PRINT ISSN 1119-8362
Electronic ISSN 2659-1502

Full-text Available Online at
https://www.ajol.info/index.php/jasem
https://www.bioline.org.br/ja

J. Appl. Sci. Environ. Manage.
Vol. 27 (10) 2355-2359 October 2023

Principal Component Analysis for Investigating the Relationship between the Semester Results and Academic Performance of Students in a Polytechnic in Niger State, Nigeria

AHMED, SS; YISA, EM; JIBRIN, M; YAHAYA, G

Department of Mathematics and Statistics, School of Natural and Applied Sciences (SNAS), Niger State Polytechnic, Zungeru, Niger State, Nigeria

**Corresponding Author Email: ahmedsule710@gmail.com; Tel: 08036154396*
Co-Authors Email: emmyyisa74@gmail.com; elhajee24@gmail.com; yahayagana963@gmail.com

ABSTRACT: Factor analysis permits the ability to simplify a set of complex variables using statistical procedures to explore the underlying dimensions that explain the relationships between the multiple variables. This paper therefore used factor analysis to investigate the relationship between the semester results and academic performance of students at a centre for continuing education and training in a Polytechnic in Niger State, Nigeria using principal component method after collecting data from 57 students. The findings of this study revealed that there is weak linear relationship between the variables. From the total variance explained table, 4 factors were extracted which accounted for 60.7% and the remaining factors only accounted for 39.3% of variation. Data obtained observed that Component 1 loads high on BAM225=0.625, Component 2 loaded as high as 0.549 for STP213, while Component 3 loaded for GNS201= 0.681 and Component 4 loads on GLTVII = 0.586. The results could signify the existence of these factors significantly contributing to the academic performance of students in polytechnic.

DOI: https://dx.doi.org/10.4314/jasem.v27i10.29

Cite this paper as: AHMED, S. S; YISA, E. M; JIBRIN, M; YAHAYA, G. (2023). Principal Component Analysis for Investigating the Relationship between the Semester Results and Academic Performance of Students in a Polytechnic in Niger State, Nigeria. *J. Appl. Sci. Environ. Manage.* 27 (10) 2355-2359

Received: 27 September 2023; Revised: 12 October 2023; Accepted: 24 October 2023 Published: 30 October 2023

Keyword: Factor Analysis; Principal Component Analysis; Statistical Significance; Academic; Performance

Factor analysis studies the variability among observed correlated variables in terms of a potentially lower number of unobserved factors that are supposed to contain essential information in a larger set of observed variables. The observed variables are modeled as linear combinations of the potential factors plus "error" terms, hence factor analysis can be thought of as a special case of errors-in-variables models (Bandalos, 2018). Factor is commonly used in market research, as well as other disciplines like Technology, Medicine, Sociology, Field Biology, Education, Psychology and many more. The broad purpose of factor analysis is to summarize data so that relationships and patterns can be easily interpreted and understood. In factor analysis, observed variables are being represented as the linear combination of the unobserved factor (Yong and Pearce, 2013). Pimpa (2013) conducted a study to analyze factors affecting academic achievement that contribute to the prediction of students' academic performance. It is useful in identifying weak students who are likely to perform poorly in their studies. Ahmad *et al*, (2018) carried out a study to examine the relationship between WhatsApp on academic performance among Saudi medical school students at the University of Albaha and the University of Dammam. A total of 160 students (79 students from Albaha medical school and 81 students from Dammam medical school) with smartphones were surveyed on their use of social media services and their academic performance (October–December 2015). A Likert scale consisting of 14 items was used to measure the use of WhatsApp and its relationship with academic achievement. Factor analysis of the self-report data of the social media addiction items was conducted. Pearson's correlations were examined to determine the relationship of WhatsApp use with academic achievement and symptoms of addiction. Nearly 99% of participants reported using WhatsApp (over 53% use for academic activities). There was no significant association between GPA and students who use WhatsApp. However, the time spent on WhatsApp usage was directly proportional to the symptoms of

**Corresponding Author Email: ahmedsule710@gmail.com; Tel: 08036154396*

addiction. Erdem *et al,* (2007) aimed to determine the probability of the factors affecting the Cumulative Grade Point Average (CGPA) of university students. They showed that academic performances are mostly measured by CGPA. It is influenced by gender, previous academic performance, living place and income level of family, social environment, the type and quality of the high school graduated, the high school Grade Point Average (GPA), the score obtained from nationwide university entrance exam, time spend for studying, learning ability and living place during the university life.

Thawabieh (2016), in his study aimed to investigate the factors that affect students' achievement, the quantitative descriptive and qualitative designs were in used in the research. The sample of 488 students (219 males and 269 females) were considered. The researcher developed a questionnaire to collect quantitative data which consist of 5 sections. In his outcome, it indicated that the following factors affect students' achievement: courses, test administration, students and faculties. Adult education and training have been focal points of labor market studies in developed economies since the 1960s. Increased attention to this topic is part of a shift towards greater emphasis on human capital in both economic and social policy in societies that are increasingly knowledge based. Several factors accounted for the growing interest in these issues. Technological change, especially advances in information and computer technologies, and globalization of production have resulted in a growing demand for highly skilled workers and changes in workplace skills requirements. In addition, because of demographic changes and reduced fertility in many countries, the labor force is both aging and growing more slowly than in the past. (Ferrer, 2010). Jinna and Maikano (2014), stated that the potential benefits of adult education are multi-dimensional adding that the contribution of adult education to development to society happens at the intersection of the social, economic, political and cultural determinants of progress in any society. The social benefits of adult education range from improved sanitation, health care and nutrition in the wider population to reduced mortality rates. The major economic benefit of adult education is the development of human capital: individuals get skill and knowledge that make them productive and thus dynamic partners in development efforts rather than just passive recipients. Yuan *et al* (2013) embarked on the analysis and evaluation of examination results which provide the theatrical basis for teaching quality and management. Base on the examination results engineering specialty students in the first term of 2011-2012 school calendar of Nuhan university of science and technology, the quantitative analysis for several parameters including difficulty, discrimination and reliability were investigated. The outcome of the analysis indicated that the distribution of examination scores approximate to normal distribution. Difficulty of the exam paper belongs to median level, and discrimination of this is qualified as well as reliability. It was concluded that the design of the examination paper was good and dependable. Hence the objective of this study is to use factor analysis to investigate the relationship between the semester results and academic performance of students at a Centre for Continuing Education and Training in a Polytechnic in Niger State, Nigeria.

MATERIALS AND METHODS

Study Area: In this research, the data used were collected from Centre for Continuing Education and Training (CCE&T) Niger State Polytechnic, Zungeru. The grades of some selected courses offered by ND2 Science and laboratory Technology were collected which comprises of first, second, and third semester of the academic session. The sampling technique used in taking the sample was simple random sampling. The technique was used because it gives all the elements of the population equal chance of being selected in to the sample. That is, it is a method without bias. The Slovene's Method was used to sample 57 from the population of 67 Students

Principal component analysis (principal factor): The covariance matrix can be factored out using the spectral decomposition theorem stated by Onyeagu, 2003, as given in equations 1 to 6

$$\sum = \lambda_1 e_1 e_1{}' + \lambda_2 e_2 e_2{}' + \cdots + \lambda_p e_p e_p{}'$$

$$= \left[\sqrt{\lambda_1}\, e_1 \,\middle|\, \sqrt{\lambda_1}\, e_2 \,\middle|\, \ldots. \,\middle|\, \sqrt{\lambda_p}\, e_p\right] \begin{bmatrix} \sqrt{\lambda_1}\, e_1{}' \\ \sqrt{\lambda_2}\, e_2{}' \\ \vdots \\ . \\ \sqrt{\lambda_1}\, e_p{}' \end{bmatrix} \qquad (1)$$

Where ($\lambda_i\ e_i$) is the eigenvalue- eigenvector pair of $\sum$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$

This fits the covariance structure of the factor analysis model having as many factors as variables ($m = p$) and specific variances $\varphi_i = 0$ for all i.

Then, $\lambda_{m+1}\, e_{m+1}\, e'_{m+1,} + \cdots + \lambda_p\, e_p e'_p$ to Σ in the spectral decomposition of Σ giving as

$$\Sigma = \left[\sqrt{\lambda_1}\, e_1 \,\middle|\, \sqrt{\lambda_2}\, e_2 \,\middle|\, \ldots. \,\middle|\, \sqrt{\lambda_m}\, e_m\right] \begin{bmatrix} \sqrt{\lambda_1}\, e_1{}' \\ \sqrt{\lambda_2}\, e_2{}' \\ \vdots \\ . \\ \sqrt{\lambda_m}\, e_m{}' \end{bmatrix} \qquad (2)$$

$$= L_{p\times m} L'_{m\times p} \qquad (3)$$

The specific factors are now included in the model and their variances are taken to be the diagonal elements.

The approximation becomes $\Sigma = LL' + \varphi$

$$\Sigma = \left[|\lambda_1 e_1| \sqrt{\lambda_2}\, e_2| \ldots . |\sqrt{\lambda_m}\, e_m|\right] \begin{bmatrix} \sqrt{\lambda_1}\, e_1' \\ \sqrt{\lambda_2}\, e_2' \\ \vdots \\ . \\ \sqrt{\lambda_m}\, e_m' \end{bmatrix} + \begin{pmatrix} \varphi_1 & 0 \ldots 0 \\ 0 & \varphi_2 \ldots 0 \\ & \vdots \\ 0 & 0 \ldots \varphi_p \end{pmatrix} \quad (4)$$

Where $\varphi_i = \sigma_{ij} - \sum_{j=1}^{m} 1_{ij}^2$ for i = 1, 2, - - -, p

The communalities are estimated as

$$\lambda_i^2 = 1_{i1}^2 + 1_{i2}^2 + - - - + 1_{im}^2 \quad (5)$$

This representation is called the principal component solution when applied to the sample covariance matrix S or the sample correlation matrix r.

In the choice of m, the number of factors can be based on the estimated eigenvalues in the same manner as with principal components. The contribution of the first few factors to the sample variances of the variables should be large. The contribution to the sample variance S_{ij} from the first common factor is 1_{ij}^2. The contribution to the total sample variance $s_{11} + s_{22} + \ldots + s_{pp}$ = tr(s) from the first common factor is then

$$\hat{l}_{11}^2 + \hat{l}_{21}^2 + \cdots + \hat{l}_{p1}^2 = (\sqrt{\hat{\lambda}_1}\hat{e}_1)'(\sqrt{\hat{\lambda}_1}\hat{e}_1) = \hat{\lambda}_1 \quad (6)$$

Since the eigenvector e_1 has unit length.

In general, the Proportion of total sample variance due to j^{th} factor

$$= \frac{\hat{\lambda}_1}{S_{11} + S_{22} + - - - + S_{pp}} \quad \text{for a factor analysis of } S$$

$$\frac{\lambda_j}{P} \quad \text{for a factor analysis of } r$$

This is used as a rule for determining the appropriate number of common factors. The number of common factors retained is increased until a "suitable proportion" of the total sample variance is explained.

Alternatively, m is set to be the number of eigenvalues of r greater than one if the sample correlation matrix is factored or equal to the positive eigenvalues of S if the sample covariance matrix is factored.

RESULTS AND DISCUSSION

The relationship that exists between the variables is presented in Table 1 and the data obtained shows a weak between the variables.

Table 1: correlation matrix

		BAM225	GLTVII	STB222	SLT221	STB212	STC222	GNS201	COM215	STM211	STP213
Correlation	BAM225	1									
	GLTVII	-.054	1								
	STB222	.267	-.268	1							
	SLT221	.101	.037	-.171	1						
	STB212	-.022	.017	-.151	.196	1					
	STC222	.414	-.172	.304	.136	.080	1				
	GNS201	-.034	-.054	.019	-.171	-.014	-.257	1			
	COM215	.184	-.289	.289	.066	.133	.241	.092	1		
	STM211	-.090	-.061	-.203	-.035	-.207	-.170	-.091	-.002	1	
	STP213	-.134	.109	.086	-.283	-.160	-.104	.102	-.062	.095	1

Determinant = .303

In table 2**,** the KMO value is large with value of 0.568 as the measure of sampling adequacy, which indicates that the sample size is adequate to be applied for the factor analysis. The Bartlett's Test Sphericity shows that p = 0.044 < 0.05 which implies that the null hypothesis is to be rejected since the correlation matrix is an identity matrix. Hence, we can run a factor analysis.

Table 2: Kaiser-Meyer-Olkin and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.568
Bartlett's Test of Sphericity	Approx. Chi-Square	81.971
	Df	45
	Sig.	.044

Knowing the standard rule of thumb for eigenvalues for factor analysis, which tells us that to select all factors with eigenvalues> 1. From the total variance displayed in Table 3, component 1 to 4 have initial eigenvalues of 2.098, 1.619, 1.214and 1.143 with total variance of 20.98%, 16.19%, 12.14%, and 11.43% respectively. This means that component 1, 2, 3 and 4 were extracted from others.

Table 3: Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.098	20.984	20.984	2.098	20.984	20.984
2	1.619	16.191	37.175	1.619	16.191	37.175
3	1.214	12.137	49.312	1.214	12.137	49.312
4	1.143	11.431	60.744	1.143	11.431	60.744
5	.879	8.792	69.536			
6	.839	8.390	77.926			
7	.678	6.775	84.701			
8	.609	6.088	90.789			
9	.495	4.949	95.738			
10	.426	4.262	100.000			

Extraction Method: Principal Component Analysis.

The scree plot is used to determine the number of factors to retain in an exploratory factor analysis (FA) or principal components to keep in a principal component analysis (PCA). The procedure of finding statistically significant factors or components using a scree plot is also known as a scree test. The scree plot of the principal component is presented in Figure 1, which is plotted by eigenvalue against the order of extraction of the component number, which indicates that component 1, 2, 3 and 4 have eigenvalues >1.

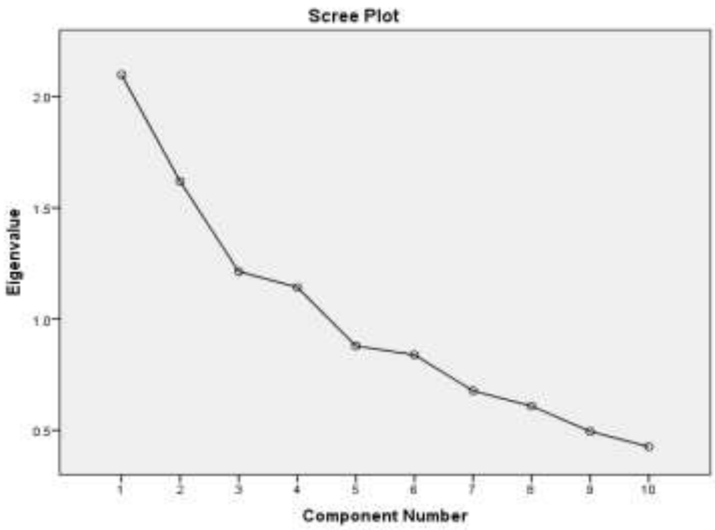

Fig. 1: Scree Plot of the Principal Components

The table 4 shows the component loadings for each variable. Looking at it critically, we will observed that there are variables loaded high and some loaded low. The variables loaded high in each component such as STC 222, STP213, GNS201 and GLTVII, are the factors that contributed significantly to the CP1, CP2 CP3 and CP4 respectively.

Table 4: Rotated Component Matrix[a]

	Component			
	1	2	3	4
BAM225	.625	.004	-.223	.224
GLTVII	-.439	-.296	-.115	.586
STB222	.591	.538	-.018	.185
SLT221	.217	-.691	-.051	-.174
STB212	.165	-.520	.538	.013
STC222	.745	-.082	-.278	.188
GNS201	-.152	.378	.681	-.042
COM215	.577	.175	.272	-.397
STM211	-.304	.115	-.484	-.655
STP213	-.296	.549	-.094	.247

Conclusion: From the analysis above, the result of the findings shows that 4 components were identified. This was determined using the Kaiser criterion that was suggested by Guttman and adopted by Kaiser which stated that "only principal components whose latent roots greater than 1($\lambda > 1$) should be retained and all others are discarded." this implies that the process of obtaining principal component should stop when we encounter principle components with latent roots less than 1($\lambda < 1$). Having seen the result of the analysis, the researcher observed that the relationships that exist between some of the courses are positive and some are negatively correlated. Therefore, it was concluded that there is a weak linear relationship that exists between the variables.

REFERENCES

Ahmed, MA; Ara, T; Yoon, SP (2018). The Impact of WhatsApp Use on the Academic Achievement among Saudi Medical Students Medical Teacher. 40 (1): S10-S14.

Bandalos, DL. (2018). Measurement Theory and Applications for the Social Sciences. The Guilford Press.

Erdem, C; Ismail, S; Cem, KA (2007). Factor Affecting Grade Point Average of University Students. *Emp. Econ. Let.* 6(5):

Ferrer, W.C. Riddell. An International Encyclopedia of Education (Third Edition), 2010.

Jinna, YJ; Maikano, PN (2014). Tthe Role of Adult Education in National Development", *Inter. Let. Soc. Hum. Sci.* 32: 35-42

Onyeagu, SI (2003). First Course in Multivariate Statistical Analysis. Mega-concept Publishers, Awka, Nigeria

Pimpa. C (2013). Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program Proceedings. *Inter. Multi-Conf. Engr. Comp. Sci.* 1: 13-15.

Thawabieh, AM (2016). Factors Affecting University Students' Achievement. *J. Educ. Soc. Behavioural Sci*. 14(4): 1-11

Yong, AG; Pearce, S (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tut. Quant. Meth. Psychol.* 9: 79-94

Yuan, W; Deng, C; Zhu, H; Li, J. (2013). The Statistical Analysis and Evaluation of Examination Results of Materials Research Method Course Creative Education. 3(7B): 162-164