COMPARISON OF CANONICAL AND GENERALIZED CANONICAL CORRELATION ANALYSIS USING SOME CONTINUOUS DISTRIBUTIONS

Abstract: This study is designed to compare canonical and generalized canonical correlation analysis for two data sets using five continuous distributions namely; Beta, Exponential, Gamma, Weibull and Normal distributions as control. Simulation studies for samples of sizes n = 10, 20, 30, 40, and 50 replicated 10,000 each were analyzed using R-programming language. Relative efficiencies of the two methods (CCA and GCCA) calculated for each of the distributions under consideration showed no significant differences in the two methods.

==========

Keywords: Canonical Correlation Analysis, Generalized Canonical Correlation Analysis, Relative Efficiencies, Simulation Experiment.

1.0 INTRODUCTION

The study of simple correlation (ρ) between two univariate random variables X and Y is of immense importance in real life circumstances because most variables in practice show some kind of relationships. For example, there is a relationship between price of items and their supply, income and expenditures etc. We can equally estimate the value of one variable given the value of another by the help of regression analysis. The $Cov(X,Y)$ between the two variables X and Y normalized by the geometric mean of the variances Var (X and Y) is given by ,

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \tag{1.1}$$

which can also be written as

$$\rho = \frac{\left\langle \left[X - E(X)\right]\left[Y - E(Y)\right]\right\rangle}{\sqrt{\left[\left(X - E[X]^2\right)\left(Y - E[Y]^2\right)\right]}} \tag{1.2}$$

where E is the expected values of the random variables. The maximum likelihood of ρ is obtained by replacing Cov and Var in Equation (1.1) by their maximum likelihood estimators. The statistic obtained is called the *Pearson's correlation coefficient.*

Canonical Correlation analysis as developed by Hotelling (1936) is a multivariate statistical technique which is concerned with the maximization of the correlation between two linear functions of two sets of

random variables. It is employed in testing for dependence among two sets of variables. It is especially useful in data reduction and could be applied in instances where a researcher may be interested in the level of association between sets of variables.

Canonical correlation analysis forms linear composites, that is, canonical variates $U = X\alpha$ and $V = Y\beta$ from each set, then develop a function that maximizes the canonical correlation coefficient ρ between the two canonical variates. These canonical variates U and V are interpreted as canonical loading, which is the correlation between the individual variables and their respective variates.

According to Onyeagu et al. (2014), "it is important to note that as the Canonical Correlation decreases in size, the relationship between the corresponding canonical variates becomes weaker and the consequent predictions become less accurate. More so, although, the technique is of some interest in the study of relationships between two sets of variables, and may even provide some useful predictive models, it can be seen that their scope is very limited. This is because they only predict linear combinations of the X_i and Y_i , and furthermore the linear combinations that they predict are determined by the data and are not under the control of the investigator". Based on the above demerits of the Canonical Correlation, Kettenring (1971) "developed and compared extensions of Canonical Correlation to three or more sets of variates, and has given iterative schemes for the computation of the correlations and coefficient that is user friendly". However, it was Van de Velden and Biljmolt (2006) that explained the work by Carroll and Chang (1970) in which they introduced the Generalized Canonical Correlation Analysis (GCCA) which allows for several sets of variables to be analyzed simultaneously. This makes the method suited for the analysis of various types of data especially in situations where subjects may be asked to rate a set of objects on a set of attributes. In this case, for each individual, a data matrix can be constructed where objects are represented row-wise and attributes column-wise. Then using Generalized Canonical Correlation Analysis, a graphical representation, sometimes referred to as a perceptual map can be made on the basis of the individuals' observation matrices. The advantage of the Caroll and Chang's approach to Generalized Canonical Correlation Analysis is that it has some attractive properties that make the method well fitted for the analysis of multiple sets of data. This is because computationally, the method is

straight forward and its solution is based on an eigen- equation and the method is closely related to several well known multivariate techniques such as principal component analysis and multiple correspondence analysis.

In this paper, Canonical and Generalized canonical correlation analysis are compared using five continuous distributions namely Beta, Exponential, Gamma, Weibull and Normal distributions for different same sizes replicated 10,000 each. The relative efficiencies of the two methods; CCA and GCCA were analyzed using their variances and standard deviations obtained from the distribution using R-programming language.

2.0 LITERATURE REVIEW

In the last two decades, researchers have shown keen interest on the subjects Canonical Correlation Analysis and Generalized Canonical Correlation Analysis due to the availability of computer programs which has facilitated quick computation of CCA and GCCA. This was not so before.

According to Cramer (1973), the "complicated manner" in which canonical correlation equations are derived in standard texts like Anderson (1958) and others contributed to the lack of understanding of the methods as expressed by practitioners. He proposed a simple approach of calculating CCA base on simple derivation which follows directly from the relation between multiple regression analysis and multiple correlations. Although, this simple approach did not yield the exact canonical variates as obtained by Hotelling (1936) who originated the method but it provided a simple derivation approach that suggested that CCA could be cast in a regression model.

Muller (1982) presented CCA as a multivariate multiple regression in which the least square approach was employed in finding the estimates β , α , and $D(\rho\kappa)$ in the model equation:

$$Y\beta = X\alpha D(\rho\kappa) + E; \qquad (2.1)$$

where β is a $q \times d$ matrix, with the kth column being the canonical weights for the set for the kth canonical variate pair. $D(\rho\kappa)$ is a $d \times d$ diagonal matrix of canonical correlation. α is $p \times d$ matrix, with the kth column being the canonical weights for the X set for the kth canonical variate pair. The matrices β , α , and $D(\rho\kappa)$ must correspond in the sense that the kth column of β and α provide the linear combinations that are correlated $\rho\kappa$, which is the (k,k) element of $D(\rho\kappa)$. E is an $n \times d$ matrix of errors. This multivariate formulation by Muller was seen as wonderful novel but if not handled very well could lead to greater complications. For instance, the equivalence of β , α , and $D(\rho\kappa)$ in the standard statement of CCA are vectors (not matrices).

Pedhazur (1997) stated the following properties of canonical functions "the first canonical function identifies linear combinations of the study's variables that yield the largest squared correlation R^2c possible. The second canonical function identifies linear combinations of the study's variables that are not correlated with the first pair of canonical variates and yield the second largest R^2c possible, given the residual variance left over from the first function, the same is true for subsequent canonical functions, such that the m^{th} canonical function identifies linear combinations of the study's variables that are not correlated with prior pairs of canonical variates and yields the m^{th} largest possible. Takane and Hwang (2002) proposed "a method for Generalized Constrained Canonical Correlation Analysis (GCCANO) which incorporates external information on both rows and columns of data matrices. They demonstrated the method with two illustrations and observed that in Canonical Correlation Analysis, the interpretation of the canonical variates obtained from GCCANO can be difficult and a simple rotation of either the canonical pattern or the structure matrix (by varimax) may be used to make it easier to interpret".

Ebenezer and Iyaniwura (2012) observed that Canonical Correlation Analysis is able to analyze two sets of data simultaneously to see if there are strong and meaningful links between the data. This was noted in their work in which they compared the correlation between poverty level and literacy level in a certain group. The aim of their paper was to investigate whether poverty level and literacy level are related to one another.

3.0 METHODOLOGY

3.1 CANONICAL CORRELATION ANALYSIS (CCA)

If we have two vectors $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_m)$ of random variables, and there are correlations among the variables, then canonical correlation analysis will find linear combination of the X_i and Y_i which have maximum correlation with each other. Given two column vectors $X = (x_1, ..., x_n)'$ and $Y = (y_1, ..., y_m)'$ of random variables with finite second moments, one may define the cross-covariance $\sum XY = \text{cov}(X, Y)$ to be the $n \times m$ matrix whose (i, j) entry is the covariance $\text{cov}(x_i, y_i)$.In practice, we would estimate the covariance matrix based on sampled data from X and Y (i.e. from a pair of data matrices).Canonical correlation analysis seeks vectors a and b such that the random variables

$a'X$ and $b'Y$ maximize the correlation $\rho = corr(a'X, b'Y)$. The random variables $U = a'X$ and $V = b'Y$ are the first pair of canonical variables. Then one seeks vectors maximizing the same correlation subject to the constraint that they are to be uncorrelated with the first pair of canonical variables; this gives the second pair of canonical variables. This procedure may be continued up to $\min(m, n)$ times.

Let $Y' = (Y_1, Y_2, ..., Y_p)$ and $X' = \left(X_1, X_2, ..., X_q\right)$

We want to find two linear functions, $U = a'X$ and $V = b'Y$ of unit variance such that the correlation between U and V is a maximum. These functions U and V are called canonical variates. They are of unit variance that is,

$a'\sum_{11} a = 1$ and $b'\sum_{22} b = 1$

The covariance between them is given as

$a'\sum_{12} b = b'\sum_{21} a$

Therefore, the correlation between U and V is,

$$\rho(U,V) = \frac{Cov(U,V)}{\sqrt{Var(U), Var(V)}}$$

$$\therefore \rho(U,V) = \frac{a'\sum_{12} b}{\sqrt{a'\sum_{11} a}\sqrt{b'\sum_{22} b}} \qquad (3.1)$$

Onyeagu (2003) observed that, the correlation will attain a maximum value when the denominator equals one. That is,

$$\sqrt{a'\sum_{11} a} = 1$$
$$\sqrt{b'\sum_{22} b} = 1$$
$$\therefore a'\sum_{11} a = 1$$
$$b'\sum_{22} b = 1$$

Thus, to obtain the maximum value of the correlation between U and V, we determine the values of a and b. Hence, we maximize

$$F(U,V) = \max_{ab} a'\sum_{12} b$$

Subject to the constraints that,

$a'\sum_{11} a = b'\sum_{22} b = 1$

where

$Var(U) = a'\sum_{11} a$

$Var(V) = b'\sum_{22} b$

Using the Lagrange's multipliers, the function

$$F = a'\sum_{12} b - \frac{\lambda_1}{2}\left(a'\sum_{11} a - 1\right) - \frac{\lambda_2}{2}\left(b'\sum_{22} b - 1\right)$$

is to be maximized.

$$\frac{\partial F}{\partial a'} = \sum_{12} b - \lambda_1 \sum_{11} a = 0 \qquad \ldots (3.2)$$

$$\frac{\partial F}{\partial b'} = \sum_{12} a - \lambda_2 \sum_{22} b = 0 \qquad \ldots (3.3)$$

$$\frac{\partial F}{\partial \lambda_1} = a'\sum_{11} a - 1 \ = 0 \qquad \ldots (3.4)$$

$$\frac{\partial F}{\partial \lambda_2} = b'\sum_{22} b - 1 \ = 0 \qquad \ldots (3.5)$$

Multiplying (3.2) by a' and (3.3) by b'

$$a'\sum_{12} b - \lambda_1 a'\sum_{11} a = 0$$

$$b'\sum_{21} a - \lambda_2 b'\sum_{22} b = 0$$

$$a'\left(\sum_{12} b - \lambda_1 \sum_{11} a\right) = 0$$

$$b'\left(\sum_{12} a - \lambda_2 \sum_{22} b\right) = 0$$

$$-\lambda \sum_{11} a + \sum_{12} b = 0 \qquad \ldots (3.6)$$

$$\sum_{21} a - \lambda \sum_{22} b = 0 \qquad \ldots (3.7)$$

Where $\lambda = \lambda_1 = \lambda_2 = a'\sum_{12} b$

Multiplying Equation (3.6) by λ and Equation (3.7) by $\sum_{12}\sum_{22}^{-1}$, we have,

$$-\lambda^2 \sum_{11} a + \lambda \sum_{12} b = 0 \qquad \ldots (3.8)$$

$$\sum_{12}\sum_{22}^{-1} \sum_{21} a - \lambda \sum_{12}\sum_{22}^{-1} \sum_{22} b = 0 \qquad \ldots (3.9)$$

Adding Equation (3.8) and Equation (3.9) yields the equation

$$\left[\sum_{12}\sum_{22}^{-1}\sum_{21} - \lambda^2 \sum_{11}\right] a = 0$$

$\lambda_1^2, \lambda_2^2, ..., \lambda_p^2$ and $a_1, a_2, ..., a_p$

are the roots and vectors respectively of the characteristic equation

$$\left|\sum_{12}\sum_{22}^{-1} \sum_{21} - \lambda^2 \sum_{11}\right| = 0 \qquad \ldots (3.10)$$

Let $A = \left[a_1, a_2, ..., a_p\right]$ then

$A'\sum_{11} A = I_p$ and

$$A'\sum_{12}\sum_{22}^{-1} A = \Lambda_1 \qquad \ldots (3.11)$$

Where Λ_1 is a diagonal matrix with roots $\lambda_1^2, \lambda_2^2, ..., \lambda_p^2$.

Similarly by multiplying Equation (3.7) by λ and Equation (3.6) by $\sum_{21}\sum_{11}^{-1}$ and adding ,we have,

$$\sum_{21}\sum_{11}^{-1}\sum_{12} b - \lambda^2 \sum_{22} b = 0 \qquad \ldots (3.12)$$

Let $B' = \left[b_1, b_2, ..., b_q \right]$ then

$B' \sum_{22} B = I_q$ and

$$B' \sum_{21}\sum_{11}^{-1}\sum_{12} B = \Lambda_2 \qquad \ldots (3.13)$$

where Λ_2 is a diagonal matrix.

The nonzero positive square roots of λ_i^2 are called the canonical correlations between the canonical variates $U_i = a_i^1 Y$ and $V_i = b_i X$ for $i = 1, 2, ..., p \le q$.

From equation (3.9)

$$\sum_{12}\sum_{22}^{-1} \sum_{21} a - \lambda \sum_{12}\sum_{22}^{-1} \sum_{22} b = 0 \qquad \ldots (3.14)$$

The relationship between a_i and b_i is given by

$$b_i = \frac{\sum_{22}^{-1} \sum_{21} a_i}{\lambda_i} \qquad \ldots (3.15)$$

U_i and V_i are clearly uncorrelated and have

$$Cov\left(U_i, U_j\right) = Cov\left(V_i, V_j\right) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad ...(3.16)$$

Furthermore the covariance between V_i and U_i is λ_i for $i = 1, 2, ..., p$ and 0 otherwise.

$$Cov\left(U_i, V_i\right) = \begin{cases} \lambda_i, & i = 1, 2, ..., p \\ 0 & i \neq j \end{cases} \qquad ...(3.17)$$

Thus, if U_1 and U_2, V_1,V_2 and V_3 are the canonical variates. The correlation matrix for U_1, U_2, V_1, V_2 and V_3 has the form

$$\begin{array}{c} \\ U_1 \\ U_2 \\ V_1 \\ V_2 \\ V_3 \end{array} \begin{array}{c} \begin{array}{ccccc} U_1 & U_2 & V_1 & V_2 & V_3 \end{array} \\ \begin{pmatrix} 1 & 0 & \lambda_1 & 0 & 0 \\ 0 & 1 & 0 & \lambda_2 & 0 \\ \lambda_1 & 0 & 1 & 0 & 0 \\ 0 & \lambda_2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

This can be transformed to the correlation matrix

$$R_{U,V} = \begin{bmatrix} 1_p & \Delta \\ \Delta' & 1_q \end{bmatrix}$$

Where Δ is a $p \times q$ matrix containing the first canonical correlations between U_i and V_i.

$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$ implies that

$$\left| S_{12} S_{22}^{-1} S_{21} - \lambda^2 S_{11} \right| = 0 \qquad \ldots (3.18)$$

$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$ implies that

$$\left| R_{12} R_{22}^{-1} R_{21} - \lambda^2 R_{11} \right| = 0 \qquad \ldots (3.19)$$

A test of the significance of the canonical correlations is provided by

$$\Lambda = \prod_{i=1}^{s} \left(1 - r_i^2\right)$$

$$= \frac{\left| S_{11} - S_{12} S_{22}^{-1} S_{11} \right|}{\left| S_{11} \right|} \quad , \quad \left| S_{11} \right| \neq 0$$

Where r_i^2 are the sample estimates of λ_i^2, that is

$$H_O : \sum\nolimits_{12} = 0 \quad \text{VS} \quad H_1 : \sum\nolimits_{12} \neq 0$$

and the Bartletts' procedure is employed for the test of significance.

3.2 GENERALIZED CANONICAL CORRELATION ANALYSIS (GCCA)

In generalized canonical correlation analysis, linear combinations are obtained in such a way that the sum of squared correlations of the linear combinations of the variables with group configuration is a maximum. Let Y denote the unknown group configuration. The order of Y is $m \times k$, where m is the number of rows for each observation matrix X_i (i.e., the ith data set) and k is the dimensionality of the solution. The data matrices X_i are first centered. Sometimes, if the variables are for example measured on different scales, they are also standardized. Note that the sizes of the observation matrices X_i are $m \times p_i$ for $i = 1, 2, ..., n$.

One way of expressing the GCCA objective is as follows:

$$\text{Min } \phi(Y, A_i) = \text{min trace } \sum_{i=1}^{n} (Y - X_i A_i)' (Y - X_i A_i) \quad \ldots (3.20)$$

Subject to: $Y'Y = I_k$ … (3.21)

It is known, (Carroll (1968)), that for observed X_i matrices, the group configuration Y can be obtained from the eigen-equation

$$\left(\sum_{i=1}^{n} X_i \left(X_i' X_i \right)^{-1} X_i' \right) Y = Y\Lambda \quad \ldots (3.22)$$

Where Λ is a diagonal matrix with diagonal elements λ_j, being the k largest eigen-values of $\sum_{i=1}^{n} X_i \left(X_i' X_i \right)^{-1} X_i'$ (where we assumed that the $X_{i's}'$ are of full column rank) and the matrices Λ_i can be calculated as: $\Lambda_i = \left(X_i' X_i \right)^{-1} X_i' Y$ … (3.23)

4.0 COMPUTATIONAL ALGORITHMS AND RESULTS

The Computational algorithms for CCA and GCCA functions in R-2.13.0 programming language were propose by Becker et al (1988); and Lahti and Huovilainer (2013) as **Cancor (x,y, xcenter = TRUE, ycenter = TRUE) and regCCA(datasets,reg=0)** respectively.

So, in order to compare the performance of CCA and GCCA, we imputed data on R-command window, calling for the CCA and GCCA function using Beta, Exponential, Gamma, Weibull and Normal distributions as control. Simulation studies for samples of sizes n = 10, 20, 30, 40, and 50 replicated 10,000 generated the following results:

Table 1: Summary of results from the Analysis

		Correlation	Eigenvalue	X mean vector		Y mean vector	
Distribution	**Sample**	**CCA**	**GCCA**	**CCA**	**GCCA**	**CCA**	**GCCA**
Beta	**10**	**1.00**	**2.00**	**0.39**	**0.39**	**0.44**	**0.44**
		0.98	**1.98**	**0.73**	**0.73**	**0.40**	**0.40**
		0.53	**1.53**	**0.59**	**0.59**	**0.59**	**0.59**
		0.44	**1.44**	**0.39**	**0.39**	**0.51**	**0.51**
		0.11	**1.11**	**0.41**	**0.41**	**0.40**	**0.40**
		Var=0.14	**Var=0.14**				
		SD=0.37	**SD=0.37**				
	20	**0.73**	**1.73**	**0.55**	**0.55**	**0.44**	**0.44**
		0.50	**1.50**	**0.59**	**0.59**	**0.43**	**0.43**
		0.29	**1.29**	**0.54**	**0.54**	**0.38**	**0.38**
		0.18	**1.18**	**0.55**	**0.55**	**0.48**	**0.48**
		0.00	**1.00**	**0.47**	**0.47**	**0.53**	**0.53**

		Var=0.08	Var=0.08				
		SD=0.28	SD=0.28				
	30	0.58	1.58	0.58	0.58	0.46	0.46
		0.40	1.40	0.44	0.44	0.56	0.56
		0.32	1.32	0.49	0.49	0.57	0.57
		0.04	1.04	0.51	0.51	0.52	0.52
		0.01	1.01	0.53	0.53	0.51	0.51
		Var=0.05	Var=0.05				
		SD=0.24	SD=0.24				
	40	0.55	1.55	0.55	0.55	0.53	0.53
		0.48	1.48	0.45	0.45	0.48	0.48
		0.36	1.36	0.58	0.58	0.43	0.43
		0.15	1.15	0.47	0.47	0.53	0.53
		0.08	1.08	0.50	0.50	0.49	0.49
		Var=0.04	Var=0.04				
		SD=0.20	SD=0.20				
	50	0.55	1.55	0.45	0.45	0.52	0.52
		0.41	1.41	0.45	0.45	0.48	0.48
		0.34	1.34	0.47	0.47	0.50	0.50
		0.21	1.21	0.55	0.55	0.42	0.42
		0.11	1.11	0.47	0.47	0.43	0.43
		Var=0.29	Var=0.29				
		SD=0.17	SD=0.17				
		Correlation	Eigenvalue	X mean vector		Y mean vector	
Distribution	Sample	CCA	GCCA	CCA	GCCA	CCA	GCCA
Exponential	10	1.00	2.00	1.07	1.07	0.81	0.81
		0.92	1.92	1.24	1.24	1.25	1.25
		0.59	1.59	1.00	1.00	1.12	1.12
		0.51	1.51	1.28	1.28	0.81	0.81
		0.07	1.07	1.03	1.03	0.65	0.65
		Var=0.14	Var=0.14				
		SD=0.37	SD=0.37				
	20	0.77	1.77	1.03	1.03	0.72	0.72
		0.63	1.63	0.91	0.91	0.18	0.18
		0.39	1.39	0.82	0.82	0.92	0.92
		0.11	1.11	0.85	0.85	0.69	0.69
		0.03	1.03	0.61	0.61	1.00	1.00
		Var=0.10	Var=0.10				
		SD=0.32	SD=0.32				
	30	0.61	1.61	0.74	0.74	1.67	1.67
		0.52	1.52	0.81	0.81	0.82	0.82
		0.44	1.44	1.00	1.00	0.67	0.67
		0.41	1.41	1.17	1.17	0.91	0.91
		0.06	1.06	0.73	0.73	1.19	1.19
		Var=0.04	Var=0.04				
		SD=0.21	SD=0.21				

	40	0.77	1.77	0.97	0.97	0.96	0.96
		0.43	1.43	0.95	0.95	1.13	1.13
		0.30	1.30	1.02	1.02	0.92	0.92
		0.16	1.16	0.89	0.89	0.80	0.80
		0.15	1.15	0.92	0.92	1.21	1.21
		Var=0.07	Var=0.07				
		SD=0.26	SD=0.26				
	50	0.38	1.38	1.00	1.00	0.70	0.70
		0.19	1.19	0.98	0.98	0.91	0.91
		0.13	1.13	1.22	1.22	1.08	1.08
		0.05	1.05	0.86	0.86	1.03	1.03
		0.03	1.03	0.84	0.84	1.26	1.26
		Var=0.02	Var=0.02				
		SD=0.14	SD=0.14				

		Correlation	Eigenvalue	X mean vector		Y mean vector	
Distribution	Sample	CCA	GCCA	CCA	GCCA	CCA	GCCA
Gamma	10	1.00	2.00	0.85	0.85	1.77	1.77
		0.69	1.69	1.40	1.40	1.72	1.72
		0.51	1.51	1.49	1.49	0.75	0.75
		0.23	1.23	1.80	1.80	1.10	1.10
		0.17	1.17	1.06	1.06	0.83	0.83
		Var=0.12	Var=0.12				
		SD=0.34	SD=0.34				
	20	0.83	1.83	0.76	0.76	1.15	1.15
		0.64	1.64	0.81	0.81	0.75	0.75
		0.49	1.49	0.88	0.88	0.79	0.79
		0.29	1.29	0.97	0.97	0.94	0.94
		0.07	1.07	0.95	0.95	0.84	0.84
		Var=0.09	Var=0.09				
		SD=0.29	SD=0.29				
	30	0.75	1.75	1.04	1.04	1.17	1.17
		0.43	1.43	1.02	1.02	0.95	0.95
		0.23	1.23	1.39	1.39	1.28	1.28
		0.05	1.05	0.95	0.95	1.04	1.04
		0.03	1.03	0.91	0.91	1.02	1.02
		Var=0.09	Var=0.09				
		SD=0.29	SD=0.29				
	40	0.40	1.40	1.08	1.08	1.24	1.24
		0.35	1.35	0.97	0.97	1.00	1.00
		0.27	1.27	0.23	0.23	1.00	1.00
		0.09	1.09	1.02	1.02	1.32	1.32
		0.04	1.04	1.03	1.03	0.86	0.86
		Var=0.02	Var=0.02				
		SD=0.15	SD=0.15				
	50	0.40	1.40	1.04	1.04	1.00	1.00
		0.34	1.34	0.68	0.68	1.14	1.14

		0.28	1.28	0.95	0.95	1.16	1.16
		0.12	1.12	1.22	1.22	0.94	0.94
		0.00	1.00	0.90	0.90	1.01	1.01
		Var=0.03	Var=0.03				
		SD=0.16	SD=0.16				
		Correlation	Eigenvalue	X mean vector		Y mean vector	
Distribution	Sample	CCA	GCCA	CCA	GCCA	CCA	GCCA
Weibull	10	1.00	2.00	1.49	1.49	1.88	1.88
		0.99	1.99	1.56	1.56	1.80	1.80
		0.73	1.73	1.58	1.58	1.99	1.99
		0.51	1.51	2.06	2.06	1.30	1.30
		0.00	1.00	1.94	1.94	1.87	1.87
		Var=0.17	Var=0.17				
		SD=0.41	SD=0.41				
	20	0.80	1.80	1.18	1.18	1.02	1.02
		0.60	1.60	0.78	0.78	1.16	1.16
		0.41	1.41	0.97	0.97	0.86	0.86
		0.20	1.20	1.05	1.05	1.27	1.27
		0.11	1.11	1.17	1.17	1.17	1.17
		Var=0.80	Var=0.80				
		SD=0.28	SD=0.28				
	30	0.61	1.61	0.88	0.88	1.07	1.07
		0.50	1.50	0.77	0.77	0.94	0.94
		0.42	1.42	0.91	0.91	1.03	1.03
		0.15	1.15	0.71	0.71	0.94	0.94
		0.02	1.02	0.95	0.95	0.94	0.94
		Var=0.06	Var=0.06				
		SD=0.24	SD=0.24				
	40	0.66	1.66	0.84	0.84	0.91	0.91
		0.44	1.44	1.23	1.23	1.25	1.25
		0.32	1.32	1.29	1.29	1.15	1.15
		0.21	1.21	1.26	1.26	1.03	1.03
		0.03	1.03	0.77	0.77	1.22	1.22
		Var=0.05	Var=0.05				
		SD=0.23	SD=0.23				
	50	0.60	1.60	1.00	1.00	1.19	1.19
		0.35	1.35	0.82	0.82	0.99	0.99
		0.21	1.21	1.09	1.09	1.12	1.12
		0.17	1.17	0.88	0.88	0.97	0.97
		0.44	1.44	1.14	1.14	1.01	1.01
		Var=0.03	Var=0.03				
		SD=0.17	SD=0.17				
		Correlation	Eigenvalue	X mean vector		Y mean vector	
Distribution	Sample	CCA	GCCA	CCA	GCCA	CCA	GCCA
Normal	10	1.00	2.00	0.34	0.34	-0.05	-0.05
		0.97	1.97	0.68	0.68	0.30	0.30

		0.71	**1.71**	**0.27**	**0.27**	**0.14**	**0.14**
		0.55	**1.55**	**-0.36**	**-0.36**	**-0.46**	**-0.46**
		0. 11	**1.11**	**-0.03**	**-0.03**	**0.21**	**0.21**
		Var=0.13	**Var=0.13**				
		SD=0.36	**SD=0.36**				
	20	**0.72**	**1.72**	**-0.26**	**-0.26**	**0.38**	**0.38**
		0.63	**1.63**	**-0.34**	**-0.34**	**-0.28**	**-0.28**
		0.45	**1.45**	**-0.37**	**-0.37**	**-0.24**	**-0.24**
		0.22	**1.22**	**0.04**	**0.04**	**-0.18**	**-0.18**
		0.05	**1.05**	**-0.24**	**-0.24**	**0.09**	**0.09**
		Var=0.07	**Var=0.07**				
		SD=0.27	**SD=0.27**				
	30	**0.56**	**1.56**	**0.28**	**0.28**	**0.00**	**0.00**
		0.43	**1.43**	**0.00**	**0.00**	**0.02**	**0.02**
		0.37	**1.37**	**-0.01**	**-0.01**	**-0.02**	**-0.02**
		0.17	**1.17**	**-0.21**	**-0.21**	**0.00**	**0.00**
		0.01	**1.01**	**-0.08**	**-0.08**	**-0.19**	**-0.19**
		Var=0.04	**Var=0.04**				
		SD=0.21	**SD=0.21**				
	40	**0.53**	**1.53**	**0.10**	**0.10**	**-0.02**	**-0.02**
		0.44	**1.44**	**-0.16**	**-0.16**	**0.07**	**0.07**
		0.32	**1.32**	**-0.26**	**-0.26**	**0.01**	**0.01**
		0.29	**1.29**	**0.12**	**0.12**	**0.19**	**0.19**
		0.01	**1.01**	**-0.02**	**-0.02**	**0.03**	**0.03**
		Var=0.03	**Var=0.03**				
		SD=0.19	**SD=0.19**				
	50	**0.36**	**1.36**	**0.05**	**0.05**	**0.00**	**0.00**
		0.34	**1.34**	**-0.12**	**-0.12**	**-0.04**	**-0.04**
		0.25	**1.25**	**-0.06**	**-0.06**	**0.00**	**0.00**
		0.16	**1.16**	**0.00**	**0.00**	**-0.14**	**-0.14**
		0.01	**1.01**	**0.24**	**0.24**	**0.02**	**0.02**
		Var=0.02	**Var=0.02**				
		SD=0.14	**SD=0.14**				

5.0 SUMMARY AND CONCLUSION

This study compared CCA and GCCA using simulated data from five distributions namely Weibull, Beta, Gamma, Exponential and Normal distributions. The aim is to determine whether there is any difference in the two methods for two data sets as well as determine the coefficients of X and Y variates. This was done by comparing the relative efficiencies of the methods using variances and standard deviations from the five distributions as control. The summary result of table 1 above shows that the relative efficiencies of CCA and GCCA is the same since the variances and standard deviations of the correlations and eigenvalues of the methods are the same for the five distributions using sample sizes 10, 20, 30, 40 and 50 replicated 10,000 times. It was equally observed from the summary result table that X and Y variates for CCA and GCCA do not defer. Hence we conclude that there is no difference in the two methods for two data set. This is also in line with the findings of Van del Velden (2011) who proved using mathematical derivations that when there are only two sets of variables, the orthogonality of the group configuration of the generalized canonical correlation implies the orthogonality of the canonical correlation with scaling as the only difference between the methods.

REFERENCES

Anderson, T.W: ***"An Introduction to Multivariate Statistical Analysis".*** John Wiley & Sons, Inc., New York, 1958.

Caroll, J.D. and Chang, J.J. (1970): ***Analysis of individual differences in multidimensional Scaling via an N-way generalization of the "Eckart-Young" decomposition.*** Psychometrika, 35,283-319.

Cramer,E (1973): ***A Simple Derivation of the Canonical Correlation Equations***. Biometrics, 29(2),379-380.

Gujarati,D. (2004): ***Basic Econometrics (4th ed.).*** The McGraw Hill Companies.

Horst, P: **"*Generalized Canonical Correlations and their applications to experimental data*".** Journal of clinical Psychology (Monograph Supplement), N0 14(October, 1961), pp331-347.

Hotelling, H. (1936): ***Simplified Calculation of Principal Components***. Psychometrika, Vol. 1, pp2735

Kettenring, J.R. (1971): ***"Canonical Correlation Analysis of several variables".*** Biometrika, vol 56, pp. 433-451.

Krzanowski, W.J ;(1993): "***Principles of Multivariate Analysis"*** Oxford University Press inc. New York.

Lahti, L. and Houvilainen (2013): ***Depending Modeling Tool Kit***. Statistical Machine learning and Bioinformatics; 14:965-1003.

Mackeon, J.J.(1966): ***Canonical Analysis: Some Relationship between Canonical Correlation, Factor Analysis, Discriminant Function Analysis and Scaling Theory"*** Psychometric Monographs, N0.13

Muller,K. (1982): ***Understanding Canonical Correlation Analysis through the General Linear Model and Principal Components.*** The American Statisticians, 36 (4), 342-346.

Onyeagu, S.1. (2003): ***A First course in Multivariate Statistical Analysis***. Mega Concept, Nigeria.

Takane, Y; and Hwang, H,(2002):"***Generalized Constrained Canonical Correlation Analysis"***. Multivariate Research,37(2), 163-195 .

Van deVelden, M. (2011).***"Generalized Canonical Correlation Analysis" International Statistical Institute***. Proceedings of the 58th world Statistical congress, Dublin (session 1PS042); pages 758-765

Vande Velden, M., and Bijmolt, T.H.(2006): ***Generalized canonical correlation analysis of matrices with missing rows***: A simulation study. Psychometrica 71, 323-33