Journal of Advances in Mathematics and Computer Science

36(3): 132-160, 2021; Article no.JAMCS.67074
ISSN: 2456-9968
(Past name: British Journal of Mathematics & Computer Science, Past ISSN: 2231-0851)

Principal Component Factor Analysis of Some Development Factors in Southern Nigeria and Its Extension to Regression Analysis

Nnaemeka Martin Eze[1*], Oluchukwu Chukwuemeka Asogwa[2] And Chinonso Michael Eze[1]

[1]Department of Statistics, University of Nigeria, Nsukka, Nigeria.
[2]Department of Mathematics, Computer Science, Statistics and Informatics, Alex Ekwueme Federal University Ndufu-Alike Ikwo, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. The contributions of authors are available included in conclusion. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JAMCS/2021/v36i330351
Editor(s):
(1) Dr. Junjie Chen, University of Texas at Arlington, USA.
Reviewers:
(1) Ewa Dacewicz, University of Agriculture in Krakow, Poland.
(2) Hua Zhang, Wuhan University of Science and Technology, China.
Complete Peer review History: http://www.sdiarticle4.com/review-history/67074

Original Research Article

Received 27 January 2021
Accepted 30 March 2021
Published 24 April 2021

Abstract

This study was conducted to evaluate some development factors in Southern Nigeria in order to ascertain common factors that explained the interrelationships among them and identify best cities for recommendation. A total sample of 250 cities from different states in three geopolitical zones in Southern Nigeria was used in this study and 11 development factors were considered. Kaiser-Meyer-Olkin (KMO) of (> 0.5) was computed to test the sampling adequacy; Bartlett's Test of Sphericity (Significant at 0.001) was conducted to test whether the correlation between the variables are sufficiently large for factor analysis; correlation matrix was computed to confirm the inter-item correlation. In this analysis, principal component factor analysis was the factor extraction method. Varimax rotation technique was used for factor rotation. The result showed that three new factors with eigenvalues greater than 1 were successfully constructed. The three new factors accounted for 71.63% of total variance in the dataset and assigned as the common factors influencing sustainable development in Southern Nigeria. The communalities results ranging from 0.32-0.88

**Corresponding author: Email: nnaemeka.eze@unn.edu.ng;*

depicted that factor model was adequate. The results of factor analysis were extended to multiple regression analysis. The multiple regression model was fitted using development scores as dependent variable and rotated factors as independent variables. The coefficient of determination,R^2, for the regression model was 99% and this shows that the model is adequate to evaluate the Southern Nigerian cities. The higher the estimated development scores, the better a city. Tolerance and VIF values showed that there was no multicollinearity in the regression model.

Keywords: Communality; sustainable development; factor analysis; principal component; multiple regression model; varimax rotation.

1 Introduction

In this study, we intend to evaluate some development factors in Southern Nigeria and identify how many unobservable factors that influences them. However, we also intend to extend the study to know the best cities in Southern Nigeria. Since the second half of the twentieth century, after the international agenda started to focus on development, the term development has been used widely and indiscriminately. The term 'development' has been defined in many ways by many researchers. Development is the process of coming into existence or creating something new or continuing growth of something so that it becomes more advanced. This development needs a physical reality and a state of mind. In the process of development, the interactions between social, economic and institutional processes must be continually sustained to meet up with increasing future demands in terms of population growth and continuous use of natural, human and material resources [1]. According to Seers [2], development is when a country experiences a reduction or elimination of poverty, inequality and unemployment. In the view of Owens [3], development is when there is development of people (i.e., human development) and not development of things. Israel [4] defined development as a process that creates growth, progress, positive change or the addition of physical, economic, environmental, social and demographic components. Human Development Report [5] stated that development is a way of improving capabilities and opportunities of people so that a good environment can be built for both present and generations to come. The major intention of development is to increase the level and quality of life of the citizens, the creation or growth of local regional income and creation of employment opportunities, without destroying the resources of the environment. Development should be noticeable and useful. It must not happen immediately and should include features of quality change and creates conditions for sustainability of that change. Thus, sustainability of development in our society is very essential because it brings development expansion and makes it more useful for future generation. Sustainable development is a term used as a way of responding to global environmental concerns, biophysical issues, fairness, equity and distribution. Brundtland Report also known as Our Common Future was the first to come out with the concept of sustainable development in 1987 and the report stated that a sustainable development must meet the needs of the present and this should not prevent the future generations from meeting their own needs [6]. In 2012, the United Nations Conference on Sustainable Development met and their agenda were to discuss and build up a set of sustainable development goals (SDGs) that would enable the world to achieve sustainable development by 2030 [7]. These SDGs include: (a) No Poverty, (b) Zero Hunger, (c) Good Health and well-being, (d) Quality Education, (e) Gender Equality, (f) Clean Water and Sanitation, (g) Affordable and Clean Energy, (h) Decent Work and Economic Growth, (i) Industry, Innovation and Infrastructure, (j) Reducing Inequality, (k) Sustainable Cities and Communities, (l) Responsible Consumption and Production, (m) Climate Action, (n) Life Below Water, (o) Life On Land, (p) Peace, Justice, and Strong Institutions, (q) Partnerships for the Goals. These SDGs were adopted by the United Nations General Assembly in September 2015 as a universal call to action to achieve the 17 aforementioned goals by 2030 [8].

In regard with this sustainable development, it has been observed that changes in the integrated approach to social, economic and environmental issues have not really facilitated the developmental goals in Nigeria and some of the problems to Nigeria development are poverty, flooding, ethnicity, environmental pollution, corruption, attitudes and lopsided income distribution. Nigeria is one of the African countries that are located on the western coast of Africa. The country was colonized by United Kingdom and got her independent on October 1, 1960. The country is now officially known as the Federal Republic of Nigeria and has mass land of approximately 923,768 square kilometers with density of around 212.04 individuals per square kilometers. She has over five hundred different ethnic groups and many different languages [9]. Nigeria has six geopolitical

zones that feature 36 states with Federal Capital Territory, which is known as Abuja. The six geopolitical zones in Nigeria are; North-Central, North-East, North-West, South-East, South-South, and South-West [10]. From these six geopolitical zones, one can say that Nigeria is divided into two protectorates; Northern and Southern Nigeria Protectorates.

This research focuses on the sustainable development in Southern Nigeria Protectorate using some of the SDGs of United Nations [7] as assessment criteria. Southern Nigeria which has a total surface area of approximately 206,888 square kilometers was created in 1900 by the British government. It was known as British protectorate and was officially renamed the Colony and Protectorate of Southern Nigeria in 1906. In 1914, Southern Nigeria Protectorate was joined with Northern Nigeria Protectorate to form a single colony of Nigeria [11].

The main purpose of this study is to evaluate some development factors in Southern Nigeria in order to ascertain common factors that explained the interrelationships among them and identify best cities to recommend for tourism, excursion, holiday and so on. The objectives of this research are as follows:

- To fit exploratory factor analysis model using the following development factors in Southern Nigeria; housing, healthcare, crime, transportation, education, arts, recreation, economy, borehole & pipe-borne water (i.e. water demand from the sources), energy, and climate (i.e. average temperature)
- To fit a multiple regression model using rotated factors scores as independent variables and development scores (i.e., principal component scores) as dependent variable. This fitted multiple regression model will be used to estimate cities development scores which will enable us to identify the best cities to recommend for tourism, excursion, holiday, and so on.

2 Literature Review

Several researchers had conducted studies using factor analysis as their method of data analysis such as Sakar et al. [12] that studied fruit length, fruit width, fruit height, fruit weight, shell thickness, kernel weight, kernel ratio, and filled-firm kernel ratio from 365 Ankara walnut samples using factor analysis. The results showed that three out of seven factors have eigenvalues greater than one and were selected for further study. Multiple regression analysis was used to extend the study by using factor scores from the three selected factors. The factor scores for the three selected factors were used as independent variables in multiple linear regression model for prediction of kernel weight. All of the selected factors were found to have significant linear relationships with kernel weight and 85.9% of variance in kernel weight was explained by the factors.

Song and Zhang [13] studied the consumer decision making in rural tourism based on factor analysis model. They considered 4 indicators which include price factors, market factors, safety factors and personal factors and in which these 4 factors consist of 18 secondary indicators. They distributed 600 copies of questionnaires in which 564 (94%) were valid. The Kaiser-Meyer-Olkin (KMO) test and Bartlett test were conducted to determine whether the data were suitable for factor analysis or not. The result shows that the test value of KMO is 0.785, and p-value of Bartlett test is less than 0.05, which illustrates that the collected data were suitable for factor analysis. The result from factor analysis showed that food prices, accommodation prices, others recommend and local security level are the most important factor that will affect rural tourist decision-making.

Adejumo and Adejumo [1] carried out a research on prospects for achieving sustainable development through the millennium development goals in Nigeria. The study looked at some theoretical and practical principles on sustainable development, the plan implementation of the world summit on sustainable development and the Nigerian case on sustainable development. The study showed that some factors could be identified as obstacles to achieving sustainable development in Nigeria and other part of the world and these include; poverty, corruption, lack of qualified people to develop and implement alternative technologies, lack of education. They suggested that sustainable development could be achieved in Nigeria and in the whole world, if a conscious step towards the achievement of the goals of sustainable development were considered as given by World Summit on Millennium Development Goals.

Onyeabor and Alimba [14] carried out a study on factor analysis of influence of host-community characteristics on ecotourism development in South-East Nigeria. From the study, they found out that host-community characteristics influence ecotourism development in South-East Nigeria. Particularly, poor states of socio-

economic infrastructure, inability to maximize ecotourism-induced economic opportunities, socio-political and economic exclusion of women and poor sanitary condition of host-community environments constitute impediments to ecotourism development in the area and they recommended that governments at state and council levels should step up the provision of socio-economic infrastructure in rural areas, particularly in ecotourism host-communities, including construction and maintenance of rural roads, supply of water and electricity, and spurring telecommunication services providers to provide quality services in host-communities.

Aldahmash et al. [15] conducted a study using factor analysis on the critical success factors (CSFs) of Agile Software Development. They used questionnaire on 131 respondents from agile practitioners from more than 28 countries. A principal component analysis method of factor analysis was used and it was run on 8-question questionnaire that explores the importance of the CSFs of agile projects. For each success factor, a question was asked base on seven-point options from strongly agree to strongly disagree (Likert scale 1-7). The result indicated that the first two components explained 45.77% and 12.83% of the total variance respectively. i.e., the two factors (or components) combined explained 58.61% of the total variance and this helped them to understand how these success factors are related to each other. This also helped them in planning or improving agile training programmes.

3 Data Collection

Data collection is a very vital feature of any research in education and with respect to that, it would be necessary to state how data were collected for this research. Thus, the data used in this research were secondary data collected from different ministries and agencies in three geopolitical zones in Southern Nigeria. These ministries and agencies include: state ministry of housing, National Health Insurance Scheme (NHIS), Law Enforcement Agency (Nigeria Police Force were considered), state ministry of transportation, state ministry of education, state ministry of tourism, arts & culture, National Bureau of Statistics, state ministry of water resources, Power Holding Company of Nigeria (PHCN), and Nigerian Meteorological Agency for Climate. The geopolitical zones and states in Southern Nigeria are as follows: South-East:- Abia State, Anambra State, Ebonyi State, Enugu State, and Imo State; South-South:- Akwa-Ibom State, Bayelsa State, Cross River State, Delta State, Edo State, and Rivers State; South-West:- Ekiti State, Lagos State, Osun State, Ondo State, Ogun State, and Oyo State.

From the aforementioned states, 250 areas or cities were selected using simple random sample technique and these include: **Abia State:-** Aba, Akwete, Arochukwu, Bende, Umuahia, Osisioma, Omoba, Okpuala-Ngwa, Oke-Ikpe, Mbalano, Isiala-Oboro, Nkwoagu-Isuochi, Ebe-Ohiafia, and Mgboko. **Akwa-Ibom State:-** Afaha Ikot Ebak, Afaha Offiong, Eket, Etinan, Eyofin, Ibiaku Ntok Okpo, Ikot Abasi, Ikot Akpa Nkuk, Ikot Edibon, Ikot Ekpene, Ikot Ibritam, Itu, Nto Edino, Odoro-Ikpe, Oko Ita, Oron, Urua Inyang, Urue Offong, Utu Etim Ekpo, and Uyo. **Anambra State:-** Abagana, Aguata, Agulu, Awka, Atani, Enugu-Ukwu, Igbo-Ukwu, Ihiala, Nkpor, Nnewi, Obosi, Ogidi, Okpogho, Onitsha, Otuocha, Ozubulu, and Umunze. **Bayelsa State:-** Kaiama, Nembe, Ogbia, Oporoma, Sagbama, Twon-Brass, and Yenagoa. **Cross River state:-** Abuochiche, Akamkpa, Akpet-Central, Boje, Calabar, Effraya, Ikom, Ikot-Nakanda, Itigidi, Ogoja, Obubra, Obudu, Odukpani, Okpoma, Sankwala, and Ugep. **Delta State:-** Aboh, Agbor, Akwukwu-Igbo, Asaba, Bomadi, Burutu, Effurun, Issele-Uku, Kwale, Mele, Obiaruku, Oghara, Ogwahi-Uku, Orerokpe, Sapele, Ughelli, and Warri. **Ebonyi State:-** Abakaliki, Afikpo, Effium, Ezillo, Ezzamgbo, Iboko, Ishieke, Isiaka, Isu, Nguzu-Edda, obiozara, Onuebonyi-Echara, Onueke, and Ugbodo. **Edo State:-** Afuze, Agenebode, Auchi, Benin city, Ekpoma, Fugar, Idogbo, Igarra, Igueben, Iguobazuwa, Irrua, Okada, Uromi, and Uselu. **Enugu State:-** Agbani, Aguobu-Owa, Amagunze, Awgu, Enugu, Enugu-Ezike, Ikem, Ndeaboh, Nkwo-Nike, Nuskka, Obollo-Afor, Ogbede, Oji-River, and Udi. **Imo State:-** Aboh, Afor-Oru, Awo-Idemili, Igbema, Isinweke, Mgbidi, Nkwerre, Nnenasa, Nwaorieubi, Oguta, Okigwe, Okwe, Orlu, Owerri, Umuguma, and Umundugba. **Lagos State:-** Agege, Ajegunle, Akodo, Apapa, Badagry, Ebute-Metta, Epe, Festac Town, Ifako, Ikeja, Ikorodu, Ikoyi, Lagos, Mushin, Ojota, Oshodi, Shomolu, Somolu, and Surulere. **Ogun State:-** Abeokuta, Atan, Ayetoro, Ifo, Ijebu-Igbo, Ijebu-Ode, Ilaro, Imeko, Ipokia, Isara, Itori, Ogbere, Ota, Owode, Sango Otta, and Shagamu. Ondo State:- Akure, Bolorunduro, Ifon, Igbara-Oke, Igbekebo, Igbokoda, Iju, Ikare, Ile-Oluji, Ita-Ogbolu, Ode-Irele, Oka-Akoko, Oke-Agbe, Ondo, Ore, and Owo. **Osun State:-** Apomu, Ede, Gbongan, Ifon, Ijebu-Jesa, Ikire, Ikire, Ikirun, Ila Orangun, Ile-Ife, Ile-Ogbo, Ilesha, Ilobu, Ipetumodu, Iwo, Oshogbo, and Osu. **Oyo State:-** Akanran, Egbeda, Ibadan, Igbeti, Igboho, Igbo-Ora, Ikoy-Ile, Iresa-Adu, Iseyin, Iwere-Ile, Iyana-Ofa, Kishi, Moniya, Ogbomosho, Oyo, and Shaki. **River State:-** Abonnema, Afam, Ahoada, Akinima, Bonny, Bori, Buguma, Degema, Emuoha, Isiokpo, Ogu, Okrika, Omoku, Opobo, Port Harcourt, Rumuodomaya, and Saakpenwa.

In regard with this data collection, an extensive review was done on development factors that Nigeria uses to develop her states, the following development factors were selected; Housing, Healthcare, Crime, Transportation, Education, Arts, Recreation, Economy, Borehole & Pipe-borne Water, Energy, and Climate. Moreover, in determining the outcomes of this research, a total of 2,750 data samples were collected for the analysis (i.e., 250 cases of cities by 11 development factors). The data were analyzed using factor analysis and multiple regression analysis with the help of R version 4.0.3.

4 Research Methodology

4.1 Data screenig

After the collection of data for conducting factor analysis, one should screen the data for the following assumptions before proceeding to factor analysis:

4.1.1 Interval data

The data use for factor analysis are usually performed on interval (continuous) or ordinal variables. Sometimes categorical and dichotomous variables may be considered [16].

4.1.2 Adequate sample size

It is assumed that the sample size should be large, that is, the case must be greater than the factor. The adequacy of sample size can be checked using Kaiser-Meyer-Olkin (KMO) statistic. Kaiser [17] introduced a Measure of Sampling Adequacy (MSA) which was later modified by Kaiser and Rice [18]. The KMO measure of sampling adequacy is a statistic use to test if the sample size is big enough for factor analysis. It ranges from 0 to 1. The KMO with more than 0.50 should be sufficient for factor analysis [19-21]. The maximum value of KMO is 1 and its value of 0.90 is considered as 'Excellent', 0.80 is 'good', 0.70 is 'moderate' and 0.60 is 'poor' [22,23]
The KMO test statistic as stated by Norusis [24] is

$$KMO = \frac{\Sigma_{j=1}^{n}\Sigma_{i=1}^{n} r_{ij}^2}{\left(\Sigma_{j=1}^{n}\Sigma_{i=1}^{n} r_{ij}^2 + \Sigma_{j=1}^{n}\Sigma_{i=1}^{n} q_{ij}^2\right)} \quad (1)$$

Where r_{ij} is the correlation coefficient between i^{th} and j^{th} of the original variables, q_{ij} is the partial correlation (anti-image) correlation coefficient

4.1.3 Linearity

The variables use in factor analysis are based on linearity assumption, that is they should be linearly related to each other or moderately correlated otherwise the number of factors will be almost the same as the number of original variables and when this happens, the purpose of factor analysis has been defeated [25]. Non-linear variables can also be used only when they have been transformed into linear variables using any transform method. This assumption can be checked by looking at scatterplots of pairs of variables or pairwise correlation method such as Pearson correlation method. One can also use Bartlett's test of sphericity to confirm if correlation exists between variables. Bartlett's test of sphericity tests the null hypothesis that the original correlation matrix is an identity matrix, that is, no correlation between the original variables or that the variables are orthogonal and therefore unsuitable for structure detection if null hypothesis is rejected. Bartlett's test is valid for large samples (N>150) [21].

Bartlett's test of sphericity as proposed by Bartlet [26] is given as

$$\chi^2 = -\left[(n-1) - \frac{(2K+5)}{6}\right].log(|R|) \quad (2)$$

Where n is the number of observations, k is the number of variables, and R is the correlation matrix of the data while $|R|$ is the determinant of R. Bartlett's χ^2 is asymptotically $\chi^2 -$ distributed with degrees of freedom $(df) = \frac{K(K-1)}{2}$

In this study, Pearson correlation r will be used to check this linearity assumption. For the Pearson correlation r, both variables should be normally distributed. The value of the correlation coefficient varies between -1 and +1. A value of ± 1 indicates a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker [27, 28].

Pearson correlation (r) is given by

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2\}\{n(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2\}}} \quad (3)$$

Where r_{xy} is the Pearson r correlation coefficient between x and y; n is the number of observations, x_i is the value of x for i^{th} observation and y_i is the value of y for i^{th} observation.

4.1.4 No Outlier

There should be no outlier in the data that will be used for factor analysis. This can be examined by using any normality test.

Note that when there is a violation of this assumption, a method of factor extraction known as "Principal Axis Factor" or "Principal Component method of factor extraction" should be considered [16].

In this research, Shapiro-Wilk (SW) test will be used to check the normality of the observations. The previous studies showed that, for all sample sizes, Shapiro-Wilk (SW) test is the most powerful test for normality [29, 30, 31].

According to Normadiah and Yap [32], given an ordered random sample, $y_1 < y_2 < \cdots < y_n$, the original Shapiro-Wilk (SW) test statistic as stated by Shapiro [33] is defined as,

$$W = \frac{(\sum_{i=1}^{n} a_i y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (4)$$

Where y_i is the i^{th} order statistic, $\overline{y}$ is the sample mean, $a_i = (a_1, \cdots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} m)^{1/2}}$ and $m = (m_1, \cdots, m_n)^T$ are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution and V is the covariance matrix of those order statistics. The value of W lies between zero and one. Small values of W lead to the rejection of normality whereas a value that is close to one or exactly one indicates normality of the data. The null hypothesis of Shapiro's test is that the population is distributed normally.

4.1.5 No perfect multicollinearity

There should not be perfect multicollinearity between the variables use for factor analysis because factor analysis is an interdependency technique. Multicollinearity occurs when independent variables in a model are correlated. In some analysis, this kind of correlation is a problem because independent variables should be independent, i.e., there should be weak or no relationship among themselves. In other words, if the degree of correlation among independent variables is high enough, it can cause problem(s) when a model is fitted and when interpreting the result(s). For instance, from regression analysis perspective, if there is presence of multicollinearity, regression estimates will be unstable and have high standard errors. To check if there is multicollinearity between variables, one can use determinant score. As a rule of thumb, determinant score of 0.0001 indicates that there is no multicollinearity and Haitovsky's test is use to test if the determinant score is significantly different from zero which indicates an absence of multicollinearity [34,16]. In addition, a researcher can also use Tolerance method or Variance Inflation Factor (VIF) method to check this multicollinearity. Tolerance is used as an indicator of multicollinearity. The high value of tolerance is an indication that there is no multicollinearity in the model while the low value of tolerance is known to affect adversely the results associated with the model. A value of 0.10 is recommended as the minimum value of

tolerance [35]. A recommended minimum value as high as 0.20 has also been suggested [36]. Other researchers suggested tolerance minimum value of 0.25 [37].

$$Tolerance = 1 - R_k^2 \tag{5}$$

Where R_k^2 is the coefficient of determination of k^{th} predictor and it is obtained by regressing the k^{th} predictor (i.e., independent variable of interest) onto the remaining independent variables included in the model.

Variance Inflation Factor (VIF) is the reciprocal of tolerance. It identifies correlation between independent variables and the strength of that correlation. The minimum value of VIF is 1 and it has no upper limit. VIF value between 1 and 4 indicates that there is no correlation between this independent variable and any other variable and it suggests absence of multicollinearity, VIF value between 5 and 9 indicates that there is a moderate correlation, but it is not sever enough to cause problem. VIF value more than 10 is said to be highly collinear and it indicates critical levels and causes problem [38-42].

$$VIF = \frac{1}{1-R_k^2} \tag{6}$$

Where R_k^2 is the coefficient of determination of k^{th} predictor order.

In other to solve the problem of multicollinearity, the following potential solutions may be considered:

- ✓ Remove some of the highly correlated independent variables.
- ✓ Linearly combine the independent variables, such as adding them together.
- ✓ Perform an analysis designed for highly correlated variables, such as principal components analysis or partial least squares regression.

4.1.6 Homoscedasticity

The assumption of homoscedasticity (constant variance) between variables is not necessary when performing factor analysis. The reason is because factor analysis is a linear function of measured variables and homogeneous samples lower the variance and factor loadings [43].

4.2 Factor model

There are number of factor extraction methods that are use in factor analysis, such as, principal component (PC), maximum likelihood, principal axis factoring (PAF), image factor analysis, and canonical factor analysis [44-46]. In this study, principal component (PC) of factor analysis method will be considered. As the name suggests, this method uses the method used to carry out a principal components analysis. The results of this method are not actually the principal components but factor loadings although the loadings for the m^{th} factor will be proportional to the coefficients of the m^{th} Principal component. This method is mostly used when the observed data violate the assumption of multivariate normality and also used to eliminate multicollinearity. Generally, there is no assumption of normality in PCA method but the data should be linearly weakly related to avoid multicollinearity. The idea of PCA is just decomposing the variation in a p-dimensional dataset into orthogonal components that are ordered according to amount of variance explained [47, 48]. Using PCA method, the original data are reconstructed in order to provide a unique solution. This method provides total variance among the variables; therefore, making it possible to generate the same number of factors as the number of the original variables. Although the PCA method of factor analysis generates the same number of factors as the number of the original variables but note that not all these factors will meet the criteria for retention (see Number of Factors in 4.3 of Section 4). The purpose of factor analysis is to represent each of the original variables as a linear combination of a smaller set of common factors plus a factor unique.

According to Johnson and Wichern [49], factor analysis model can be written algebraically as follows:

$$\begin{array}{rcccccccccc} X_1-\mu_1 & = & l_{11}F_1 & + & l_{12}F_2 & + & \cdots & + & l_{1m}F_m & + & \varepsilon_1 \\ X_2-\mu_2 & = & l_{21}F_1 & + & l_{22}F_2 & + & \cdots & + & l_{2m}F_m & + & \varepsilon_2 \\ \vdots & & & & & & \vdots & & & & \\ X_p-\mu_p & = & l_{p1}F_1 & + & l_{p2}F_2 & + & \cdots & + & l_{pm}F_m & + & \varepsilon_p \end{array} \quad (7)$$

Where

$X_i(i=1,2,\cdots,p)$ is the observable trait *i*, i.e., the i^{th} original variable which is the data from each subject.
$\mu_i(i=1,2,\cdots,p)$ is the mean of the i^{th} original variable which is independent of the j^{th} factor.
$l_{ij}(i=1,2,\cdots,p\ and\ j=1,2,\cdots,m)$ is the coefficient of the j^{th} factor in the i^{th} original variable. They are called factor loadings.
$F_j(j=1,2,\cdots,m)$ is the common factor.
$\varepsilon_i(i=1,2,\cdots,p)$ is the unique factor associated with the i^{th} original variable. They are called specific error terms.

The estimator for the factor loadings is given by

$$\hat{l}_{ij}=\hat{e}_{ij}\sqrt{\hat{\lambda}_j} \quad (8)$$

Where $\hat{e}_{ij}$ is the estimated eigenvector of the i^{th} variable in the j^{th} principal components and $\hat{\lambda}_j$ is the estimated eigenvalue of the j^{th} principal components.

Note that this method of estimating factor loadings is known as principal component method. The estimated eigenvectors $\hat{e}_{ij}$ are the coefficients of principal components.

Moreover, the estimated principal components model using standardized data is given by

$$\begin{array}{rcccccccc} \hat{Y}_1 & = & \hat{e}_{11}Z_1 & + & \hat{e}_{12}Z_2 & + & \cdots & + & \hat{e}_{1p}Z_p \\ \hat{Y}_2 & = & \hat{e}_{21}Z_1 & + & \hat{e}_{22}Z_2 & + & \cdots & + & \hat{e}_{2p}Z_p \\ \vdots & & & & & & \vdots & & \\ \hat{Y}_p & = & \hat{e}_{p1}Z_1 & + & \hat{e}_{p2}Z_2 & + & \cdots & + & \hat{e}_{pp}Z_p \end{array} \quad (9)$$

Where $\hat{e}_{ij}$ is the estimated eigenvector for i^{th} variable in j^{th} principal components and $Z_{ij}=\frac{X_{ij}-\bar{x}_j}{s_j}$. Where Z_{ij} is the standardized data for i^{th} sample unit in j^{th} variable, X_{ij} is the original data for i^{th} sample unit in j^{th} variable, $\bar{x}_j$ is the sample mean for j^{th} variable and s_j is the sample standard deviation for j^{th} variable.

To compute eigenvalues and eigenvectors, one can use variance-covariance matrix **Σ** if the variables of interest have the same measurement units or one can use correlation matrix R if the variables have different measurement units and also if we want each variable to be given equal weight in the analysis. It is always advisable to use correlation matrix R in order to assign equal weight to all the variables. If correlation matrix R is used in the computation of eigenvalues and eigenvectors, we said that the data have been standardized.

The eigenvalues and eigenvectors for the correlation matrix are obtained using the following:

For eigenvalue:

$$|R-\lambda I|=0 \quad (10)$$

The eigenvalues obtained should be arranged in descending order i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ before computing for eigenvector.

For eigenvector:

$$(R-\lambda I)X=0 \quad (11)$$

Where R is the correlation matrix of the original variables, I is the unit matrix, λ is the characteristic root (eigenvalue), and coefficient of X is the eigenvector.

4.3 Number of factors

In factor analysis, it has been known that there are several methods for estimating common factors and each of these methods generates a certain number of common factors. However, not all these common factors will be retained prior to rotation of the factors. There are several criteria which have been proposed for determining the number of common factors to be retained. Unfortunately, various criterion rules used by researchers often lead to different solutions [50, 51].

4.3.1 Kaiser criterion

Kaiser suggested that the number of eigenvalues of the correlation matrix that is greater than 1 should given an appropriate number of common factors. Eigenvalue for a given common factor is defined as the measurement of the variance in the entire variable which is accounted for by that particular given common factor. Thus, if a factor has a low eigenvalue its contribution to the variable can be ignored [52]. Jollife [53, 54] criticized Kaiser's idea by proposing a cutoff value of eigenvalues to be 0.7 when correlation matrices are analyzed. Jollife suggested that Kaiser's proposal for cutoff point is too large especially when non eigenvalues is up to 1. Some researchers suggested that if the largest eigenvalue is close to one, then holding to a cutoff of 1 may cause useful factors to be dropped but if there are too many eigenvalues greater than 1, then those that are close to 1 may be dropped.

4.3.2 Scree plot criterion

One can also use scree plot criterion which was proposed by Cattell [55] to determine the number of common factors to retain. This criterion may cause many researchers to analyze the same data with different results. It plots eigenvalues $(\hat{\lambda}_i)$ against the number of components (i). The eigenvalues drop as one moves towards right on components (x-axis) and one should cutoff at the point it starts to curve. Cattell said that all further components after the curve point provides less information for the factors and they should be ignored.

4.3.3 Variance explained criterion

Some researchers use cumulative proportion of eigenvalues to determine the number of common factors to be retained. The cumulative percentage explained is obtained by adding the successive proportions of eigenvalue/variations explained by a given factor. In most cases, the required cutoff point is pre-specified, that is, how much of the variation to be explained is pre-determined. For instance, a researcher might state that s/he would be satisfied if s/he could explain 50% or 60% or 70% and so on of the variation. So, through doing this, a researcher would select the eigenvalues necessary until s/he gets up to his/her pre-specified cutoff percentage.

4.4 Communality

Communality is defined as the proportion of variation in a particular variable that is explained by the selected number of common factors. In other words, communality is the sum of squared loadings for a particular variable. It ranges from 0 to 1. One can think of this value as multiple R^2 value for regression models predicting the variables of interest from a certain number of factors. A variable without any unique variance at all, i.e., one with explained variance that is 100% as a result of other variables has a communality of 1 while a variable with variance that is completely unexplained by any other variables has a communality of 0 [56, 57].

Communality is given by

$$\hat{h}_i^2 = \sum_{j=1}^{m} \hat{l}_{ij}^2 \tag{12}$$

Where,

h_i is the communality of the i^{th} variable.
l_{ij} is the loading (or correlation) between j^{th} common factor and i^{th} variable.

4.5 Specific variance and error

Specific error (Ψ) is the unique factor associated with any particular original variable which is not explained by common factors. If the data are standardized, the variance for the standardized data is equal to 1. These specific errors (variances) are computed by subtracting the communality from the variance.

That is,

$$var(X_i) = \sum_{j=1}^{m} \hat{l}_{ij}^2 + var(e_i)$$
$$var(X_i) = (l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2) + \Psi_{\mathrm{i}}$$
Since $var(X_i) = 1$

$$1 = Commu\ nalities + Specific\ Variances$$

$$1 = \sum_{i=1}^{p} \hat{h}_i^2 + \Psi_{\mathrm{i}}$$
$$\Psi_{\mathrm{i}} = 1 - \sum_{i=1}^{p} \hat{h}_i^2 \quad (13)$$

This specific variances (Ψ) can also be estimated using variance-covariance matrix $\Sigma = LL' + \Psi$. This is the matrix of factor loadings times its transpose, plus a diagonal matrix containing the specific variances. Johnson and Wichern [49] stated that the estimated specific variances are provided by the diagonal elements of the matrix $\Sigma - LL'$

$$\Psi = \begin{pmatrix} \Psi_1 & 0 & \cdots & 0 \\ 0 & \Psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \Psi_{\mathrm{p}} \end{pmatrix} \quad (14)$$

4.6 Complexity

Complexity (C^*) is another statistic that examines the number of factors on which a variable has moderate or high loadings [58]. The complexity of all the variables is greater than or equal to 1. This complexity is reduced by carrying out rotational computation on factors, that is, by rotating factors.

Complexity is given by

$$\widehat{C_i^*} = \frac{\left(\sum_{j=1}^{m} \hat{l}_{ij}^2\right)^2}{\sum_{j=1}^{m} \hat{l}_{ij}^4} \quad (15)$$

Where,

C_i^* is the complexity of the i^{th} variable.
l_i is the loading (or correlation) between j^{th} common factor and i^{th} variable.

4.7 Factor rotation

After obtaining number of initial factor loadings using any of the criteria for determining number of factors to be retained, the next is to interpret the factors and their loadings. The interpretation of factors is easy to be done when factors are rotated [15]. There are two types of factor rotation methods, namely; orthogonal rotation method and Oblique rotation method. In orthogonal rotation method, the rotated factors remain uncorrelated while in oblique rotation method, the rotated factors are correlated. Both the orthogonal and oblique rotation method has different types of rotation techniques. The most common orthogonal method is called varimax rotation technique. It has been known that varimax rotation technique illustrates the extracted components clearer and easier for interpretation. It minimizes the number of variables that have high loadings on each factor and works to make small loadings even smaller [59]. This varimax rotation technique will be applied in this research work. According to Mohamad et al. [23], the varimax factors values which are greater than 0.75 (> 0.75) is considered as strong, the values range from 0.50 - 0.75 (0.50 ≤ factor loading ≤ 0.75) is considered as moderate and the values range from 0.30 - 0.49 (0.30 ≤ factor loading ≤ 0.49) is considered as weak factor loadings. Varimax rotation technique involves scaling the common factor loadings by dividing them by the corresponding communality as follows:

$$\tilde{l}_{ij}^{*} = \frac{\hat{l}_{ij}}{\hat{h}_i} \tag{16}$$

Where,

$\tilde{l}_{ij}^{*}$ is the quantity maximizes by varimax rotation.
l_{ij} is the initial factor loadings (or correlation) between j^{th} common factor and i^{th} variable.
h_i is the communality of i^{th} variable.

Hence, varimax rotation technique is given as

$$V = \frac{1}{p}\sum_{j=1}^{m}\left\{\sum_{i=1}^{p}\left(\tilde{l}_{ij}^{*}\right)^4 - \frac{1}{p}\left(\sum_{i=1}^{p}\left(\tilde{l}_{ij}^{*}\right)^2\right)\right\} \tag{17}$$

4.8 Factor score and multiple linear regression analysis

In the extension of factor analysis to regression analysis, factor score values obtained from factor coefficients of the selected common factors are used as independent variables. Factor scores can be computed such that they are nearly uncorrelated or orthogonal. It can be regarded as a variable explaining how much a sample unit would score on a factor. The use of these factor score values in multiple regression analysis helps to solve the problem of multicollinearity, therefore, helping a researcher to make a good prediction [35]. Although, it may even be of interest of a researcher to use these factor score values as the dependent variables in a future analysis [12]. There are several methods that have been proposed for estimating factor scores from the data, namely; ordinary least squares, weighted least squares and regression method [60, 61]. The method for estimating these factor scores depends on the method used to obtain common factor loadings as mentioned in this section (see Factor Model in 4.2 of Section 4). In regard with factor scores method, weighted least squares method will be considered in this research. In this method squared residuals are divided by the specific variances and this gives more weight to variables that have low specific variances. The factor model fits the data best for variables with low specific variances and these variables should give more information regarding the true values for the specific factors [62].

Given the factor model: $X_i = \mu + Lf_i + \epsilon_i$

We want to find f_i that minimizes

$$\sum_{j=1}^{p}\frac{\epsilon_{ij}^2}{\Psi_i} = \sum_{j=1}^{p}\frac{\left(x_{ij} - \mu_i - l_{j1}f_1 - l_{j2}f_2 - \cdots - l_{jm}f_m\right)^2}{\Psi} = (X_i - \mu - Lf_i)'\Psi^{-1}(X_i - \mu - Lf_i)$$

Where Ψ is the diagonal matrix whose diagonal elements are equal to the specific variances.

$$\text{Hence, } \hat{f}_i = \left(L'\Psi^{-1}L\right)^{-1}L'\Psi^{-1}(X_i - \mu) = \left(L'\Psi^{-1}L\right)^{-1}L'\Psi^{-1}(X_i - \overline{x}) \tag{18}$$

After obtaining factor scores, because of standardization, a factor score of zero (0) represents an average contribution of that factor to a specific observation or sample. A factor score of 1.0 indicates that the contribution of that factor is one standard deviation higher than the average while a factor score of -1.0 indicates one standard deviation lower than the average [63].

Having computed factor score values and using them as independent variables, the general multiple regression model is given by

$$Y_i = \alpha + \beta_1 FS_{i1} + \beta_2 FS_{i2} + \beta_3 FS_{i3} + \cdots + \beta_m FS_{i\,m} + \varepsilon_i \tag{19}$$

Where $Y_i(\mathrm{i} = 1, 2, \cdots, \mathrm{n})$ is the dependent varible for ith unit, $FS_{i1}, FS_{i2}, FS_{i3}, \cdots, FS_{i\,m}$ $(i = 1, 2, 3, \cdots, n)$ are the factor scores for i^{th} unit on m number of factor scores (i.e., m is the number of factor scores (FS) we considered), α is regression constant (it is the value of intercept and its value is zero); $\beta_1, \beta_2, \beta_3, \cdots, \beta_m$ are regression coefficients of factor scores (FS) on m number of factor scores. FS is factor score and ε is the error term. Regression coefficients are tested with a t-statistic. The coefficient of determination, R^2, is used as an indicator of the quality of the regression [64].

In regard to the multiple regression model in equation (19), a researcher can estimate dependent variable Y_i using principal components model in equation (9) if there is no dependent variable Y_i. In principal components analysis, we create new variables that are linear combinations of the observed variables but in factor analysis, we model the observed variables as linear functions of the factors. Therefore, in practice, principal component model (Y_i) is used to create new variable that serves as dependent variable instead of factor model. This principal components (Y_i) in equation (9) is a function of observed random data, and so the data create from it are also random. Moreover, the first principal component will be used to create dependent variable for our study because it always explains maximum variance among all linear combinations. It accounts for as much variation in the data as possible.

This first principal component is given by

$$\hat{Y}_{i1} = \hat{e}_{11}Z_{i1} + \hat{e}_{12}Z_{i2} + \cdots + \hat{e}_{1p}Z_{ip}. \tag{20}$$

Where $\hat{Y}_{i1}$ $(\mathrm{i} = 1, 2, \cdots, \mathrm{n})$ is the first principal component for i^{th} unit, $\hat{e}_{11}, \hat{e}_{12}, \cdots, \hat{e}_{1p}$ are the estimated eigenvectors for the first principal component on p number of standardized variables and $Z_{i1}, Z_{i2}, \cdots, Z_{ip}$ $(i = 1, 2, \cdots, n)$ are the standardized data for i^{th} sample unit on p number of standardized variables.

In addition, a researcher may also have interest in using factor scores (FS) as dependent variables. In this case, standardized variables Z_{ij} will now be the independent variables. Since the first factor score (FS) is computed through first eignevalue which always explains maximum variance among other eigenvalues, it is more acceptable to use as dependent variable. In regard to this, the multiple regression model is given as

$$FS_{i1} = \alpha + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \cdots + \beta_k Z_{ik} + \varepsilon_i \tag{21}$$

Where FS_{i1} $(\mathrm{i} = 1, 2, \cdots, \mathrm{n})$ is the first factor scores for i^{th} unit, $\beta_1, \beta_2, \beta_3, \cdots, \beta_k$ are the regression coefficients on k number of standardized variables, $Z_{i1}, Z_{i2}, \cdots, Z_{ip}$ $(i = 1, 2, \cdots, n)$ are the standardized data for i^{th} sample unit on p number of standardized variables.

5 Results and Discussion

In this section, we presented results and discussion from data analysis. The results were divided into two which are **(a)** factor analysis and **(b)** multiple linear regression analysis. With regard to factor analysis, we first used Kaiser-Meyer-Olkin (KMO) test and Bartlett test to determine whether the data were suitable for factor analysis

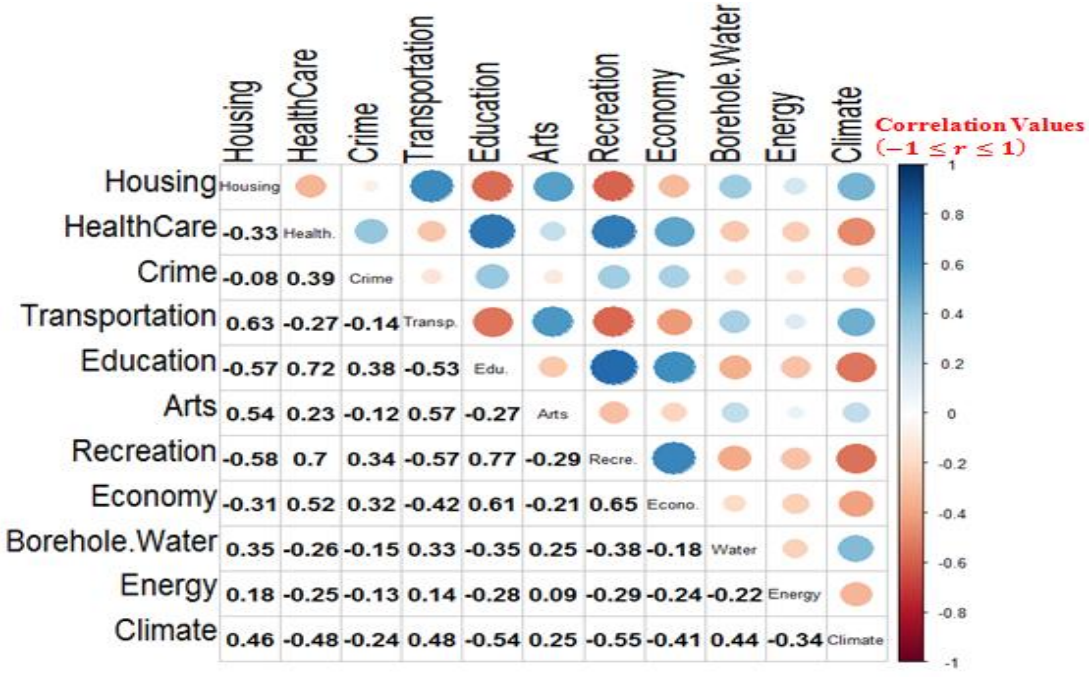
or not, as showed in Table 1. The statistical tool used to obtain KMO and Bartlett tests was SPSS version 23. The result showed that the test value of KMO was 0.777 which exceeded the factor analysis validity threshold value of 0.5 that was recommended [19-22]. The result of KMO was supported by Bartlett's test of Sphericity which was significant (Chi-square = 1668.376; P<0.001 *i.e.*, P-value of Bartlett test is less than 0.05). The KMO result indicates that the sample size is large enough for factor analysis and Bartlett's test of Sphericity result shows that the original correlation matrix is not an identity matrix, therefore, the data are suitable for factor analysis, that is, the variables are correlated highly enough to provide a reasonable basis for factor analysis [26] (see Adequate Sample Size and Linearity in Section 4).

Table 1. The result of KMO statistical test and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.777
Bartlett's Test of Sphericity	Approx. Chi-Square	1668.376
	df	55
	Sig.	0.000

In addition to investigate further that Bartlett's test of Sphericity was significant (P<0.001) and also to check if multicollinearity exists among the variables, we computed a correlation matrix (Table 2). From the result in Table 2, which was obtained using R version 4.0.3, we can see that there is correlation between each pair of variables (*i.e.*, the considered development factors). In this Table 2, the faint darkred and blue circles indicate weak correlation values whereas the bold darkred and blue circles indicate strong correlation values. Since the correlations between each pair of the variables were significant (P<0.01 or P<0.05); the correlation coefficients may be factorable. Again, there is high correlation between some variables and this might be resulting in multicollinearity in the model. In other to solve this problem of multicollinearity, principal components analysis was conducted (see Multicollinearity in 4.1.5. of Section 4).

Table 2. Pearson correlation coefficients among all development factors



The normality of data was examined by plotting histogram (Fig. 1). This normality test enabled us to examine if there is an outlier in the sample.

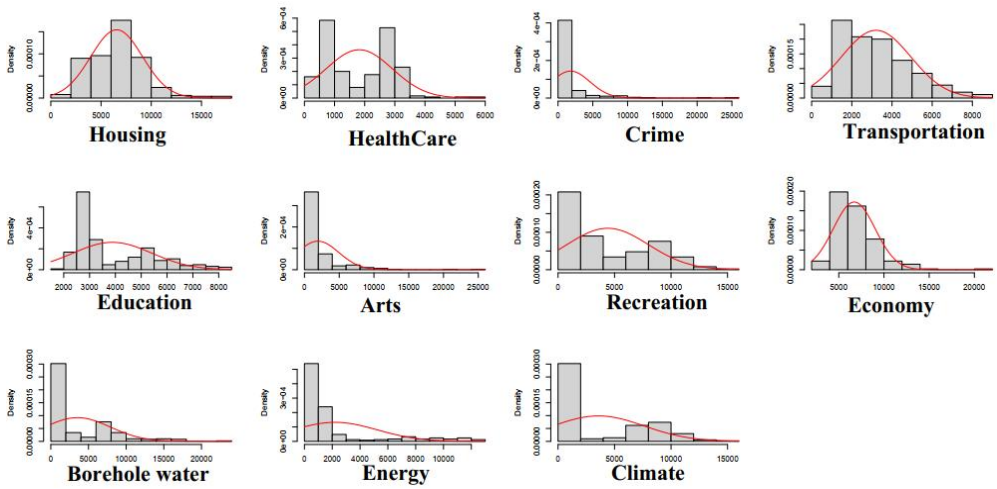


Fig. 1. Histogram plot

The histogram plot in Fig. 1 showed that the data from each variable were not normally distributed and this means that there is an outlier in the samples. Since graph can be interpreted or viewed in different ways, the numerical method of normality test can be used to confirm the normality test of the sample observations. In this case, Shapiro-Wilk normality test was applied for the data (see Table 3). As can be seen from column 4 of Table 3, the p-value of each variable is less than 0.001 (*i.e.*, p<0.001) and this indicates that the data were not normally distributed.

Table 3. Shapiro-wilk test of normality

Test	Variable	Statistic	p-value	Normality
Shapiro-Wilk	Housing	0.9624	<0.001	NO
Shapiro-Wilk	HealthCare	0.8971	<0.001	NO
Shapiro-Wilk	Crime	0.4828	<0.001	NO
Shapiro-Wilk	Transportation	0.9488	<0.001	NO
Shapiro-Wilk	Education	0.8485	<0.001	NO
Shapiro-Wilk	Arts	0.5702	<0.001	NO
Shapiro-Wilk	Recreation	0.845	<0.001	NO
Shapiro-Wilk	Economy	0.8847	<0.001	NO
Shapiro-Wilk	Borehole Water	0.7494	<0.001	NO
Shapiro-Wilk	Energy	0.6113	<0.001	NO
Shapiro-Wilk	Climate	0.745	<0.001	NO

5.1 Factor analysis result

Having screened the data for factor analysis and found out that they were appropriate (see Table 1); factor analysis was performed using principal component analysis (PCA) method. The PCA method of factor analysis was used to extract the factors due to the violation in the assumption of normality (see Table 3) and also to solve multicollinearity problems (see Table 2). The extraction of the factors was done by computing eigenvalues of the correlation matrix in Table 2. However, correlation matrix was used because we want each variable to be given equal weight in the analysis. In practice, eigenvalues that are greater than 1 are mostly considered [52]. From the result in Table 4, the first three components have eigenvalues greater than 1 and this implies that the first three common factors are required. Also, the percentage of variance explained by each of the components as well as the cumulative percentage of variance explained is provided in column 3 & 4 of this Table 4. The proportion of variance in the set of variables accounted for by a factor is the sum of square loading (i.e., sum of eigenvalues) for the factor divided by the number of variables. The percentage of variance is the multiplication of the proportion of variance by 100. For instance, the proportion of variance explained by the first component is (4.7845/11.0001) = 0.43 whereas the percentage of variance explained by the first component is (0.434951x100) = 43.49%. The second component accounted for 15.24% and the third component accounted for

12.89%. The three selected components explained 71.63% of the total variation of variables in this analysis. Moreover, we obtained variance rotation using varimax rotation method and because varimax rotation is orthogonal, the cumulative variance proportion of the rotated three components together accounted for (7.88/11.0001) = 0.72 and its cumulative variance percentage is (0.71629 x 100) = 71.63%. The cumulative variance percentage can also be used to determine the number of common factors to compute. The cumulative variance percentage shows the amount of variances that components explain. The numbers of components that explain an acceptable level of variance are retained and the acceptable level depends on the application. In our result, the first three components explained 71.63% of the total variation. This is an acceptably large percentage (see Table 4).

Table 4. Total eigenvalues and the total variance explained

	Initial sums of squared loadings			Rotation sums of squared loadings		
Component	**Eigenvalue**	**Variance Percent**	**Cumulative Variance Percent**	**Eigenvalue**	**Variance Percent**	**Cumulative Variance Percent**
Component 1	4.7845	43.4951	43.4951	3.529	32.081	32.081
Component 2	1.6769	15.2444	58.7395	2.702	24.560	56.641
Component 3	1.4179	12.8899	71.6295	1.649	14.989	71.629
Component 4	0.8435	7.6682	79.2976			
Component 5	0.6576	5.9777	85.2754			
Component 6	0.5253	4.7753	90.0506			
Component 7	0.3453	3.1393	93.1900			
Component 8	0.2491	2.2647	95.4547			
Component 9	0.2193	1.9939	97.4486			
Component 10	0.1883	1.7114	99.1600			
Component 11	0.0924	0.8400	100.0			
Total	**11.0001**			**11.0001**		

Moreover, the scree plot which was proposed by Cattell [55] (see Number of Factors in 4.3 of Section 4) can be used to investigate further if the first three eigenvalues should be used to obtain the required common factors. As shown in Fig. 2, the abscissa represents the number of components ordered from largest to the smallest and the vertical axis represents the variance (*i.e.*, eigenvalues) percentage explained. The number of components is a unique number to identify each eigenvalue during analysis and this Fig. 2 displays only the eigenvalues that the percentage is greater than or equal to 1. This indicates that any eigenvalue that the percentage is less than 1 is very negligible to contribute in analysis (see Table 4). The red dashed line on the scree plot indicates the cutoff point. The eigenvalues greater than or equal to 1 are above the red dashed line while the eigenvalues less than 1 are below the line.

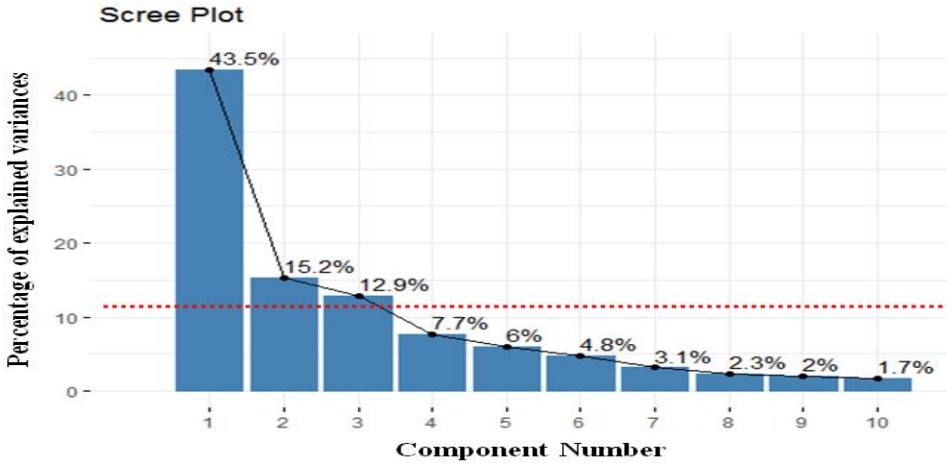


Fig. 2. Scree plot for initial eigenvalues

Having decided the number of factors to be extracted, it is the time to interpret the factors and their loadings. Since PCA method of factor analysis had been considered in this study, the first-stage was to compute PCA using equation (9) and its results were used to obtain the initial required common factors in Table 5 (see factor model in 4.2 of section 4).

Table 5 presented initial factor loadings, communalities, specific variances and complexities. These initial factor loadings were rotated using varimax rotation technique (see factor rotation in 4.7 of section 4). The application of the varimax rotation technique was to find factors that are easier to interpret (see Table 6 for the interpretation of factor loadings). The communalities are the proportion of variance that each variable has in common with other variables. If communality of a variable is high, it means that the extracted factors explained a big proportion of the variables variance. The range of communality values from 0.3 – 1 indicates that the data are conformable to factor analysis. As can see from column 3 of Table 5, all the communalities ranging from 0.31 – 0.88 depicted that factor model is a good model (see communality in 4.4 of section 4). Column 4 shows specific variance which is the variance that is specific to a particular variable, that is, variance explained by a particular variable. Complexities in column 5 examine the number of factors on which a variable has moderate or high loadings. This complexity is reduced by carrying out rotational computation on factors, that is, by rotating factors (see complexity in 4.6 of section 4).

Table 5. Initial common factor loadings matrix and its communality, specific variance and complexity

Variable	Initial factor (F) loadings (l_{ij})			Communalities $(\hat{h}_i^2)$	Specific variances (Ψ_i)	Complexities $(\hat{C}_i)$
	F1	F2	F3			
Housing	-0.7256	0.3361	0.2536	0.7037	0.2963	1.6841
HealthCare	0.7010	0.5481	0.2930	0.8776	0.1224	2.2715
Crime	0.4299	0.3339	0.1352	0.3146	0.6854	2.1092
Transportation	-0.7322	0.3562	0.2614	0.7314	0.2686	1.7355
Education	0.8793	0.2017	-0.0020	0.8138	0.1862	1.1049
Arts	-0.4391	0.6319	0.4792	0.8217	0.1783	2.7083
Recreation	0.8934	0.1600	-0.0250	0.8245	0.1755	1.0657
Economy	0.7081	0.2681	-0.0438	0.5752	0.4248	1.2894
Borehole Water	-0.4982	0.3720	-0.4027	0.5488	0.4512	2.8129
Energy	-0.2207	-0.5266	0.7398	0.8733	0.1267	2.0133
Climate	-0.6951	0.2684	-0.4893	0.7946	0.2054	2.1336

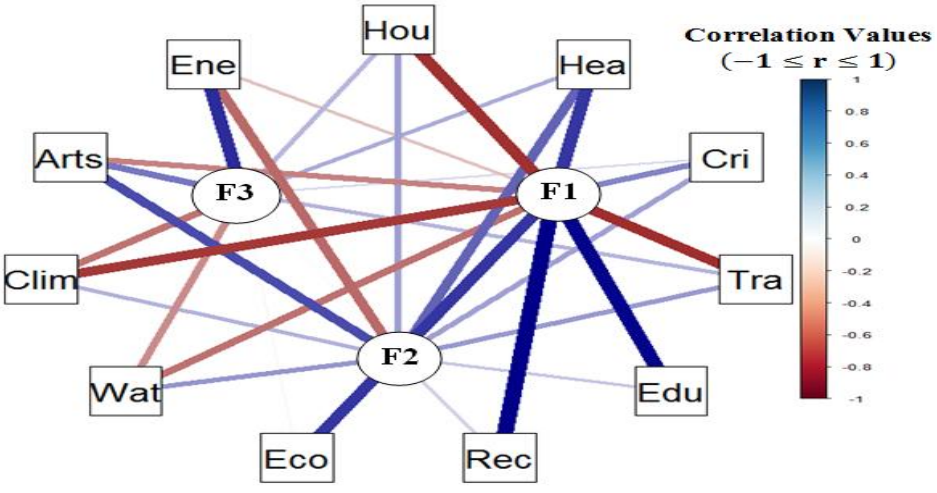


Fig. 3. Initial factor loadings plot as a network of edges (i.e., links) between nodes (i.e., variables)

The graph in Fig. 3 is a way of showing the relationships between initial factors and variables as a network of line segments. It arranges the nodes in a way that locates more highly correlated variables closer to one another. This graph shows the strength and sign of the correlations between factors and variables. The sign of the correlations are indicated by color. The blue color indicates those variables that correlated positively with factors and the darkred color indicates those that are negatively correlated with factors (see Table 5). The strength or thickness of the line increases as correlation value tends to ± 1. The legend calibrated in different colors shows the point where the correlation value falls and this can be traced using the network line color.

Table 6 depicted the result of rotated factor loadings which was obtained from the rotation of the initial factor loadings (see Table 5) using varimax rotation technique. The application of the varimax rotation technique had optimized the initial factor loadings and this was done to find factors that are easier to interpret. Factor loadings are the correlations between variables and common factors [65]. To explain each of the common factors, examine the magnitude and direction of the correlation between the original variable and the corresponding factor. The acceptance of the correlation between a variable and common factor depends on the absolute value of the correlation. The larger the absolute value of the correlation, the more important the corresponding variable is in determining the common factor. In this analysis, the absolute value of the correlation greater than or equal to 0.5 is accepted and it implies that the correlation between the common factor and variable is significant. After varimax rotation of the initial factor axes, three factors were extracted which accounted for 71.63% of the total variance of the original 11 variables (see Table 4). From the result in Table 6, factor pattern correlations of the rotated factors showed the relative contribution of each variable (i.e., development factor) to a particular factor. The bold marked loads indicate the highest correlation between variables and corresponding factors. The first rotated factor explained 32.08% of the variance in the dataset (see rotation sums of squared loadings in Table 4). This first rotated factor is strongly correlated with five of the original variables; Healthcare, Crime, Education, Recreation and Economy with correlation values of 0.93, 0.56, 0.80, 0.78, and 0.70 respectively (Table 6). This suggests that the five variables vary together, that is, if one of these variables increases, then the remaining ones tends to increase as well. Therefore, the first rotated factor can be viewed as a measure of Healthcare, Crime, Education, Recreation and Economy since they have high correlation values. The second rotated factor explained 24.56% of the variance in the dataset. This second rotated factor is highly associated with Housing, Transportation, and Arts with correlation values 0.77, 0.79, and 0.89 respectively. Furthermore, the second rotated factor primary measures Housing, Transportation, and Arts and it implies that the second rotated factor increases with them. The third rotated factor explained 14.99% of the variance in the dataset. The third rotated factor highly associated with three of the original variables, namely; Borehole water, Energy, and Climate with correlation values of 0.64, -0.84, and 0.69 respectively. This suggests that the five variables vary together but in this case, as Borehole water and Climate increase with the third factor, Energy decreases with it. Hence, the third rotated factor can be viewed as a measure of Borehole water, Energy, and Climate.

Table 6. Rotated common factor loadings matrix and its communality, specific variance and complexity (Factor Loadings < 0.50 are Excluded)

Variable	Rotated factor (RF) loadings (l_{ij})			Communalities $(\hat{h}_i^2)$	Specific variances $(\boldsymbol{\Psi}_i)$	Complexities $(\hat{C}_i)$
	RF1	RF2	RF3			
Housing	-0.3117	**0.7688**	0.1247	0.7037	0.2963	1.3794
HealthCare	**0.9280**	0.0684	-0.1083	0.8776	0.1224	1.0383
Crime	**0.5598**	0.0168	-0.0310	0.3146	0.6854	1.0079
Transportation	-0.3031	**0.7890**	0.1302	0.7314	0.2686	1.3498
Education	**0.8025**	-0.4032	-0.0849	0.8138	0.1862	1.5010
Arts	0.1302	**0.8966**	0.0306	0.8217	0.1783	1.0446
Recreation	**0.7840**	-0.4490	-0.0908	0.8245	0.1755	1.6243
Economy	**0.7029**	-0.2841	0.0216	0.5752	0.4248	1.3205
Borehole Water	-0.2320	0.3030	**0.6350**	0.5488	0.4512	1.7321
Energy	-0.3575	0.2129	**-0.8368**	0.8733	0.1267	1.4995
Climate	-0.4628	0.3116	**0.6952**	0.7946	0.2054	2.1853

The communalities are the proportion of variance that each variable has in common with other variables. If communality of a variable is high, it means that the extracted factors explained a big proportion of the variables

variance. The range of communality values from 0.3 – 1 indicates that the data are conformable to factor analysis. As can see from column 3 of Table 6, all the communalities ranging from 0.31 – 0.88 depicted that factor model is a good model (see communality in section 4). Column 4 of Table 6 shows specific variance which is the variance that is specific to a particular variable, that is, variance explained by a particular variable. From the results in Table 6, we noticed that the communalities and specific variances of the rotated factors are the same with the communalities and specific variances of the initial factors in Table 5. This means that rotation of factors does not affect the proportion of variance that each variable has in common with other variables and the variance that is specific to a particular variable.

Complexities in column 5 of Table 6 examine the number of factors on which a variable has moderate or high loadings. This complexity is reduced by carrying out rotational computation on factors, that is, by rotating factors. For instance, the rotated complexity for Housing in Table 6 is 1.38 while the initial complexity for Housing in Table 5 is 1.68; rotated complexity for Healthcare is 1.04 while initial complexity for Healthcare is 2.27 and so on.

The graph in Fig. 4 is a way of showing the relationships between rotated factors and variables as a network of line segments. It arranges the nodes in a way that locates more highly correlated variables closer to one another. This graph shows the strength and sign of the correlations between rotated factors and variables. The sign of the correlations are indicated by color. The blue color indicates those variables that correlated positively with factors and the darkred color indicates those that are negatively correlated with factors (see Table 6). The strength or thickness of the line increases as correlation value tends to ± 1. The legend calibrated in different colors showed the point where the correlation value falls and this can be traced using the network line color.

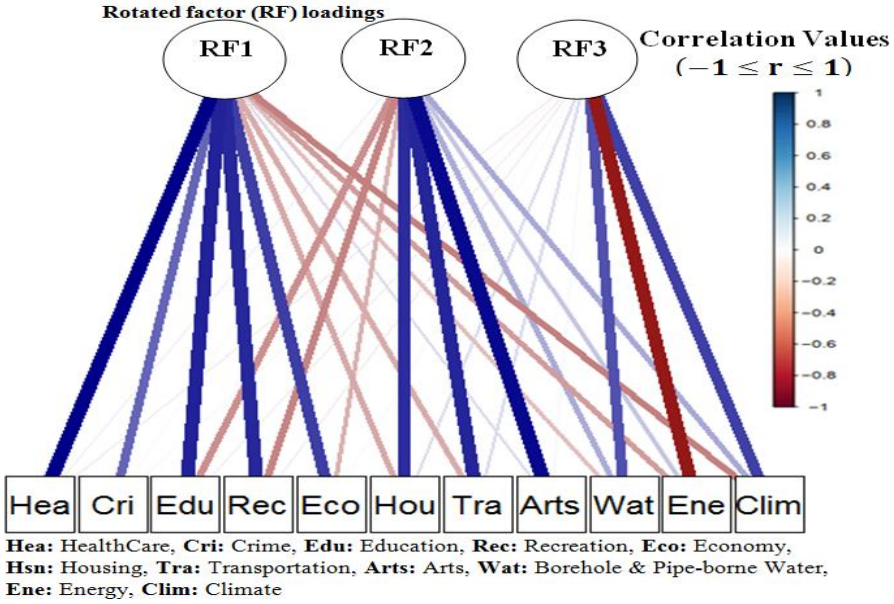


Fig. 4. Rotated factor loadings plot as a network of links (i.e., edges) between variables (i.e., nodes)

The 3D plot in Fig. 5 is known as factor loadings plot in 3-dimensional space. The values used in this plot are the values obtained in Table 6. These values are the correlation values between the factors and original variables. Values closest *to* ± 1 represent the strongest relationships and with zero being uncorrelated. This Fig. 5 shows the three rotated factors in 3-dimensional space pointing the correlations between each variable and corresponding rotated factor. The percentages 32.08%, 24.56% and 14.99% in rotated factor 1, 2 and 3 respectively are the percentage of variance in the set of variables accounted for by the factors (see Table 4). The dashed red circles in Fig. 5 depicted how the first rotated factor (RF 1) in Table 6 put Healthcare, Crime, Education, Recreation and Economy in the same group; the second rotated factor (RF 2) put Housing, Transportation and Arts in the same group; the third rotated factor (RF 3) put Borehole water, Energy and

Climate in the same group. The variables in each group imply that they are more related to each other and being influenced by the same factor than other variables in another group. The grouping of variables using 3D plot in this Fig. 5 showed that rotated factors are easier to interpret than unrotated/initial factors (see Table 5) which we could not group their variables using 3D plot. The legend calibrated in different colors showed the point where the correlation value falls and this can be traced using the variable color as shown in the 3D space.

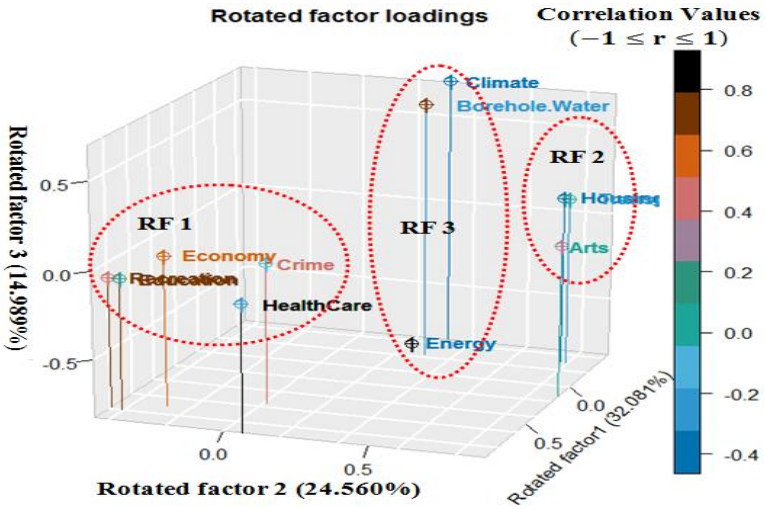


Fig. 5. 3D plot for rotated factor loadings of rotated factor 1, rotated factor 2 and rotated factor 3

5.2 Multiple linear regression analysis result

In this section, we presented results of multiple regression analysis computed using standardized values, development scores (DS) (i.e., PC scores) and rotated factor scores (RFS). Development scores were obtained through the multiplication of standard values of the considered variables by PC matrix (see Table 7). Rotated factor scores were obtained through the multiplication of standard values of the considered variables by weight of rotated factor matrix, that is, rotated factor score coefficient matrix (see Table 8).

After development scores and factor scores were obtained, the first development scores (DS 1) in Table 7 was used as a dependent variable and the three rotated factor scores (RFS 1, RFS 2, and RFS 3) in Table 8 were used as independent variables; then we obtained the results of multiple regression analysis as shown in Table 9. Moreover, the DS 1 was used as a dependent variable because it explained more variance than DS 2 and DS 3, that is, it accounts for as much variation in the data as possible more than DS 2 and DS 3. The use of interdependent explanatory variables should be treated with caution, since multicollinearity has been shown to be associated with unstable estimates of regression coefficients rendering the estimation of unique effects of these predictors impossible [66]. This justifies the use of factor scores for prediction. These factors are orthogonal to each other if rotated using varimax technique and are more reliable in development estimation. From the results in Table 9, when probabilities were taken into consideration, the regressions of DS 1 on RFS 1 ($P<0.001$), RFS 2 ($P<0.001$) and RFS 3 ($P<0.001$) were statistically significant. The three rotated factors had a positive effect on development factors (DS 1) and this means that the considered development factors increased with increasing score values of the three rotated factors. The adequacy of the regression model was examined using multiple coefficient of determination, R^2. The result showed R^2 was 99% and this means that 99% of variation in development factors (DS 1) was explained by all the three rotated factors. Tolerance and VIF were used to check if multicollinearity exists among the independent variables (RFS 1, RFS 2 and RFS 3). The result showed Tolerance and VIF values were reduced to 1 which is the VIF threshold value and it exceeded the threshold value of Tolerance [38, 37]. This result showed that multicollinearity problems were eliminated.

Table 7. Principal components matrix and development scores (i.e., PC scores)

	Principal Component (PC)			Development Scores (DS) (i.e., Principal Component Scores)			
Variable	**PC1**	**PC2**	**PC3**	**Sample No.**	**DS 1**	**DS 2**	**DS 3**
Housing	0.3317	-0.2595	0.213	1	2.3634	-0.3861	-0.7919
HealthCare	-0.3205	-0.4233	0.246	2	2.1227	-0.8233	-0.3183
Crime	-0.1965	-0.2578	0.1136	3	2.5747	-1.3720	-1.2962
Transportation	0.3348	-0.2751	0.2196	4	2.2134	-0.0116	-1.6206
Education	-0.402	-0.1557	-0.0017	5	3.1471	-7.0564	3.3893
Arts	0.2008	-0.4879	0.4025	.	.	.	.
Recreation	-0.4085	-0.1236	-0.021	.	.	.	.
Economy	-0.3237	-0.207	-0.0368	.	.	.	.
Borehole Water	0.2278	-0.2873	-0.3382	246	1.7412	0.3234	-1.2541
Energy	0.1009	0.4067	0.6213	247	1.6504	0.6263	-1.5720
Climate	0.3178	-0.2073	-0.4109	248	2.0293	-2.4522	0.5250
				249	2.5406	-0.3654	-1.2016
				250	1.5601	-0.1414	-0.6498

Table 8. Rotated factor score coefficient matrix and factor score

	Weight of Rotated Factor (WRF)			Rotated Factor Scores (RFS)				Absolute value of Development Scores
Variable	**WRF 1**	**WRF 2**	**WRF 3**	**Sample No.**	**RFS 1**	**RFS 2**	**RFS 3**	
Housing	0.0366	0.3061	-0.0097	1	-0.7745	0.4642	0.9399	0.3893
HealthCare	0.3491	0.2194	-0.0317	2	-0.4127	0.8143	0.7642	0.8287
Crime	0.2076	0.1169	0.0057	3	-0.463	0.7503	1.7075	1.3806
Transportation	0.0438	0.3171	-0.0077	4	-1.0223	-0.1231	1.3477	0.0128
Education	0.215	-0.038	0.0239	5	2.7088	5.645	0.8124	7.0955
Arts	0.2187	0.4615	-0.0619	.	.	.	.	.
Recreation	0.1993	-0.0634	0.0236	.	.	.	.	.
Economy	0.2061	-0.0089	0.0766	.	.	.	.	.
Borehole Water	0.0031	0.0424	0.3727	246	-0.957	-0.2412	0.9117	0.3242
Energy	-0.1326	0.1193	-0.5841	247	-1.1151	-0.5491	1.0007	0.6287
Climate	-0.0773	-0.0029	0.3998	248	0.5117	1.922	0.8277	2.4663
				249	-0.9088	0.3182	1.2328	0.3686
				250	-0.5837	0.1967	0.6626	0.1429

Table 9. Results of multiple regression analysis using first development scores (DS 1) as dependent variable and rotated factor scores (RFS 1, RFS 2, and RFS 3) as independent variables

Model					**Collinearity Statistics**	
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**	**Tolerance**	**VIF**
(Intercept)	-4.83e-07	1.93e-04	-0.209	0.835		
Rotated Factor Score1	7.88e-01	1.93e-04	8759.336	<2e-16 **	1.0	1.0
Rotated Factor Score2	7.77e-01	1.93e-04	6734.898	<2e-16 **	1.0	1.0
Rotated Factor Score3	6.77e-01	1.93e-04	2469.965	<2e-16 **	1.0	1.0
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
Residual standard error: 0.003049 on 246 degrees of freedom						
Multiple R-squared: 0.99, Adjusted R-squared: 0.98						
F-statistic: 4.273e+07 on 3 and 246 DF, p-value: < 2.2e-16						

Relationship between dependent variable (DS 1) and independent variables (RFS 1, RFS 2, and RFS 3)

Variable	**Zero Order value**	**Partial value**	**Part value**
Rotated Factor Score 1	0.774	1.000	0.774
Rotated Factor Score 2	0.595	1.000	0.595
Rotated Factor Score 3	0.218	1.000	0.218

$$\hat{Y} = -0.00000048 + 0.79(\text{RFS 1}) + 0.78(\text{RFS 2}) + 0.68(\text{RFS 3}) \quad (22)$$

Equation (22) is a fitted multiple regression model for prediction of development scores.

Column 4 of Table 8 showed the absolute value of development scores. This development scores were used to evaluate the considered cities for recommendation and they were obtained using the fitted multiple regression model in equation (22). From the results in this Column 4 of Table 8, we noticed that the seven cities that have highest development scores are Ikeja, Badagry, Calabar, Ibadan, Port Harcourt, Bonny and Obudu with development scores of 7.1, 5.81, 3.45, 3.35, 3.2, 3.15 and 3.06 respectively. Moreover, Table 9 showed relationship between dependent variable (DS 1) and independent variables (RFS 1, RFS 2 and RFS 3) in terms of zero order, partial and part values. Zero order value is Pearson correlation coefficient between dependent variable and independent variables. The result showed that dependent variable (DS 1) correlated strongly with RFS 1 (r = 0.77) and RFS 2 (r = 0.60) and weakly correlated with RFS 3 (r = 0.22). Partial value shows how much of the variance of dependent variable which is not estimated by the other independent variables in the model. In other words, it is the variance estimated by a specific variable. As can see, the three factors have equal partial values of 1. Part value shows how much multiple coefficient of determination R^2 will decrease if a particular independent variable is removed from the model. It is also known as a unique contribution of independent variables. The result shows that R^2 decreases if 0.77 of rotated factor score 1 is removed, 0.60 of rotated factor score 2 is removed and 0.22 of rotated factor score 3 is removed.

6 Conclusion

In this paper, the relationships among some development factors in Southern Nigeria were explored using principal component factor analysis method. In our analysis, the results showed that three new factors were successfully constructed and assigned as the common factors that influence sustainable development in Southern Nigeria. The three new factors showed how the 11 considered development factors related to each other to influence development by putting them in groups. The rotated first factor (RF 1) in Table 6 put Healthcare, Crime, Education, Recreation and Economy in the same group and this means that they are more related to each other than other variables and being influenced by the same factor. The second rotated factor (RF 2) put Housing, Transportation and Arts in the same group and this means that they are influenced by the same factor and more related to each other. The third rotated factor (RF 3) put Borehole water, Energy and Climate in the same group and also means that they are influenced by the same factor and more related to each other. The communality values obtained ranging from 0.31 – 0.88 exceeded threshold value and it showed that the common factor model was a good one. The results were extended to multiple regression analysis in order to fit a model for development scores prediction. Thus, The coefficient of determination,R^2, for the multiple regression model is 99% and this showed that the model is adequate to evaluate the areas or cities in the three geopolitical zones in Southern Nigeria. The higher the estimated development scores, the better a city. The results of the regression showed that the use of the three new rotated factors as independent variables helped to eliminate multicollinearity problems. Development scores increased with increasing score values of rotated factors with positive effect. This implies that development scores increased with increasing Healthcare, Crime, Education, Recreation and Economy in rotated factor 1 (RFS 1). Also, development score values increased with increasing Housing, Transportation and Arts in rotated factor 2 (RFS 2) and It increased with increasing Borehole water and Climate and decreasing Energy in rotated factor 3 (RFS 3). The development scores obtained from the fitted multiple regression model showed that Ikeja, Badagry, Calabar, Ibadan, Port Harcourt, Bonny and Obudu have the highest scores among others and they are regarded as the best cities.

In summary, this study enabled us to identify three common factors that explained the interrelationships among some development factors in Southern Nigeria and also fitted multiple regression model that was used to estimate areas or cities development scores. This development scores were used to identify the best areas or cities in southern Nigeria.

Disclaimer

The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the authors and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge.

Also, the research was not funded by the producing company rather it was funded by personal efforts of the authors.

Declarations

Availability of data and material

The data are available.

Code availability

The R codes used in this manuscript are available.

Acknowledgement

We wish to thank the ministries and agencies from different states in Southern Nigeria that assisted us by making the data available for this research.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Adejumo AV, Adejumo OO. Prospects for achieving sustainable development through the millennium development goals in Nigeria. European Journal of Sustainable Development. 2014;3(1):33-46.
DOI: 10.14207/ejsd.2014.v3nlp33

[2] Seers D. The meaning of development. Paper presented at the Eleventh World Conference of the Society for International Development, New Delhi; 1969.
Accessed: 18 June, 2020.
Available: http://kokminglee.125mb.com/economics/development.html

[3] Owens E. The future of freedom in the developing world: Economic development as political reform. New York: Pergamom Press; 1987.

[4] Israel S. Development Issues; 2018.
Accessed: 15 June, 2020.
Available: https://www.sid-israel.org/en/Development-Issues/What-is-Development.

[5] Human Development Report; 1994.
Accessed: 18 June, 2020.
Available: http://kokminglee.125mb.com/economics/development.html

[6] United Nations General Assembly. Report of the world commission on environment and development: Our common future. Oslo, Norway: United Nations General Assembly, Development and International Co-operation: Environment; 1987.

[7] United Nations Conference on the Sustainable Development (Rio+20). Rio Declaration on Sustainable Development. Rio de Janiero, Brazil: United Nations; 2012.

[8] Scarlat, N. and Dallemand, J. The role of bioenergy in the bioeconomy; 2019.
Accessed: 7 August, 2020.
Available:https://www.sciencedirect.com/topics/earth-and-planetary-sciences/sustainable-development.

[9] Asogwa OC, Eze NM, Eze CM, Okonkwo CI, Onwuamaeze CU. On the modeling of the effects of COVID-19 outbreak on the welfare of nigerian citizens, using network model. American Journal of Applied Mathematics and Statistics. 2020;8(2):58-63.
DOI: 10.12691/ajams-8-2-4.

[10] Yahaya, A. List of six geopolitical zones in Nigeria and their states. 2019.
Accessed: 10 October, 2020.
Available:https://nigerianinfopedia.com.ng/six-geopolitical-zones-in-nigeria-and-their-states/

[11] Wikipedia. Southern Nigeria Protectorate; 2020.
Accessed 22 September, 2020.
Availble:https://en.wikipedia.org/wiki/Southern_Nigeria_Protectorate

[12] Sakar E, Keskin S, Unver H. Using of factor analysis scores in multiple linear regression model for prediction of kernel weight in ankara walnuts. The Journal of Animal & Plant Sciences. 2011;21(2):182-185.

[13] Song H, Zhang N. Study on consumer decision making in rural tourism based on factor analysis model. Journal of Chemical and Pharmaceutical Research. 2014;6(10):722-726.
ISSN: 0975-7384.

[14] Onyeabor EN, Alimba JO. Factor analysis of influence of host-community characteristics on ecotourism development in South East Nigeria. International Journal of Development and Economic Sustainability. 2016;4(1):11-20.

[15] Aldahmash A, Gravell A, Howard Y. Using factor analysis to study the critical success factors of agile software development. Journal of Software. 2017;12(12):957-963.
DOI: 10.17706/jsw.12.12.957-963

[16] Yong AG, Pearce S. A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in Quantitative Methods for Psychology. 2013;9(2):79-94.
DOI: 10.20982/tqmp.09.2.p079

[17] Kaiser HF. A second generation little jiffy. Psychometrika.1970;35(4):401-415.

[18] Kaiser HF, Rice J. Little jiffy, mark IV. Educational and psychological measurement. 1974;34(1):111-117.

[19] Eyduran EM, Topal, Sonmez AY. Use of factor scores in multiple regression analysis for estimation of body weight by several body measurements in brown trouts (Salmo trutta fario). International Journal of Agri. and Biology. 2010;12:611-615.

[20] Hamza S, Mustaamal AHB, Kamin YB. Exploratory factor analysis of green innovative skill elements in building construction programme for economic sustainability. International Journal of Recent Technology and Engineering (IJRTE). 2019;8(2).
ISSN: 2277-3878.

[21] Ifeanyichukwu U. Use of factor scores for determining the relationship between body measurements and semen traits of cocks. Open Journal of Animal Sciences. 2012;2(1):41-44.
DIO: http://dx.doi.org/10.4236/ojas.2012.21006

[22] Kaiser HF. An index of factorial simplicity. Psychometrika. 1974;39(1):31-36.

[23] Mohamad Juahir MH, Ali NA, Kamarudin MKA, Karim F, Badarilah N. Developing health status index using factor analysis. Journal of Fundamental and Applied Sciences. 2017;9(2S):82-92.
DOI: http://dx.doi.org/10.4314/jfas.v9i2s

[24] Norusis MJ. SPSS 6.1 Base System User's Guide Part 2. SPSS, Inc, Chicago; 1994.

[25] Gorsuch RL. Factor analysis (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates; 1983.

[26] Bartlett MS. The effect of standardization on a chi-square approximation in factor analysis. Biometrika. 1951;38(3/4):337-344.

[27] Algina J, Keselman HJ. Comparing squared multiple correlation coefficients: Examination of a confidence interval and a test significance. Psychological Methods.1999;4(1):76-83.

[28] Cheung MWL, Chan W. Testing dependent correlation coefficients via structural equation modeling. Organizational Research Methods. 2004;7(2):206-223.

[29] Ghasemi A, Zahediasl, S. Normality tests for statistical analysis: A guide for non-statisticians. Int J Endocrinol Metab. 2012;10(2):486-9.
DOI: 10.5812/ijem.3505.

[30] Ogunleye LI, Oyejola BA, Obisesan KO. Comparison of some common tests for normality. International Journal of Probability and Statistics. 2018;7(5):130-137.
DOI: 10.5923/j.ijps.20180705.02.

[31] Yap BW, Sim CH. Comparisons of various types of normality tests. Journal of Statistical Computation and Simulation. 2011;81(12):2141-2155.
DOI: 10.1080/00949655.2010.520163

[32] Normadiah MR, Yap BW. Power comparisons of some selected normality tests. Proceedings of the Regional Conference on Statistical Sciences. 2010;126-138.

[33] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3/4):591-611.

[34] Haitovsky Y. Multicollinearity in regression analysis: A comment. Review of Economics and Statistics. 1969;5 (4):486-489.

[35] Tabachnick BG, Fidell LS. Using Multivariate Statististics (4th ed.). Boston, MA: Allyn and Bacon. **A** Pearson Education Company Boston, U.S.A. 2001;966.

[36] Menard S. Applied logistic regression analysis: Sage university series on quantitative applications in the social sciences. Thousand Oaks, CA: Sage; 1995.

[37] Huber E, Stephens JD. Political parties and public pensions: A quantitative analysis. Acta Sociologica. 1993;36:309-325.

[38] Hair JF Jr, Anderson RE, Tatham RL, Black WC. Multivariate data analysis (3rd ed.) New York: Macmillan; 1995.

[39] Kennedy P. A guide to econometrics. Oxford: Blackwell; 1992.

[40] Marquardt, D. W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics.1970;12:591-256.

[41] Pan Y, Jackson RT. Ethnic difference in the relationship between acute inflammation and serum ferritin in US adult males. Epidemiology and Infection. 2008;136:421-431.

[42] Rogerson PA. Statistical methods for geography. London: Sage; 2001.

[43] Kline P. An easy guide to factor analysis. New York, NY: Routledge; 1994.

[44] Guilford JP. Psychological measurement a hundred and twenty-five years later. Psychometrika. 1961;26:101-127.

[45] Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. Prentice-Hall, Englewood Cliffs; 1992.

[46] Rencher AC. Methods of Multivariate Analysis. John Wiley & Sons, Inc., New York; 1995.

[47] Costello AB, Osborne JW. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research and Evaluation. 2005;10(7):1-9.

[48] Santos R, Gorgulho BM, Alessandra M, Fisberg RM, Marchioni DM, Baltar VT. Principal component analysis and factor analysis: Differences and similarities in nutritional epidemiology application. Revista Brasileira de Epidemiologia. 2019;22.
Available:https://doi.org/10.1590/1980-549720190041

[49] Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. Fifth Edition. Prentice-Hall, Inc, Upple Saddle River; 2002.

[50] Ford JK, Mac Callum RC, Tait M. The application of exploratory factor analysis in applied psychology: A critical review and analysis. Personnel Psychology.1986;B(2):291-314.

[51] Humphreys LG, Montanelli RG. An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioral Research.1974;2:193-205.

[52] Kaiser HF. The application of electronic computers to factor analysis. Educational and Psychological Measurement. 1960;20:141-151.
DOI: 10.1177/001316446002000116.

[53] Jolliffe IT. Discarding variables in a principal component analysis. II: Real data. Journal of the Royal Statistical Society. Series C (Applied Statistics). 1973;22(1):21-31.
Available:http://www.jstor.org/stable/2346300

[54] Jolliffe IT. Discarding variables in a principal component analysis. I: Artificial data. Applied Statistics.1972;21:160-173.

[55] Cattell RB. The Scree Plot Test for the Number of Factors. Multivariate Behavioral Research.1966;1: 140-161.
DOI: http://dx.doi.org/10.1207/s15327906mbr0102_1

[56] Field A. Discovering statistics using IBM SPSS statistics. SAGE; 2013.

[57] Vogt WP. Dictionary of statistics and methodology a non-technical guide for the social sciences (2nd ed.). London Sage Publications; 1999.

[58] Rencher AC. Methods of multivariate analysis. New York: J. Wiley; 2002.

[59] Field A. Discovering statistics using SPSS. Sage publications; 2009.

[60] Stat 505, Lesson 11.5: Applied multivariate statistical analysis. Department of Statistics, Pennsylvania State University- Eberly College of Science.
Accessed: 4 October, 2020.
Available: https://online.stat.psu.edu/stat505/lesson/11/11.5.

[61] DiStefano C, Zhu M, Mîndrilã D. Understanding and using factor scores: considerations for the applied researcher. Practical Assessment, Research, and Evaluation. 2009;14(20).
DOI: https://doi.org/10.7275/da8t-4g52

[62] Stat 505, Lesson 12.12: Applied multivariate statistical analysis. Department of Statistics, Pennsylvania State University- Eberly College of Science.
Accessed: 8 October, 2020.
Available:https://online.stat.psu.edu/stat505/lesson/12/12.12.

[63] Bradley AR., Philip KH, Stuart LD, Jenks JM. The use of principal component factor analysis to interpret particulate compositional data sets. Journal of the Air Pollution Control Association.1982;32(6):637-642.
DOI: 10.1080/00022470.1982.10465439

[64] Draper NR, Smith H. Applied regression analysis. John Wiley and Sons, Inc., New York. 1998;706.

[65] De Coster J. Overview of factor analysis; 1998.
Accessed 14 February, 2020.
Available: http://www.stat-help.com/factor.pdf

[66] Yakubu A, Idahor KO, Agade Y. Using factor scores in multiple linear regression model for predicting the carcass weight of broiler chickens using body measurements. Revista UDO Agrícola.2009;9 (4): 963-967.

APPENDIX

R codes for principal component method of factor analysis

```
#"setwd ()" is a function uses to set working directory
#setwd("C:/Users/C.JAY NATECH/Documents/FACTOR ANALYSIS")
#read.csv is a function uses to import data from any spreadsheet environment to R environment
Places.Data <- read.csv("Places Data.csv", sep = ",", header = TRUE); Places.Data #To print/call
#"Places.Data"
# computation of correlation using pearson correlation method
#library(corrplot) loads "corrplot 0.84" package to create corrplot.
library(corrplot)
corrplot.mixed(cor(Places.Data), number.cex = 1.2,tl.col="black", tl.cex=0.8, lower.col = "black")
#library(MVN) loads "MVN" package which handles multivariate normality test
library(MVN)
# create univariate histograms for checking normality
hist.plot <- mvn(data = Places.Data, mvnTest = "royston", univariatePlot = "histogram")
#Normality test on the complete data
#Shapiro-Wilk Normality test
Normality_result <- mvn(data = Places.Data, mvnTest = "royston", univariateTest = "SW", desc = TRUE)
Normality_result$univariateNormality
write.csv(Normality_result$univariateNormality, file = "Shapiro-Wilk Test Result.csv")#To save/export
#"Normality_result$univariateNormality" from R to MS-Excel
# library(FactoMineR) loads "FactoMineR" package.
#This package carries out PCA method of factor analysis using "PCA ()" function
library(FactoMineR)
unrot.Places.Factors <- PCA(Places.Data, scale.unit = TRUE, graph = FALSE)
# library(factoextra) loads "factoextra" package.factoextra" is a package uses to extract PCA results #from
"FactoMineR" package. It plots different diagram/graph using results extracted from the same #"FactoMineR"
package.
library(ggplot2)#"ggplot2" is a package that helps "factoextra" in creation of diagrams
library(factoextra)
## matrix with eigenvalues extracted from "FactoMineR" package using Factoextra package
Factors.eig.val <- get_eigenvalue(unrot.Places.Factors)
print(Factors.eig.val, digit=4)# digit=4 is used to make the result in 4 s.f
write.csv(Factors.eig.val, file = "Eigenvalue Result.csv")#To export "Factors.eig.val" from R to MS-Excel
# Visualization of eigenvalues/variances
fviz_screeplot(unrot.Places.Factors, addlabels = TRUE, ylim = c(0, 46), xlim=c(0, 11), xlab = "Component
Number", main = " Scree Plot")
## matrix of factor loadings computed using "PCA ()" function in "FactoMineR" package loaded earlier
# unrot.Factor.Loadings is a variable that holds data.frame (i.e., table format) of matrix of factor loadings
unrot.Factor.Loadings <-data.frame(round(unrot.Places.Factors$var$cor[,1:3],4))
names(unrot.Factor.Loadings) <- c("F1", "F2", "F3")  #renaming of the columns of unrot.Factor.Loadings"
 unrot.Factor.Loadings #To print "unrot.Factor.Loadings"
#library(psych) loads "psych" package.
##"psych" package uses "principal ()" function to carry out PCA method of factor analysis.
##"principal ()" function uses correlation matrix by default and covariance matrix when "covar=TRUE" is
#stated
library(psych)
#unrot.Place.factor is a variable that holds results computed by "principal ()" function.
unrot.Place.factor <- principal(Places.Data, nfactors = 3, scores=TRUE, normalize=TRUE,
                        oblique.scores =FALSE,rotate = 'none', cor = "cor", method = "wls", fm="pc")
communalities <- apply(unrot.Place.factor$loadings^2,1,sum) # communality
specific.variance <- 1 - apply(unrot.Place.factor$loadings^2,1,sum) # uniqueness
complexity <- (apply(unrot.Place.factor$loadings^2,1,sum))^2/apply(unrot.Place.factor$loadings^4,1,sum)
```

```
unrot.Facto.Load.Variances <- data.frame(unrot.Factor.Loadings, communalities, specific.variance, complexity)
#data.frame
print(unrot.Facto.Load.Variances, digits = 4) #To print data frame "unrot.Facto.Load.Variances" with 4s.f
write.csv(unrot.Facto.Load.Variances, file = "unrotated Factor Loading and Variances.csv")
#rot.Place.factor is a variable that holds results computed by "principal ()" function.
rot.Place.factor <- principal(Places.Data, nfactors = 3, scores=TRUE, normalize=TRUE,
oblique.scores =FALSE,rotate = 'varimax', cor = "cor", method = "wls")
rot.Place.factor  #To print rotated factor loadings in "rot.Place.factor" from "principal ()" function
win.graph()
#library(qgrap) loads "qgraph" package. This package is use to plot qgraph which shows the lines of
#correlations
library(qgraph)
rot.Factor.Loadings_qgraph <- function(loadings_in, title) {
  ld <- loadings(loadings_in)
  qg_factor <- qgraph(ld, title=title,
layout = "spring", node.height=1.5, node.width=1.5,label.cex=1.5, posCol = "darkgreen",
negCol = "darkmagenta", arrows = FALSE,labels=attr(ld, "dimnames")[[1]])
  qgraph(qg_factor, title=title,
        posCol = "darkblue", negCol = "darkred", arrows = FALSE, node.height=1.5, node.width=1.5,
vTrans=255, edge.width=1, label.cex=1.5, width=2, height=2, normalize=TRUE)
}
rot.Factor.Loadings_qgraph(rot.Place.factor, "  ") #To print the qgraph of factor loadings matrix from the
#function "rot.Factor.Loadings_qgraph". Here this factor loadings matrix is computed using #"principal ()"
function in "psych" package
# rot.Factor.Loadings is a variable that holds data.frame (i.e., table format) of matrix of factor loadings
rot.Factor.Loadings <-data.frame(rot.Place.factor$loadings[,1:3])
names(rot.Factor.Loadings) <- c("RF1", "RF2", "RF3")  #renaming of the columns of # "unrot.Factor.Loadings"
table
rot.Factor.Loadings #To print "unrot.Factor.Loadings" communalities <
apply(rot.Place.factor$loadings^2,1,sum) # communality
specific.variance <- 1 - apply(rot.Place.factor$loadings^2,1,sum) # uniqueness
complexity <- (apply(rot.Place.factor$loadings^2,1,sum))^2/apply(rot.Place.factor$loadings^4,1,sum)
rot.Facto.Load.Variances <- data.frame(rot.Factor.Loadings,communalities, specific.variance,complexity)
print(rot.Facto.Load.Variances, digits = 4)#To print data frame "unrot.Facto.Load.Variances" with 4 s.f
write.csv(rot.Facto.Load.Variances, file = "Rotated Factor Loading and Variances.csv")#To save/export
#"unrot.Facto.Load.Variance" from R to MS-Excel
win.graph()
#libray("plot3D") loads "plot3D" package. This package is used to draw 3D plot
library("plot3D")
with(rot.Factor.Loadings, scatter3D(rot.Factor.Loadings$RF1, rot.Factor.Loadings$RF2,
                          rot.Factor.Loadings$RF3, pch = 10,
                          xlab = "Rotated factor1 (32.081%)", ylab = "Rotated factor2 (24.560%) ",
                          zlab = "Rotated factor3 (14.989%) ",
                          labels = rownames(rot.Factor.Loadings), colvar = rot.Factor.Loadings$FR1,
                          theta = 110, phi = 20,col = gg.col(10),
                          main = "Rotated factor loadings", cex = 1.5, colkey = TRUE,
                          bty = "g", ticktype = "detailed", d = 10,
                          clab = c("Correlation","Values","(-1 < r < 1 )"), adj = 0.5, font = 2))
# Add text
text3D(rot.Factor.Loadings$RF1, rot.Factor.Loadings$RF2, rot.Factor.Loadings$RF3,
      labels = rownames(rot.Factor.Loadings), add = TRUE, colkey = FALSE, cex = 1,
      colvar = rot.Factor.Loadings$RF1, col = gg.col(10),
      clab = c("Correlation","Values","(-1 < r < 1 )"), adj = -0.2, font = 2)

# Add points
with(rot.Factor.Loadings, scatter3D(rot.Factor.Loadings$RF1, rot.Factor.Loadings$RF2,
                          rot.Factor.Loadings$RF3 - 0.05,
```

```
                    colvar = rot.Factor.Loadings$RF1, col = gg.col(10),
                    type = "h", pch = " ",add = TRUE))
#rot.facto.weight is a variable that holds the weights of rotated factor
rot.facto.weight <- rot.Place.factor$weights
rot.facto.weight
write.csv(rot.facto.weight, file = "weight of rotated factors.csv") #To save "rot.facto.weight" into Excel
#Factor.Score is a variable that holds rotated factor scores
Factor.Score <- rot.Place.factor$scores

Factor.Score

write.csv(Factor.Score, file = "rotated factor scores.csv")
############################## REGRESSION ANALYSIS ##############################
#fitting multiple regression using development score (i.e., Pc Score 1) as dependent variable and factor
#scores as independent variable
DevScores_FactorScores <-read.csv("Dev Scores and Factor Scores.csv", sep = ",", header = TRUE)
head(DevScores_FactorScores) #To print/call "Factor and PC Scores.csv"
PCScores_FactorScores <-read.csv("PC SCORE AND FACTOR SCORES.csv", sep = ",", header = TRUE)
head(PCScores_FactorScores) #To print/call "Factor and PC Scores.csv"
#Reg_Dev_Factor_Scores is a variable that holds regression results
Reg_Dev_Factor_Scores <- lm(DSScore ~ FactorScore1 + FactorScore2 + FactorScore3,
data=PCScores_FactorScores)
summary(Reg_Dev_Factor_Scores)
#library(olsrr) is a package used to compute Tolerance and VIF
library(olsrr)
ols_vif_tol(Reg_Dev_Factor_Scores)
```


Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/67074