Dimensionality in canonical correlation of settlements characteristics and levels of infrastructure in rural Akwa Ibom State, Nigeria in 2014

Jacob Atser*[1] and Uwem Ituen[2]

1) Department of Urban and Regional Planning, University of Uyo, Uyo, Nigeria
2) Department of Geography and Natural Resource Management, University of Uyo, Uyo, Nigeria

**Author for Correspondence:* jacobatser@yahoo.com

Received: 7 Jul. 2015, Revised: 10 Oct. 2015, Accepted: 25 OCt. 2015

ABSTRACT

This study examines the relationship between rural settlements characteristics and access levels to infrastructural facilities in Akwa Ibom State. A total of 50 rural settlements were randomly selected and utilized for this study. Access levels to five basic social infrastructural facilities and six variables of settlements characteristics were examined simultaneously using multivariate method of canonical correlation analysis. The question this research set to address is: along how many dimensions are the settlements characteristics related to the levels of infrastructural facilities? Five canonical functions were produced out of which the first two were found to be significant at 0.001 levels. The first canonical correlation coefficient of 0.88 (0.86 adjusted), representing 78% overlapping variance for the first pair of canonical functions or variates was obtained while the second canonical correlation coefficient was 0.59 (0.50 adjusted) and thus, represent 34% overlapping variance for the second pair of canonical variates. These canonical correlations are highly significant and represent a substantial relationship between pairs of canonical variates. The result showed that in Akwa Ibom State, rural settlements characteristics and basic social infrastructure can be related at least in two major dimensions. This study therefore reaffirms the potentials of canonical correlation analysis as a useful tool for establishing empirically based linkages between two data sets.

Keywords: Rural Settlements, Characteristics, Infrastructural Facilities

INTRODUCTION

Social infrastructure cover such basic services as education, health, water, electricity, communication and transportation services, housing and other social services needed to facilitate industrial, agricultural and other socioeconomic development[1]. Social infrastructure is an umbrella term for many activities referred to as social overhead capital by development economists such as Paul Rosenstein Rodan, Ragnar Nurske and Albert Hirschman[2, 3, 4 and 5]. Social infrastructural facilities are the arteries and channels of rural development. There is a very close relationship between rural social infrastructural facilities and socio-economic development. Social infrastructure services are central to the activities of households and economic production. This reality becomes painfully evident when natural disaster or civil disturbances destroys roads, culverts, bridges, electricity lines, water mains etc. In such circumstances, communities' quality of life and productivity becomes radically reduced. Conversely, adequate provision of social infrastructure services enhances welfare and fosters economic growth. Thus, providing infrastructure services to meet the demands of households, businesses and other users is central in contemporary development discourse. This is because adequacy of social infrastructure helps determine one country's success and another's failure in diversifying production, expanding trade, coping with population growth, reducing poverty, improving standards of living and environmental conditions[6, 2, 7 and 8].

Although, access to basic social infrastructure is vital to human settlements existence and sustainability, very few studies exist to investigate the relationship between settlements characteristics and infra-structural provision in an appropriate and rigorous way. In Nigeria, most studies on social infrastructure are addressed in piece meal fashion instead of groups of facilities considering their linkage disposition and complex interrelationships. This could constrain generalization of results. In some cases, descriptive analytical procedures are utilized. It is against this background that this study utilized several social infrastructural facilities and several spatial factors that defined the characteristics of rural settlements and try to link them together simultaneously using multivariate canonical correlation analysis. The emphasis of canonical correlation analysis is the identification of the structure of multivariate relationships and the generation of the maximum amount of correlation between the linear combinations of two groups of variables [9]. Few works exist that try to investigate the relationship using canonical correlations especially in Nigeria. Olafiaji [10] applied canonical correlation analysis to investigate land use and trip generation patterns in Akure, Nigeria and was able to associate 32.43 % of trip generations with land use activities. The goal of canonical correlation is to rigorously analyze the relationships between two sets of variables. This study has as its objective to investigate the dimensions in which the settlements spatial characteristics relate to their levels of facilities. This is because social infrastructural facilities are the arteries and channels of development. There is a very close relationship between rural social infrastructure facilities and socio-economic development.

MATERIAL AND METHODS

The study is focused on Akwa Ibom State. Data were collected on social infrastructural facilities from spatially sampled rural areas across the state using both primary and secondary sources. To facilitate the selection of samples, the map of Akwa Ibom State drawn on a scale of 1cm to represent 2.5km was divided into quadrates of 0.5cm2 and numbered serially. A table of random numbers was used to select rural communities as units of observation from sampled quadrates. Because the grid map contains names of localities (villages) on it, it was easy to know the localities within each quadrate. It was possible for a quadrate to contain a number of communities; nevertheless only one community per quadrate was selected, with preference given to the communities with the largest population. Where a sampled quadrate was an urban area, which in this case is the local government or state headquarters, such a quadrate was skipped purposefully considering the rural focus of the study. Area (Spatial) sampling was preferred to point sampling because in rural areas, facilities are often provided for groups of villages, which spread across larger areas, rather than for just one village. Thus, the rural nature of this study justifies the adoption of spatial sampling framework.

The data requirements for this study were obtained primarily from the field using a structured questionnaire and field observations. A set of questionnaires was administered to the village heads. Official records from establishments were useful sources of data while relevant information from published and unpublished sources was also useful secondary sources. Thus, the data requirements which cut across socio-economic and environmental attributes were related to the following measurable dependent variables:

1.Main source of drinking water supply in sampled communities
2.Number of water boreholes in sampled areas
3.Distance to nearest major source of water supply
4.Presence of electricity power supplies in the community
5.Presence of telephone service in the community
6.Most common means of transportation
7.Type of road leading to the community(tarred or untarred)
8.Category of road leading to the community (federal, state or local)
9.Intensity of usage of the major road leading to the community
10.Number of primary schools in the community
11.Number of secondary schools in the community
12.Distance to the nearest primary school in the community
13.Types of health facilities in the community
14.Ownership status of health care facilities
15.Number of hospital beds
16.Number of Doctors
17.Number of Nurses
18.Distance to nearest health care facility
19.Distance to nearest market
20.Distance to nearest police station
21.Distance to nearest bank

Performance of sampled communities from all these variables were measured, scored and summed up as total performance in social infrastructure (infrastructure stock) in each sampled community. Six independent variables were used. These are: population size of sampled community; distance of sampled community to nearest urban centres; population size of the nearest urban centre; distance of sampled community to nearest highway; topological accessibility index of communities; and level of nucleation of communities. To determine and analyze the level of access to social infrastructure development in each sampled community, the following procedures were employed to evolve an index for each of the five social sectors considered in this study:

Potable water supply infrastructure: Here, three indicators (Surrogates/variables) were used for determining and analyzing the level of access to potable water supply in sampled communities: the major mode of drinking water supply, distance to nearest source and number of water boreholes provision. Scores derived from these indicators were summed up to obtain an index of potable water supply performance.

Road infrastructure: the quality of the road leading to sample communities was measured in terms of type of roads (paved or un-paved), categories of roads, major means of transportation and usage intensity of roads as indicators or surrogates for assessing road infrastructure development. Scores from the four variables were summed up to obtain an index of level of rural road infrastructure development in each sampled community.

Education infrastructure: The indicators in this sector include the number of primary schools, number of secondary schools, and distance to nearest primary school. The scores on these three variables were summed up to obtain an index of level of access of education infrastructure development in each sampled community.

Health facilities: under health sector, a total of six indicators were considered: Type of health care facility, ownership of healthcare facility, number of hospital beds, number of medical personnel (Doctors and Nurses) and distance to the nearest health care facility. The scores from these surrogates were summed up to obtain an index of level of development in the health sector.

Other facilities: Five indicator variables/surrogates were used in this sector. These are electricity supply, telephone service, banking facility, police and market services. A nominal scale of '1' for availability and 'o' for non-availability was employed to measure electricity and telephone services in the study area. Accessibility to bank, police and market services was measured in terms of the distance (km) it takes to access the services. The scores from these indicators were summed up to evolve an index for other facilities. Because these variables relate to issues of availability and accessibility, the initial concern was to determine whether or not a particular facility is available while the next consideration was on its level of accessibility, in terms of distance measured in kilometers or time spent in accessing a facility as well as the supportive population. The study also made use of secondary data obtained from official records of government establishments such as Akwa Ibom Rural Water and Sanitation Agency (AKRUWATSAN), Akwa Ibom State Ministries of Health, Education, Works and Transport and Economic Planning as well as Power Holding Company. Information relating to number of public water boreholes in rural areas, number of health facilities, number of educational infrastructure, total length of federal, state and local government roads was obtained. The data requirements for this study were obtained primarily from the listed dependent and independent variables. Table 1 provides a summary on the dependent variables. Data on the spatial factors of settlement characteristics were obtained from map work analysis and official records of relevant establishments.

The data from the two sets of variables were analyzed using canonical correlation. Canonical correlation analysis is the multivariate statistical technique generally used for investigating two different sets of multivariate data derived from a single population, with the goal of discerning the inter-relationship existing between the linear combinations of the first data set and the linear combinations of the second data set. The essence of the technique is to derive a linear combination for each of the multivariate data sets in such a way that the correlation between these two linear combinations is maximized. The first data set Y1, Y2, Y3 … Yn is called the Criteria variables or dependent variables while the second data set X1, X2, X3 … Xn is called the Predictor Variables or independent variables. This technique was applied to extract linear combinations for the criteria and linear combinations for the predictor in such a way that when these two sets of linear combinations are correlated a maximum amount of correlation between these two data sets are obtained.

RESULTS AND DISCUSSION

Table 2 provides the field data on the dependent and independent variables. the data as presented in Table 2, variations in settlement characteristics as well as in levels of infrastructure are discernable. While some communities recorded low index values for some facilities especially water supply and basic education, others had negative index values which by extension, signifies adversities. In terms of settlement characteristics, some communities are more favourably disposed than others. The implication of this development is that a discernable pattern of relationship is likely to emerge to offer further explanation on observable pattern of development in rural Akwa Ibom State. Thus canonical correlation was employed to examine the relative contribution of settlement characteristics in predicting changes in levels of settlement facilities.

Interpretation of Canonical Variates or functions

Interpretation of significant pairs of canonical variates is based on the loading matrices. Each pair of canonical variates is interpreted as a pair, with a variate from one set of variables interpreted vis-a-vis the variate from the other set. A variate is interpreted by considering the pattern of variables highly correlated (loaded) with it. Because the loading matrices contain correlations, and because squared correlations measure overlapping variance, variables with correlations of more than 0.30 (9% of variance) are usually interpreted as part of the variate, and variables with loadings of 0.30 and below are not. Deciding on a cutoff for interpreting loadings is, however, somewhat a matter of taste, although there are clear guidelines. Comrey and Lee [13] suggest that loadings in excess of 0.71 (50% overlapping variance) are considered excellent, 0.63 (40% overlapping variance) very good, 0.55 (30% overlapping variance) good, 0.45 (20% overlapping variance) fair, and 0.32 (10% overlapping variance) poor. Choice of the cutoff for size of loading to be interpreted is a matter of preference [14].

Most researchers do not interpret pairs with a canonical correlation lower than 0.30 because rc values of 0.30 or lower represent less than a 10% overlap in variance. In this study, loadings in excess of 0.45(20% overlapping variance) were interpreted. Canonical correlation analysis investigates two different sets of multivariate data derived from a single population in order to discern the inter-correlation between the linear combinations of the first data set and those of the second data set. The number and importance of canonical variates or functions are determined using the output from Table 3. The significance of the relationships between sets of variables is reported directly by SAS CANCORR, as shown in Tables 4 and 5 respectively. With all four canonical correlations included, $F(30, 158) = 3.94$, $P< 0.001$. With the first and second canonical correlations removed, F values are not significant: $F(12, 108.77) = 1.08$, $p=.38$. Therefore, only significant relationships are in the first two pairs of canonical variates and these are interpreted. Canonical correlations (rc) and eigenvalues (r2c) are also in Table 4. The first canonical correlation is .88 (0.86 adjusted), representing 78% overlapping variance for the first pair of canonical variates . The second canonical correlation is 0.59 (0.50 adjusted), representing 34% overlapping variance for the second pair of canonical variates. These canonical correlations are highly significant and represent a substantial relationship between pairs of canonical variates.

Table 1: Measurable indicators of the Variables used for Social Infrastructure Assessment

Sector		Variables	Unit of measurement	Standard required (expected)
1	Water (W)	(a)Major source	Borehole(3), well(2), stream/river/pond(1)	Borehole [11]
		(b)Distance	Time	30 minutes [11]
		(c)Borehole	Number / community	1/250 population [11]
2	Health (H)	(A)Types	Hospital(4), Health centre(3), Clinic(2), Disp.(1)	Base on population of community
		(B)Doctors	Number/health facility	Base on population of community
		(C)Nurses	Number / health facility	Base on population of community
		(D)Ownership status	Government(3), community(2), private(1	Government ownership
		(E)Hospital beds	Number / health facility	Base on population of community
		(F)Distance	Kilometers	Base on type of health facility/community
3	Education (E)	(a)Primary	Number	1/3000population [12]
		(b)Secondary	Number	1/12000population [12]
		(c)Distance to primary	Kilometers	2.5 kilometers as maximum
4	Road (R)	(a)Category	Federal(3), State(2), Local(1)	Federal
		(b)Types	Paved(1), unpaved(0)	Paved
		c)Mode of transport	Motorized(3), bicycle(2), on foot(1)	Motorized
		(d) Usage intensity	High(3), Moderate(2), Low(1)	High
5	Others (O)	(B)Nearness to bank		
		(P)Nearness to police	<500m(5), 500-1km(4), 1.1km-3kms(3),	
		(M)Nearness to market	3.1kms-5kms(2), >5kms(1)	<500m
		(E)Electricity supply	Available(1), not available (0)	Availability
		(T)Telephone (GSM)	Available(1), not available (0)	Availability
Water supply Index		Level of achievement for a, b, c = observed ÷ expected x 1. Index =levels of achievement for a , b, c ÷ 3		
Health Index		Level of achievement for a, b, c, d, e, f = observed ÷ expected x 1. Index = levels of achievement for a, b, c, d, e, f ÷ 6		
Education Index		Level of achievement for a, b, c = observed ÷ expected x 1. Index =levels of achievement for a , b, c ÷ 3		
Road Index		Index = summation of levels of achievement for a, b, c ÷ 10		
Others Index		Index = summation of scores for B, P, M, T, E ÷ 17		

Table 2: Data for X and Y variables

s/n	Y1	Y2	Y3	Y4	Y5	X1	X2	X3	X4	X5	X6
1	0.49	5.40	6.20	1.0	0.5	2190	30.5	42851	3.9	178	0.76
2	-0.56	2.10	-1.02	0.60	0.6	3190	8.6	74273	0.01	169	0.84
3	1.59	1.50	2.70	0.60	0.5	2937	13.5	361761	0.5	194	1.46
4	-0.67	3.90	3.20	0.60	0.1	5097	14.9	42581	8.5	195	1.38
5	-0.67	0.60	-1.10	1.0	0.1	2289	25.7	37368	0.01	205	1.7
6	0.44	2.20	6.80	0.70	0.4	1944	9.4	143767	20.3	182	1.31
7	-0.41	1.30	-1.20	0.70	0.4	4679	24.4	71012	11.1	193	0.86
8	0.47	1.70	8.50	0.70	0.5	3475	7.3	65867	9.3	241	0.65
9	2.40	2.00	3.70	0.70	0.7	1869	9.6	71012	5.5	197	0.45
10	-1.67	3.40	-2.50	0.70	0.4	1501	21.7	143767	15.1	163	1.74
11	1.67	3.40	-0.30	0.60	0.4	2533	24.1	42581	2.5	193	1.72
12	2.30	0.60	4.10	1.0	0.7	4205	9.7	143767	0.01	220	1.26
13	-1.67	1.70	4.20	1.60	0.4	12266	17.5	74273	17.6	236	1.54
14	1.64	3.20	-0.50	0.70	0.5	798	16.2	71012	7.6	159	0.84
15	1.89	2.70	1.50	0.70	0.7	893	15.6	143767	6.4	234	1.32
16	-1.67	0.60	1.10	1.0	0.4	3213	23.2	71012	0.01	187	1.68
17	2.70	5.0	6.42	0.70	0.7	5408	4.1	361761	0.02	182	0.31
18	1.96	2.90	4.10	1.0	0.4	793	9.2	42581	0.01	155	0.58
19	1.50	4.80	7.60	1.0	0.5	4347	7.5	65867	0.01	184	1.12
20	0.45	3.9	0.10	1.0	0.4	10582	17.3	65867	0.02	235	1.38
21	-1.67	2.9	0.30	0.60	0.4	10362	12.4	37368	2.5	210	1.66
22	1.47	2.00	6.50	0.90	0.5	6986	8.7	74273	0.01	199	1.22
23	-0.60	1.40	8.30	0.60	0.4	2772	8.6	361761	3.1	201	0.72
24	-1.67	2.20	-2.00	1.0	0.4	1063	25.5	143767	2.6	218	1.76
25	-0.32	0.10	0.50	0.60	0.4	3624	19	65867	12.5	272	1.68
26	1.80	1.10	4.50	0.80	0.7	2112	9.7	74273	6.4	204	1.35
27	2.38	2.10	1.40	0.60	0.5	3849	13.6	74273	1.8	194	1.34
28	1.21	2.40	-2.00	0.70	0.5	567	19.3	143767	8.8	239	1.44
29	1.18	5.50	8.20	1.0	0.8	7049	6.5	361761	0.01	196	0.32
30	0.64	1.50	-0.00	0.90	0.4	3234	11.4	143767	0.01	215	1.62
31	-1.67	1.20	0.50	0.90	0.3	2645	25.2	42581	6.6	182	1.54
32	1.78	2.60	8.80	1.0	0.6	1672	7.4	71012	0.01	193	0.64
33	0.51	2.10	6.50	0.80	0.4	4256	8.7	42581	0.01	183	1.26
34	0.93	1.70	14.00	0.80	0.7	2096	4.8	361761	0.01	165	0.33
35	-1.67	1.50	-0.9	0.60	0.5	1721	26.9	143767	3.5	205	1.74
36	-1.67	1.90	8.10	0.60	0.4	1509	14.6	143767	5.4	238	1.22
37	1.86	3.30	1.40	0.80	0.6	583	14.1	37368	10.4	256	1.18
38	1.60	1.90	7.20	0.90	0.4	2803	11.8	42581	2.6	200	1.12
39	1.61	2.80	9.10	0.90	0.6	2454	6.6	71012	0.01	168	0.38
40	-0.37	2.80	0.60	0.90	0.4	839	21.5	65867	0.2	286	1.48
41	2.27	5.10	6.70	1.0	0.6	4659	5.3	74273	0.01	196	0.58
42	1.44	4.1	8.30	0.70	0.4	4538	8.9	74273	8.4	193	0.62
43	0.22	3.90	8.00	0.70	0.4	5773	8.5	361761	8.6	184	0.72
44	2.60	1.40	-0.50	0.70	0.4	1663	22.7	71012	13.1	173	1.38
45	-0.60	1.20	6.20	0.50	0.4	3492	18.5	143767	11.5	172	1.32
46	1.78	1.50	-2.50	1.0	0.6	1674	23.5	143767	0.5	208	1.67
47	0.43	6.9	9.50	0.90	0.7	5148	4.5	361761	0.1	184	0.3
48	1.41	1.40	-0.00	1.0	0.6	3296	22.5	74273	0.5	190	1.46
49	-1.67	1.70	3.20	0.50	0.2	2208	23.1	42581	5.6	184	1.64
50	-0.64	4.40	8.40	0.80	0.7	8884	14.2	65867	8.5	240	0.65

Key to Table 2: X1= Population size of settlements; X2= Distance of settlements to nearest urban centres; X3= Population size of nearest urban centres; X4= Distance of settlements to nearest highway; X5= Accessibility index of settlements; X6= Level of nucleation of settlements; Water supply (Y1); Healthcare services (Y2); Education services (Y3); Road infrastructure (Y4); other infrastructural facilities (Y5)

Table 3: Canonical correlation Analysis

	Canonical correlation	Adjusted canonical correlation	Appro. Std. Error	Squared canonical correlation
1	0.882890	0.860574	0.031501	0.779495
2	0.586835	0.496140	0.093661	0.344375
3	0.384416	0.215577	0.121746	0.14776
4	0.309221	.	0.129198	0.095617
5	0.193135	.	0.137528	0.037301

Table 4: Test of Hypothesis

	Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)				Test of Ho: the canonical correlations in the current row and all that follow are zero				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**	**Likelihood Ratio**	**Appro. F Value**	**Num DF**	**Den DF**	**Pr >**
1	3.5350	3.0098	0.8074	0.8074	0.10726829	3.94	30	158	<.0001
2	0.5253	0.3519	0.1200	0.9274	0.48646571	1.62	20	133.61	0 .0562
3	0.1734	0.0677	0.0396	0.9670	0.74198756	1.08	12	108.77	0 .3822
4	0.1057	0.0670	0.0241	0.9912	0.87064812	1.00	6	84	0.4283
5	0.0387	.	0.0088	1.0000	0.96269899	0.83	2	43	0.4416

Table 5: Multivariate Statistics and F Approximations: S=5 M=0 N=18.5

Statistic	**Value**	**F value**	**Num DF**	**DenDF**	**Pr >F**
Wilks' Lambda	0.10726829	3.94	30	158	<.0001
Pillai's Trace	1.40456362	2.80	30	215	<.0001
Hotelling-Lawley Trace	4.37817223	5.52	30	93.412	<.0001
Roy's Greatest Root	3.53503743	25.33	6	43	<.0001

Note: F statistic for Roy's Greatest Root is an upper bound but F statistic for Wilks' Lambda is exact**.**

The multivariate test for all the canonical roots was used to evaluate the significance of the canonical correlation using a cut-off mark of 0.45 loading. The canonical correlations of selected canonical functions were significant on Wilk's Lambda, Pillars Trace, Hotelling Lawley Trace and Roy's Greatest Root statistics (Table 5) at probability level of ≥ 0.001.

Loading matrices between canonical variates and original variables are in the table on the canonical structure (Table 6). Interpretation of the two significant pairs of canonical functions or variates is based on their loadings. Correlations between variables and variates (loadings) in excess of .45 are interpreted. Both the direction of correlations in the loading matrices and the direction of scales of measurement is considered when interpreting the canonical variates. The first pair of canonical variates has high loadings on Y1(0.516), Y2(0.554), Y3(0.874), and Y5(0.664) respectively on the infrastructure data set and on X2(-0.869), X3(0.502) and X6(-0.945) on the spatial factors side. Thus, access to water supply, health facilities, basic education and other basic facilities is related to distance from the nearest urban centre, population of the nearest urban centre and level of nucleation of rural communities. The second pair of canonical variates has high loadings on Y1(-0.490), Y2(0.533), and Y4 (0.582) on the infrastructure data set and X1(0.777) on the spatial factors data set. This implies that access to water supply, health facilities, and the quality of the road network is related to the population of rural communities. The canonical functions selected represent models that explain the pattern of linkage between the two sets of data (Figs 1 and 2).

Canonical Varietes/Functions

These pairs of linear combinations for both the X and Y variables are known as canonical varietes and are similar to components in principal components analysis. The canonical varietes are extracted to help account for the maximum amount of correlation between the two sets of data while the first canonical varietes from the two sets of data (X and Y variables) provide the highest inter correlation that could be possible, the second pair of canonical variates are obtained on the basis of residual variance. The same goes for the third, fourth and so on thereby making canonical correlation coefficients to become successfully smaller.

While the first pair of canonical variates explain the highest possible inter correlations between the two data sets, the second pair of canonical variates accounts for the maximum amount of correlation between the two data sets left un-accounted for by the first pair of canonical variates, and so on. This important attributes of canonical correlation analysis enables it to produce pairs of variates that are orthogonal or independent of the preceding pairs of variates [9].

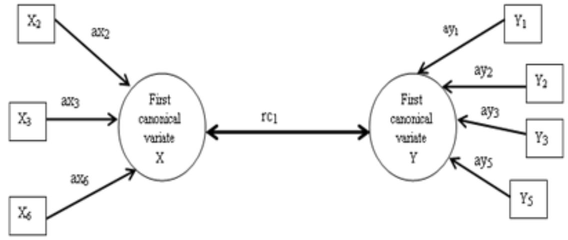
Table 6: Canonical Structure

Correlations between the Stock of infrastructure(Y variables) and Their Canonical Variables					
	V1	**V2**	V3	V4	V5
Y1	0.5163	-0.4895	-0.4681	0.4473	-0.2732
Y2	0.5542	0.5325	-0.1449	-0.1179	-0.6119
Y3	0.8744	0.0569	-0.0363	-0.3375	0.3421
Y4	0.0100	0.5820	-0.5955	0.3781	0.4046
Y5	0.6637	0.0410	0.2404	0.7061	-0.0382
Correlations between the Spatial factors(X variables) and Their Canonical Variables					
	W1	**W2**	**W3**	**W4**	**W5**
X1	0.1368	0.7770	-0.1795	-0.2791	0.5002
X2	-0.8693	0.2034	0.1278	0.0849	-0.3939
X3	0.5021	0.0937	0.7130	0.0391	-0.0352
X4	-0.2318	-0.0687	0.3849	-0.6840	0.1172
X5	-0.2142	0.1747	0.2455	0.5209	0.5793
X6	-0.9448	-0.1098	0.0686	0.0322	0.2671

Canonical function consists of pairs of variate; one as dependent variable and the other as the independent variable. The number of canonical functions generated corresponds with the number of variables in the smaller data set. However, it is the first canonical function that is of the utmost importance to the analysis as it provides the highest possible inter-correlation between the two data sets. The second, third and so forth are diminished in importance as they take care of the residual variance. For this study and as Table 4 shows, two canonical functions are selected out of the maximum number of 5 extracted. This was on the basis of the magnitude of the canonical correlation and level of statistical significance of the functions. Canonical structure matrix provides a better option for interpreting the correlation of the original variates and canonical variates extracted. A cut-off mark of 0.45 was considered for interpreting the canonical loadings. All the variables in the first data set and all those in the second data set with the exception of y4, x1, x4, and x5 load highly on the first canonical variate. This canonical variate can be described as relating wholly the basic infrastructure to the settlement attributes.

How much variance does each of the canonical variates extract from the variables on its own side is shown in the output in Table 7. How much variance the canonical variates from the independent variables extract from the dependent variables and vice versa is known as redundancy. Thus redundancy in canonical variate is the percentage of variance it extracts from its own set of variables times the squared canonical correlation for the pair. From Table 7, the values for the first pair of canonical variates are 0.356 for the first set of variables and 0.337 for the second set of variables. That is, the first canonical variate pair extracts 36% of variance from the infrastructure variables and 34% of variance from the spatial factors set of variables. The values for the second pair of canonical variates are 0.173 for the first set of variables and .117 for the second set; the second canonical variate pair extracts 17% of the variance from the first set of variables and 12% of variance from the spatial factors set of variables. Together, the two canonical variates account for 53% of the variance (36% plus 17%) in the infrastructure set of data, and 46% of the variance (34% and 12%) in the spatial factors set.

Redundancies for the canonical variates are found in Table 7. The first infrastructure variate accounts for .277 (28%) of the variance in the spatial variables, and the second infrastructure variate accounts for .060 (6%) of the variance. Together, two infrastructure variates explain 34% of the variance in the spatial factors set of variables. The first settlement spatial factors variate accounts for 0.262 (26%) and the second 0.040 (4%) of the variance in the infrastructure data set. Together, two spatial factors variates overlap the variance in the infrastructure set by 30%. Table 8 shows summary of information from this analysis appropriate for inclusion in a journal article. Shown in the table are correlations between the variables and the canonical variates, standardized canonical variate coefficients, within-set variance accounted for by the canonical variates (percent of variance), redundancies, and canonical correlations. Standardized canonical coefficients are derived from Table 9.



x_2 Variable in X data set
y_1 Variable in y data set
ax_2 Loading of correlation with ith x variable on canonical variate x
ay_1 Loading of correlation with ith y variable on canonical variable y
rc_1 Canonical correlation for the first pair of canonical variates

Fig 1: Relationships among variables, canonical variates and the first pair of canonical variates.

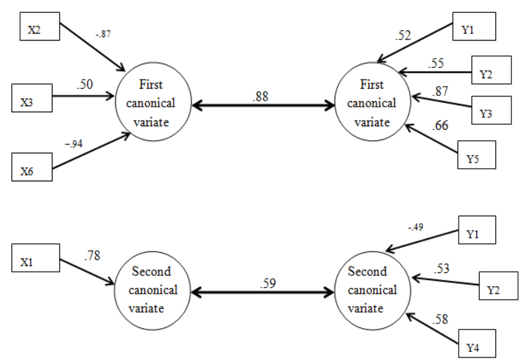


Fig 2: Loadings and canonical correlations for both canonical variate pairs for the data in Table 6.

The first four canonical functions or variates for the first set of data are labeled V1 through V5 while the canonical functions or variates for the second set are labeled W1 through W5. The two significant functions or variates are displayed graphically in Fig. 3. Thus, the graph shows the scatter plots that are between the first and second pairs of canonical functions or variates respectively. V1 is canonical variate scores for the first set in the first variate and W1 is canonical variate scores for the second set in the first variate. V2 is the canonical variate or function scores for the first set in the second variate, and W2 is the canonical variate scores for the second set in the second variate (Fig. 3). The shape of the scatter plots reflects the high canonical correlations for the solution except the second of variates where the nearly circular pattern is distorted by extreme values. From the graphs, assumptions underlying the applicability of canonical correlation could be assessed. Linearity of relationship between variables and normality of their distributions were assessed. Assumptions regarding within-set multicollinearity were met. There were no obvious departures from linearity or homoscedasticity because the overall shapes of the graphs do not curve. Deviations from normality are evident for both pairs of canonical variates. On both plots, the zero to zero point departs from the centre of the horizontal axes; but reflects on the vertical axes. This implies that normality occurs on the infrastructure component of the data set only. For the first plot (first canonical function), there is a pile up of cases at low scores than at high scores on both axes, indicating positive skewness. In the second plot, there are widely scattered cases with extremely high scores indicating positive skewness thereby confirming departures from normality.

Table 7: Canonical Redundancy Analysis

Standardized Variance of the Stock of infrastructure(Y variables) Explained by					
	Their own canonical variables			The opposite canonical variables	
Canonical variable number	Proportion	Cumulative proportion	Canonical R-square	Proportion	Cumulative proportion
1	0.3558	0.3558	0.7795	0.2773	0.2773
2	0.1734	0.5291	0.3444	0.0597	0.3370
3	0.1308	0.6599	0.1478	0.0193	0.3563
4	0.1939	0.8538	0.0956	0.0185	0.3749
5	0.1462	1.0000	0.0373	0.0055	0.3803
Standardized Variance of the Spatial factors(X variables) Explained by					
1	0.3365	0.3365	0.7795	0.2623	0.2623
2	0.1169	0.4533	0.3444	0.0402	0.3025
3	0.1283	0.5817	0.1478	0.0190	0.3215
4	0.1378	0.7195	0.0956	0.0132	0.3347
5	0.1379	0.8574	0.0373	0.0051	0.3398

Table 8: Canonical correlations, Standardized canonical coefficient, Percentage variance and Redundancies between the x and y variables and their canonical variates

	First canonical variate		**Second canonical variate**	
	Correlation	**Coefficient**	**Correlation**	**coefficient**
Y1	.52	.19	-.49	-.74
Y2	.55	.20	.53	.60
Y3	.87	.68	.06	-.12
Y4	.01	-.12	.58	.54
Y5	.66	.30	.04	.23
% variance	**.36**		**.17**	**Total= .53**
Redundancy	**.28**		**.06**	**Total= .34**
X1	.14	-.02	.78	.90
X2	-.87	-.39	.20	.89
X3	.50	.11	.09	.17
X4	-.23	-.02	-.07	-.17
X5	-.21	.06	.17	.14
X6	-.94	-.65	-.11	-.59
% variance	**.34**		**.12**	**Total= .46**
Redundancy	**.26**		**-.04**	**Total= .30**
Canonical correlation	**.88**		**.57**	

Table 9: Standardized canonical coefficients

Standardized Canonical Coefficients for the Stock of infrastructure(Y variables)					
	V1	V2	V3	V4	V5
Y1	0.1863	-0.7434	-0.8429	0.1385	-0.3053
Y2	0.2012	0.6041	-0.1289	-0.2265	-0.8521
Y3	0.6786	-0.1219	-0.0474	-0.5805	0.6315
Y4	-0.1151	0.5362	-0.6453	0.3212	0.4533
Y5	0.3015	0.2264	0.8354	0.8412	0.1102
Standardized Canonical Coefficients for the Spatial factors(X variables)					
	W1	W2	W3	W4	W5
X1	-0.0201	0.9031	-0.2133	-0.3263	0.2880
X2	-0.3890	0.8910	0.2716	0.2429	-0.9759
X3	0.1137	0.1657	0.9592	0.0686	-0.0376
X4	-0.0154	-0.1673	0.4198	-0.7706	0.1175
X5	0.0643	0.1431	0.3172	0.7009	0.4213
X6	-0.6538	-0.5925	0.0541	-0.2041	0.7953

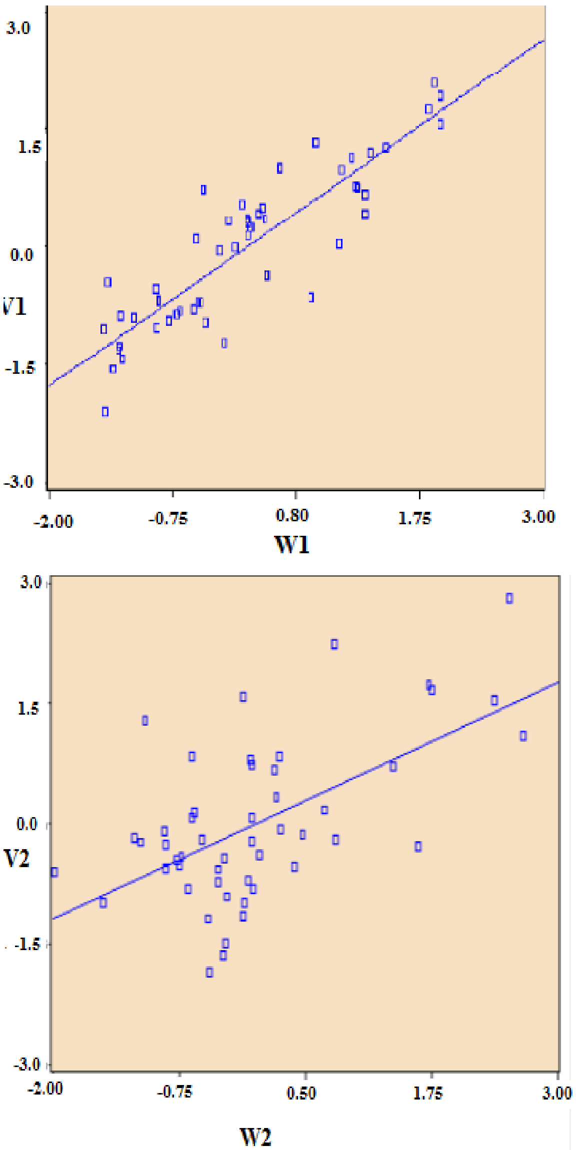


Fig.3: The scatterplots of the first and second pairs of canonical functions

The Millennium Development Goals (MDGs) are the most recent attempts at improving the standards of living. According to NISER [15], most of the indicators of the millennium development goals are not likely to be achieved by the target date of 2015 in Nigeria. The result of this study has further supported this assertion. The observed deficiency in the distribution of social infrastructure in the study area is counterproductive towards poverty alleviation. Given that there is a high level of poverty among the people in various parts of the world [16], poverty is seen as a multidimensional problem whose definition also emphasizes social dimensions. The problem of poverty becomes very worrisome when viewed from the perspective of basic social infrastructure.

This implies that both inequality and poverty are characteristics of the study area. Because, inequality presents a unique form of poverty [17] through mass deprivation, it is emphasized that the provision of basic social infrastructure would ensure that growth is consistent with poverty reduction. In other words, the poor can be identified as those who are unable to consume a basic quantity of social infrastructure, while the rich are those who have adequate access to basic social services for sustainable living. Historically, poverty has been concentrated in rural areas. The result of this study implies that poverty is still prevalent in the rural areas of Akwa Ibom State. It is common to associate development with the levels of access to social infrastructure. This study set out to achieve its aim of analyzing spatial patterns of social infrastructure in rural areas of Akwa Ibom State by collecting data on 21 social indicator variables and 6 spatial factors for 50 rural communities. The social indicator variables covered most aspects of social infrastructure such as water supply, education, health, road network and others (access to electricity, telephone, security, banks and markets). The analysis of the levels of access to social infrastructure revealed that access to most of them was either inadequate or lacking. The field data and analysis revealed the existence of inequalities among the communities. Eighteen (18) communities had negative scores on water supply while 13 communities performed negatively in the health sector. A total of 24 communities had unpaved road networks while 27 and 9 communities lacked access to telephone and electricity services respectively. Only 3 communities had no primary schools while a good proportion of the communities suffer from poor access to bank (76%) and security (50%) services.

CONCLUSION

Multivariate method of canonical correlation analysis was performed on spatial factors relating to rural settlements characteristics and variables of basic social infrastructure for the purpose of understanding the underlying dimensions of the relationship in the two sets of variables. The result showed that two out of the five canonical correlations were significant. The first canonical correlation was 88 with 78 percent overlapping variance while the second was 59 which represents 34 percent overlapping variance. The remaining three canonical correlations were effectively low. The first two pairs of canonical variates, therefore, accounted for the significant relationships between the two sets of variables. Total percent of variance and total redundancy indicate that the first pair of canonical variates was highly related while the second pair was only moderately related. With a cutoff correlation of 0.4, the variables in the infrastructure set that was correlated with the first canonical variate were water supply, health facilities, basic education and other facilities. Among the spatial factors variables, distance from the nearest urban centre, population of the nearest urban centre and level of nucleation of rural settlements correlated with the first canonical variate.

The first pair of canonical variates indicate that the level of water supply (0.516), access level to health facilities (0.554), access level to basic educational facilities (0.874), and other facilities (0.664) are negatively associated with distance of rural settlements to nearest urban centres (-.869) and levels of rural settlements nucleation (-0.945) but positively correlated with the population of the nearest urban centres (0.502). Increasingly large distances from the nearest urban centres and the highway reflected poorer access to basic infrastructural facilities. In the same vein, highly nucleated communities reflected better access to infrastructural facilities than highly dispersed communities. Rural settlement pattern is critical to the rural economy. This is because it has implication to the provision of basic social infrastructure in an area. In settlement studies, it is of interest to know the nature of the settlement pattern. This is because the pattern of distribution of facilities and households within and between settlements has a direct relationship with the existing settlement pattern. The second pair of canonical variates taken as a pair, suggest that a combination of levels of water supply (-0.490), access to health facilities (0.533), and the quality road network (0.582) is associated with more favorable population of rural settlements (0.777).

From the second canonical variate, it is revealed that the pattern of linkage is discernable. The loadings of all the settlements spatial factors set of variables were quite low with only the population of rural settlements (x1) presenting the highest loadings of 0.78. Similarly, canonical loadings of the infrastructure set of variables was equally low except water supply (y1), access to health facilities (y2) and quality of rural road network (y4). Therefore this canonical variate expressed the interrelationship between populations of rural settlements on the one hand and access levels to water supply, health facilities and quality of the rural road network on the other hand. Thus, a meaningful linkage between canonical loadings of the second variate for both variable sets is discernable.

The analysis implies that there is unequal concentration of stock of social infrastructure in the study area. On the whole, the overall performance of the communities in terms of social infrastructure stock is low however, the unequal concentration of the stock among the communities indicated that some communities are more vulnerable than others. The consequence is that many individuals and families in the most vulnerable communities may not attain minimum standards of living due to inadequate or lack of access to supportive social infrastructure. Poverty is seen as a multidimensional problem which involves social issues and as such, viewing it from the perspective of levels of access to social infrastructure, the prevalence of poverty in the study area becomes more glaring. In other words, both inequality and poverty are prevalent; manifesting through mass deprivation of access to basic social infrastructure. The relationships between the levels of social infrastructure as a dependent variable and a set of 6 independent variables were investigated using canonical correlation. The pattern that emerged again confirmed the existence of inequality in the study area. Social infrastructural facilities are the arteries and channels of rural development. There is a very close relationship between rural social infrastructure facilities and socio-economic development. The way facilities function, makes them inter-related. For instance, transportation and electricity power lines followed transportation routes while water and health have health education as their interaction. Location interaction node emphasizes that facilities be located where they may be accessible to the people at minimal cost in terms of distance and time. In this case, desirable sites are necessary for the facility location. In the case of territorial interactive node, the catchments area (the range) for a facility is considered. For instance, in planning for a group of villages, the territory may be identified by its socio-economic characteristics, geographical contiguity and other local needs. Since all the facilities interact at different nodes, they should be planned under the same planning framework known as regional planning. These results show that the rural settlements in Akwa Ibom State relates to social infrastructure in at least two major ways. The canonical correlation method therefore offers a better model of understanding this relationship.

ETICAL ISSUES

Ethical issues involved in scientific research were considered and observed during the conduct of the study. Proper permission was obtained from village heads and elders of the communities before the field work was embarked on. Participation in the study was not by force but on the willingness of respondents to participate. Anonymity of respondents was respected. During the field work all forms of identification including names, addresses and telephone numbers of respondents were avoided. This research report has not been published anywhere and it is in its original form.

CONFLIC OF INTEREST

Conflict of interest is not envisaged as the lead author and co-author contributed towards the conduct of the study in terms of human, materials and financial resources without external funding and support. The authors therefore declare that they have no competing interest relating to this work.

REFERENCES

[1] Atser J. Spatial patterns of social infrastructure in Nigeria. LAP LAMBERT Academic Publishing, Av Akademikerverlag GmbH & Co. KG, 2012

[2] World Bank. Infrastructure for Development: World Development Report. OUP, New York, 1994

[3] Adger W. N. New Indicators of Vulnerability and Adaptive Capacity. Centre for climate change research, Norwich, 2004

[4] Adger W. N. Social capital and climatic change. Centre for climate change research, Norwich, 2001

[5] Woolcock M, Narayan. D. Social Capital: Implication for Development Theory, Research and Policy. World Bank Research Observer, 2000*;* 15(2): 225-49

[6] UN. Indicators of Sustainable Development: Guidelines and Methodologies. United Nations, New York, 2001

[7] Lusting N. Economics with a Social Face. *Finance and Development*; a Quarterly Publication of the International Monetary Fund, 2005; 42(4): 4-7

[8] Maria R. Foundation for progress in Physical and Social Infrastructure. Proceedings of the World Economic Forum on Africa. Friendly Version Press, Cape Town, SA. 31 May-2 June, 2006

[9] Udofia E.P. Applied Statistics with Multivariate Methods. Immaculate Publications Ltd. Enugu, Nigeria, 2011

[10] Olafiaji E. Analysis of landuse and trip generation in Akure Metropolis, Nigeria. Unpublished MURP dissertation, Faculty of Environmental Studies, University of Uyo, 2014

[11] FGN. National Water Supply and Sanitation Policy", Federal Ministry of Water Resources, Department of Water Supply and Quality Control, First Edition. Abuja, 2000

[12] Mabogunje A.L. Cities and Social Order. Inaugural Lectures 1973-74 University of Ibadan, 1974

[13] Comrey A., Lee H.B. A first course in factor analysis (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates, 1992

[14] Tabachnick B.G, Fidell L.S. Using Multivariate Statistics (5th edition). New York, Pearson Education, Inc., 2007

[15] NISER. Millennium Development Goals: Nigeria's First Progress Report. NISER, Ibadan, 2003

[16] World Bank. World Development Report: Equity and Development. The World Bank, Washington D.C., 2005

[17] Sanusi, Y. A. Inequity in Environmental welfare in Nigerian urban centres: A case study of Minna. *NITP, 2007;* XX (1): 87-98.