FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 6, December, (Special Issue) 2024, pp 315 - 320
DOI: https://doi.org/10.33003/fjs-2024-0806-3111

EVALUATING THE RELATIONSHIP BETWEEN VARIABLES: A CANONICAL CORRELATION ANALYSIS OF ACADEMIC PERFORMANCE IN NIGER STATE POLYTECHNIC, ZUNGERU

***[1]Ahmed, S. S., [2]Sani, U. M., [1]Santali, S. and [1]Saidu, U.**

[1]Department of Mathematics and Statistics, Niger State Polytechnic, Zungeru
[2]Department of Statistics, Kogi State Polytechnic, Lokoja

*Corresponding authors' email: ahmedsule710@gmail.com Phone: +2348036154396

ABSTRACT

Canonical Correlation Analysis (CCA) is a statistical technique used to investigate the relationship between two set of variables. CCA is particularly useful when dealing with multiple outcome variables that are intercorrelated. In situations where multiple regression analysis would be applicable, but there are multiple correlated dependent variables, CCA provides a more suitable approach. In this research, we used Canonical Correlation Analysis to investigate the level of correlation between some departmental and non-departmental courses, taken ND1 Estate Management and Valuation department, Niger State Polytechnic, Zungeru, 2022/2023 session as case study. Slovin's formula was used to determine the appropriate sample size to be used in this study. The researchers sampled 48 from the population in ND1 class. The analysis carried out using the SPSS package. Results obtained from the analysis shows that the correlation of V_1 on V_2 (EST111 on EST114) is 0.708. Also, the correlation of U_2 on V_2 (GNS111 on EST114) is 0.552. Y variables are the results of GNS101 and GNS111 and also represented by U_1 and U_2 respectively. X variables are the results for EST111 and EST114 and represented as V_1 and V_2 respectively. The extent to which departmental courses correlate with non-departmental courses is stronger than how non-departmental courses correlate with departmental courses this is in line with the outcome of the analysis. Based on the results obtained, it was recommended that there should be more efforts by the lecturers teaching non-departmental courses in the department concerned and the institution entirely.

Keywords: Canonical Correlation, *Slovin's formula*, Linear Combination, Canonical Variate, Intercorrelation

INTRODUCTION

Students' academic performance is a complex phenomenon influenced by a multitude of factors, including academic background, demographic characteristics, and individual attributes. Understanding the relationships between these factors is crucial for educators, policymakers, and researchers seeking to improve educational outcomes and optimize resource allocation.

The Canonical Correlation Analysis (CCA) is describe as a statistical technique particularly suited for exploring the relationships between the given sets of variables. By identifying patterns of correlation between academic and demographic factors, CCA can provide valuable insights into the factors influencing students' performance. CCA generates a set of canonical variates, which are orthogonal linear combinations of the variables within each set. These variates are specifically designed to maximize the explanation of variability both within each set and between the two sets, thereby revealing the underlying relationships between the variables.

According to Guo (2019), Canonical Correlation Analysis (CCA), also known as Canonical Variates Analysis (CVA), is a statistical technique that extracts valuable information from cross-covariance matrices. Given two vectors, X = $(X_1,.., X_m)$ and Y = $(Y_1 \ , ..., Y_m)$, comprising random variables with inherent correlations, CCA identifies linear combinations of these variables that maximize their mutual correlation. As noted by Knapp (1978): CCA serves as a comprehensive framework for investigating relationships between two sets of variables, encompassing various parametric tests of significance as special cases. The concept of CCA was first introduced by Harold Hotelling in 1936, building upon the mathematical foundations laid by Jordan in 1875 regarding angles between flats.

CCA is a versatile family of multivariate statistical techniques designed for analyzing paired sets of variables. Since its inception, it has undergone significant extensions to address various challenges, including: Handling situations where sample sizes are limited compared to the high dimensionality of the data. Capturing non-linear relationships between variables. And managing extremely high-dimensional data that surpasses human interpretability.

Canonical correlation (CANCOR) is describe as a correlation analysis involving multiple X and multiple Y. CANCOR seeks to find the correlation between set of X variables and set of Y variables. The term "canonical" is the statistical terms for analyzing latent variables (which are not directly observed) that represent multiple variables (which are directly observed). CANCOR measures the strength of association between two canonical variates (latent variables). A canonical variate is the weighted sum of the variables in the analysis (Usman, 2023).

Consider a scenario where you want to investigate the relationships between exercise habits and overall health. In this context, you have two distinct sets of variables. Exercise Variables such as; Climbing rate on a stair stepper, Running speed over a certain distance, Weight lifted on bench press and Number of push-ups per minute. Health Variables, these comprise metrics that assess overall health, including; Blood pressure, Cholesterol levels, Glucose levels, Body mass index (BMI). By examining the relationships between these two sets of variables, you can gain valuable insights into how exercise habits impact overall health. This is precisely the type of scenario where canonical correlation analysis (CCA) can be applied to uncover the underlying relationships between the exercise and health variables.

Canonical Correlation Analysis (CCA) offers a powerful means of distilling complex relationships between two sets of variables into a more manageable and interpretable form. By reducing the dimensionality of the data, CCA enables us to capture the essential features of the relationships while discarding less important details. In this sense, CCA shares a similar motivation with Principal Component Analysis (PCA), another dimension reduction technique. While PCA focuses on reducing the dimensionality of a single dataset, CCA extends this concept to two or more datasets, uncovering the underlying correlations and relationships between them.

CCA provides a valuable approach for examining the relationships between two sets of variables. This method describes the interconnections between the first set of variables and the second set, highlighting the correlations and patterns that exist between them. It does not inherently imply a cause-and-effect relationship between the two sets of variables. Instead, it treats both sets as interconnected, without designating one as independent and the other as dependent.

Canonical Correlation Analysis (CCA) has emerged as a versatile tool in various cutting-edge scientific disciplines, including; Neuroscience, Machine Learning and Bioinformatics. Relations have been explored for developing brain-computer interfaces (Cao *et al.* 2015; Nakanishi *et al.* 2015) and in the field imaging genetics (Fang *et al*. 2016). Examples of application studies conducted in the fields of bioinformatics and computational biology include (Seoane *et al.* 2014; Baur and Bozdag 2015; Sarkar and Chakraborty 2015; Cichonska *et al*. 2016). The wide range of application domains explain the utility of CCA in discovering relations between variables.

In a recent study published by Sevinç (2022) aimed to explore the relationship between nutrition and psychological status in adolescent students. To achieve this, Sevinç employed canonical correlation analysis (CCA) to identify statistically significant correlations between these two variables. The study further investigated whether the nutritional and psychological status of adolescents, differentiated by gender, had an impact on their educational success. Logistic regression analysis was used to examine this relationship.

A study conducted by Nouri *et al*. (2022) investigated the impact of access to and utilization of Information and Communication Technology (ICT) on academic benefits and motivation. The research focused on 300 students aged 12-16, comprising 160 boys, from Sanandaj, Iran. The researchers employed the ICT Familiarity Questionnaire, which evaluated both in-school and out-of-school ICT usage. The study's findings revealed a notable disparity in ICT access and usage between home and school environments. Specifically, the results indicated that students had significantly greater access to and utilization of ICT at home compared to school.

Despite the crucial role of Science, Technology, Engineering, and Mathematics (STEM) education in driving economic growth and development, the levels of performance and participation in STEM subjects remain persistently low in Kenya. This study focuses specifically on modeling the impact of school factors on student performance in mathematics and science in Kenyan secondary schools, utilizing Canonical Correlation Analysis (CCA) as the primary analytical technique. The core objectives of this research endeavor include; determining the magnitude of the relationship between school factors and performance in STEM education. Identifying the specific school factors that exert the most significant influence on student performance in mathematics and science as investigated by Mucunu (2018).

Akour *et al*. (2023) conducted a research study involving a sample of students from the Faculty of Business at Al-Balqa Applied University. To facilitate data analysis and extract meaningful results, the researchers employed statistical software programs, including SPSS version 26 and Stata graphics version 11. The study yielded several key findings; the first, second, and third canonical correlations were found to be statistically significant at a significance level of $\alpha \leq 0.05$. This comprehensive review paper provides an in-depth examination of multidimensional data analysis methods, with a specific focus on canonical analysis and its various variants. The paper also explores the applications of canonical analysis in omits data research, a field that has experienced rapid growth in recent years. The advent of high-throughput technologies has led to an explosion in the availability of large and complex datasets, which in turn has created a pressing need for innovative analytical approaches. These new methods must be capable of integrating data from diverse levels of biological organization, ranging from molecular and cellular levels to entire organisms and ecosystems (Wróbel, *et al* 2024).

A study by Laleh *et al.* (2015) highlights the importance of understanding the structural dependencies among a protein's side chains, which can provide valuable insights into their coupled motions. These coupled fluctuations play a crucial role in facilitating communication and information propagation within a molecule, ultimately influencing its function. Traditionally, side-chain conformations are represented by multivariate angular variables. However, existing partial correlation methods, which are commonly employed to infer structural dependencies, are limited in their ability to handle multivariate angular data.

A recent study conducted by Tang *et al.* (2022) investigated the relationship between nursing undergraduates' perceptions of their learning environment and their self-directed learning (SDL) abilities. This cross-sectional study, which took place in December 2020, involved a sample of 1096 junior and senior undergraduate nursing students aged 16-22 from Wannan Medical College in Anhui Province, China. The study's findings revealed that the total score for the learning environment was 120.60, corresponding to a scoring rate of 60.3%. Similarly, the total score for SDL ability was 89.25, with a scoring rate of 63.8%.

Correlation analysis cuts across all spheres of life and our everyday living as it involves knowing what will be the outcome of event A given B occurs or does not occur based on their correlation. Meanwhile, CCA tries to study the correlation of several variables grouped into two (2) major groups to know how well correlated they are or are not. In this research, the researchers limit the variables to the results of the ND1 students of Department of Estate Management and Valuation, Niger State polytechnic in some selected departmental and non-departmental courses to know if there is a reasonable correlation between the result of the student in departmental and non-departmental courses.

There are several methods through which canonical correlation analysis can be calculated ranging from manual calculation to running the process with the aid of statistical packages, but for the sake of this research, the statistical package SPSS (Version 23) will be used to carry out the analysis. The SPSS is a statistical package used for statistical analysis software to researchers, businesses, and academia.

The researchers aim is to use canonical correlation analysis to carry out the analysis and the objectives are; to determine the correlation between the result of the students in some selected departmental and non-departmental courses. And, to identified if there is correlation between the performances of students in departmental and non-departmental courses

MATERIALS AND METHODS

Data are important ingredient needed in statistics. It is the information that comes up as a result of statistical inquiries. It can also mean information, especially facts or numbers, collected to be examined and considered and used to run analysis which helps in decision-making.

In this research, the data used were obtained secondarily from the Estate Management and Valuation Department, Niger State Polytechnic, Zungeru 2022/2023 session. After the data was obtained, we used Slovin's formula to determine the appropriate sample size to be used from the data obtained.

After the appropriate sample size has been determined with the help of Slovin's formula, the sample space was selected using the random digit number to avoid any form of bias in the selection of the samples. The data collected was grouped into X and Y variates. The variables for Y are GNS101 and GNS111 and also denoted by U_1 and U_2 respectively. And, variables in X are the EST111 and EST114 represented by V_1 and V_2 respectively.

The CCA is performed when there are p variables in Z_1 and q variables in Z_2 and $(p+q)$ variables in all. Then, the correlation matrix R of order $(p+q)\times(p+q)$ is partitioned in to four parts. Thus, you have the following partition:

$$R = \begin{pmatrix} R_{11} & \vdots & R_{12} \\ \dots\dots & \vdots & \dots\dots \\ R_{21} & \vdots & R_{22} \end{pmatrix} \qquad (1)$$

Such that, R_{11} contains the intercorrelations among the elements of Z_1 of order$(p\times p)$, R_{22} contains the intercorrelations among the elements of Z_2 of order $(q\times q)$ and $R_{12} = R_{21}^T$, contains the cross-correlations between elements of Z_1 and Z_2 of order $(p\times q)$ (Usman, 2023).

Canonical Variates

By the notation, two group of variables; $X\ and\ Y$ are given as:

Assuming we have p-variables in group 1:

$$\boldsymbol{X} = \begin{Bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{Bmatrix} \qquad (2)$$

and we have q-variables in group 2:

$$\boldsymbol{Y} = \begin{Bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{Bmatrix} \qquad (3)$$

We select $2\ and\ 3$ on the bases of number of variables that exist in group so that $p \le q$, this is done for computational purposes.

In this approach, we examine linear combinations of the data, similar to principal components analysis. We define two sets of linear combinations, denoted as U and V. The set U corresponds to the linear combinations derived from the first set of variables (X), while the set V corresponds to the linear combinations derived from the second set of variables (Y). Each member of the set U is paired with a corresponding member of the set V. For instance, u_1 below is a linear combination of the pX variables and v_1 is the corresponding linear combination of the qX variables. Similarly, u_2 is a linear combination of the pX variables, and v_2 is the corresponding linear combination of the qY variables. And, so on....

$$\begin{aligned} U_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\ U_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \qquad (4) \\ &\vdots \\ U_p &= a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p \end{aligned}$$

Similarly:

$$\begin{aligned} V_1 &= b_{11}Y_1 + b_{12}Y_2 + \cdots + b_{1q}Y_q \\ V_2 &= b_{21}Y_1 + b_{22}Y_2 + \cdots + b_{2q}Y_q \qquad (5) \\ &\vdots \\ V_p &= b_{p1}Y_1 + b_{p2}Y_2 + \cdots + b_{pq}Y_q \end{aligned}$$

In equation 4 & 5, we define (U_i, V_i) as the i^{th} canonical variate pair. (u_1, v_1), is the first canonical variate pair, similarly (u_2, v_2), would be the second canonical variate pair and so on. With $p \le q$ there are p canonical covariate pairs, our objective is to identify linear combinations that maximize the correlations between the members of each pair. To achieve this, we compute the variance of u_i variables using the following expression:

$$var(U_i) = \sum_{k=1}^{p}\sum_{l=1}^{q} a_{ik}a_{il}cov(X_k,\ X_l) \qquad (6)$$

The coefficients a^{i1} through a^{ip} which appear in the double sum have the same coefficients that appear in equation 6 above. The covariances between the k^{th} and i^{th} $X-variables$ are multiplied by the corresponding coefficients a^{ik} and a^{il} for the variate.

Similar calculations can be made for the variance of v_j as shown below:

$$var(V_j) = \sum_{k=1}^{p}\sum_{l=1}^{q} b_{jk}b_{jl}cov(Y_k, Y_l) \qquad (7)$$

The covariance between 5 and 6 is $cov(U_i, V_j) = \sum_{k=1}^{p}\sum_{l=1}^{q} a_{ik}b_{jk}cov(X_k, Y_l)$ (8)

The correlation between u_i and v_j is calculated using the usual formula. We take 8 and divide by the square root of the product of the variances U_i & V_j as shown below:

$$\frac{cov(U_i, V_j)}{\sqrt{var(U_i)var(V_j)}} \qquad (9)$$

The canonical correlation is a specific type of correlation. The canonical correlation for the i^{th} canonical variate pair is simply the correlation between u_i and v_j:

$$\rho_i{}^* = \frac{cov(U_i, V_j)}{\sqrt{var(U_i)var(V_j)}} \qquad (10)$$

The correlation coefficient $\rho(U, V)$ represents the quantity to be maximized. Our objective is to identify the optimal linear combinations of the variables $X's$ and the variables $Y's$ that maximize the correlation coefficient $\rho(U, V)$.

To establish the validity of the canonical correlation analysis, it is essential to test for the presence of a relationship between the canonical variate pairs. This preliminary step enables us to determine whether there exists any statistically significant relationship between the two sets of variables.

To assess whether the Sales Performance variables and the Test Score variables are independent, we can initiate the analysis by formulating a multivariate multiple regression model. In this model, the Sales Performance variables serve as the outcome or response variables, while the Test Score variables act as the predictor variables. In this case, we have p multiple regressions, as such, each multiple regression predicting one of the variables in the first group $(X\ variables)$ from the q variables in the second group $(Y\ variables)$.

$$\begin{aligned} X_1 &= \beta_{10} + \beta_{11}Y_1 + \beta_{12}Y_2 + \cdots + \beta_{1q}Y_q + \epsilon_1 \\ X_2 &= \beta_{20} + \beta_{21}Y_1 + \beta_{22}Y_2 + \cdots + \beta_{2q}Y_q + \epsilon_2 \qquad (11) \\ \vdots\ &\quad \vdots \qquad \vdots \qquad\quad \vdots \quad \vdots \qquad\quad \vdots \\ X_p &= \beta_{p0} + \beta_{p1}Y_1 + \beta_{p2}Y_2 + \cdots + \beta_{pq}Y_q + \epsilon_p \end{aligned}$$

RESULTS AND DISCUSSION

Y variables are the results of GNS101 and GNS111 and are also denoted by U_1 and U_2 respectively. While, X variables are the results for EST111 and EST114 and are also represented by V_1 and V_2 respectively. This illustrated in table 1:

Table 1: Canonical Correlations Settings

	Values
Set 1: Y-Variables	GNS101(Y1) and GNS111(Y2)
Set 2: X-Variables	EST111(X1) and EST114(X2)
Correlations Used for Scoring	2

Table 2: Correlations[a]

		GNS 101	GNS 111	EST 111	EST 114
GNS 101	Pearson Correlation	1			
GNS 111	Pearson Correlation	.553	1		
EST 111	Pearson Correlation	.619	.578	1	
EST 114	Pearson Correlation	.602	.552	.708	1

a. Listwise N=48

Table 2: Shows the correlations which measure the magnitude of the association between the two sets of variables. The correlation of $V_1\, on\, V_2$ is 0.708 also, the correlation of $U_2\, on\, V_2$ is 0.552 which means they both have a positive correlation which also happens to be the strongest and weak correlation respectively.

Table 3: Canonical Correlations

	Correlation	Eigenvalue	Wilks Statistic	F	Num D.F	Denom D.F.	Sig.
1	.724	1.102	.476	9.903	4.000	88.000	.000
2	.014	.000	1.000	.009	1.000	45.000	.927

H0 for Wilks test is that the correlations in the current and following rows are zero

Table 3: In Canonical Correlation Coefficient, a high value (close to 1) indicates a strong relationship between the two sets of variables. The above result, it is shown that the canonical correlation of the first variate is 0.724 which means that 72.4% of variations in U_1 is explained by V_1 while the remaining 27.6% of variations in U_1 is accountable by other variables that are not in the model. Also, the correlation of the second variate is 0.014 means that 1.4% of variations in U_2 is explained by V_2 while the remaining 98.6% of variations in U_2 is accountable by other variables not captured in the model. Furthermore, the eigenvalue of 1.102 represents the amount of variance explained by the first canonical variate. A high eigenvalue indicates that the first canonical variate explains a substantial proportion of variance. While, 0.000 indicating that the second canonical variate does not explain any variance. 0.476 of Wilks' statistic represents the proportion of variance not explained by the canonical variate, a small Wilks' statistic indicates that the canonical variate explains a substantial proportion of variance. and 1.000 indicates that the second canonical variate does not explain any variance. The p-value for the first variate is .000 which implies that H_0 at 5% ls is significant since p-value < 0.05 and we conclude that $U_1 \neq V_1, and\ H_0$ at 5% ls is insignificant since p-value > 0.05 and we can conclude that $U_2 = V_2$.

Table 4: Set 1 & 2 Canonical Loadings

Variable	1	2
Y1	-.913	-.409
Y2	-.845	.534
X1	-.940	.341
X2	-.907	-.422

Table 4: Similar to factor loadings, this outcomes indicating the contribution of each original variable to the canonical variates. It indicate that Y_2 contribute higher to the 2 canonical variates

Table 5: Set 1 & 2 Cross Loadings

Variable	1	2
Y1	-.661	-.006
Y2	-.612	.007
X1	-.681	.005
X2	-.656	-.006

Table 5: Measure the correlation between the original variables and the opposite canonical variate. This shows extreme low positive and negative correlation.

Table 6: Proportion of Variance Explained

Canonical Variable	Set 1 by Self	Set 1 by Set 2	Set 2 by Self	Set 2 by Set 1
1	.774	.406	.853	.447
2	.226	.000	.147	.000

Table 6: Identify Set 1 by self, this denotes the proportion of variance explained by the first canonical variate of Set 1, Value: 0.774, this means that the first canonical variate of Set 1 explains approximately 77.4% of the variance in Set 1. However, Set 1 by Set 2: it represents the quantity of variance explained by the first canonical variate of Set 1, as predicted by Set 2, Value: 0.406, it indicate that first canonical variate of Set 1 explains approximately 40.6% of the variance in Set 1, as predicted by Set 2. Set 2 by self: This represents the variance explained proportionally by the first canonical variate of Set 2, Value: 0.853, it shows that first canonical variate of Set 2 explains approximately 85.3% of the variance in Set 2. Furthermore, Set 2 by Set 1: this represents the certain proportion of variance explained by the first canonical variate of Set 2, as predicted by Set 1 with Value: 0.447, meaning the first canonical variate of Set 2 explains approximately 44.7% of the variance in Set 2, as predicted by Set 1. For the second canonical variate; Set 1 by self: 0.226 (22.6%). Set 1 by Set 2: 0.000 (0%). Set 2 by self: 0.147 (14.7%). Set 2 by Set 1: 0.000 (0%). The second canonical variate explains a smaller proportion of variance in both sets, and the cross-set predictions are not significant (0%). Overall, the first canonical variate explains a substantial proportion of variance in both sets, disclosing a strong relationship between the two sets. Also, the second canonical variate explains a smaller proportion of variance and does not provide significant cross-set predictions.

The outcome of this research is compare to the research carried out by Akour *et al*. (2023), where they conducted a research study involving a sample of students from the Faculty of Business at Al-Balqa Applied University. The study yielded several key findings; the first, second, and third canonical correlations were found to be statistically significant at a significance level of $\alpha \leq 0.05$. The finding of this work can also be compare to the research by Mucunu *et al*. (2018), Their study focuses specifically on modeling the impact of school factors on student performance in mathematics and science in Kenyan secondary schools, utilizing Canonical Correlation Analysis (CCA) as the primary analytical technique. The core objectives of their research endeavor include; determining the magnitude of the relationship between school factors and performance in Science, Technology, Engineering, and Mathematics STEM education. Identifying the specific school factors that exert the most significant influence on student performance in mathematics and science as investigated.

CONCLUSION

After the analysis was carried out using a statistical package (SPSS), the results obtained showed that all the variables have positive correlations but the strongest correlation as shown in Table 2, is the correlation of $V_1\ on\ V_2$ *(EST111 on EST114) is* 0.708. Also, the correlation of $U_2\ on\ V_2$ *(GNS111 on EST114) is 0.552* which is a fairly positive correlation. While our $V_1\ and\ V_2$ are departmental courses EST111 and EST114 respectively. it means these two courses have higher influence on each other than any of the non-departmental courses in the model have on either of them. It also showed that the extent to which departmental courses correlate with other departmental courses is stronger than how any non-departmental courses correlate with departmental courses based the results of the analysis.

The researchers recommended that, there should be more efforts by the lecturers teaching non-departmental courses in the department of Estate Management. And, Students should be encouraged to see non-departmental courses as important as departmental courses.

REFERENCES

Akour, I., Rahamneh, A. A. L., Al Kurdi, B., Alhamad, A., Al-Makhariz, I., Alshurideh, M., and Al-Hawary, S. (2023). Using the canonical correlation analysis method to study students' levels in face-to-face and online education in Jordan. *Inf. Sci. Lett*, 12(2), 901-910. https://dx.doi.org/10.18576/isl/120229 .

Baur B and Bozdag S (2015).A canonical correlation analysis based dynamic Bayesian network Prior to infer gene regulatory networks from multiple types of biological data. *Journal of Computational Biology **22**, 289–299.* https://doi.org/10.1089/cmb.2014.0296

Cao L, Ju Z., Li J., Jian R., and Jiang C. (2015). Sequence detection analysis based oncanonical correlation for steady-state visual evoked potential brain computer interfaces. *Journal of neuroscience methods 253,, 10–17.* https://doi.org/10.1016/j.jneumeth.2015.05.014

Chen X., He C., and Peng H. (2014). Removal of Muscle Artifacts from Single-Channel EEG Based on Ensemble Empirical Mode Decomposition and Multiset Canonical Correlation Analysis. *Journal of Applied Mathematics.* https://doi.org/10.1155/2014/26/347.

Cichonska A, Rousu J, Marttinen P, Kangas A.J, Soininen P, Lehtim¨aki T, Raitakari O.T, J¨arvelin M.R, Salomaa V, and Ala-Korpela M, (2016). Meta Cannonical Correlation Analysis (CCA): Summary Statistics Based Multivariate Meta- Analysis of Genome Wide Association Studies Using Canonical Correlation Analysis. *Bioinformatics.* https://doi.org/10.1093/bioinformatics/btw052.

Fang J., Lin D., Schulz S.C, Xu Z, Calhoun V.D, and Wang Y-P. (2016). Joint sparse canonical correlation analysis for detecting differential imaging genetics modules. *Bioinformatics 32, 22 3480–3488.* https://doi.org/10.1093/bioinformatics/btw485

Guo, C., & Wu, D. (2019). Canonical correlation analysis (CCA) based multi-view learning: An overview. arXiv preprint arXiv:1907.01693. https://doi.org/10.48550/arXiv.1907.01693

Kabir A, Merrill R.D, Shamim A.A, Klemn R.D.W, Labrique A.B, Christian P, West Jr K.P, and Nasser M (2014). Canonical Correlation Analysis of Infant's Size at Birth and Maternal Factors: A Study In Rural Northwest Bangladesh. *PloS one 9(4), e94243.* https://doi.org/10.1371/journal.pone.0094243

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance-testing system. Psychological Bulletin, 85(2), 410. https://doi.org/10.1037/0033-2909.85.2.410

Laleh Soltan Goraie, Forbes Burkowski, and Mu Zhu, (2015): Using kernelized partial canonical correlation analysis to study directly coupled side chains and allosteric in small G proteins, Bioinformatics, Vol. 31(12), P. i124–i132. https://doi.org/10.1093/bioinformatics/btw241

Mucunu, J. M and George M. (2018). Modelling School Factors and Performance in Mathematics and Science in Kenyan Secondary Schools Using Canonical Correlation

Analysis. *Int. J. Comp. Theo Stat. 5(2).* https://dx.doi.org/10.12785/ijcts/050201

Nakanishi M., Wang Y, Wang Y.T, and Jung T.P. (2015). A Comparison Study of CanonicalCorrelation Analysis Based Methods for Detecting Steady-State Visual Evoked Potentials. *PloS one 10, e0140703.* https://doi.org/10.1371/journal.pone.0140703

Nouri, A., Zandi, T., and Etemadizade, H. (2022). A Canonical Correlation Analysis of the Influence of Access to and Use of ICT on Secondary School Students' Academic Performance. Research in Learning Technology, 30. https://doi.org/10.25304/rlt.v30.2679

Sarkar B.K and Chakraborty C. (2015). DNA pattern recognition using canonical correlation algorithm. *Journal of biosciences 40(4), 709–719.* https://doi.org/10.1007/s12038-015-9555-z

Sevinç, E. (2022). The effects of the relationship between psychological status and nutritional status on success in adolescent students with canonical correlation analysis. *Journal of Experimental and Clinical Medicine*, 39(2), 388-392. https://doi.org/10.52142/omujecm.39.2.15

Seoane J.A, Campbell C., Day I.N.M, Casas J.P, and Gaunt T.R. (2014). Canonical correlation analysis for gene based pleiotropy discovery. *PLoS Comput Biol 10(10), e1003876.* https://doi.org/10.1371/journal.pcbi.1003876

Tang, L. Q., Zhu, L. J., Wen, L. Y., Wang, A. S., Jin, Y. L., and Chang, W. W. (2022). Association of learning environment and self-directed learning ability among nursing undergraduates: a cross-sectional study using canonical correlation analysis. Bmj Open, 12(8), e058224. https://doi.org/10.1136/bmjopen-2021-058224

Tenenhaus A., Philippe C, and Frouin V. (2015). Kernel generalized canonical correlation analysis. *Computational Statistics & Data Analysis 90, 114–131.* https://doi.org/10.1016/j.csda.2015.04.004

Wróbel S, Turek C, Stępień E, and Piwowar M.(2024): Data integration through canonical correlation analysis and its application to OMICs research. *J Biomed Inform. 151:104575. 44, 1031–1040.* https://doi.org/10.1016/j.jbi.2023.104575