Principal Component Analysis of Students' Academic Performance in Mathematics and Statistics

Author's Details:
[(1)]**E. A Aboagye-** Department of Mathematics and Statistics, Takoradi polytechnic P. O.BOX 256, Takoradi Ghana- West Africa [(2)]**C. A Mensah-**Department of Mathematics and Statistics Takoradi Polytechnic P. O.BOX 256, Takoradi Ghana- West Africa

Abstract:
This study seeks to identify a basis of assessments of students' performance in the Department of Mathematics and Statistics of University of Cape Coast. Data on level 300 students in the Department of Mathematics and Statistics of the 2013/2014 academic year was obtained. These covered ten courses, four of which are mathematics courses and the remaining six (6), statistics. The ten subjects served as variables to be studied each with several observations that are the grades of students in the various courses. Principal Component Analysis was used to analyse the data. This technique is used since the Principal components generated could serve as indices of measuring the students' performance. From the analysis, three principal components were retained as rules or indices for classification of students' performance. The first principal component was used to classify overall performance of students as good, average, below average or excellent. The second principal component was used to classify students on semester basis. That is whether or not the students perform well in either of the semester or both. It was observed that the third principal component can be used to classify the performance of students on subjects' basis. This can either be mathematics or statistics inclined. Thus using the three principal components it was further observed that majority of students exhibited uniform but just average performance in both subjects and in both semesters. A few showed specific strengths in either of the two subjects. This study confirms the notion that only a few students could major in only one subject in the department with greater success.
Key Words: *Principal Component Analysis,Academic Performance, Mathematics and Statistics*

1. Introduction

.

Determinants of students' performance have been the subject of ongoing debate among educators, academics, and policy makers. There have been many studies that sought to examine this issue and the findings of these studies point out to hard work and discipline, previous schooling, parents' education, family income and self-motivation as factors that can explain differences in students' grades. For example, Siegfried and Fels (1979) concluded that the student's attitude is the most important determinant of his/her learning. In a study of high school students who are in an Economics class and want to take another Economics course, Beron (1990) found that there was a link between the perceived usefulness of an additional course in Economics and the performance of the students in a current Economics course. Gender wise, Williams'*et.al* (1992) found no evidence to support the hypothesis that significant and consistent gender differences exist in college students' performance on Economic examinations. Romer (1993) found that class attendance is reflected significantly on the students' GPA. Anderson and Benjamin (1994) found that the most important factors that affect students' performance in university introductory Economics course were the overall achievement level and taking a course in Calculus. With regard to gender, they found that male students outperform their female counterparts. Kennedy and Tay (1994) concluded in their survey that the research on the factors affecting students' performance in Economics points out to student's attitude as the most important determinant of learning. Study effort, age of student, and a good match between student's learning style and instructor's teaching style all have positive effect on student's performance.

Cohn *et al* (1995) found that memory and note-taking affect learning in the introductory courses in Economics. Devadoss and Foltz (1996) studied the effects of previous GPA, class attendance and financial status on the performance of students of some Agriculture Economics related courses. They concluded that previous GPA and motivation affected positively the current GPA. They also found that students who supported themselves financially were likely to have better performance. Zimmer and Fuller (1996) in their survey of the factors affecting students' performance in statistics found that statistics anxiety and attitude, and computer experience were linked to students' performance in statistics courses. A multivariate analysis of students' performance in large Engineering classes was conducted by Sullivan *et.al* (1996) at Virginia Tech. The study was aimed at describing statistics results by the application of multiple linear regression to students' records and performance. The final weighted score was the dependent variable. The independent variables include gender, academic level, grade point average, SAT Maths score, SAT verbal score and high school class standing. Further delineations regarding particular engineering major, morning versus afternoon session and instructor were also made in the student records data base. Linear regression was discovered to provide the most accurate predictions of the final weighted score in the Engineering Economy course. Because of the greater variability in the final examination scores, it was determined that the final weighted score was the more appropriate dependent variable to use.

A web based data analysis decision support tool titled "Concert Inform" from Pearson Digital Learning provided the following outline for assessing performance.

The need for information that yields a better understanding of students and schools; provision of a comparative analysis of students achievement and classroom performance across a district or school; accessibility to a list of students in any performance category and see the number and percentage of students above, below or attaining proficiency. This is more convenient in this computer age; a tool can be helpful in analyzing students' performance. However, since the required tools fo various objectives were not specified, researchers wgo do not have knowledge in statistics may misapply the statistical tools. Also those who do not have access to internet facility may not be capable of employing his outline.

The University of Cape Coast has, in recent years, introduced new and innovative programmes. Accordingly, various departments have been modifying their existing course structures to meet the current trend. One of such departments is the Departments of Mathematics and Statistics. The department in recent years has introduced programmes such as Actuarial science, Business with Mathematics, Statistics, Mathematics and Statistics. There are six different programmes available to students in the department. Due to the introduction of these new programmes, the department has witnessed an unusual increase in students' enrolments, just as almost all other departments have had their fair share of increasing student numbers. Some students are offered direct admission into the programmes in the department while others are offered admission as physical sciences students. Those who are offered admission as Physical sciences students may end up completing their degrees programmes in Physics, Chemistry or Mathematics and Statistics Departments. Those who opt for the Department of Mathematics and Statistics may further major in Mathematics only, or Statistics only, or in combined Mathematics and Statistics.

Most often students opt for certain programmes based on their performance in the area. They often tend to offer programmes in which they think they perform well. Thus this study seeks to identify relative strength of students in semester and subject basis.

The following objectives have been set:

- To classify the general performance of students
- To classify the relative performance of students by subjects
- To classify the relative performance of students on semester basis.

It focuses on performance of students in mathematics and statistics. This study will afford the students in the Mathematics and statistics department the opportunity to assess themselves well in deciding on which subject area to major in.

2. Methodology

The data for the study were secondary data obtained from the Department of Mathematics and Statistics. The data consist of Level 300 results for the first and second semesters of the 2013/2014 academic year. Level 300 courses were chosen since it is the performance in Level 300 courses which are normally used by students to finally decide on which subject area to major in. In the Department of Mathematics and Statistics, the courses have been coded. The code is in two parts: the letter part and the number part. The letter usually consists of the first three letters of the subject. We have MAT representing mathematics subjects and STA representing statistics subjects. The number part also consists of three digits. The first digit to the left represent the level of the course. For example, 301 representing a level 300 code. The numbers are assigned such that the odd number codes usually represent courses taken in the first semester while the even number codes usually represent courses in the second semester.

Also some of the courses have been grouped into two parts and taken in two semesters. A course like Advanced Calculus is taken in two parts in level 300. Advanced Calculus 1 is taken in first semester and Advanced Calculus 2 taken in second semester of Level 300. However, courses like Data Analysis, Modern Algebra and Statistical Methods are taken in two parts in both Levels 300 and 400. Data Analysis 1, Modern Algebra 1 and Statistical Methods 1 are taken in the Level 300 while Data Analysis 2, Modern Algebra 2 and Statistical Methods are taken in the Level 400. Thus, in this study the following courses are the variables under study and it has been grouped into first and second semesters. The first semester courses are as follows:

MAT 301: Advanced Calculus1
MAT 303: Introductory Analysis
STA 301: Probability Distributions
STA 303: Statistical Methods 1
STA 399: Research Methods

The second semester courses are as follows:

MAT 302: Advanced Calculus 2
MAT 304: Modern Algebra 1
STA 302: Sampling technique and Survey Methods
STA 304: Design and Analysis of Experiment
STA 305: Data Analysis 1

A sample size of one hundred and eighty –three students (183) was used and the source of the data was secondary. Minitab software was used for the analysis of data.

2.1 Principal Component Analysis

Principal component analysis is a technique for forming new variables which are linear combinations of the original variables. The new variables are uncorrelated among themselves. The maximum number of new variables that can be formed is equal to the number of original variables.

2.2 Analytical approach

Consider the original random vector $X = (x_1, x_2, \cdots, x_p)$ having covariance matrix Σ. Let $y_1, y_2, \cdots, y_p$ represent the linear combinations of the original variables. Then

$$y_1 = w_{11}x_1 + w_{12}x_2 + \cdots + w_{1p}x_p$$
$$y_2 = w_{21}x_1 + w_{22}x_2 \cdots + w_{2p}x_p$$
$$\vdots$$
$$y_p = w_{p1}x_1 + w_{p2}x_2 + \cdots + w_{pp}x_p$$

Where w_{ij} is the weight of the j^{th} variables where $(j = 1, 2, \cdots, p)$ for the i^{th} new variable and is the linear combination of all the original variables.

Assuming that there are p variables, we are interested in forming the following p linear combinations:

$$\sum_{j=1}^{p} w_{ij}^2 = 1 \qquad \sum_i \sum_j w_i w_j = 0 \qquad (2.1)$$

This condition is used to fix the scale of the new variables in order to increase the variance of a linear combination by changing the scale of the weights. The second condition ensures that the new axes are orthogonal to each other. The matrix form of equation (2.1) such that those conditions specified above are satisfied.

Geometrically, the objective of principal component analysis is to identify a new set of orthogonal axes such that

- The coordinates of the observations with respect to each of the axes give the values for the new variables. The new axes or the variables are called principal components and the values of the new variables are called principal components scores.
- Each new variable is a linear combination of the original variables.
- The first new variable accounts for the maximum variance in the data.
- The second new variable accounts for the maximum variance that has not been accounted for by the first variable.
- The third new variable accounts for the maximum variance that has not been accounted for by the first two variables.
- Generally, the pth new variable accounts for the variance that has not been accounted for by the first $p-1$ variables.

Now if a substantial amount of the total variance in the data is accounted for by a few (m) principal components, then the researcher can use these few principal components for interpretational purposes or in further analysis of the data instead of the original p variables. This would result in a substantial amount of data reduction if the values of p is large. Thus, principal components analysis is a data reduction technique.

2.3 Interpreting Principal Components

Since the principal components are linear combinations of the original variables, it is often necessary to interpret or provide a meaning to the linear combination. One way of interpreting is to use the loadings of the principal components. The loadings are given by $l_{ij} = \frac{w_{ij}}{\hat{s}_j}\sqrt{\lambda_i}$ where l_{ij} is the loading of the j^{th} variable for the i^{th} principal components, w_{ij} is the weight of the j^{th} variable for the i^{th} principal component, λ_i is the eigenvalue or variance of the i^{th} principal components and $\hat{s}_j$ is the standard deviation of the j^{th} variable. The higher the loading of a variable, the more influence it has in the formation of the principal components score and vice versa. Therefore, one can use the loading to determine which variables are influential in the formation of principal components. Based on these loadings one can assign a meaning or label to the principal component. The objective of principal component as listed on page eight (8) is a guide to interpretation of the principal components. The first principal component accounts for the maximum variation in the data. The second principal component accounts for the maximum variation in the data that has not been accounted for by the first principal components. Then the p^{th} principal component accounts for the remaining variation in the data that has not been accounted for by the $p-1$ principal components.

However, it may not be of practical use to interpret all p new components. A number of rules have been proposed (Zwick and Veliicer, 1986) to aid in retaining the first few principal components. These are

The principal components with eigenvalues greater than one are considered influential

Given that the principal components are orthogonal, a high loading (w_{ij}) of a variable (x_j) on a particular component y_i is an indication of the variable's importance in the formation of principal component score.

A scree plot of eigenvalues against number of components reveals the influential components that can be retained.

The final interpretation of the few principal components retained is then mainly dependent on the object – matter sense of the study.

2.4 Principal Component Loadings

The simple correlation between the original variables and the principal components are called loadings. They give an indication of the extent to which the original variables are influential in forming the principal components. That is, the higher the magnitude of the loading the more influential the variable is in forming the principal components score and vice versa. When the correlation matrix is used for the eigenanalysis, the principal component loadings is defined as follows:

$$s_{ij} = w_{ij}\sqrt{\lambda_i}$$

Where s_{ij} the correlation between the i^{th} principal component and the j^{th} variable , w_{ij} is the principal component weight of the j^{th} variable in the i^{th} principal component; λ_i is the eigenvalue associated with the i^{th} principal component. The range of the loadings is between -1 and 1.

An important decision in principal component analysis is determining which principal component loadings are worth considering in the interpretation of each component. These are some general rules to consider.

The firs rule is not based on any mathematical proposition; it is a rule of thumb frequently employed by researchers. Principal component loadings greater than 0.30 or less than -0.30 are considered significant. Loadings greater than 0.40 or less than -0.40 are considered more important and loadings greater than 0.50 or less than – 0.50 are considered very significant. Thus, the larger the absolute size of the loadings, the more significant the loading is in interpreting the principal component structure. These guidelines are considered useful when the sample size is greater than or equal to 50 Sharma (1996). He is also of the view that a loading of 0.5 or above as a cutoff point is appropriate Hair et.al (1987). There are views that loadings greater than 0.32 or less than – 0.32 are poor, greater than 0.45 or less than – 0.45 are fair, greater than 0.55 or less than – 0.55 are good, greater than 0.63 or less than – 0.63 are very good and greater than 0.71 or less than – 0.71 are excellent. These benchmarks are intuitively appealing because they are tied to the percentage of variance accounted for by the component; specifically these benchmarks identify the variables that account, respectively, for more than 10%, 20%, 30%, 40% and 50% of the variance in the component. Tabachnik and Fidell (1989).

Hair et.al (1987) recommend loadings greater than or equal to ± 0.19 and ± 0.26 for the 5 percent and 1 percent significance levels, respectively, when the sample is 100. When the sample size is 200, ± 0.14 and ± 0.18 are recommended for the 5% and 1% levels and when the sample size is greater than or equal to 300, loadings of ± 0.11 and ± 0.15 are recommended.

The major setback in applying these rules is that the dimension of the data set being analyzed and the specific components being examined are not taken into consideration. To address this, the following general guidelines are recommended.

The larger the sample size, the smaller the loadings to be considered significant;

The larger the number of variables to be analyzed, the smaller the loading to be considered significant; and

The larger the number of components, the larger the size of the loading on later components to be considered significant for interpretation.

3. Results

The analysis of this study is in two parts:; preliminary analysis and further analysis. The preliminary analysis consist eigen analysis and scree plot.

3.1 Eigenanalysis

Table 1: Eigenvalues

Eigen Value	5.3919	1.0341	0.7947	0.6494	0.4676	0.4073	0.3791	0.3190	0.2887	0.2681
Proportion	0.539	0.103	0.079	0.065	0.047	0.041	0.038	0.032	0.029	0.027
Cumulative	0.539	0.643	0.722	0.787	0.834	0.875	0.912	0.944	0.973	1.000

The table 1 above presents Eigen analysis of the ten courses. Eigen values are weights of principal components obtained, Eigen values greater than one rule state that when the correlation matrix is used for Eigen analysis, only values greater than one should be included in the analysis. Using Eigen values greater than one rule, it can be seen that two principal components should be retained. The number of components to be retained is further supported with the scree plot below.

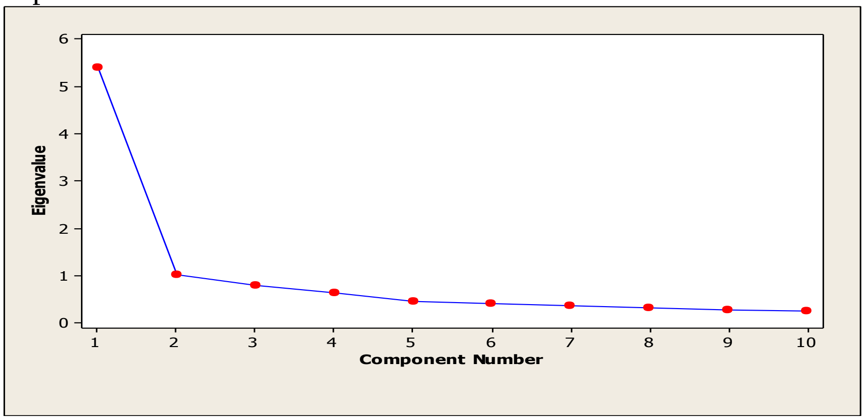

Figure 1: scree plot

The scree plot is a plot of eigen values against the number of components. The rule here is to examine the plot for an "elbow". From the scree plot the elbow suggests that the number of principal components that can be retained for further analysis is two (2).

3.2 Further Analysis

The students' performance is further analyzed using principal components.

3.2.1Principal component analysis

From the principal component analysis results, we have ten principal components produced since there are ten variables involved. It will be recalled that both the Eigen analysis and the scree plot suggested that two components be retained but in this work, three principal components will be retained for further analysis due to the relevance of these components to the topic under discussion

Table 2: principal components and their eigen values

VARIABLE	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
MAT 301	0.335	-0.308	0.177	-0.165	-0.169	0.389	-0.296	-0.639	0.088	-0.224
MAT 303	0.338	-0.325	0.100	-0.244	-0.253	-0.026	-0.219	0.529	0.466	0.318
STA 301	0.333	-0.324	0.246	-0.006	0.143	0.199	0.561	0.331	-0.267	-0.410
STA 303	0.288	-0.415	0.005	0.569	0.019	-0.589	-0.014	-0.213	-0.087	0.144
STA 399	0.244	-0.189	-0.793	-0.352	0.296	-0.005	0.172	-0.124	-0.035	0.137
MAT 302	0.333	0.290	0.368	-0.208	0.099	0.034	0.223	-0.192	-0.302	0.662
MAT 304	0.315	0.357	0.186	-0329	0.348	-0.519	-0.136	-0.040	0.229	-0.409
STA 302	0.347	0.187	-0.156	0.207	0.121	0.199	-0.604	0.318	-0.503	-0.067
STA 304	0.311	0.339	-0.108	0.521	0.221	0.360	0.184	-0.030	0.538	0.042
STA 305	0.304	0.353	-0.252	0.002	-0.778	-0.150	0.229	-0.027	-0.096	-0.172
EIGEN VALUE	5.3919	1.0341	0.7947	0.6494	0.4676	0.4073	0.3791	0.3190	0.2887	0.2681
PROPORTION	0.539	0.103	0.079	0.065	0.047	0.041	0.038	0.032	0.029	0.027
CUMMULATIVE	0.539	0.643	0.722	0.787	0.834	0.875	0.912	0.944	0.973	1.000

3.2.2 Interpretation of the principal components

It can be seen from the table that the first principal component, PC1, accounts for 53.9% of the total variance in the data. It can be seen that the first principal component, PC1 is a weighed sum of all the various courses. The second principal component account for 10.3% of the total variance in the data and together with the first principal component accounts for 64.2%. it can be seen from the second principal component that the coefficient of variables representing the first semester courses are positive while those for the second semester courses are negative. Thus on the second principal component, the first semester courses contrast with the second semester courses. The third principal component accounts for 7.9% of the total variation in the data. This principal component together with the first and second principal components accounts for 72.2% of the total variance in the data. It is evident that the third principal component is contrast between subjects. That is, the mathematics courses contrast with the statistics subjects' courses. This is so since the coefficients of variables representing the mathematics courses are all positive while those representing the statistics courses are all negative. The first three components have been retained in the research due to their relevance to the topic under discussion.

3.2.3Computation of the principal components scores

Table 3 shows the principal component scores for the first three components of the first 10 students.

Table 3: principal component scores

Observation	PC1	PC2	PC3
1.	1.08083	-1.53413	-2.28176
2	-3.14939	0.41027	-0.83602
3	0.93748	-1.33499	0.75925
4	-2.50944	-0.25828	-0.97521
5	0.77798	-0.0807	-0.51187
6	1.63682	-0.36891	0.82195
7	0.62021	0.46556	0.159
8	-0.86827	-1.06317	-0.69156
9	-2.70839	0.39206	-0.0414
10	-3.49559	1.71897	0.04552

A critical examination of the component scores for PC1 reveals that the first rank value is for the least negative score and highest ranked is for the score with the highest positive score. Thus for any set of values of the courses, PC1 can result in a very small negative number if the values of the scores are large. Also a very small component score represent an excellent performance or very high scores in all courses while a very high component score represent a very low score in all the courses, and hence a poor performance. Also for PC1, a component score closer to 0 (or a zero score) means the student performs very well in some courses, and in the remaining courses, performs below average. Thus, the first ranked student has the least component of -5.4663 and corresponds to the student with the best grades. Similarly, the last ranked (180) student has a component score of 7.2743 and represent the student with the least grades in most courses. The component closer to zero (0) is 0.0199. This represents a student ranked 96. From the original grades, this student obtained high marks in a few (3) of the courses and performed just averagely in other courses. Thus a critical examination of the original grades and their corresponding component suggest a classification in table 4.

Table 4: Classification criteria for students s' general performance

Component score(s)	Level of performance
$S < -4$	Excellent
$-4 \le S \le -2$	Good
$-2 \le S \le 3$	Average
$S \ge 3$	Below average

From table 4, it can be seen that a student with a component score (S) less 4 is deemed as an excellent while a student with a component score greater than 3 is seen as performing below average. Based on the criteria set above, classification of students overall performance have been done in table 5.

Table 5: Classification of students overall performance

performance	No of students	Percentage (%)
Excellent	6	3
Good	48	27
Average	116	64
Below average	10	6

Table 5 shows that for the students understudy, only 3% of them can be classified as excellent while 6% of them too can be classified as performing below average. Majority (64%) of the students can be classify as performing averagely.
Similarly, critical examination of the component scores for PC2 and PC3, suggest the following classification criteria which can be used to classify students based on semester and subject performances in tables 6 and 7.

Table 6: Classification of performance of students by semester

Performance per semester	No of students	Percentage
First semester	18	10
Second semester	22	12
Both semester	140	78

From table 6, it can be seen that 78% of the students performs relatively equal in both semesters with 10% of the being first semester performers. 12% of them were found to be secod semester performers.

Table 7: Classification of students by subjects

Area of strength	No of students	Percentage (%)
Mathematics	23	13
Statistics	22	12
Mathematics and statistics	135	75

Table 7 shows that 13% of the students are mathematics incline while 12% of them are statistics incline with the majority (75%) of them being uniform(that is have equal strength in both courses). This implies that only a handful of the students can major in a single subject.

4. Discussion

From the analysis, it was realized that the general performance of most of the students was average and this is reflected in both the first and second semesters. Special mention should be made of students who are generally good in both semesters but show no with specific strength. These students usually perform uniformly well in both subjects. Thus the perception that most students tend to major in subjects areas they think they perform well may be true as most of these students under study majored in combined Mathematics and Statistics. This is so because the evaluation of the individual performance shows that most students perform uniformly (almost the same level of performance) in both mathematics and statistics courses. It is therefore justifiable that the number of students who opt for single major should be few. The identification of three main dimensions for assessing students' performance conforms to facilities available in the web-based analysis tool "concert inform". This is so because a student can be categorize by general ability, semester performance or subject area, using the three principal components. By this, specific information can be obtained on any student regarding his or her academic performance.
It was further observed that twenty two (22) students have inclinations towards statistics while twenty three (23) of them have inclinations towards mathematics. These observations show that only a few of the students could

specialize in a single subject with greater success. The remaining majority can specialize in a combination of the two subjects. However, only a few of this large number is expected to excel in a combined major.

5. Conclusion

The focus of this research is to find the bases of assessment of students' performance in the Department of Mathematics and Statistics. Thus the following conclusions were drawn from the research:

- It was found that majority (64%) of the students have been identified to be performing just averagely.
- It was also found that students can be classified by semester basis. It was realized that performance of most of the students in both the first and second semesters are uniform. That is, if the performance of a student is good or otherwise, it is often reflected in both semesters.
- Finally performance of students can be classified on subject basis. Thu, the performance of students can be classified according to their specific strength in either mathematics or statistics. It was found that, most students tend to have almost equal strengths in both subjects. That is most of them tend to have similar performance in both mathematics and statistics.

6. Recommendation

From the results base on this study, it is recommended that only a few of the students should be allowed to specialize in only a single subject while majority should be allowed to specialize in a combination of subjects..

References

1. Anderson, G., and D. Benjamin. (1994). The determinants of success in university introductory economics courses. *Journal of Economic Education* 25(2): 99-119.

2. Beron, K. (1990). Joint determination of current classroom performance and additional economics classes: A Binary/continuous model. *Journal of Economic Education* 21(3): 255-264.
3. Cohn, E., S. Cohn, and J. Bradley. 1995. Note taking, working memory, and learning in principles of economics. *Journal of Economic Education* 26(4): 291-308.

4. Devadoss, S., and J. Foltz. (1996). Evaluation of factors influencing students attendance and performance. *American Journal of Agricultural Economics* August: 499-507.

5. Hair,J.F et al.(2006). *Multivariate Data Analysis*. 6^{th} ed. Upper Saddle River, New Jersey, USA Prentices –Hall
6. Kennedy, P., and Tay, R. (1994). Students' performance in economics: Does the norm hold across cultural and institutional settings? *Journal of Economic Education* 25(4): 291-301.
7. Pearson, J. " Concert Inform" *www.Pearsondigital.com/pres/pres/2003/pr-1106*
8. Romer, D. (1993). Do Students go to class? Should they? *Journal of Economic Perspectives* 7 (3): 167-74.
9. Sharma S(1996). *Applied Multivariate Techniques*; John Wiley & Sons, Inc. pp 58-70, 84-87

10. Siegfried, J., and R. Fels. (1979). Research on teaching college economics: A Survey. *Journal of Economic Literature*: 17(3): 923-69.

11. Sullivan, W.G, Daghestani, S.F.(1996). Multivariate Analysis of Large Engineering Economy Classes, *proceedings of the ASEE Annual Conference-succeed.ufl.edu.*

12. Tabachnick, B.G and Fidell, L.S. (1989). Using Multivariate Statistics, 3^{rd} ed. New York: Harper and Row.

13. Williams, M., L. Waldauer, C. Duggal, and G. Vijaya. 1992. Gender differences in economic knowledge: An extension of the analysis. *Journal of Economic Education* 23(3): 219- 231.

14. Zimmer, J., and D. Fuller. 1996. Factors affecting undergraduate performance in statistics: A Review of Literature. Paper presented at the Annual Meeting of the Mid- South Educational Research Association (Tuscaloosa, AL, November 1996). The ERIC Database 1992-2003.ED406424

15. Zwick, W.R, and Veliicer, W.F: Comparison of five rules for determining the number of components to retain, *Psychological Bulletin*, **99(3)**, pp 432-442