

# Canonical Correlation Analysis via Gaussian and Gamma Distributed Data

J. Opara<sup>1</sup>, P. A. Esemokumo<sup>2</sup> and E. U. Oti<sup>3</sup>

<sup>1</sup>Department of Mathematics and Statistics, Ignatius Ajuru University of Education,  
PMB 5047, Port Harcourt, NIGERIA  
Email : judend88@yahoo.com

<sup>2</sup>Department of Statistics, Federal Polytechnic Ekowe,  
PMB 110, Yenagoa, NIGERIA  
Email : contactperes4good@gmail.com

<sup>3</sup>Department of Statistics, Federal Polytechnic Ekowe,  
PMB 110, Yenagoa, NIGERIA  
Email : eluchcollections@gmail.com

## ABSTRACT

*The study examined the canonical correlation analysis (CCA) via Gaussian and Gamma distributed data. Medical (vital signs and anthropometric indicators) data were collected for the study. Data of different sample sizes (10, 25, 40, 70, 150, 200, 250, 500, 650, 800, 1000, 1200, 1500, 2100, 2500, 2700 and 3000) were simulated from the medical data and used to perform the analysis. A computer programming language codes was written in RStudio for simulation and solution of CCA via Gaussian and Gamma data distribution. The result of the study revealed that the standard error of mean (SEM) of correlation figures for Gamma distribution is 0.04238, which was less than that of SEM of correlation figures of 0.04888 for Gaussian distribution. Further analysis using Mann-Whitney U test statistic showed that the difference in the SEM of correlation figures between the distributions was insignificant. The study thereby concluded that for CCA, the relative efficiency of the Gaussian and Gamma distributed data is the same.*

**Keywords:** Canonical Correlation Analysis, Gaussian, Gamma, Relative Efficiency, Simulation.  
**Mathematics Subject Classification Number:** 62xx

## 1. INTRODUCTION

Canonical Correlation Analysis (CCA), which was created by Hotelling in 1936, is often employed for the study concerning interrelations between two sets of variables (Góreck *et al*, 2020). CCA is the aspect of statistical analysis which is concerned with relationships among sets of dependent variates (Makino, 2022). Corresponding pairs of new variables are named canonical variates, and correlation coefficients within pairs are known as canonical correlations (Akano & Adams, 2019). Numerous public health datasets contain many variables collected from subject which is the same. One method, like CCA, helps summarize the entire data set into a small number of linear combinations of data.

These low-dimensional representations of data can help medical and scientific researchers comprehend the structure of complex datasets, potentially leading to new scientific hypotheses and discoveries. This is particularly useful when researchers are interested in the degree of association between variable sets or relationships within a single set (Affi, 2019).

More importantly, CCA is not only used in testing the independence of two sets of variables, but is also employed as a technique for data reduction to ease the researcher's work and clarify interpretation in coordinate systems. It is imperative however to be reminded that as the magnitude of the canonical correlation decreases, the relationship between the corresponding canonical variables weakens and the resulting predictions become less accurate. Moreover, although this technique is interesting for examining the relationship between variables of two sets and can even provide useful predictive models, we find that its scope is very limited. This is because it only predicts linear combinations of the dependent variables ( $Y_i$ ) and the linear combinations to predict are determined by the data and not under the control of the researcher (Sakamotoa *et al*, 2021).

The classical CCA, though limited to symmetric and linear relationships, has been used indiscriminately in most applications requiring relationship between two sets of variables. A CCA is a generic parametric model employed in the statistical analysis of data requiring interrelated or interdependent input and output variables; hence, it requires that the variables in the data sets are from a Gaussian distribution. It is as a result of the symmetric limitation that this research is being embarked on to examine the correlations from Gaussian and Gamma distributed data.

## 2. MATERIALS AND METHODS

### 2.1. Canonical variates and canonical correlations

CCA measures the relationship in terms of strength between two variable sets. Its maximization nature of the approach denotes attempts to pay a close attention of a high-dimensional association between variables of two sets on a few pairs of canonical variables (Iheagwara *et al*, 2018). The 1st category of  $p$  variables is denoted by  $(p \times 1)$  random vector  $\mathbf{X}^{(1)}$ , while 2nd category of  $q$  variables is  $(q \times 1)$  random vector  $\mathbf{X}^{(2)}$ . It will be considered, in the theoretical development, that  $\mathbf{X}^{(1)}$  represents the lesser set, such that  $p \leq q$ .

In the case of  $\mathbf{X}^{(1)}$  and  $\mathbf{X}^{(2)}$ , let

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

Conveniently considering  $\mathbf{X}^{(1)}$  and  $\mathbf{X}^{(2)}$  jointly, the random vector

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

has mean vector

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

and matrix of covariance

$$\begin{aligned} \boldsymbol{\Sigma}_{(p+q) \times (p+q)} &= E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \\ &= \begin{bmatrix} E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' \\ E(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \end{aligned} \quad (4)$$

If there are relatively big  $p$  and  $q$ , it is usually futile to interpret the elements of  $\boldsymbol{\Sigma}_{12}$  together (Ekezie *et al*, 2013). In addition, it is usually a linear combination of variables that presents encouraging and important predictive desires. The primary assignment of CCA is to analyze  $\mathbf{X}^{(1)}$  and  $\mathbf{X}^{(2)}$  sets in terms of summarization the relationships in carefully selecting a little correlations (or covariance) preferably than  $pq$  covariance in  $\boldsymbol{\Sigma}_{12}$ .

## 2.2. 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  $\mu$  and  $\sigma^2$  as its parameters if the probability density function (pdf) is defined by

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, & -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

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

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).

The pdf of a standard normal distribution is

$$f(z) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, & -\infty < z < \infty \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

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

$$E(X) = \mu \quad (7)$$

and

$$Var(X) = \sigma^2 \quad (8)$$

### 2.3. Gamma distribution

A R.V  $X$  is said to follow a gamma R.V with parameters  $\alpha$  and  $\beta$ , if its pdf is given by:

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)}, & x \geq 0, \alpha > 0, \beta > 0 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

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

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \quad (10)$$

The gamma probability density function as given in equation (9) is a normal or legitimate pdf.

Thus, we have shown that the gamma distribution satisfies the properties of a pdf.

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

$$E(X) = \alpha\beta \quad (11)$$

and

$$Var(X) = \alpha\beta^2 \quad (12)$$

To obtain the scale ( $\beta$ ) and shape ( $\alpha$ ) parameters of a gamma distribution, we have

$$E(X) = \mu = \alpha\beta \quad (13)$$

$$Var(X) = \sigma^2 = \alpha\beta^2 \quad (14)$$

From equation (13), put  $\alpha = \frac{\mu}{\beta}$  into equation (14) to obtain

$$\sigma^2 = \mu\beta \quad (15)$$

From equation (15), the scale parameter is obtained as

$$\beta = \frac{\sigma^2}{\mu} \quad (16)$$

Substitute equation (16) into equation (13) to obtain the shape parameter as

$$\alpha = \frac{\mu^2}{\sigma^2} \quad (17)$$

Equations (16) and (17) were employed in the simulation of data for gamma distribution in this study.

#### 2.4. Mann-Whitney U test

This is a non-parametric test statistic which claims that given two sets of data say X and Y from independent continuous distributions, the X's values ranks in the merged arrangement order of two sets would normally be greater than Y's values ranks if the median for population X is greater than that of population Y.

For large samples, the normal approximation to the distribution or rejection regions for W can be used because of the asymptotic normality of the general linear rank statistic. This approximation is shown to be accurate enough for most practical applications for combined sample sizes  $N \geq 12$  (Gibbons, 2003). The normal distribution approximates the Mann-Whitney U statistic T as:

$$Z = \frac{|R - \mu| - 0.5}{\sigma} \quad (18)$$

Given  $H_0$ ,

$$\mu = \frac{m(N+1)}{2} \quad (19)$$

and

$$\sigma^2 = \frac{mn(N+1)}{12} \quad (20)$$

respectively.  $R$  is the sum of ranks for smaller sample size ( $n$ ),  $m$  is the larger of sample sizes,

$$N = m + n.$$

A table of critical values corresponding to Mann-Whitney U statistic is contained in any standard text. The table gives the rejection regions for level of significance  $\alpha$ , and the sample sizes  $m$  and  $n$ . Of course, if  $Z$  is less than or greater than the critical values, the decision is to reject the null hypothesis in favour of the alternative.

## 2.5. RStudio codes for CCA for gamma distribution of different sample sizes

```

data=read.csv(file.choose())
attach(data)
X1bar=mean(X1)
X2bar=mean(X2)
X3bar=mean(X3)
Y1bar=mean(Y1)
Y2bar=mean(Y2)
Y3bar=mean(Y3)
X1Stddev=sd(X1)
X2Stddev=sd(X2)
X3Stddev=sd(X3)
Y1Stddev=sd(Y1)
Y2Stddev=sd(Y2)
Y3Stddev=sd(Y3)
X1Variance=var(X1)
X2Variance=var(X2)
X3Variance=var(X3)
Y1Variance=var(Y1)
Y2Variance=var(Y2)
Y3Variance=var(Y3)
#Scale
X1theta<-(X1Variance/X1bar)
X1theta
X2theta<-(X2Variance/X2bar)
X2theta
X3theta<-(X3Variance/X3bar)
X3theta
Y1theta<-(Y1Variance/Y1bar)
Y1theta
Y2theta<-(Y2Variance/Y2bar)
Y2theta
Y3theta<-(Y3Variance/Y3bar)
Y3theta
#Scale
X1alpha<-((X1bar^2)/X1Variance)
X1alpha
X2alpha<-((X2bar^2)/X2Variance)
X2alpha
X3alpha<-((X3bar^2)/X3Variance)
X3alpha
Y1alpha<-((Y1bar^2)/Y1Variance)
Y1alpha
Y2alpha<-((Y2bar^2)/Y2Variance)
Y2alpha
Y3alpha<-((Y3bar^2)/Y3Variance)
Y3alpha
#Simulation of Gamma Variables
set.seed(2023)
n=15
Z1<-rgamma(n,X1alpha,scale=X1theta)
Z2<-rgamma(n,X2alpha,scale=X2theta)
Z3<-rgamma(n,X3alpha,scale=X3theta)
W1<-rgamma(n,Y1alpha,scale=Y1theta)
W2<-rgamma(n,Y2alpha,scale=Y2theta)
W3<-rgamma(n,Y3alpha,scale=Y3theta)
#Performing CCA
library(CCP)
X=cbind(Z1,Z2,Z3)

```

```

Y=cbind(W1,W2,W3)
cc=cancor(X,Y)
rho=cc$cor
rho
XMeanVector=colMeans(X)
XMeanVector
YMeanVector=colMeans(Y)
YMeanVector

```

## 2.6. RStudio codes of CCA for Gaussian distribution for different sample sizes

```

data=read.csv(file.choose())
attach(data)
X1bar=mean(X1)
X2bar=mean(X2)
X3bar=mean(X3)
Y1bar=mean(Y1)
Y2bar=mean(Y2)
Y3bar=mean(Y3)
X1Stddev=sd(X1)
X2Stddev=sd(X2)
X3Stddev=sd(X3)
Y1Stddev=sd(Y1)
Y2Stddev=sd(Y2)
Y3Stddev=sd(Y3)
#Simulation of Gaussian Variables
set.seed(2023)
n=15
Z1<-rnorm(n,X1bar,X1Stddev)
Z2<-rnorm(n,X2bar,X2Stddev)
Z3<-rnorm(n,X3bar,X3Stddev)
W1<-rnorm(n,Y1bar,Y1Stddev)
W2<-rnorm(n,Y2bar,Y2Stddev)
W3<-rnorm(n,Y3bar,Y3Stddev)
#Performing CCA
library(CCP)
X=cbind(Z1,Z2,Z3)
Y=cbind(W1,W2,W3)
cc=cancor(X,Y)
rho=cc$cor
rho

```

## 3. RESULTS

Table 1: Data on vital signs (VS) – Y and anthropometric indicators (AI) – X

| VS  |                |                |                |                |                |                | AI  |                |                |                |                |                |                |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|-----|----------------|----------------|----------------|----------------|----------------|----------------|
| S/N | Y <sub>1</sub> | Y <sub>2</sub> | Y <sub>3</sub> | X <sub>1</sub> | X <sub>2</sub> | X <sub>3</sub> | S/N | Y <sub>1</sub> | Y <sub>2</sub> | Y <sub>3</sub> | X <sub>1</sub> | X <sub>2</sub> | X <sub>3</sub> |
| 1   | 16             | 77             | 35.6           | 68             | 77.3           | 169            | 26  | 14             | 52             | 36.4           | 48             | 60.5           | 159            |
| 2   | 14             | 70             | 36.1           | 59             | 79.2           | 165            | 27  | 18             | 57             | 36.5           | 51             | 65.8           | 173            |
| 3   | 18             | 60             | 35.5           | 46             | 66.5           | 161            | 28  | 16             | 56             | 35.6           | 48             | 64.7           | 169            |
| 4   | 19             | 67             | 38.5           | 52             | 76.2           | 165            | 29  | 15             | 51             | 36.2           | 43             | 60.4           | 163            |
| 5   | 20             | 69             | 37.5           | 54             | 78.3           | 167            | 30  | 18             | 56             | 36.4           | 54             | 65.3           | 172            |
| 6   | 10             | 51             | 35.7           | 49             | 60.2           | 164            | 31  | 14             | 50             | 37.4           | 43             | 58.7           | 165            |
| 7   | 18             | 54             | 36.7           | 52             | 63.3           | 171            | 32  | 13             | 50             | 36.4           | 43             | 58.5           | 163            |
| 8   | 17             | 58             | 35.8           | 58             | 67.3           | 172            | 33  | 12             | 60             | 36.4           | 48             | 57.8           | 171            |
| 9   | 16             | 55             | 36.9           | 61             | 64.4           | 168            | 34  | 19             | 68             | 39.1           | 51             | 76.5           | 174            |
| 10  | 15             | 53             | 36.5           | 51             | 61.5           | 163            | 35  | 16             | 58             | 36.2           | 47             | 66.5           | 162            |
| 11  | 16             | 56             | 37.6           | 54             | 65.1           | 166            | 36  | 17             | 61             | 36.4           | 51             | 69.6           | 172            |

|    |    |    |      |    |      |     |    |    |    |      |    |      |     |
|----|----|----|------|----|------|-----|----|----|----|------|----|------|-----|
| 12 | 18 | 76 | 35.7 | 61 | 68.3 | 169 | 37 | 15 | 60 | 36.4 | 46 | 69.4 | 167 |
| 13 | 14 | 63 | 36.5 | 51 | 72.3 | 165 | 38 | 17 | 68 | 37.4 | 50 | 76.5 | 169 |
| 14 | 17 | 40 | 36.3 | 47 | 60.2 | 163 | 39 | 18 | 71 | 38.4 | 57 | 79.5 | 170 |
| 15 | 9  | 59 | 36.9 | 49 | 68.4 | 165 | 40 | 14 | 45 | 37.4 | 46 | 54.4 | 165 |
| 16 | 18 | 52 | 36.5 | 42 | 60.5 | 160 | 41 | 16 | 56 | 37.4 | 51 | 64.5 | 165 |
| 17 | 16 | 57 | 35.8 | 51 | 65.6 | 168 | 42 | 19 | 77 | 35.8 | 61 | 85.5 | 161 |
| 18 | 19 | 60 | 37.8 | 59 | 65.8 | 170 | 43 | 17 | 56 | 37.5 | 49 | 65.4 | 165 |
| 19 | 14 | 54 | 35.4 | 47 | 62.9 | 168 | 44 | 19 | 67 | 36.6 | 52 | 76.4 | 167 |
| 20 | 16 | 53 | 36.2 | 45 | 62.2 | 158 | 45 | 13 | 54 | 36.8 | 51 | 62.8 | 162 |
| 21 | 13 | 61 | 35.8 | 49 | 70.4 | 165 | 46 | 12 | 52 | 37.2 | 47 | 60.9 | 171 |
| 22 | 18 | 61 | 39.1 | 49 | 70.3 | 166 | 47 | 17 | 60 | 36.4 | 51 | 68.5 | 172 |
| 23 | 16 | 62 | 36.2 | 51 | 71.4 | 167 | 48 | 18 | 62 | 36.4 | 61 | 70.5 | 171 |
| 24 | 15 | 63 | 37.1 | 53 | 72.3 | 176 | 49 | 15 | 57 | 38.4 | 39 | 65.5 | 163 |
| 25 | 14 | 57 | 36.2 | 48 | 65.5 | 171 | 50 | 18 | 57 | 37.4 | 40 | 65.6 | 166 |

Table 2 shows the standard error of mean of the correlations of CCA for Gamma and Gaussian distributions respectively. It is discovered that the X and Y-variables of CCA differ. It is also observed that the average standard error of the mean for Gamma distribution is lower than that of Gaussian distribution, but there was need to examine if the difference is significant. Hence, Mann-Whitney U statistic test was employed to justify it.

Table 2: Summary results from the two distributions for different sample sizes

| Sample | CCA           |               | X-Mean Vector |           | Y-Mean Vector |          |
|--------|---------------|---------------|---------------|-----------|---------------|----------|
|        | Gamma         | Gaussian      | Gamma         | Gaussian  | Gamma         | Gaussian |
| 10     | 0.58358       | 0.90370       | 49.23181      | 48.64731  | 17.42602      | 16.29429 |
|        | 0.47279       | 0.43993       | 66.83595      | 69.78251  | 58.69962      | 64.05236 |
|        | 0.07722       | 0.01136       | 165.69121     | 167.28211 | 36.63617      | 36.97472 |
|        | SEM = 0.15369 | SEM = 0.25766 |               |           |               |          |
| 25     | 0.45775       | 0.36372       | 49.80044      | 50.82457  | 15.41576      | 15.53940 |
|        | 0.29920       | 0.27293       | 68.27498      | 69.70341  | 61.00457      | 60.88605 |
|        | 0.16907       | 0.10816       | 167.20255     | 166.94606 | 37.03244      | 36.86259 |
|        | SEM = 0.08347 | SEM = 0.07480 |               |           |               |          |
| 40     | 0.55005       | 0.40179       | 50.70820      | 51.13387  | 16.63939      | 16.47618 |
|        | 0.39847       | 0.14445       | 67.15204      | 68.28058  | 56.58147      | 59.92806 |
|        | 0.17927       | 0.02004       | 166.59783     | 167.22995 | 36.77350      | 36.42717 |
|        | SEM = 0.10763 | SEM = 0.11241 |               |           |               |          |
| 70     | 0.29381       | 0.25335       | 50.88620      | 51.80648  | 15.97843      | 15.14362 |
|        | 0.20112       | 0.13561       | 67.41537      | 67.70621  | 58.82998      | 59.64830 |
|        | 0.01400       | 0.02547       | 166.02249     | 167.58680 | 36.51684      | 36.66154 |
|        | SEM = 0.08229 | SEM = 0.06580 |               |           |               |          |
| 150    | 0.12465       | 0.16596       | 50.96110      | 51.33511  | 15.70896      | 15.85945 |
|        | 0.12010       | 0.13125       | 66.96129      | 67.07074  | 58.69485      | 58.04235 |
|        | 0.04560       | 0.04734       | 166.56148     | 166.60596 | 36.69960      | 36.69399 |
|        | SEM = 0.02563 | SEM = 0.03521 |               |           |               |          |
| 200    | 0.16428       | 0.21459       | 50.57533      | 51.49245  | 15.88936      | 15.75262 |
|        | 0.11587       | 0.16500       | 66.88327      | 66.80400  | 59.20606      | 58.54438 |
|        | 0.02729       | 0.00229       | 166.70745     | 166.54984 | 36.71540      | 36.68410 |
|        | SEM =         | SEM =         |               |           |               |          |

|         |                  |                  |           |           |          |          |
|---------|------------------|------------------|-----------|-----------|----------|----------|
| 350     | 0.04011          | 0.06412          |           |           |          |          |
|         | 0.21845          | 0.17314          | 50.61437  | 50.93154  | 15.87833 | 16.11143 |
|         | 0.07308          | 0.03746          | 66.73694  | 66.88663  | 59.07769 | 58.81425 |
|         | 0.03331          | 0.00292          | 166.79871 | 166.52775 | 36.70022 | 36.68592 |
| 500     | SEM =            | SEM =            |           |           |          |          |
|         | 0.05627          | 0.05195          |           |           |          |          |
|         | 0.09237          | 0.09704          | 50.50598  | 50.70824  | 16.00244 | 15.88627 |
|         | 0.02638          | 0.03667          | 67.14006  | 66.95326  | 58.39705 | 58.62447 |
| 650     | 0.01071          | 0.01537          | 166.56798 | 166.95784 | 36.72705 | 36.66603 |
|         | SEM =            | SEM =            |           |           |          |          |
|         | 0.02502          | 0.02446          |           |           |          |          |
|         | 0.10484          | 0.11210          | 50.35308  | 50.65721  | 15.75088 | 15.82491 |
| 800     | 0.02448          | 0.06286          | 67.19174  | 67.27340  | 58.92530 | 58.65568 |
|         | 0.01763          | 0.01070          | 166.90409 | 166.76525 | 36.69911 | 36.78579 |
|         | SEM =            | SEM =            |           |           |          |          |
|         | 0.02800          | 0.02928          |           |           |          |          |
| 1000    | 0.05524          | 0.08266          | 50.44280  | 50.58706  | 15.89131 | 15.86912 |
|         | 0.03383          | 0.04390          | 67.27835  | 67.30480  | 58.93952 | 59.43417 |
|         | 0.00773          | 0.018331         | 166.64102 | 166.65511 | 36.74218 | 36.74165 |
|         | SEM =            | SEM = 0.01870    |           |           |          |          |
| 1200    | 0.01374          |                  |           |           |          |          |
|         | 0.09175          | 0.08251          | 50.50746  | 50.53873  | 15.87531 | 16.06559 |
|         | 0.06139          | 0.06819          | 67.27207  | 67.39897  | 59.10558 | 58.97092 |
|         | 0.01176          | 0.01718          | 166.66661 | 166.54460 | 36.71932 | 36.66716 |
| 1500    | SEM =            | SEM =            |           |           |          |          |
|         | 0.02331          | 0.01983          |           |           |          |          |
|         | 0.04562          | 0.06331          | 50.50281  | 50.51468  | 15.91433 | 16.01238 |
|         | 0.02992          | 0.02746          | 67.10327  | 67.28566  | 59.03682 | 58.46603 |
| 2100    | 0.00570          | 0.00530          | 166.74046 | 166.76782 | 36.72646 | 36.72821 |
|         | SEM =            | SEM =            |           |           |          |          |
|         | 0.01161          | 0.01690          |           |           |          |          |
|         | 0.08862          | 0.06272          | 50.47693  | 50.67096  | 15.93131 | 15.79109 |
| 2500    | 0.03339          | 0.05413          | 67.20337  | 67.01820  | 59.03893 | 58.97617 |
|         | 0.01812          | 0.00276          | 166.73681 | 166.97834 | 36.72792 | 36.73684 |
|         | SEM =            | SEM =            |           |           |          |          |
|         | 0.02141          | 0.01872          |           |           |          |          |
| 2700    | 0.05135          | 0.06293          | 50.52834  | 50.61113  | 15.94799 | 15.90547 |
|         | 0.03480          | 0.04601          | 67.21437  | 67.39790  | 59.05002 | 59.08607 |
|         | 0.01037          | 0.00210          | 166.76758 | 166.62463 | 36.71682 | 36.72613 |
|         | SEM =            | SEM =            |           |           |          |          |
| 3000    | 0.01190          | 0.01813          |           |           |          |          |
|         | 0.03906          | 0.03944          | 50.50681  | 50.60344  | 15.90875 | 15.91753 |
|         | 0.01733          | 0.01301          | 67.30631  | 67.36842  | 58.93009 | 59.04549 |
|         | 0.00887          | 0.00414          | 166.79732 | 166.66856 | 36.73926 | 36.72802 |
| 3300    | SEM =            | SEM =            |           |           |          |          |
|         | 0.00899          | 0.01060          |           |           |          |          |
|         | 0.04307          | 0.04171          | 50.48835  | 50.58211  | 15.91738 | 15.93182 |
|         | 0.02025          | 0.02764          | 67.36584  | 67.32392  | 58.95503 | 58.99693 |
| 3600    | 0.00622          | 0.02618          | 166.81688 | 166.71481 | 36.75661 | 36.74263 |
|         | SEM =            | SEM =            |           |           |          |          |
|         | 0.01074          | 0.00495          |           |           |          |          |
|         | 0.06049          | 0.04211          | 50.52832  | 50.54902  | 15.90975 | 15.93532 |
| 3900    | 0.03213          | 0.02591          | 67.31721  | 67.28472  | 59.11311 | 58.94190 |
|         | 0.00281          | 0.01648          | 166.80472 | 166.79899 | 36.74773 | 36.74709 |
|         | SEM =            | SEM =            |           |           |          |          |
|         | 0.01665          | 0.00748          |           |           |          |          |
| Average | SEM =<br>0.04238 | SEM =<br>0.04888 |           |           |          |          |

### 3.1. Distribution performance adequacy of correlation values

This section examined if there was a significant difference in the average standard error of mean (SEM) value of 0.04238 for Gamma distribution and the average SEM value of 0.04888 for Gaussian distribution using the Mann-Whitney U statistic test. The ranks of SEM values for Gamma and Gaussian distribution data are displayed in table 3.

Table 3: Summary of decision for testing SEM values for gamma and Gaussian distributed data

|        | SEM Values |          | Ranks |          | Z | Decision               |
|--------|------------|----------|-------|----------|---|------------------------|
| Sample | Gamma      | Gaussian | Gamma | Gaussian |   |                        |
| 10     | 0.15369    | 0.25766  | 33    | 34       | 0 | Do not Reject<br>$H_0$ |
| 25     | 0.08347    | 0.07480  | 30    | 28       |   |                        |
| 40     | 0.10763    | 0.11241  | 31    | 32       |   |                        |
| 70     | 0.08229    | 0.06580  | 29    | 27       |   |                        |
| 150    | 0.02563    | 0.03521  | 19    | 22       |   |                        |
| 200    | 0.04011    | 0.06412  | 23    | 26       |   |                        |
| 350    | 0.05627    | 0.05195  | 25    | 24       |   |                        |
| 500    | 0.02502    | 0.02446  | 18    | 17       |   |                        |
| 650    | 0.02800    | 0.02928  | 20    | 21       |   |                        |
| 800    | 0.01374    | 0.01870  | 8     | 12       |   |                        |
| 1000   | 0.02331    | 0.01983  | 16    | 14       |   |                        |
| 1200   | 0.01161    | 0.01690  | 6     | 10       |   |                        |
| 1500   | 0.02141    | 0.01872  | 15    | 13       |   |                        |
| 2100   | 0.01190    | 0.01813  | 7     | 11       |   |                        |
| 2500   | 0.00899    | 0.01060  | 3     | 4        |   |                        |
| 2700   | 0.01074    | 0.00495  | 5     | 1        |   |                        |
| 3000   | 0.01665    | 0.00748  | 9     | 2        |   |                        |
| Total  |            |          | 297   | 298      |   |                        |

### 3.2. Hypothesis testing for standard error of mean values

In line with the observations in Table 3, the hypotheses are stated

$H_0$ : There is no significant difference between standard error of mean values of CCA for Gamma and Gaussian distributed data.

$H_1$ : There is no significant difference between standard error of mean values of CCA for Gamma and Gaussian distributed data

$$m = n = 17.$$

Using equation (19),

$$\mu = \frac{17(34+1)}{2} = 297.5 \text{ and using equation (20),}$$

$$\sigma^2 = \frac{17(17)(34+1)}{12} = 842.9167, \therefore \sigma = 29.0330.$$

$$R = 297 \text{ or } 298.$$

$$\text{Using equation (18), } Z = \frac{|298 - 297.5| - 0.5}{29.0330} = 0$$

Since  $Z = 0 < Z_{0.025} = 1.96$ , the test concludes that there is no significant difference between correlation values of CCA for Gamma and Gaussian distributed data.

#### 4. DISCUSSION OF THE FINDINGS

This study examined CCA via Gaussian and Gamma distributed data in order to assess their performances. A real-life medical data were used to simulate data of sample sizes of 10, 25, 40, 70, 150, 200, 250, 500, 650, 800, 1000, 1200, 1500, 2100, 2500, 2700 and 3000 for two continuous distributions. A computer programming language codes was written in RStudio for simulation and solution of CCA via Gaussian and Gamma data distribution. The result of the study revealed that the SEM of correlation values for Gamma distribution was 0.04238, which was less than that of SEM of correlation values of 0.04888 for Gaussian distribution. Further analysis indicated that the difference in the standard error of mean of correlation values between the distributions was insignificant using the Mann-Whitney U test statistic. The study thereby concluded that for CCA, the relative efficiency of the Gaussian and Gamma distributed data is the same. This result is similar to the findings of Onyeagu *et al.* (2014) whose result found out that the relative efficiency of the CCA and generalized canonical correlation analysis (GCCA) was the same for the lognormal and Cauchy distributed data since the standard deviation of the correlations and eigenvalues were the same for GCCA and CCA.

#### 5. CONCLUSION

Using the Gaussian and Gamma distributed data for canonical correlation analysis with simulated data of sample sizes of 10, 25, 40, 70, 150, 200, 250, 500, 650, 800, 1000, 1200, 1500, 2100, 2500, 2700 and 3000 from a real-life medical data, the study concluded that for CCA, the relative efficiency of the Gaussian and Gamma distributed data is the same.

#### 6. RECOMMENDATIONS

Having examined the performance of canonical correlation analysis via Gaussian and Gamma distributed data with simulated data of sample sizes of 10, 25, 40, 70, 150, 200, 250, 500, 650, 800, 1000, 1200, 1500, 2100, 2500, 2700 and 3000 from real life medical data, further studies should be done on the following:

1. Investigating the performance of CCA and nonparametric CCA using at least six continuous distributions.
2. In this study, RStudio was employed to write a programming code to tackle the numerical problems posed, but the time of execution was not incorporated. Hence, further works, irrespective of any programming language package to be employed should ensure that the CPU time of execution is added, as it will contribute in performance judgment.
3. In this study, multicollinearity and homoscedasticity assumptions were not diagnosed for the simulated data. Hence, further studies should examine it.

## 7. REFERENCES

- Affi, P. O. 2019, *Canonical correlation analysis of neonatal anthropometric indicators and maternal socio-demographic factors*. *Int. J. Sci. Technol. Res.* **8**, 93–97.
- Akano, R., Adams, S. 2019, *Nigeria government's expenditure on economic and social service development: a canonical correlation analysis approach*. *Arch. Curr. Res.* **18**, 1-12.
- Ekezie, D. D., Opara, J., Mbachu, H. I. 2013, Multivariate analysis of chemical components of tobacco leaves. *Int. J. Emerg. Knowl.* **1**, 69-89.
- 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**, 1-24.
- Gibbons, J. 2003, *Nonparametric statistical inference (4th ed.)*. Marcel Dekker, Inc. New York, 34-39.
- Góreck, T., Krzysko, M., Wołynski, W., 2020, Generalized canonical correlation analysis for functional data. *Biomed. Lett.* **57**, 1-12.
- Iheagwara, A. I., Okenwe, I., Ononogbu, A. C. 2018, *Canonical correlation analysis on vital signs and demographic measures of patients*. *Am. J. Comput. Sci. Eng.* **5**, 121-131.
- Makino, N. 2022, *Rotation in correspondence analysis from the canonical correlation perspective*. *Psychometrika*, **5**, 18-28.
- Onyeagu, S. I., Osuji, G. A., Onyia, O. M. 2014, *Comparison of canonical correlation analysis and the generalized canonical correlation analysis using the lognormal and Cauchy distributed data*. *Math. Theory Model.* **4**, 177-185.
- Sakamotoa, K., Okabea, M., Yadoshisa, H. 2021, *Generalized canonical correlation analysis for labeled data*. *Procedia Comput. Sci.*, **192**, 517-525.
- 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 Explor. Exploit.* **39**, 2214-2239.