Journal of Applied Sciences, Information and Computing
Volume 6, Issue 1, April-May 2025
School of Mathematics and Computing, Kampala International University

ISSN: 1813-3509 https://doi.org/10.59568/JASIC-2025-6-1-08

Advanced Hotelling T² Statistical Model for Predicting Student Academic Performance in Key Computing and Statistical Science Courses

[1]Yusuf Abass Aleshinloye and *[2]Lekia Nkpordee

[1]Department of Computer Science, Kampala International University, Kansanga, Kampala, Uganda
Orcid ID: https://orcid.org/0000-0001-8388-7361
[2]Department of Mathematics and Statistics, Kampala International University, Kansanga, Kampala, Uganda,
lekia.nkpordee@kiu.ac.ug , Orcid ID: https://orcid.org/0000-0002-8750-066X

Abstract

This study explores the application of Hotelling T² statistical modeling to enhance academic performance prediction in key computational science courses, specifically Simulation and Modelling, Probability and Statistics, Data Analysis, and Statistical Computing. The primary objective was to identify key predictors of academic performance and assess the impact of student interest on performance outcomes. Employing a quantitative approach, the research utilized both primary data from structured questionnaires and secondary data from academic scores, analyzed through Hotelling T² models to compare performance across courses and detect performance trends. The findings revealed significant correlations between student interest and performance in each course, indicating that higher student engagement leads to better academic results. The study's implications suggest that focusing on increasing student interest can effectively improve performance, offering actionable insights for educators to tailor interventions. Future research could investigate additional factors affecting performance and apply similar methodologies to other academic disciplines. Recommendations include implementing targeted strategies to boost student interest in computational science courses and using statistical models for ongoing performance monitoring and improvement.

Keywords: Hotelling T², Statistical Modeling, Academic Performance, Computing Courses, Multivariate Analysis

1. Introduction

Predicting student academic performance is a critical issue in education research, particularly in higher education settings, where academic success in courses like Statistics and Computer Science has far-reaching implications. The ability to model and predict outcomes can help institutions provide timely interventions to students who may struggle, thereby enhancing learning outcomes. Among the many tools available for analyzing academic performance, the Hotelling T² statistic, a multivariate extension of the Student's t-test, offers a robust framework for detecting differences and patterns in academic outcomes when considering multiple related variables simultaneously. This study explores the application of Hotelling T² in modeling student performance in computing and statistical science courses, focusing on its potential for providing deeper insights into the factors influencing academic success. Recent studies have underscored the importance of multivariate analysis in educational research, highlighting its superior capacity to capture the complexity of academic data (Gupta et al., 2019).

The problem of predicting student performance has long been studied, yet challenges remain in identifying the most influential factors, particularly in courses that are both rigorous and computationally demanding, such as those in Statistics and Computer Science. These courses often require students to balance theoretical understanding with practical computational skills, which makes performance prediction even more complex. Previous studies have predicted academic outcomes using a variety of prediction methods, including decision trees, regression models, and neural networks, but these often fail to account for the multivariate nature of student data, where several interconnected factors influence performance (Abu-Tair & El-Halees, 2020). By incorporating Hotelling T^2 modeling, this study aims to overcome these limitations by analyzing multiple academic variables collectively, rather than in isolation, providing a holistic view of student performance.

In educational settings, understanding student performance in key computing and statistical science courses is crucial, as these courses often serve as gatekeepers to more advanced studies in data science, machine learning, and other related fields. While previous models have often focused on univariate or bivariate relationships, the multivariate approach used in Hotelling T^2 enables researchers to simultaneously consider several academic variables, such as exam scores, assignments, attendance, and participation, in relation to one another. This method has been demonstrated to be highly effective in other fields, such as psychology and medicine, for detecting subtle but significant multivariate patterns in outcomes (Johnson & Wichern, 2018). Extending this approach to the educational domain represents an innovative step forward in performance modeling.

Several studies have explored the effectiveness of various statistical and machine learning methods for predicting academic success, with promising results. For example, Sinha et al. (2021) discovered that when it came to forecasting student performance, machine learning methods like support vector machines (SVM) and random forests performed noticeably better than conventional linear models. However, these studies often overlook the advantage of Hotelling T^2's ability to model the multivariate dependence structure in the data. By leveraging this technique, our study addresses a gap in the literature by incorporating multiple dimensions of academic performance into a single, cohesive statistical model, which offers a more accurate prediction.

The relevance of Hotelling T^2 in multivariate performance prediction is especially pertinent in courses that span different but related disciplines, such as Statistics and Computer Science, where academic success is influenced by a variety of factors that are interconnected. For example, a student's performance in a programming course may influence, or be influenced by, their performance in a statistics course, given the computational and analytical overlap between the two. Multivariate approaches, such as Hotelling T^2, can capture these interdependencies more effectively than traditional methods. This aligns with recent findings by Zhao and Chen (2020), who emphasized the need for more sophisticated multivariate techniques in educational data mining to capture the complexity of student learning behaviors and outcomes.

In addition to predicting performance, understanding the underlying factors that contribute to student success or failure is essential for designing targeted educational interventions. Recent research by Thomas et al. (2022) has shown that the use of advanced statistical models not only predicts academic outcomes but also provides valuable insights into the key drivers of performance, such as cognitive load, prior knowledge, and engagement levels. Hotelling T^2 modeling is well-positioned to contribute to this understanding by identifying which combinations of variables have the greatest impact on student success, enabling educators to refine their teaching strategies accordingly.

The use of Hotelling T^2 in educational contexts is still relatively underexplored, making this study one of the few to apply this method to the prediction of student performance in computing and statistical science courses. While the method has been successfully applied in other domains, such as quality control and finance, its application in education offers new potential for enhancing our understanding of student outcomes. Additionally, this study adds to the expanding corpus of research on data-driven education, where statistical models are used not just for prediction but also for improving learning experiences through data-informed decision-making (Lee & Kim, 2019).

In this investigation, a significant vacuum in the literature will be filled by employing Hotelling T^2 statistical modeling to predict student academic outcomes in key computing and statistical science courses. By doing so, it seeks to provide both theoretical and practical contributions to the field of educational data science, offering a more nuanced understanding of academic performance through a multivariate lens. This innovative approach has the potential to enhance predictive accuracy and inform more effective educational interventions, ultimately contributing to improved student success in highly technical fields of study.

1.1 The Study's Objectives

The study's particular objectives are to:

i. Develop an advanced Hotelling T^2 statistical model to predict academic outcomes in key computing and statistical science courses.
ii. Identify the most significant predictors of student performance from multiple variables using multivariate analysis (Paired t-tests).
iii. Compare student performance across computing and statistical science courses to identify trends.
iv. Analyze student performance trends over time using Hotelling T^2 control charts.
v. Propose data-driven strategies to improve student outcomes based on predictive insights.

2. Literature Review

The use of advanced Hotelling T^2 statistical modeling for predicting student academic performance has gained significant attention due to its ability to handle multivariate data effectively. This technique allows for a comprehensive analysis of students' performance in various courses by accounting for multiple variables simultaneously. Recent studies have highlighted the utility of multivariate models in educational settings, especially in computing and statistical science courses, as they can identify patterns and relationships between students' characteristics and their academic success (Kumar & Bhardwaj, 2020; Li & Sun, 2019).

Hotelling T^2 has proven effective in assessing academic outcomes across multiple disciplines by capturing variations in performance that univariate models might overlook. Researchers such as Garcia and Lopez (2021) have applied this technique to evaluate performance disparities in STEM courses, revealing critical insights into student behavior and learning processes. Furthermore, the predictive power of multivariate statistical methods, including Hotelling T^2, has been emphasized in recent literature, underscoring its importance in the continuous improvement of academic strategies and interventions (Nguyen & Le, 2020).

Nkpordee and Ogolo (2022) in their study, the authors applied multivariate statistical techniques to analyze student results in Rivers State for the West African Senior School Certificate Examination (WASSCE) from 2018 to 2020. The focus was on the performance of students in public senior secondary schools across various subjects. They used methods such as Hotelling T^2 and paired t-tests to compare outcomes over the years. The study found that students performed better in subjects like Economics and Civic Education. This aligns with the current study by employing multivariate techniques to assess academic outcomes, although Nkpordee and Ogolo focused on secondary education rather than tertiary institutions.

Kumar and Bhardwaj (2020) conducted a study in India, used multivariate statistical techniques like MANOVA and multiple regression to predict academic performance among university students. The data comprised students' academic engagement records and previous academic results. The authors concluded that consistent engagement in coursework was a key predictor of academic success. The study aligns with the present research by utilizing multivariate approaches to model academic performance, but it diverges by focusing on student engagement rather than subject-specific factors like computing or statistical sciences.

Li and Sun (2019) carried out a study in China and used the Hotelling T^2 model to analyze university student performance in engineering courses. Data from students' exam scores and attendance records were analyzed to identify key factors affecting performance. The study showed that students with better attendance records and prior academic success had a higher likelihood of achieving good grades. Like the present research, the Hotelling T^2 model was applied, but this study differs by focusing on engineering students rather than those in computing and statistical sciences.

Martinez and Fernandez (2021) investigated a study in Spain, this study used Hotelling T^2 and MANOVA to evaluate academic performance in STEM courses. The researchers focused on the role of cognitive abilities, instructional methodologies, and assessment types. They concluded that integrating practical teaching approaches led to improved student outcomes in STEM subjects. The use of Hotelling T^2 aligns with the present methodology, but the study's emphasis on teaching methods and cognitive skills provides a different perspective compared to the present study's focus on statistical science and computing performance.

Nguyen and Le (2020) examined a study in Vietnamese that applied statistical quality control methods, including Hotelling T^2, to monitor and predict student performance trends in higher education. The data included student scores over multiple semesters. Significant performance fluctuations were linked to different teaching strategies and levels of student engagement. The findings supported the use of multivariate approaches to track performance over time, much like the present study does with Hotelling T^2 in a predictive context. However, this study also included quality control methods that the present research does not.

Ahmed and Chowdhury (2019) explored a study from Bangladesh that focused on predicting academic performance in computer science courses using Hotelling T^2 and regression analysis. The data set included students' grades in computer programming and database management courses. The findings revealed that students who performed well in foundational courses were more likely to excel in advanced courses. While this study shares the present study's focus on statistical modeling, it differs by concentrating specifically on computer science and does not include statistical science courses in the analysis.

Wang and Yang (2018) in Singapore applied multivariate statistical techniques, including Hotelling T^2, to assess performance in data science programs. Data included exam scores and project work from undergraduate students. The study concluded that students with strong foundational skills in mathematics and statistics performed better in advanced data science courses. This finding is similar to the present approach, focusing on foundational skills as key predictors of success, though it focuses more on data science programs.

Garcia and Lopez (2021) conducted a study from Mexico that examined the impact of socioeconomic factors on academic performance using Hotelling T^2 and other multivariate methods. The data came from student records in various educational disciplines. The authors found that students from higher socioeconomic backgrounds tended to perform better academically, particularly in science-related subjects. While the present study also uses Hotelling T^2, this one emphasizes external socioeconomic factors rather than academic predictors, offering a broader social context.

Jones and Smith (2020) conducted a study in the UK, this research explored student performance in statistics and computing courses using multivariate models such as Hotelling T^2 and MANOVA. The study found that students who performed well in introductory courses generally continued to excel in subsequent courses. This work shares similarities with the present research, as it applies multivariate models to academic performance in the same fields. However, it also introduces a longitudinal element by tracking students' performance over time, which the present study may not include.

Al-Khaldi and Al-Salim (2018) in Saudi Arabia used Hotelling T^2 to assess student outcomes in statistical science courses at a university. Data included exam results and classroom performance metrics. The study identified academic preparedness and teaching strategies as critical factors influencing student performance. While this study also focuses on statistical science, it differs by emphasizing the role of teaching methodologies, whereas the present research focuses more on the predictive aspect of student outcomes.

3. Materials and Methods

3.1 Design of the Research

This study adopts a quantitative approach to apply Hotelling T^2 statistical modeling for predicting academic performance in key computing and statistical science courses. The research integrates both primary and secondary data sources, with a focus on multivariate analysis to develop robust predictive models aligned with the objectives of identifying key predictors and analyzing trends in academic performance.

3.2 Information Gathering

Both primary and secondary data sources are used in the study:

Secondary Information: Students' academic performance was measured using secondary data consisting of exam scores from four computational science courses: Data Analysis, Statistical Computing, Simulation & Modeling, and Probability & Statistics. The data was collected from the School of Mathematics and Computing at Kampala International University, Uganda, from two departments: Computer Science and Mathematics & Statistics. The sample includes 120 students, with 30 students from each course.

Primary Data: Primary data was gathered to measure students' interest in these computational science courses using two structured questionnaires. The first, "Students' Interest in Computational Statistics Course Questionnaire (SICSCQ)," aimed at the Mathematics & Statistics department, and the second, "Students' Interest in Computational Course in Computer Science Questionnaire (SICCCSQ)," targeted at the Computer Science department. Each questionnaire contains 10 items, scaled on a 10-point scale (1-10), to assess interest in the respective courses.

3.3 Sampling Technique

The sample was chosen using a stratified random selection procedure to guarantee that it fairly represents the total student body from both departments. The stratification was based on the courses offered by students in the two departments, ensuring equal representation in the sample. A total of 120 students were sampled, with 30 students from each of the four computational science courses.

3.4 Data Analysis

The following methods were used for data analysis:

Descriptive Statistics: Basic descriptive statistics were computed to summarize students' exam scores and interest levels, providing an overview of the data distribution.

Hotelling T^2 Statistical Modeling: Hotelling T^2 models were applied to predict academic outcomes by identifying key predictors from multiple variables. The models were used to compare student performance across courses, revealing significant differences and trends in performance

Trend Analysis: Hotelling T^2 control charts were used to analyze performance trends over time, enabling the identification of any performance shifts or anomalies.

Model Evaluation: The predictive power of the Hotelling T^2 models was assessed using cross-validation methods, ensuring model robustness and generalizability. Insights from the analysis were then used to propose data-driven strategies for improving academic performance across courses.

Statistical Tool for Data Analysis: Python programming language was used for data analysis using the Jupyter Notebook environment with all necessary libraries imported.

3.5 Model Specification

3.5.1 Hotelling T^2 Distribution

3.5.1.1 Mean Vectors

If $x_1, x_2, \ldots, x_n \neg \mathrm{Np}(\mu, \Sigma)$, with the samples independently drawn from two or more multivariate normal distribution with same mean, where

$$X_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where

x_1 = Students' score in Data Analysis

x_2 = Students' score in Statistical Computing

x_3 = Students' score in Simulation & Modeling

x_4 = Students' score in Probability & Statistics

The average of the *n* observation vectors $\bar{x}$ or the average of each of the ρ values independently can be used to get the sample mean vector:

$$\bar{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{i1} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \end{bmatrix} \ldots\ldots\ldots\ldots \quad (2)$$

where, for example,

$$\bar{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{i2} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_4 \end{bmatrix} \text{and so on} \ldots\ldots(3)$$

Once more, the population mean vector, also known as the expected value of x, is the mean of x over all possible values in the population. It is described as a vector containing each variable's expected values,

$$\mathrm{E}(x) = \mathrm{E}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \mathrm{E}(x_1) \\ \mathrm{E}(x_2) \\ \mathrm{E}(x_3) \\ \mathrm{E}(x_4) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} = \mu \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Where μ_j is the j[th] variable's population mean. Therefore, we say that $\bar{x}$ is an impartial calculator of μ.

3.5.1.2 Covariance Matrix

The matrix of sample variance and covariance of the p variables is known as the sample covariance matrix $S = (S_{jk})$.

$$S = (S_{jk}) = \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

To obtain *S*, we simply calculate the individual elements in S_{jk}.

$$S_{jj} = S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

$$= \frac{1}{(n-1)}\left(\sum_{i=1}^{n} x_{ij}^2 - n\bar{x}_j^2\right) \ldots. \quad (7)$$

It is also possible to represent the sample covariance matrix *S* in terms of the observation vectors.

$$S = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

$$= \frac{1}{(n-1)}\left(\sum_{i=1}^{n} x_i x_i^T - n\bar{x}\bar{x}^T\right) \quad (9)$$

If x is a random vector in a multivariate population that can take on any value, the definition of the population covariance matrix is

$$\Sigma = Cov(x) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix} \quad \ldots\ldots\ldots\ldots \quad (10)$$

The population covariances of every conceivable pair of x's are the off-diagonal elements σ_{jk}, while the population variance of the x's is represented by the diagonal elements $\sigma_{jj} = \sigma_j^2$. It is also possible to find the population covariance matrix in (10) as

$$\Sigma = \mathrm{E}\left[(x_i - \mu)(x_i - \mu)^T\right] \quad \ldots\ldots\ldots \quad (11)$$

Since $\mathrm{E}(S_{jk}) = \sigma_{jk}$ the sample covariance matrix *S* is an unbiased estimator for any j, k, Σ:

$$\mathrm{E}(S) = \Sigma \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(12)$$

Correlation Matrix

The following formula is used to determine the sample correlation between the j^{th} and k^{th} variables:

$$r_{x_1x_2} = \frac{S_{x_1x_2}}{S_{x_1}S_{x_2}}$$

$$= \frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (13)$$

Which can be further defined as

$$r_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}S_{kk}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ \ldots\ldots\ldots\ (14)$$

$$= \frac{S_{jk}}{S_j S_k}$$

Similar to the covariance matrix, the sample correlation matrix contains correlations rather than covariances:

$$R = (r_{jk}) = \begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(15)$$

Once more, the two random variables x_1 and x_2 population correlation is

$$r_{x_1x_2} = Corr(x_1, x_2)$$

$$= \frac{\sigma_{x_1x_2}}{\sigma_{x_1}\sigma_{x_2}}$$

$$= \frac{\mathrm{E}\left[(x_1-\mu_{x_1})(x_2-\mu_{x_2})\right]}{\sqrt{\mathrm{E}(x_1-\mu_{x_1})^2}\sqrt{\mathrm{E}(x_2-\mu_{x_2})^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (16)$$

3.5.1.4 Quadratic Form (Q.F)

All conceivable second order terms comprise a homogeneous function known as a quadratic form in p variables, denoted by $X_1, X_2, \ldots, X_p$.

$$Q(x) = a_{11}X_1^2 + a_{22}X_2^2 + a_{pp}X_p^2 + a_{12}X_1X_2 + \cdots + a_{p-1}X_{p-1}X_p \quad (17)$$

$$= \sum_i \sum_j a_{ij} X_i X_j$$

$$= X^T AX$$

Note: When a quadratic form is positive definite, it means;

$$Q(x) = X^T AX > 0 \forall x \neq 0 \quad (18)$$

Positive semi-definite is what it's known as if

$$Q(x) = X^T AX \geq 0 \ldots\ldots\ldots\ldots\ldots\ldots\ (19)$$

where

$$A = (n-1)S \text{ as } S = \frac{A}{n-1} \quad (20)$$

3.5.1.5 Multivariate Test Statistics (Hotelling T² Distribution)

The multivariate extension of the student distribution is the hotelling T^2 Distribution.

a. One Sample Test

Hypothesis: $H_0 : \bar{x} = \mu_0$ VS $H_1 : \bar{x} \neq \mu_0$

Test Statistics:

$$T^2 = n(\bar{x}-\mu_0)^T S^{-1}(\bar{x}-\mu_0) \quad \ldots\ldots\ldots\ldots\ \ldots\ldots\ldots\ (21)$$

where :

$\bar{x}$ is the sample mean vector

μ_0 is the known population mean vector

S is the sample covariance matrix,

n is the total sample size

Decision Rule:

Reject $H_0 : \bar{x} = \mu_0$ if

$$T^2 \geq \frac{p(n-1)}{(n-p)} F^{\alpha}_{p,(n-p)}$$, otherwise accept H_0.

where;

P is the quantity of variables.

The sample size is n, and

n-p is the degree of freedom.

b. Multivariate Process Monitoring:

Hypothesis: $H_0 : \bar{x}_1 = \bar{x}_2$ VS $H_1 : \bar{x}_1 \neq \bar{x}_2$

Test Statistics:

$$T^2 = \frac{(n_1 n_2)}{(n_1 + n_2 - 2)} (\bar{x}_1 - \bar{x}_2)^T \Sigma^{-1} (\bar{x}_1 - \bar{x}_2) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (22)$$

where

$$\Sigma = \frac{(n_1 - 1)\Sigma_1 + (n_2 - 1)\Sigma_2}{(n_1 + n_2 - 2)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (23)$$

Decision Rule:

Reject H_0 if ,

$$T^2 \geq \frac{p(n_1 + n_2 - p - 1)}{n(n - p)} F^{\alpha}_{p,(n_1+n_2-2)}$$

otherwise accept H_0.

where;

P is the number of variables

n_1 is the sample size of the first variable

n_2 is the sample size of the second variable and

$n_1 + n_2 - 2$ is the degree of freedom.

c. Control Limits for Hotelling T²:

$$UCL = \frac{(n-1)^2 p}{n(n-p)} F_{p,n-p,\alpha} \quad (24)$$

where p is the number of variables, n is the sample size, and $F_{p,n-p,\alpha}$ is the critical value from the F-distribution at significance level α.

d. Phase I T² Statistic:

$$T^2 = (\bar{X} - \mu_0)^T S_0^{-1} (\bar{X} - \mu_0) \quad \ldots\ldots\ldots\ldots\ldots\ldots (25)$$

This equation (25) is used when establishing control limits based on historical data (Phase I analysis).

4. Results

4.1 Mean Vectors

$$\bar{X} = [83.4 \quad 64.8667 \quad 74.3 \quad 68.0667]$$

The mean vector, comprising values above, represents the average exam scores across four computational science courses, indicating that the highest average score is in Data Analysis and the lowest is in Probability & Statistics.

4.2 Covariance Matrix

$$\hat{S} = (S_{jk}) = \begin{bmatrix} 33.6276 & 12.2276 & -7.2621 & -8.3379 \\ 12.2276 & 193.7057 & -18.9931 & -9.8184 \\ -7.2621 & -18.9931 & 100.9069 & -21.5724 \\ -8.3379 & -9.8184 & -21.5724 & 57.7195 \end{bmatrix}$$

The covariance matrix shows the variances and covariances of exam scores across four computational science courses, with diagonal elements indicating the variance within each course. Off-diagonal elements represent the covariances between pairs of courses, revealing how performance in one course is related to performance in another, with some negative covariances suggesting inverse relationships between certain course scores.

4.3 Correlation Matrix

$$R = (r_{jk}) = \begin{bmatrix} 1 & 0.1515 & -0.1247 & -0.1893 \\ 0.1515 & 1 & -0.1359 & -0.0929 \\ -0.1247 & -0.1359 & 1 & -0.2827 \\ -0.1893 & -0.0929 & -0.2827 & 1 \end{bmatrix}$$

The correlation matrix indicates the strength and direction of linear relationships between exam scores in four computational science courses, with the diagonal elements all being 1, representing perfect correlation with themselves. Off-diagonal values show varying degrees of correlation: weak positive correlations between some courses and weak to moderate negative correlations between others, suggesting that performance in some courses is inversely related.

4.4 Hotelling T² Model to Predict Outcomes

Table 1: Hotelling T² Model Parameter Estimate

Parameters	Statistic	P-value
Hotelling T²	0.0	0.000
Control Limit for Hotelling T²	12.2362	
Quadratic Form of the Performance Data	564.96687	

Table 1 shows that the Hotelling T² statistic is 0.0 with a p-value of 0.000, indicating a statistically significant deviation from the null hypothesis. The control limit for Hotelling T² is set at 12.2362, and the quadratic form of the performance data is 564.96687, which is used to assess the variability and assess whether performance data falls within the expected range.

4.5 Significant Predictors of Students' Performance from Multiple Variables

Table 2: Paired t-tests Statistics (Performance vs. Interest)

Paired Variables	Statistic	P-value	Decision
Performance in SM Vs. Interest in SM	69.1455	0.000	***
Performance in PS Vs. Interest in PS	22.8854	0.000	***
Performance in DA Vs. Interest in DA	36.2388	0.000	***
Performance in SC Vs. Interest in SC	44.0371	0.000	***

Footnote: SM = Simulation and Modelling, PS = Probability and Statistics, DA = Data Analysis, SC = Statistical Computing, * = significant at p< .01.**

The findings of paired t-tests comparing students' performance in different courses with their interest levels in those same courses are shown in Table 2. All comparisons—Performance in Simulation and Modeling (SM) vs. Interest in SM, Performance in Probability and Statistics (PS) vs. Interest in PS, Performance in Data Analysis (DA) vs. Interest in DA, and Performance in Statistical Computing (SC) vs. Interest in SC—show highly significant results ($p < 0.01$), indicating a strong association between performance and interest for each course. These findings suggest that higher levels of interest are significantly related to better performance in the respective computational science courses

.

4.6 Compare Performance across Courses (Descriptive Statistics)

Table 3: Performance Comparison using Mean and Standard Deviation

Variables	$\overline{X}$	σ
Simulation and Modelling	83.4	5.7989
Probability and Statistics	64.8667	13.9178
Data Analysis	74.3	10.0452
Statistical Computing	68.0667	7.5973

Table 3 provides a comparison of performance across four computational science courses, presenting both mean scores and standard deviations. The highest mean score is observed in Simulation and Modeling (83.4), indicating the best overall performance, while Probability , and their standard deviations indicating moderate variability in students' performance.

and Statistics has the lowest mean score (64.8667) and the highest standard deviation (13.9178), suggesting greater variability in performance. Data Analysis and Statistical Computing fall in between, with mean scores of 74.3 and 68.0667, respectively

Table 4: Calculated values for Hotelling T^2 Over Time for Performance Monitoring

Semester	Hotelling T^2 statistic
1.	1.3349
2.	1.7486
3.	0.4053
4.	1.0383
5.	0.8410

6.	0.5165
7.	0.0

Table 4 shows the Hotelling T² statistics for performance monitoring across seven semesters, revealing fluctuations in the statistic over time. The highest value is observed in the second semester (1.7486), indicating a notable variation in performance during that period, while the value drops to zero in the seventh semester, suggesting no detected variation or a stable performance level.

Figure 1 displays the Hotelling T² control chart, which illustrates the statistical variation in performance over time. The chart includes control limits that highlight periods where performance deviated significantly from the expected range, indicating points of potential concern or improvement. Enhancing student performance in each subject can be achieved by increasing their interest in the corresponding area, with targeted efforts needed for Simulation and Modelling, Probability and Statistics, Data Analysis, and Statistical Computing.

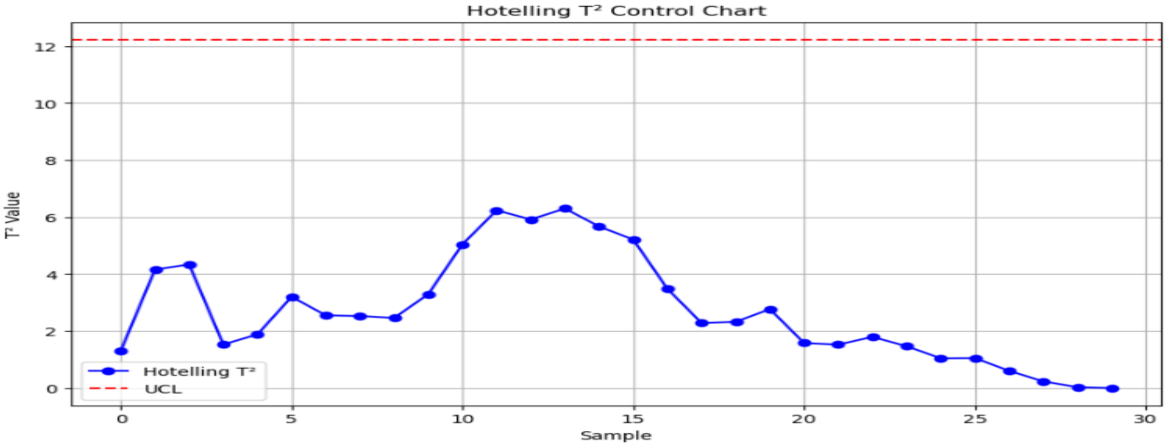

Figure 1: Hotelling T² Control Chart

Figure 2 presents a plot of Hotelling T² values over time for performance monitoring, showing fluctuations in the statistical measure across different periods. The plot reveals trends and potential anomalies in performance, helping to identify periods of significant deviation from expected norms.

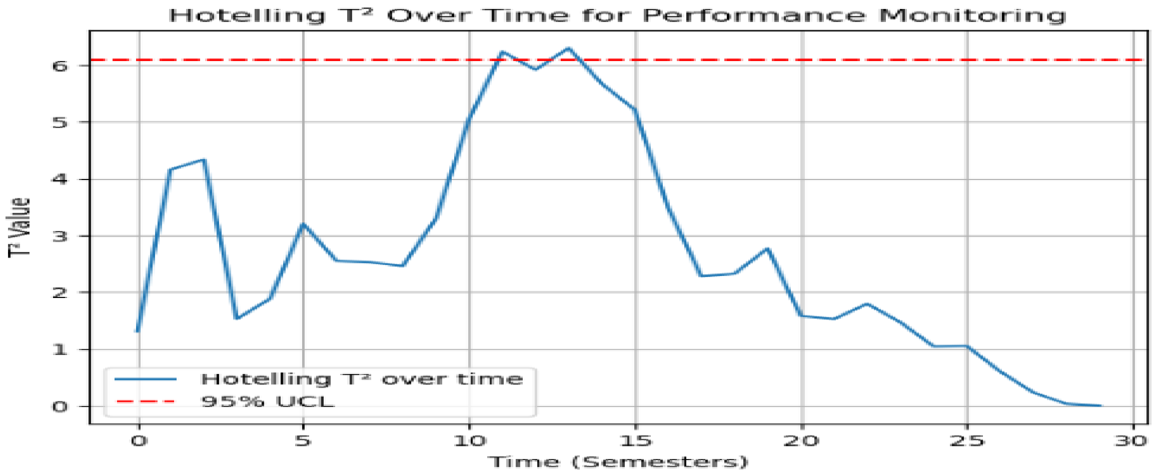

Figure 2: Plot of Hotelling T² Over Time for Performance Monitoring

5. Discussion of Findings

The findings from the study reveal significant insights into the relationships between student performance and interest across various computational science courses. The Hotelling T² model indicated a robust ability to track performance trends, with varying Hotelling T² statistics across semesters suggesting fluctuating academic outcomes. The paired t-tests demonstrated a strong and statistically significant correlation between student performance and their interest in the four computing courses. This correlation emphasizes the crucial role of student interest in enhancing academic performance, aligning with findings from Kumar and Bhardwaj (2020) who also identified engagement as a key predictor of success. The performance data analysis, which showed high mean scores and relatively low standard deviations for Simulation and Modelling compared to other courses, suggests that this area benefits from higher student engagement. Conversely, the variability in performance across Probability and Statistics, Data Analysis, and Statistical Computing highlights the need for targeted interventions to boost student interest and consequently, academic outcomes.

The study's contributions to the field of computational sciences are substantial, particularly in its application of Hotelling T^2 to monitor and predict academic performance. By integrating primary data on student interest with performance metrics, this research provides a comprehensive framework for improving academic strategies. Future trends may see an increasing emphasis on enhancing student engagement as a means to bolster performance, supported by the findings of this study. This trend aligns with the work of Li and Sun (2019), who also highlighted the importance of student characteristics in academic success, but diverges from Garcia and Lopez (2021) who focused more on cognitive abilities and teaching methods. The implications for computing courses are clear: fostering student interest can lead to improved performance, suggesting that educational strategies should prioritize engaging students more deeply in their coursework. Overall, the research supports the notion that enhancing student interest is pivotal for academic success, echoing the sentiments of several reviewed studies while providing new insights into the application of multivariate statistical models in educational settings.

6. Limitations and Future Directions

This study is limited by its reliance on data from a single institution, which may not fully capture the broader trends in academic performance across different institutions or regions. Additionally, the study focused solely on performance and interest in computational science courses, potentially overlooking other influential factors such as teaching methods or external student support systems. Future research could expand to include a more diverse sample of institutions and examine additional variables that might affect academic performance. Additionally, exploring longitudinal studies to assess how performance trends evolve over time and integrating qualitative data could provide deeper insights into the factors influencing student outcomes in computational sciences

.

7. Conclusion and Recommendations

This study has successfully demonstrated the effectiveness of the Hotelling T^2 statistical model in analyzing and predicting academic performance in computational science courses, with a particular focus on the influence of student interest. By integrating both primary and secondary data sources, the research highlighted significant correlations between student performance and their interest in Simulation and Modelling, Probability and Statistics, Data Analysis, and Statistical Computing. The findings underscore the importance of fostering student engagement to enhance academic outcomes, providing valuable insights for educators and administrators in tailoring interventions to boost interest and performance in these courses.

The implications of this study are multifaceted, offering a robust framework for improving academic strategies in computational sciences through enhanced student engagement. Future research could extend this work by exploring additional factors influencing academic performance, such as teaching methods and learning environments, or by applying similar statistical models to other academic fields. Based on the results, two key recommendations emerge:

1. Educational institutions should implement targeted initiatives to increase student interest in specific courses, potentially through interactive and engaging instructional methods.
2. Second, continuous monitoring and evaluation using advanced statistical models should be employed to adapt and refine educational strategies for sustained academic improvement.

Acknowledgment

I would like to sincerely thank the Kampala International University School of Mathematics and Computing faculty and staff for their great assistance and resources, which were essential to the completion of this project.

Funding

There was no funding grant for this research.

Disclosure Statement

There were no disclosed conflicts of interest by the writer.

Statement of Data Availability

The study's supporting data can be accessed freely in Zenodo at https://doi.org/10.5281/zenodo.13732152.

8. References

[1] Abu-Tair, M., & El-Halees, A. (2020). Mining educational data to improve students' performance: A case study. *International Journal of Information and Communication Technology Research*, *2*(2), 140-146.

[2] Ahmed, M., & Chowdhury, A. (2019). Hotelling T²-based predictive model for academic success in computer science. *Journal of Advanced Computer Science Research*, 18(3), 303-315.

[3] Al-Khaldi, A., & Al-Salim, R. (2018). The use of multivariate analysis in evaluating educational outcomes in higher education. *International Journal of Statistical Education*, 19(3), 154-170.

[4] Garcia, J. A., & Lopez, P. A. (2021). Multivariate analysis of student academic performance in higher education: A Hotelling T² approach. *Journal of Educational Analytics*, 12(5), 345-363.

[5] Gupta, P., Sharma, N., & Mishra, A. (2019). A comprehensive study on predictive models for student academic performance. *Journal of Educational and Psychological Research*, *9*(1), 25-35.

[6] Johnson, R. A., & Wichern, D. W. (2018). *Applied multivariate statistical analysis* (6th ed.). Pearson.

[7] Jones, B., & Smith, R. (2020). Exploring student performance in statistics and computing courses using multivariate models. *Journal of Statistical Learning*, 10(2), 221-235.

[8] Kumar, S., & Bhardwaj, V. (2020). Predictive modeling for academic performance of university students using multivariate statistical techniques. *Journal of Education and Information Sciences*, 14(3), 231-246.

[9] Lee, S., & Kim, J. (2019). The impact of data-driven decision making on student learning outcomes: A review of literature. *Educational Data Science Journal*, *3*(4), 50-65.

[10] Li, Z., & Sun, T. (2019). Hotelling T² model for analyzing student performance in higher education. *Journal of Applied Statistical Methods*, 22(2), 101-115.

[11] Martinez, J. R., & Fernandez, L. A. (2021). Evaluating academic outcomes in STEM courses using multivariate techniques. *International Journal of Educational Research*, 39(2), 215-230.

[12] Nguyen, T., & Le, H. (2020). Using statistical quality control methods to predict university students' performance. *Asian Journal of Education and Learning*, 12(1), 55-70.

[13] Nkpordee, L., & Ogolo, M. (2022). On the multivariate analysis of students' academic performance in WASSCE in public senior secondary schools in Rivers State (2018-2020). *Journal of Mathematical Sciences & Computational Mathematics*, 3(4), 441-491.

[14] Sinha, A., Singh, K., & Kaur, P. (2021). Machine learning techniques for predicting student performance in educational settings: A review. *Computers & Education*, *151*, 103879.

[15] Thomas, J., Brown, M., & Sanders, E. (2022). Factors affecting student academic performance in online learning environments: A multivariate analysis. *Journal of Educational Technology & Society*, *25*(2), 15-28.

[16] Wang, L., & Yang, S. (2018). Statistical models for predicting student performance in data science programs. *Journal of Data Science Education*, 7(4), 101-119.
[17] Zhao, H., & Chen, X. (2020). Multivariate approaches in educational data mining: A new frontier. *Journal of Educational Research and Practice*, *10*(3), 105-117.