Principal Component Analysis of Students Academic Performance

F. B. K. Twenefour[1], E. N. N. Nortey[2], E. M. Baah[1]

ABSTRACT

The purpose of this study was to identify a metric for measuring students' performance in the Department of Mathematics and Statistics of a public university in Ghana. Some of the students of the department are of the view that the current grading system used by the Department does not do a good job of distinguishing between the performances of students, as at times students of different academic performance could end up with the same Grade Point Average (GPA), a performance measure. Data for the research which covers the 2012/2013 third year students of the Department were obtained from the university's student records unit. Principal Component Analysis (PCA) was used to analyze the data. Three principal components were retained as rules or indices for the classification of students' performance. A derivative of the first principal component, RSI, could serve as a new performance measure for the Department as it takes into consideration differences in the raw scores of the students.

Keywords: Academic performance, principal component analysis, relative score index (RSI).
JEL Codes: C1, C3, C4, I2.
Available Online: 18th Feb, 2015
MIR Centre for Socio-Economic Research, USA.

1.0 INTRODUCTION

The performance of a student in a given semester is gauged by calculating what is referred to as Grade Point Average (GPA). The grading system for the raw exams scores used by the Department is as shown in Table A1 of Appendix A.

The GPA is found by dividing the sum of the products of the numerical equivalents of the letter grade obtained in a course and the corresponding credit hours by the total number of credit hours. Some students argue that the system is not fair because two students with different performance could end up with the same GPA. To see their point, we examine the information in Table A2 (see Appendix A)

[1] Department of Mathematics and Statistics, Takoradi Polytechnic, P. O. Box 256, Takoradi, Western Region, Ghana. E-mail: frank.twenefour@tpoly.edu.gh
[2] Department of Statistics, University of Ghana, P. O. Box LG 115, Legon, Accra, Ghana.

concerning Student 1 and Student 2. Suppose the raw scores obtained by the two students in three courses, MM, EM and TM and the associated credit hours are as shown in Table A2.

From the information presented in Table A2, it is clear that Student 2 is better in performance than Student 1, yet both students have the same GPA. The students argue that a suitable grading system should take into consideration any differences in the raw scores. There is therefore the need to explore the possibility evolving an alternative performance measure for the Department that could take into consideration the concerns of the students. We attempted to come up with an alternative to the current grading system by applying the statistical technique of Principal Component Analysis to the raw course scores of the student which resulted in the creation of a new instrument referred to as RSI, which could potentially address the concerns of the students.

Some students are offered straight admission into a programme of their choice while others are offered admission as physical sciences students. Those who are offered admissions as physical sciences students may end up finalizing their degree programmes in Physics, Chemistry or Mathematics and Statistics Departments. Those who opt for programmes in the Department of Mathematics and Statistics may major in mathematics only or statistics only, or in the combined subject of mathematics and statistics. Most often, students opt for certain programmes based on their performance in related courses they have taken previously.

Our primary aim in the research reported in this paper was to develop a system by which one could report the overall performance of a student relative to that of the student's class mates, tell which subject area a student's aptitude lie and the semester in which a student did relatively better. To accomplish this, the following objectives were pursued: rank the overall performance of students, determine the subject area in which a student performs better and which of the semesters, first or second or both, in which a student's performance is relatively better.

Section 2 explores the mathematical basis of the major statistical technique, Principal Component Analysis, used in analyzing the data and Section 3 presents and discusses the results of the analysis. A summary of the research, the findings and the implications of the findings are presented in the last section, Section 4.

2.0 MATERIALS AND METHODS

In Principal Component Analysis (PCA) new uncorrelated variables (called Principal Components (PC)) are formed which are a linear combination of the original (observable) variables, and the number of new variables is equal to the number of old variables. However, the new variables are so formed that the first principal component accounts for the highest variance in the data, the second principal component accounts for the highest of the remaining variance in the data, the third principal component accounts for the highest of the remaining variance not accounted for by the first and second components, and so on. Ideally, one would want a situation where the first few principal components account for much of the variance in the original data and thereby achieving data reduction by replacing the original variables by the first few principal components, for further analysis or interpretation of the correlation amongst the indicator variables (Everitt and Dunn 2001; Johnson and Wichern 1992; Sharma, 1996).

Given the observed variables $X_1, X_2, \dots, X_p$ and the coefficients (weights) w_{ij}, $i = 1, \dots, p$, $j = 1, \dots, p$, the principal component $C_1, C_2, \dots, C_p$ are given by

$$\begin{aligned} C_1 &= w_{11}X_1 + w_{12}X_2 + \dots + w_{1p}X_p \\ C_2 &= w_{21}X_1 + w_{22}X_2 + \dots + w_{2p}X_p \\ &\vdots \\ C_p &= w_{p1}X_1 + w_{p2}X_2 + \dots + w_{pp}X_p \end{aligned}$$

To place a limit on the variance of the C_is, $i = 1, \ldots, p$ and to guarantee that the new axes representing the C_is are uncorrelated, the weights w_{ij}, $i = 1, \ldots, p$, $j = 1, \ldots, p$ are estimated subject to the conditions given by Equations 1 and 2 respectively (Sharma, 1996; Everitt and Dunn 2001; Johnson and Wichern, 1992).

$$w_i' \cdot w_i = 1 \qquad \ldots\ldots\ldots 1$$

and

$$w_i' \cdot w_j = 0 \quad \text{for all } i \neq j \qquad \ldots\ldots\ldots 2$$

where

$$w_i' = (w_{i1}, w_{i2}, \ldots, w_{ip})$$

Given the mean μ_i and the standard deviation σ_{ii} of the variable X_i, the transformed variables Z_i, $i = 1, \ldots, p$ given by

$$Z_i = \frac{X_i - \mu_i}{\sigma_{ii}}$$

could be used to form the principal components (Johnson and Wichern, 1992). Expressed in matrix notation, the vector of standardized variables could be written as

$$Z = \left(V^{1/2}\right)^{-1}(X - \mu)$$

where $\mu' = (\mu_1, \mu_2, \ldots, \mu_p)$ and $V^{1/2}$ is the standard deviation matrix given by

$$V^{1/2} = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}$$

$E[Z_i] = 0$, $\text{Var}[Z_i] = 1$, $i = 1, \ldots, p$ and $Cov(Z) = \left(V^{1/2}\right)^{-1}\Sigma\left(V^{1/2}\right)^{-1} = \rho$ where the variance-covariance matrix Σ and the correlation matrix ρ of X are given by

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1p}^2 \\ \sigma_{21}^2 & \sigma_{22} & \cdots & \sigma_{2p}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}^2 & \sigma_{p2}^2 & \cdots & \sigma_{pp}^2 \end{bmatrix}$$

$$\rho = \begin{bmatrix} \frac{\sigma_{11}^2}{\sigma_{11}\sigma_{11}} & \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} & \cdots & \frac{\sigma_{1p}^2}{\sigma_{11}\sigma_{pp}} \\ \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} & \frac{\sigma_{22}^2}{\sigma_{22}\sigma_{22}} & \cdots & \frac{\sigma_{2p}^2}{\sigma_{22}\sigma_{pp}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}^2}{\sigma_{11}\sigma_{pp}} & \frac{\sigma_{2p}^2}{\sigma_{22}\sigma_{pp}} & \cdots & \frac{\sigma_{pp}^2}{\sigma_{pp}\sigma_{pp}} \end{bmatrix}$$

and

$$\rho_{ij} = \frac{\sum_{k=1}^{n}(x_{ki} - \mu_i)(x_{kj} - \mu_j)}{n}, \quad i \neq j$$

is the covariance between variables X_i and X_j, each of which has n observations.

The p principal components $C' = [C_1, C_2, \ldots, C_p]$ are then given by

$$C = A'Z$$

where $A = [e_1 e_2, \ldots, e_p]$ and the e_i s $i = 1, 2, \ldots, p$ are the eigenvectors of ρ. The eigenvalue-eigenvector pairs $(\lambda_1, e_1), (\lambda_2, e_2), \ldots, (\lambda_p, e_p)$ of ρ are such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$, $e_i' \cdot e_i = 1$ and $e_i' \cdot e_j = 0$.

$$\mathrm{Var}(C_i) = e_i' \rho e_i = \lambda_i$$

and

$$\sum_{i=1}^{p} Var(C_i) = \sum_{i=1}^{p} Var(Z_i) = p$$

Thus the proportion of the variance in the data that is accounted for by the C_j is given by λ_j / p.

2.01 PRINCIPAL COMPONENT LOADINGS

The correlations between a given PC C_i and a given standardized variable Z_j, referred to as the loading of variable Z_j on C_i, is given by

$$Corr(C_i, Z_j) = e_{ij} \cdot \lambda_j^{(1/2)}$$

The loading, which lies between -1 and 1 inclusive, reflects the degree to which each Z_j influences each C_i given the effect of the other variables Z_k, $j \neq k$ (Johnson and Wichern, 1992). The higher the absolute value of a loading of a variable on a component, the more influential the variable is in interpreting and naming the component. The magnitude of a loading considered to be significant depends on the sample size. Hair et al. (2006) recommends a loading of absolute value of at least 0.30 for a sample size of 350 and 0.35 for a sample size of 250, with lower sample sizes requiring higher absolute loading values. However loadings of magnitude greater or equal to 0.50 are thought to be more appropriate (Sharma, 1996).

2.02 THE PRINCIPAL COMPONENT AS AN INDEX

Using the principal component as an index requires the determination of principal component scores and factor loadings. By substituting the standardized observed values of the variables into the equation

$$C_i = e_i' Z$$

We obtain the values of the i^{th} principal component, referred to as the principal component scores.

Since the first few components almost invariably account for a greater percentage of the variance in the original data they can represent most of the information in the original data. Principal components analysis is thus regarded as a dimensionality reduction technique as the original *p*-dimensional data set could be represented in a lower *m*-dimensional space, where $m < p$. In particular, given that each principal component is a weighted sum of the original variables and the first few principal components account for a reasonably large proportion of the variance in the data, it may be possible to use each of these principal components as an index, with the factor loadings providing guidance as to what kind of index a given principal component is, as the higher the loading of a variable the greater the influence it has on the component.

2.0 RESULTS AND DISCUSSION

The data which is the subject of the analysis whose results follow were obtained from the university's student records unit. The data covers the 2012/2013 third year students and is made up of ten courses, six of which are Statistics courses and the remaining four, Mathematics courses. The ten courses are the variables considered; each with several observations which are the grades of students in the various courses.

3.01 SUMMARY STATISTICS

Table 1 shows the summary statistics of the Data. The mean grades of the Mathematics courses are quite close to each other. Probability Distributions (STA 301) and Research Methods (STA 399) have their mean grades quite higher than the other four Statistics courses, namely: Statistical Methods I (STA 303), Sampling Techniques and Survey Methods (STA 302), Design and Analysis of Experiment (STA 305) and Data Analysis I (STA 304). The values in Table 1 shows that Advanced Calculus II (MAT 302) has the most widely spread out performance in the case of mathematics courses while Design and Analysis of Experiments (STA 305) has the most widely spread out performance in the case of the statistics courses. It is also observed that the maximum grade score was recorded in Introductory Analysis (MAT 303) while the Minimum grade score was recorded in Advanced Calculus II (MAT 302). It can further be seen that the performance in the mathematics courses are relatively better compared to performance in the statistics courses.

The data was also explored on both semester and subject bases using graphical representations as shown in Figures 1, 2, 3 and 4. Analyzing the performance by semester, it can be deduced from Figure 1 that for the first semester courses, the mean grade of Probability Distributions and Research Methods are higher than that of the rest of the other courses especially in the case of Research Methods, while for the second semester (Figure 2), the mean grade of the courses are quite close to each other. Figure 3 shows, as was observed above, that the performance in Research Methods (STA 399) is

Table 1: Summary statistics of the data.

Variable	Code	N	Minimum (%)	Maximum (%)	Mean (%)	Standard deviation (%)
Advanced Calculus I	MAT 301	258	21	92	63.19	11.946
Introductory Analysis	MAT 303	258	17	94	68.02	10.993
Probability Distributions	STA 301	258	26	87	58.23	10.970
Statistical Methods I	STA 303	258	35	86	63.21	10.563
Research Methods	STA 399	258	45	89	72.09	8.179
Advance Calculus II	MAT 302	258	6	88	65.61	12.176
Modern Algebra	MAT 304	258	26	97	68.03	11.417
Sampling Techniques & Survey Meth.	STA 302	258	33	87	64.98	10.325
Data Analysis I	STA 304	258	38	85	64.94	8.423
Design & Analysis of Experiment	STA 305	258	21	90	65.09	11.463

Source: Result from analysis of data, 2014.

relatively higher as its approximating normal curve is shifted to the right relative to that of the other statistics courses. Also Figure 4 reveals that the performance in Modern Algebra (MAT 304) and Introductory Analysis (MAT 303) are higher than that of Advanced Calculus I (MAT 301) and Advance Calculus II (MAT 302).

3.02 CORRELATION ANALYSIS

This section gives the result of the correlation analysis. The correlation matrix is a numerical evaluation of the relationships among the ten courses. Table 2 shows that there are some fairly high correlations among the courses. The lowest correlation (0.250) is between the performance in Advanced Calculus II (MAT 302) and Research Methods (STA 399). This may be due to the fact that Research Methods is not as numerate/algebraic as the other mathematics and statistics courses. The highest correlation (0.618) is between performance in Sample Techniques and Survey Methods (STA 302) and Data Analysis I (STA 304). This could be explained by the fact that they are both statistics courses. There is also an appreciable level of correlation (greater than 0.5) between Advanced Calculus 1 (MAT 301) and Probability Distribution (STA 301). Indeed there are altogether 13 pairs of courses with such correlations.

Also Table 2 reveals that Research Methods have a weak correlation with both mathematics and other statistics courses. This may be due to the fact that, the content of Research Methods is not numerate or algebraic in nature -- it is unlike the other mathematics and statistics courses.

3.03 EIGEN ANALYSIS OF THE CORRELATION MATRIX

Table 3 shows the eigenvalues obtained from the Eigen analysis of the correlation matrix of the ten courses. The number of principal components produced is always equal to the number of the original variables, thus ten principal components were generated. The eigen-value-greater-than-one rule suggests that two principal components should be retained for further analysis or interpretation but three components (the first three) were retained because there appear to be another major change in the direction of the curve of the scree plot of Figure 5 at PC3. From Table 3, the first principal component, PC1, accounts for 49.2% of the total variance in the data. The second principal component accounts for 12.9% of the total variance in the data. Together, the first and second components account for 62.1% of the variance in the data. The third principal component accounts for 8.0% of the total variation in the data.

Figure1: Performance in first semester

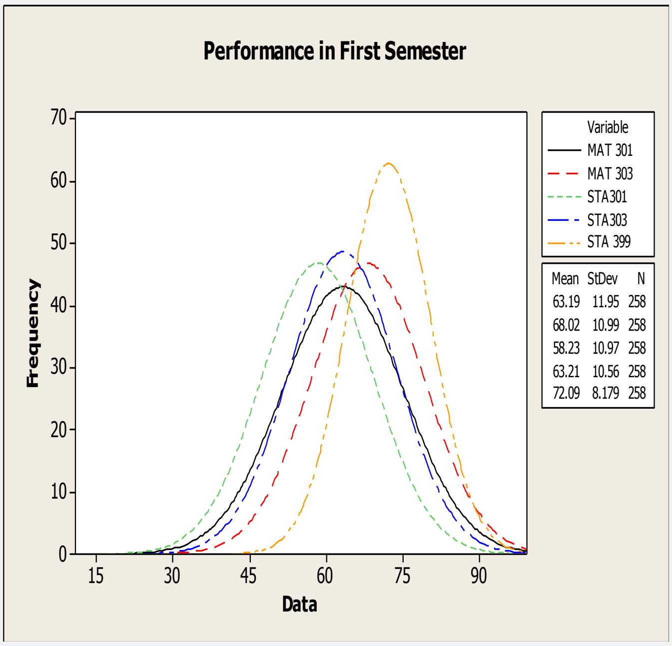


Figure 2: Performance in second semester.

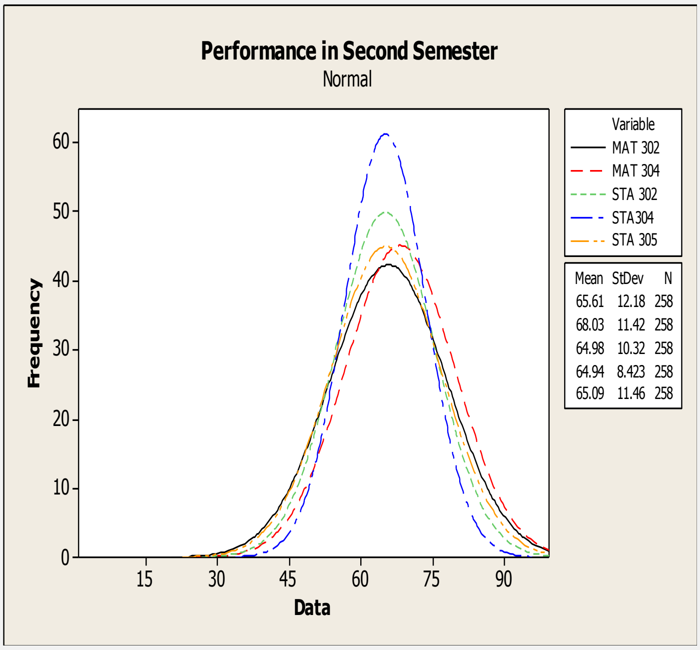
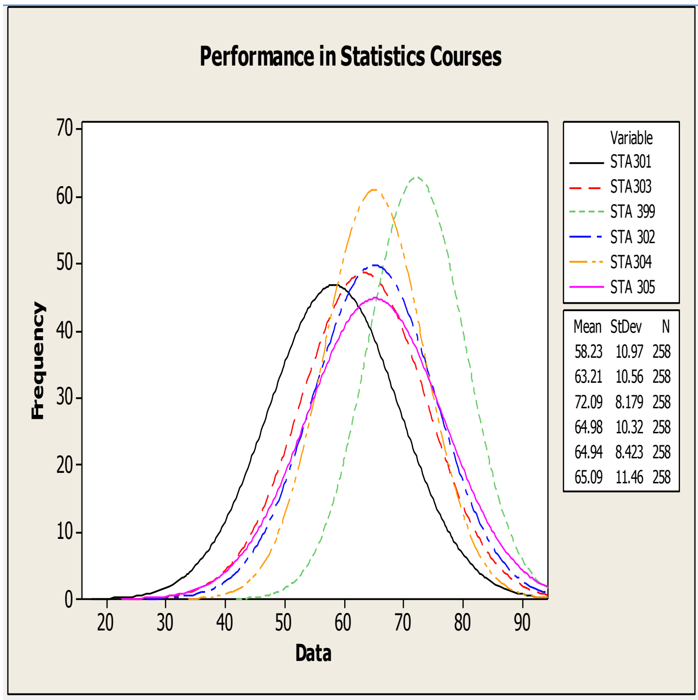
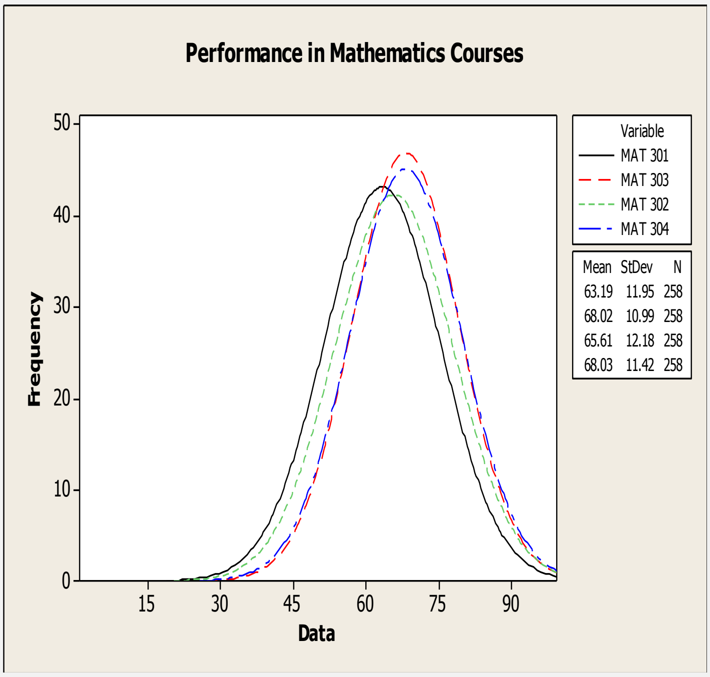


Figure 3: Performance in statistics courses

Figure 4: Performance in mathematics Courses





The 3rd principal component together with the first and second principal components accounts for 70.1% of the total variance in the data, which is quite substantial as it is higher than the suggested 60% threshold (Hair et al., 2006).

Table 2: Pearson correlation coefficients (correlation matrix).

Courses	Correlation between vectors of values									
	MAT 301	MAT 303	STA 301	STA 303	STA 399	MAT 302	MAT 304	STA 302	STA 304	STA 305
MAT 301	1									
MAT 303	0.615	1								
STA 301	0.607	0.594	1							
STA 303	0.542	0.477	0.576	1						
STA 399	0.368	0.405	0.291	0.358	1					
MAT 302	0.423	0.389	0.396	0.362	0.250	1				
MAT 304	0.553	0.364	0.259	0.301	0.320	0.605	1			
STA 302	0.460	0.406	0.379	0.465	0.386	0.536	0.597	1		
STA 304	0.415	0.324	0.357	0.423	0.395	0.489	0.444	0.618	1	
STA 305	0.364	0.370	0.320	0.350	0.404	0.473	0.528	0.601	0.562	1

Source: Result from analysis of data, 2014.

Table 3: Eigenvalues of the principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Eigen Values	4.9163	1.2926	0.8019	0.6428	0.4934	0.4564	0.4098	0.3605	0.3306	0.2956
Proportion	0.492	0.129	0.080	0.064	0.049	0.046	0.041	0.036	0.033	0.030
Cummulative	0.492	0.621	0.701	0.765	0.815	0.860	0.901	0.937	0.970	1.00

Source: Result from analysis of data, 2014.

Figure 5: Profile Plot (scree plot)

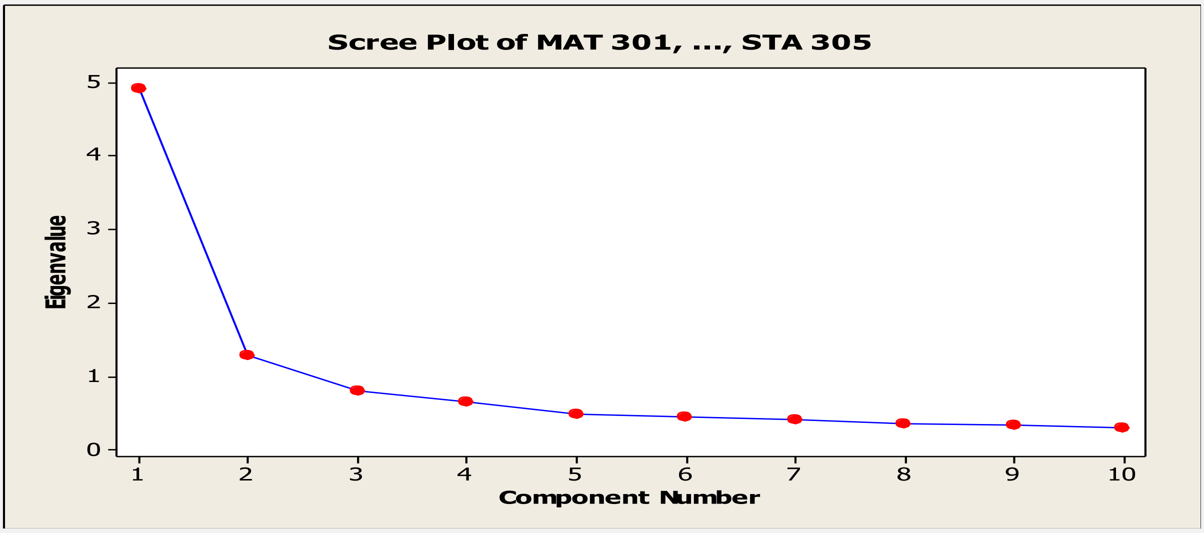


Table 4: The first three principal components and their eigenvalues.

Variable	PC1	PC2	PC3
MAT 301	-0.330	0.363	-0.076
MAT 303	-0.318	0.356	-0.024
STA301	-0.307	0.447	0.207
STA303	-0.312	0.308	0.009
STA 399	-0.261	0.019	0.800
MAT 302	-0.319	-0.233	-0.459
MAT 304	-0.304	-0.404	-0.225
STA 302	-0.355	-0.261	0.008
STA304	-0.326	-0.242	0.149
STA 305	-0.322	-0.320	0.166
Eigenvalue	4.9163	1.2926	0.8019
Proportion	0.492	0.129	0.080
Cumulative	0.492	0.621	0.701

Source: Result from analysis of data, 2014.

3.04 INTERPRETATION OF THE FIRST PRINCIPAL COMPONENT, PC1

Table 4 shows the eigenvectors (which provides the weights) of the first three principal components and their corresponding eigenvalues.

3.4.1 Computation of Relative Score Index (RSI)

Some of the weights of principal components are negative in sign and some are positive in sign. The pattern in the occurrence of the signs could have a bearing on the way a given principal component is interpreted. For PC1, all the weights are of the same sign, negative, and hardly any deduction can be made by way of interpretation relative to the scores of the course. Again some of the scores of PC1 are negative and some positive. While the scores of PC1 are interpretable relative to the original scores of the courses, the interpretation was facilitated by computing a measure that transforms the scores of PC1 into percentages, thereby eliminating the negative signs. The measure referred to as the Relative Score Index (RSI) (Nortey & Aguh, 2012) is given by

$$RSI_i = 100\left[1 - \frac{S_i - S_{min}}{S_{max} - S_{min}}\right]$$

It expresses every score as a value between 0 and 100. RSI_i is the i^{th} RSI value corresponding to i^{th} score. S_i is the i^{th} score and S_{min} and S_{max} are the respective minimum and maximum scores obtained for a principal component from the analysis. The Relative Score Index was computed based on the principal component scores estimated for each student.

Table 5 shows the principal components scores/relative score index for PC1 and their corresponding ranks for the first 30 observations. A critical examination of the PC1 scores and the RSI values reveals that the first ranked RSI value has the most negative PC1 score and last ranked RSI value has the most positive PC1 score. Thus the first ranked student has the least component score of -5.30332 and has the best course scores or grades and the last ranked (258th) student has the highest component score of 5.57678 and has the worse course scores or grades.

High course scores are associated with a negative PC1 score and low course scores are associated with a positive PCI score. Thus the better the performance of a student, the more negative the PC1 score and the higher the RSI value, and the poorer the performance of a student the more positive the PC1 score and the lower the RSI value. A PC1 score value close to zero (0) is an indication of an average performance. The PC1 score closest to zero (0) is 0.02986 which corresponds to an RSI ranking of 144. This student obtained high marks in four (4) courses and performed just averagely in the other six (6) courses. For such students the RSI values is close to 50%. The above enumerated attributes of RSI lends it as a suitable alternative to the GPA which is currently used by the Department.

Table 6 presents the classification of the performance of students based on the RSI values, using a criterion that divides the interval 0 – 100 into four equal parts, and Table 7 gives the number of student falling within each of the categories resulting from the classification presented in Table 6. From Table 7, thirty-two (32) students performed excellently, hundred and fifteen (115) had very good performance, seventy-seven (77) had good performance and thirty-four (34) students performed averagely.

3.05 INTERPRETATION OF THE SECOND PRINCIPAL COMPONENTS, PC2

It can be seen from Table 4 that the weights of the second principal component PC2 corresponding to first semester courses are positive and those for the second semester courses are negative. Thus the second principal component, PC2, serves to differentiate first semester courses from the second semester courses. For a positive PC2 score, the value of the scores for most of the first semester courses must not only be greater than their respective average scores in the semester but also greater than the scores achieved in the second semester courses. On the other hand, for a negative PC2 score, value of the scores of most of the second semester courses must not only be greater than their respective average scores in the semester but also substantially greater than the scores achieved in the first semester courses as the weights of second semester courses are lower than that of the first in magnitude. Thus, using the scores of PC2, a student can be classified as performing well in a given semester depending on the sign (negative or positive) of the student's PC2 score. For instance, the highest PC2 score of 4.95636, which is for the 214th person, indicates that the student performed better in the first semester than in the second semester. The 202nd person has a PC2 score of -4.59136, which indicates that the student did better in the second semester than in the first semester. Also, a score closer to zero (0) indicates the student has similar levels of performance in both the first and second semesters. For example the component score of -0.01115, which is for the 221st person shows the student had comparable levels of performance in both the first and second semester.

A critical look at both the original course scores and the PC2 scores suggests a classification criterion for categorizing the performance of students based on the values PC2. The classification criterion is shown in Table 8. A student with a component scores (S) less than negative one ($S < -1$) is considered as performing well in the second semester than in the first semester, while a student with a component score (S) greater than 1 ($S > 1$) is seen as performing well in the first semester than in the second semester. A student whose component score (S) falls in the range of $-1 \leq S \leq 1$ is seen as having similar levels of performance in both semesters.

Table 5: Distribution of Principal Component Score/Relative Score Index and their ranks for the first 30 observations.

OBS	PC1	RSI	RANK OF RSI
1	-3.32701	81.8355	15
2	-3.16666	80.3618	17
3	-0.79202	58.5363	104
4	-2.63511	75.4763	31
5	0.16440	49.7457	150
6	-0.09645	52.1432	137
7	-0.00380	51.2917	143
8	-1.07967	61.1801	90
9	-2.78535	76.8572	24
10	-3.91342	87.2253	8
11	-0.61101	56.8726	111
12	2.94548	24.1845	229
13	-0.21842	53.2643	130
14	-1.32983	63.4793	72
15	-1.16790	61.9910	83
16	-0.19201	53.0215	131
17	-0.43915	55.2930	120
18	-0.43100	55.2181	122
19	-0.16095	52.7361	132
20	-3.63991	84.7115	11
21	1.99770	32.8957	204
22	-1.12278	61.5763	87
23	-1.25772	62.8165	79
24	-0.72356	57.9070	108
25	-3.11170	79.8566	18
26	-1.74730	67.3164	51
27	-2.69458	76.0229	29
28	2.41450	29.0649	214
29	-1.00978	60.5377	95
30	-1.29128	63.1250	76

Source: Result from Analysis of Data, 2014.

Using the criteria set in Table 8, the number and corresponding percentage of students who would be considered as performing well in a particular semester is given in Table 9.

3.06 INTERPRETATION OF THE THIRD PRINCIPAL COMPONENT, PC3

The third principal component serves to indicate the subject (mathematics or statistics) in which a student has a better performance. This is evidenced by the fact that the weights corresponding to mathematics courses are all negative while those corresponding to the statistics courses are all positive (Table 4).

Table 6: Classification criteria for students' general performance.

Relative Score Index	Level of Performance
$75 \leq S \leq 100.00$	Excellent
$50 \leq S \leq 74.99$	Very Good
$25 \leq S \leq 49.99$	Good
$S \leq 25$	Average

Source: Result from analysis of data, 2014.

Table 7: Classification of students overall performance.

Performance	No of Students	Percentage (%)
Excellent	32	12.40
Very Good	115	44.57
Good	77	29.85
Average	34	13.18
Total	258	100.00

Source: Result from analysis of data, 2014.

Table 8: Classification criteria for students based on semester performance.

Component Score	Level of Performance
$S > 1$	High in First semester
$S < -1$	High in Second semester
$-1 \leq S \leq 1$	The similar levels of performance in both semesters

Source: Result from analysis of data, 2014.

Table 9: Classification criteria of performance of students by semester.

Performance	No of Students	Percentage (%)
Better in First Semester	38	14.7
Better in Second Semester	39	15.1
Similar levels of performance in both semesters	181	70.2
Total	258	100

Source: Result from analysis of data, 2014.

The performance in statistics is better than that in mathematics if the PC3 score is positive, while for a negative PC3 score, the performance in mathematics is better than that in statistics.

For instance, the 108th person obtained the least PC3 score of -2.21426 denoting a better performance in mathematics than in statistics. Also the 52nd person has the highest PC3 score of 2.74865 and represents a good performance in statistics than in mathematics. A PC3 value close to zero (0) reflects a situation where the performances in the mathematics courses are on par with those of the statistics courses. For example, 147th person has PC3 score of -0.02001. This student performed well in both the statistics and the mathematics courses. Thus, using PC3, a student can be classified as mathematics or statistics inclined or having comparable levels of performance in both mathematics and statistics courses.

Table 10: Classification criterion for students based on subject performance.

Component Score	Level of Performance
$S < -1$	Mathematics
$S > 1$	Statistics
$-1 \leq S \leq 1$	The similar levels of performance in Mathematics and Statistics

Source: Result from analysis of data, 2014.

Table 11: Classification of student's performance by subjects.

Performance	No of Students	Percentage (%)
Mathematics	34	13.1
Statistics	36	14.0
Mathematics and Statistics	188	72.9
Total	258	100

Source: Result from analysis of data, 2014.

A critical examination of both the scores for original courses and the corresponding PC3 scores suggests the classification criteria given in Table 10. A students with a PC3 score (S) less than negative one ($S < -1$) is deemed to be Mathematics inclined while a student with component score (S) greater than one ($S > 1$) is deemed as Statistics inclined. A student whose PC3 score (S) falls into the range

$-1 \leq S \leq 1$ is seen to perform at comparable levels in both Mathematics and Statistics courses. Thus using the criteria set in Table 10, the number and corresponding percentage of students who could be considered as performing well in either of the two subjects areas or performing at similar levels in both subjects are as given in Table 11.

4.0 CONCLUSION

The study sought to find an objective basis for assessing students' performance in the Department of Mathematics and Statistics of a public higher institution in Ghana. The multivariate tool of Principal Component Analysis proved useful in this regard. The first three principal components were retained for interpretational purposes. The study established that the overall performance of students can be assessed using RSI, a derivative of the first principal component PC1, which is a weighted sum of all courses offered. Going by the values of RSI, and dividing the interval 0 – 100 into four equal parts, the performances of the students were categorized as Excellent (32), Very Good (115), Good (77) and Average (34). It was also found that one can determine the semester, first or second, in which a student did relatively better by using the scores of the second principal component, PC2. It was found that 38 students performed relatively better in the first semester, 39 students performed relatively better in the second semester and 181 had comparable levels of performance in both semesters. The third principal component, PC3, was found to be useful in determining which subject area, mathematics or statistics, a student's academic strength lie. The numbers of students suitable for each of the subject areas of Mathematics, Statistics and the double major of Mathematics and Statistics, going by the values of PC3, are 34, 36 and 118 respectively.

We conclude from the results that most students perform at comparable levels in both semesters (one and two) and in the two subject areas of mathematics and statistics, with a few doing relatively better in either of the semesters or in either of the subject areas. The Department could use the foregoing findings to guide the students in making the choice as to which subject area to major in at level 400, the final year. Also RSI could replace GPA as it does not suffer from the weakness of assigning the same value to different levels of performance as does GPA some times. The Department could pilot RSI alongside GPA for about two (2) years to fully appreciate its behaviour and then replace GPA by RSI once it has been found to be wholesome. Barring any unforeseen problems with RSI, the replacement of GPA with RSI will encourage the student to put up their best performance as it will reflect the true performance of the students.

REFERENCES

Everitt, B. S. and Dunn, G. (2001). *Applied Multivariate Data Analysis*. Arnold/Hodder Headline Group, London, UK.

Hair, J. F. et al. (2006). *Multivariate Data Analysis*. 6th ed. Pearson Prentice Hall, Upper Saddle River, New Jersey, USA.

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. 3rd ed. Prentice-Hall, Inc, New Jersey, USA.

Nortey, E. N. N. and Aguh, N. (2012). *Re-working of Means Testing Tool Report for Student Loan Trust Fund (SLTF)*. SLTF, Accra, Ghana.

Sharma, Subhash (1996). *Applied Multivariate Techniques*. John Wiley and Sons, Inc., USA.

APPENDIX A

Table A1: Grading system for examination scores.

Exams score	Letter grade	Numerical equivalent
80-100	A	4.0
75-79	B+	3.5
70-74	B	3.0
65-69	C+	2.5
60-64	C	2.0
55-59	D+	1.5
50-54	D	1.0
<50	E	0.0

Table A2: Example on GPA computation for Student 1 and Student 2.

Student 1			**Student 2**		
	Score	**Credit**		**Score**	**Credit**
MM	**60**	**3**	**MM**	**64**	**3**
EM	**71**	**2**	**EM**	**73**	**2**
TM	**55**	**3**	**TM**	**59**	**3**
$GPA = \frac{3\times 2.0 + 2\times 3.0 + 3\times 1.5}{8} = \frac{16.5}{8} = 2.06$			$GPA = \frac{3\times 2.0 + 2\times 3.0 + 3\times 1.5}{8} = \frac{16.5}{8} = 2.06$		