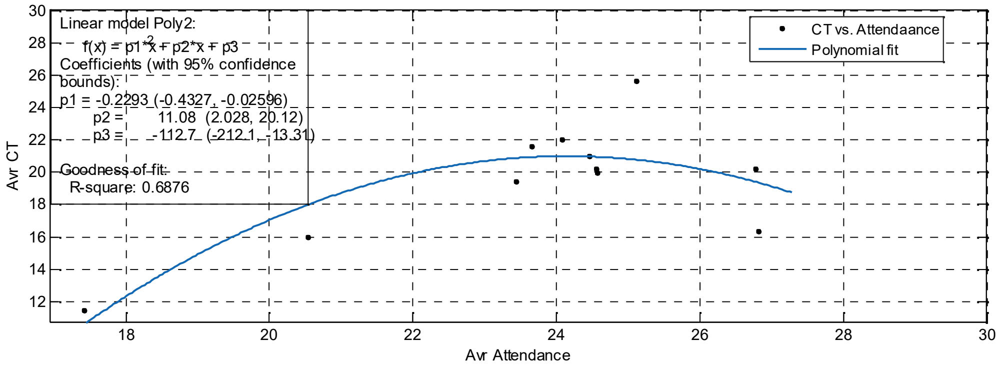

Figure 5.4(b) Polynomial fitting of average Attendance vs. average CT of CE

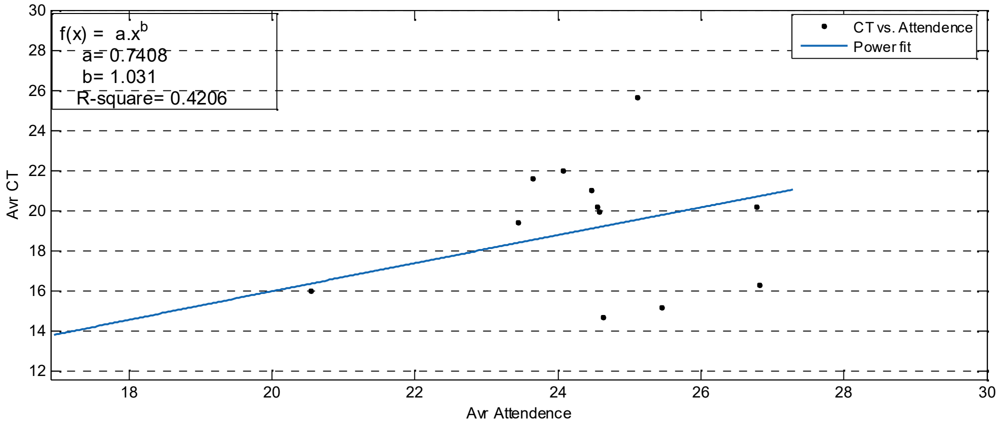

Figure 5.4(c) Power fitting of average Attendance vs. average CT of CE

Therefore according to the data of class Attendance and Class Test marks in perspective of CE department, the estimated regression model be a linear Meta-model in which proxy variables are ignored. Moreover according to the average data the R^2 value is 0.42 which is of no doubt a significant value. Therefore the estimated simple Meta linear regression model of CT marks on class Attendance is as follow:

$$f(x) = 0.9 \cdot x + 1.9 \tag{5.4}$$

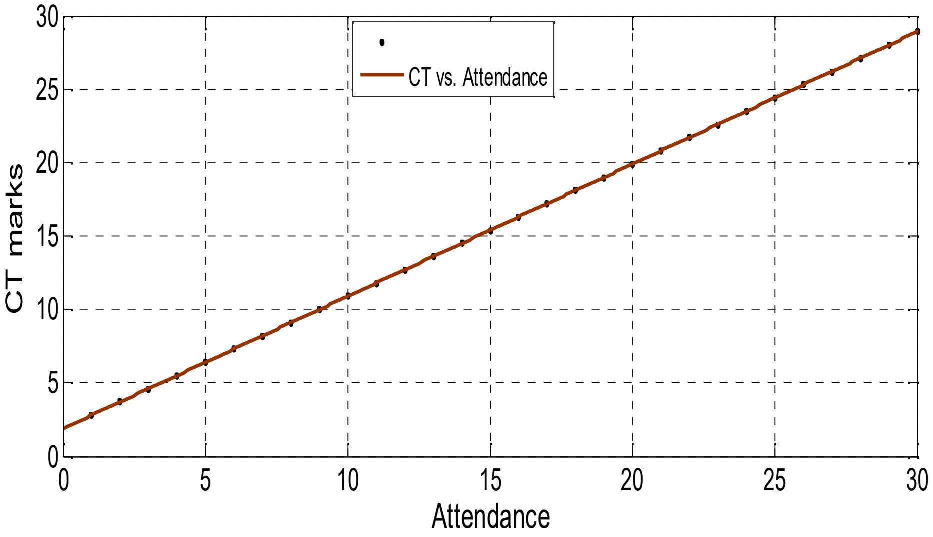

Figure 5.5 Simple Meta Linear Regression Model of CT marks on class Attendance for CE

The graphical view of simple Meta Linear Regression model of CT marks on class Attendance for CE is given in Figure 5.5. Similarly we will able to find out simple linear regression model of CT marks on class attendance for each Department. But for the lack of sufficient data we could not establish Meta regression model for other departments as well as a general Meta model for the whole university, KUET.

5.2.2 Regression analysis between Class Attendance and Final Grade

Now we will investigate regressions between class attendance and Final Grade. For this experimental study we will again consider Civil Engineering Department. For the regression model we consider three type of regression model as considered earlier. Now we consider the data of 1st year 2nd semester, CE: 2001. The data are plotted and then fitted by using MatLab fitting tools. The plotted data as well as fitting curves are displayed in Figures 5.6(a), 5.6(b) and 5.6(c).

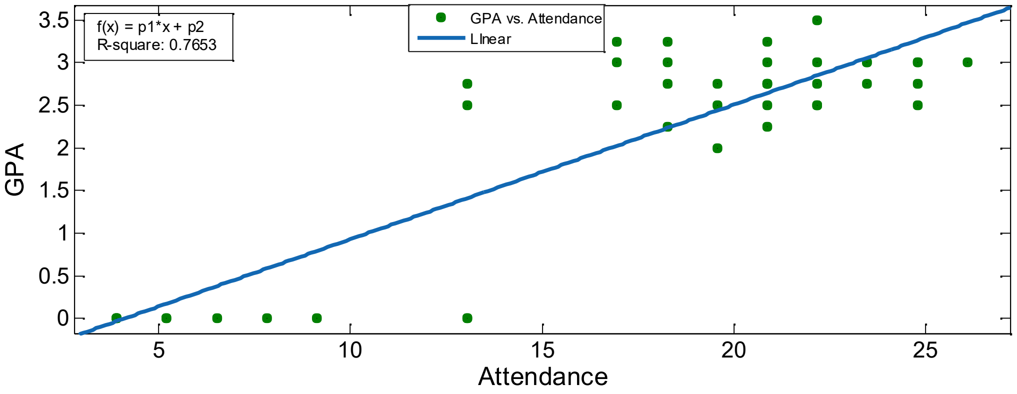

Figure 5.6 (a) Linear model of GPA on Attendance; CE: 1201, 2001

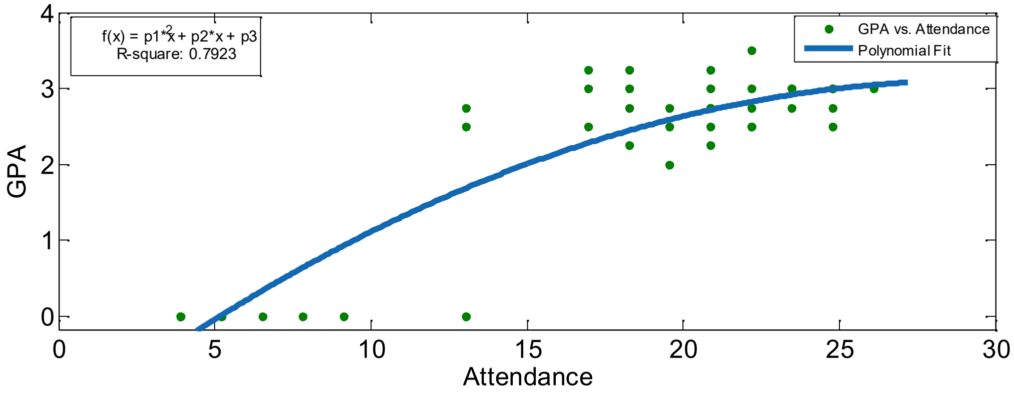

Figure 5.6 (b) Polynomial model of GPA on Attendance; CE: 1201, 2001

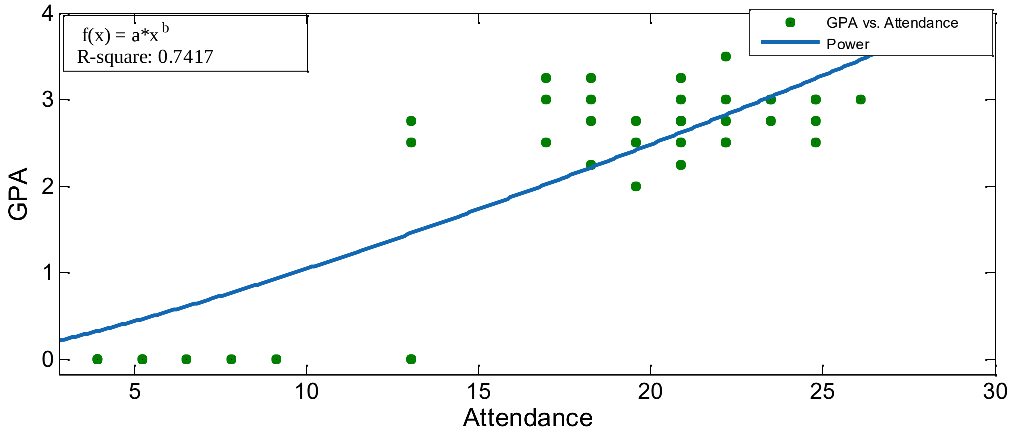

Figure 5.6 (c) Power model of GPA on Attendance; CE: 1201, 2001

It is observed that the GPA is 0 when number of attendance is 15. Because greater that 50% (actually 60%) attendance is mandatory to appear in the final examination. Now Figure 5.6(a) represents the linear fitting whereas Figure 5.6(b) represents polynomial fitting and Figure 5.6(c) represents exponential fitting for the same data. The value of coefficient of determination (R^2) of (a) linear fitting is 0.77, (b) polynomial fitting is 0.79 and (c) exponential fitting is 0.74. In this data set, we observe that the R^2 values of nonlinear fitting are a bit better than that of linear model as well as Power model. Note that the polynomial model is almost looked like a logarithmic curve.

Similarly we have tested several model equations with all data regarding CE and tried to fit each of the above mentioned models. The characteristics of the fitted models are summarized in Table 5.2. It is observed in the Table 5.2 that on average the value of R^2 is almost identical. The values of R^2 of Linear model, Polynomial Model and Power model are 0.37, 0.38 and 0.36 respectively. Since the R^2 value of Power model is relatively small so we will consider here only Liner model and Polynomial model. Now, on average, the p1 and p2 values of linear model are 0.138 and -0.899 respectively. Therefore according to the linear model the slop of regression line is about 8^0 and Eq. (5.5) be the corresponding equation.

- Linear model: $f(x) = 0.138 \cdot x - 0.899$ (5.5)
- Polynomial model $f(x) = 0.0028 \cdot x^2 + 0.125 \cdot x - 0.487$ (5.6)

On the other hand, on average, the *p*1, *p*2 and *p*3 values of Polynomial model are 0.002808, 0.125212 and -0.48677 respectively. Therefore Eq. (5.6) represents corresponding Polynomial regression model.

Table 5.2: Parameter values of fitted model for Attendance vs. GPA and corresponding R^2 values in CE department

Conducted Year	**Course Math-**	**Model**	**R^2**	**p1 (or a)**	**p2 (or b)**	**p3**
2001	1101	Linear	0.42	0.1457	-1.716	
		Polynomial	0.4462	0.003834	-0.002011	-0.5065
		Power	0.4345	0.0007034	2.444	
2002	1101	Linear	0.48	0.1405	-0.9961	

		Polynomial	0.5127	0.003626	0.02479	-0.3534
2001	1201	Linear	0.77	0.1579	-0.6514	
		Polynomial	0.7923	-0.005254	0.3099	-1.461
		Power	0.7417	0.0592	1.247	
2002	1201	Linear	0.13	0.1221	-0.3941	
		Polynomial	0.1787	0.01942	-0.8017	10.4
		Power	0.1347	0.04636	1.256	
2004	1201	Linear	0.12	0.05701	1.213	
		Polynomial	0.1205	-0.0007972	0.09284	0.8264
		Power	0.1202	0.5191	0.506	
2011	1201	Linear	0.3386	0.1982	-2.899	
		Polynomial	0.3398	0.002444	0.08017	-1.519
		Power	0.337	0.0005665	2.533	
2003	1201	Linear	0.12	0.1071	0.1636	
		Polynomial	0.124	0.0006411	0.07825	0.478
		Power	0.1237	0.1359	0.9443	
2002	2101	Linear	0.79	0.1629	-1.011	
		Polynomial	0.8454	-0.01124	0.528	-3.301
		Power	0.7505	0.03119	1.423	
2004	2101	Linear	0.16	0.09662	-0.2661	
		Polynomial	0.1672	-0.003318	0.2441	-1.847
		Power	0.1612	0.0605	1.108	
2012	2101	Linear	0.34	0.1626	-1.376	
		Polynomial	0.3492	-0.004822	0.3911	-4.005
		Power	0.3381	0.02029	1.515	
2013	2101	Linear	0.29	0.1064	0.5954	
		Polynomial	0.287	0.0008946	0.06663	1.011
		Power	0.2847	0.2477	0.8011	
2009	2101	Linear	0.45	0.127	-1.238	
		Polynomial	0.4573	0.001719	0.05496	-0.5558
		Power	0.4554	0.003558	1.946	
2010	2201	Linear	0.40	0.2157	-3.118	
		Polynomial	0.4035	-0.004105	0.4166	-5.513
		Power	0.3868	0.001315	2.31	
Average		Linear	0.36989	0.13844	-0.8995	
		Polynomial	0.38038	0.002808	0.125212	-0.48677
		Power	0.35570	0.09387	1.51004	

Now according to Polynomial model Eq. (5.6), we have plotted the graph which is shown in Figure 5.7. It is observed in the figure that when attendance is greater or equal to 25 then estimated GPA becomes greater than 4 (out of 4) which is inconsistent. So we should regret the polynomial Model.

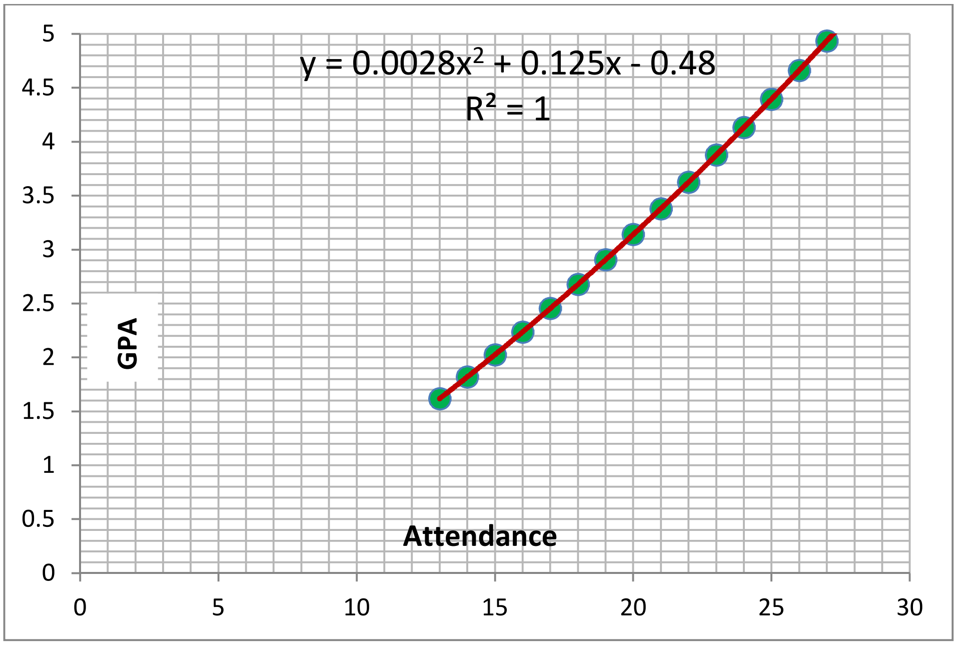

Figure 5.7 Polynomial regression of GPA on Attendance by taking average parameter values

Again we have plotted equation (5.5) which is corresponding to linear model. The graphical representation of the linear model is shown in Figure 5.8. It is observed in the figure that when attendance is, on average, 30 than on average the final Grade i.e. GPA becomes about 3.25. Now let us modify the above Linear model given by equation 5.5 as 5.7. This has been done under the following consideration.

) Linear model:

$$\begin{aligned} G &= 0.138 \cdot x - 0.155\,;\text{when } 13 < x \leq 30 \\ &= 0;\text{when } 0 \leq x \leq 13 \end{aligned} \qquad (5.7)$$

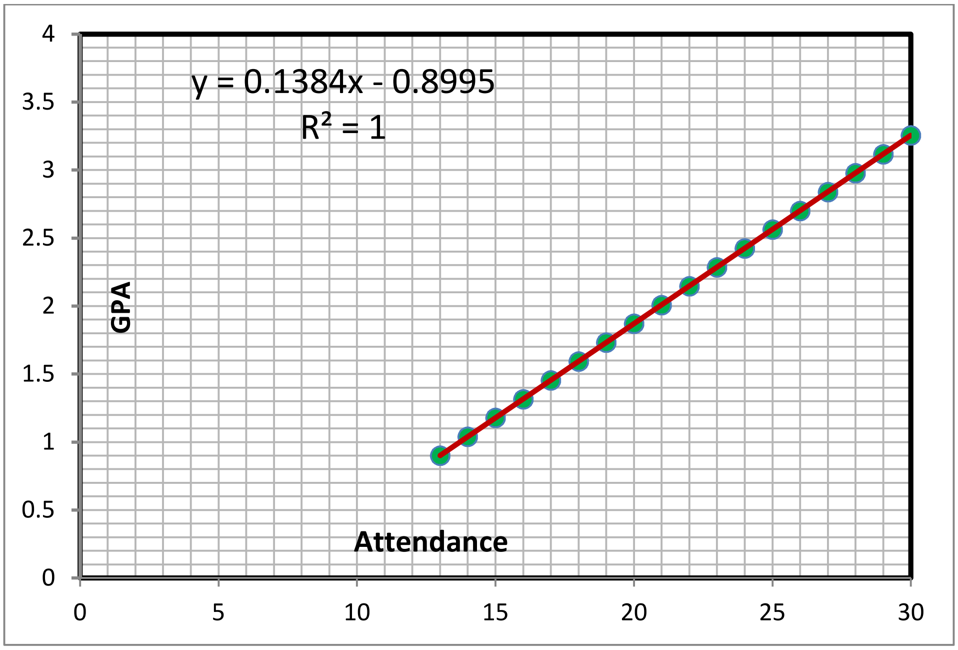

Figure 5.8 Regression Line of GPA on Attendance by taking average parameter values

As we mentioned earlier that attendances are normalized to 30 and 60% (above 45% attendance was allowed in some instances) attendance is mandatory to appear in the final examination. So number of attendance less than 13 will get 0 as GPA. Moreover kept unchanged the regression coefficient p1 and replaced the constant term $p2$ = -0.889 to $p2$= -0.155, we have the new Linear regression model. Now we have plotted the modified linear model which is displayed in the Figure 5.9. It is noticed in the Figure 5.9 that when value of attendance is 30 than corresponding estimated GPA becomes about 4 (out of 4). Note that in the Eq. (5.7) variable x represent normalized class attendance.

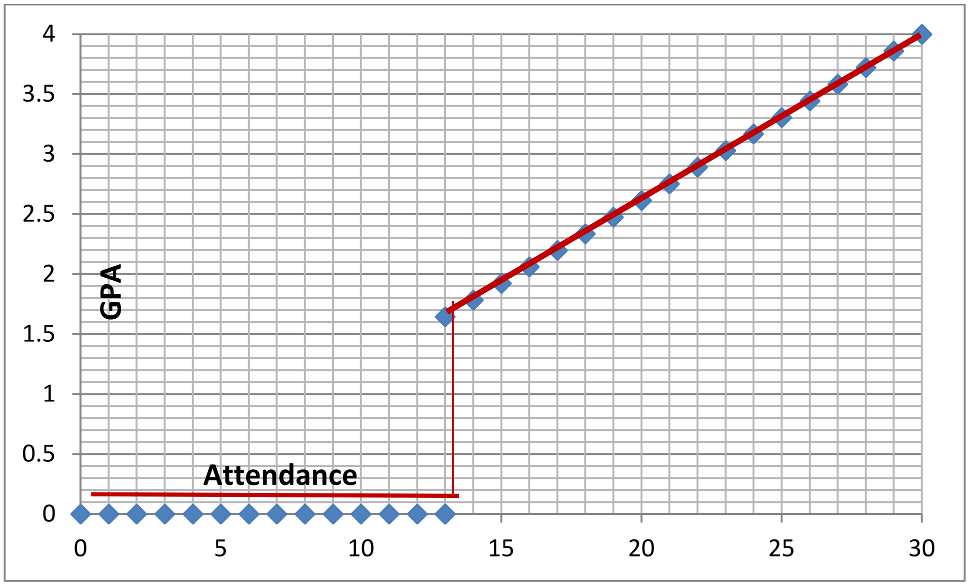

Figure 5.9 Modified Linear Regression model of GPA on Attendance of CE department

5.2.3 Multiple Regression analysis among Attendance, CT marks and Final Grade

Now we will investigate multiple linear regression models on Final Grade on class attendance and CT marks. The mathematical form of the multiple linear model and multiple polynomial model are described by the equation (5.8) and (5.9).

- Multiple Linear model: $G = p00 + p10 \cdot x + p01 \cdot y$ (5.8)
- Multiple Poly. model: $G = p00 + p10 \cdot x + p01 \cdot y + p20 \cdot x^2 + p11 \cdot x$ (5.9)

At first, in this aspect, we consider CE: Math 1201, year 2001. We have plotted the data and fit by a linear multiple regression model. The estimated multiple linear regression model is displayed in the Figure 5.10. It is observed that the R^2 value of the model is 0.81 which implies that this model sufficiently significant. According to the estimated model the mathematical equation of the multiple linear regression model is as follow:

$$G = -0.45 + 0.1142x + 0.0488y \qquad (5.8)$$

where *x* represents for attendance and *y* represents for CT marks.

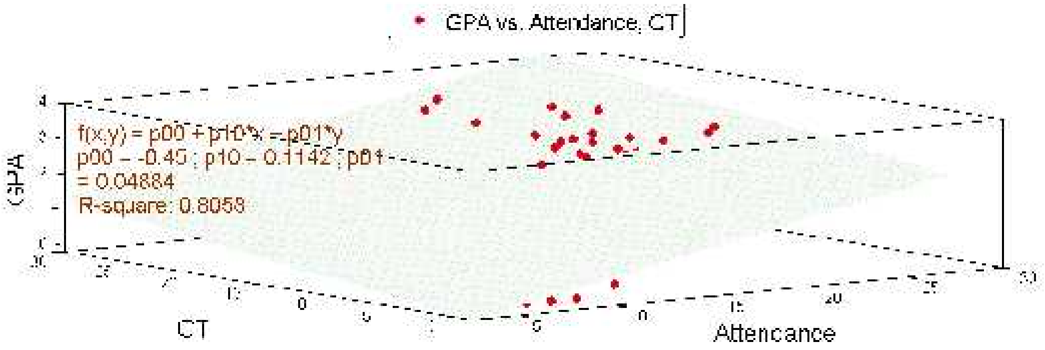

Figure 5.10 Multiple linear Regression model of GPA on Attendance and CT
of CE: Math 1201, Year: 2001

Again the estimated polynomial multiple regression of the above data are displayed in Figure 5.11. It is observed in the Figure 5.11 that the R^2 value is only 0.202 which in not very significant. So we regret polynomial multiple regression model.

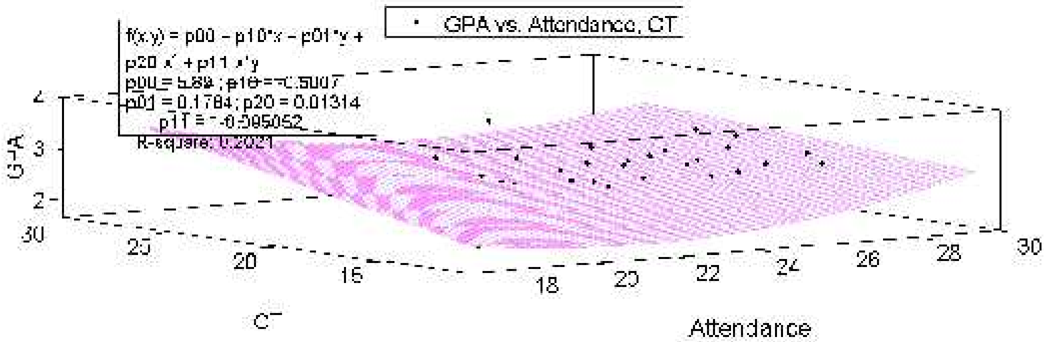

Figure 5.11 Polynomial Multiple regression Model of GPA on Attendance and CT
of CE: Math 1201, Year: 2001

Similarly we have estimated the multiple regression models for other departments especially we have considered the data which have significant R^2 values. The estimated models are summarized in the Table 5.3. From the experimental study it is observed that linear models and Polynomial (only of Attendance value) model are almost identical regarding R^2 values.

Moreover it also worthwhile to be mentioned here that the patterns of the models are vary year to year as well department to departments.

Table 5.3 Some estimated multiple regression of GPA on Attendance and CT marks of different departments which have good R^2 values.

Year	**Course Math-**	**Model**	**R^2**	**p00**	**p10**	**p01**	**p20**	**p11**
2001	1201	Linear	0.8058	-0.45	0.1142	0.04884		
		Polynom	0.8674	-0.1937	-0.006572	0.3364	0.005891	-0.01456
2001	2103	Linear	0.6566	-0.2459	0.1141	0.03618		
		Polynom	0.6649	-0.4008	0.1113	0.09006	0.0005617	-0.002521
2005	2107	Linear	0.5631	-3.859	0.1873	0.04976		
		Polynom	0.5665	-6.864	0.4999	-0.08193	-0.007258	0.00463
2009	2109	Linear	0.6733	-1.156	0.121	0.03976		
		Polynom	0.739	2.599	-0.4492	0.401	0.01544	-0.01289

CHAPTER VI

CONCLUSION

As it has been seen from the earlier researches, the performance of students depend on many factors such as, attendance, motivation, level of engagement etc. which may be considered as the students attributes towards learning and some other attributes of the teachers on the process. No such research has been conducted and is published in case of universities of our country, especially of KUET in which at least 60% attendance is imposed. Though it has been seen from other researches, many factors may be considered but we are interested and selected the factors namely class attendance for the assessment of the academic attainment.

Table 6.1 Number of courses of different departments considered for the study

Name of Departments	Number of Courses
CE	11
EEE	07
ME	01
CSE	06
ECE	05
IEM	01
LE	01
TOTAL	32

There are many departments in this university (KUET) and a lot of subjects are taught. But Mathematics is common to all. Hence Mathematics is chosen for this study. This study is done among the students of first year and second year in several engineering departments regarding mathematics courses during the period 2000 -2013. It is of no doubt that for the existence of proxy variables (Teacher’s attribute, student’s attribute, subjects, socio-economic environment etc.), it is very difficult and hard working to find out the impact of class attendance on academic attainment. Moreover, due to the imposed of mandatory percentage of class attendance, it is very difficult to find out the impact of attendance on final grade. In this study we have considered only existing old data (student attendance and academic performance), where the effect of proxy variables are ignored. Moreover, for better comparison as well as for finding some test statistics, all the data are normalized.

Table 6.2 Different parameters of student's academic performance (in average)

Name of Departments	**AT (30)**	**CT(30)**	**GPA(4)**
CE	24.58	19.70	2.50
EEE	25.46	18.90	2.76
ME	24.76	16.62	2.56
CSE	23.80	18.60	2.68
ECE	25.78	17.73	2.86
IEM	27.11	20.20	2.27
LE	24.88	20.32	2.61

In this study we have considered 32 set of courses, within 2000 – 2013, which are displayed in the Table 6.1. We also have summarized the student average statistics namely average number of attendance, average CT marks and average GPA and displayed in the Table 6.2. Anyway, in CHAPTER IV, we have rigorously studied about the correlation between class Attendance and Class Test Marks for each course. From these experimental studies, it may conclude that there exist correlation between class Attendance and Class Test marks in each department. In this we have also investigated thoroughly in each course for the existence of correlation between class Attendance and Final Grade. From the experimental study, it has been seen that the values of correlation coefficients vary from year to year as well as department to department. Perhaps it was the results of proxy variables and socio-economic effects.

The experimental studies reveal that there exist strong correlations between Attendance and Final Grade in perceptive of all departments. Though there exist correlation between Class Attendance and CT marks as well as correlation between Class Attendance and Final marks, but correlation between Class Attendance and final Grade is stronger than that of Class Attendance and CT marks. It may reveals that class attendance grow some intuitive knowledge to the students which effect on his final Grade.

After finding the existence of correlations among attendance, CT and Final Grade, we have performed intensive experiments for regression analysis. It is observed from the experimental study that though there exist regression model of CT on attendance as well as GPA on attendance but according to the R^2 value the regression models of GPA on attendance are more significant than that of CT on attendance. It is also noticed that regression coefficients are varying from year to year as well as department to department. We have also estimated a Meta linear model of CT as well as a Meta linear model of GPA on attendance in perspective CE department. Finally we have investigated a multiple regression model of GPA on attendance and CT marks. It is observed that though multiple linear models shows good result but some instances show very poor results regarding R^2 values.

In spite of mandatory of 60% attendance, from the experimental results, it is revealed that attendance has a great effect on academic attainments. But the effect is varying department to department as well as semester to semester. To establish a general Meta model regarding academic attainments, class attendance and socio-economic changes further experiments need to perform especially sufficiently much more data will be required. In this study though we have much information regarding CE departments but we have no sufficient data in other departments.

References

1. Ali, N., K. Jusoff, S. Ali, N. Mokhtar and A.S.A. Salamat (2009), The factors influencing students' performance at University Teknologi MARA Kedah, Malaysia. Management Science and Engineering, Vol. 3, No. 4, pp. 81-90.
2. Applegate, K. (2003), The relationship of attendance, socio-economic status, and ability and the achievement of seventh graders, doctoral dissertation, Saint Louis University, St. Louis, MO.
3. Arthur, W., W. Bennett, P. S. Eden, and S. T. Bell (2003), Effectiveness of training in organizations: A meta-analysis of design and evaluation features, Journal of Applied Psychology, Vol. 88, pp. 234–245.
4. Bornstein, M. C. and R. H. Bradley (Eds.) (2003), Socio-economic status, parenting, and child development, Mahwah, NJ: Lawrence Erlbaum.
5. Brooks-Gunn, J. and G. J. Duncan (1997), The effects of poverty on children, The future of children, Vol. 7, No. 2, pp. 55–71.
6. Cohall, D. H. (2009), Course outline for Fundamentals of Disease & Treatment, Faculty of Medical Sciences, Cave Hill, Barbados: The University of the West Indies.
7. Coleman, J. S. (1988), Social capital in the creation of human capital, American Journal of Sociology, Vol. 94,S95–S120.
8. Crede, M., S. G. Roch and U. M. Kieszczynka (2010), Class attendance of undergraduate students: A meta-analytic review of the relationship of class attendance with grades and student characteristics, Review of Educational Research, Vol. 80 No. 2, pp. 272–295.
9. Damian H. Cohall and Desiree Skeete (2012), The impact of an attendance policy on the academic performance of first year medical students taking the Fundamentals of Disease and Treatment course, Caribbean Teaching Scholar, era educational research association, Vol. 2, No. 2, pp.115–123.

10. Devadoss, S. and J. Foltz (1996), Evaluation of Factors Influencing Student Class Attendance and Performance, American journal of Agriculture Economics, Vol. 78, pp. 499- 507.
11. Durden, G.C. and L.V. Ellis (1995), “The effects of attendance on student learning in principles of economics”, American Economic Review Papers and Proceedings, Vol. 85, No. 2, pp. 343-346.
12. Entwisle, D. R. and N. M. Astone (1994), Some practical guidelines for measuring youth’s race/ethnicity and socioeconomic status, Child Development, Vol. 65, No. 6, pp. 1521–1540.
13. Friedman, P., F. Rodriguez, and J. McComb (2001), Why students do and do not attend class, College Teaching, Vol. 49, pp. 124-133.
14. Gamble, Z. P. (2004), The effect of student mobility on achievement and gain-score test results (Unpublished doctoral dissertation), University of Tennessee, Memphis, TN.
15. Guleker, R. and J. Keci (2014), The Effect of Attendance on Academic Performance, Mediterranean Journal of Social Sciences MCSER Publishing, Rome-Italy, Vol. 5, No. 3.
16. Hancock, T.M. (1994), Effects of mandatory attendance on student performance, College Student Journal, Vol. 28, pp. 326-329.
17. Haveman, R. and B. Wolfe (1994), Succeeding generations: On the effects of investment in children, NY: Russell Sage.
18. Hyde, R. M. and D. J. Flournoy (1986), A case against mandatory lecture attendance, Journal of Medical Education, Vol. 61, pp. 175–176.
19. Jenne, F. H. (1973), Attendance and student proficiency change in a health science class, Journal of School Health, Vol. 43, pp. 135–126.
20. Jones, D. J. (2006), The Impact of Student Attendance , Socio-Economic Status and Mobility on Student Achievement of The Grade Students, Ph. D., Dissertation, Blacksburg, Virginia.
21. Kirby, A. and B. McElroy (2003), The Effect of Attendance on Grade for First Year Economics Students in University college Cork", The Economic and Social Reviev, Vol. 34, No. 3, pp. 311-326.

22. Kleinbaum, D., L. L. Kupper, K. E. Muller (1988), *Applied Regression Analysis and Other Multivariable Methods (*2nd edition*)*, PWS-Kent, Boston, MA.
23. Lamdin, D. J. (1996), Evidence of student attendance as an independent variable in education production functions. Journal of Educational Research, Vol. 89, No. 3, pp. 155–162.
24. Marburger, D. R. (2001), Absenteeism and Undergraduate Exam Performance, Journal of Economic Education, pp. 99-110.
25. McLoyd, V. (1998), Socio-economic disadvantage and child development, American Psychologist, Vol. 53, pp. 185–204.
26. Mitchell (1993), A comparison of achievement and attendance of fifth grade African American male and female students attending same-gender classes and coeducational classes in Polytechnic Institute and State University, doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.
27. Moore, R. (2003), Attendance and performance: How important is it for students to attend class? Journal of College Science Teaching, Vol. 32, pp. 367–371.
28. Ripple, C. H. and S. S. Luthar (2000), Academic risk among inner-city adolescents: The role of personal attributes, Journal of School Psychology, Vol. 38, No. 3, pp. 277–298.
29. Rocca, A. K. (2003), Student Attendance: A Comprehensive Literature Review, Journal on Excellence in College Teaching, Vol. 14, pp. 85-107.
30. Romer, D. (1993), Do students go to class? Should they? Journal of Economic Perspectives, Vol. 7, pp. 167–174.
31. Rothman, S. (2001), School absence and student background factors: A multilevel analysis, International Education Journal, Vol. 2, No. 1, pp. 59-68.
32. Sander, P., K. Stevenson, M. King and D. Coates (2000), University student's expectations of teaching, Studies in Higher Education, Vol. 25, No. 3, pp. 309-329.
33. Sexton M. (2003), A case study of the effect of year round education on attendance, academic performance, and behavior patterns, Ph.D. Dissertation, Blacksburg, Virginia.
34. Seyfried, S. F. (1998), Academic achievement of African American preadolescents: The influence of teacher perceptions, American Journal of Community Psychology, Vol. 26, No. 3, pp. 381–402.

35. Shimoff, E. and A.C. Catania (2001), Effects of recording attendance on grades in introductory psychology, Teaching Psychology, Vol. 28, No. 3, pp. 192-195.

36. Sirin, S. R. (2005), Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research, Review of Educational Research, Vol. 75, No. 3, pp. 417–453.

37. St. Clair, K. L. (1999), A case against compulsory class attendance policies in higher education, Innovative Higher Education, Vol. 23, pp. 171–180.

38. Stanca, L. (2006), The Effects of Attendance on Academic Performance: Panel data evidence from introductory microeconomics, Journal of Economic Education, Vol. 37, No. 3, pp. 251-266.

39. Suleiman, O., A. Bashir, W. A. Jadayil (2012), The Importance of Class Attendance and Cumulative GPA for Academic Success in Industrial Engineering Classes World Academy of Science, Engineering and Technology 61.

40. Sutton, A. and I. Soderstrom (1999), Predicting elementary and secondary school achievement with school-related and demographic factors. Journal of Educational Research, Vol. 92, No. 6, pp. 330–338.

41. U.S. Department of Education (1996), The problem of truancy in America's communities: Washington, DC: Government Printing Office.

42. Vanblerkon, M. L. (1992), Class attendance in undergraduate courses. The Journal of Psychology, Vol. 126, No. 5, pp. 487-494.

43. Wang, M. C., G. D. Haertel and H. J. Walberg (1993), Toward a knowledge base for school learning, Review of Educational Research, Vol. 63, No. 3, pp. 249–294.

44. White, K. (1982), The relation between socioeconomic status and academic achievement. Psychological Bulletin, Vol. 91, pp. 461–481.

45. Zamudio, G. (2004), Student mobility: the relationship between student population stability and academic achievement, doctoral dissertation, University of Arizona, Arizona.

46. Ziegler, C. W. (1972), School attendance as a factor in school progress (Rev. ed.), New York, NY: AMS Press, Inc.

Appendix A

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										