American Journal of Computer Science and Engineering
2018; 5(5): 121-131
http://www.openscienceonline.com/journal/ajcse

Canonical Correlation Analysis on Vital Signs and Demographic Measures of Patients

Iheagwara Andrew Ihuoma[1, *], Okenwe Idochi[2], Ononogbu Amarachi Christiana[3]

[1]Department of Statistics, Imo State University, Owerri, Nigeria

[2]Department of Statistics, Ken Saro Wiwa Polytechnic, Bori, Rivers, Nigeria

[3]Department of Statistics, Michael Okpara University of Agriculture Umudike, Umuahia, Nigeria

Email address

andyiheagwara@yahoo.com (I. A. Ihuoma)
[*]Corresponding author

To cite this article

Iheagwara Andrew Ihuoma, Okenwe Idochi, Ononogbu Amarachi Christiana. Canonical Correlation Analysis on Vital Signs and Demographic Measures of Patients. *American Journal of Computer Science and Engineering.* Vol. 5, No. 5, 2018, pp. 121-131.

Received: July 29, 2018; **Accepted**: August 23, 2018; **Published**: October 10, 2018

Abstract

This study examined the canonical correlation analysis on vital signs and demographic measures of 200 patients in the selected government owned hospital in Imo State Nigeria. Canonical correlation analysis focuses on the correlation between a linear combination of the variables in one set and a linear combination of the variables in another set. The objective of the study is to identify and quantify the association between two sets of variables. The data for this study were collected in an arrangement with the nurses of the selected government owned hospital in Imo State Nigeria. A total number of 200 patients were randomly selected and used for this study. An ethical approval was obtained from the Ethical committee of the selected hospitals. The consent of the subjects were sought and obtained and the consenting subjects were used by the patients, using the following inclusion criteria: They were normally balanced (i.e. the subjects were able to answer questions on name, age, sex, date of birth, marital status etc., consciously and correctly; All the patients (both in- and out-patients) within April 1, to April 18 2018 participated in this exercise, as the nurses did not allow them to know the rationale behind the vital signs and demographic measurements; Both male and female were recruited. The data for this study, which contain four criterion measures and three predictor variables, were analyzed using the "Stata" statistical software package. Based on the results obtained, and the hypotheses carried out, it was revealed that out of the three sample canonical correlations, the first two are significant, while the third one is insignificant. Finally, it was concluded that relationship exists between the vital signs and the demographic measures of the 200 patients.

Keywords

Canonical Correlations, Standardized Coefficients, Canonical Loadings, Correlation Matrix

1. Introduction

In many research settings, the social scientist encounters a phenomenon that is best described not in terms of a single criterion but, because of its complexity, in terms of a number of response measures [1]. In such cases, interest may center on the relationship between the set of criterion measures and the set of explanatory factors. In a manufacturing process, for instance, we might be concerned with the relationship between a set of organic chemical constituent variables, on the one hand, and various inorganic chemical constituent variables on the other hand, as it is applicable in this paper. In the business or economic fields, we might be interested in the relationship between a set of price indices and a set of production indices, with a view towards (say) predicting one from the other. The study of the relationship between a set of predictor variables and a set of response measures is known as canonical correlation analysis.

With a growing number of large scales genomic data the focus these days have been in finding the relationship between two or more sets of variables. One of the classical methods that can be used in cases when we have two set of variables from the same subject is Canonical Correlation Analysis (CCA) but it lacks biological interpretation for situations in which each set of variables has more than thousands of variables. This issue was first addressed by Parkhomenko et al [2] who proposed a novel method for Sparse Canonical Correlation Analysis (SCCA).

Canonical correlation analysis seeks to identify and quantify the associations between two sets of variables [3]. Canonical correlation analysis focuses on the correlation between a linear combination of the variables in one set and a linear combination of the variables in another set. The idea is first to determine the pair of linear combinations having the largest correlation. Next, we determine the pair of linear combinations having the largest correlation among all parts uncorrelated with the initially selected pair. The process continues. The pairs of linear combinations are called the canonical variables, and their correlations are called canonical correlations.

2. Literature Review

Lin et al [4] carried out a research on Real Time Electroencephalogram (EEG) Signal Enhancement Using Canonical Correlation Analysis and Gaussian Mixture Clustering. The study proposed a real-time artefact removal algorithm that was based on canonical correlation analysis (CCA), feature extraction, and the Gaussian Mixture Model (GMM) to improve the quality of EEG signals. The CCA was used to decompose EEG signals into components followed by feature extraction to extract representative features and GMM to cluster these features into groups to recognize and remove artefacts. The feasibility of the proposed algorithm was demonstrated by effectively removing artefacts caused by blinks, head/body movement, and chewing from EEG recordings while preserving the temporal and spectral characteristics of the signals that were important to cognitive research. The findings revealed that common artefacts in EEG were successfully suppressed by the proposed artefact removal algorithm. EEG signals appeared much cleaner after artefact removal than before it.

Jayadevan [5] worked on a Canonical Correlation Analysis of Sector-al Composition of GDP and Development in Asia. The study identified the factors that influence percentage contribution of sectors to Gross Domestic Product (GDP) for a group of 32 Asian countries for two cross-section points 1994-96 and 2014-16. The study employed canonical correlation analysis for 32 Asian countries. The analysis revealed that the structural changes in sector-al GDP composition in the selected Asian countries were significantly determined by the factors like employee productivity, employment growth in services sector, rising life expectancy, growth of value added in manufacturing and gross capital formation.

Mussina and Bachisse [6] researched on Canonical Correlation Analysis between Business Sophistication and Macroeconomic Environment: A Secondary study of Countries Global competitiveness. The purpose of the study was to provide information for improving national competitiveness in different countries around the world through identifying the most impactful indexes affecting its level. The sample size of 102 countries out of 138, omitting countries at two transition stages was analyzed. Canonical correlation analysis was employed to investigate the interaction between the two pillars. The findings of the study showed that there was a significant and positive relationship between the set of "Business Sophistication" and "Macroeconomic Environment".

Idowu and Titilola [7] did a work on Development of a canonical correlation model involving non linearity and asymmetric variables. The study formulated a new model which generalized CCA to non linear associations and asymmetric distributions. Special cases of the proposed model were discussed. The behavior of canonical solutions under varying mixtures of skewness and non linearity (NL) was also examined in a simulation study. In addition, these solutions were compared with some commonly used methods of Hotelling, Spearman, and Kendall. The findings of the study revealed, among others, that for a fixed level of NL, the canonical correlation (ρ) increased as skewness increased. Whether by ρ, likelihood, Akaike information criterion and Bayesian information criterion, the proposed method performed better than the other methods in all degrees of skewness and NL considered. It was further confirmed with real-life data application as Hotelling, Spearman, and Kendall overestimated ρ by 2.08%, 37.81%, and 22.15%, respectively, compared to the proposed technique.

3. Materials and Methods

The method of analysis used in this study is the Canonical Correlation Analysis. This paper shall focus on how to analyze 200 patients of vital signs and demographic measures in government owned hospitals in Imo State Nigeria via Stata Statistical Software Package.

3.1. Canonical Variates and Canonical Correlations

The canonical correlations measure the strength of association between the two sets of variables. The maximization aspect of the technique represents an attempt to concentrate a high-dimensional relationship between two sets of variables into a few pairs of canonical variables. This study shall focus on measures of association between two groups of variables. The first group of p variables is represented by the (p × 1) random vector $X^{(1)}$, while the second group of q variables is represented by the (q × 1) random vector $X^{(2)}$. It will be assumed, in the theoretical development, that $X^{(1)}$ represents the smaller set, so that $p \leq q$.

For the random vectors $X^{(1)}$ and $X^{(2)}$, let

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

It will be convenient to consider $X^{(1)}$ and $X^{(2)}$ 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

$$\begin{aligned} \underset{(p+q)\times(p+q)}{\boldsymbol{\Sigma}} &= \mathrm{E}(\mathbf{X}-\mu)\mathrm{E}(\mathbf{X}-\mu)' \\ &= \left[\begin{array}{c:c} E(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})' \\ \hdashline E(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})' \end{array}\right] \\ & \left[\begin{array}{c:c} \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}} & \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}} \\ \hdashline \underset{(q\times p)}{\boldsymbol{\Sigma}_{21}} & \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}} \end{array}\right] \end{aligned} \tag{4}$$

The co-variances between pairs of variables from different sets – one variable from $X^{(1)}$, one variable from $X^{(2)}$ – are contained in $\sum_{12}$ or, equivalent, in $\sum_{12}$. That is, the pq elements of $\sum_{12}$ measure the association between the two sets. When p and q are relatively large, interpreting the elements of $\sum_{12}$ collectively is ordinarily hopeless [8]. Moreover, it is often linear combinations of variables that are interesting and useful predictive or comparative purposes. The main task of canonical correlation analysis is to summarize the associations between the $X^{(1)}$ and $X^{(2)}$ sets in terms of a few carefully chosen covariance (or correlations) rather than the pq covariance in $\sum_{12}$.

Linear combinations Z = CX have

$$\left.\begin{aligned} \boldsymbol{\mu}_z &= E(\mathbf{Z}) = E(\mathbf{CX}) = \mathbf{C}\boldsymbol{\mu}_X \\ \Sigma_Z &= Cov(\mathbf{Z}) = Cov(\mathbf{CX}) = \mathbf{C}\Sigma_X C' \end{aligned}\right\} \tag{5}$$

and provide simple summary measures of a set of variables.

$$\text{Let} \quad \begin{aligned} \mathbf{U} &= \mathbf{a}'\mathbf{X}^{(1)} \\ \mathbf{V} &= \mathbf{b}'\mathbf{X}^{(2)} \end{aligned} \tag{6}$$

for some pair of coefficient vectors a and b. Using Equations (5) and (6),

$$\left.\begin{aligned} Var(\mathbf{U}) &= \mathbf{a}'Cov(\mathbf{X}^{(1)})\mathbf{a} = \mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a} \\ Var(\mathbf{V}) &= \mathbf{b}'Cov(\mathbf{X}^{(1)})\mathbf{b} = \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b} \\ Cov(\mathbf{U},\mathbf{V}) &= \mathbf{a}'Cov(\mathbf{X}^{(1)},\mathbf{X}^{(2)})\mathbf{b} = \mathbf{a}'\Sigma_{12}\mathbf{b} \end{aligned}\right\} \tag{7}$$

Let us seek coefficient vectors a and b such that

$$Corr(\mathbf{U},\mathbf{V}) = \frac{\mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a}}\sqrt{\mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}}} \tag{8}$$

is as large as possible. We then define:

The first pair of canonical variables are the pair of linear combinations U_1, V_1 having unit variances, which maximize the correlation in Equation (8);

The second pair of canonical variables are the linear combinations U_2, V_2 having unit variances, which maximize the correlation in Equation (8) among all choices which are uncorrelated with the first pair of canonical variables.

At the kth step:

The kth pair of canonical variables are the linear combinations U_k, V_k having unit variances, which maximize the correlation in Equation (8) among all choices uncorrelated with the previous k – 1 canonical variable pairs.

The correlation between the kth pair of canonical variables is called the kth canonical correlation. If the original variables are standardized with $Z^{(1)} = \left[\mathbf{Z}_1^{(1)}, \mathbf{Z}_2^{(2)}, \ldots, \mathbf{Z}_p^{(1)}\right]'$ and $\mathbf{Z}^{(2)} = \left[\mathbf{Z}_1^{(1)}, \mathbf{Z}_2^{(2)}, \ldots, \mathbf{Z}_q^{(1)}\right]'$ from first principles, the canonical varaites are of the form

$$\left.\begin{aligned} \mathbf{U}_k &= \mathbf{a}_k'\mathbf{Z}^{(1)} = \mathbf{e}_k'\boldsymbol{\rho}_{11}^{-1/2}\mathbf{Z}^{(1)} \\ \mathbf{V}_k &= \mathbf{b}_k'\mathbf{Z}^{(2)} = \mathbf{f}_k'\boldsymbol{\rho}_{22}^{-1/2}\mathbf{Z}^{(2)} \end{aligned}\right\} \tag{9}$$

Here $Cov(Z^{(1)}) = \rho_{11}$, $Cov(Z^{(2)}) = \rho_{22}$, $Cov(Z^{(1)}, Z^{(2)}) = \rho_{12} = \rho_{21}$ and e_k and f_k are the eigenvectors of $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ and $\boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{22}\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1/2}$, respectively. The canonical correlations, ρ_k^*, satisfy

$$Corr(\mathbf{U}_k,\mathbf{V}_k) = \boldsymbol{\rho}_k^*,\ \mathrm{k} = 1, 2, \ldots, \mathrm{p} \tag{10}$$

where $\rho_1^{*2} \geq \rho_2^{*2} \geq \ldots \geq \rho_p^{*2}$ are the nonzero eigenvalues of the matrix $\boldsymbol{\rho}_{11}^{-1/2}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1}\boldsymbol{\rho}_{21}\boldsymbol{\rho}_{11}^{-1/2}$ (or equivalently, of $\boldsymbol{\rho}_{22}^{-1/2}\boldsymbol{\rho}_{22}\boldsymbol{\rho}_{11}^{-1}\boldsymbol{\rho}_{12}\boldsymbol{\rho}_{22}^{-1/2}$).

It should be noted that:

$$\mathrm{a}_k'(\mathrm{X}^{(1)} - \mu^{(1)}) = \mathrm{a}_{k1}(\mathrm{X}_1^{(1)} - \mu_1^{(1)}) + \mathrm{a}_{k2}(\mathrm{X}_2^{(1)} - \mu_2^{(1)}) + \cdots + \mathrm{a}_{kp}(\mathrm{X}_p^{(1)} - \mu_p^{(1)})$$

$$= a_{k1}\sqrt{\sigma_{11}}\frac{(X_1^{(1)}-\mu_1^{(1)})}{\sqrt{\sigma_{11}}} + a_{k2}\sqrt{\sigma_{22}}\frac{(X_2^{(1)}-\mu_2^{(1)})}{\sqrt{\sigma_{22}}} + \ldots + a_{kp}\sqrt{\sigma_{pp}}\frac{(X_p^{(1)}-\mu_p^{(1)})}{\sqrt{\sigma_{pp}}}$$

where $\mathrm{Var}\left(X_i^{(1)}\right) = \sigma_{ii}$, i = 1, 2, …, p. Therefore, the canonical coefficients for the standardized variables, $Z_i^{(1)} = (X_i^{(1)} - \mu_i^{(1)})/\sqrt{\sigma_{ii}}$, are simply related to the canonical coefficients attached to the original variables $X_i^{(1)}$. Specifically, if a'_k is the coefficient vector for the kth canonical variate U_k, then $a'_k V_{11}^{1/2}$ is the coefficient vector for the canonical variate constructed from the standardized variables $Z^{(1)}$. Here $V_{11}^{1/2}$ is the diagonal matrix with ith diagonal element $\sqrt{\sigma_{ii}}$. Similarly, $b'_k V_{22}^{1/2}$ is the coefficient vector for the canonical varaite constructed from the set of standardized variables $Z^{(2)}$. In this case $V_{22}^{1/2}$ is the diagonal matrix with ith diagonal element $\sqrt{\sigma_{ii}} = \sqrt{\mathrm{Var}(X_i^{(2)})}$. The canonical correlations are unchanged by the standardization. However, the choice of the coefficient vectors a_k, b_k will not be unique if $\rho_k^* = \rho_{k+1}^{*2}$.

3.2. Identifying the Canonical Variables

Even though the canonical variables are artificial, they can often be "identified" in terms of the subject matter variables [9]. This identification is often aided by computing the correlations between the canonical variates and the original variables. These correlations, however, must be interpreted with caution. They only provide univariate information in the sense that they do not indicate how the original variables contribute jointly to the canonical analyses [10]. For this reason, many investigators prefer to assess the contributions of the original variables directly from the standardized coefficients in Equation (9).

Let $\underset{(p\times p)}{\mathbf{A}} = [a_1, a_2, \ldots, a_p]'$ and $\underset{(q\times q)}{\mathbf{B}} = [b_1, b_2, \ldots, b_q]'$, so that the vectors of canonical variables are

$$\underset{(p\times 1)}{\mathbf{U}} = \mathbf{AX}^{(1)}, \quad \underset{(q\times 1)}{\mathbf{V}} = \mathbf{BX}^{(2)}, \tag{11}$$

where we are primarily interested in the first p canonical variables in V. Then

$$\mathrm{Cov}(U, X^{(1)}) = \mathrm{Cov}(AX^{(1)}, X^{(1)}) = A\Sigma_{11} \tag{12}$$

Because $\mathrm{Var}(U_i) = 1$, $\mathrm{Corr}(U_i, X_k^{(1)})$ is obtained by dividing $\mathrm{Cov}(U_i, X_k^{(1)})$ by $\sqrt{\mathrm{var}(X_k^{(1)})} = \sigma_{kk}^{1/2}$. Equivalently, $\mathrm{Corr}(U_i, X_k^{(1)}) = \mathrm{Cov}(U_i, \sigma_{kk}^{1/2} X_k^{(1)})$. Introducing the (p × p) diagonal matrix $V_{kk}^{-1/2}$ with kth diagonal element $\sigma_{kk}^{-1/2}$, we have, in matrix terms,

$$\underset{(p\times p)}{\rho_{U,X^{(1)}}} = \mathrm{Corr}(U, X^{(1)}) = \mathrm{Cov}(U, V_{11}^{-1/2} X^{(1)}) = \mathrm{Cov}(AX^{(1)}, V_{11}^{-1/2} X^{(1)})$$

$$A\Sigma_{11} V_{11}^{-1/2}$$

Similar calculations for the pairs $(U, X^{(2)})$, $(V, X^{(2)})$ and $(V, X^{(1)})$ yield

$$\left.\begin{array}{ll} \underset{(p\times p)}{\rho_{U,X^{(1)}}} = A\Sigma_{11}V_{11}^{-1/2} & \underset{(q\times q)}{\rho_{V,X^{(2)}}} = A\Sigma_{22}V_{22}^{-1/2} \\ \underset{(p\times q)}{\rho_{U,X^{(1)}}} = A\Sigma_{12}V_{22}^{-1/2} & \underset{(q\times p)}{\rho_{V,X^{(2)}}} = A\Sigma_{21}V_{11}^{-1/2} \end{array}\right\} \tag{13}$$

where $V_{22}^{-1/2}$ is the (q × q) diagonal matrix with ith diagonal element $\sqrt{\mathrm{var}(X_i^{(2)})}$.

Canonical variables derived from standardized variables are sometimes interpreted by computing the correlations.

$$\left.\begin{array}{ll} \rho_{\mathbf{U},\mathbf{Z}^{(1)}} = \mathbf{A}_z\rho_{11}; & \rho_{V,Z^{(2)}} = \mathbf{B}_z\rho_{22} \\ \rho_{\mathbf{U},\mathbf{Z}^{(2)}} = \mathbf{A}_z\rho_{12}; & \rho_{V,Z^{(1)}} = \mathbf{B}_z\rho_{21} \end{array}\right\} \tag{14}$$

where $\underset{(p\times p)}{A_z}$ and $\underset{(q\times q)}{B_z}$ are the matrices whose rows contain the canonical coefficients for the $Z^{(1)}$ and $Z^{(2)}$ sets, respectively [11]. The correlations in the matrices displayed is Equation (14) have the same numerical values as those appearing in Equation (13), that is, $\rho_{U,X^{(1)}} = \rho_{V,Z^{(1)}}$ and so forth [12]. This follows because, for example,

$\rho_{U,X^{(1)}} = A\Sigma_{11}V_{11}^{-1/2} = AV_{11}^{-1/2}V_{11}^{-1/2}\Sigma_{11}V_{11}^{-1/2} = A_z\rho_{11} = \rho_{U,Z^{(1)}}$. The correlations are unaffected by the standardization [13].

3.3. The Sample Canonical Variates and Sample Canonical Correlations

A random sample of n observations on each of the (p + q) variables $X^{(1)}$, $X^{(2)}$ can be assembled into the ((p + q) × n) data matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \hdashline \mathbf{X}^{(2)} \end{bmatrix} = \begin{bmatrix} x_{11}^{(1)} & x_{12}^{(1)} & \cdots & x_{1n}^{(1)} \\ x_{21}^{(1)} & x_{22}^{(1)} & \cdots & x_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1}^{(1)} & x_{p2}^{(1)} & \cdots & x_{pn}^{(1)} \\ \hdashline x_{11}^{(2)} & x_{12}^{(2)} & \cdots & x_{1n}^{(2)} \\ x_{21}^{(2)} & x_{22}^{(2)} & \cdots & x_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{q1}^{(2)} & x_{q2}^{(2)} & \cdots & x_{qn}^{(2)} \end{bmatrix} = [x_1, x_2, \ldots, x_n] \text{where } x_j = \begin{bmatrix} x_j^{(1)} \\ x_j^{(2)} \end{bmatrix} \tag{15}$$

The vector of sample means can be organized as

$$\underset{(p+q)\times 1}{\bar{x}} = \begin{bmatrix} \bar{x}^{(1)} \\ \hdashline \bar{x}^{(2)} \end{bmatrix} \text{ where } \left.\begin{aligned} \bar{x}^{(1)} &= \frac{1}{n}\sum_{j=1}^{n} x_j^{(1)} \\ \bar{x}^{(2)} &= \frac{1}{n}\sum_{j=1}^{n} x_j^{(2)} \end{aligned}\right\} \tag{16}$$

Similarly, the sample covariance matrix can be arranged analogous to the representation in Equation (4). Thus

$$\underset{(p+q)(p+q)}{\mathbf{S}} = \left[\begin{array}{c:c} \underset{(p\times p)}{\mathbf{S}_{11}} & \underset{(p\times q)}{\mathbf{S}_{12}} \\ \hdashline \underset{(q\times p)}{\mathbf{S}_{21}} & \underset{(q\times q)}{\mathbf{S}_{22}} \end{array}\right]$$

where

$$\mathbf{S}_{kl} = \frac{1}{n-1}\sum_{j=1}^{n}\left(x_j^{(k)} - \bar{x}^{(k)}\right)\left(X_j^{(l)} - \bar{x}^{(l)}\right)' \quad k, l = 1, 2, \tag{17}$$

The linear combinations

$$\hat{U} = \hat{a}'x^{(1)}; \quad \hat{V} = \hat{b}'x^{(2)} \tag{18}$$

have sample correlation

$$r_{\hat{U},\hat{V}} = \frac{\hat{\mathbf{a}}'\mathbf{S}_{12}\hat{\mathbf{b}}}{\sqrt{\hat{\mathbf{a}}'\mathbf{S}_{11}\hat{\mathbf{a}}}\sqrt{\hat{\mathbf{b}}'\mathbf{S}_{22}\hat{\mathbf{b}}}} \tag{19}$$

The first pair of sample canonical variates is the pair of linear combinations $\hat{U}_1, \hat{V}_1$ having unit sample variances that maximize the ration in Equation (19).

In general: the kth pair of sample canonical variates is the pair of linear combinations $\hat{U}_k, \hat{V}_k$ having unit sample variances that maximize the ratio in Equation (19) among those linear combinations uncorrelated with the previous k – 1 sample canonical variates. The sample correlation between $\hat{U}_k$ and $\hat{V}_k$ is called the kth sample canonical correlation [14].

3.4. Data Collection

The data for this study were collected in an arrangement with the nurses of the selected government owned hospital in Imo State Nigeria. A total number of 200 patients were randomly selected and used for this study. An ethical approval was obtained from the Ethical committee of the selected hospitals. The consent of the subjects were sought and obtained and the consenting subjects were used by the patients, using the following inclusion criteria:

i. They were normally balanced (i.e. the subjects were able to answer questions on name, age, sex, date of birth, marital status etc., consciously and correctly.
ii. All the patients (both in- and out-patients) within April 1, to April 18 2018 participated in this exercise, as the nurses did not allow them to know the rationale behind the vital signs and demographic measurements.
iii. Both male and female were recruited.

3.5. Sampling Procedure

No sampling procedure was involved in taking the vital signs measurements, as it involves all the patients within the time of study. For the purpose of this study, a simple random sampling of 200 patients was used for this study, out of the patients visited in the selected hospitals. Table 1 shows the data collected for this study.

The dependent variables considered are defined as follows:

Y_1: Pulse Rate

Y_2: Systolic Blood Pressure
Y_3: Body Temperature
Y_4: Respiration Rate
The independent variables are defined as follows.
X_1: Age
X_2: Weight
X_3: Sex

Table 1 shows the four dependent variables (Vital signs) and three independent variables (demographic measures) of 200 patients chosen from the selected government owned hospital in Imo State Nigeria.

Table 1. *Data for Vital Signs and Demographic Measures.*

S/N	Dependent variables				Independent variables			S/N	Dependent variables				Independent variables		
	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3	S/N	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3
1	65	145	35.6	12	25	95	1	101	73	163	35.8	16	34	95	0
2	68	134	37.5	15	19	69	1	102	77	146	35.7	19	43	69	0
3	77	153	35.6	14	45	77	1	103	76	184	37.9	10	23	77	1
4	76	145	36.1	13	54	53	1	104	87	196	36.8	17	43	53	1
5	89	154	35.5	16	34	66	1	105	67	185	35.9	19	23	66	1
6	58	137	38.5	16	54	53	0	106	76	195	36.9	18	42	53	1
7	65	146	37.5	19	23	70	0	107	78	163	36.5	15	23	70	0
8	70	147	35.7	10	45	60	1	108	82	174	35.9	10	23	60	0
9	57	154	36.7	17	43	63	0	109	69	186	35.9	18	43	63	0
10	65	164	35.8	19	23	54	0	110	86	153	36.8	19	53	54	0
11	76	165	36.9	18	54	64	0	111	84	174	35.7	9	32	64	0
12	84	156	36.5	15	24	85	1	112	75	175	35.8	18	45	85	0
13	68	145	37.6	10	54	65	1	113	67	157	36.1	17	42	65	0
14	65	134	35.7	18	34	50	1	114	87	174	36.9	21	43	50	0
15	76	209	36.5	19	54	62	1	115	67	154	35	22	23	62	0
16	45	156	36.3	9	34	60	1	116	66	185	37.1	19	42	60	1
17	67	122	36.9	18	54	68	0	117	76	164	36.1	16	43	68	1
18	87	118	36.5	17	34	60	0	118	58	174	38.9	17	45	60	0
19	67	123	35.8	21	43	65	0	119	68	153	36.7	14	23	65	1
20	89	152	37.8	22	43	65	0	120	69	184	36.9	15	43	65	1
21	59	123	35.4	19	43	62	0	121	66	198	35.8	16	23	62	0
22	87	123	36.2	16	54	65	1	122	86	165	38.8	14	26	65	0
23	56	132	35.8	17	34	70	1	123	76	175	36.3	18	36	70	1
24	67	123	39.1	14	43	70	1	124	74	164	35.2	19	34	70	1
25	87	145	36.2	15	23	105	1	125	63	186	37.4	20	32	105	0
26	80	125	37.1	16	65	70	0	126	69	154	35.8	10	24	70	0
27	67	154	36.2	14	45	65	0	127	64	167	37.9	18	35	65	1
28	87	156	36	18	54	105	0	128	66	174	35.8	17	42	105	0
29	67	134	38.6	19	65	65	0	129	73	172	35.8	16	28	65	1
30	98	153	35.7	20	45	65	1	130	77	154	37.6	15	35	65	0
31	60	135	38.1	10	54	88	1	131	76	175	35.9	16	46	88	1
32	69	143	38.6	18	65	65	1	132	87	164	37.7	18	35	65	1
33	75	143	35.8	17	45	100	0	133	67	145	39.7	14	43	100	1
34	76	163	36.8	16	63	40	0	134	76	190	35.8	17	35	40	0
35	78	184	36.2	15	67	80	0	135	78	153	35.8	9	35	80	0
36	56	143	36.4	16	47	70	0	136	58	134	35.9	18	24	70	0
37	86	185	37.1	18	63	50	1	137	76	163	35.8	16	43	50	1
38	65	146	35.8	14	45	80	1	138	87	185	36.8	19	28	80	1
39	69	174	36	17	45	65	1	139	56	145	35.9	14	36	65	1
40	76	145	36.4	9	54	55	0	140	78	135	36.6	16	38	55	1
41	56	185	37.4	18	64	60	0	141	74	143	36.9	13	45	60	0
42	76	153	37.1	16	54	60	0	142	79	174	35.9	18	43	60	0
43	71	175	36.4	19	46	50	0	143	56	185	35.8	16	53	50	1
44	69	185	37.1	14	54	70	0	144	78	145	36.8	15	45	70	1
45	58	146	36.5	16	45	65	0	145	79	155	36.1	18	54	65	1
46	68	174	36.4	13	64	60	1	146	67	185	37.8	14	35	60	0
47	66	145	36.5	18	54	55	1	147	87	145	37.6	13	45	55	0
48	77	176	35.6	16	43	65	1	148	56	143	36	12	34	65	0
49	87	146	36.2	15	34	60	1	149	89	185	36.9	19	43	60	1

S/N	Dependent variables				Independent variables			S/N	Dependent variables				Independent variables		
	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3	S/N	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3
50	75	164	36.4	18	24	65	0	150	76	145	36.5	16	54	65	1
51	67	195	37.4	14	43	95	1	151	87	175	36.8	17	43	95	1
52	66	157	36.4	13	46	69	1	152	67	164	35.2	15	43	69	0
53	60	186	36.4	12	33	77	0	153	89	174	36.9	17	35	77	0
54	67	156	39.1	19	56	53	0	154	76	146	37.4	18	43	53	0
55	65	186	36.2	16	43	66	0	155	68	145	36.7	12	54	66	1
56	78	175	36.4	17	34	53	0	156	76	186	38.6	15	34	53	1
57	77	175	36.4	15	54	70	0	157	65	146	38.9	14	24	70	0
58	89	153	37.4	17	45	60	1	158	87	165	35.9	13	43	60	0
59	76	169	38.4	18	56	63	1	159	66	185	36.8	16	54	63	1
60	67	163	37.4	13	63	54	1	160	56	185	36.8	16	35	54	1
61	87	164	37.4	16	45	64	1	161	76	134	37.8	19	43	64	1
62	66	186	35.8	19	55	85	1	162	86	154	37.5	10	54	85	0
63	63	165	37.5	17	34	65	0	163	56	186	36.8	17	35	65	0
64	78	146	36.6	19	54	50	0	164	78	145	35.9	19	35	50	1
65	67	154	36.8	13	34	62	0	165	87	166	35.8	18	43	62	1
66	65	194	37.2	12	54	60	0	166	68	176	38.6	15	35	60	1
67	67	146	36.4	17	45	68	0	167	64	165	35.9	10	43	68	0
68	58	185	36.4	18	53	60	0	168	66	164	36.3	18	25	60	0
69	58	164	38.4	15	54	65	1	169	73	186	37.7	19	54	65	0
70	86	186	37.4	18	28	65	1	170	77	145	39.6	9	46	65	0
71	78	145	36.5	15	46	62	0	171	76	209	35.9	18	54	62	0
72	86	174	37.4	17	64	65	0	172	87	185	37.6	17	54	65	1
73	67	187	36.5	14	54	70	0	173	67	134	36.2	21	45	70	0
74	77	175	37.8	16	53	70	0	174	76	146	37.5	22	45	70	0
75	98	153	36.5	18	45	105	1	175	78	174	35.8	19	46	105	1
76	56	174	38.5	14	65	70	1	176	82	135	36.8	16	54	70	1
77	78	185	36.6	13	56	65	1	177	69	165	37.7	17	24	65	0
78	87	184	36.7	16	45	105	1	178	86	158	35.2	14	42	105	0
79	88	153	38.5	18	64	65	0	179	84	165	37.6	15	34	65	0
80	85	163	37.5	15	54	65	1	180	75	154	35.8	16	36	65	1
81	76	184	38.5	14	75	88	1	181	67	186	35.8	14	37	88	0
82	88	143	36.6	17	45	65	1	182	87	163	38.2	18	27	65	0
83	67	164	38.5	14	76	100	0	183	67	156	37.4	19	46	100	0
84	78	158	36.4	17	55	40	0	184	69	185	37.4	20	33	40	0
85	76	175	36.1	15	45	80	0	185	78	174	36.8	10	46	80	1
86	56	184	36.2	18	64	70	0	186	87	185	36.9	18	46	70	0
87	65	196	38.2	19	56	50	1	187	98	124	35.5	17	46	50	0
88	66	155	38.4	15	64	80	1	188	76	203	37.9	16	35	80	0
89	76	156	36.4	15	24	65	1	189	78	126	36.4	15	34	65	0
90	58	201	35.5	14	19	55	1	190	58	185	37.5	16	54	55	1
91	68	189	36.6	14	43	60	0	191	78	132	36.9	18	35	60	1
92	69	136	38.4	13	34	60	0	192	98	167	36.2	14	46	60	0
93	66	96	36.5	19	34	50	0	193	87	124	37.9	17	54	50	0
94	86	157	37.7	17	45	70	0	194	67	164	35.9	9	35	70	1
95	76	185	35.6	18	43	65	1	195	78	142	35.8	18	54	65	1
96	74	134	35	12	23	60	1	196	70	174	35.9	16	24	60	0
97	63	136	36.4	15	53	55	1	197	55	163	35.8	19	54	55	0
98	69	183	36.1	14	34	65	1	198	58	165	37.8	14	32	65	0
99	64	163	35.9	13	24	60	0	199	67	143	35.8	16	36	60	1
100	66	129	36.8	16	53	65	0	200	69	174	36.9	13	34	65	1

4. Data Analysis

The vital signs, $X^{(1)}$, and the demographic measures $X^{(2)}$, are defined as:

$$\mathbf{X}^{(1)} = \begin{pmatrix} X_1^{(1)} \\ X_2^{(1)} \\ X_3^{(1)} \\ X_4^{(1)} \end{pmatrix} = \begin{pmatrix} \text{Pulse Rate} \\ \text{Systolic Blood Pressure} \\ \text{Blood Temperature} \\ \text{Respiration Rate} \end{pmatrix}$$

$$\mathbf{X}^{(2)} = \begin{pmatrix} X_1^{(2)} \\ X_2^{(2)} \\ X_3^{(2)} \end{pmatrix} = \begin{pmatrix} \text{Age} \\ \text{Weight} \\ \text{Sex} \end{pmatrix}$$

Responses for variables $X^{(1)}$ and $X^{(2)}$ were recorded on a scale and then standardized. The sample correlation matrix based on 25 responses is:

$$\mathbf{R} = \left(\begin{array}{c|c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right)$$

$$= \left(\begin{array}{cccc|ccc} 1.0000 & & & & 0.4356 & 0.4017 & 0.1558 \\ 0.3654 & 1.0000 & & & 0.4852 & 0.3093 & 0.0758 \\ 0.3406 & 0.4298 & 1.0000 & & 0.6119 & 0.4491 & 0.1175 \\ 0.4961 & 0.3065 & 0.5164 & 1.0000 & 0.4548 & 0.6305 & -0.0878 \\ \hline 0.4356 & 0.4852 & 0.6119 & 0.4548 & 1.0000 & & \\ 0.4017 & 0.3093 & 0.4491 & 0.6305 & 0.4502 & 1.0000 & \\ 0.1558 & 0.0758 & 0.1175 & -0.0878 & 0.1527 & 0.0831 & 1.0000 \end{array}\right)$$

Table 2. *Output of Canonical Correlations and Multivariate Statistics.*

canon (Pulserate systolicBP bodytemp respirationrate)(age weight sex), test (1 2 3)
Tests of significance of all canonical correlations

	Statistic	df1	df2	F	Prob>F
Wilks' lambda	.337524	12	510.922	21.6117	0.0000 a
Pillai's trace	.78349	12	585	17.2321	0.0000 a
Lawley-Hotelling trace	1.61033	12	575	25.7206	0.0000 a
Roy's largest root	1.35877	4	195	66.2399	0.0000 u

Test of significance of canonical correlations 1-3

	Statistic	df1	df2	F	Prob>F
Wilks' lambda	.337524	12	510.922	21.6117	0.0000 a

Test of significance of canonical correlations 2-3

	Statistic	df1	df2	F	Prob>F
Wilks' lambda	.79614	6	388	7.8079	0.0000 e

Test of significance of canonical correlation 3

	Statistic	df1	df2	F	Prob>F
Wilks' lambda	.981001	2	195	1.8883	0.1541 e
e = exact, a = approximate, u = upper bound on F					

As Table 2 shows, the first canonical correlation is 0.7590, which would appear to be substantially larger than any of the between-set correlations. The probability level for the null hypothesis that all the canonical correlations are 0 in the population is only 0.0000, so firm conclusions can be drawn. Hence, the remaining canonical correlations are worthy of consideration.

For this particular model, there are three canonical dimensions of which only the first two are statistically significant. The first test of dimensions tests whether all three dimensions combined are significant which revealed that they are, the next test tests whether dimensions 2 and 3 combined are significant combined are significant which revealed that they are. Finally, the last test tests whether dimension 3, by itself, is significant which showed that it is not.

Table 3. *Output of Linear Combinations for Canonical Correlations.*

canon, test (1 2 3)
Linear combinations for canonical correlations
Number of obs = 200

		Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
u1							
	pulserate	.0175712	.0071157	2.47	0.014	.0035394	.031603
	systolicbp	.0121003	.0036762	3.29	0.001	.0048511	.0193496
	bodytemp	.2609796	.0467027	5.59	0.000	.1688839	.3530753
	respiratio~e	.1475122	.0249728	5.91	0.000	.098267	.1967575
v1							
	age	.0555526	.0060334	9.21	0.000	.0436551	.0674501
	weight	.0446671	.0056305	7.93	0.000	.0335641	.0557701
	sex	-.139978	.1242196	-1.13	0.261	-.3849338	.1049777
u2							
	pulserate	.043588	.0172132	2.53	0.012	.0096444	.0775317
	systolicbp	.0115943	.0088929	1.30	0.194	-.0059421	.0291307
	bodytemp	.420179	.1129764	3.72	0.000	.1973944	.6429636
	respiratio~e	-.3755301	.0604106	-6.22	0.000	-.4946571	-.2564031
v2							
	age	.0608586	.014595	4.17	0.000	.0320778	.0896393
	weight	-.0646738	.0136204	-4.75	0.000	-.0915326	-.037815
	sex	1.165916	.3004943	3.88	0.000	.5733542	1.758477
u3							
	pulserate	.101297	.0596009	1.70	0.091	-.0162334	.2188274
	systolicbp	-.0366158	.0307918	-1.19	0.236	-.097336	.0241043
	bodytemp	-.0787116	.3911829	-0.20	0.841	-.8501072	.692684
	respiratio~e	-.0473251	.2091726	-0.23	0.821	-.4598045	.3651543
v3							
	age	-.0538934	.0505355	-1.07	0.288	-.1535472	.0457604
	weight	.0475918	.0471608	1.01	0.314	-.0454072	.1405908
	sex	1.652417	1.040467	1.59	0.114	-.3993386	3.704172

(Standard errors estimated conditionally)

Canonical correlations:

0.7590	0.4341	0.1378

Tests of significance of all canonical correlations

	Statistic	df1	df2	F	Prob>F
Wilks' lambda	.337524	12	510.922	21.6117	0.0000 a
Pillai's trace	.78349	12	585	17.2321	0.0000 a
Lawley-Hotelling trace	1.61033	12	575	25.7206	0.0000 a
Roy's largest root	1.35877	4	195	66.2399	0.0000 u

For the first dimension, all the variables except for sex are statistically significant along with the dimension as a whole. Thus, pulse rate, systolic blood pressure, body temperature, and respiration rate share some variability with one another, as well as with age and weight, which also share variability among each other. For the second dimension, all the variables except systolic blood pressure are statistically significant along with the dimension as a whole. The third dimension is not significant and therefore no attention will be given to its coefficient or to the Wald tests.

Since the variables in the model have very different standard deviations, the standardized coefficients allow for easier comparisons among the variables. Hence, the standardized canonical coefficients for the first two dimensions that are significant are displayed in output 3

Table 4. *Standardized coefficients for the first variable set.*

canon (Pulserate systolicBP bodytemp respirationrate)(age weight sex), first(2) stdcoef notest

Canonical correlation analysis

Number of obs = 200

Standardized coefficients for the first variable set

	1	2
Pulserate	0.1809	0.4488
systolicBP	0.2318	0.2221
bodytemp	0.4254	0.6848
respiratio~e	0.4619	-1.1758

Standardized coefficients for the second variable set

	1	2
age	0.6373	0.6981
weight	0.5445	-0.7884
sex	-0.0699	0.5820

Canonical correlations:			
	0.7590	0.4341	0.1378

The technique of interpreting standard regression coefficients is applicable in standard canonical coefficients. For instance, consider the age, a one standard deviation increase in aging leads to a 0.64 standard deviation increase in the score on the first canonical variate for set 2 when the other variables in the model are held constant. The min.(p,q) = min.(4,3) = 3 sample canonical correlations and the sample canonical variate coefficient vectors are displayed in output 3.

The next is to obtain all the correlations within and between sets of variables as displayed in Table 5.

Table 5. *Output of Correlations within and between sets of variables.*

. estat correlations
Correlations for variable list 1

	Pulser~e	systol~P	bodytemp	respir~e
Pulserate	1.0000			
systolicBP	0.3654	1.0000		
bodytemp	0.3406	0.4298	1.0000	
respiratio~e	0.4961	0.3065	0.5164	1.0000

Correlations for variable list 2

	age	weight	sex
age	1.0000		
weight	0.4502	1.0000	
sex	0.1527	0.0831	1.0000

Correlations between variable lists 1 and 2

	Pulser~e	systol~P	bodytemp	respir~e
age	0.4356	0.4852	0.6119	0.4548
weight	0.4017	0.3093	0.4491	0.6305
sex	0.1558	0.0758	0.1175	-0.0878

Table 5 shows the correlations among the original variables.

The correlations between the vital signs and demographic measures are moderate, the largest being 0.6305 between weight and respiration rate. There are no larger within-set correlations as the highest among the vital signs occurred between respiration rate and body temperature (0.5164), while the highest among the demographic measures occurred between age and weight (0.4502).

Lastly, it is necessary to display the loadings of the variables on the canonical dimensions (variates). These loadings are correlations between variables and the canonical variates, and it is displayed in Table 6.

Table 6. *Output of Canonical loadings between variables and the canonical variates.*

estat loadings
Canonical loadings for variable list 1

	1	2
Pulserate	0.6396	0.1799
systolicBP	0.6223	0.3201
bodytemp	0.8251	0.3260
respiratio~e	0.8423	-0.5314

Canonical loadings for variable list 2

	1	2
age	0.8717	0.4321
weight	0.8256	-0.4257
sex	0.0727	0.6232

Correlation between variable list 1 and canonical variates from list 2

	1	2
Pulserate	0.4855	0.0781
systolicBP	0.4723	0.1390
bodytemp	0.6263	0.1415
respiratio~e	0.6393	-0.2307

Correlation between variable list 2 and canonical variates from list 1

	1	2
age	0.6616	0.1876
weight	0.6266	-0.1848
sex	0.0552	0.2705

5. Conclusion

It can be concluded from the analysis that canonical relations exhibited by the vital signs-demographical measures' data proved statistically significant in the first two canonical correlations, and statistically insignificant in the last (third) canonical correlation. Finally, the study concluded that relationship exists between the vital signs and the demographic measures of the 200 patients.

References

[1] William, R. D. and Matthew, G. (1984): Multivariate Analysis, Methods and Applications. John Wiley & Sons, New York.

[2] Parkhomenko, E., Tritchler, D., and Beyene, J. (2009): Sparse Canonical Correlation Analysis with Application to Genomic Data Integration. Statistical Applications in Genetics and Molecular Biology: Vol. 8: Iss. 1, Article 1.60.

[3] Johnson, R. A. and Wichern, D. W. (1992): Applied Multivariate Statistical Analysis, prentice Hall, Englewood Cliffs, New Jersey.

[4] Lin, C. T., Huang, C. S., Yang, W. Y., Singh, A. K., Chuang, C. H. and Wang, Y. K. (2018). Real Time EEG Signal Enhancement Using Canonical Correlation Analysis and Gaussian Mixture Clustering. Journal of Healthcare Engineering Volume 2018, Article ID 5081258.

[5] Jayadevan, C. M. (2018). A Canonical Correlation Analysis of Sectoral Composition of GDP and Development in Asia. Modern Economy, 9, 379-392.

[6] Mussina, K. P. and Bachisse, M. (2018). Canonical Correlation Analysis between Business Sophistication and Macroeconomic Environment: A Secondary study of Countries Global competitiveness. Revista ESPACIOUS. Vol. 39, Number 18.

[7] Idowu, O. A. and Titilola, O. O. (2018). Development of a canonical correlation model involving non linearity and asymmetric variables, Communications in Statistics - Theory and Methods, 47: 2, 399-414, DOI: 10.1080/03610926.2017.1303731

[8] Johnson, R. A. and Wichern, D. W. (2007): Applied Multivariate Statistical Analysis, Sixth Edition. Pearson Prentice Hall, New Jersey.

[9] Neil, H. T. (2002): Applied Multivariate Analysis. Springer-Verlag, New York Berlin Heidelberg.

[10] Witten, D. M. and Tibshirani, R. (2009) Extensions of Sparse Canonical Correlation Analysis with Applications to Genomic Data. Statistical Applications in Genetics and Molecular Biology: Vol. 8: Iss. 1, Article 28.

[11] Hoerl, A. E. (1962): Application of Ridge Analysis to Regression Problems. Chemical Engineering Progress. 58, 54-59.

[12] Lee, W., Lee D, Lee Y., and Pawitan, Y. (2011): Sparse Canonical Covariance Analysis for High-throughput Data. Statistical Applications in Genetics and Molecular Biology: Vol. 10: Iss. 1, Article 30.

[13] Lee, W., Lee, D, Lee Y., and Pawitan, Y. (2011): Sparse canonical covariance analysis. R package. http://www.meb.ki.se/yudpaw/.

[14] Waaijenborg S., Verselewel. D. E., Witt, Hamer P. C., and Zwinderman, A. H. (2008): Quantifying the Association between Gene Expressions and DNA-Markers by Penalized Canonical Correlation Analysis. Statistical Applications in Genetics and Molecular Biology: Volume 7: Issue 1, Article 3.