MULTIVARIATE ANALYSIS ON HND II STATISTICS STUDENTS SEMESTER ACADEMIC PERFORMANCE 2023/2024 ACADEMIC SESSION

(A CASE STUDY OF KWARA STATE POLYTECHNIC, ILORIN)

BY

ABDULLAHI, KAOSARAH OLAJUMOKE

HND/23/STA/FT/0034

BEING A RESEARCH PROJECT SUBMITTED TO THE DEPARTMENT OF STATISTICS,

INSTITUTE OF APPLIED SCIENCES,

KWARA STATE POLYTECHNIC, ILORIN.

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF HIGHER NATIONAL DIPLOMA (HND) IN STATISTICS

JULY, 2025

CERTIFICATION

I certify that this project was carried out by ABDULLAHI KAOSARAH OLAJUMOKE with matriculation number HND/23/STA/FT/0034 as meeting the requirement for the award of Higher National Diploma in the department of Statistics, Kwara State Polytechnic, Ilorin.

______________________________ ________________

DR. BEKI, D.O **Date**
Project Supervisor

______________________________ ________________

MISS. AJIBOYE, R.A **Date**
Project Co-ordinator

______________________________ ________________

MRS.ELEPO, T.A **Date**
Head of Department

______________________________ ________________

External Supervisor **Date**

DEDICATION

This project is dedicated to Almighty God who has seen me through from the beginning to the end of the course. And to my lovely parents who has always been supportive.

ACKNOWLEDGMENT

All thanks due to Almighty Allah, the owner and the creator of Universe for His divine protection and guidance throughout my life in school and this research project work.

First and foremost, I give thanks and I appreciate the efforts of my distinguished project supervisor, Dr. Beki D.O who made sure I acquired all the necessary information and Knowledge during the course of this project work may Allah reward him abundantly.

My gratitude goes to my beloved parents Mr and Mrs Abdullahi, I can't thank you enough for your parental advice and financial support since the beginning of this program, Jazaaakumllah Khayr. I pray Allah (SWT) spare their lives and make them reap the fruit of their labour.

My special thanks and appreciation goes to my siblings, who stood behind my back, in person of Alhaji Saheed, Mrs Adegoke, Sister Nasmat, Alfa Musbaudeen, Officer Abdulrahman, Toheer and Sofiyah. Thank you all for the love and supports.

I also give thanks to Mr. Babatunde Toheeb, thank you for the love and supports, I really do appreciate each and everything. Thank you all Jazakmullah Khayr.

This acknowledgement will be incomplete if I fail to appreciate all my friends Rofiat, Maryam Adebayo, Barakat, Hawau, Taiwo and my entire course mate in the department of Statistics 2025set. Thank you all. Jazaakullah Khayr. I pray Allah (SWT) guide and protect us till we reach the peak we're all eyeing.

Lastly, my sincere gratitude also goes to the Head of Statistics Department, in person of Mrs. Elepo T.A and all the lecturers of Statistics Department, Kwara State Polytechnic, Ilorin in person of Dr. (Mrs.) Aiyelabegan, A.B; Miss Ajiboye, R.A; Dr. Beki, D.O; Mr. Shafihi, F.G; Mr. Kuranga, J.O; Yusuf Saad; Mrs. Yusuf, G.A; Mrs. Abdulrahman; Mr Sulyman; and Mr Olarinoye, S.B for the good academic support given since the beginning of our program till point on time. I won't fail to appreciate DMLS-Rock Cyberworld located @Westend Market (+234 817365 1164) for the typeset of this project work.

ABSTRACT

This study examines the academic performance of HND II Statistics students of Kwara State Polytechnic, Ilorin, during the 2023/2024 academic session using multivariate analysis techniques. The primary objective was to determine whether significant differences exist in student performance across semesters and CGPA groupings. Data collected from semester results were analyzed using SPSS version 23, with statistical tools including descriptive statistics, paired samples t-tests, ANOVA, and Multivariate Analysis of Variance (MANOVA). The findings revealed that students classified under the 'distinction' CGPA group consistently outperformed their counterparts in the 'credit' and 'pass' categories in both semesters. A significant improvement in academic performance was observed between the first and second semesters ($p < 0.001$), indicating an overall upward trend. MANOVA results confirmed a strong effect of CGPA grouping on semester performance, with partial eta squared values ranging from 0.634 to 0.774, signifying substantial variance explained by CGPA classification. The study concludes that multivariate analysis is a robust method for evaluating academic performance and recommends its application for continuous academic assessment and policy development within tertiary institutions.

Keywords: Students, Grade, Results, Semester, MANOVA.

TABLE OF CONTENTS

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER ONE

1.0 Introduction

Mathematics has been viewed as one of the difficult subjects to learn and teach in elementary, junior, and senior high schools. Students' mathematics performance has been consistently low in comparison to different subjects within the curricula. From the national level down to the school level, different methods were utilized to address the factors that contribute to the decline of student's performance (Zacal, 2019). In the K-12 Statistics curriculum, one of the core subjects offered in the senior high school is statistics and probability. Elements of statistics and probability reasoning have emerged as requisite for a wide variety of applications especially in research in all fields of endeavors. Thus, without enough foundation of the essential ideas to understand data, readers are left confused. As mentioned in the study of Chiesi, Pirmi, and Morsanyi (2010), many students find it hard to understand statistics and probability concepts as documented in the different educational contexts and it is perhaps true to all secondary and tertiary students. For this reason, to better understand the underlying mechanism of statistics achievement, more attention has been paid to student's beliefs and feelings about statistics, and focused on the identification of models with non-cognitive factors such as beliefs and feeling about statistics (Zieffer, et.al., 2008 as cited by Chiesi, Pirmi, and Morsanyi, 2010). As mentioned by Batanero and Diaz (2010), statistics is becoming increasingly important in modern society; the relevance of developing statistical thinking in students across all levels of education has grown. Statistics is offered in secondary education. Unfortunately, many students fail to recognize its importance (Peters, Smith, Middledorp, Karpin, Sin, & Kilgore, 2013). In order for students to succeed and to use statistics, they should think that statistics is valuable in their lives (Emmioglu & Capa-Aydin, 2012). However, as mentioned by Peters et al., (2013), acquiring statistical skills and knowledge poses significant challenges for many students; a statistics course is challenging because it is abstract and requires logical reasoning, critical thinking, and the skills of interpretation and drawing conclusions. In addition, since teaching Statistics is integrative to secondary year levels, some students find it difficult to understand especially statistics (Salvan, 2014). Also, when it comes to rational number concepts and proportional reasoning, most of the students have an underlying difficulty necessary for calculating, reporting, and interpreting probabilities (Doyle et al., 2015). These are confirmed by Hansen and Myers (2012) also reported students to have low percentages of correct responses.

In addition, before the implementation of the current curriculum, students typically have an inadequate background in statistics before they enter college (Lemana, 2012). There was, however, the introduction of the concepts of statistics and probability have vigorously growing movement into the secondary school curriculum as basic literacy in statistics and probability, that mandates all citizens in today's world to have (Swift, 2012). Although the K-12 curriculum provides the concepts in statistics and probability and teaches these concepts from Grades 1 to 8 and in Grade 10, the depth at which learners absorb and fully grasp them may need reinforcement and consideration (Salvan, 2014). In the Philippine educational system, statistics and probability ideas frequently seem to struggle in students' experiences and how they perceive the world (Herrera, 2011). Most of the students have developed a distance of learning statistics and probability in the most concrete way. For this reason, Prado and Gravoso (2011) cautioned teachers to introduce topics not by abstractions but through activities and simulations. In the Division of Davao City, the integration of statistics and probability as one of the subjects taught in senior high school has brought concerns but also to the students as well. This concern has rooted in the fact that the majority of the teachers who are assigned to teach this subject had no formal training in statistics nd probability and some other factors contributing to the low students' performance include parents, teachers, and schools (Jamisola, 2014). Nevertheless, teachers with or without enough preparations are required to be competent and knowledgeable in presenting and developing the topics (Salvan, 2014). Moreover, Bakhshalizadeh, Pasha and Rastgar (2013) stated that mathematics has a close relationship with statistics in a manner and statistics is the scientific application of mathematical basics for the collection, analysis, and presentation of numerical data. Therefore, learning mathematics helps a lot to learn and succeed in statistics. Thus, different factors could influence Statistics performance in statistics. The study of Chapagain (2021) has identified socio-demographic and student-related factors such as type of school, type of local government, nature of examination, gender, age, and ethnicity that influence students' academic performance. In addition, parent's education, family income, and family size are some of the variables to be included. Students' attitudes toward statistics are prevalent. To this, Chinn (2012) supports those bad experiences in statistics specifically in learning statistics, math anxiety, and lack of support from adults may have caused the negative culture around Statistics. Furthermore, as mentioned by McGrath et al. (2015), to be successful in their statistics courses, students must possess high levels of self-efficacy. According to Bandura (1986), having high self-efficacy helps students to feel that they can develop the skills they need to master a given concept, even if they have to work through setbacks. These beliefs, in turn, prompt students to engage in effective studying and learning behaviors. Other researchers have similarly found self-efficacy to be a significant predictor of course performance in statistics (Byrne et al., 2014; McGrath et al., 2015). Similarly, as mentioned in the study of Huynh, Baglin, and Bedford (2014), in statistics education, there exists a consistent positive relationship between statistics attitudes and achievement. Improving student's attitudes towards statistics and statistics remains a challenging task with many students finding the subjects uninteresting, difficult, and tedious. According to Murray (2011), there are reasons for the decline in students taking statistics as subjects in high school. There is a need to improve student's attitudes towards quantitative subjects like statistics. Furthermore, improving student's attitudes in statistics in secondary school may lead to improved tertiary and career pathways into mathematical and statistical disciplines. Junior and senior high school students have been noted to have low performance in statistics and probability (Zacal, 2014). With this, it is very imperative to identify those factors leading to students' low performances in statistics. Thus, educational authorities adopt some policies improving the students' performance. Moreover, there is still a lack of studies investigating factors affecting students' selfefficacy, attitude change towards statistics, and performance in statistics, particularly at the secondary school level.

1.2 Statement of the Problems

Despite the growing emphasis on academic excellence in Nigerian tertiary institutions, student performance remains inconsistent across semesters. At the Kwara State Polytechnic, Ilorin, evidence suggests disparities in academic achievement among students in the Higher National Diploma (HND) II Statistics program. These discrepancies may stem from a range of cognitive and non-cognitive factors such as background knowledge, learning attitudes, self-efficacy, and instructional methods.

Previous studies have shown a significant influence of demographic and academic-related variables on student outcomes, yet limited research has applied multivariate statistical techniques to examine semester-wise performance within the same student group. The absence of such data-driven analysis hinders informed academic planning and targeted interventions. Therefore, it is crucial to investigate whether students' performance significantly varies between semesters and among different CGPA classifications using multivariate analysis.

1.3 Aim and Objectives

This project work aims to examine the multivariate analysis of students' semester results with respect to academic performance

Objectives are;

i. to determinants whether there is significant difference in the results between the semesters
ii. to determinants whether there is significant difference in the results within the semesters

1.4 Research Hypothesis

The study is guided by the following hypotheses:

H_{01}: There is no significant difference in academic performance between the first and second semesters among HND II Statistics students.

H_{02}: There is no significant difference in academic performance among students across different CGPA groupings (pass, credit, distinction).

H_{03}: There is no significant multivariate effect of CGPA grouping on academic performance across semesters.

1.5 Scope of the Study

This research is restricted to the academic performance of HND II Statistics students at the Department of Statistics, Kwara State Polytechnic, Ilorin, during the 2023/2024 academic session. It focuses specifically on the semester total marks (T.M), total credit points (TCP), and CGPA obtained across the first and second semesters. The study applies Multivariate Analysis of Variance (MANOVA) and related statistical techniques to identify patterns and differences in student performance.

1.6 Limitations of the Study

The study is limited to a single academic session (2023/2024) and does not account for longitudinal performance over multiple sessions. It excludes non-statistical departments, hence the findings may not be generalizable to students in other programs.

The data used are secondary, relying solely on recorded semester results, which may not fully reflect all the factors influencing academic performance such as attendance, class participation, or learning environments. Non-academic variables such as socio-economic status, teaching methods, and student motivation were not quantitatively assessed.

1.7 Definition of Terms

1. **CGPA (Cumulative Grade Point Average):** A numerical measure that reflects the overall academic performance of a student, computed as the average of total grade points earned over total credit units.

2. **MANOVA (Multivariate Analysis of Variance):** A statistical technique used to examine the effect of one or more independent variables on multiple dependent variables simultaneously.
3. **Semester Total Mark (T.M):** The aggregate of all scores obtained by a student across courses registered in a given semester.
4. **Credit Point:** A weighted score assigned based on the grade obtained and the unit of the course.
5. **Distinction/Credit/Pass:** Academic classifications of performance based on CGPA ranges, typically defined as:

- **Distinction:** CGPA ≥ 3.50
- **Credit:** CGPA between 2.50 and 3.49
- **Pass:** CGPA between 2.00 and 2.49

6. **Academic Performance:** The outcome of education—the extent to which a student has achieved their learning goals, often measured through grades, test scores, and CGPA.

CHAPTER TWO

2.0 Literature review

Teaching and learning in educational institutions from the nursery school to the University are assessed. Whatever is the system of education, the learning outcome is determined by assessing performance. The essence of certification in our daily life is not only of its usage to enter labour market but also for vertical and horizontal mobility. However, the most important fact tied to examination process is the need for study techniques which act as a tool for examination success. Assessment of the learning outcome is done for determining what has been learned and for decision-making with respect to selection of candidates for higher studies or job placement. It is

a great joy to achieve what one aims at. If we define achievement as having good results in anything, we put in more efforts so that we can be happy and receive commendation, then it became necessary to learn how to succeed, putting in place various ways from those who have used similar ways successfully. From a psychological point of view, Morrison and Macintyre (1993) defined academic success in terms of acquisition of different kinds of knowledge and cognitive skills. They went further to say that in some respects the multidimensional description would be theoretically most satisfactory. This approach is not all well applicable because evidence usually available is only on attainments on one or two aspects of school work or ‘average’ attainments with particular skills assessed not generally specified.

Pillow, (2008) has examined the gender differences among student on their academic performance has reveal that in individuals background characteristic affect his/her cognitive and non-cognitive is one of the most significant and influential characteristics in academic performance. Nori, (2002) studies the sex differences and the relationship between creativity and self-concept on academic performance among high school students.

There were 306 high school students (150 boys and 156 girls) to measure the rate of creativity questionnaire and cumulative grade point average Cumulative Grade Point Average (CGPA) the result were analyzed by cumulative grade point average Cumulative Grade Point Average (CGPA) for academic performance the analysis revealed that there was no significant relationship between creativity and self-concept on academic performance. In his study, the students were randomly selected from 68 schools (2,264 students, 38% where boys and 62% were girls). The academic performance of students was assessed using a self-reported achievement in some subject area English, Natural science, Mathematics, and Social science. A canonical correlation analysis found that when operationalized by their grade creativity and self-concept was related to academic achievement for both girls and boys. For girls, elaboration related to two of the academic subject

(social science and English language) and fluency related to natural science and mathematics. For boys, flexibility was the pre-dominant factor that related to all four (4) academic subject areas. When Operationalized thinking the Torrance creativity test (TTCT) Abedi-Schumacher creativity test (ACT) and villa and Auzmendi creativity test (VAT) on the other hand, creativity and self-concept was scarcely related to academic achievement. Yet, several other researchers also have attended to the idea that creativity is related to academic performance.

In Nigeria, education is emerging to be one of the biggest and largest industries, and the government continues to ensure that funds, instructional material and teaching personnel are made available for the sector.

Despite all the effort put in by the government and stakeholders of educational industry in Nigeria, the academic performance of Nigerian university students is still below expectation. Academic performance of a student is the extent to which he achieves specific academic goals. This is commonly measured by examinations or continuous assessment, but there is no general agreement on how it is best tested or which aspects are most important (Wikipedia 2017).

University education is mostly suited for providing the socioeconomic development that Nigeria yearns for. This is because, it is the development of the human capital that invariably leads to the development of other sectors of the economy. For this, effort has been on how to improve the qualities of the university education to ensure sustainable growth and development. In our effort to add to what other researchers have done toward ensuring quality university training, we decided to study the academic performance of students through numerical calculations.

Multivariate analysis of variance is a statistical technique used to determine if the categorical independent variable(s) with two or more levels affect the continuous independent variables (Ying Li et al., 2012). Aykut, Esra and Alperen (2014) carried out a research on the relationship between the academic achievement and performance assignment achievement scores of students in science courses with regard to different variables using MANOVA and correlation analysis and observed no significant difference between the grade levels and the students' academic achievement scores and performance scores whereas a significant difference was found between the gender variable and performance scores, which was in favour of females. Hussein, Gabriel and Adamu (2017) studied the influence of students' sex, age, and course of study on the performance of Senior High School students on mathematic course using MANOVA and discovered no significant difference in the performance of student across sex and age but significant difference across course of study. In this paper, we compared the performance of university students across the faculties using their cumulative grade point average (CGPA).

Academic performance of students may be adversely affected by many factors, some of which include poor location of the school, incessant changes in government policies, closure of schools, teachers strike action, me-school distance, inadequate supervision, monitoring, and evaluation machinery, lack of good textbooks, poor content and context of instructional materials, poor and non-conducive learning environment (Adepoju 1995 and Adepoju, 2003). Chansarkar and Mishaeloudis (2001) studied the effects of age, qualification, and distance from learning place etc.

on student performance. According to them the performance of students on the module is not affected by the factors like age, sex and place of residence, but is associated with qualification in quantitative subjects. They also found that those who live near the University perform better than other students. Yvonne and Kola, (1998) elaborated that the student performance is very much dependent on SEB (socio economic background). High school students' variation in the levels of the performance is linked to their gender, grade level, school location, school type, student type and socioeconomic background (SEB) they later commented. Learning preferences is the way by which an individual prefers to acquire and process different forms of information. In the account of Omrod (2008), some students seem to learn better when information is presented through words (verbal learners), whereas others seem to learn better when it is presented in the form of pictures

(visual learners). According to him, in a class where only one instructional method is employed, there is a strong possibility that a number of students will find the learning environment less optimal and this could affect their academic performance. Felder (1993) established that alignment between students' learning preferences and an instructor's teaching style leads to better recall and understanding. The learning preferences approach, according to him has gained significant mileage despite the lack of experimental evidence to support the utility of this approach. He stated further that there are a number of methods used to assess the learning preferences/styles of students but they all typically ask students to evaluate the kind of information presentation they are most at ease with. One of these approaches being used widely is the Visual/Aural/Read and Write/Kinesthetic (VARK) questionnaire, pioneered by Neil Flemming in 1987, which categorized

learners into a minimum of four major learning preference classes which includes: Visual learners: These are learners who process information better when it is visually displayed. They prefer information to be presented on the whiteboard or screen, with charts, graphs, diagrams, maps,

plans and colour. Aural (or oral)/auditory learners: These are learners who process information better when it is presented through discussions, stories, guest speakers, and chats. They do not like making a lot of notes and may prefer to record lectures for later playbacks and reference.

Read/write learners: These are learners who prefer information better when it is written down and are made available for reading. They write a lot of notes and text. Kinesthetic (or tactile) learners: These are learners who prefer practical exercises, examples, cases, trial and errors and use of senses in learning. They prefer to be actively involved in their learning and thus would benefit from active learning strategies in class (Flemming 2011). A number of studies have found positive effects of class attendance on academic performance of student. Lukkarinen, Koivukangas and Seppala (2016) investigated the relationship between university students' class attendance and learning performance using cluster and regression analyses and discovered that attendance is positively and significantly related to performance of students. Durden and Ellis, (1995) in their study reported a nonlinear effect of attendance on learning. According to them a few absences to class do not lead to poor grades but excessive absenteeism does. Newman-Ford, Lloyd & Thomas (2009) expressed a contrary view when they remarked that by the use of information technology, information that used to be obtained through lectures can be obtained at the click of a mouse.

According to them web-based learning approaches have become the order of the day. Other determinants of academics performance such as age and gender had been studied by Haist, et al

(2000), who observed that men perform better than women in certain settings while women outperform men in other settings. Borde (1998), on the other hand, found no evidence of academic performance being influenced by gender. Woodfield and Earl-Novell (2006) in study involving a close to two million graduating students found that female students outperformed male students and attributed this partly to female students being more conscientious and thus less likely to miss lectures. La Paro and Pianta (2000) and Crosser (1991) presented evidence that older children fare better academically than their younger age appropriate peers.

Okeke et al.(2018) suggests that students' performance across the three Faculties does not significantly differ. However, that does not mean that their performances are all up to expectation in the various departments across Faculties. Other measures can still be employed to enhance general academic performance of students.

Mildin and Marilyn (2021) in their paper drawn the following conclusion, first, since the level of student's self-efficacy beliefs and attitudes towards statistics is neither agree nor disagree, the students cannot make a clear-cut decision whether they are favorable or not with the statements on self-efficacy and the students have an uncertain opinion whether they dislike or like statistics. Level of student's performance in statistics is fairly satisfactory, this implies that most of the students possess the minimum knowledge and skills and core understanding in statistics but need help throughout the performance of tasks. Second, the variables such gender, family income, family size, parents' educational level, and SHS track preference do not affect the students' self-efficacy, attitudes toward mathematics, and their performance in mathematics, while the type of school does. Third, since the type of school is a key determinant of student's self-efficacy, attitudes towards statistics and performance in statistics, thus, class size and facilities, as well as other factors that substantially differentiate private from public schools, significantly affect the dependent variables. Lastly, due to the limited studies in Mindanao, future studies may take into consideration to include the schools in the rural areas in the sampling to provide and establish strong statistical analysis and solid evidence of the factors that influence the determinants significantly affect the performance of senior high school students in statistics. Further research should be done in order to confirm or refute the findings of this study.

CHAPTER THREE

3.0 Methodology

This chapter gives details on the various methods of data collection and the various methods applied in the data analysis.

3.1 Sources of Data

The data used for this project is secondary data based on HND II statistics graduating students' results for the 2023/2024 session.

3.2 Statistical methods

3.2.1 Multivariate Analysis of Variance (MANOVA) Basic Concepts

Univariate case

One-way ANOVA investigates the effects of a categorical variable (the groups, i.e. independent variables) on a continuous outcome (the dependent variable). In one-way ANOVA, we have m random variables $x_1, \dots, x_m$ (also called groups or treatments). For each group, we have a sample, where we denote the jth group sample as $\{x_{1j}, \dots, x_{n_j j}\}$. Group j is said to have n_j subjects in its sample. We also define $n = \sum_{j=1}^{m} n_j$.

The objective is to test the null hypothesis H_0: $\mu_1 = \mu_2 = \cdots = \mu_m$.

We use the following definitions for the total (T), between groups (B) and within groups (W) sum of squares (SS), degrees of freedom (df) and mean square (MS):

	df	SS	MS
T	$n-1$	$\sum_j \sum_i (x_{ij}-\bar{x})^2$	SS_T/df_T
B	$k-1$	$\sum_j n_j(\bar{x}_j-\bar{x})^2$	SS_B/df_B
W	$n-k$	$\sum_j \sum_i (x_{ij}-\bar{x}_j)^2$	SS_W/df_W

The test statistic F is defined as follows and has an F distribution with df_B, df_W degrees of freedom:

$$F = \frac{MS_B}{MS_W} \sim F(df_B, df_W)$$

We reject the null hypothesis if $F > F_{crit}$.

Multivariate case

MANOVA also investigates the effects of a categorical variable (the groups, i.e. independent variables) on a continuous outcome, but in this case, the outcome is represented by a vector of dependent variables.

We could simply perform multiple ANOVA's, one for each dependent variable, but this would have two disadvantages: it would introduce additional experiment-wise error and it would not account for the correlations between the dependent variables. It is, therefore, possible that MANOVA shows a significant difference between the means while the individual ANOVA do not.

Also MANOVA can be used in place of ANOVA with repeated measures; in which case no sphericity assumption needs to be met when using MANOVA. In this case, you treat the repeated levels as dependent variables.

Definition 1: In One-way MANOVA, we have m random vectors X_1, …, X_m (representing **groups** or **treatments**). Each X_j is a $k \times 1$ column vector of form

$$\begin{bmatrix} x_{j1} \\ \cdots \\ x_{jk} \end{bmatrix}$$

where each x_{jp} is a random variable.

For each random vector X_j we collect a sample $\{X_{1j}, \ldots, X_{n_j j}\}$ of size n_j. We also define $n = \sum_{j=1}^{m} n_j$. Each sample X_{ij} is a $k \times 1$ vector of form

$$\begin{bmatrix} x_{ij1} \\ \cdots \\ x_{ijk} \end{bmatrix}$$

where each x_{ijp} is a data element (not a random variable), where index i refers to the **subject** in the experiment ($1 \leq i \leq n_j$), index j refers to the group ($1 \leq j \leq m$) and index p refers to the position (i.e. dependent variable) within the random vector ($1 \leq p \leq k$).

Our objective is to test the **null hypothesis** H_0: $\mu_1 = \mu_2 = \cdots = \mu_m$ where the μ_j are vectors

$$\begin{bmatrix} \mu_{j1} \\ \cdots \\ \mu_{jk} \end{bmatrix}$$

and so the null hypothesis is equivalent to H_0: $\mu_{1p} = \mu_{2p} = \cdots = \mu_{mp}$ for all p such that $1 \leq p \leq k$. The alternative hypothesis is, therefore, H_1: $\mu_r \neq \mu_j$ for some r, j such that $1 \leq r, j \leq m$, or equivalently, $\mu_{rp} \neq \mu_{jp}$ for some r, j, p such that $1 \leq r, j \leq m$ and $1 \leq p \leq k$.

Now we define the various means as in the univariate case, except that now these means become $k \times 1$ vectors. The **total** (or **grand**) **mean vector** is the column vector

$$\bar{X}_T = \begin{bmatrix} \bar{x}_1 \\ \cdots \\ \bar{x}_k \end{bmatrix}$$

where $\bar{x}_p = \frac{1}{n} \sum_{j=1}^{m} \sum_{i=1}^{n_j} x_{ijp}$

The **sample group mean vector** for group j is a column vector

$$\bar{X}_j = \begin{bmatrix} \bar{x}_{j1} \\ \cdots \\ \bar{x}_{jk} \end{bmatrix} \text{ where } \bar{x}_{jp} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ijp}$$

Definition 2: Using the terminology from Definition 1, we define the following **total cross products** for *p* and *q*.

$$CP_{pq} = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(x_{ijp} - \bar{x}_p\right)\left(x_{ijq} - \bar{x}_q\right)$$

When $p = q$, we have

$$CP_{pp} = S_p = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(x_{ijp} - \bar{x}_p\right)^2$$

which is the total sum of squares (as in ANOVA) and measures the total variation in the *p*th dependent variable. When $p \neq q$, we have the total cross-product terms, which measure the dependence between the *p*th and *q*th variables across all observations.

The multivariate equivalent of the total sum of squares is the **total sum of squares and cross products**, i.e. the $SSCP_T$ matrix, which is abbreviated a *T*, and is defined as

$$T = \begin{bmatrix} CP_{11} & \cdots & CP_{1k} \\ \cdots & \cdots & \cdots \\ CP_{k1} & \cdots & CP_{kk} \end{bmatrix}$$

Note that the diagonal terms are SS_1, ..., SS_k. An alternative way of expressing T is as follows:

$$T = \sum_{j=1}^{m} \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_T)(X_{ij} - \bar{X}_T)^T$$

The sample covariance matrix plays the role of MS_T since $MS_T = SSCP_T / df_T$ where the degrees of freedom is given by $df_T = n - 1$.

Definition 3: We define the **hypothesis cross products** for p and q as follows:

$$CP_{pq} = \sum_{j=1}^{m} n_j(\bar{x}_{jp} - \bar{x}_p)(\bar{x}_{jq} - \bar{x}_q)$$

We define the **hypothesis sum of squares and cross products** as the matrix H where

$$H = \begin{bmatrix} CP_{11} & \dots & CP_{1k} \\ \dots & \dots & \dots \\ CP_{k1} & \dots & CP_{kk} \end{bmatrix}$$

Alternatively, H can be defined as

$$H = \sum_{j=1}^{m} n_j(\bar{X}_j - \bar{X}_T)(\bar{X}_j - \bar{X}_T)^T$$

Definition 4: We define the **error** (or **residual**) **cross products** for p and q as follows:

$$CP_{pq} = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(x_{ijp} - \bar{x}_{jp}\right)\left(x_{ijq} - \bar{x}_{jq}\right)$$

We define the **error** (or **residual**) **sum of squares and cross products** as the matrix E where

$$E = \begin{bmatrix} CP_{11} & \cdots & CP_{1k} \\ \cdots & \cdots & \cdots \\ CP_{k1} & \cdots & CP_{kk} \end{bmatrix}$$

Alternatively, E can be defined as

$$E = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_j\right)\left(X_{ij} - \bar{X}_j\right)^T$$

Property 1: $T = H + E$

Also for any p and q, $CP_T = CP_H + CP_E$

<u>Proof</u>:

$$T = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_T\right)\left(X_{ij} - \bar{X}_T\right)^T$$

$$= \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left[(X_{ij}-\bar{X}_j)+(\bar{X}_j-\bar{X}_T)\right]\left[(X_{ij}-\bar{X}_j)+(\bar{X}_j-\bar{X}_T)\right]^T$$

$$= \sum_{j=1}^{m} n_j(\bar{X}_j-\bar{X}_T)(\bar{X}_j-\bar{X}_T)^T + \sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_j)(X_{ij}-\bar{X}_j)^T = H + E$$

Observation: The three **sum of squares and cross product** (**SSCP**) terms play the role of the *SS* in ANOVA. The degrees of freedom terms are $df_T = n - 1$, $df_H = m - 1$, $df_E = n - m$. As usual $df_T = df_H + df_E$. How these terms are used to create the appropriate F test is more complicated than in ANOVA. We will look at this shortly.

Definition 5:

Wilk's Lambda: $\Lambda = \frac{|E|}{|H+E|}$

H is large compared to E when the numerator of the above is small compared to the denominator. Thus we reject the null hypothesis when Wilk's Lambda is close to zero.

Hotelling-Lawley Trace: T_0^2 = trace(HE^{-1})

H is large compared to E when Hotelling-Lawley Trace is large. In this case, we reject the null hypothesis.

Pillai-Bartlett Trace: V = trace($H(H+E)^{-1}$)

If H is large compared to E then this statistic will be large. Thus, we reject the null hypothesis when this value is large.

Roy's Largest Root: Θ = largest eigenvalue of HE^{-1}

Again we reject the null hypothesis if this statistic is large. The following alternative version of Roy's Largest Root is also sometimes used:

$\frac{\lambda_p}{1+\lambda_p}$ where λ_p largest eigenvalue of HE^{-1}

Property 2: $\Lambda = \frac{1}{|I+HE^{-1}|}$

Proof: Note that E and H are both symmetric matrices, and so E^{-1} is also symmetric. This means that $HE^{-1} = E^{-1}H$. But then $E(I+HE^{-1}) = E(I+E^{-1}H) = EI + EE^{-1}H = E + H$. By Property 2 of Determinants and Simultaneous Linear Equations, the determinant of a product of two square matrices is equal to the product of the determinants of each matrix, and so $|E| \cdot |I + HE^{-1}| = |E(I+HE^{-1})| = |E + H|$. Thus

$$\Lambda = \frac{|E|}{|H+E|} = \frac{1}{|I+HE^{-1}|}$$

Wilk's Lambda: $\Lambda = \prod_{p=1}^{k} \frac{1}{1+\lambda_p}$

Hotelling-Lawley Trace: $T_0^2 = \sum_{p=1}^{k} \lambda_p$

Pillai-Bartlett Trace: $V = \sum_{p=1}^{k} \frac{\lambda_p}{1+\lambda_p}$

The Pillai-Barlett Trace is similar to SS_B/SS_T, which is the percentage of the variance explained by the model, which is similar to R^2. It is the most conservative of the tests but is most robust in cases of violation of the assumptions at least for balanced models.

Wilk's Lambda is similar to $SS_E\ /SS_T = 1 - R^2$. It is the most commonly used of the tests.

The Hotelling-Lawley Trace is similar to SS_B/SS_E which is the F-test used in univariate ANOVA.

Property 4 (Wilk's Lambda Test)

Let

$$a = n - m - \frac{k - m + 2}{2}$$

$$b = \begin{cases} \sqrt{\dfrac{k^2(m-1)^2 - 4}{k^2 + (m-1)^2 - 5}}, & k^2 + (m-1)^2 - 5 > 0 \\ 1, & \text{otherwise} \end{cases} \qquad c = \frac{k(m-1) - 2}{2}$$

If the null hypothesis is true, then

$$df_1 = k(m-1) \qquad df_2 = ab - c \qquad F = \frac{1-\Lambda^{1/b}}{\Lambda^{1/b}} \cdot \frac{df_2}{df_1} \sim F(df_1, df_2)$$

Property 5 (Hotelling-Lawley Trace Test)

Let s = min(k, m -1) = # of non-zero eigenvalues in HE^{-1}

$$t = \frac{|k-m+1|-1}{2} \qquad u = \frac{n-m-k-1}{2}$$

If the null hypothesis is true, then

$$df_1 = s(2t+s+1) \qquad df_2 = 2(su+1) \qquad F = \frac{T_0^2}{s} \cdot \frac{df_2}{df_1} \sim F(df_1, df_2)$$

Note that $df_1 = s \cdot \max(k, m-1)$

Property 6 (Pillai-Barlett Trace Test)

Under the same assumption as Property 5

$$df_1 = s(2t+s+1) \qquad df_2 = s(2u+s+1) \qquad F = \frac{V}{s-V} \cdot \frac{df_2}{df_1} \sim F(df_1, df_2)$$

Two-way MANOVA

Two-way MANOVA can be considered to be an extension of one-way MANOVA to support two factors and their interaction or as an extension to two-way ANOVA to support multiple dependent variables.

Univariate case

Two-way ANOVA investigates the effects of two categorical variables on a continuous outcome (the dependent variable). In two-way ANOVA, we have *r* random variables (the levels) for Factor A (the row factor) and *c* random variables for Factor B (the column factor). For each interaction between the factor levels, we have a sample with *m* elements (we restrict ourselves to the balanced model here). We can represent the sample for the i^{th} level from Factor A and the j^{th} level from Factor B by $X_{ij} = \{x_{ij1}, \ldots, x_{ijm}\}$.

Our objective is to test three types of null hypotheses:

H_0: $\mu_{1.} = \mu_{2.} = \ldots = \mu_{r.}$ (Factor A)

H_0: $\mu_{.1} = \mu_{.2} = \ldots = \mu_{.c}$ (Factor B)

H_0: $\delta_{ij} = 0$ for all *i, j* (Interaction between A and B, where the δ_{ij} are the interaction effects

We define the following where $n = mrc$.

$SS_T = \sum_k \sum_j \sum_i (x_{ijk} - \bar{x})^2$	$df_T = n - 1$	$MS_T = SS_T / df_T$
$SS_A = mc \sum_i (\bar{x}_i - \bar{x})^2$	$df_A = r - 1$	$MS_A = SS_A / df_A$
$SS_B = mr \sum_j (\bar{x}_j - \bar{x})^2$	$df_B = c - 1$	$MS_B = SS_B / df_B$
$SS_{AB} = m \sum_j \sum_i (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$	$df_{AB} = (r-1)(c-1)$	$MS_{AB} = SS_{AB} / df_{AB}$
$SS_W = \sum_k \sum_j \sum_i (x_{ijk} - \bar{x}_{ij})^2$	$df_W = n - rc$	$MS_W = SS_W / df_W$

The test statistics F are defined as follows:

$$F_A = \frac{MS_A}{MS_W} \sim F(df_A, df_W) \qquad F_B = \frac{MS_B}{MS_W} \sim F(df_B, df_W) \qquad F_{AB} = \frac{MS_{AB}}{MS_W} \sim F(df_{AB}, df_W)$$

For each of the three null hypotheses, we reject the null hypothesis if $F > F_{crit}$.

Multivariate case

As for one factor MANOVA, two-factor MANOVA is similar to two-factor ANOVA except that in place of simple variables (for each factor) we have random vectors and in place of sample data $\{x_{ij1}, \ldots, x_{ijm}\}$ for the i^{th} level in Factor A and the j^{th} level in Factor

B, we have sample data $\{X_{ij1}, \ldots, X_{ijm}\}$ where each X_{ijk} is a $p \times 1$ column vector of data elements of the form x_{ijkh}. Here, there are p dependent variables.

The additive model takes the form:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$$

This is exactly as for a two-factor ANOVA model except that x_{ijk} is replaced by X_{ijk} and the other terms are column vectors instead of scalars. With this change, the null hypotheses are as for ANOVA. In place of the ANOVA normality assumption, assume multivariate normality, namely

$$\varepsilon_{ijk} \sim N(0, \Sigma)$$

The df terms are as for ANOVA, as defined above, and the SS terms are defined in terms of the following $p \times p$ cross-product matrices.

$$SS_A = mc \sum_i (\bar{x}_{i.} - \bar{x})(\bar{x}_{i.} - \bar{x})^T$$

$$SS_B = mr \sum_i (\bar{x}_{.j} - \bar{x})(\bar{x}_{.j} - \bar{x})^T$$

$$SS_W = \sum_k \sum_j \sum_i (x_{ijk} - \bar{x}_{ij})(x_{ijk} - \bar{x}_{ij})^T$$

$$SS_{AB} = SS_T - SS_A - SS_B - SS_W$$

$$SS_T = \sum_k \sum_j \sum_i (x_{ijk} - \bar{x})(x_{ijk} - \bar{x})^T$$

Recall that each of these *SS* matrices takes the form

$$SS = \begin{bmatrix} CP_{11} & \dots & CP_{1k} \\ \dots & \dots & \dots \\ CP_{k1} & \dots & CP_{kk} \end{bmatrix}$$

where for SS_A

$$CP_{pq} = mc \sum_{i=1}^{r} (\bar{x}_{jp} - \bar{x}_p)(\bar{x}_{jq} - \bar{x}_q)$$

and similarly for the other *SS* matrices.

Wilk's Lambda

Each of the three null hypotheses is tested using the appropriate Wilk's Lambda.

$$\Lambda_A = \frac{|SS_W|}{|SS_A + SS_W|} \qquad \Lambda_B = \frac{|SS_W|}{|SS_B + SS_W|} \qquad \Lambda_{AB} = \frac{|SS_W|}{|SS_{AB} + SS_W|}$$

If the Factor *H* null hypothesis is true (for *H* = *A*, *B* or *AB*) then

$$\left[\frac{p + 1 - df_H}{2} - df_W\right] \ln \Lambda_H \sim \chi^2\,(p \cdot df_H)$$

As for one factor MANOVA, a more accurate test is based on the F test, described as follows:

$$F_H = \frac{1-\Lambda_H^{1/b}}{\Lambda_H^{1/b}} \cdot \frac{df_2}{df_1} \sim F(df_1, df_2)$$

$$df_1 = p \cdot df_H \qquad df_2 = ab - c$$

where

$$a = df_H + df_W - \frac{p + df_H + 1}{2} = df_W + \frac{df_H - p - 1}{2} b = \sqrt{\frac{p^2 df_H^2 - 4}{p^2 + df_H^2 - 5}}$$

$$c = \frac{p \cdot df_H}{2} - 1$$

If either the numerator or denominator of *b* equals 0 or results in a negative value inside the square root, then *b* takes the value of 1.

Hotelling-Lawley Trace

Each of the three null hypotheses is tested using the appropriate Hotelling-Lawley Trace test.

$$F = \frac{T_0^2}{s} \cdot \frac{df_2}{df_1} \sim F(df_1, df_2)$$

$$T_0^2 = \text{trace}(HE^{-1})\quad t = \frac{|p - df_H| - 1}{2} \qquad u = \frac{df_E - p - 1}{2} \qquad s = \min(p, df_H)$$

where

$$df_1 = s(2t + s + 1) \qquad df_2 = 2(su + 1)$$

Pillai-Bartlett Trace

Each of the three null hypotheses is tested using the appropriate Pillai-Bartlett Trace test.

$$F = \frac{V}{s - V} \cdot \frac{df_2}{df_1} \sim F(df_1, df_2)$$

$$V = \text{trace}(H(H+E)^{-1}\quad t = \frac{|p - df_H| - 1}{2} \qquad u = \frac{df_E - p - 1}{2} \qquad s = \min(p, df_H)$$

where

$$df_1 = s(2t + s + 1) \qquad df_2 = s(2u + s + 1)$$

Roy's Largest Root

Each of the three null hypotheses is tested using the appropriate Roy's Largest Root test.

$$F = \varphi \cdot \frac{df_2}{df_1} \sim F(df_1, df_2)$$

$$\varphi = \text{largest eigenvalue of } HE^{-1}$$

where

$$df_1 = 2(t+1) \qquad df_2 = 2(u+1)$$

and *s, t, u* are as for the Pillai-Bartlett Trace test.

3.2.2 Assumption of MANOVA

i. One of the assumptions of MANOVA is that the response variables come from group populations that are multivariate normal distributed. This means that each of the dependent variables is normally distributed within group, that any linear combination of the dependent variables is normally distributed, and that all subsets of the variables must be multivariate normal. With respect to Type I error rate, MANOVA tends to be robust to minor violations of the multivariate normality assumption.

ii. The homogeneity of population covariance matrices (a.k.a. sphericity) is another assumption. This implies that the population variances and covariances of all dependent variables must be equal in all groups formed by the independent variables.

iii. Small samples can have low power, but if the multivariate normality assumption is met, the MANOVA is generally more powerful than separate univariate tests.

3.3 Data Presentation

	INSTITUTE OF APPLIED SCIENCES						
	DEPARTMENT OF STATISTICS						
	HNDII FT STATISTICS						
	2023/2024 SESSIONAL RESULT						
S NO	1 stSEM T.M	2 ndSEM T.M	TCP	TCF	CHF %	C.T.P 62	CGPA
	31	31				62	
1	78.5	90.5	23	0	0	169	2.73
2	80.75	94.75	23	0	0	175.5	2.84
3	95.5	90.75	23	0	0	186.25	3
4	69.5	78	23	0	0	147.5	2.39
5	78.25	85	23	0	0	163.25	2.63
6	74	82	23	0	0	156	2.52
7	115	107.75	23	0	0	222.75	3.6
8	105.75	103.5	23	0	0	209.25	3.37
9	105.75	106.75	23	0	0	212.5	3.44
10	77	93.75	23	0	0	170.75	2.76
11	114.75	121	23	0	0	235.75	3.81
12	104	98.25	23	0	0	202.25	3.26
13	90.5	93.25	23	0	0	183.75	2.97
14	80	87.5	23	0	0	167.5	2.71
15	93	95.25	23	0	0	188.25	3.03
16	105.5	110.5	23	0	0	216	3.48
17	75.75	88.75	23	0	0	164.5	2.66
18	76.25	92.25	23	0	0	168.5	2.73
19	95	96.75	23	0	0	191.75	3.1
20	101	106	23	0	0	207	3.34
21	83	88.75	23	0	0	171.75	2.77
22	95	96.5	23	0	0	191.5	3.1
23	81.5	89.75	23	0	0	171.25	2.76
24	76.5	88.5	23	0	0	165	2.66
25	69.25	83.75	23	0	0	153	2.47
26	92	97	23	0	0	189	3.05
27	81	88.75	23	0	0	169.75	2.74
28	105.5	100.25	23	0	0	205.75	3.32
29	81.75	87.25	23	0	0	169	2.73
30	94.5	95	23	0	0	189.5	3.06
31	109.75	118.25	23	0	0	228	3.68
32	85.5	86.5	23	0	0	172	2.77

33	80	82	23	0	0	162	2.61
34	74.5	85	23	0	0	159.5	2.58
35	80	86	23	0	0	166	2.68
36	89.75	91.75	23	0	0	181.5	2.94
37	102.75	99	23	0	0	201.75	3.26
38	86.25	95.25	23	0	0	181.5	2.94
39	99.5	109	23	0	0	208.5	3.37
40	82	92.75	23	0	0	174.75	2.82
41	74	83	23	0	0	157	2.53
42	82.5	89.25	23	0	0	171.75	2.77
43	79.75	84	23	0	0	163.75	2.65
44	78.25	86.5	23	0	0	164.75	2.66
45	88	89.25	23	0	0	177.25	2.85
46	80.75	81.75	23	0	0	162.5	2.63
47	87.75	82.5	23	0	0	170.25	2.74
48	92	86	23	0	0	178	2.87
49	85	84.75	23	0	0	169.75	2.74
50	73.75	75.25	23	0	0	149	2.4
51	69.75	90.5	23	0	0	160.25	2.58
52	78.5	88	23	0	0	166.5	2.69
53	76.75	88.25	23	0	0	165	2.66
54	86.5	88.5	23	0	0	175	2.82
55	92.25	93.75	23	0	0	186	3
56	79	86.5	23	0	0	165.5	2.68
57	79	97	23	0	0	176	2.84
58	80.75	91.75	23	0	0	172.5	2.79
59	76.5	72.5	23	0	0	149	2.4
60	82.5	89.75	23	0	0	172.25	2.77
61	71.5	80.75	23	0	0	152.25	2.45
62	83	96.25	23	0	0	179.25	2.89
63	91.75	89	23	0	0	180.75	2.92
64	82.75	89.5	23	0	0	172.25	2.77
65	94.25	94	23	0	0	188.25	3.03
66	82.75	95	23	0	0	177.75	2.87
67	81	88.25	23	0	0	169.25	2.73
68	87.5	86.5	23	0	0	174	2.81
69	92.5	90.5	23	0	0	183	2.95
70	88	90	23	0	0	178	2.87
71	88.25	100.75	23	0	0	189	3.05
72	82	84.75	23	0	0	166.75	2.69

73	75.75	89.75	23	0	0	165.5	2.68
74	83.25	87.25	23	0	0	170.5	2.76
75	72.25	86.25	23	0	0	158.5	2.56
76	91.25	0	12	11	50	91.25	1.47
77	84.75	88.75	23	0	0	173.5	2.81
78	86.75	90.75	23	0	0	177.5	2.87
79	92.75	93.5	23	0	0	186.25	3
80	90.25	85.5	23	0	0	175.75	2.84
81	74.75	82.5	23	0	0	157.25	2.53
82	79.25	85.75	23	0	0	165	2.66
83	80	84.5	23	0	0	164.5	2.66
84	81.5	89.25	23	0	0	170.75	2.76
85	85.25	93.5	23	0	0	178.75	2.89
86	88.25	89.5	23	0	0	177.75	2.87
87	87.75	85	23	0	0	172.75	2.79
88	83.75	85.75	23	0	0	169.5	2.74
89	78	83	23	0	0	161	2.6
90	83.75	86.5	23	0	0	170.25	2.74
91	85.75	92	23	0	0	177.75	2.87
92	82.5	87	23	0	0	169.5	2.74
93	76	80.5	23	0	0	156.5	2.53
94	82	85.75	23	0	0	167.75	2.71
95	94	104.25	23	0	0	198.25	3.19
96	85.5	85.5	23	0	0	171	2.76
97	79.5	86.5	23	0	0	166	2.68
98	72.25	91.25	23	0	0	163.5	2.65
99	77.25	84.25	23	0	0	161.5	2.61
100	89.5	93	23	0	0	182.5	2.95
101	81.75	87.75	23	0	0	169.5	2.74
102	70	86.25	23	0	0	156.25	2.52
103	77.5	83	23	0	0	160.5	2.6
104	80.5	92.25	23	0	0	172.75	2.79
105	83.25	88.5	23	0	0	171.75	2.77
106	79.75	84.75	23	0	0	164.5	2.66
107	85.25	87	23	0	0	172.25	2.77
108	92	101.25	23	0	0	193.25	3.11
109	83.25	96.5	23	0	0	179.75	2.9
110	77.75	85.75	23	0	0	163.5	2.65
111	91.5	94.75	23	0	0	186.25	3
112	89	90.75	23	0	0	179.75	2.9

113	71.25	87.25	23	0	0	158.5	2.56
114	81.5	91	23	0	0	172.5	2.79
115	81.75	92.25	23	0	0	174	2.81
116	83.75	86.25	23	0	0	170	2.74
117	77	88.75	23	0	0	165.75	2.68
118	80.75	94.75	23	0	0	175.5	2.84
119	76	94	23	0	0	170	2.74
120	68.75	81.75	23	0	0	150.5	2.44
121	77.25	88.5	23	0	0	165.75	2.68
122	72.5	74.5	23	0	0	147	2.37
123	68.5	104.25	21	2	8	172.75	2.79
124	81	86.25	23	0	0	167.25	2.69
125	80.75	87.25	23	0	0	168	2.71
126	74.75	82	23	0	0	156.75	2.53
127	79.25	95.5	23	0	0	174.75	2.82
128	87	88.5	23	0	0	175.5	2.84
129	79	86.5	23	0	0	165.5	2.68
130	69	45.5	20	3	15	114.5	1.85
131	74.25	81.5	23	0	0	155.75	2.52
132	84.25	89.25	23	0	0	173.5	2.81
133	74.5	90.75	23	0	0	165.25	2.66
134	78.25	90	23	0	0	168.25	2.71
135	69.75	81.25	23	0	0	151	2.44
136	67.5	89.75	23	0	0	157.25	2.53
137	81.5	80.5	23	0	0	162	2.61
138	87.5	83.25	23	0	0	170.75	2.76
139	76	83.25	23	0	0	159.25	2.56
140	78.25	88.75	23	0	0	167	2.69
141	69	72.25	23	0	0	141.25	2.27
142	69	97.25	23	0	0	166.25	2.68
143	69.25	88.5	23	0	0	157.75	2.55
144	73.25	91.25	23	0	0	164.5	2.66
145	73.25	83.25	23	0	0	156.5	2.53
146	77.5	85.5	23	0	0	163	2.63

CHAPTER FOUR

4.0 ANALYSIS OF DATA

This chapter deals with data presentation, description, analysis and results. The analysis was carried out using Statistical Package for Social Sciences (SPSS) v.23.

4.1 Data presentation

The data used for this project work can be found in the data presentation in Chapter Three

4.2 Data Analysis and Results

Table 4.1 shows the descriptive statistics

Descriptive Statistics

	TCF	CHF	CGPA_GROUP1	Mean	Std. Deviation	N
1st SEM T.M	.00	.00	Pass	71.3437	2.77082	8
			Credit	80.5698	5.45489	111
			distinction	98.4565	7.02659	23
			44.00	114.7500	.	1
			Total	83.1696	9.44007	143
		Total	Pass	71.3437	2.77082	8
			Credit	80.5698	5.45489	111
			distinction	98.4565	7.02659	23
			44.00	114.7500	.	1
			Total	83.1696	9.44007	143
	2.00	8.00	Credit	68.5000	.	1
			Total	68.5000	.	1
		Total	Credit	68.5000	.	1
			Total	68.5000	.	1
	3.00	15.00	Fail	69.0000	.	1
			Total	69.0000	.	1

			Total	Fail	69.0000	.	1
				Total	69.0000	.	1
		11.00	50.00	Fail	91.2500	.	1
				Total	91.2500	.	1
			Total	Fail	91.2500	.	1
				Total	91.2500	.	1
		Total	.00	Pass	71.3437	2.77082	8
				Credit	80.5698	5.45489	111
				distinction	98.4565	7.02659	23
				44.00	114.7500	.	1
				Total	83.1696	9.44007	143
			8.00	Credit	68.5000	.	1
				Total	68.5000	.	1
			15.00	Fail	69.0000	.	1
				Total	69.0000	.	1
			50.00	Fail	91.2500	.	1
				Total	91.2500	.	1
			Total	Fail	80.1250	15.73313	2
				Pass	71.3437	2.77082	8
				Credit	80.4621	5.54874	112
				distinction	98.4565	7.02659	23
				44.00	114.7500	.	1
				Total	83.0274	9.51682	146
2nd T.M	SEM	.00	.00	Pass	77.3438	4.38125	8
				Credit	88.1216	3.89243	111
				distinction	100.5543	6.75965	23
				44.00	121.0000	.	1
				Total	89.7483	7.50273	143
			Total	Pass	77.3438	4.38125	8
				Credit	88.1216	3.89243	111
				distinction	100.5543	6.75965	23
				44.00	121.0000	.	1

		Total	89.7483	7.50273	143
2.00	8.00	Credit	104.2500	.	1
		Total	104.2500	.	1
	Total	Credit	104.2500	.	1
		Total	104.2500	.	1
3.00	15.00	Fail	45.5000	.	1
		Total	45.5000	.	1
	Total	Fail	45.5000	.	1
		Total	45.5000	.	1
11.00	50.00	Fail	.0000	.	1
		Total	.0000	.	1
	Total	Fail	.0000	.	1
		Total	.0000	.	1
Total	.00	Pass	77.3438	4.38125	8
		Credit	88.1216	3.89243	111
		distinction	100.5543	6.75965	23
		44.00	121.0000	.	1
		Total	89.7483	7.50273	143
	8.00	Credit	104.2500	.	1
		Total	104.2500	.	1
	15.00	Fail	45.5000	.	1
		Total	45.5000	.	1
	50.00	Fail	.0000	.	1
		Total	.0000	.	1
	Total	Fail	22.7500	32.17336	2
		Pass	77.3438	4.38125	8
		Credit	88.2656	4.16378	112
		distinction	100.5543	6.75965	23
		44.00	121.0000	.	1
		Total	88.9298	11.17833	146

Interpretation: This table presents the mean, standard deviation, and sample size (N) of students' performance across CGPA groups in the first and second semesters. It shows that 'distinction' students performed best in both semesters, followed by 'credit' and 'pass' students. The mean score increased from 83.03 in the first semester to 88.93 in the second semester, indicating overall academic improvement.

Table 4.2: Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	1st SEM T.M	83.0274	146	9.51682	.78762
	2nd SEM T.M	88.9298	146	11.17833	.92512

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	1st SEM T.M & 2nd SEM T.M	146	.461	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	1st SEM T.M - 2nd SEM T.M	-5.90240	10.83284	.89653	-7.67436	-4.13044	-6.584	145	.000

Interpretation: The paired samples statistics compared students' average performance across semesters. A significant increase in mean score (mean difference = -5.90, $p < .001$) indicates better performance in the second semester. The correlation ($r = .461$, $p <$

.001) between both semesters suggests a moderate positive relationship in performance trends.

Table 4.3: Multivariate Tests

Effect		alue	F	Hypot hesis df	E rror df	ig.	Partial Eta Squared
Intercept	Pillai's Trace	979	3 202.622 b	2.000	1 40.000	000	.979
	Wilks' Lambda	021	3 202.622 b	2.000	1 40.000	000	.979
	Hotelli ng's Trace	5.75 2	3 202.622 b	2.000	1 40.000	000	.979
	Roy's Largest Root	5.75 2	3 202.622 b	2.000	1 40.000	000	.979
CGPA_ GROUP1	Pillai's Trace	.272	6 1.566	8.000	2 82.000	000	.636
	Wilks' Lambda	105	7 3.202b	8.000	2 80.000	000	.677
	Hotelli ng's Trace	.959	8 6.171	8.000	2 78.000	000	.713
	Roy's Largest Root	.077	1 43.716c	4.000	1 41.000	000	.803

a. Design: Intercept + CGPA_GROUP1

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Interpretation: Multivariate tests using Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root all indicate a highly significant effect of CGPA group on academic performance ($p < .001$). The partial Eta Squared values (ranging from .636 to

.803) indicate that a large proportion of variance in performance is explained by CGPA grouping

1st SEM T.M

	Sum of Squares	f	Mean Square		Sig.
Between Groups	8327.634		2081.909	1.092	.000
Within Groups	4805.006	41	34.078		
Total	13132.640	45			

ANOVA

2nd SEM T.M

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	14019.321	4	3504.830	120.557	.000
Within Groups	4099.147	141	29.072		
Total	18118.468	145			

Interpretation: The ANOVA results confirm significant differences in academic performance across CGPA groups for both semesters ($p < .001$). The F-values (61.092 and 120.557) and associated large partial Eta Squared values (.634 and .774) suggest strong between-group effects.

Table 4.4: Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1st SEM T.M	8327.634[a]	4	2081.909	61.092	.000	.634
	2nd SEM T.M	14019.321[b]	4	3504.830	120.557	.000	.774
Intercept	1st SEM T.M	118127.120	1	118127.120	3466.369	.000	.961
	2nd SEM T.M	100172.038	1	100172.038	3445.658	.000	.961
CGPA_GROUP1	1st SEM T.M	8327.634	4	2081.909	61.092	.000	.634
	2nd SEM T.M	14019.321	4	3504.830	120.557	.000	.774
Error	1st SEM T.M	4805.006	141	34.078			
	2nd SEM T.M	4099.147	141	29.072			
Total	1st SEM T.M	1019590.750	146				
	2nd SEM T.M	1172760.688	146				
Corrected Total	1st SEM T.M	13132.640	145				
	2nd SEM T.M	18118.468	145				

a. R Squared = .634 (Adjusted R Squared = .624)

b. R Squared = .774 (Adjusted R Squared = .767)

Interpretation: This test further confirms that CGPA group significantly influences student performance in both semesters. R-squared values indicate that 63.4% and 77.4% of the variability in 1st and 2nd semester marks, respectively, are explained by the CGPA group.

Table 4.5: Estimated Marginal Means

1. Grand Mean

Dependent Variable	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1st SEM T.M	89.027	1.512	86.038	92.017
2nd SEM T.M	81.983	1.397	79.222	84.744

Interpretation: The estimated marginal means provide adjusted group means. They reinforce that 'distinction' students consistently outperformed other groups. The data shows increasing performance trends from 'pass' to 'credit' to 'distinction' categories

2. CGPA_GROUP1

Dependent Variable	CGPA_GROUP1	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1st SEM T.M	Fail	80.125	4.128	71.965	88.285
	Pass	71.344	2.064	67.264	75.424
	Credit	80.462	.552	79.372	81.553
	distinction	98.457	1.217	96.050	100.863
	44.00	114.750	5.838	103.209	126.291
2nd SEM T.M	Fail	22.750	3.813	15.213	30.287
	Pass	77.344	1.906	73.575	81.112
	Credit	88.266	.509	87.258	89.273
	distinction	100.554	1.124	98.332	102.777
	44.00	121.000	5.392	110.341	131.659

Table 4.5: Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.947	1238.540[b]	2.000	138.000	.000
	Wilks' Lambda	.053	1238.540[b]	2.000	138.000	.000

	Hotelling's Trace	17.950	1238.540[b]	2.000	138.000	.000
	Roy's Largest Root	17.950	1238.540[b]	2.000	138.000	.000
TCF	Pillai's Trace	.000	.[b]	.000	.000	.
	Wilks' Lambda	1.000	.[b]	.000	138.500	.
	Hotelling's Trace	.000	.[b]	.000	2.000	.
	Roy's Largest Root	.000	.000[b]	2.000	137.000	1.000
CHF	Pillai's Trace	.000	.[b]	.000	.000	.
	Wilks' Lambda	1.000	.[b]	.000	138.500	.
	Hotelling's Trace	.000	.[b]	.000	2.000	.
	Roy's Largest Root	.000	.000[b]	2.000	137.000	1.000
CGPA_GROUP1	Pillai's Trace	.783	29.783	6.000	278.000	.000
	Wilks' Lambda	.248	46.311[b]	6.000	276.000	.000
	Hotelling's Trace	2.903	66.279	6.000	274.000	.000
	Roy's Largest Root	2.859	132.478[c]	3.000	139.000	.000
TCF * CHF	Pillai's Trace	.000	.[b]	.000	.000	.
	Wilks' Lambda	1.000	.[b]	.000	138.500	.
	Hotelling's Trace	.000	.[b]	.000	2.000	.
	Roy's Largest Root	.000	.000[b]	2.000	137.000	1.000
TCF * CGPA_GROUP1	Pillai's Trace	.000	.[b]	.000	.000	.
	Wilks' Lambda	1.000	.[b]	.000	138.500	.
	Hotelling's Trace	.000	.[b]	.000	2.000	.
	Roy's Largest Root	.000	.000[b]	2.000	137.000	1.000
CHF *	Pillai's Trace	.000	.[b]	.000	.000	.

CGPA_GROUP1	Wilks' Lambda	1.000	.[b]	.000	138.500	.
	Hotelling's Trace	.000	.[b]	.000	2.000	.
	Roy's Largest Root	.000	.000[b]	2.000	137.000	1.000
TCF * CHF * CGPA_GROUP1	Pillai's Trace	.000	.[b]	.000	.000	.
	Wilks' Lambda	1.000	.[b]	.000	138.500	.
	Hotelling's Trace	.000	.[b]	.000	2.000	.
	Roy's Largest Root	.000	.000[b]	2.000	137.000	1.000

a. Design: Intercept + TCF + CHF + CGPA_GROUP1 + TCF * CHF + TCF * CGPA_GROUP1 + CHF * CGPA_GROUP1 + TCF * CHF * CGPA_GROUP1

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Interpretation: Interaction effects (e.g., TCF*CHF, CHF*CGPA) were all non-significant ($p > .05$), indicating that CGPA group independently explains variance in scores, without significant interaction with other variables.

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary of Findings

This study explored the multivariate analysis of academic performance among HND II Statistics students at Kwara State Polytechnic, Ilorin, during the 2023/2024 academic session. The focus was on determining the influence of CGPA groupings on first and second semester performance, using robust multivariate statistical tools. SPSS version 23 was utilized to analyze data obtained, and the study incorporated a variety of descriptive, inferential, and multivariate tests.

The descriptive statistics showed that students in the 'distinction' category consistently performed better in both semesters than those in the 'credit' and 'pass' categories. The mean total score for the first semester was 83.03 while the second semester increased to 88.93, indicating notable academic improvement.

Paired sample t-tests revealed a statistically significant increase in performance from the first to second semester with a mean difference of -5.902 ($p < .001$). The correlation coefficient of .461 also signified a moderate positive relationship between both semesters.

Multivariate tests, including Wilks' Lambda and Pillai's Trace, indicated significant effects of CGPA group on semester scores. The partial eta squared values for CGPA_GROUP1 ranged from .636 to .803, suggesting that CGPA classification explained a substantial portion of variance in academic performance.

ANOVA results confirmed statistically significant differences across CGPA groups in both semesters. Between-subjects effects showed that 63.4% and 77.4% of variance in first and second semester scores respectively were explained by CGPA groups.

Estimated marginal means showed that students with higher CGPA (i.e., distinction category) had higher performance means. Interaction effects among other variables (TCF, CHF) were found to be non-significant, reinforcing that CGPA group had an independent and dominant effect on academic performance.

5.2 Conclusion

The results from this study provide compelling evidence that CGPA grouping has a significant and independent influence on student academic performance in both the first and second semesters. Students in higher CGPA categories consistently outperformed their peers in lower categories. The improvement from first to second semester indicates not just academic growth but also possible effectiveness in learning and teaching strategies across the academic session.

The study confirms that multivariate analysis is a potent statistical approach for assessing group differences in educational research. It also highlights the importance of using robust statistical models to understand how academic categorization impacts student outcomes over time.

5.3 Recommendations

Based on the findings and conclusion of the study, the following recommendations are proposed:

1. Targeted Academic Support: Students in the lower CGPA groups, particularly those in the 'pass' category, should be provided with additional academic support, such as tutoring, counseling, and skills workshops.

2. Performance Monitoring: Academic performance should be continuously monitored across semesters to identify early warning signs and intervene before students fall below expected performance thresholds.

3. Faculty Development: Lecturers and course facilitators should receive periodic training in pedagogical approaches that align with diverse learning styles, especially to support students in the lower CGPA brackets.

4. Data-Informed Decisions: The Polytechnic should utilize similar multivariate statistical techniques periodically to assess academic trends and the effectiveness of academic programs and policies.

REFERENCES

Aykut, E.B., Esra G., and Alperen (2004). An example of secondary school students' academic achievement in science course and achievement scores in performance assignment with regard to different variables: A boarding school example, Participatory Education Research, 1(2): 95-105.

Baksh Ali, Z. S., Pasaha, E. & Rastegar, A. (2013). Statistics and modeling. Tehran: Iranina textbook Publishing Company. Bandura, A. (1986). Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, N.J. Prentice Hall.

Batanero, C. & Diaz, C. (2010). Training teachers to teach statistic:what can we learn from research? Statistique et enseignement, 1(1), 5-20.

Borde, S. F. (1998). Predictors of student academic performance in the introductory marketing course. Journal of Education for Business, 73 (5), 302 – 307.

Byrne, M., Flood, B., & Griffin, J. (2014). Measuring the academic self-efficacy of first-year accounting students. Accounting Education, 23(5), 407-423. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/09639284.2014.931240

Chansarkar B. A. and Michaeloudis, A. (2001) Student profiles and factors affecting performance Int. j. math.educ. sci. technol., 2001, vol. 32, no. 1, 97–104, Pp 103-104.

Chapagain, Y. (2021). School student academic performance in Nepal: An analysis using the School Education Exam (SEE) results. International Journal on Studies in Education, 3(1), 22-36. https://doi.org/10.46328/ijonse.34

Chiesi, F., Pirmi. C., & Morsanyi, K. (2010). Learning Probability and Statistics:Cognitive and Non-cognitive Factors Related to Psychology Students' Achievement. International Association of Statistical Education (ISAE).

Chinn, S. (2012). Beliefs, anxiety and avoiding fear in Statistics. Child Development Research, 2012. 366 Journal of Research and Advances in Mathematics Education, 6(4), October 2021, 352-367 http://journals.ums.ac.id/index.php/jramathedu

Crosser, S. L. (1991). Summer birth date of children, Kindergarten entrance age and academic achievement, Journal of Educational Research, 84(3):140-146

Doyle, K. M., Dias, O., Kennis, J. R., Czarnocha, B. & Baker, W. (2015). The ratonal number subconstructs as a foundation for problem solving. Adults Learning Statistics: An International Journal, 11(1), 21-42. Retrieved from https://files.eric.ed.gov/fulltext/EJ1091996.pdf

Durden, G. C., & Ellis, L. V. (1995). The effects of attendance on student Learning in principles of economics. American Economic Review, 85(2), 343–346.

Emmioglu, E., & Capa-Aydin, Y. (2012). Attitudes and achievement in statistics: A metaanalysis study. Statistics Education Research Journal, 11(2) 95 – 102.

Felder, R. M. (1993). Reaching the second tier: Learning and teaching styles in college science education.

Flemming, N. (2011) Vark a guide to learning styles. Accessed on November 02, 2011 from http://www.varklearn.com/english/page.asp?p=categories.

Haist, S. A., Wilson, J. F., Elam, C. L., Blue, A. V., & Fosson, S. E. (2000). The effect of gender and age on

Hussein S., Gabriel, N., and Adam I. (2017), Effect of some performance indicator of mathematics in the Nalerigu senior High School., European Scientific Journal, 13(3):429-437

Huynh, M., Baglin, J., & Bedford, A. (2014). Improving the attitudes of high school students towards statistics: An island-based approach. In Sustainability in Statistics Education. Proceedingd of the Ninth International Conference on Teaching Statistics

(ICOTS9), Flagstaff, Arizona, USA. Voorburg; International Association of Statistics Association> Retrieved from https://icots.info/9/proceedings/pdfs/ICOTS9_9G@_HUYNH.pdf

Jamisola, N. B. (2014). Predictors of Statistics Performance of Fourth Year Students in Davao City [Unpublished Thesis]. University of Southeastern Philippines, Davao City, Philippines.

La Paro, K. M. and Pianta, R. C. (2010). Predicting childre's competence in the early school years:

Lemana, L. G. (2012). A Structural Model Predicting Performance of Freshmen Statistics Unpublished Thesis. University of Southeastern Philippines, Davao City, Philippines.

Lukkarinen, Koivukangas, and Seppala (2016). Relationship between class attendance and student performance, Procedia-Social and Behavioral Sciences 228:341-347

McGrath, A. L., Ferns, A., Greiner, L., Wanamaker, K., & Brown, S. (2015). Reducing anxiety and increasing self-efficacy within an advanced graduate psychology statistics course. Canadian Journal for the Scholarship of Teaching and Learning, 6(1), 5. Retrieved from https://eric.ed.gov/?id=EJ1057728 medical school performance: An important interaction. Advances in Health Sciences Education, 5(3), 197 – 205.

Mildin Jimenez Retutas, Marilyn Torela Rubio (2021). Multivariate analysis on performance in statistics, self-efficacy and attitudes of senior high school students, Journal of Research and Advances in Mathematics Education Volume 6, Issue 4, October 2021, pp. 352-367 DOI: 10.23917/jramathedu.v6i4.14368 p-ISSN: 2503-3697, e-ISSN: 2541-2590

Morrison, C. & Machintyre, W.P. (1993). Aspiration, motivation and achievement: Perspective in Education, 3 (2), 102-110

Murray, S. (2011). Declining participation in post-compulsory secondary school Statistics: students' views and solutions to the problem. Research in Statistics Education, 13(3), 269-285. https://doi.org/10.1080/14794802.2011.624731

New Jersey: Pearson Education. Wikipedia, http//en.m.wikipedia.org

Newman-Ford, L., Lloyd, S., & Thomas, S. (2009). An investigation in the effects of gender, prior academic achievement, place of residence, age and attendance on first-year undergraduate attainment. Journal of Applied Research in Higher Education, 1(1),13.

Nori, Z. (2002). Gender differences creativity, academic achievement (mathematics, sciences and language of literature) among high school in City of Shiraz, Iran: University of Shiraz, Shiraz.

Okeke, Evelyn Nkiruka, Okeke, Joseph Uchenna & Adashu Daniel (2018). Multivariate Analysis Of Variance Of University Students' Academic Performance, *International Journal of Applied Mathematics & Statistical Sciences (IJAMSS)* ISSN(P): 2319-3972; ISSN(E): 2319-3980 Vol. 7, Issue 1, Dec - Jan 2018; 13 – 20.

Omrod, J. E. (2008). Educational psychology: developing learners. Sixth Edition. Upper Saddle River,

Peters, P., Smith, A., Middledorp, J., Karpin, A., Sin, S., & Kilgore, A. (2013). Learning essential terms and concepts in statistics and accounting. Higher Education Research & Development, 33(4), 742 – 756. https://doi.org/10.1080/07294360.2013.863838

Pillow, B. (2008). A comparison of academic performance in A- level economics between two years:

Prado, M. & Gravoso, R. (2011). Improving high school students' statistical reasoning skills: A case of applying anchored instruction. The Asia-Pacific Education Researcher, 20(1), 61 – 72.

Salvan, E. (2016). Demographic Characteristics, Socio-Psychological Attributes, and Performance of College Students in Statistics. [Unpublished Dissertation]. Central Mindanao University, Bukidnon Philippines.

Salvan, N. L. (2014). Determinants of Grade 7 Mathematical Achievement: Basis for Developing a Tool for Selecting Special Program of the Science, Technology and Engineering (STE) Students [Unpublished master's thesis]. University of Southeastern Philippines, Davao City, Philippines

Swift, J. (2012). Challenges for enriching the curriculum: Statistics and probability. Statistics Teachers, 76, 268 – 269. https://doi.org/10.5951/MT.76.4.0268

Woodfield, R., & Earl-Novell, S. (2006). An assessment of the extent to which subject variation in relation to the award of first class degree between the arts and sciences can explain the 'gender gap'. British Journal of Sociology of Education, 27(3), 355.

Yvonne B., and Kola S. (1998) "An Analysis of High School Students' Performance on Five Integrated Science Process Skills" Research in Science & Technical Education, Volume 19, Number 2 / November 1, 2001Pp 133 – 145

Zacal, S. G. (2019). Psycho-social Attributes, Reading Comprehension and Problem-Solving Skills of Junior High School: A Structural Model on Statistics Performance [Unpublished Dissertation]. Bukidnon State University, Malaybalay City, Bukidnon, Philippines.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What Does Research Suggest About the Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature. Journal of Statistics Education, 16(2).