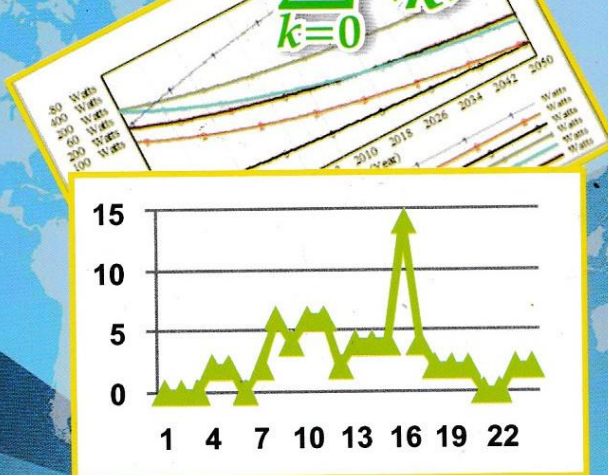


$$(x+a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k}$$



# MODERN QUANTITATIVE RESEARCH DESIGN & APPLICATIONS IN ENVIRONMENTAL STUDIES

Editor - DR. JACOB ATSER

Published 2016 by

**Immaculate Publications Limited**

4 Aku Street, Ogui New Layout, Enugu.

Cell Phone: 08052506848; 08037156007

© **Dr. Jacob Atser Editor**

e-mail: [jacobatser@uniuyo.edu.ng](mailto:jacobatser@uniuyo.edu.ng)

All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronics, mechanical, photocopying, recording or otherwise, without prior permission of the author.

ISBN: 978-036-095-6

Printed and Bound in the Federal Republic of Nigeria by  
IMMACULATE BUSINESS SUPPORT SERVICES

## CHAPTER EIGHT

### **Dimensionality in Canonical Correlation of Settlements Characteristics and Levels of Infrastructure in Rural Akwa Ibom State, Nigeria**

Jacob Atser and Uwem Ituen

#### **8.1 Abstract**

*This study examines the relationship between rural settlements characteristics and access levels to infrastructural facilities in Akwa Ibom State, Nigeria. 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 is .88 (.86 adjusted), representing 78% overlapping variance for the first pair of canonical functions or variates while the second canonical correlation coefficient is .59 (.50 adjusted) and thus, represents 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 useful tool for establishing empirically based linkages between two data sets.*

**Key words:** Rural settlements, characteristics, infrastructural facilities, canonical correlation.



## 8.2 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 (Atser, 2012). 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 (World Bank 1994, Adger 2004, Adger 2001, Woolcock and Narayan 2000). 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 to economic production. This reality becomes painfully evident when natural disaster or civil disturbances destroy 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 (UN 2001, World Bank 1994, Lusting 2005, Maria 2006).

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 infrastructural 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 a multivariate relationships and the generation of the maximum amount of correlation between the linear combinations of two groups of variables (Udofia, 2011).

## 8.3 Justification for the choice of canonical correlation

The goal of canonical correlation is to analyze the relationships between two sets of variables. It may be useful to think of one set of variables as independent variables and the other set as dependent variables. In any event, canonical correlation provides a statistical analysis for research in which each subject is measured on two sets of variables with a view to know if and how the two sets relate to each other. Canonical analysis might reveal that there are two statistically significant ways that the two sets of variables are related. In canonical correlation, sets of variables on each side are combined to produce, for each side, a predicted value that has the highest correlation with the predicted value on the other side. The combination of variables on each side can be thought of as a dimension that relates the variables on one side to the variables on the other. Although there are potentially as many ways to recombine the variables as there are variables in the smaller

set, usually only the first two or three combinations are statistically significant and need to be interpreted. This study has as its objective to investigate the dimensions in which the variables in one set are related to the variables in the other set.

#### 8.4 Data Requirements and Sources

The data requirements for this study were obtained primarily from 21 social infrastructure indicator variables as summarized in Table 1. 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 is presented in Table 2.

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 (FGN, 2000)
	(b)Distance	Time	30 minutes (FGN, 2000)
	(c)Borehole	Number / community	1/250 population (FGN, 2000)
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 (Mabogunje, 1974)
	(b)Secondary	Number	1/12000population (Mabogunje, 1974)
	(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: Input 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 1: 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)

### 8.5 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 .30 (9% of variance) are usually interpreted as part of the variate, and variables with loadings of .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 (1992) suggest that loadings in excess of .71 (50% overlapping variance) are considered excellent, .63 (40% overlapping variance) very good, .55 (30% overlapping variance) good, .45 (20% overlapping variance) fair, and .32 (10% overlapping variance) poor. Choice of the cutoff for size of loading to be interpreted is a matter of preference (Tabachnick and Fidell, 2007). Most researchers do not interpret pairs with a canonical correlation lower than .30 because  $r_c$  values of .30 or lower represent less than a 10% overlap in variance. In this study, loadings in excess of .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. Significance of the relationships between the sets of variables is reported directly by SAS CANCORR, as shown in Tables 4 and 5 respectively. With all four canonical correlations included,  $F(30, 158) = 3.94$ ,  $P < .001$ . 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 ( $r_c$ ) and eigenvalues ( $r^2_c$ ) are also in Table 4. The first canonical correlation is .88 (.86 adjusted), representing 78% overlapping variance for the first pair of canonical variates. The second canonical correlation is .59 (.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 3: Canonical correlation Analysis

	Canonical correlation	Adjusted canonical correlation	Approx. Std. Error	Squared canonical correlation
<b>1</b>	<b>0.882890</b>	<b>0.860574</b>	<b>0.031501</b>	<b>0.779495</b>
<b>2</b>	<b>0.586835</b>	<b>0.496140</b>	<b>0.093661</b>	<b>0.344375</b>
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 $\text{Inv}(E) * H = \text{CanRsq} / (1 - \text{CanRsq})$					Test of $H_0$ : the canonical correlations in the current row and all that follow are zero				
	Eigenvalue	Difference	Proportion	Cumulative	Likelihood Ratio	Approx. 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.

Canonical correlation actually measures the correlation between pairs of canonical variates in the two data sets. 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 the selected

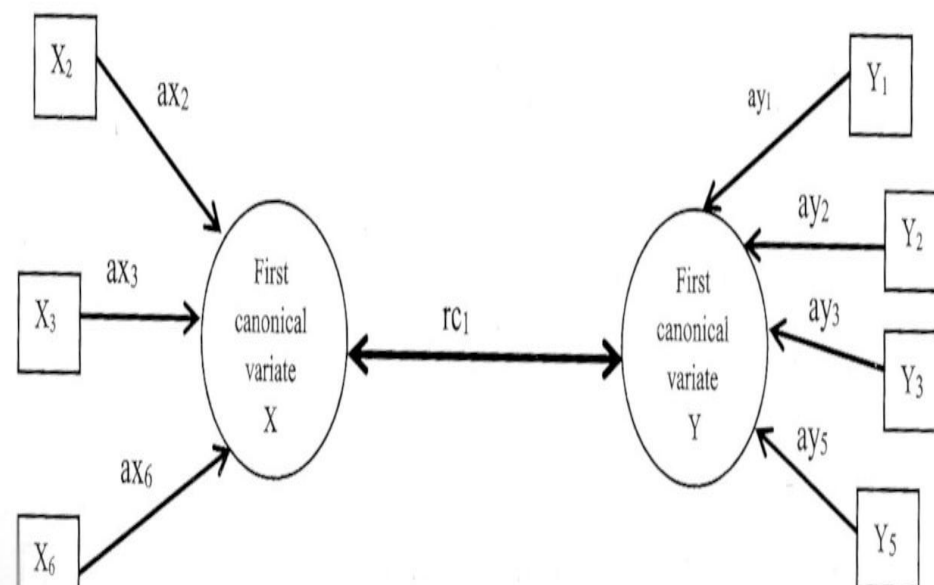
canonical functions were significant on Wilk's Lambda, Pillars Trace, Hotelling Lawley Trace and Roy's Greatest Root Statistics (Table 5) at probability level of  $\geq 0.001$ .

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

Loading matrices between canonical variates and original variables are in the Table on 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 are considered when interpreting the canonical variates. The first pair of canonical variates has high loadings on Y1(.516), Y2(.554), Y3(.874), and Y5(.664) respectively on the infrastructure data set and on X2(-.869), X3(.502) and X6(-.945) on the spatial factors side. Thus, access to water supply, health facilities, basic education and other basic facilities are related to distance from nearest urban centre, population of nearest urban centre and level of nucleation of rural communities. The second pair of canonical variates has high loadings on Y1(-.490), Y2(.533), and Y4(.582) on the

infrastructure data set and X1(.777) on the spatial factors data set. This implies that access to water supply, health facilities, and the quality of road network are 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 (Figures 1 and 2).



- X2 Variable in X data set
- y1 Variable in y data set
- ax2 Loading of correlation with ith x variable on canonical variate x
- ay1 Loading of correlation with ith y variable on canonical variable y
- rc1 Canonical correlation for the first pair of canonical variates

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



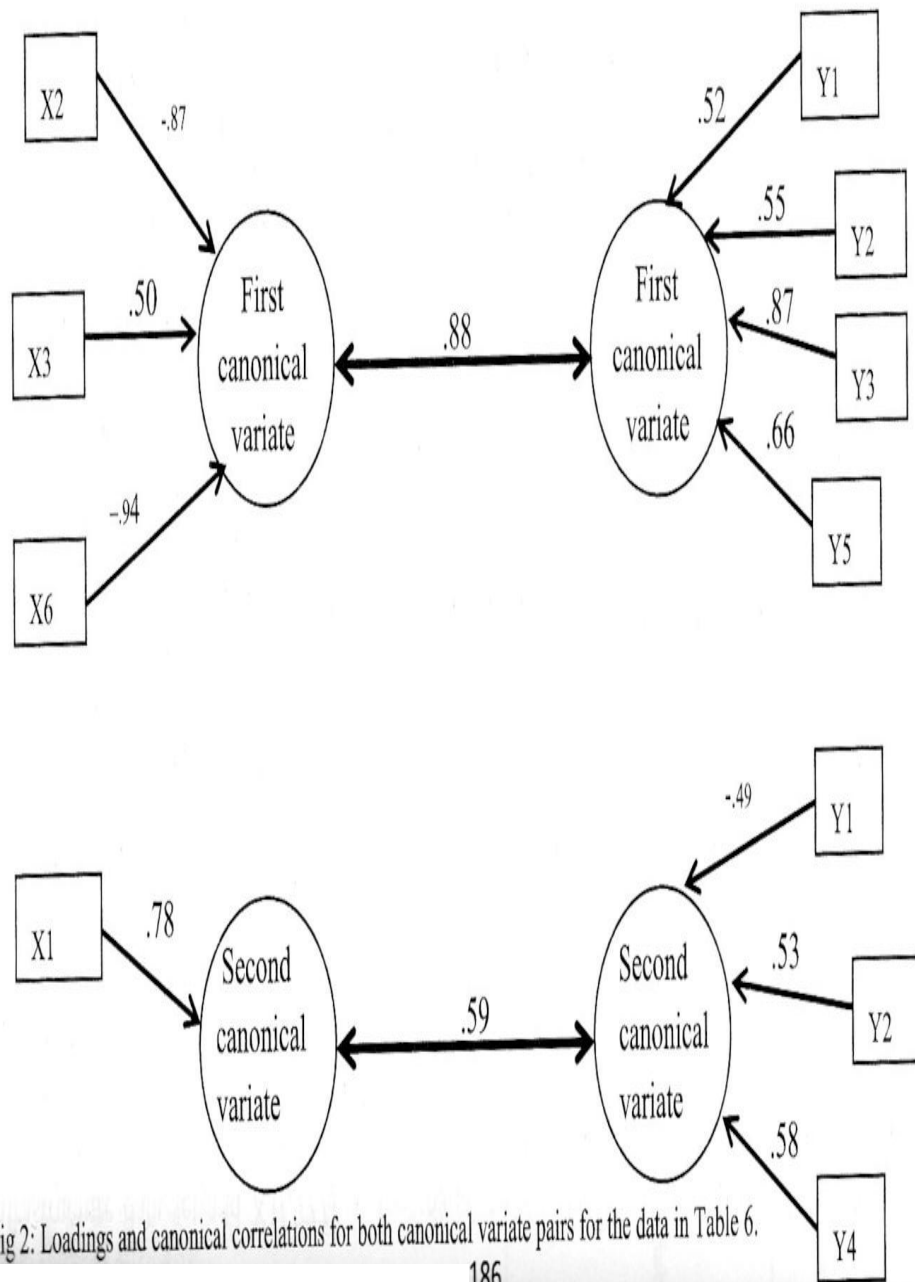


Fig 2: Loadings and canonical correlations for both canonical variate pairs for the data in Table 6.

These pairs of linear combinations for both the X and Y variables are known as canonical variates and are similar to components in principal components analysis. The canonical variates are extracted to help account for the maximum amount of correlation between the two sets of data while the first canonical variates 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 is 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 unaccounted 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 (Udofia, 2011).

Canonical function consists of pairs of variate; one as dependent variable and the other as 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 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 .356 for the first set of variables and .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 .173 for the first set of variables and .117 for the second set; the second canonical variate pair extracts 17% of 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 variance (36% plus 17%) in the infrastructure set of data, and 46% of variance (34% and 12%) in the spatial factors set.

Table 7: Canonical Redundancy Analysis

Standardized Variance of the Stock of infrastructure(Y variables) Explained by					
Canonical variable number	Their own canonical variables		Canonical R-square	The opposite canonical variables	
	Proportion	Cumulative proportion		Proportion	Cumulative proportion
1	<b>0.3558</b>	<b>0.3558</b>	<b>0.7795</b>	<b>0.2773</b>	<b>0.2773</b>
2	<b>0.1734</b>	<b>0.5291</b>	<b>0.3444</b>	<b>0.0597</b>	<b>0.3370</b>
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	<b>0.3365</b>	<b>0.3365</b>	<b>0.7795</b>	<b>0.2623</b>	<b>0.2623</b>
2	<b>0.1169</b>	<b>0.4533</b>	<b>0.3444</b>	<b>0.0402</b>	<b>0.3025</b>
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

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, the two infrastructure variates explain 34% of the variance in the spatial factors set of variables. The first settlement spatial factors variate accounts for .262 (26%) and the second .040 (4%) of the variance in the infrastructure data set. Together, the 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. The standardized canonical coefficients are derived from Table 9.

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

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 (Figure 3). The shape of the scatter plots reflect 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 second plot, there are widely scattered cases with extremely high scores indicating positive skewness thereby confirming departures from normality.

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 display graphically in Figure 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

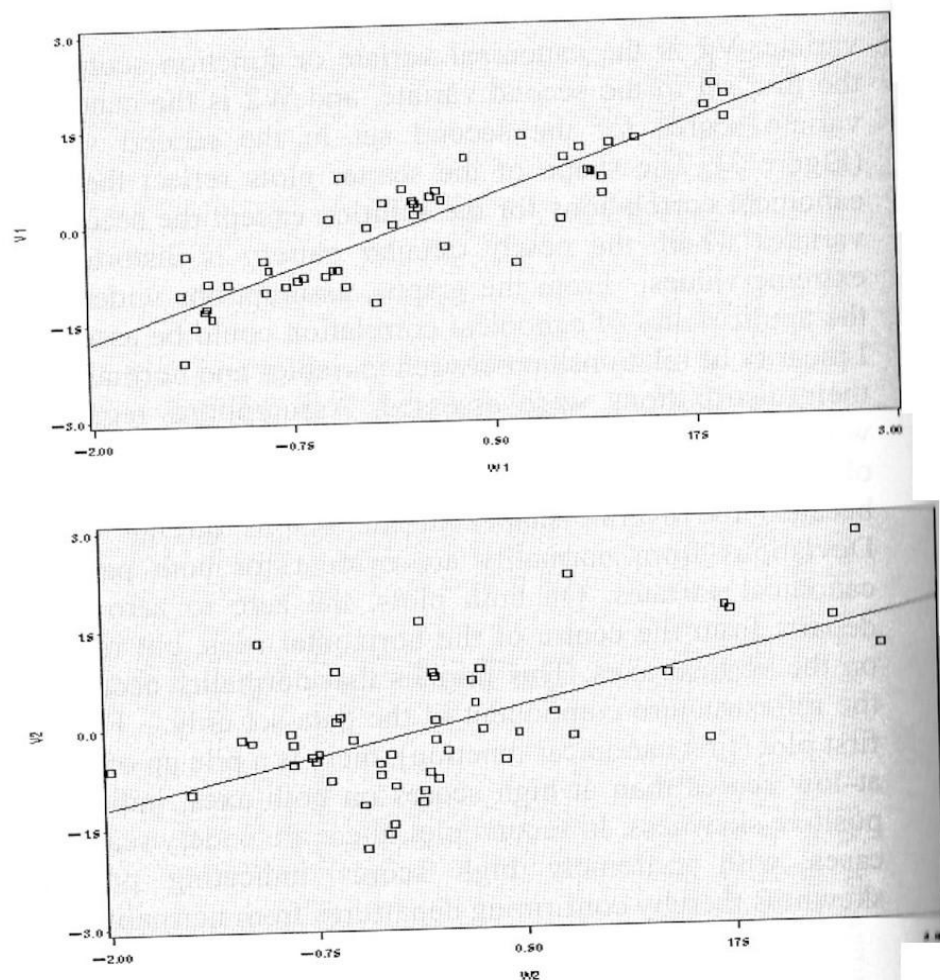


Figure 3: The scatterplots of the first and second pairs of canonical functions

### 8.6 Conclusion and Recommendations

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 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 were correlated with the first canonical variate were water supply, health facilities, basic education and other facilities. Among the spatial factors variables, distance from nearest urban centre, population of 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 (.516), access level to health facilities (.554), access level to basic educational facilities (.874), and other facilities (.664) are negatively associated with distance of rural settlements to nearest urban centres (-.869) and levels of rural settlements nucleation (-.945) but positively correlated with the population of nearest urban centres (.502). Increasingly large distances from 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 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 (-.490), access to health facilities (.533), and the quality road network (.582) is associated with more favorable population of rural settlements (.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, the canonical loadings of the infrastructure set of variables were 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 rural road network on the other hand. Thus, a meaningful linkage between canonical loadings of the second variate for both variable sets is discernable.

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, make 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 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.

## References

- Adger, W. N (2004) "New Indicators of Vulnerability and Adaptive Capacity". Centre for climate change research, Norwich, U.K.
- Adger, W. N (2001) "Social capital and climatic change". *Centre for climate change research*, Norwich, U.K
- Atser J (2012). Spatial patterns of social infrastructure in Nigeria. LAP LAMBERT Academic Publishing, Av Akademikerverlag GmbH & Co. KG
- Comrey A.L. and Lee H.B (1992). A first course in factor analysis (2<sup>nd</sup> edition). Hillsdale, NJ: Lawrence Erlbaum Associates
- FGN (2000) "National Water Supply and Sanitation Policy", Federal Ministry of Water Resources, Department of Water Supply and Quality Control, First Edition. Abuja.
- Lusting, N. (2005) "Economics with a Social Face". *Finance and Development*, A Quarterly Publication of the International Monetary Fund, Vol. 42(4) pp 4-7.
- Mabogunje A.L.(1974) *Cities and Social Order*. Inaugural Lectures 1973-74 University of Ibadan.

- Maria, R. (2006) "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
- Tabachnick B.G. and Fidell L.S. (2007) Using Multivariate Statistics (5<sup>th</sup> edition). New York, Pearson Education, Inc.
- Udofia, E.P. (2011) *Applied Statistics with Multivariate Methods*. Immaculate Publications Ltd. Enugu, Nigeria.
- UN (2001) *Indicators of Sustainable Development: Guidelines and Methodologies*. United Nations, New York.
- Woolcock, M. and Narayan, D. (2000) "Social Capital: Implication for Development Theory, Research and Policy". *World Bank Research Observer* 15 (2) pp. 225-249.
- World Bank (1994) *Infrastructure for Development: World Development Report*. OUP, New York.