USMANU DANFODIYO UNIVERSITY, SOKOTO

(POST GRADUATE SCHOOL)

A STUDY OF THE PREVALENCE OF HIV/AIDS USING CANONICAL CORRELATION ANALYSIS
(A CASE STUDY OF GENERAL HOSPITAL MINNA)

A Dissertation

Submitted to the

Postgraduate School

USMANU DANFODIYO UNIVERSITY, SOKOTO

In partial fulfillment of the Requirements

For the Award of Degree of

MASTER OF SCIENCE (STATISTICS)

BY

MUHAMMAD, ABBAS AMINA

ADMIN NO. 112103061477

DEPARTMENT OF MATHEMATICS

JUNE, 2015

DEDICATION

This research dissertation is dedicated to Almighty Allah (S.W.T), The Most High.

CERTIFICATION

This Dissertation by Muhammad Abbas Amina Admission N0.1120103061477 has met the requirements for the award of the Degree of Masters of Science (STATISTICS) of the Usmanu Danfodiyo University, Sokoto, and is approved for its contribution to knowledge.

--------------------------------- ---------------------------
(External Examiner) Date

------------------------------- ---------------------------
Dr. Singh R.V Kumar Date
Major Supervisor

------------------------------------ ---------------------------
Dr. B.K Asare Date
Co-Supervisor

----------------------------------- ----------------------------
Dr. Ahmed Uba Date
Co-Supervisor

-------------------------------- ---------------------------
Prof. I.J. Uwanta Date
Head of Mathematics Department

ACKNOWLEDGMENTS

I attributed the success of this research work foremost to Allah for the courage and guidance He bestowed on to me, which beyond any reasonable doubt takes me through the hardship and adversity encountered during this research and the years of the study in general.

Special acknowledgment is also due to my supervisory team Dr.R.V.K.Singh, Dr. B.K. Asare, Dr. Ahmad Uba, Who did everything that an able supervisor would do to bring the best out of this project. Acknowledgement is also due to my head-of statistics Units Prof. S.U. Gulumbe, HOD Mathematics Prof I. J Uwanta.

My profound gratitude goes to Dr. A.D Koko, Dr.M.A.Aminu, Dr. Umar Usman, Dr. Yakubu Musa, Mal. Ahmad Audu and Shamsu Suleiman for their contribution. My gratitude also goes to Tasi'u Musa for his contribution and patience through this research work.

I remain forever indebted to my parents Alh. Muhammad Tako and Hajiya Rahmatu Tako, Hajiya Zainab Tako for their care and immense contribution towards my living and studies. Verily they have been of great help, may Allah reward them with Jannatul Firdausi.

This project would not have appeared like this without the assistance, encouragement advice, prayers, support and most especially love I received from my best friend and the love of my life, Abbas Jimada, my sweet son, Muhammadul-Ameen Abbas Jimada, I thank you for the patience and endurance you exhibited throughout my year of study.

My profound gratitude also goes to Alh, Musa Tako, Prof.Y.M. Auna, (Deputy Vice Chancellor Ibbul, and Prof. M.T. Mustapha (Director Academy Planning Ibbul) thank you advice and support.

Finally, I register my sincere gratitude to all my siblings and other relations for their moral support throughout the period of my study. May Allah always be your guide in your endeavors. To my sweet little sister Khadijat M. Tako, Thank you for your time out of school just to be supportive.

TABLE OF CONTENTS

TITLE PAGE i
CERTIFICATION iii
ACKNOWLEDGMENTS iv
TABLE OF CONTENTS vi
LIST OF TABLES viii
ABSTRACT ix

CHAPTER ONE
1.0 GENERAL INTRODUCTION 1
1.1 Statement of the Problem 4
1.2 Significance of the Study 5
1.3 Aim and Objectives 5
1.4 Scope and Limitation of the Study 5

CHAPTER TWO
2.0 REVIEW OF LITERATURES 6
2.1 Information Impact on HIV/AIDS Awareness 12
2.2 Concept of Artisans 13
2.3 Government Intervention in Nigeria 13

CHAPTER THREE
3.0 SOURCES OF DATA 15
3.1 Definition 15
3.2 Assumptions in Canonical Correlation 16
3.3 Theoretical Foundations 17
3.4 The Difference between Canonical Correlation Analysis and Ordinary Correlation Analysis 23
3.5 Correlation and Covariance Matrices 24
3.6 Tests for Significance 25
3.6.1 Wilk's lambda test 25
3.6.2 Bartlett's test 26
3.7 Coding and Importing of Data 27

3.8 Software for the Analysis 27
CHAPTER FOUR
4.0 ANALYSIS AND DISCUSSIONS 28
4.1 Introduction.. 28
4.2 Bartlett's Test.. 31

CHAPTER FIVE
5.0 SUMMARY, CONCLUSION AND RECOMMENDATION........................... 32
5.1 Introduction.. 32
5.2 Summary.. 32
5.3 Conclusion 33
5.4 Recommendation 33
REFERENCES 34
APPENDIX.. 38

LIST OF TABLES

Table 4.1: Canonical correlation coefficient of Set-A and Set-B 28

Table 4.2: To test that the canonical correlations are zero 29

Table 4.3: Canonical loading for Set-A and Set-B 29

ABSTRACT

This research aimed at fitting canonical correlation model that is capable of determining whether literacy level, age, marital status and gender are risk factors for HIV/AIDs. The data used is obtained from Heart to Heart center situated in General Hospital Minna, Niger State in June 2013. The Statistical package used to analyze the data is NCSS (Number Cruncher Statistical System) 2007 package and the result shows that the first set of variables measured the correlation of 0.2972 with the proportion of variability of 50.7% and the second set measured the correlation of 0.1412 with the proportion of variability of 49.3% . Hence, concluded that marital status and gender are risk factors of HIV/AIDs and implies that the prevalence of HIV/AIDS is higher among the married men.. I finally recommended that there is a need for the government to enact a Law that will make it mandatory for the two partners to undergo the HIV/AIDS test.

CHAPTER ONE

1.0 GENERAL INTRODUCTION

Since the discovery of Human immunodeficiency virus (HIV) as the causative organism of Acquired Immune Deficiency Syndrome (AIDS) in 1983, the infection has attained epidemic proportion globally. HIV/AIDS is an extraordinary kind of crisis; it is both an emergency and a long-term development issue. Tumer and Unal (2000) assert that (HIV/AIDS) is one of the most complex health problems of the 21^{st} century. Despite increased funding, political commitments and progress in expanding access to HIV treatment, the AIDS epidemic continues to outpace every global response.

Today the AIDS epidemic has become a pandemic disease that is threatening the world population. As the HIV/AIDS pandemic continues to spread around the world at an alarming rate, the number of people with this disease is been expected to grow significantly by the end of this decade. Moreover according to UNAIDS (2006); an estimated 24.7million people are living with HIV/AIDS in sub Saharan Africa. Meyer (2003) claims that HIV/AIDS which is acclaimed the fourth- leading cause of death worldwide is estimated to have claimed 25million lives since the beginning of the epidemic.

Acquired immune Deficiency syndrome (AIDS) is a viral disease caused by human Immunodeficiency virus (HIV) that is usually found in body fluids like blood, semen, vagina fluid, and breast milk of infected persons. The virus can be transferred from one infected person to another, mostly through sexual intercourse and sharing of unsterilized instruments like blades, knives, and syringes which had once been used by infected persons. (Olaleye 2003)

AIDS has rendered many children orphans, many of which were born with HIV infection. AIDS is killing the most productive people in the population, widening the level of development between developed and developing nations. It is also taking toll on the health sector since a lot of fund is channeled towards HIV/AIDS prevention and control. It has been observed that despite the many programmes organized to inform people about the problem of HIV/AIDS, the rate of it infection continues to be on the increase. (Omoniyi and Tayo-Olajubu 2006)

However, Cichocki (2010) Insists that, HIV testing is the first step to take when trying to find out a person's status. Never should one rely on symptoms of HIV to decide whether one is infected. HIV testing is the only way to know for sure. The importance of early diagnosis of HIV cannot be overstated. Decades of HIV and AIDS researchers have proven that the earlier HIV is diagnosed, the better the prognosis and the likelihood of a long and healthy life. Meanwhile, certain risk behaviors have been associated with high HIV infection rate. These behaviors according to Anochie and Eneh (2001) are either life style related or health-care provider risk. The life style related risk behaviors include multiple sexual partners, prostitution, sex with prostitute or casual partners, unprotected sex, intravenous, drug abuse and commercial blood donation among others.

Moreover, various campaigns have been mounted by both governmental and non - governmental association (NGOs) to curtail the spread of HIV/AIDS. Olaleye (2003) posits that these campaigns focused on measures to prevent HIV/AIDS infection. The measures include total abstainess from sex, use of condom to avoid infections from unprotected sexual intercourse, screening of blood meant for transfusion, keeping to one sexual partner, use of sterilized sharp object like blades, knives, needles / syringe, shaving and barbing instruments, Intending couples are also advised to do HIV/AIDS test before being joined in marriage.

Omoniyi and Tayo-Olajubu (2006) submit that People diagnosed with AIDS may get life-threatening diseases called opportunistic infection which are caused by microbes such as viruses and bacterial that usually does not make healthy people sick. However, What the HIV does is to gradually damage the immune system so that an infected person would be vulnerable to all sorts of diseases and illnesses, which may eventually lead to the total collapse of the immune system. It is at this point a person is said to be suffering from AIDS.

As of 2012 in Nigeria, the HIV prevalence rate among adults ages 15–49 was 3.1 percent. Nigeria has the second-largest number of people living with HIV. The HIV epidemic in Nigeria is complex and varies widely by region. In some states, the epidemic is more concentrated and driven by high-risk behaviors, while other states have more generalized epidemics that are sustained primarily by multiple sexual partnerships in the general population. Youth and young adults in Nigeria are particularly vulnerable to HIV, with young women at higher risk than young men. There are many risk factors that contribute to the spread of HIV, including prostitution, high-risk practices among itinerant workers, high prevalence of sexually transmitted infections (STI), clandestine high-risk heterosexual and homosexual practices, international trafficking of women, and irregular blood screening.

Nigeria is emerging from a period of military rule that accounted for almost 28 of the 47 years since independence in 1960. Consequently, the policy environment is not fully democratized. Civil society was weak during the military era, and its role in advocacy and lobbying remains weak. The size of the population and the nation pose logistical and political challenges particularly due to the political determination of the Nigerian

Government to achieve health care equity across geopolitical zones. The necessity to coordinate programs simultaneously at the federal, state and local levels introduces complexity into planning. The large private sector is largely unregulated and, more importantly, has no formal connection to the public health system where most HIV interventions are delivered. Training and human resource development is severely limited in all sectors and will hamper program implementation at all levels. Care and support is limited because existing staff are overstretched and most have insufficient training in key technical areas to provide complete HIV services.

Epidemiologically, UNAIDS estimates worldwide that 40,000,000 persons are living with HIV/AIDS, 18,500,000 (44%) of whom are women, and 3,000,000 (7.1%) of whom are children. The most heavily affected area of the world is sub-Saharan Africa, with almost 30,000,000 people infected with HIV.

South Africa has the largest population of HIV patients in the world, followed by Nigeria and India. South & South East Asia are second worst affected; in 2007 this region contained an estimated 18% of all people living with AIDS, and an estimated 300,000 deaths from AIDS.

1.1 Statement of the Problem

HIV/AIDS is being classified as dangerous disease and contribute to global health crisis. The rate of HIV/AIDS infection is increasing despite measures taken by government at federal, state and local levels. As a result of the danger of the disease, this research work is carried out to investigate whether canonical correlation model is capable of determining whether literacy level and gender are risk factors for HIV/ AIDs in Niger State.

1.2 Significance of the Study

This study is design to contribute knowledge to the understanding of HIV/AIDS. It is also useful to government and nongovernmental organization for them to rate strict measures to curtail the spread of the disease.

1.3 Aim and Objectives

The aim of this research work is to fit a canonical correlation model that is capable of determining whether literacy level and gender are risk factors for HIV /AIDS in Niger State.

The above aim is achieved through the following objectives;

- To determine the contribution of the risk factors to the prevalence of HIV/AIDS using canonical correlation.
- To determine the level of association between the canonical variates using Wilk's Lambda test.

To test for homogeneity of variances among the risk factors using Bartlet's test.

1.4 Scope and Limitation of the Study

This research work is based on data obtain from General Hospital Minna. (Heart to Heart centre, established, 2007) on HIV/AIDS and analyzed using Canonical correlation model that is capable of establishing the relationship between two sets of variables (Literacy level and Gender on one hand and Age, weight and marital status on the other hand).

CHAPTER TWO

2.0 REVIEW OF LITERATURES

In canonical correlation analysis (Hotelling, 1936), linear combinations of two sets of variables are obtained in such a way that the correlation between the linear combinations is a maximum. Generalizations to a similar approach for more sets of variables have been the topic of several studies (Horst 1961; Carroll 1968; Kettenring 1971).Consequently, several different approaches have been proposed. Kettenring (1971) provides an overview of four different generalizations. In the framework of homogeneity analysis Van der Burg (1988) and Gifi (1990) introduced nonlinear canonical correlation analysis, also referred by the algorithm name OVERALS, which takes Carroll (1968) generalized canonical correlation analysis as a special case. Using generalized canonical correlation analysis, a graphical representation, sometimes referred to as a perceptual map, can be made on the basis of the individuals' observation matrices. Note that, the observation matrices do not necessarily contain the same attributes. Steenkamp and Wittink (1994) focused on this flexibility in their analysis of idiosyncratic sets of attributes.

Another type of application, considered by Green and Carroll (1988), concerns the derivation of a composite configuration from a set of configurations. For example, multidimensional scaling solutions (perceptual maps) for the same objects from different countries can be used as input data. Generalized canonical correlation analysis can then be applied to the coordinate matrices to obtain a composite configuration. Finally, generalized canonical correlation analysis can be used when, for the same set of subjects, we have data on sets of variables. For example, in their analysis of socio-economic determinants of HIV pandemic and nations efficiencies, Zanakis *et al*. (2007) used a set of 50 explanatory variables which could be divided into different sets (e.g.

economic indicators, education related variables, etc.). For such multiple set data, generalized canonical correlation analysis can be used to obtain a configuration depicting the cases. Since generalized canonical correlation analysis deals with possibly large sets of data, the possibility of the occurrence of missing values is significant. Some procedures to deal with missing in generalized canonical correlation analysis have been proposed, however, no attempt has been made to compare and evaluate the alternatives .In this dissertation .We shall only concern ourselves with methods specifically aimed at dealing with missing values in generalized canonical correlation analysis. General methods (e.g. multiple imputation, Rubin 1987) that require distributional assumptions, are beyond the scope of this dissertation. The performance of the proposed methods under various conditions will be assessed by means of a simulation study. The results of this simulation study clearly indicate the validity and, in some cases, superiority of the new methods.

The Human immunodeficiency virus infection / acquired immunodeficiency syndrome (HIV/AIDS) is a disease of the human immune system caused by infection with human immunodeficiency virus (HIV). During the initial infection, a person may experience a brief period of influenza-like illness. This is typically followed by a prolonged period without symptoms. As the illness progresses, it interferes more and more with the immune system, making the person much more likely to get infections, including opportunistic infections and tumors that do not usually affect people who have working immune systems. HIV is transmitted primarily via unprotected sexual intercourse(including anal and even oral sex), contaminated blood transfusions, hypodermic needles, and from mother to child during pregnancy, delivery, or breastfeeding. Some bodily fluids, such as saliva and tears, do not transmit HIV. Prevention of HIV infection, primarily through safe sex and needle-exchange programs, is a key strategy to control the spread of the disease. There is no cure or vaccine; however, antiretroviral treatment can slow the course of the disease and may lead to a near-normal life expectancy. While antiretroviral treatment reduces the risk of death and complications from the disease, these medications are expensive and may be associated with side effects.

Genetic research indicates that HIV originated in west-central Africa during the early twentieth century. AIDS was first recognized by the Centers for Disease Control and Prevention (CDC) in 1981 and its cause—HIV infection—was identified in the early part of the decade. Since its discovery, AIDS has caused nearly 30 million deaths approximately 34 million people are living with HIV globally .AIDS is considered a pandemic—a disease outbreak which is present over a large area and is actively spreading HIV/AIDS has had a great impact on society, both as an illness and as a source of discrimination. The disease also has significant economic impacts. There are many misconceptions about HIV/AIDS such as the belief that it can be transmitted by casual non-sexual contact. The disease has also become subject to many controversies involving religion.

The Center for Disease Control and Prevention (CDC) defined a set of guidelines and recommendations for HIV-infected adolescents and adults on the basis of clinical conditions associated with the HIV infection and CD4+ T-lymphocyte counts [CDC 1997]. The system is based on three ranges of $CD4^+$ T-lymphocyte counts or CD4% and three clinical categories and is represented by a matrix of nine mutually exclusive categories.

These CDC guidelines have been based on studies done mostly in developed countries. A few studies have been carried out in developing countries on the basis of the present staging and monitoring system [Kam *et al.*, 1998]

With the pandemic spread of AIDs, a universally applicable staging system for HIV infection and disease is needed. The GPAids of WHO to first conduct a worldwide cross-sectional study, where clinical conditions were correlated with laboratory markers already known to reflect disease progression (particularly CD4 counts), using the later as surrogates of survival. The validation exercise involved data on 907 HIV naïve patients collected in 26 clinical centers from each of the 5 continents. The results of this validation exercise were reviewed by a TWG that met in Geneva on 21-23 February 1990, and which developed the present proposal (weekly epidemiological report 1990).

Vajpayee *et al.,* (2005) conducted a research that tried to classify HIV-seropositive antiretroviral treatment (ART)-naïve Indian individuals on CDC criteria of clinical symptoms and CD4% and CD4 counts. The optimum cut-off values of CD4 counts and CD4% obtained were compared with the CDC recommended values. The study also aimed to investigate the CDC staging of HIV-1 patients, on the basis of CD4 counts and CD4%, and the clinical implications in terms of HIV treatment and prophylaxis of these two staging criteria in an Indian population.

Ohno-Machado (1992) used Discriminant analysis to study the survival rates and the prognosis variables that corresponded to death during hospitalization in 312 AIDS CDC group IV patients in Sao Paulo. He asserted that Discriminant Analysis proved to be a good tool to perform exploratory data analysis that guided the survival analysis groups based on the nine predictor variables selected crucial to the progression of HIV/AIDS.

Suzy (2006) of centre for Public Health, Liverpool University investigated relationship between CD4 Count and viral load in HIV positive people in the North- West of England. He also identify whether a low CD4 count or a detectable viral load can be predicted using some socio-demographic variables. He specifically uses Binary Logistic Regression models to predict the stages of HIV/AIDS based on CDC staging format.

Mobili *et al.* (2010) used both principal component Analysis (PCA) and partial least square – discriminant analysis (PLS -DA) to analysis and interpretation of the Raman spectra collected from microorganism of different species recorded in the spectral range of 2000 to 200 cm^{-1}. To develop a classification rule, they use PLS-DA in a LOOCV method for the calibration and validation of a classification model, they asserted that, results obtained showed an acceptable classification among the strains under study; thereby suggested it to be useful tools for the classification and discrimination of similar samples.

Basta *et al.* (2008) used a cross sectional self report data collected from 208 HIV sero-positive individual to determine the accuracy of Transtheoritical model (TTM) constructs to predict the stages of change for exercise behaviors in individual living with HIV/ AIDS. They discovered based on their sample that predictive discriminant analysis classified HIV- naive individuals into the correct stages substantially better than chance alone except that no one was accurately predicted in one of the stages out of four.

Rubsamen – Waigman *et al.* (1991) work on CD4 - cell count, HIV antigen, β_2- micro globulin and serum cholesterol to evaluate the predictive values of clinical state of HIV patients with or without antiviral therapy, viral cultures an lymphocytes and monocytes . In their work, they concluded that Multivariant Discriminant Analysis showed that the combination of β_2- micro globulin, viral antigen, CD4+ cell count and HDL cholesterol predicted the outcome of viral cultures with 80% accuracy.

Ramayah *et al.* (2010) demonstrated an illustrated approach in presenting how the discriminant analysis can be carried out and how the output can be interpreted using knowledge sharing in an organizational context as a case study. They also presented 3 testing criteria to test whether the model developed has good predictive accuracy. In the work, they discovered based on their data analysis by advising case study organization to leverage on creating a more positive attitude among employees which will also result in a stronger subjective norm to share knowledge.

Mu Zhu and Trevor (2003) developed a general method for finding important discriminant directions without assuming the class densities belong to any particular parametric family, they also show that their method can be easily integrated with projection pursuit density estimation to produce a powerful procedure for (reduced-rank) nonparametric discriminant analysis.

Michael (2000) in his paper titled ‘Discriminant analysis and its application in DNA sequence’ review both Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) as the two basic parametric methods, he demonstrated their usage in recognition of splice sites and exons in the human genome (DNA).

Doss *et al* (2011) conducted a cross-sectional study of adults newly diagnosed with oral cancer in Malaysia to assess the cross-sectional construct validity of the Malay-translated and cross-culturally adapted FACT-H&N (v 4.0) for discriminative use in the sampled population with oral cancer.

2.1 Information Impact on HIV/AIDS Awareness

Information increases the level of certainty in any human decision process; little wonder, Edewor (2010) posits that information is indispensable for human development. Likewise , Nwafor-Orizu (2003) while describing sources of information dissemination in the rural areas in Nigeria, avers that, oral sources like face-to-face interaction, radio, television, traditional institutions, associations, and written sources like newspapers and magazines aims to facilitate rural information transfer as a way of eliminating ignorance and superstition. The present information and education campaign to forestall the spread of the disease should be pursued with vigor but some energy has to be dissipated to the care of people already afflicted. (Akanmu and Akinsete 2006)

Mooko and Aina (2007) opine that every individual, whether literate or illiterate, needs information for a variety of issues essential for his or her survival. It is therefore, not surprising that information is needed for awareness, increase productivity, health and so on.

They further assert that users of information are complex, while some are homogenous such as professionals, students, policy makers, researchers, some could be heterogeneous like rural inhabitants, artisans and so on. Ilo and Adeyemi (2010) in their own opinion submit that information is the most potent weapon available for the prevention and cure of HIV & AIDS.

HIV is a daily companion, In order to control the HIV epidemic, we all need to learn as much as possible about the disease. As for those living with HIV, comprehensive and up-to-date information is an essential part of a healthy life. There is no better place to start the education than at the beginning.

2.2 Concept of Artisans

According to Mooko and Aina (2007), artisans are those who perform skilled work with their hands. They are equipped with vocational education that may be acquired formally or informally. In most cases, they serve as apprentices before they become adept in their vocation. They are involved in all kinds of occupations, especially those in construction and motor industry; hence we have carpenters, plumbers, bricklayers, welders, painters, panel-beaters, electricians, mechanics, and so on. Moreover, other activities artisans are involved including hairdressing, tailoring, and dressmaking. They have limited education. Most of them would have primary education and a few have attained secondary education. Artisans are manual workers who labour with their hands, often in an urban setting and originally learned through apprenticeship. Artisans are the producers of hand-crafted goods that require an advanced level of knowledge and training.

The need for awareness programs to be extended to the artisans stems from the fact that this category of people are rarely found at home during the time of most educational activities on HIV/AIDS awareness take place. Hence, they harvest relevant and irrelevant information from their immediate environment and share same with their folks during conversation.

2.3 Government Intervention in Nigeria

In Nigeria an estimated 3.1 percent of adult between ages 15-49 are living with HIV/AIDS, approximately 170,000 people died from AIDS in 2007 alone (UNAID 2008). The first two cases of HIV/AIDS in Nigeria were identified in 1985 and were reported at an international AIDS conference in 1986. (Adeyi *et al.*, 2006). In 1987, the Nigerian health sector established the National AIDS Advisory Committee which was shortly followed by the establishment of the National Expert Advisory Committee on AIDS (NEACA).

Initially the Nigerian government was slow to respond to the increasing rates of HIV transmission and it was only in 1991 that the federal ministry of health made their first attempt to Nigeria's AIDS situation (Kanki and Adeyi 2006). The result then showed that 1.8 percent of the Nigerian population was infected with the deadly disease. However, when Olusegun Obasanjo became the president in 1999, HIV prevention, treatment and care became one of the government primary concerns. The National Action Committee on AIDS (NACA) was created and in 2001, the government set up a three-year HIV/AIDS Emergency Action Plan (HEAP). Subsequently, the president hosted the organization of Africa Unity's first African summit on HIV/AIDS, Tuberculosis and other related infection diseases (Adeyi *et al.,* 2006). A recent report from the Director General of the National

Agency for the Control of AIDS (NACA) during a stake holders meeting with the National Steering Committee on Orphans and Vulnerable Children (OVC), revealed that about one thousand (1,000) fresh cases of Human Immune Virus (HIV) is being recorded daily in Nigeria and that it was prevalent among the youths. (NACA 2010)

CHAPTER THREE

3.0 SOURCES OF DATA

The data used in this study is classified under the secondary source of data having been obtained from the Heart to Heart Centre, General Hospital Minna, Niger state on the 11^{th} Jun, 2013.

3.1 Definition

Canonical correlation analysis can be defined as the problem of finding two sets of basic vectors, one for x and the other for y, such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. Canonical correlation analysis (CCA) can also be defined as a way of measuring the linear relationship between two multidimensional variables. It finds two bases, one for each variable, that are optimal with respect to correlations and, at the same time, it finds the corresponding correlations. In other words, it finds the two bases in which the correlation matrix between the variables is diagonal and the correlations on the diagonal are maximized. The dimensionality of these new bases is equal to or less than the smallest dimensionality of the two variables.

An important property of canonical correlations is that they are invariant with respect to affine transformations of the variables. This is the most important difference between CCA and ordinary correlation analysis which highly depend on the basis in which the variables are described.

The appropriate data for canonical correlation analysis are two sets of variables. We assume that each set can be given some theoretical meaning, at least to the extent that one set could be defined as the independent variables and the other as the dependent variables. Once this distinction has been made, canonical correlation can address a wide range of objectives. These objects may be any or all of the following:

- Determining whether two sets of variables (measurements made on the same objects) are independent of one another or, conversely, determining the magnitude of the relationships that may exist between the two sets.
- Deriving a set of weights for each set of dependent and independent variables so that the linear combinations of each set are maximally correlated. Additional linear functions that maximize the remaining correlation are independent of the preceding set(s) of linear combinations.
- Explaining the nature of whatever relationships exist between the sets of dependent and independent variables, generally by measuring the relative contribution of each variable to the canonical functions (relationships) that are extracted.

The inherent flexibility of canonical correlation in terms of the number and types of variables handled, both dependent and independent, makes it a logical candidate for many of the more complex problems addressed with multivariate techniques.

3.2 Assumptions in Canonical Correlation

The generality of canonical correlation analysis also extends to its underlying statistical assumptions. The assumption of linearity affects two aspects of canonical correlation results. First, the correlation coefficient between any two variables is based on a linear relationship. If the relationship is nonlinear, then one or both variables should be transformed, if possible. Second, the canonical correlation is the linear relationship between the variates. If the variates relate in a nonlinear manner, the relationship will not be captured by canonical correlation. Thus, while canonical correlation analysis is the most generalized multivariate method, it is still constrained to identifying linear relationships.

Canonical correlation analysis can accommodate any metric variable without the strict assumption of normality. Normality is desirable because it standardizes a distribution to allow for a higher correlation among the variables. But in the strictest sense, canonical correlation analysis can accommodate even non - normal variables if the distributional form (e.g., highly skewed) does not decrease the correlation with other variables. This allows for transformed non - metric data (in the form of dummy variables) to be used as well.

Because tests for multivariate normality are not readily available, the prevailing guideline is to ensure that each variable has univariate normality. Thus, although normality is not strictly required, it is highly recommended that all variables be evaluated for normality and transformed if necessary.

Homoscedasticity, to the extent that it decreases the correlation between variables, should also be remedied. Finally, multicollinearity among either variable set will confound the ability of the technique to isolate the impact of any single variable, making interpretation less reliable.

3.3 Theoretical Foundations

Proposed by Hotelling in (1936) Canonical correlation analysis can be seen as the problem of finding basic vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors are mutually maximized. Correlation analysis is dependent on the co-ordinate system in which the variables are described, so even if there is a very strong linear relationship between two sets of multidimensional variables, depending on the co -ordinate system used, this relationship might not be visible as a correlation. Canonical correlation analysis seeks a pair of linear transformations one for each of the sets of variables such that when the set of variables are transformed the corresponding co-ordinates are maximally correlated.

Consider a multivariate random vector of the form (X, Y). Suppose we are given a sample of instances S $= [(X_1, Y_1), \ldots\ldots, (X_n, Y_n)]\ of\ (X, Y)$, we used S_x to denote $(X_1, \ldots, X_n)$ and similarly S_y to denote $(Y_1, \ldots., Y_n)$. We can consider defining a new co-ordinate by choosing a direction w_x and projecting X onto that direction

$$\mathrm{X} \rightarrow \langle w_x, X\rangle \qquad 3.1$$

If we do the same for **Y** by choosing a direction w_y we obtain a sample of the new **x** co-ordinate as

$$w_x, S_x = (\langle w_x, X_1\rangle, \ldots, \langle w_x, X_n\rangle) \qquad 3.2$$

With the corresponding values of the new y co-ordinate being

$$w_y, S_y = \left(\langle w_y, Y_1\rangle, \ldots, \langle w_y, Y_n\rangle\right) \qquad 3.3$$

The first stage of canonical correlation is to choose w_x and w_y to maximize the correlation between the two vectors. In other words the function to be maximized is $\rho =$

$$\underset{w_x, w_y}{\mathrm{maxcorr}}\left(w_x S_x, w_y S_y\right) = \max_{w_x, w_y} \frac{\langle w_x S_x, w_y S_y\rangle}{\|w_x S_x\| \|w_y S_y\|} \qquad 3.4$$

If we use $\hat{\mathbf{E}}[\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{Y})]$ to denote the empirical expectation of the function were

$$\hat{\mathrm{E}}\ [f\ (\mathrm{x},\ \mathrm{y})] = \frac{1}{m} \sum_i^m = 1\, f(X_i, Y_i) \qquad 3.5$$

We can rewrite the correlation expression as

$$\rho = \max_{w_x, w_y} \frac{\hat{E}[\langle w_x, X\rangle\langle w_y, Y\rangle]}{\sqrt{[\hat{E}\langle w_x, X\rangle^2][\hat{E}\langle w_y, Y\rangle^2]}} \qquad \rho =$$

$$\max_{w_x, w_y} \frac{\hat{E}[w_x' XY' w_y]}{\sqrt{\hat{E}[w_x' XX' w_x]\hat{E}[w_y' YY' w_y]}} \qquad 3.6$$

Which follows that

$$\rho = \max_{W_x, W_y} \frac{w'_{x}\, \hat{E}[XY']\, wy}{\sqrt{[w_x' \hat{E}\,(XX')\, w_x][w_y' \hat{E}(YY')w_y]}} \qquad 3.7$$

Where we used A' to denote the transpose of a vector or matrix **A**.

Now observed that the covariance matrix of $(X, Y)$is

$$C(X, Y) = \hat{E}\left[\begin{pmatrix} X \\ Y \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}^I\right] = \begin{bmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{bmatrix} = C \qquad 3.8$$

The total covariance matrix **C** is a block matrix where the within –sets covariance matrices are C_{xx} and C_{yy} and the between –sets covariance matrices are $C_{xy} = C_{yx}'$

The canonical correlations between X and Y can be found by solving the eigen value equations.

$$\begin{cases} C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} w_x = \rho^2 w_x \\ C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} w_y = \rho^2 w_y \end{cases} \qquad 3.9$$

Where the eigen values are the squared canonical correlations and the eigen vectors and are normalized canonical correlation basis vectors.

The number of non-zero solutions of these equations are limited to the smallest dimensionality of X and Y. Example, if the dimensionality of X and Y is 8 and 5 respectively, the maximum number of canonical correlation is 5.

Only one of the eigen value equations needs to be solved since the solution are related by

$$\begin{cases} C_{xy}w_y = \rho\lambda_x C_{xx}w_x \\ C_{yx}w_x = \rho\lambda_y C_{yy}w_y \end{cases} \qquad 3.10$$

Where

$$\lambda_x = \lambda_y^{-1} = \sqrt{\frac{w_y^{'}C_{yy}w_y}{w_x^{'}C_{xx}w_x}} \qquad 3.11$$

Hence, we can rewrite the function ρ as

$$\rho = \max_{w_x w_y} \frac{w_x^{'}C_{xy}w_y}{\sqrt{(w_x^{'}C_{xx}w_x)(w_y^{'}C_{yy}w_y)}} \qquad 3.12$$

The maximum canonical correlation is the maximum of ρ with respect to $w_x\ and w_y$

Observe that the solution of equation (3.12) is not affected by re-scaling $w_x\ or\ w_y$ either together or independently, so that for example replacing $w_x$by αw_x gives the quotient.

$$\frac{\alpha w_x^{'}C_{xy}w_y}{\sqrt{\alpha^2 w_x^{'}C_{xx}w_x w_y^{'}C_{yy}w_y}} = \frac{w_x^{'}C_{xy}w_y}{\sqrt{w_x^{'}C_{xx}w_x w_y^{'}C_{yy}w_y}} \qquad 3.13$$

Since the choice of re-scaling is therefore arbitrary, the CCA optimization problem formulated in equation (3.13) is equivalent to maximizing the numerator.

Subject to

$$w_x' C_{xx} w_x = 1 \qquad 3.14$$

$$w_y' C_{yy} w_y = 1 \qquad 3.15$$

The corresponding Lagrangian function is

$$\mathrm{L}(\lambda, w_x, w_y) = w_x' C_{XY} w_y - \frac{\lambda_x}{2}(w_x' C_{XX} w_x - 1) - \frac{\lambda_y}{2}\left(w_y' C_{YY} w_y - 1\right) \qquad 3.16$$

Taking derivatives in respect to w_x and w_y we obtain

$$\frac{\partial f}{\partial w_x} = C_{XY} w_y - \lambda_x C_{XX} w_X = 0 \qquad 3.17$$

$$\frac{\partial f}{\partial w_y} = C_{YX} w_x - \lambda_y C_{YY} w_y = 0 \qquad 3.18$$

Subtracting w_y' times the second equation from w_x' times the first we have

$$0 = w_x' C_{XY} w_y - w_x' \lambda_x C_{XX} w_x - w_y' C_{YX} w_x + w_y' \lambda_y C_{YY} w_y$$

$$= \lambda_y w_y' C_{yy} w_y - \lambda_x w_x' C_{XX} w_x \qquad 3.19$$

which together with the constraints implies that

$$\lambda_y - \lambda_x = 0$$

Let λ= $\lambda_x = \lambda_y$.

Assuming C_{yy} is invertible we have

$$w_y = \frac{C_{YY}^{-1} C_{YX} w_x}{\lambda} \qquad 3.20$$

and so substituting in equation (3.19) gives

$$\frac{C_{XY} C_{YY}^{-1} C_{YX} w_x}{\lambda} - \lambda C_{XX} w_x = 0$$

Or

$$C_{XY}C_{YY}^{-1}C_{YX}w_x = \lambda^2 C_{XX}w_x \qquad 3.21$$

We are left with a generalized eigen problem of the form$= \lambda BX$. We can therefore find the co-ordinate system that optimizes the correlation between corresponding co-ordinates by first solving for the generalized eigenvectors of equation (3.21) to obtain the sequence of w_x and then using equation (3.20) to find the corresponding $w_y's$.

As the covariance matricesC_{XX} and C_{YY} are symmetric positive definite we are able to decompose them using a complete Cholesky Decomposition.

$$C_{XX} = R_{XX}R_{XX}^{'}$$

Where R_{XX} is a lower triangular matrix. If we let $U_x = R_{XX}^{'}.w_x$we have equation (3.21) as follows.

$$C_{XY}C_{YY}^{'}C_{YX}R_{XX}^{-1}U_x = \lambda^2 R_{XX}U_x$$

$$R_{XX}^{-1}C_{XY}C_{YY}^{-1}C_{YX}R_{XX}^{-1'}U_x = \lambda^2 U_x$$

We are therefore left with a symmetric eigen problem of the form $AX = \lambda X$

3.4 The Difference between Canonical Correlation Analysis and Ordinary Correlation Analysis

Ordinary correlation analysis is dependent on the coordinate system in which the variables are described. This means that even if there is a very strong linear relationship between two multidimensional signals, this relationship may not be visible in a ordinary correlation analysis if one coordinate system is used, while in another coordinate system this linear relationship would give a very high correlation.

Canonical correlation analysis finds the coordinate system that is optimal for correlation analysis, and the eigen vectors of equation (3.9) defines this coordinate system .For example, Considering two normally distributed two dimensional variables X and Y with unit variance.

Let $y_1 + y_2 = x_1 + x_2$ .It is easy to confirm that the correlation matrix between X and Y is

$$R_{xy} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

This indicates a relatively weak correlation of 0.5 despite the fact that there is a perfect linear relationship (in one dimension) between X and Y.

A Canonical correlation analysis on this data shows that the largest and only canonical correlation is one and it also gives the direction $[1 \quad 1]'$ in which this perfect linear relationship lies. If the variables are described in the bases given by the canonical correlation basis vectors (i.e the eigen vectors of equation (3.9) the correlation matrix between the variables is

$$R_{xy} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

3.5 Correlation and Covariance Matrices

In neural network literature, the matrix C_{xx} in equation (3.8) is often called a correlation matrix. This can be a bit confusing, since C_{xx} does not contain the correlations between the variables in a Statistical sense, but rather the expected values of the products between them. The correlation between x_i and x_j is defined as

$$\rho_{ij} = \frac{E[(x_i - \bar{x}_i)(x_j - \bar{x}_j)]}{\sqrt{E[(x_i - \bar{x}_i)^2]E\left[(x_j - \bar{x}_j)^2\right]}} \qquad 3.22$$

From the above example, the covariance between x_i and $x_j$normalized by the geometric mean of the variances of x_i and x_j, $(= E[x])$Hence, the correlation is bounded.

$$-1 \leq \rho_{ij} \leq 1.$$

The diagonal terms of C_{xx} are the second order origin moments,$E[x_i^2]$, of x_i .

The diagonal terms in a covariance matrix are the variances or the second order central moments,

E$[(x_i - \bar{x}_i)^2]$, of x_i .

The maximum likelihood estimator of ρ is obtained by replacing the expectation operator in equation (3.22) by a sum over the samples. This estimator is sometimes called the pearson correlation coefficient after K. Pearson.

3.6 Tests for Significance

3.6.1 Wilk's lambda test

Wilk's lambda is a test statistic used in multivariate analysis of variance (MANOVA) to test whether there are differences between the means the means of identified groups of subjects on a combination of dependent variables. It is a direct measure of the proportion of variance in the combination of dependent variables that is unaccounted for by the independent variable (the group variable or component). Wilk's lambda statistic can be transformed (mathematically adjusted) to a statistic which has approximately on F distribution. This makes it easier to calculate p-value (Everitt and Dunn, 1991).

The Wilk's lambda test of hypothesis is given as:

H_0:$\sum_{xy} = 0$, i.e. there is no relationship between the canonical variates.

H_1:$\sum_{xy} \neq 0$, i.e. there is relationship between the canonical variates.

Test statistic:

$$\lambda = \frac{|R|}{|R_{YY}||R_{XX}|} \qquad 3.19$$

Where:

R is the correlation between x`s and y`s

$R_{XX}$is the correlation between x`s

R_{YY} is the correlation between y`s

Significance level:$\propto$ = 0.05

Decision rule: Reject H_0 if $p < 0.05$ and otherwise accept. (Rencher, 2012)

3.6.2 Bartlett's test

Bartlett's test is used to test the homogeneity of variance in the factors. Bartlett's test is used to test if K group have equal variances. Equal variances across groups or samples are called homogeneity of variance. Bartlett's Test is sensitive to departures from normality. That is, if your groups/samples come from non-normality distribution.

The Bartlett's test of hypothesis is given as:

H_0: $\delta_1 = \delta_2 = \ldots = \delta_k$

H_1: $\delta_i \neq \delta_j$ for at least one pair (i, j)

Test statistic:

$$T = \frac{(N-k)\ln S_p^2 - \sum_{i=1}^{k}(N_i - 1)\ln S_i}{1 + \left(1/3(k-1)\right)\left(\left(\sum_{i=1}^{k} \frac{1}{(N_i - 1)}\right) - \frac{1}{N-k}\right)} \quad 3.20$$

In the above, S_i^2 is the variance of the i^{th} group, N is the total sample size, N_i is the sample of the i^{th} groups, and S^2_p is the pooled variance. The pooled variance is weighted average of the group variance and is defined as:

$$S_P^2 = \sum_{i=1}^{k} {(N_i - 1)S_i^2}/{(N-k)} \quad 3.21$$

Significance Level: $\propto = 0.05$

Critical Region: the variances are judged to be unequal if, T $> x^2_{(\alpha,k-1)}$

Where $x^2_{(\alpha,k-1)}$ is the upper critical value of the chi-square distribution with k-1 degree of freedom and a significance level of .

3.7 Coding and Importing of Data

Coding was done in line with the NCSS package requirements. Our chosen variables were accordingly coded as follows.

Sex (Gender) _______ Male 1 Female 0

Age _______Numerical

Weight _______ Numerical

Marital Status: married 1, singled 2, divorced 3 , widow 4

3.8 Software for the Analysis

The software used for the analysis of the data collected is NCSS (Number Cruncher Statistical System, 2007 version) .

CHAPTER FOUR

4.0 ANALYSIS AND DISCUSSIONS

4.1 Introduction

This chapter deals with the analysis and discussion of the studied data considered in the research work. The analysis of canonical correlation was adopted. An initial step in canonical correlation is an inspection of the correlation matrix of the given data.

Let S denote the data such that:

S = {set- A, set- B}

Where

Set-A= {literacy level, Gender}

Set-B= {Age, Marital Status ,Weight}

Proper analysis begins with a simple examination of the correlation significance. Dunn and Deokson, (1977)

Table 4.1: Canonical correlation coefficient of Set-A and Set-B

Canonical function	Canonical correlation	Eigen value	% of variance
1	0.2972	0.0883	50.7
2	0.1412	0.0199	49.3

Table 4.1 shows the canonical correlation of the two canonical varieties and their corresponding eigen values. The eigen values of the canonical variates can be tested by

employing Wilk's lambda criterion to test for the significant by using Wilk's lambda test,(Rencher, 1998)

H_o: $\sum_{xy} = 0$ against H_1: $\sum_{xy} \neq 0$ at $\alpha = 0.05$

Reject H_o if p<= 0.05. We have the following table

Table 4.2: To test that the canonical correlations are zero

S/NO	N	P	O	DF	P-Value	α-value
1	1045	4	2	6	0.0000	0.05
2	1045	3	1	2	0.0000	0.05

Table 4.2 shows that the canonical correlations tested are significant i.e. p-value <0.05, which implies that, the null hypothesis is rejected, it also indicates that, the first and the second canonical correlations are significantly different from zero where p is the number of variables considered in a certain canonical variates, while Q is the number of variables considered in the opposite canonical variates, and df is the degree of freedom used at each level of canonical function. Refer to Table A1 in Appendix .

Table 4.3: Canonical loading for Set-A and Set-B

Sets	Risk factors	R_1	R_2
Set-A	L/level	0.2708	0.9631
	Gender	0.9555	-0.2962
Set-B	Age	0.5711	-0.7793
	Weight	0.3841	0.0201
	Mar-status	-0.8584	-0.4667

The canonical loading in Table 4.3 provides information about the relative contribution of variable to each independent canonical relationship, the first pair of canonical variate can be written as follows.

U_1=0.2708 lit level + 0.955 Gender

V_1=0.5718 Age + 0.3841weight - 0.8584 Marital status

Ø=0.2972.

The correlation $Ø_1$ between U_1 and V_1 is called the first canonical correlation coefficient of the individual variables. Gender loading is heaviest with the value 0.9555 followed by marital status -0.8584 loading for the ordering for the criterion variables. This implies that the prevalence of HIV/AIDS is higher among the married men.

The values attached to each risk factor are their partial correlation to their corresponding canonical variable and indicating the individual contribution to the canonical pair. Refer to Table A3 and A4 in Appendix.

Table 4.4: Variable-variants correlations section for Set A and Set B

Sets	Subject	R_1	R_2
Set A	l/level	0.0881	0.1349
	Gender	0.2861	-0.0382
Set B	Age	0.1262	-0.1259
	Weight	0.1271	-0.0208
	Martial	-.2053	-0.0131

Tables 4.4 shows the canonical cross loading of two canonical functions, in the first canonical function, we see that Gender slightly has high correlation with independent canonical variate 0.2861 and -0.0382 respectively while the weakest correlation came from set B, i.e. Marital Status with 0.2053 and Age with-0.1259 refer to Table A5 in Appendix.

However the canonical correlation which examine the linear relationship between set-A and set-B variables is by creating the combinations. The first canonical correlation explains the maximum relationship between the canonical variate and each successive canonical correlation is estimated so as to be orthogonal, yet still explain the maximum relationship not accounted for by the previous canonical correlation. This reflects the high variance among these variable by squaring the terms in the canonical loading, we find percentage of the variance for each of the variable explained by function.

4.2 Bartlett's Test

The Bartlett's requires measuring the Homogeneity of variance across variables, i.e. risk factors.

Hypothesis

H_0: $\delta_1 = \delta_2 = \cdots = \delta_k$

H_1: $\delta_i \neq \delta_j$ for at least one pair (i, j)

Reject H_o if p-value <= 0.05

Form the Bartlett's test approximation chi-square value is 120.01 with degree of freedom of 18, and probability level of 0.000 at α =0.05. We therefore reject H_o and conclude that the variates across the variables are not equal. Refer to Table A6 in Appendix.

CHAPTER FIVE

5.0 SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Introduction

This chapter shows the summary, conclusion and recommendations reached in the analysis of canonical correlation used on the entire variables / data.

5.2 Summary

The aim of the study is to fit a canonical correlation model that is capable of determining whether literacy level and gender are risk factors for HIV/AIDS in Niger State and hence to predict future occurrence.

The canonical correlation analysis generated two correlation coefficients, which are tested. It was found that both of the correlations were statistically different from zero.

The first canonical pair captured the variability of about 50.7%, and the second canonical pair captured the variability of about 49.3%. Hence the total variability captured by two canonical pairs is 100%. The 50.7% variability is due to the individual contribution given by the risk factors (literacy level, gender, age, marital status and weight); showing that there is a slight relationship among listed factors.

Considering the first canonical variate pair U_1 and V_1 with canonical correlation coefficient $r_1=0.2972$, that the proportion of variance common to the first canonical variate pair is $r_1^2 = 0.0883$ showing about 8.83% of the proportion of variance captured by the first canonical variate.

Similarly r_2=0.1412 is the canonical correlation coefficient between the second canonical variate pair and so r_2^2= 0.0199 which indicates about 1.99% of the proportion of variance captured.

Although; canonical correlation analysis has many table for interpretation, further interpretation of the canonical correlation coefficient will be done as suggested by Dunn and Deokson, (1997), using canonical loadings and canonical cross loadings.

5.3 Conclusion

It can be seen that set-A and set-B are slightly correlated as sought for. Canonical correlation analysis measured the strength of relationship of the canonical pair and the factors that strongly contributed. The first pair has a measure of correlation 0.2972 with the proportion of variability of about 50.7% and the second pair has a measure of correlation of 0.1412 with the proportion of variability of about 49.3%.

The model shows that marital status and gender are the risk factors with higher contribution for the prevalence of HIV/ AIDS. Conclusively, by combining the results of the two major risk factors, the prevalence of HIV/AIDS is higher among married men.

5.4 Recommendation

I finally recommended that there is a need for the government to enact a Law that will make it mandatory for the two partners to undergo the HIV/AIDS test.

REFERENCES

Adeyi, O., Kanki , P.J., Odutolu, O., &Idoko, J.A (2006). Aids in Nigeria: A nation on the threshold Cambridge, *MA: Harvard centre for population and development studie*s.

Akanmu , A.S., &Akinsete, A. (2006). Current thought in the management of HIV/AIDs *Journal of Clinical Sciences 1(1): 8-14*

Anderson ,T. W. . *An introduction to Multivariate Statistical Analysis.* John Wiley & Sons, second edition

Anochie, I.C., &Eneh, A.U (2001).Survey of HIV Risk Behavior among Health Workers in Port-Harcourt.*The Nigerian Medical Practitioner 39(3&4):49*

Basta *et al* (2008): *Predictors of exercise stage of change among individuals living with HIV/AIDS;* Med Sci Sports Exerc. (2008) Vol. 40 (9) 1700 -6

Carroll, J. D., Green, P. E., & Schaffer, C. M. (1968). Reply to Greenacre's commentary on the Carroll-Green-Schaffer scaling of two-way correspondence analysis solutions. *Journal of Marketing Research*, 366-368.

CDC. (1997): Revised guidelines for performing CD4 + T – cell determinations in persons infected with HIV. MMWR 1997, vol 46 (no RR -2).

Cichocki,R.N., (2010). Symptoms and prevention of HIV/AIDS.*Retrieved July 14,2010 from http://aids.about .com/od/hiv101/u/basics.htm.*

Country Profile (2008): Nigeria (http://2006-2009. Pepfar.gov/press/81548.htm). *U. S. Department of State.2008*. Archived from the original (http://www.pepfar.gov/pepfar/press/81548.htm) on 27 June 2014.

Das ,S. and P. K. Sen. *Restricted canonical correlations. Linear Algebra and its Applications,*

Doss, J. G., Thomson, W. M., Drummond, B. K., & Latifah, R. R. (2011). Validity of the FACT- H&N (v 4.0) among Malaysian oral cancer patients. *Oral oncology*, *47*(7), 648-652

Dunn, J.W., &Doekson, G.A., (1977). *Canonical Correlation Analysis of Selected Demographic and Health Personal Economics. 2: 565-570.*

Edewor, E. (2010). Access to health information by people living with HIV/AIDs In Nigeria. Library philosophy and practice. *Retrieved 22 march 2010 from http://unllib.unl.edu/llp/edewor .htm* edition,1984.

Everitt B.S and Dunn G 1991: *Applied Multivariate data Analysis*.

Green, P. E., & Carroll, J. D. (1988).A simple procedure for finding a composite of several multidimensional scaling solutions. *Journal of the Academy of Marketing Science*, *16*(1), 25-35.

Health Protection Agency. The National CD4 Surveillance Scheme; 2007. www.hpa.org.uk/infections/topics_az/hiv_and_sti/hiv/epidemiology/cd4.htm.

HIV/AIDS–adult prevalence rate*(https://www.cia.gov/library/publications/the-world-factbook/rankorder/ 2155rank.html)* CIA World Factbook (2012) Accessed February 20, 2014.

HIV/AIDS–People Living with HIV/AIDS *(https://www.cia.gov/library/publications/the-world-factbook/rankorder/ 2156rank.html)* CIA World Factbook(2012) Accessed February 20, 2014.

Horst, P. (1961).Generalized canonical correlations and their applications to experimental data.*Journal of Clinical Psychology*, *17*(4), 331-347.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 321-377.

Ilo, P.I., &Adeyemi, A.(2010). HIV/AIDs information and awareness among market women: A study of Olofinmoye Market Sango –Ota, Ogun state, *Nigeria. Library Philosophy and Practice. Retrieved 22 march 2010 from htt://unllib.Unl.edu/LLp/ilo.adeyemi.htm*

In Unal S. (Ed) AIDS: *Modern Medical Seminars Ankara : Gunes:1-10*

Joint United Nations Programme on HIV/AIDS Staff, World Health Organization, & World Health Organization Staff.(2006). *AIDS epidemic update, December 2006*.World Health Organization.

Kam KM, Wong K.H. (1998) : *Proposed cd4 + T- cell criteria for staging HIV – infected Chinese adults*. Clin immunol immunopattrol vol 89, pg 11-22.

Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, *58*(3), 433-451.

McLachlan, G. J. (1992), '*Discriminant Analysis and Statistical Pattern Recognition'*, John Wiley,NewYork.

Meyer, D. (2003). HIV/AIDs and Education in Africa. Rand Afrika'ansuniversity. *Retrieved 22 march 2010 from : http:// www.unep.org*

Michael Q. Zhang (2000): *Discriminant Analysis and its application in DNA sequence, Briefings in Bioinformatics* Vol 1 no 4 00-00

Mobili P, Landero A .De-Antoni G. Araujo Andraote C, Avila- Donoso H, Moreno J. (2010): *Multivariate Analysis of Raman Spectra applied to microbiology: Discriminant of Microorganism at the species level*. REVISTA MEXICANA DE FESICA 56 (5) 378- 385

Mooko, N., &Aina, L.O. (2007).*Information Environment of artisans in Botswana Libri 57: 27-33.*

Mu Zhu and Trevor J. Haste (2011) : *Feature extraction for nonparametric discriminant analysis*, Journal of Computational and Graphical Statistical Computing vol 12, no 1

NACA (2010).Nigeria record 1000 fresh cases of HIV infection daily.*The Nigerian Tribune (August 18)*

Nwafor –Orizu, U.E (2003). Do something about HIV/AIDs now. *New Horizon Volunteers on AIDs (HOVA).*

Olaleye, R.S (2003). Level of HIV/AIDS Awareness among rural farmers and its implication for food security in Ondo State, Nigeria.*International Journal of Gender & Health Studies 1 (1): 16-21*

Omoniyi, M.B. I, & Tayo-Olajuba , O. (2006). Relief from HIV / AIDs: A Psychological Approach. *International Journal of Gender & Health Studies 4 (1&2) : 33-42*

Ramayah T, Ahmed N, AbdulHalim H, Rohaida M.Z May – Chium L.(2012): *Discriminant Analysis* ; An illustrate example; *African Journal of Bussiness Management* Vol. 4 (9) , pp 1654-1667. Online; http: //www.academic journals.org / AJBM 4th August 2012

Rencher, A.C (1998), *Multivariate Statistical Inference and Applications*, New York: John

Rencher, A.C and Christensen, W.F (2012) *Method of Multivariate Analysis,* third Edition, John

Rubin, D. B. (1987). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91*(434), 473-489.

Rubsamen -Waigman, Schroder B, Biesert L Baiermeister D, VonBriesen H, Suhartono, F, Zimmerman H.D, Brede A and Regeniter A. Staszewski (1991): *Makers for HIV - disease progression in untreated patients and patients receiving AZT: Evaluation of viral* activity , AZT resistance, serum cholesterol, B_2 micro – globulin, cd4+ cell counts and *HIV antigen.* Infection ,Vol 19, supplement -2, pp 577- 582

Steenkamp, J. B. E., &Wittink, D. R. (1994).The metric quality of full-profile judgments and the number-of-attribute-levels effect in conjoint analysis.*International Journal of Research in Marketing, 11*(3), 275-286.

The UK Collaborative Group for HIV & STI Surveillance. (2006) : A Complex Picture: HIV and other Sexually Transmitted Infections in the United Kingdom 2006. London; Health Protection Agency, Centre for Infections;

Tumer, A., Unal,S. (2000). Epiedemiology of HIV infection in the world & Turkey.

UNAIDs (2008). Report on the global AIDS epidemic .*Retrieved 22 march 2010 from http://www.Unaids.org/en/dataanlysis /epidemiology /2008 report on the global aids epidemic.*

Van der Burg, E., de Leeuw, J., & Dijksterhuis, G. (1994). OVERALS: Nonlinear canonical correlation with k sets of variables. *Computational Statistics & Data Analysis*, *18*(1), 141-163.

Vajpayee M., Amatya R. and Kauslik S etal (2004) : Lymphocyte Immunophenotypes reference ranges in healthy Indian Adults : Implications for Management of HIV/AIDS in India Clin Immonol, vol 112, pg 290 – 295.

Zanakis, S. H., Alvarez, C., & Li, V. (2007). Socio-economic determinants of HIV/AIDS pandemic and nations efficiencies. *European Journal of Operational Research*, *176*(3), 1811-1838.

APPENDIX

Descriptive Statistics Section

Type	Variable	Mean	Standard Deviation	Non-Missing Rows
X	AGE	33.14737	9.676536	1045
X	WEIGHT	54.56172	15.67153	1045
X	MARITAL_STATUS	1.050718	0.7429824	1045
Y	LITERACY_LEVEL	0.737799	0.440042	1045
Y	GENDER	0.2641148	0.4410718	1045

Correlation Section

	AGE	WEIGHT	MARITAL_STATUS	LITERACY_LEVEL	GENDER
AGE	1.000000	0.175510	0.249032	-0.082921	0.155599
WEIGHT	0.175510	1.000000	0.065830	0.017767	0.128037
MARITAL_STATUS	0.249032	0.065830	1.000000	-0.149719	-0.172445
LITERACY_LEVEL	-0.082921	0.017767	-0.149719	1.000000	0.026489
GENDER	0.155599	0.128037	-0.172445	0.026489	1.000000

Canonical Correlations Section

Variate Number	Canonical Correlation	R-Squared	F-Value	Num DF	Den DF	Prob Level	Wilks' Lambda
1	0.297167	0.088308	20.07	6	2080	0.000000	0.893519
2	0.141184	0.019933	10.59	2	1041	0.000028	0.980067

F-value tests whether this canonical correlation and those following are zero.

Variation Explained Section

Canonical Variate Number	Variation in these Variables	Explained by these Variates	Individual Percent Explained	Cumulative Percent Explained	Canonical Correlation Squared
1	X	X	28.0	28.0	0.0883
2	X	X	41.7	69.8	0.0199
1	X	Y	2.5	2.5	0.0883
2	X	Y	0.8	3.3	0.0199
1	Y	X	4.5	4.5	0.0883
2	Y	X	1.0	5.5	0.0199
1	Y	Y	50.7	50.7	0.0883
2	Y	Y	49.3	100.0	0.0199

Canonical Correlation Report

Database

X Variables	AGE, WEIGHT, MARITAL_STATUS
Y Variables	LITERACY_LEVEL, GENDER

Standardized X Canonical Coefficients Section

	X1	**X2**
AGE	0.571096	-0.779341
WEIGHT	0.384146	0.020079
MARITAL_STATUS	-0.858402	-0.466690

Standardized Y Canonical Coefficients Section

	Y1	**Y2**
LITERACY_LEVEL	0.270778	0.963006
GENDER	0.955496	-0.296192

Variable - Variate Correlations Section

	X1	**X2**	**Y1**	**Y2**
AGE	0.424748	-0.892038	0.126221	-0.125941
WEIGHT	0.427871	-0.147426	0.127149	-0.020814
MARITAL_STATUS	-0.690892	-0.659449	-0.205311	-0.093103
LITERACY_LEVEL	0.087988	0.134853	0.296089	0.955160

GENDER	0.286074	-0.038216	0.962668	-0.270683

Plots Section

Canonical Correlation Report

Database

X Variables AGE, WEIGHT, MARITAL_STATUS

Y Variables LITERACY_LEVEL, GENDER