IJO - INTERNATIONAL JOURNAL OF MATHEMATICS

(ISSN: 2992-4421) Osuagwu, Chidimma Udo.*

https://ijojournals.com/ Volume 07 || Issue 08 || August, 2024 ||

PERFORMANCE EVALUATION OF CANONICAL CORRELATION ANALYSIS AND REDUNDANCY ANALYSISUSING GAUSSIAN, GAMMA, EXPONENTIAL AND BETA DISTRIBUTED DATA

Osuagwu, Chidimma Udo
Department of Statistics, Federal University of Technology, Owerri, Imo State Nigeria

Okenwe Idochi
Department of Statistics, School of Applied Sciences, Ken Saro Wiwa Polytechnic PMB 20, Bori, Rivers State Nigeria

Abstract

This study was embarked to examine the performance evaluation of canonical correlation and redundancy analysis with some continuous distributed data (Gaussian, Gamma, Exponential and Beta). The objectives of the study were to: obtain the relative efficiency of CCA and RDA techniques for four continuous distributed simulated data; and determine the model performance adequacy of CCA and RDA techniques. Three variates of the response variable (Y_1, Y_2, Y_3) and three variates of independent variables (X_1, X_2, X_3) were used for the simulation. The means used for response and independent variables for the Gaussian distribution were 80, 85 and 90, whereas their standard deviations were 10, 12 and 15. The alpha values used for response and independent variables for the Gamma distribution were 80, 85 and 90 whereas their theta values were 40, 43 and 45. The rates parameters used for response and independent variables for the Exponential distribution were 0.5. 0.7 and 0.9; whereas the shape parameters used for the Beta distribution were taking from 2 to 5 values. The adequacy of the CCA and RDA was evaluated with Wilcoxon rank sum test; and the study concluded thatRDA was more efficient than that of CCA for the Beta distributed data, while for Gaussian, Gamma and Exponential distributed data, the relative efficiency of the CCA and RDA was the same. The study also concluded that the X-variates of the CCA and RDA did not differ.

Keywords: Canonical correlation analysis,Redundancy analysis, Gaussian, Gamma, Exponential, Beta,Performance evaluation, Simulated data.

1 Introduction

Canonical Correlation Analysis (CCA) and Redundancy Analysis (RDA) are multivariate statistical techniques used to analyze the relationships between two or more sets of variables.CCA is a method for analyzing the relationships between two sets of variables, X and Y, by finding the linear combinations of variables that maximize the correlation between the two sets, which was developed by Hotelling in 1936 (Górecki et al, 2020). Canonical Correlation Analysis (CCA) involves finding a linear transformation that converts the original variables from

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS
(ISSN: 2992-4421)
https://ijojournals.com/
Osuagwu, Chidimma Udo.*
Volume 07 || Issue 08 || August, 2024 ||

two sets into new sets of variables (Wang et al., 2022). These new variables have the property of being uncorrelated within each set, but maximally correlated between sets. The resulting pairs of new variables are called canonical variates, and the correlation coefficients between these pairs are known as canonical correlations. By identifying these canonical variates and correlations, CCA reveals the underlying relationships between the two sets of variables (Li et al., 2020).

RDA is a statistical method that summarizes the linear relationships between two sets of variables, one set being the explanatory variables and the other being the response variables (Ramette, 2017). It's an extension of multiple linear regression; allowing for multiple response variables to be regressed on multiple explanatory variables. RDA produces an ordination that summarizes the main patterns of variation in the response matrix, which can be explained by a matrix of explanatory variables (Hui&Warton, 2022).

The results of RDA include the total variance of the data set, partitioned into constrained and unconstrained variances, which shows how much variation in the response variables was redundant with the variation in the explanatory variables (Székely et al., 2020). RDA also produces scores for objects, response variables, and explanatory variables, which can be used to ordinate points and vectors. RDA is often used in ecological studies to relate environmental variables to species composition. For example, Ramette (2007) used RDA to analyze the relationships between microbial community composition and environmental variables in coastal sands

This study therefore was aimed to: ascertain the relative efficiency of CCA and RDA techniques for four continuous distributed simulated data; and determine the model performance adequacy of CCA and RDA techniques.

2 Review of Related Literature

Makino (2022) explored the application of rotation in correspondence analysis (CA) from a canonical correlation perspective. CA is a statistical method used to visualize the relationship between two categorical variables, typically emphasizing graphical representations. Makino's study introduced a CA formulation based on canonical correlation analysis (CCA), where correlations within and between row/column categories in a reduced dimensional space can be expressed through canonical variables. However, existing CCA-based formulations only allowed for orthogonal rotation. Makino proposed an alternative CCA-based formulation that permits oblique rotation, defining the CA loss function as maximizing the generalized coefficient of determination, which measures the proximity between two variables. The study demonstrated the benefits of the proposed formulation through simulation studies and real data examples.

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS

(ISSN: 2992-4421) Osuagwu, Chidimma Udo.*

https://ijojournals.com/ Volume 07 || Issue 08 || August, 2024 ||

PERFORMANCE EVALUATION OF CANONICAL CORRELATION ANALYSIS AND REDUNDANCY ANALYSISUSING GAUSSIAN, GAMMA, EXPONENTIAL AND BETA DISTRIBUTED DATA

Nayir and Saridas (2022) investigated the relationship between culturally responsive teacher roles and innovative work behavior using canonical correlation analysis. The study aimed to identify the relationship between these two constructs based on teachers' views. The results showed that the first canonical function, which maximized the relationship between the two datasets, shared approximately 77% variance. Furthermore, the analysis revealed a positive relationship between the culturally regulating teacher (CRT) and culturally mediating teacher (CMT) variables in the culturally responsive teacher roles dataset and the GII and FSI variables in the innovative work behavior dataset.

McKeague and Zhang (2021) investigated significance testing for canonical correlation analysis in high-dimensional settings. They addressed the challenge of testing for linear relationships between large sets of random variables using post-selection inference techniques. The authors developed a stabilized one-step estimator for the Euclidean norm of canonical correlations, which was shown to be consistent and asymptotically normal under certain conditions. They also proposed a greedy search algorithm for computing the estimator, leading to a computationally tractable omnibus test for the global null hypothesis. Additionally, they constructed a confidence interval that accounted for variable selection.

García-Valdés et al. (2020) conducted a study using Redundancy Analysis (RDA) to examine the impacts of climate change on species distribution in a Mediterranean ecosystem, incorporating 155 plant species and 15 environmental variables. The analysis revealed that climate variables, including temperature, precipitation, and drought, explained a significant portion of the variation in species distribution, accounting for 24.5% of the variation. Additionally, soil and topographic variables played important roles, explaining 20.1% and 15.4% of the variation, respectively. The study's findings suggested that climate change led to shifts in species distribution, resulting in some species expanding their ranges while others contract. The RDA framework provided a powerful tool for understanding the complex relationships between climate change and species distribution, with important implications for conservation and management efforts in the face of climate change.

3 Materials and Methods

3.1 Canonical Variates and Canonical Correlations

The canonical correlations measure the strength of association between the two sets of variables (Wang & Liu, 2022).The first group of p variables is represented by the $(p \times 1)$ random vector $\mathbf{X}^{(1)}$, while the second group of q variables is represented by the $(q \times 1)$ random vector $\mathbf{X}^{(2)}$. It will be assumed, in the theoretical development, that $\mathbf{X}^{(1)}$ represents the smaller set, so that $p \leq q$. For the random vectors $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, let

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS
(ISSN: 2992-4421)
https://ijojournals.com/
Osuagwu, Chidimma Udo.*
Volume 07 || Issue 08 || August, 2024 ||

$$\left.\begin{aligned} &E(\mathbf{X}^{(1)}) = \boldsymbol{\mu}^{(1)}; \quad Cov(\mathbf{X}^{(1)}) = \Sigma_{11} \\ &E(\mathbf{X}^{(2)}) = \boldsymbol{\mu}^{(2)}; \quad Cov(\mathbf{X}^{(2)}) = \Sigma_{22} \\ &Cov(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \Sigma_{11} = \Sigma'_{22} \end{aligned}\right\} \tag{1}$$

It will be convenient to consider $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ jointly, so, the random vector

$$\underset{((p+q)\times 1)}{\mathbf{X}} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \hdashline \mathbf{X}^{(2)} \end{bmatrix} = \begin{bmatrix} X_1^{(1)} \\ X_2^{(1)} \\ \vdots \\ X_p^{(1)} \\ \hdashline X_1^{(2)} \\ X_2^{(2)} \\ \vdots \\ X_q^{(2)} \end{bmatrix} \tag{2}$$

has mean vector

$$\underset{((p+q)\times 1)}{\boldsymbol{\mu}} = E(\mathbf{X}) = \begin{bmatrix} E(\mathbf{X}^{(1)}) \\ \hdashline E(\mathbf{X}^{(2)}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \hdashline \boldsymbol{\mu}^{(2)} \end{bmatrix} \tag{3}$$

and covariance matrix

$$\underset{(p+q)\times(p+q)}{\boldsymbol{\Sigma}} = \mathrm{E}(\mathbf{X} - \mu)\mathrm{E}(\mathbf{X} - \mu)'$$

$$= \left[\begin{array}{c:c} E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' \\ \hdashline E(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' \end{array}\right]$$

$$\left[\begin{array}{c:c} \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}} & \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}} \\ \hdashline \underset{(q\times p)}{\boldsymbol{\Sigma}_{21}} & \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}} \end{array}\right] \tag{4}$$

3.2 Matrices and Computational Procedures of Redundancy Analysis

Let $\mathbf{x}$ be a $p \times 1$ vector that includes p predictor variables in the first set and $\mathbf{y}$ be a $q \times 1$ vectorthat includes q criterion variables in the second set. According to Van Den Wollenberg in 1977, all variables in $\mathbf{x}$ and $\mathbf{y}$ should be standardized variableswith zero mean and unit variance (Gua et al., 2023). Thus, the $(p+q) \times (p+q)$ covariance matrix of the $(p+q) \times 1$ vector $(\mathbf{x}'\mathbf{y}')'$ is a correlation matrix, denoted by $\mathbf{R}$, which can be partitioned as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{xx} & \mathbf{R}_{xy} \\ \mathbf{R}_{yx} & \mathbf{R}_{yy} \end{pmatrix}, \tag{4}$$

where $\mathbf{R}_{xx}$ is the $p \times p$ correlation matrix of $\mathbf{x}$, $\mathbf{R}_{yy}$ is the $q \times q$ correlation matrix of $\mathbf{y}$, and $\mathbf{R}_{yx} = \mathbf{R}'_{xy}$ is a $q \times p$ matrix that includes the inter-set correlations between $\mathbf{x}$ and $\mathbf{y}$.

To construct p redundancy variates, denoted by $\xi_i (i = 1,2,\cdots, p)$, with p predictor variables in $\mathbf{x}$, the characteristic equation is evaluated as shown in Equation (5):

$$(\mathbf{R}_{xy}\mathbf{R}_{yx} - \boldsymbol{\mu}_i \mathbf{R}_{xx})\mathbf{w}_i = 0, \tag{5}$$

where $\boldsymbol{\mu}_i$ is the ith eigenvalue and $\mathbf{w}_i$ is the ith eigen-vector. Then, one can employ the weight coefficients that are the elements of the scaled eigenvector $\mathbf{w}_i$ to construct ξ_i, such that:

(a) ξ_i is uncorrelated with $\xi_j (i \neq j)$, and (b) ξ_i has unit variance $(i = 1,2,\cdots, p)$.

3.3 Continuous Probability Distributions

Four probability distributions known as the Gaussian, Gamma, Beta and Exponential are discussed in this study.

3.3.1 The Gaussian Distribution

A random variable (R.V.) in continuous form say X, choosing the whole real values in intervals $(-\infty, \infty)$ is known to be a Gaussian (also known as normal) distribution with μ and σ^2 as its parameters if the probability density function (pdf) is defined by

$$f(x) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}} \quad e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0 \\ 0 \quad otherwise \end{cases} \tag{6}$$

Where the study used the notation $N(\mu;\sigma^2)$ to show that X is normal with mean μ and variance σ^2 (Sumair, et al., 2021). This pdf is bell-shaped, symmetrical, and centered at its mean value μ.

The entire area bounded by this function $f(x)$ and x-axis is 1 and therefore the area beneath the curve across two values of X, say, a and b with $a < b$, constitutes the probability that the R.V X lies across a and b, which we write as $P(a < X < b)$. An example of a normal R.V is height of students at a specified age for a specified sex in a specified racial group even though heights must be positive (El-Morshedy et al., 2021).

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS
(ISSN: 2992-4421)
https://ijojournals.com/
Volume 07 || Issue 08 || August, 2024 ||
Osuagwu, Chidimma Udo.*

The pdf of a standard normal distribution is

$$f(z)=\begin{cases}1\Big/\sqrt{2\pi}\,e^{-\frac{z^2}{2}}, & -\infty<z<\infty\\ 0, & otherwise\end{cases} \tag{7}$$

The respective mean and variance of a Gaussian distribution are respectively given as;

$$E(X)=\mu \tag{8}$$

and

$$Var(X)=\sigma^2 \tag{9}$$

3.3.2 Gamma Distribution

A R.V X is said to follow a gamma R.V with parameters αand β, if its pdf is given by:

$$f(x)=\begin{cases}\dfrac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}, & x\ge 0,\ \alpha>0,\ \beta>0\\ 0, \text{ otherwise}\end{cases} \tag{10}$$

where $\Gamma(\alpha)$ is the gamma function defined as;

$$\Gamma(\alpha)=\int_0^{\infty}t^{\alpha-1}e^{-t}\,dt \tag{11}$$

The gamma probability density function as given in Equation (10) is a normal or legitimate pdf.

The respective mean and variance of a Gamma distribution are respectively given as;

$$E(X)=\alpha\beta \tag{12}$$

and

$$Var(X)=\alpha\beta^2 \tag{13}$$

To obtain the scale (β) and shape (α) parameters of a gamma distribution, we have

$$E(X)=\mu=\alpha\beta \tag{14}$$

$$Var(X)=\sigma^2=\alpha\beta^2 \tag{15}$$

From Equation (14), put $\alpha=\dfrac{\mu}{\beta}$ into Equation (15) to obtain

$$\sigma^2=\mu\beta \tag{16}$$

From Equation (16), the scale parameter is obtained as

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS

(ISSN: 2992-4421)

https://ijojournals.com/

Osuagwu, Chidimma Udo.*

Volume 07 || Issue 08 || August, 2024 ||

$$\beta = \frac{\sigma^2}{\mu} \tag{17}$$

Substitute Equation (17) into Equation (14) to obtain the shape parameter as

$$\alpha = \frac{\mu^2}{\sigma^2} \tag{18}$$

Equations (17) and (18) were employed in the simulation of data for gamma distribution in this study.

3.2.3 The Exponential Distribution

A continuous random variable X, is said to have an exponential distribution with parameter $\lambda > 0$ if it has a probability density function defined by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & otherwise \end{cases} \tag{19}$$

The respective mean and variance of an Exponential distribution are respectively given as;

$$E(X) = \frac{1}{\lambda} \tag{20}$$

and

$$Var(X) = \frac{1}{\lambda^2} \tag{21}$$

To obtain the rate parameter (λ) of an exponential distribution, we have

$$E(X) = \mu = \frac{1}{\lambda} \Rightarrow \lambda = \frac{1}{\mu} \tag{22}$$

Equations(22) was employed in the simulation of data for an exponential distribution in this study.

3.2.4 Beta Distribution

A standard beta distribution is a two-parameter family of distribution for a continuous random variable Y , defined in a finite interval on a real line with its density function given by

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS
(ISSN: 2992-4421)
https://ijojournals.com/
Volume 07 || Issue 08 || August, 2024 ||
Osuagwu, Chidimma Udo.*

$$f(y) = \begin{cases} \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)}, & 0 \le y \le 1 \\ 0, \text{otherwise} \end{cases} \tag{23}$$

where $\alpha, \beta > 0$ and $B(\alpha,\beta)$ is the beta function; its formula is given by

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt \tag{24}$$

The mean and variance of Y are

$$E(X) = \mu = \frac{\alpha}{\alpha+\beta} \tag{25}$$

and

$$Var(X) = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} \tag{26}$$

The scale (β) and shape (α) parameters of a beta distribution are obtained as;

$$\alpha = \frac{-\mu(\sigma^2+\mu^2-\mu)}{\sigma^2} \tag{27}$$

and

$$\beta = \frac{(\sigma^2+\mu^2-\mu)(\mu-1)}{\sigma^2} \tag{28}$$

4 Results

4.1 Simulated Data of Different Sample Sizes for CCA and RDA

Data were simulated on R-Studio command window, calling for the CCA and RDA function for Gaussian distribution, Gamma distribution, Exponential distribution and Beta distribution for samples of sizes 10, 20, 30, 40, 50, 60 and 70. Three variates of the response variable (Y_1, Y_2, Y_3) and three variates of independent variables (X_1, X_2, X_3) were used for the simulation. The means used for response and independent variables for the Gaussian distribution were 80, 85 and 90,whereas their standard deviations were 10, 12 and 15. The alpha values used for response and independent variables for the Gamma distribution were 80, 85 and 90 whereas their theta values were 40, 43 and 45. The rates parameters used for response and independent variables for the Exponential distribution were 0.5. 0.7 and 0.9; whereas the shape parameters used for the Beta distribution were taking from values from 2 to 5 and the results obtained are summarized in Table 1.

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS
(ISSN: 2992-4421)
https://ijojournals.com/
Osuagwu, Chidimma Udo.*
Volume 07 || Issue 08 || August, 2024 ||

Table 1: Summary Results from the Four Distributions for Different Sample Sizes

Distribution		Correlation	Eigen-value	X-Mean Vector		Y-Mean Vector	
	Sample	CCA	RDA	CCA	RDA	CCA	RDA
Gaussian	10	0.9609	0.7180	81.6417	81.6417	78.5301	78.5301
		0.5303	0.2717	75.8878	75.8878	89.9215	89.9215
		0.4498	0.0020	89.3103	89.3103	84.7729	84.7729
		SD = 0.2748	**SD = 0.3616**				
	20	0.5343	0.5715	81.5175	81.5175	76.8798	76.8798
		0.3713	0.2015	82.4420	82.4420	87.3907	87.3907
		0.1763	0.0098	91.7642	91.7642	95.7642	95.7642
		SD = 0.1792	**SD = 0.2855**				
	30	0.4399	0.3159	82.1643	82.1643	80.9557	80.9557
		0.2426	0.0852	85.5123	85.5123	86.2550	86.2550
		0.0318	0.0050	91.6238	91.6238	89.4687	89.4687
		SD = 0.2041	**SD = 0.1614**				
	40	0.2583	0.1442	80.8160	80.8160	78.6760	78.6760
		0.1231	0.1146	87.3596	87.3596	83.7756	83.7756
		0.0028	0.0044	90.5419	90.5419	89.0918	89.0918
		SD = 0.1278	**SD = 0.0737**				
	50	0.3583	0.0809	78.4572	78.4572	79.9256	79.9256
		0.1483	0.0211	85.5144	85.5144	84.7669	84.7669
		0.0103	0.0057	89.8407	89.8407	92.1156	92.1156
		SD = 0.1752	**SD = 0.0397**				
	60	0.3444	0.0599	81.1089	81.1089	80.7678	80.7678
		0.1502	0.0105	79.5728	79.5728	86.6178	86.6178
		0.031	0.0012	88.9582	88.9582	89.5820	89.5820
		SD = 0.158	**SD = 0.0316**				
	70	0.2150	0.1753	81.4539	81.4539	81.7998	81.7998
		0.0973	0.0724	85.5524	85.5524	82.6863	82.6863
		0.0012	0.0082	91.8743	91.8743	90.5477	90.5477
		SD = 0.1071	**SD = 0.0843**				
Gamma	10	0.8263	0.8768	3311.883	3311.883	3117.559	3117.559
		0.4987	0.5569	3691.902	3691.902	3520.850	3520.850
		0.2967	0.3712	3967.700	3967.700	4020.455	4020.455
		SD = 0.2673	**SD = 0.2558**				
	20	0.3916	0.2781	3145.036	3145.036	3329.500	3329.500
		0.1921	0.0870	3703.089	3703.089	3672.906	3672.906
		0.0523	0.0022	4019.466	4019.466	4122.456	4122.456
		SD = 0.1705	**SD = 0.1413**				
	30	0.5626	0.1461	3316.971	3316.971	3139.415	3139.415
		0.3175	0.0736	3601.826	3601.826	3687.641	3687.641
		0.0309	0.0065	4074.426	4074.426	4072.765	4072.765
		SD = 0.2661	**SD = 0.0698**				
	40	0.4502	0.0491	3151.013	3151.013	3127.473	3127.473
		0.2016	0.0176	3618.440	3618.440	3702.444	3702.444
		0.0370	0.0014	4011.102	4011.102	4057.979	4057.979
		SD = 0.2080	**SD = 0.0243**				
	50	0.3273	0.1721	3288.111	3288.111	3169.010	3169.010
		0.2901	0.0077	3565.821	3565.821	3641.781	3641.781
		0.0622	0.0005	4136.478	4136.478	4100.952	4100.952
		SD = 0.1435	**SD = 0.0971**				
	60	0.2258	0.0436	3158.815	3158.815	3195.705	3195.705
		0.1697	0.0121	3631.133	3631.133	3751.599	3751.599
		0.0145	0.0001	4115.701	4115.701	4001.559	4001.559
		SD = 0.1095	**SD = 0.0225**				

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS
(ISSN: 2992-4421)
https://ijojournals.com/
Osuagwu, Chidimma Udo.*
Volume 07 || Issue 08 || August, 2024 ||

	70	0.2659	0.2709	3211.757	3211.757	3239.607	3239.607
		0.2170	0.1029	3666.475	3666.475	3686.072	3686.072
		0.1373	0.0066	4075.567	4075.567	3998.146	3998.146
		SD = 0.0649	**SD = 0.1338**				
Exponential	10	0.4361	0.8230	1.1191	1.1191	2.4481	2.4481
		0.3081	0.0420	2.8346	2.8346	1.5564	1.5564
		0.1447	0.0237	0.8268	0.8268	1.1561	1.1561
		SD = 0.1461	**SD = 0.4563**				
	20	0.4120	0.3147	2.2670	2.2670	1.9106	1.9106
		0.3517	0.1538	1.3586	1.3586	1.6637	1.6637
		0.0190	0.0013	1.6114	1.6114	1.1930	1.1930
		SD = 0.2117	**SD = 0.1567**				
	30	0.4655	0.2104	1.5786	1.5786	1.5809	1.5809
		0.2066	0.0256	1.2137	1.2137	1.1047	1.1047
		0.0284	0.0042	0.8006	0.8006	0.7525	0.7525
		SD = 0.2198	**SD = 0.1134**				
	40	0.4752	0.0573	1.4892	1.4892	1.8839	1.8839
		0.2934	0.0276	1.2863	1.2863	1.4824	1.4824
		0.1192	0.0041	0.9426	0.9426	1.0180	1.0180
		SD = 0.1780	**SD = 0.0267**				
	50	0.2732	0.2847	2.0737	2.0737	2.3757	2.3757
		0.2422	0.1221	1.5447	1.5447	1.5730	1.5730
		0.0173	0.0188	0.9702	0.9702	1.0287	1.0287
		SD = 0.1397	**SD = 0.1340**				
	60	0.3197	0.1642	2.0932	2.0932	1.6467	1.6467
		0.2416	0.0383	1.4325	1.4325	1.6291	1.6291
		0.1147	0.0150	1.0182	1.0182	1.1243	1.1243
		SD = 0.1035	**SD = 0.0803**				
	70	0.3419	0.0333	1.8292	1.8292	1.9513	1.9513
		0.2172	0.0152	1.2600	1.2600	1.2107	1.2107
		0.0537	0.0081	0.9883	0.9883	0.9481	0.9481
		SD = 0.1445	**SD = 0.0130**				
Beta	10	0.7963	0.2434	0.2932	0.2932	0.2610	0.2610
		0.4646	0.0627	0.4356	0.4356	0.3669	0.3669
		0.0710	0.0031	0.6326	0.6326	0.5026	0.5026
		SD = 0.3631	**SD = 0.1251**				
	20	0.5851	0.5324	0.3218	0.3218	0.2944	0.2944
		0.5173	0.1768	0.4216	0.4216	0.4652	0.4652
		0.3069	0.0034	0.6327	0.6327	0.5084	0.5084
		SD = 0.1451	**SD = 0.2697**				
	30	0.4148	0.3035	0.3263	0.3263	0.2616	0.2616
		0.1741	0.0262	0.4498	0.4498	0.4344	0.4344
		0.0256	0.0080	0.5727	0.5727	0.5483	0.5483
		SD = 0.1964	**SD = 0.1656**				
	40	0.4796	0.1522	0.3285	0.3285	0.2641	0.2641
		0.1904	0.0629	0.4468	0.4468	0.4209	0.4209
		0.1156	0.0511	0.6191	0.6191	0.6008	0.6008
		SD = 0.1922	**SD = 0.0553**				
	50	0.5701	0.0681	0.2805	0.2805	0.2751	0.2751
		0.2277	0.0133	0.4535	0.4535	0.4555	0.4555
		0.0613	0.0001	0.6037	0.6037	0.5730	0.5730
		SD = 0.2594	**SD = 0.0361**				
	60	0.5456	0.1719	0.2424	0.2424	0.2994	0.2994
		0.2209	0.0387	0.4367	0.4367	0.4373	0.4373
		0.1515	0.0067	0.5737	0.5737	0.5745	0.5745
		SD = 0.2104	**SD = 0.0876**				
		0.2990	0.0806	0.2953	0.2953	0.3030	0.3030
		0.0762	0.0192	0.4589	0.4589	0.4352	0.4352

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS
(ISSN: 2992-4421)
https://ijojournals.com/
Volume 07 || Issue 08 || August, 2024 ||
Osuagwu, Chidimma Udo.*

	70	0.0525	0.0003	0.6028	0.6028	0.5580	0.5580
		SD = 0.1360	**SD = 0.0420**				

Table 1 shows the standard deviation of the correlations and eigenvalues for CCA and RDA respectively.It can be observed that the standard deviation of the RDA is lower than that of CCA except for the cases of sample sizes 10 and 20 for Gaussian distribution; sample size 70 for Gamma distribution, sample size 10 for Exponential distribution and sample size 20 for Beta distribution, but there is need to examine if the differencesare significant. It is also observed that the X and Y-variates of the CCA and RDA do not differ.

4.2 Model Performance Adequacy of CCA and RDA Techniques

Table 2: Summary of Decision for Testing SD Values for CCA and RDA

		SD Values		Ranks		Z	p-value	Decision
Distribution	**Sample**	**CCA**	**RDA**	**CCA**	**RDA**			
Gaussian	10	0.2748	0.3616	12	14	0.958	0.338	Do not Reject H_0
	20	0.1792	0.2855	10	13			
	30	0.2041	0.1614	11	8			
	40	0.1278	0.0737	6	3			
	50	0.1752	0.0397	9	2			
	60	0.158	0.0316	7	1			
	70	0.1071	0.0843	5	4			
Gamma	10	0.2673	0.2558	14	12	1.725	0.085	Do not Reject H_0
	20	0.1705	0.1413	10	8			
	30	0.2661	0.0698	13	4			
	40	0.2080	0.0243	11	2			
	50	0.1435	0.0971	9	5			
	60	0.1095	0.0225	6	1			
	70	0.0649	0.1338	3	7			
Exponential	10	0.1461	0.4563	9	14	1.469	0.142	Do not Reject H_0
	20	0.2117	0.1567	12	10			
	30	0.2198	0.1134	13	5			
	40	0.1780	0.0267	11	2			
	50	0.1397	0.1340	7	6			
	60	0.1035	0.0803	4	3			
	70	0.1445	0.0130	8	1			
Beta	10	0.3631	0.1251	14	5	2.108	0.035	Reject H_0
	20	0.1451	0.2697	7	13			
	30	0.1964	0.1656	10	8			
	40	0.1922	0.0553	9	3			
	50	0.2594	0.0361	12	1			
	60	0.2104	0.0876	11	4			
	70	0.1360	0.0420	6	2			

Table 2 shows the Wilcoxon rank sum testsignificance difference result for the four continuous distributions employed in this study. The result reveals that there is no significant difference in the standard deviation of the correlations and eigenvalues for the methods for Gaussian, Gamma and Exponential distributions. This implies that the relative efficiency of the CCA and RDA is the same for the Gaussian, Gamma and Exponential distributed data. On the other hand, the result reveals thatthere is significant difference in the standard deviation of the correlations and eigenvalues for the methods for Beta distribution. This implies that RDA is more efficient than that of CCA for the Beta distributed data.

4 Conclusion

This study used canonical correlation and redundancy analysis via four continuous distributions (Gaussian, Gamma, Exponential and Beta) in order to assess their performances. The adequacy of the CCA and RDA was evaluated with Wilcoxon rank sum test; and the study concluded thatRDA is more efficient than that of CCA for the Beta distributed data, while for Gaussian, Gamma and Exponential distributed data, the relative efficiency of the CCA and RDA is the same. The study also concluded that the X-variates of the CCA and RDA do not differ.

References

El-Morshedy, M., Alshammari, F.S., Hamed, Y. S., Eliwa, M.S. &Yousof, H.M. (2021).A New Family of Continuous Probability Distributions.*Entropy*, 23(194), 1–24.

García-Valdés, R., Sánchez, A. M., Fernández-Palacios, J. M., Padrón, R. P., & Rodríguez-Rodríguez, M. A. (2020). Climate change impacts on species distribution: A Redundancy Analysis approach. Ecography, 43(1), 141-152.

Góreck, T., Krzysko, M. &Wołynski, W. (2020).Generalized canonical correlation analysis for functional data.*Biometrical Letters*, *57*(2020), 1 – 12.

Gua, F., Yungb, Y., Cheungc, M. W. L., Jood, B. K. &Nimon, K. (2023). Statistical inference in redundancy analysis: a direct covariance structure modeling approach. *Multivariate Behavioral Research*, 58(5), 877–893.

Hui, F. K. C., &Warton, D. I. (2022).Redundancy analysis and related methods for multivariate ecological data.*Methods in Ecology and Evolution*, 13(1), 15-28. doi: 10.1111/2041-210X.13734

Li, M., Xu, X., & Chen, J. (2020).Integrative analysis of gene expression and disease outcomes using canonical correlation analysis.*Bioinformatics*, 36(10), 2911-2918.

Makino, N. (2022). Rotation in correspondence analysis from the canonical correlation perspective.*Psychometrika,5(2022)*, 18–28.

McKeague, I. W. & Zhang, X. (2021).Significance testing for canonical correlation analysis in high dimensions. *Biometrika*, 2021.

Nayir, F. &Saridas, G. (2022). The relationship between culturally responsive teacher roles and innovative work behavior: canonical correlation analysis. *Journal of Educational Research and Practice*, *12(2022)*, 36 –50.

Ramette, A. (2007). Multivariate analyses in microbial ecology.*FEMS Microbiology Ecology*, 62(2), 142-160.

IJO - INTERNATIONAL JOURNAL OF MATHEMATICS
(ISSN: 2992-4421)
https://ijojournals.com/
Volume 07 || Issue 08 || August, 2024 ||

Osuagwu, Chidimma Udo.*

Ramette, A. (2017). Multivariate analyses in microbial ecology: A decade of progress. *FEMS Microbiology Ecology*, 93(12), fix106.doi: 10.1093/femsec/fix106

Sumair, M., Aized, T., Gardezi, S.A.R., Bhutta, M.M.A., Rehman, S.M.S. &Rehman, S.U. (2021). Application of five continuous distributions and evaluation of wind potential at five stations using normal distribution. *Energy Exploration & Exploitation*, 39(6), 2214–2239.

Székely, E., Botta-Dukát, Z., &Lengyel, A. (2020).Redundancy analysis as a tool for identifying drivers of community composition in vegetation ecology.*Journal of Vegetation Science*, 31(3), 537-546. doi: 10.1111/jvs.12854

Van de Velden. M. (2011).On generalized canonical correlation analysis.Proc.58th World Statistical Congress. Dublin, 758–765.

Wang, H., Zhang, Y., & Singh, R. (2022). Climate-crop yield relationships: a canonical correlation analysis. *Agricultural and Forest Meteorology*, 313, 108702.

Wang, Y., & Liu, X. (2022).Canonical correlation analysis for identifying relationships between climate variables and crop yields.*Journal of Agricultural Science*, *160*(3), 257-265.