International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Factor Analysis of Students Performance and Relationship Between Applied and Theoretical Courses

Lawal Mohammed Kinta[1], Lawal Adekunle Yusuf[2], Yusufu Ojodomo Emmanuel[3], Omojudi Razaq Daniel[4], Sulaiman Yunus Aduagba[5], Achanya Sunday Edward[6], Ali Audu Baidu[7], Bayedo Bartholomew Busuyi[8], Kilani Lukman Oladimeji[9]

[1]Department of Academic Planning Unit, Federal Polytechnic, Bida.
[2] Nigerian Army College of Education, Ilorin.
[3] Department of Mathematics & Statistics, Dorben Polytechnic, Bwari Abuja.
[4, 8] Department of Statistics, University of Ilorin, Ilorin.
[5]Department of Academic Planning Unit, Kwara State Polytechnic, Ilorin
[6, 9] Department of Statistics, Ahmadu Bello University Zaria.
[7] Department of Mathematics, Nigerian Army University Biu

ABSTRACT

The data used for this study was 400-level students' grade scores from the 2018/2019 session of the Department of Statistics at Ahmadu Bello University. It consists of 50 students' results in six courses of two sets (applied and theoretical courses): Applied courses consist of Regression analysis, Demography, and Econometrics; while Theoretical courses consist of Multivariate analysis, stochastic processes, and Statistical inference. The data was structured and analyzed using NCSS 2018 version. The aim of this research is to evaluate the students' academic performance by studying the relationship between applied and theoretical courses and to find out the significant contribution among the variables. The methodologies used were canonical correlation analysis to analyze the relationship between applied and theoretical courses, and factor analysis to investigate the variability among the courses and find out the variables that contribute significantly to the percentage of variance obtained. Wilk's Lambda and Bartlett's tests were obtained to respectively test the significance of canonical variate and the homogeneity of variance among the variables. The results indicated that there are fairly positive correlations among some variables as shown by the correlation matrix. It was further revealed that less-mathematical courses (applied courses) have a significant impact on determining students' academic performance. Two canonical roots were obtained and one is statistically significant showing a strong correlation between the two sets.

Keywords: correlation, factor analysis, loading factor, Academic performance, Wilk's Lambda, Bartlett's test, Applied courses and theoretical courses.

1.0 Introduction

However, as a measure of academic performance, teacher-given grades have well-known limitations. Grades are composite measures that account not only for students' content mastery but often for other factors such as their class participation, attitudes, progress over time, and attendance *(Blackorby, Wagner, Levine, Cameto, & Guzman, 2003).* This study presents canonical correlation analysis on the type of relationship that may exist between theoretical courses and applied courses. Performance indicators are a means to focus on specific expectations of a program. They facilitate the curriculum delivery strategies and assessment procedures. There is an important first step that must come before the development of performance indicators, and that is deciding on student outcomes. These are usually communicated to students in the program description, and are stated in terms that inform the students about the general purpose of the program and the expectations of the faculty. The primary difference between student outcomes and performance indicators is that student outcomes are intended to provide general information about the focus of student learning and are broadly stated of the outcome, not measurable, while performance indicators are concrete measurable performances, students must meet as indicators of achievement. Performance indicators are developed from program outcomes. This research work aims to study the relationship between applied and theoretical courses using canonical correlation procedures.

1.1 The concept of academic performance

The very concept of academic failure varies in its definition. Bonaciet *et.al* (2010) is the measurement of student achievement across various academic subjects. Teachers and education official typically measure achievement using classroom performance, graduation rate and result from standardized test. A student does not attain the expected achievement according to his or her abilities, resulting in an altered personality which affects all other aspects of life. Similarly, Tapia (2002) noted that, while the current Educational System perceives that the student fails if he or she does not pass, a more appropriate way of determining academic failure is whether the student performs below his or her potential.

Wooten (1998) undertook a study of 271 students taking introductory accounting at a major South Eastern American University of which there were 74 students equal to or older than 25 years of age identified as non-traditional, while 127 students were under 25 years of age identified as traditional. He found that, for the traditional cohort grade history, motivations and family responsibilities all influenced the amount of effort these students made. However, neither extracurricular activities nor work responsibilities influenced their effort. However, for the non-traditional students, motivation was the only variable that significantly influenced effort. Neither grade history nor extracurricular activities, work responsibilities, nor family responsibilities affected motivations. Family activities had a significant negative impact on effort for the traditional students, but not for the non-traditional students. It is conjectured by the authors of this paper that these age differences may also capture different socio-economic circumstances.

Much research has been done on common predictive factors of academic performance in accounting courses, including gender, prior knowledge of accounting, academic aptitude, mathematical background, previous working experience, age, class size, lecturer attributes and student effort, as documented by Naser, K. and Peel, M. (1998) and Koh, M. Y. and Koh H.C. (1999). The findings are not definitive.

Mc Kenzie and Schweitzer (2001) investigated academic, psychosocial, cognitive and demographic predictors of academic performance to improve interventions and support services for a student at risk of academic problems. They recommended implementing stringent record-keeping procedures at the university level to enable researchers to fully examine the relationship between age, previous academic performance and university achievement. Nonis and Hudson (2006) noted that the Higher Education Research Institute at the University of California Los Angeles UCLA's Graduate School of Education found that since 1987, the time students spend studying outside of class has declined each year, with only 47% spending six or more hours per week studying outside of class compared with 34% in 2003. This corresponds with the findings of Gose (1998) who found an increase in the number of students employed with 39% of students working 16 or more hours per week in 1998 compared with 35% working in 1993. Nonis and Hudson (2006) identified a need for empirical research to determine the impact of student work on academic performance, and its impact on the design of academic programs. Their study found a lack of evidence for a direct relationship between times spent working and academic performance.

2.0 Methodologies

This section discusses the method of canonical correlation analysis which is the approach used in this research to determine the type of relationship that exists between the performance of students in Applied and Theoretical courses.

2.1 Mathematical computation of canonical correlation analysis

Anderson (1958) gave a detailed mathematical concept of canonical correlation analysis. Let X be a q-dimensional random vector and Y be a p-dimensional random vector. Suppose that X and Y have mean μ and v respectively and that

$$E[(X-\mu)(X-\mu)'] = \sum_{XX} \quad (3.1)$$

$$E[(Y-v)(Y-v)'] = \sum_{YY} \quad (3.2)$$

$$E[(X-\mu)(Y-v)'] = \sum_{XY} \quad (3.3)$$

Let us now consider the two linear combinations

$$G = a'X = \sum_{i=1}^{n} a_i x_i \quad (3.4)$$

and

$$F = b'Y = \sum_{i=1}^{n} b_i y_i \quad (3.5)$$

Where a_i and b_i are the canonical weights to maximize the correlation between the canonical variates.
The correlation between a and b is defined in (3.6)

$$\rho(a,b) = \frac{a'\sum_{XY} b}{[(a'\sum_{XX} a)(b'\sum_{YY} b)]^{\frac{1}{2}}} \quad (3.6)$$

Our canonical variables are

$$g = a_0'\sum_{XX}^{-1/2} X \quad (3.7)$$

and

$$f = b_0'\sum_{YY}^{-1/2} Y \quad (3.8)$$

which can be used to find the canonical correlation coefficient $corr(g,f)$), which is the measure of the association between g and f.

2.2 Method of computation of canonical coefficient

Consider two sets of variables $Y = Y_{n\times p}$ and $X = X_{n\times q}$ where $p \leq q$.
The construction of the linear combinations will be as defined above in 3.4 and 3.5 such that r_{gf} is a maximum.
The XY matrix is:

$$XY = \begin{bmatrix} X_{11} & X_{1q} & Y_{11} \cdots & Y_{1p} \\ X_{21} & X_{2q} & Y_{21} & Y_{2p} \\ \vdots & & \vdots & \vdots \\ X_{n1} & X_{nq} & Y_{n1} \cdots & Y_{np} \end{bmatrix} \quad (3.9)$$

Let R denotes the correlation matrix, then the partitioned correlation matrix R is given by:

$$R = \begin{bmatrix} R_{XX} & R_{XY} \\ R_{YX} & R_{YY} \end{bmatrix} \quad (3.10)$$

Where R_{XX} and R_{YY} are the correlation matrices within X and Y respectively and $R_{YX} = R'_{XY}$ which is the correlation matrix between X and Y.

To obtain the eigenvalues and their corresponding eigenvectors, it is required to form a symmetric matrix $R_{XX}^{-\frac{1}{2}} R_{XY} R_{YY}^{-1} R_{YX} R_{XX}^{-\frac{1}{2}}$. The eigenvalues and the corresponding eigenvectors of $R_{XX}^{-\frac{1}{2}} R_{XY} R_{YY}^{-1} R_{YX} R_{XX}^{-\frac{1}{2}}$ give the canonical correlations and the canonical coefficients of the independent variable X, we use:

$$a_1 = R_{XX}^{-\frac{1}{2}} \ell_1 \quad (3.11)$$

Where ℓ_1 is the corresponding eigenvector for the independent variable.

To find the corresponding eigenvectors for Y, we use:

$$h_1 = \frac{1}{\lambda} R_{YY}^{-\frac{1}{2}} \sum_{YX} a_1 \quad (3.12)$$

We can now obtain the canonical coefficients for the dependent variable Y, by applying (3.13)

$$b_1 = \frac{1}{\lambda} \sum_{YY}^{-1} \sum_{YX} a_1 \quad (3.13)$$

3.0 Results and Discussion

This section discusses data analyses and results as follows:

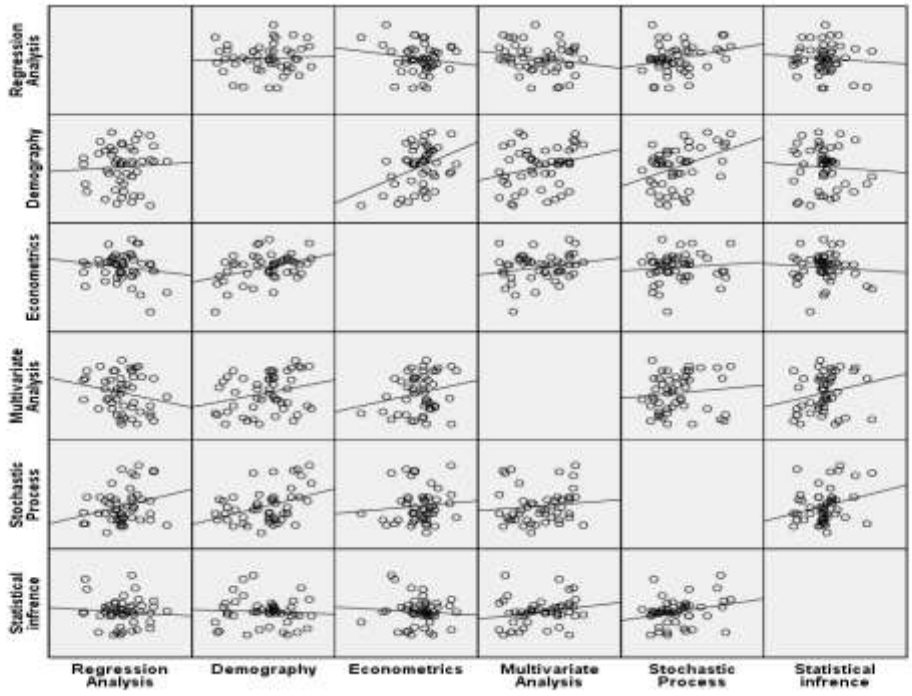

Figure 1: Correlation matrix for the set-X and set-Y

Figure 1 shows the relationship between the set of variables, X and Y using a Scatter plot. The fit line indicates the direction of the relationship and the results show that there is a positive relationship between the set of variables X and Y (i.e. both the set of variables are moving in the same direction). We also observed that there are some points close to the fitted line and some points far from the fitted line which indicates the strength of the relationship between the set of variables.

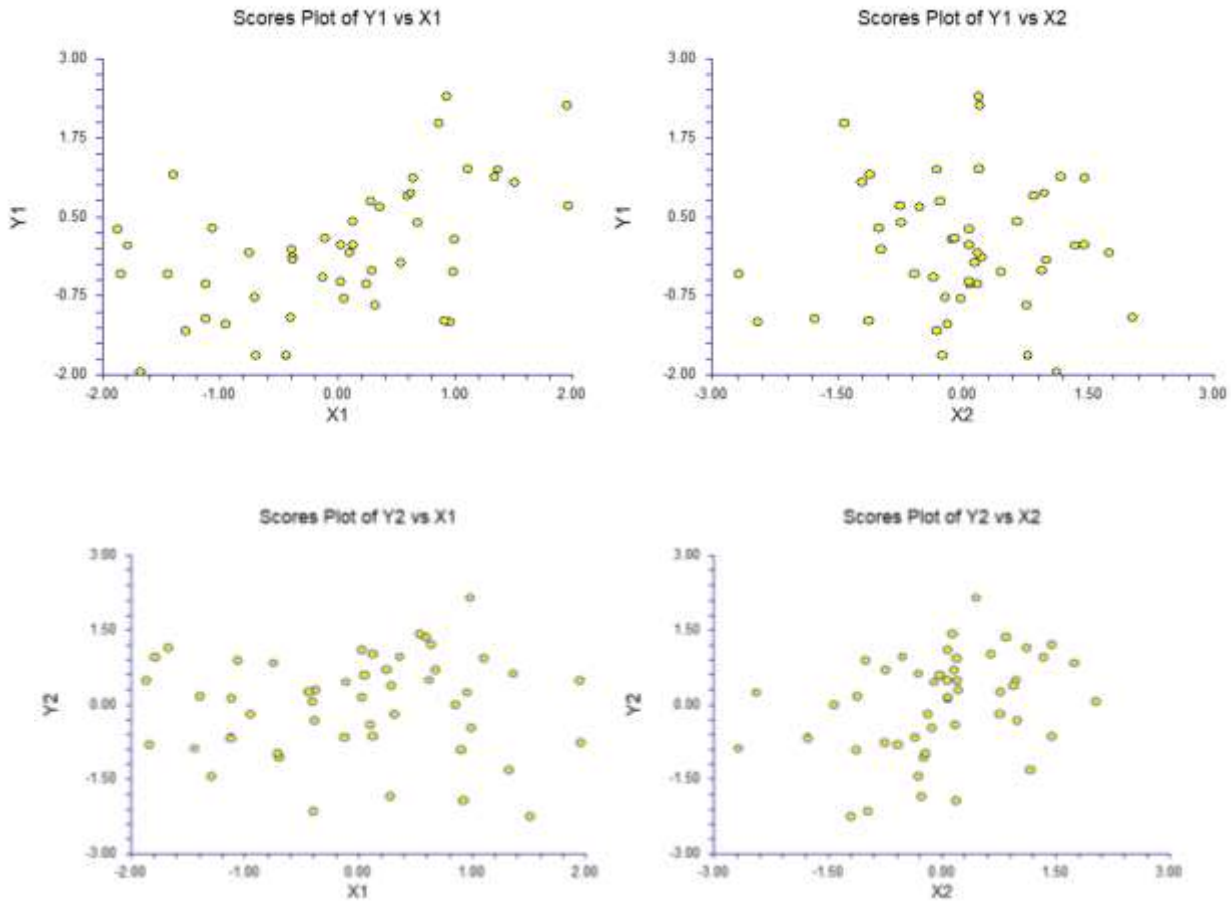

Figure 2: Score plots for canonical correlation

Figure 2 shows the relationship between each pair of canonical variates. The correlation coefficient of the data in the first plot (Y1 versus X1) is the first canonical correlation coefficient; the second plot (Y2 versus X2) is the second canonical correlation coefficient and so on. The results show that there are relationships between their pairs.

3.1 Correlation matrix for set of variables and significant of correlation

Table 1: Correlation matrix for set X and set Y

Variables	Regression Analysis (X_1)	Demography (X_2)	Econometric (X_3)	Multivariate analysis (Y_1)	Stochastic processes (Y_2)	Statistical inference (Y_3)
Regression analysis (X_1) Sig. value	1.000	0.059 0.685	– 0.157 0.280	– 0.209 0.149	0.266 0.044*	– 0.082 0.573
Demography (X_2) Sig. value		1.000	0.387 0.006**	0.269 0.042*	0.386 0.006**	– 0.059 0.686
Econometric (X_3) Sig. value			1.000	0.215 0.137	0.096 0.512	– 0.078 0.596
Multivariate analysis (Y_1) Sig. value				1.000	0.110 0.451	0.217 0.134
Stochastic processes (Y_2) Sig. value					1.000	0.258 0.046*
Statistical inference (Y_3) Sig. value						1.000

**Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the 0.05 level (2-tailed)

Table 1 shows the correlation matrix between each pair of variables. The results showed that there is significant and fairly positive correlation between stochastic process and regression analysis with (r = 0.266, P-value = 0.044), there is significant and fairly positive correlation between demography and

econometrics with (r = 0.387, P-value = 0.006), there is significant and fairly positive correlation between demography and multivariate analysis with (r = 0.269, P-value = 0.042), there is significant and fairly positive correlation between demography and stochastic process with (r = 0.386, P-value = 0.006), there is significant and fairly positive correlation between stochastic process and statistical inference with (r = 0.258, P-value = 0.046). There is negative correlation between regression analysis and each of econometric (r = – 0.157, P-value = 0.280), multivariate analysis (r = – 0.209, P-value = 0.149) and statistical inference (r = – 0.082, P-value = 0.573); but the correlation is not significantly different from zero. There exist no significant difference from zero and positive correlation between econometric and each of multivariate analysis (r = 0.215, P-value = 0.137) and stochastic processes (r = 0.096, P-value = 0.512); multivariate analysis and each of stochastic processes (r = 0.110, P-value = 0.451) and statistical inferences (r = 0.217, P-value = 0.134); and regression analysis and demography (r = 0.059, P-value = 0.685).

3.2 Canonical correlations analysis

An initial step in canonical correlation analysis is an inspection of the correlation matrix of the given data.

Let S denotes the data such that

S = {set X, set Y}

Where:

Set X = {Regression analysis, Demography and Econometrics}

Set Y = {Multivariate analysis, Stochastic processes and Statistical inference}

Table 2: Canonical correlation coefficient of set X and set Y

Canonical Function	Canonical Correlation	Eigen value	% of variance Explained
First pair of canonical variate	0.5423	0.2941	70.8
Second pair of canonical variate	0.3485	0.1214	29.2

Table 2 shows the canonical correlation of the canonical variates and their corresponding eigenvalues. Now, consider the first canonical variate pair X_1 and Y_1 with canonical correlation coefficient r_1 = 0.5423, such that the proportion of variance common to the first pair of canonical variate is r_1^2 = 0.2941 showing about 70.8% of the proportion of variance captured by the first canonical variate. Similarly, r_2 = 0.3485 is the canonical correlation coefficient between the second pair of the canonical variate, such that r_2^2 = 0.1214, which indicates about 29.2% of the proportion of variance captured.

The eigenvalues of the canonical variates can be tested by employing Wilk's Lambda criterion to test for significance, Rencher (1998).

Hypothesis: H_0: $\sum_{XY} = 0$ against H_1: $\sum_{XY} \neq 0$

Decision Rule: Reject H_0, if P-value $< \alpha$, at $\alpha = 0.05$.

Table 3: Wilk's Lambda test results

Pairs	N	P	Q	DF1	DF2	P-value
First	50	3	3	9	107	0.0079
Second	50	2	2	4	90	0.1643

In Table 3, P is the number of variables considered in a certain pair of canonical variate, Q is the number of variables considered in the opposite canonical variate and DF is the degree of freedom used at each level of canonical function. The result shows that only the first pair of canonical correlation tested is significant with P-value (0.0079) $< \alpha$ (0.05), this implies that the null hypothesis is rejected. This is an indication that one out of the two canonical correlations is significantly different from zero.

Table 4: Canonical loadings for set X and set Y

Sets	Courses	r_1	r_2
X	Regression analysis	0.4473	– 0.8688
	Demography	0.8009	0.3906
	Econometrics	0.1335	0.2215
Y	Multivariate analysis	0.2751	0.9859
	Stochastic processes	0.9352	– 0.2888
	Statistical inference	– 0.5497	0.0116

The canonical loadings in Table 4, provided information about the relative contribution of variables to each independent canonical relationship. The first pair of canonical variates can be written as follows:

$U_1 = 0.4473 * Regression\ analysis\ +\ 0.8009 * Demography\ +\ 0.1335 * Econometrics$

$V_1 = 0.2751 * Multivariate\ analysis\ +\ 0.9352 * Stochastic\ \ processes\ -0.5497 * Statistical\ inference$

The correlation $\emptyset_1 = \ 0.5423$ between $U_1$and $V_1$is called the first canonical correlation coefficient. Of the individual variable, Demography loading is the highest with the value (0.8009) followed by Regression analysis (0.4473) and Econometrics (0.1335) loading for the ordering of the criterion variables.

The values attached to each course are the correlation to their corresponding canonical variables and indicate the individual contribution to the canonical pair.

Table 5: Canonical cross loading for set X and set Y

Sets	Courses	r_1	r_2
X	Regression Analysis	0.5083	– 0.8538
	Demography	0.8957	0.4001
	Econometrics	0.4139	0.4604
Y	Multivariate Analysis	0.2483	0.9585
	Stochastic process	0.8367	– 0.1836
	Statistical inference	– 0.2714	0.1673

Table 5 shows the canonical cross loading of the two canonical functions. In the first canonical function, we discovered that Demography has the highest correlations (0.8957) with independent canonical variate. While, the weak correlation came from set Y i.e. Multivariate analysis with 0.2483.

The first canonical correlation explains the maximum relationship between the canonical variates while each successive canonical correlation is estimated to be orthogonal and yet explains the maximum relationship not accounted for by the previous canonical correlation. This reflects the high variance among these variables. By squaring the terms in the canonical loading, we find the percentage of the variance for each of the variables explained by function 1.

3.3 Factor Analysis

We wish to determine the hidden factors behind the variables to identify the natural groupings (factors that are highly correlated with each other and those that are weakly correlated with others). The correlations between the independent variables are in the range of – 0.781 to 0.501. Kaiser (1974) recommends accepting values greater than 0.5 which means the result for this research is accepted with the value of Keiser-Meyer-Olkin (KMO) to be 0.502. Bartlett's test is highly significant (P-value < 0.01) and therefore, factor analysis is appropriate for this data, to see the variables that possess high variability contribution to the set of data

Table 6: KMO statistics for sampling adequate and Bartlett's test for homogeneity

Tests	**DF**	**Approx. Chi-Square**	**P-values**
Keiser-Meyer-Olkin Measure of Sampling Adequate	–	–	0.502
Bartlett's Test of Sphericity	15	37.288	0.001

Table 7: Total variance explained

Components	Initial Eigenvalues			Extraction sum of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	1.776	29.608	29.608	1.776	29.608	29.608
2	1.374	22.893	52.501	1.374	22.893	52.501
3	1.232	20.536	73.037	1.232	20.536	73.037
4	0.659	10.990	84.027	0.659	10.990	84.027
5	0.548	9.132	93.159	0.548	9.132	93.159
6	0.410	6.841	100.000			

Table 7 lists the eigenvalues associated with each linear component (factor) before extraction, after extraction and after rotation. Before extraction, it has identified six (6) linear components within the data set. The eigenvalues associated with each factor represent the variance explained by the particular linear component and also displays their eigenvalues in term of the percentage of variance explained (so, factor 1 explains 29.608% of total variance). Principle Component Analysis (PCA) extracts all factors with eigenvalues greater than 0.5. In this study, we used the common decision in which we retain only the factor with about 93.159% of variance explained. Therefore, from the extraction sum of squared loading column, we observed that five factors are retained together with their percentage of variance explained by each factor. The cumulative variance given as well shows that the first five factors accounted for 93.159% of the total variance in the data. A factor's eigenvalue may be computed as the sum of its squared factor loadings for the entire variable (Rencher, 2002).

Table 8: Communalities extracted by each variable

Variables	Initial	Extraction
Regression analysis	1.000	0.984

Demography	1.000	0.853
Econometrics	1.000	0.995
Multivariate analysis	1.000	0.982
Stochastic processes	1.000	0.851
Statistical inference	1.000	0.924

Table 8 shows the communalities which measure the percentage of variance explained by all the components. That is, communality is the squared multiple correlations for the variable using the components as predictors. Communalities for variables are the sum of squared components loadings for that variable (row) and are the per cent of variance due to the variable explained by all the components. For full orthogonal factor analysis, the communality will be 1.0 and all the variance in the variable will be explained by all the factors, with their number equal to that of the variables and is written under initial. Extraction communalities are estimates of the variance in each variable accounted for by the components. From Table 8, it can be seen that all the courses are well represented because all variables extracted are high. If any communality is very low in the extraction of a principal component, you may need to extract another component.

Conclusion

Canonical correlation analysis was employed in this study to measure the strength of the relationship of the canonical pairs and to identify the courses that contributed strongly. The canonical correlation analysis generated three correlation coefficients, which were tested and found one of the correlations statistically different from zero. A set of weights for each of theoretical and applied courses were determined so that the linear combination of each set is maximally correlated and explained the nature of whatever relationship exists between the two variable sets. It was revealed in Table 2 that the measure of correlation of the first pair is 0.2941 with a variability proportion of about 70.8%, whereas, the second pair had 0.1214 as its measure of the correlation with the variability proportion of about 29.2%. Hence, the total variability captured by the two canonical pairs is 100%. The 70.8% variability is due to the individual contribution of the composites of demography, regression analysis, econometrics, multivariate analysis, stochastic processes and statistical inference.

Factor analysis was also applied and showed four groups of closely inter-related courses based on the fact that four factors were used which indicates variable reduction. The strongest inter-related courses are found in the beginning column and decrease through the last column.

Therefore, we conclude that the student's performance in the applied courses has influence on their performance in the theoretical courses. It was clearly shown that set-X and set-Y are were directly related, that is, an increase in performance of students in theoretical courses resulted to an increase in performance of students in applied courses.

References

Anderson, (1958). Relations between Two Sets of Variables, Biometrika, 28:312 - 377

Blackorby, Wagner, Levine, Cameto, & Guzman (2003). Canonical Correlation Analysis of Data on Human Automation Interaction, Proceeding of the 41st Annual Meeting of the Human Factors and Ergnomics Society.

Bonaciet et.al (2010). Influential Factors of Accounting Students' Academic Performance: A Romanian Case Study, Accounting and Management Information Systems; 9(4), 558-580.

Gose, B. (1998). Tutoring Companies Take over Remedial Teaching at Some Colleges: Can Kaplan and Sylvan Help Students Erase Educational Deficiencies More Quickly? The Chronicle of Higher Education, 35 (12).

Koh, M. Y. and Koh H.C. (1999). These Determinants of Performance in an Accountancy

Degrees Programme Accounting Education 8(1), 13-29.

Mc Kenzie, K. and R. Schweitzer (2001). Who succeeds at university? Factors predicting academic performance in first-year Australian university students. Higher Educ. Res. Dev., 20: 21-33.

Naser, K. and Peel, M. (1998). An exploratory study of the impact of intervening variables on student performance in a Principles of Accounting Course. Accounting Education, 7(3): 209-223.

Nonis, S. and Hudson, G. (2006). "Academic performance of college students; the influence of time spent studying and working" Journal ofEducation for Business 81(3), 151- 160.

Wooten (1998). Non-Parametric and Statistical Method based on ranks, Englewood Cliffs.

Tapia (2002). Common Factor Analysis versus Principal Component Analysis, University of California, Los Angeles.