

## Canonical Correlation Analysis, A Sine Quanon for Multivariant Analysis in Educational Research.

Iweka, Fedelis Ph.D<sup>\*</sup>, Magnus-Arewa Anthonia

Department of Educational Psychology, Guidance and Counselling, University of Port Harcourt, Rivers State, Nigeria

**\*Corresponding Author:** Iweka, Fedelis, Department of Educational Psychology, Guidance and Counselling, University of Port Harcourt, Rivers State, Nigeria

**Abstract:** This paper explains the application of Canonical correlation analysis on unavoidable necessity for multivariant analysis in educational research. Canonical analysis is the largest correlation that can be found between linear combination and the goal is to maximize the correlation (not the variance extracted) as in most other techniques. It is a method for exploring the relation between two multivariant sets of variables (vectors), all measured on the same individual. The paper highlighted when to use the canonical correlation analysis, the important assumption for the analysis and the procedure for the analysis. The objectives for the canonical correlation analysis were extensively and intensively discussed. Finally, software's to be used for canonical correlation analysis was outlined and explained.

**Keywords:** Canonical correlation, Multivariant analysis, Discriminant analysis, Marmality.

### 1. INTRODUCTION

Canonical correlation analysis is a multivariate analysis of correlation, it's a method used to ascertain the relationships between two multivariate sets of variables, and the measure is carried out on same individual, it is used to describe the relationship between the first set of variables(x) and the second set of variables (Y).

Consider, as an example, variables related to exercise and health. On one hand, you have variables associated with exercise, observations such as the climbing rate on a stair stepper, how fast you can run a certain distance, the amount of weight lifted on bench press, the number of push-ups per minute, etc. On the other hand, you have variables that attempt to measure overall health, such as blood pressure, cholesterol levels, glucose levels, body mass index, etc. Two types of variables are measured and the relationships between the exercise variables and the health variables are of interest.

The data set is split into two groups  $X$  and  $Y$ , based on some common characteristics. The purpose of canonical analysis is then to find the relationship between  $X$  and  $Y$ , i.e. can some form of  $X$  represent  $Y$ . It works by finding the linear combination of  $X$  variables, i.e.  $X_1, X_2$  etc., and linear combination of  $Y$  variables, i.e.  $Y_1, Y_2$  etc., which are most highly correlated. This combination is known as the "first canonical variates" which are usually denoted  $U_1$  and  $V_1$ , with the pair of  $U_1$  and  $V_1$  being called a "canonical function" or canonical root. The next canonical functions  $U_2$  and  $V_2$  etc are uncorrelated with  $U_1$  and  $V_1$ . PennStateEberly College of Science (n.d)

### 2. PROCEDURE

#### 2.1. Variates and Variables

There is a clear distinction in between the words variables and variates in Canonical correlation. Variables are original variables being analyzed, while variates are variables that are constructed as *weighted averages of the original variables*. Thus a set of  $Y$  variates is constructed from the original  $Y$  variables. Likewise, a set of  $X$  variates is constructed from the original  $X$  variables.

Joshua (2016) defines *Canonical Variate* as the linear combination created from the set of independent variate and dependent variate. (The term independent variable and dependent variable are arbitrary not causation).

##### 2.1.1. Canonical Weight

These are weight used to create the linear combination interpreted like regression coefficient.

### 2.1.2. Canonical Loadings

This is the correlation between each variable and its variate, interpreted like Principal Component Analysis.

### 2.1.3. Canonical Cross-Loadings

This is the correlation of each observed independent or dependent variable with opposite canonical variate.

## 2.2. Interpreting Canonical Variate

### 2.2.1. Canonical Weight

- Larger weight contributes more to the function
- Negative weights indicate an inverse relationship with other variable.
- Multicollinearity can make estimate unstable.

### 2.2.2. Canonical Loadings:

- The canonical loading gives direct assessment of each variable contribution to its respective canonical variate.
- Larger loading is important in deriving the canonical variate.

### 2.2.3. Canonical Cross-linking:

- Canonical cross-loading measures the correlation of original dependent variable with the independent canonical variate.
- Direct assessment of the relationship between each dependent variable and the independent variate.
- Provide a more pure measure of the dependent and independent variable relationship,

## 3. CONSIDERATIONS OF CANONICAL CORRELATION ANALYSIS

Small sample sizes may have an adverse effect (suggested minimum size should be 10 and above number of variable.)

### 3.1. Selection of variable to be included

- It should be on conceptual or theoretical basics,
- Inclusion of irrelevant or deletion of relevant variables may adversely affect the entire canonical solution,
- All independent variables must be interrelated and all dependent variable's must be interrelated,
- Composition of dependent variable and independent variates is critical to producing practical results.

### 3.2. Deriving Canonical Functions & Assessing Overall Fit

- Maximum number of variate functions or roots is equal to the number of variables in the smallest set of the independent variable or dependent variable.
- Variates are extracted in steps. Factor which accounts for max residual variance is selected.
  - The first pair of canonical variates has the highest intercorrelation possible.
  - Successive pairs of variates are orthogonal and independent of all variates (Joshua, 2016).

### 3.3. Assumptions for Canonical Correlation

Priya(2018) states some important Assumptions for Canonical Correlation as follows:

- One of the key assumptions that Canonical Correlation Analysis is based on is that the variables in the population should have Multivariate Normal or Gaussian distribution from which the sample was taken.

- Canonical Correlation Analysis requires a large sample size same as the multivariate regression to generate a robust model.
- Canonical correlations analysis cannot be performed if multicollinearity is found among one or more variable sets. In other words, none of the variables should have correlations equal to 1 among each other.

### 3.4. Application of Canonical Correlation Analysis

- A credit card company can apply CCA to find out the association between bank account type (Current, Savings, or Fixed Deposits) with credit cards taken.
- A healthcare research Centre can apply CCA to test the association between predictors of any disease based on medical history of patients.
- Insurance companies use CCA to test association between type of insurance policies taken like life insurance, health insurance etc. and characteristics of individuals like income, age, gender, medical background.
- Marketers use CCA in examining association between customers' demographic factors and preferences for different products.(NCSS Statistical Software n.d)

### 3.5. When to use Canonical Correlation Analysis

- It is use for descriptive techniques which can define structures in both the dependent and independent variables simultaneously.
- It is use where series of measures are used for both dependable and independable variable
- Canonical correlation is used where there is need to define structure in each variate, which are derived to maximize their correlation.(Joshua,2016)

### 3.6. Canonical Correlation Analysis Checklist

#### 3.6.1. Missing Data/Outlier

You should begin by screening your data for outliers. (*An outlier is an observation that lies in an abnormal distance from other values in a random sample from a population*). Pay particular attention to patterns of missing values. The program ignores rows with missing values. If it appears that most of the missing values occur in one or two variables, you might want to leave these out of the analysis in order to obtain more data on the remaining variables. This may flaw the result.

#### 3.6.2. Linearity

Canonical correlation analysis assumes linear relations among the variables. Scatter plots of each pair of variables should be studied, watching carefully for curvilinear patterns and for outliers. The occurrence of curvilinear relationship will reduce the effectiveness of the analysis.

#### 3.6.3. Multicollinearity and Singularity

Multicollinearity occurs when one variable is almost a weighted average of the others. Singularity occurs when this relationship is exact. Since inverse matrices are needed during the analysis, you must check for this. Try running a principal component analysis on each set of variables, separately. If you have eigenvalues at or near zero, you have multicollinearity problems. You must omit the offending variables.

#### 3.6.4. Multivariate Normality

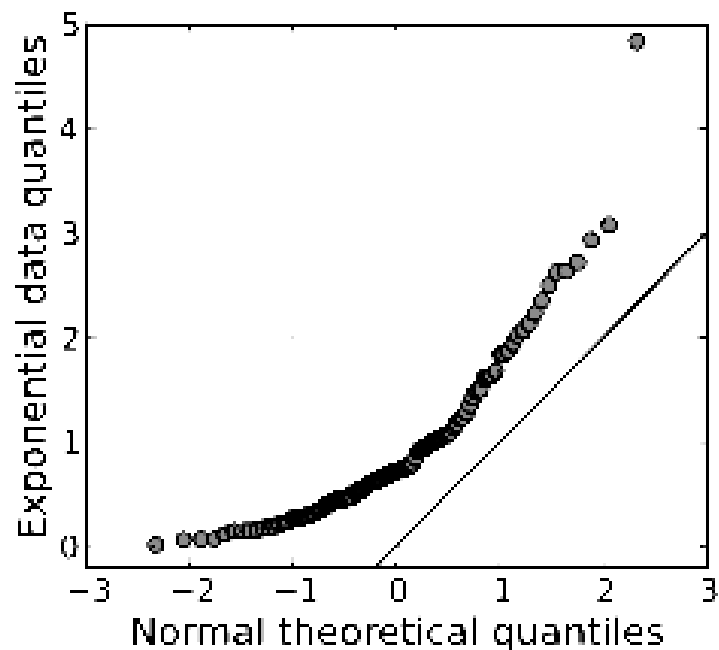
Canonical correlation analysis does not make strong normality assumptions. Assumption of normality means that you should make sure your data roughly fits a bell curve shape before running certain statistical tests or regression. It is therefore important to run a normality test. (NCSS,n.d).

## 4. HOW TO TEST FOR NORMALITY

There are two main ways to test for normality: eyeball a graph, or run a test that's specifically designed to test for normality. The data doesn't have to be perfectly normal. However, data that definitely does not meet the assumption of normality is going to give you poor results for certain types of test (i.e. ones that state that the assumption must be met).

Some of the tests can be cumbersome to use and involve finding test statistics and critical values. One of the easiest graph is the histogram. The easiest test to run is probably the Jarque-Bera Test (Stephanie, 2018).

#### 4.1. Using a Graph for a Normality Test

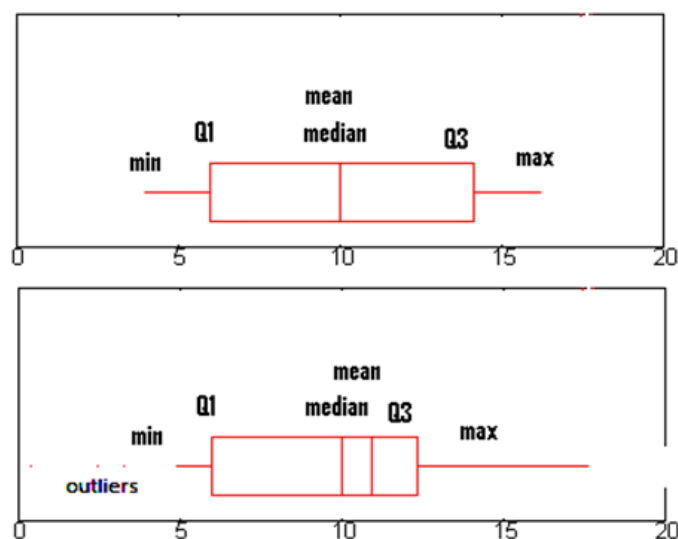


A Q Q plot showing the 45 degree reference line

A Q Q plot compares two different distributions. If the two sets of data came from the same distribution, the points will fall on a 45 degree reference line. To use this type of graph for the assumption of normality, compare your data to data from a distribution with *known* normality.

#### 4.2. Box plot

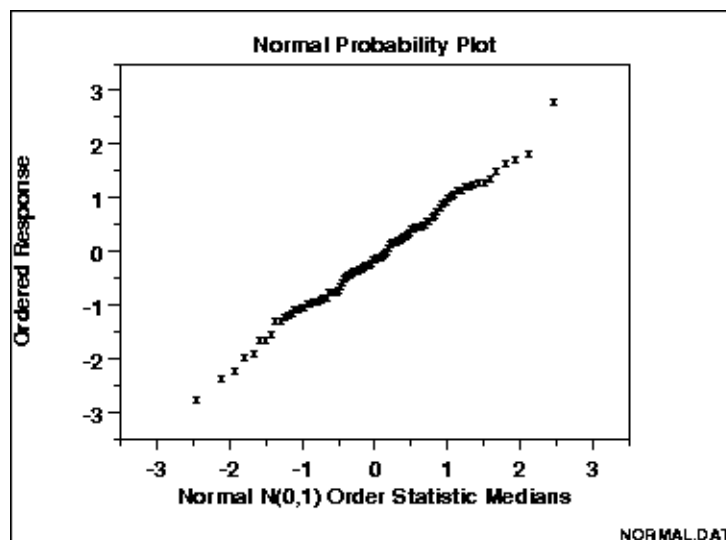
Draw a boxplot of your data. If your data comes from a normal distribution, the box will be symmetrical with the mean and median in the center. If the data meets the assumption of normality, there should also be few outliers.



Box plots for normally distributed data (top) and non-normal data (bottom).

#### 4.3. Normal Probability Plot

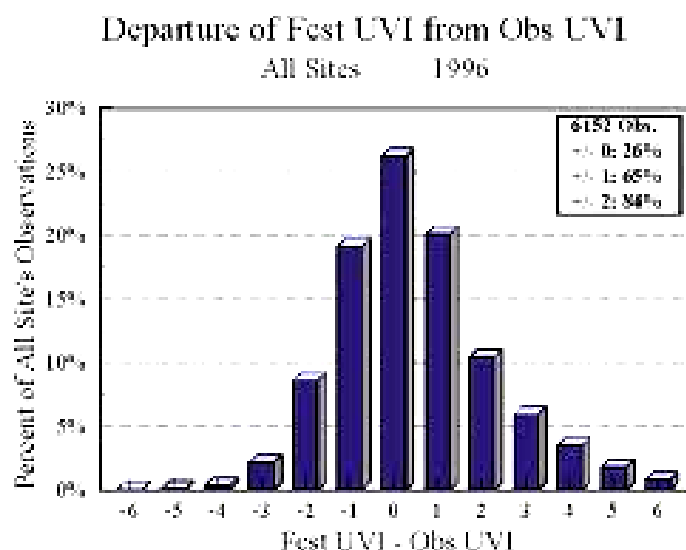
The normal probability plot was designed specifically to test for the assumption of normality. If your data comes from a normal distribution, the points on the graph will form a line.



A normal probability plot showing data that's approximately normal

#### 4.4. Histogram

The popular histogram can give you a good idea about whether your data meets the assumption. If your data looks like a bell curve: then it's probably normal.



Normally distributed data in a classic bell curve.

#### 4.5. Statistical Tests for Normality and Outliers

Stephanie (2018) states the options to test for normality. Most of these are included with statistical packages like SPSS.

- Chi-square normality test. You can use a chi square test for normality. The advantage is that it's relatively easy to use, but it isn't a very strong test. If you have a small sample (under 20), it may be the *only* test you can use. For larger samples, you're much better off choosing another option.
- D'Agostino-Pearson Test. This uses skewness and kurtosis to see if your data matches normal data. It requires your sample size to be over 20.
- Jarque-Bera Test. This common test is also relatively straightforward. Like the D'Agostino-Pearson, the basic idea is that it tests the skew and kurtosis of your data to see if it matches what you would expect from a normal distribution. The larger the JB statistic, the more the data deviates from the normal.
- Kolmogorov-Smirnov Goodness of Fit Test. This compares your data with a known distribution (i.e. a normal distribution).

- Lilliefors Test. The Lilliefors test calculates a test statistic T which you can compare to a critical value. If the test statistic is bigger than the critical value, it's a sign that your data isn't normal. It also computes a p-value for your distribution, which you compare to a significance level.
- Shapiro-Wilk. Test This test will tell you if a random sample came from a normal distribution. The test gives you a W value; small values indicate your sample is *not* normally distributed.

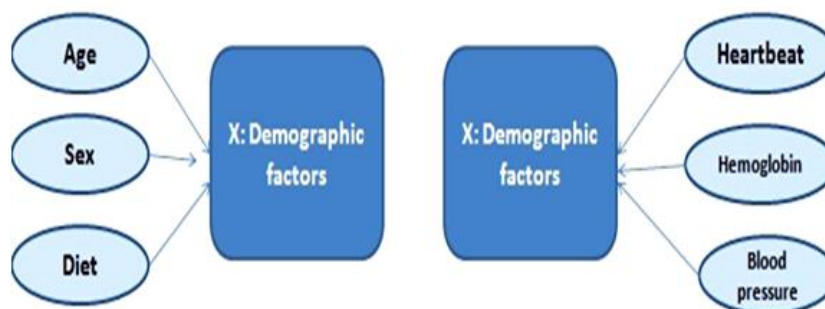
### 5. LIMITATION OF CANONICAL CORRELATION ANALYSIS

- Canonical correlation analysis reflects only the variance shared by the linear composites, not the variances extracted from variable.
- Precise statistics have not been developed to interpret canonical analysis.
- Interpretation is difficult because rotation is not possible.
- Canonical weight are subject to a great deal of instability.(Joshua,2016).

### 6. HOW CANONICAL CORRELATION ANALYSIS WORKS

Priya (2018) view Canonical Correlation Analysis (CCA),as a measure that deals with correlation among variables which are in different datasets.

For example, below figure shows two datasets:



*Graph showing canonical correlation from two different data sets*

Here, Canonical Correlation Analysis helps to estimate the possible association of different variables (like, age, sex and diet) with variables of other dataset (like heartbeat rate, hemoglobin and blood pressure). Canonical correlation analysis is based on the correlations between two sets of variables which we call Y and X

Canonical correlation analysis is based on the correlations between two sets of variables which we call Y and X.

The correlation matrix of all the variables is divided into four parts:

- $R_{xx}$  .The correlations among the X variables.
- $R_{yy}$ . The correlations among the Y variables.
- $R_{xy}$  .The correlations between the X and Y variables.
- $R_{yx}$ . The correlations between the Y and X variables.(NCSS,n.d)

#### 6.1. Canonical Variates

According to (PSEC, n. d), canonical variates have two sets of variables X and Y

Suppose we have p variables in set 1,  $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{pmatrix}$

And suppose we have q variables in set 2:  $Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_p \end{pmatrix}$

We selected X and Y based on the number of variables that exist in each set so that  $p \leq q$ . This is done for computational convenience.

We look at linear combinations of the data, similar to principal components analysis. We define a set of linear combinations named U and V. U corresponds to the linear combinations from the first set of variables, X, and V corresponds to the second set of variables, Y. Each member of U is paired with a member of V. For example,  $U_1$  below is a linear combination of the  $p$  X variables and  $V_1$  is the corresponding linear combination of the  $q$  Y variables. Similarly,  $U_2$  is a linear combination of the  $p$  X variables, and  $V_2$  is the corresponding linear combination of the  $q$  Y variables. And, so on....

$$\begin{aligned}
 U_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\
 U_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\
 U_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \\
 V_1 &= b_{11}Y_1 + b_{12}Y_2 + \dots + b_{1q}Y_q \\
 V_2 &= b_{21}Y_1 + b_{22}Y_2 + \dots + b_{2q}Y_q \\
 V_p &= b_{p1}Y_1 + b_{p2}Y_2 + \dots + b_{pq}Y_q
 \end{aligned}$$

Thus define  $(U_i, V_i)$

As the  $i^{th}$  canonical 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  $p \leq q$  there are  $p$  canonical covariate pairs.

To find the linear combinations that maximizes the correlations between the members of each canonical variate pair.

We compute the variance of  $U_i$  variables with the following expression:

$$\text{Var}(U_i) = \sum_{k=1}^p \sum_{l=1}^p a_{ik} a_{il} \text{cov}(X_k, X_l)$$

The coefficients  $a_{i1}$  through  $a_{ip}$  that appear in the double sum are the same coefficients that appear in the definition of  $U_i$ . The covariance between the  $k$ th and  $l$ th X-variables are multiplied by the corresponding coefficients  $a_{ik}$  and  $a_{il}$  for the variate  $U_i$ .

Similar calculation can be made for the variance of  $V_j$  as shown below

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

The covariance between  $U_i$  and  $V_j$  is:

$$\text{cov}(U_i, V_j) = \sum_{k=1}^p \sum_{l=1}^q a_{ik} b_{jl} \text{cov}(X_k, Y_l)$$

The correlation between  $U_i$  and  $V_j$  is calculated using the usual formula. We take the covariance between the two variables and divide it by the square root of the product of the variances:

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

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  $V_i$ :

$$P_i^* = \frac{\text{cov}(U_i, V_i)}{\sqrt{\text{var}(U_i) \text{var}(V_i)}}$$

This is the quantity to maximize. We want to find linear combinations of the X's and linear combinations of the Y's that maximize the above correlation.

## 6.2. Canonical Variates Defined

Let us look at each of the  $p$  canonical variates pair one by one.

### 6.2.1. First canonical variate pair: $(U_1, V_1)$ :

The coefficients  $a_{11}, a_{12}, \dots, a_{1p}$  and  $b_{11}, b_{12}, \dots, b_{1q}$  are selected to maximize the canonical correlation  $p_1^*$  of the first canonical variate pair. This is subject to the constraint that variances of the two canonical variates in that pair are equal to one.

$$\text{Var}(U_1) = \text{Var}(V_1) = 1$$

This is required to obtain unique values for the coefficients.

### Second canonical variate pair: $(U_2, V_2)$

Similarly we want to find the coefficients  $a_{21}, a_{22}, \dots, a_{2p}$  and  $b_{21}, b_{22}, \dots, b_{2q}$  that maximize the canonical correlation  $p_2^*$  of the second canonical variate pair,  $(U_2, V_2)$ . Again, we will maximize this canonical correlation subject to the constraints that the variances of the individual canonical variates are both equal to one. Furthermore, we require the additional constraints that  $(U_1, U_2)$ , and  $(V_1, V_2)$  are uncorrelated. In addition, the combinations  $(U_1, V_2)$  and  $(U_2, V_1)$  must be uncorrelated. In summary, our constraints are:

$$\begin{aligned} \text{var}U_2 &= \text{var}V_2 = 1 \\ \text{cov}(U_1, U_2) &= \text{cov}(V_1, V_2) = 0 \\ \text{cov}(U_1, V_2) &= \text{cov}(U_2, V_1) = 0 \end{aligned}$$

Basically we require that all of the remaining correlations equal zero.

This procedure is repeated for each pair of canonical variates. In general,

### 6.2.2. $i$ th canonical variate pair: $(U_i, V_i)$

We want to find the coefficients  $a_{i1}, a_{i2}, \dots, a_{ip}$  and  $b_{i1}, b_{i2}, \dots, b_{iq}$  that maximize the canonical correlation  $p_i^*$  subject to the constraints that

$$\begin{aligned} \text{var}(U_i) &= \text{var}(V_i) = 1, \\ \text{cov}(U_1, U_i) &= \text{cov}(V_1, V_i) = 0, \\ \text{cov}(U_2, U_i) &= \text{cov}(V_2, V_i) = 0, \\ &\vdots \\ \text{cov}(U_{i-1}, U_i) &= \text{cov}(V_{i-1}, V_i) = 0 \\ \text{cov}(U_1, V_i) &= \text{cov}(U_i, V_1) = 0 \\ \text{cov}(U_2, V_i) &= \text{cov}(U_i, V_1) = 0 \\ &\vdots \\ \text{cov}(U_{i-1}, V_i) &= \text{cov}(U_i, V_{i-1}) = 0 \end{aligned}$$

Again, requiring all of the remaining correlations to be equal zero.

(PSEC, n.d)

According to James (n.d), canonical correlation coefficients test for the existence of **overall relationships** between two sets of variables, Wilk's lambda (also called U value) and Bartlett's V are used as a *Test of Significance of the canonical correlation coefficient*.

Typically Wilk's lambda is used to test the *significance of the first canonical correlation coefficient*

Bartlett's V is used to test the *significance of all canonical correlation coefficients*

## 6.3. Software use for Canonical Correlation

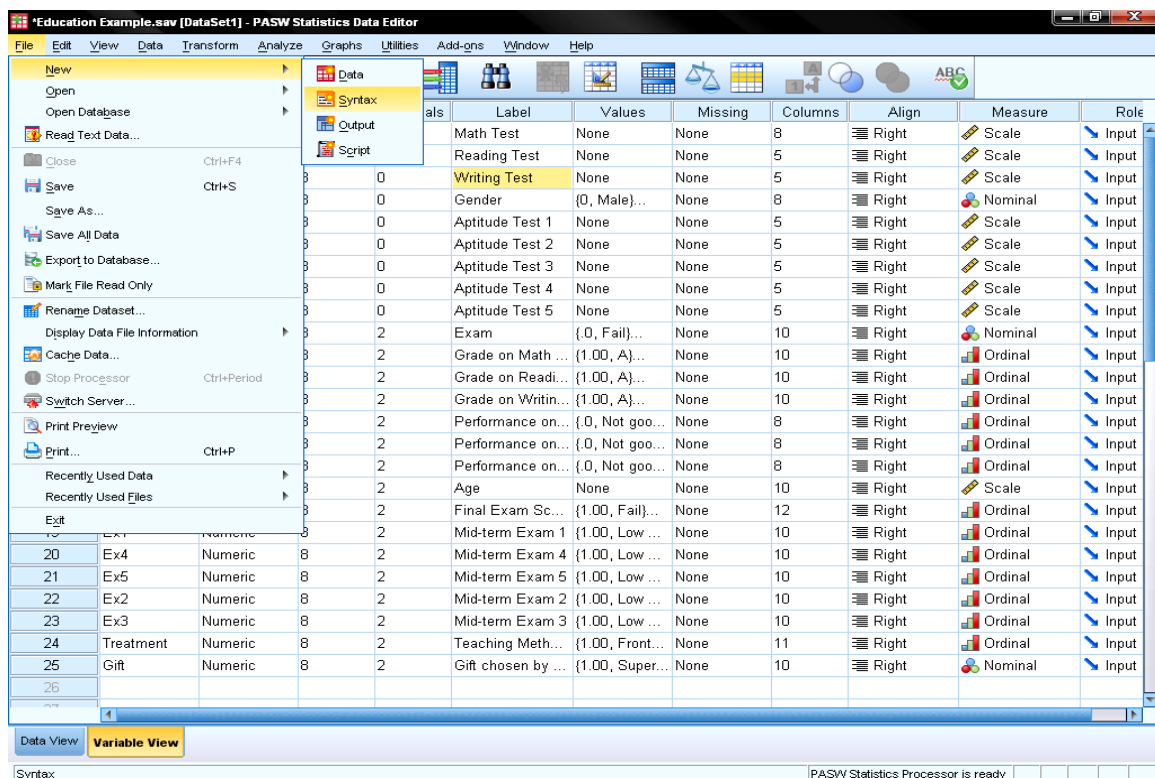
Software that supports CCA with multiple independent variables is R, SAS, MATLAB, STATA and SPSS. However, SPSS does not include a separate command for CCA; in that case, it is carried out using Syntax. (Priya, 2018).

### 6.3.1. Canonical Correlation Analysis in SPSS

James (n.d) showed the strength of association between the five aptitude tests and the three tests on math, reading, and writing using SPSS. Unfortunately, SPSS does not have a menu for canonical

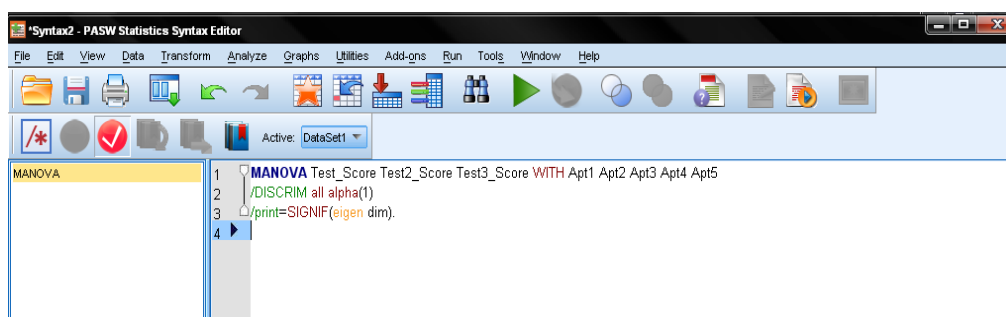


correlation analysis. So we need to run a couple of syntax commands. First, we need to open the syntax window. Click on *File/New/Syntax*.

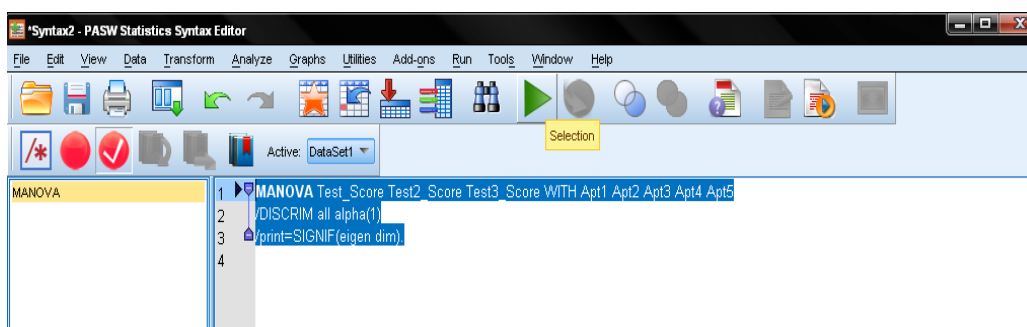


In the SPSS syntax we need to use the command for MANOVA and the subcommand `/discrim` in a one factorial design. We need to include all independent variables in one single factor separating the two groups by the `WITH` command. The list of variables in the MANOVA command contains the dependent variables first, followed by the independent variables (Please do not use the command `BY` instead of `WITH` because that would cause the factors to be separated as in a MANOVA analysis).

The subcommand `/discrim` produces a canonical correlation analysis for all covariates. Covariates are specified after the keyword `WITH`. `ALPHA` specifies the significance level required before a canonical variable is extracted, default is 0.25; it is typically set to 1.0 so that all discriminant functions are reported. Your syntax should look like this:



To execute the syntax, just highlight the code you just wrote and click on the big green Play button.



6.3.2. The Output of the Canonical Correlation Analysis

The syntax creates an overwhelmingly large output. The output starts with a sample description and then shows the general fit of the model reporting Pillai's, Helling's, Wilk's and Roy's multivariate criteria. The commonly used test is Wilk's lambda, but we find that all of these tests are significant with  $p < .05$ .

```

***** Analysis of Variance--Design 1 *****
EFFECT .. WITHIN CELLS Regression
Multivariate Tests of Significance (S = 3, M = 1/2, N = 48 1/2)

```

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.71752	6.35012	15.00	303.00	.000
Hotellings	1.98475	12.92290	15.00	293.00	.000
Wilks	.32195	9.26289	15.00	273.70	.000
Roys	.65784				

The next section reports the canonical correlation coefficients and the eigenvalues of the canonical roots. The first canonical correlation coefficients and the eigenvalues of the canonical roots. The first canonical correlation coefficient is .81108 with an explained variance of the correlation of 96.87% and an eigenvalue of 1.92265. Thus indicating that our hypothesis is correct – generally the standardized test scores and the aptitude test scores are positively correlated.

Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	1.92265	96.87114	96.87114	.81108	.65784
2	.04840	2.43863	99.30977	.21486	.04617
3	.01370	.69023	100.00000	.11625	.01351

So far the output only showed overall model fit. The next part tests the significance of each of the roots. We find that of the three possible roots only the first root is significant with  $p < .05$ . Since our model contains the three test scores (math, reading, writing) and five aptitude tests, SPSS extracts three canonical roots or dimensions. The first test of significance tests all three canonical roots of significance ( $f = 9.26$   $p < .05$ ), the second test excludes the first root and tests roots two to three, the last test tests root three by itself. In our example only the first root is significant  $p < .05$ .

Dimension Reduction Analysis

Roots	Wilks L.	F	Hypoth. DF	Error DF	Sig. of F
1 TO 3	.32195	9.26289	15.00	273.70	.000
2 TO 3	.94094	.77260	8.00	200.00	.627
3 TO 3	.98649	.46121	3.00	101.00	.710

In the next parts of the output SPSS presents the results separately for each of the two sets of variables. Within each set, SPSS gives the raw canonical coefficients, standardized coefficients, correlations between observed variables, the canonical variant, and the percent of variance explained by the canonical variant. Below are the results for the 3 Test variables.

```

EFFECT .. WITHIN CELLS Regression (Cont.)
Univariate F-tests with (5,101) D. F.

```

Variable	Sq. Mul. R	Adj. R-sq.	Hypoth. MS	Error MS	F	Sig. of F
Test_Sco	.23789	.20016	1607.60519	254.95838	6.30536	.000
Test2_Sc	.64316	.62549	2333.49702	64.09337	36.40777	.000
Test3_Sc	.37145	.34033	1035.16713	86.71635	11.93739	.000

The raw canonical coefficients are similar to the coefficients in linear regression; they can be used to calculate the canonical scores.

-----  
 Raw canonical coefficients for DEPENDENT variables  
 Function No.

Variable	1	2	3
Test_Sco	.00792	.02522	.08179
Test2_Sc	-.06959	-.07412	.00839
Test3_Sc	-.02168	.08361	-.11318

-----

Easier to interpret are the standardized coefficients (mean = 0, st.dev. = 1). Only the *first root is relevant since root two and three are not significant*. The strongest influence on the first root is variable Test\_Score (which represents the math score).

-----  
 Standardized canonical coefficients for DEPENDENT variables  
 Function No.

Variable	1	2	3
Test_Sco	.14136	.45026	1.46034
Test2_Sc	-.91034	-.96965	.10970
Test3_Sc	-.24862	.95865	-1.29763

-----  
 Correlations between DEPENDENT and canonical variables  
 Function No.

Variable	1	2	3
Test_Sco	-.57434	.60197	.55476
Test2_Sc	-.98833	-.08947	.12326
Test3_Sc	-.72991	.66990	-.13589

-----  
 Variance in dependent variables explained by canonical variables

CAN. VAR.	Pct Var DEP	Cum Pct DEP	Pct Var COV	Cum Pct COV
1	61.31476	61.31476	40.33556	40.33556
2	27.30463	88.61939	1.26055	41.59611
3	11.38061	100.00000	.15380	41.74991

-----

The next section shows the same information (raw canonical coefficients, standardized coefficients, correlations between observed variables and the canonical variate, and the percent of variance explained by the canonical variate) for the aptitude test variables.

**REFERENCES**

[1] James.L,(n,d). *Statistics Solutions*, Retrieved June 17,2018 from <http://www.statisticssolutions.com/canonical-correlation/>

[2] Joshua C. (2016). *Introduction to Canonical Correlation Analysis (CCA)* Video 1 [https://www.youtube.com/watch?v=yz5JO\\_fnGMA](https://www.youtube.com/watch?v=yz5JO_fnGMA)

[3] Joshua C.(2016). *Introduction to Canonical Correlation Analysis (CCA)* Video 2 [https://www.youtube.com/watch?v=59\\_zANrbDpQ](https://www.youtube.com/watch?v=59_zANrbDpQ)

[4] NCSS Statistical Software (n.d) *Chapter 400 Canonical Correlation*[https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Canonical\\_Correlation.pdf](https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Canonical_Correlation.pdf)

[5] PennStateEberly College of Science (n.d) *Applied Multivariate Statistical Analysis STAT 505 Lesson 13, Canonical Correlation Analysis*, Retrieved June 9,2018 from <https://newonlinecourses.science.psu.edu/stat505/node/63/>

[6] Priya C. (2018). *Performing Canonical Correlation Anaylsis (CCA)*, Retrieved June 15,2018 from <https://www.projectguru.in/publications/performing-canonical-correlation-analysis-cca/>

[7] Stephanie G. (2018). *Probability and Statistics Topic Index*, Retrieved June 16, 2018 from <http://www.statisticshowto.com/assumption-of-normality-test/>

**Citation:** Iweka, Fedelis Ph.D, Magnus-Arewa Anthonia. "Canonical Correlation Analysis, A Sin Quanon for Multivariant Analysis in Educational Research." *International Journal of Humanities Social Sciences and Education (IJHSSE)*, vol 5, no. 7, 2018, pp. 116-126. doi: <http://dx.doi.org/10.20431/2349-0381.0507013>.

**Copyright:** © 2018 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.