Study Some Risk Factors of Spontaneous Abortion by Using Canonical Correlation Analysis

Dr. Abbas Gulmurad Beg Murad

University of Sulaimaniyah / Statistics & Informatics Dept.

Iraq-Sulaimaniyah

Email: abbas.beg@univsul.edu.iq

Dr. Sozan Saber Haider

University of Sulaimaniyah/ Statistics & Informatic Dept.

Iraq-Sulaimaniyah

Email: sozan.haider@univsul.edu.iq

دراسة بعض عوامل الخطر للإجهاض التلقائي باستخدام تحليل الارتباط الكنسي

د.عباس كولمراد بك مراد

جامعة السليمانية/ قسم الإحصاء والمعلوماتية/ العراق – السليمانية

البريد إلالكتروني: abbas.beg@univsul.edu.iq

د.سوزان صابر حيدر

جامعة السليمانية / قسم الإحصاء والمعلوماتية/ العراق – السليمانية

البريد إلالكتروني: sozan.haider@univsul.edu.iq

Abstract:

Canonical correlation analysis (CCA) means a involving several dimensions investigative statistical procedure that operates on the same concept as the principal component analysis. The primary occasion of it is the expedition of correlations between two sets of variables on the identical exploratory items. Furthermore, the PCA technique deals with one case study data only. It tries to decrease the general dimensional of the sample data using several linear collections of the primary variables.

Some more packages can be downloaded and installed likewise (mainly the R programming codes insides packages CCA, CCP, insides CC).

The case study be formed of two sample data where each dataset represents (100) pregnant women localities in the Sulaimani General Hospital. The first dataset contains measurements of biological metrics of two variables (Type of abortion, duration of abortion in weeks), and the second dataset contains measurements of three variables (wife's age, Occupation, and residence).

The main idea is to by some means correlate both sets to explain what biometrics can correlate which each other. The way of replying this is to implement canonical correlation procedure described with additional tools to expedite the understanding of the outcomes in this study.

Keyword: Canonical Correlation Analysis (CCA), canonical variates, Canonical variation, Sample Correlation Matrix, Principal Components Analysis (PCA).

الملخص:

تحليل الارتباط القويم هو نهج إحصائي استكشافي متعدد الأبعاد يعمل بنفس مبدأ تحليل المركبات الرئيسة. والغرض الرئيسي منه هو كشف الارتباطات بين مجموعتين من المتغيرات في نفس الوحدات التجريبية. من ناحية أخرى، تتعامل طريقة PCA مع مجموعة بيانات واحدة فقط وتحاول تقليل الأبعاد الكلية لمجموعة البيانات باستخدام مجموعة خطية من المتغيرات او عوامل الخطورة، في برمجة البرمجة الإحصائية باستعمال لغة الـ(R)، يوجد رمز وظيفي متاح في إطار تثبيت الحزم القياسية، تتضمن الدراسة مجموعتي بيانات حيث تمثل كل مجموعة بيانات قياسات لعينة من (١٠٠) تشمل بيانات للنساء الحوامل في مستشفى السليمانية العام، حيث تحتوي مجموعة البيانات الأولى على متغيرات (عوامل خطورة) لمتغيرين

(نوع الإجهاض، مدة الإجهاض بالأسابيع) بينما مجموعة البيانات الثانية تحتوي على قياسات لثلاثة متغيرات (عمر الزوجة ، الوظيفة والإقامة داخل المدينة او خارجها).

الكلمات المفتاحية: تحليل الارتباط الكنسي (CCA) ، المتغيرات الأساسية ، الاختلاف الكنسي ، نموذج مصفوفة الارتباط ، تحليل المكونات الرئيسية (PCA).

1. Introduction

Canonical correlation analysis means the seeking of the linear association among two bulks of variables [11], Canonical Correlation Analysis (CCA), proposed by Hotelling in 1936 [3], Canonical Correlation Analysis is a statistical Approach to discover correlations between two data sets acquired on the same experimental units. Some statistical packages like cancor, CC, CCP functions in R (R Development Core Team 2007) with related other parts performing the root of manipulating but more work was needed to support the users with more tools to facilitate the explanation of the outcomes. We performed an R package, CCA obtainable from the R Archive Network (CRAN, http://CRAN.R-project.org/). Canonical correlation analysis packages also integrates version of CCA to work with case study data collections with several variables than units we practiced its.

Canonical correlation analysis is a multi-dimensional statistical technique is the same with Principal Components Analysis (PCA); both techniques have the same mathematic bases (matrices algebra and Eigen (latent) values derivation), and outcome can be explained through analogous graphic representation, The main aim of it is the exploration of sample correlations between two sets of Scale variables on the experimental units itself [5], as the figure in bellow:

Fig.1
The System of Classical CCA
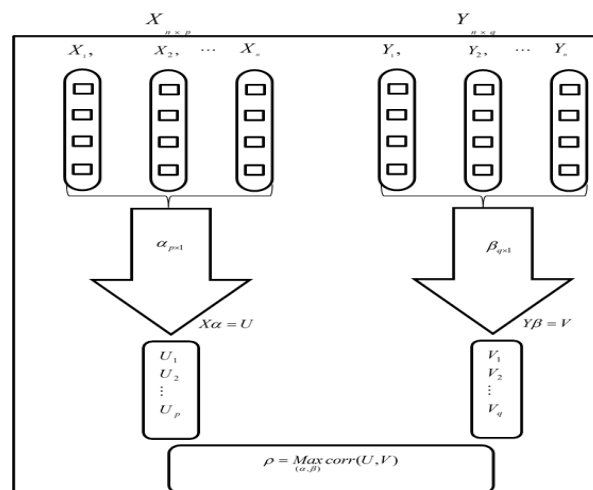
Practically, number of experimental units has to more than the total aggregate of variables in both data samples, which is not constantly potentially. The other variables are gained on these same experimental units to appear the correlations; the Classical CCA programming package in R results cannot be performed. Canonical Correlation Analysis has been implemented favorably in several tasks such as dimensional reduction. That means we need other packages to use it to make the statistical results more transparent. [3].

Furthermore, the target of Canonical Correlation Analysis is to discover the correlations between two data samples of variables to a common dimension space in which they are farthest correlated whose functions are accurately symmetric in the statistical analysis, As a result, the mathematical bases of the two techniques are completely different between (CCA & PCA) [5], and statistical software usually proposes Canonical correlation analysis calculation (proc cancorr in SAS, cancor in R,. . .). We observed the cancor results are restricted and prerequisite to be completed to be used in active techniques (used codes). We used an R package to achieve numerical as well as graphic outputs; in this paper, we used codes by details (step by step) to write the program with some packages to analyze the data set; this dataset included two sets, the first set included variables (duration of abortion, type of abortion), while the second set included variables (wife's age, Occupation, and residence).

2. An Introduction to Canonical correlation analysis [4],[2]

Let we have supplied a two sets of multivariate variables $S = \{(x_1, y_1), \cdot \cdot \cdot, (x_n, y_n)\}$ of (x, y), since $xi \in R\ d_x$, $y_i \in R\ d_y$ and n is the size of the sample. We do S_x to denote $(x_1, \cdot \cdot \cdot, x_n)$ and similarly Sy to denote $(y_1, \cdot \cdot \cdot, y_n)$. The construction of the suggested model is displayed in Figure Let we have two matrices M and N of dimensions (n × p) and (n × q), respectively. Each column of M and N match with the variables, and the rows match with the experimental units. column jth of the matrix X is expressed by Xj; as well the kth column Yk is indicate to Y. in the absence of a lack of generality, let we assumed the columns of M and N are standardized (zero mean & constant variance); by it assumptions that $p \leq q$ (in another meaning, the sample contains the minor variables is represent by X). And S_{XX} and S_{YY} represent by the sample covariance of the two samples for variable samples M and N, respectively, with mean vectors μ_1 and μ_2, with our assumptions $\mu_1 = \mu_2 = 0$. And the sample

covariance matrix between X and Y represent by SXY; and the figure in below explains that. [11]
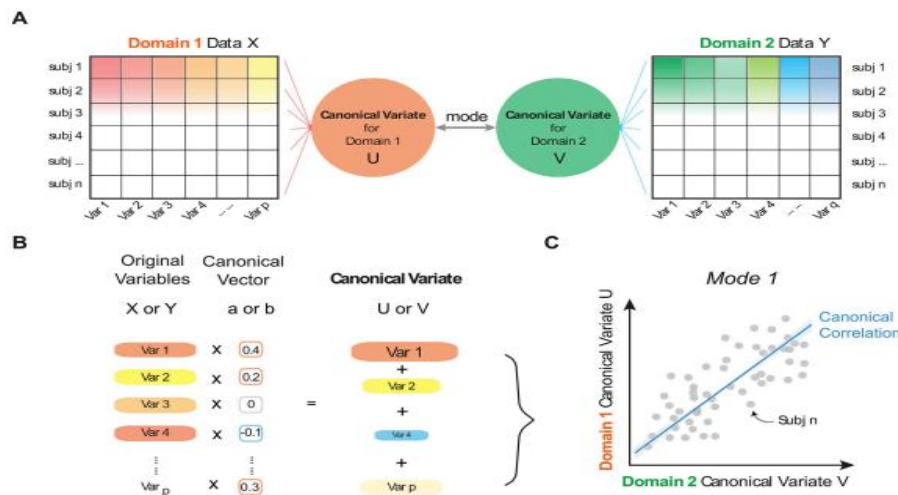

Fig. 1.
A general schematic for (CCA) [1]

With mean vectors μ_1 and μ_2 and Σ_{11} and Σ_{22} as a covariance matrices, with our assumptions $\mu_1 = \mu_2 = 0$.

Symmetry was an essential feature of CCA leads the two variable sets are mutually exchangeable. Swapping the left and right sides yields exact estimates of canonical variates and canonical correlations. Put differently, the classical notions of dependent variable or (response variable) usually denoting the model output and independent variables (explanatory variables) usually represent the model input; if we exchange the context of CCA; the canonical correlation remains identical because the underlying linear correlation is asymmetric bi-variate metric; One unit something change in one variable is related to another variable in another set of observations [11]

CCA has several advantages; it limits the probability of rejecting the null hypothesis (H_0) when it is true. It means it reduces the probability of doing Type I errors. The danger of a Type I error is linked to the probability of obtaining a decision was statistically significant When it does not there. Risen danger of error from I type results from the selfsame variables in the data are used also for several statistical tests.

It can identify two or more unique relationships; if they exist, CCA develops multiple canonical functions; these functions are orthogonal with respect to the other functions so that they depict different relations among variables. [7]

3. The Difference between CCA and PCA [11], [7]

There is a modest variation between CCA and PCA; CCA is applied to correlate simultaneously some metric dependent variables and several metric independent variables measured on the experimental units itself, PCA is often concerned with reducing the data dimensionality of a sole dataset by having a few linear combinations of the variables, The technique is an expansion of multi correlation analysis and is often applicable in the same situations in which multivariate regression analysis methods would be appropriate.

Before we get started with the analysis in R, let's review some concepts related to Canonical correlation analysis.

4. Purposes of Canonical Correlation Analysis [5]

There are two main purposes of Canonical Correlation Analysis:

i. Data Reduction

Canonical correlation analysis presents a way for describing the relationship between 2 sets of variables applying linear combinations of these variables.

ii. Data interpretation

In general multivariate analysis, the technique is also of particular interest when it comes to finding features (canonical variates) that are statistically significant in terms of explaining the covariance between the given sets of variables; CCA was never changing to changes of scale on the response variables (y) and the explanatory variables (x), changing the scale of measurement of the two sets of variables of interest in the analysis, for instance, from inches to centimeters does not interfere with the canonical correlations that follow. Also, the first canonical correlation (r_1) is the maximum correlation between the linear functions of Y and X.

Fig. 2.

Scheme explains the uses of CCA types [5]

5. Assumptions of Canonical Correlations

There are some assumptions about Canonical Correlation we can explain it briefly as follow:

If canonical correlation analysis has multicolinearity problem, we can used Partial Least Squares (PLS) regression to reduce the spaces, in same way for Hetroscedasticity problem, and transformation the data if Normality isn't found, or re-sampling again.

7. Principle of Classical Canonical Correlation derivation [9], [7]

Where first $p \leq n$ and $q \leq n$, the two matrixes M and N are of full rank equal to p and q respectively. In the following, the basis of CCA is approaching as a problem solved along iterative algorithm. First stage of Canonical correlation be formed to detection two vectors $a_1 = (a_{11} , \ldots , a_{1p})$ and $b_1 = (b_{11} , \ldots , b_{1q})$, We attempt to maximize the correlation between the two linear combinations $U_1 = Xa_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p$ and $V_1 = Y b_1 = b_{11}Y_1 + b_{12}Y_2 + \cdots + b_{1q}Y_q$, and the statistical analysis employ (var-cov) matrixes of the two variables. For example, the analysis is based on a matrix of the form:

$$\begin{bmatrix} \sum_x & \sum_{xy} \\ \sum_{yx} & \sum_y \end{bmatrix}$$

By assumptions the two vectors a_1 and b_1 are normalized vectors (have orthogonality property) that means var $(U_1) = $ var$(V_1) = 1$. In another meaning, the problem be formed in solving

$\rho_1 = \text{cor}(U_1, V_1) = \text{max a,b cor}(Xa, Yb)$, correlated with the two constraint var(Xa) = var(Yb) = 1 . The two variables U_1 and V_1 are named the first canonical variates and ρ_1 is make mention of the first canonical correlation [12].

Canonical correlation (ρ) between two sets of multivariate random variables X and Y is the covariance (Cov) between the two variables normalized by the geometric mean of the variances (Var) of X_α and Y_β; we beginning with:

$$\rho = \frac{Cov(X\alpha, Y\beta)}{\sqrt{V\,ar(X\alpha)V\,ar(Y\beta)}} \qquad \ldots (1)$$

A multivariate multiple regression representation of CCA amounts to finding an estimate of β, α and D(ρk) in the following model equation:

$$Y_\beta = X_\alpha D(\rho_k) + E \qquad \ldots(2)$$

Where D(ρk) is a d × d diagonal matrix of canonical correlation, Consider N observations on two sets of standardized variables X and Y:

$$X = \{x_{ij}\}; \qquad i = 1, 2, 3, \ldots , N, j = 1, 2, 3, \ldots , p$$

$$Y = \{y_{ij}\}; \qquad i = 1, 2, 3, \ldots , N, j = 1, 2, 3, \ldots , q$$

The matrices of correlations R among the X variables and the Y variables, and between the two sets of variables are R_{XX}, R_{YY} and R_{XY}, respectively.

CCA seeks to maximize the correlation (ρ) between Xα and Y β; ($\alpha_{p\times1}$ and $\beta_{q\times1}$) s the variance , Var(Xα) = Var(Yβ) = 1, Covariance between U and V can be equivalently expressed as:

$$\rho = \frac{\alpha^j R_{XY} \beta}{((\alpha^j R_{XX} \alpha)(\beta^j R_{YY} \beta))^{1/2}} \qquad \ldots (3)$$

Maximize the quantity αj RXY β subject to the constraints αj RXXα = βj RY Y β = 1: Introduce Lagrangian multipliers, then compute matrix derivatives, set them to zero and simplify. The solution leads to the normal equations:

$$R_{XY} \beta - KR_{XX}\alpha = 0 \qquad \ldots (4)$$
$$R_{YX}\alpha - KR_{YY} \beta = 0 \qquad \ldots (5)$$

The values of K are obtained by solving the multivariate Eigen value problem:

$$\lfloor R_{XX}^{-1} R_{XY}R_{YY}^{-1} R_{YX} - K^2 I\rfloor = 0 \qquad \ldots (6)$$

Or

$$[R_{YY}^{-1} R_{YX} R_{XX}^{-1} R_{XY} - K^2 I] = 0 \qquad \qquad \dots (7)$$

The positive squared root of the highest Eigen value gives the largest correlation, Further, the values of α and β can be obtained from Equations (7) and (8) as follows:

$$\beta = \frac{1}{\rho} R_{XX}^{-1} R_{YX}\, \alpha \qquad \qquad \dots (8)$$

Substituting (8) in Equation (5) and rearranging terms we get:

$$(R_{XX}^{-1} R_{XY} \textstyle\sum_{YY}^{-1} R_{YX}\, \alpha - K^2 I)\alpha = 0 \qquad \qquad \dots (9)$$

Similarly

$$(R_{YY}^{-1} R_{YX} \textstyle\sum_{YY}^{-1} R_{XY}\, \alpha - K^2 I)\beta = 0 \qquad \qquad \dots (10)$$

The primary CCA form is given back to the linear combination between of the variables in X with linear combination for variables in Y as follow:

$$U = a' X \qquad ; \quad a \in R^p$$
$$V = b' Y \qquad ; \quad b \in R^q$$

in order to maximizes the first mode linear association, that is, correlation

$$\rho = corr(U, V) = corr\,(a^T X, b^T Y). \qquad \qquad \dots (11)$$

8. Types of Spontaneous Abortion (miscarriage) [10], [1]

The Miscarriage problems are very common complex problem, happening near from 30% an estimated of all pregnancies in general. But the loss of pregnancy is often imperfectly explained and combines with motivated abortion problems; for instance, terms "induced miscarriage" used to denote intent to terminate the pregnancy, that "spontaneous abortion" is medical synonym of miscarriage. There are various categories of Abortion (miscarriage) approached, complete, incomplete, inevitable, or missed; we covered it in this research.

i. Complete miscarriage

A complete spontaneous abortion occurs when the pregnancy tissue has passed out of the uterus. Vaginal bleeding may be go on with for multi

days. Colic pain, like labor or severe menstrual pain, is common, this is the uterus convening to empty.

ii. Incomplete miscarriage

Some pregnancy tissue will remain in the uterus. Vaginal bleeding and lower abdominal cramping may continue as the uterus continues trying to empty itself. This is known as an 'incomplete miscarriage.

In the uterus will remain several of the pregnancy tissue. Lower abdominal and Vaginal bleeding cramps may go on with uterus continues to try to empty itself. This is recognized as an "incomplete" miscarriage The physician or obstetrics will want to be towed to evaluate if a short procedure called cervical dilation and curettage (often recognized as a "D, C") is requisite to remove any pregnancy tissue that remaining. This is significant medical procedure that is performed in the operating room.

iii. Missed miscarriage

At times the baby dies but remains in uterus. This is named (missed miscarriage). Some of the miscarriage, like (nausea and fatigue) may have faded. There is may be no any notification remarkable.

iv. Threatened miscarriage

Sometimes the body shows signs of a possible spontaneous abortion, this is named (threat of miscarriage). It has low vaginal bleeding or light abdominal pain. It can latest weeks, and the cervix nevertheless locked.

The bleeding with the pain may step out, and can carry on with pregnancy and the baby in good health. Or things could go wrong and lead to a spontaneous abortion.

v. Inevitable miscarriage

Inevitable spontaneous abortion can came next a threatened miscarriage or in the absence of excitation. In general a lot more vaginal bleeding and strong lower stomach cramps. During the spontaneous abortion, the cervix opens, and the developing fetus will come away in the bleeding.

vi. Recurrent miscarriage

A few women experienced recurrent spontaneous abortion. If the third or more miscarriage in a row, it is best to discussing it with physician, who may be qualified for examine the risk factors and refers to a specialist.

9. Data analysis by R

We can bring out a CCA in R insides CC, CCP and heplots packages, It is freely available from the extensive R archive network (CRAN,

http://CRAN. R-project.org/). Once loaded into R, but we depend on writing a program by user R codes and packages provides the user with functions to perform Canonical correlation analysis of the data set for illustration purposes; the data measurements were gathered for (100) pregnant women from Sulaimani General Hospital. was cleaned of 5 variables, and these variables separated two sets of measures; the first set included variables (duration of abortion (by weeks), type of abortion (complete, incomplete, Inevitable miscarriage and missed), while the second set included variables (wife's age(less than 30 years, bounded between 30 and 40 or more than 40 years), Occupation (Housewife or Employee) and residence (downtown or Rural), We performed a CCA, which makes us deep meaning to comprehend the canonical variable pair that we obtained as results, thus, a common latent variable beyond the two measurements. The main objective of CCA analysis is to discover the covariate that has more correlation, before conducting the analysis; we present some descriptive statistics in the following tables and figures.
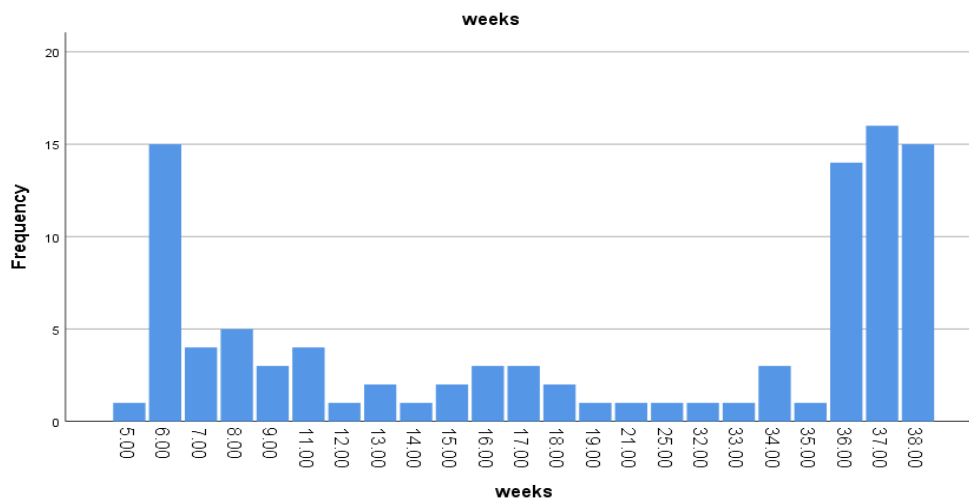

Fig. (4)

Describes spontaneous abortion during the lumen period (in weeks)

From the above figure, we note that the highest miscarriage appears in the sixth week of pregnancy, as well as between weeks (36-38) through the study sample.
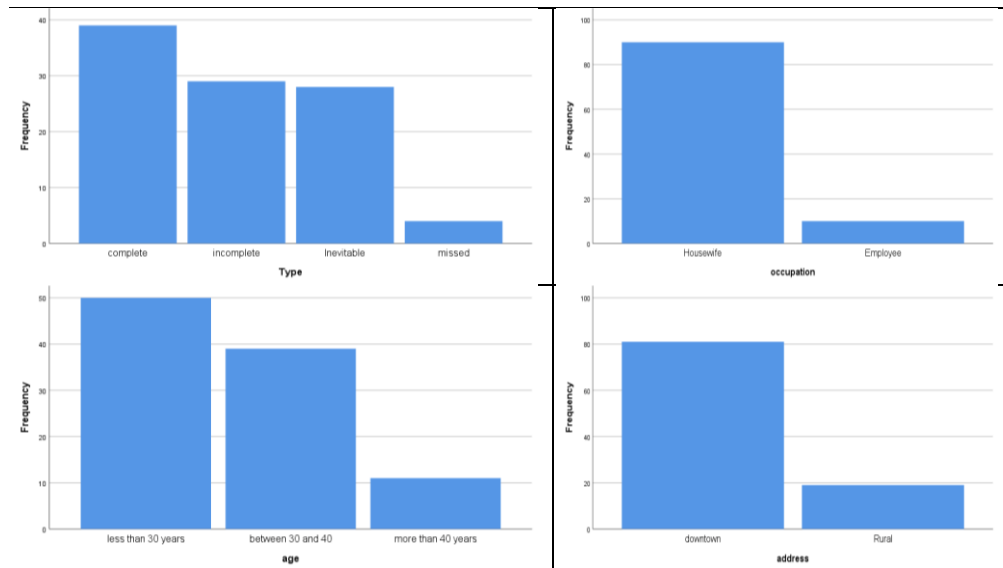
Fig. 5

Show the bar-chart for (Type of abortion, Occupation, Age and The Address)

we note that complete abortion is the highest compared to the rest of the three types of abortion type variable, as well as for the Occupation variable, the housewife is much higher than the employee, as for the age variable, women who are less than 30 years old are the highest in relation to the rest, finally With regard to housing, the women who live inside the city are the highest in concern for the women who live outside the city for the our sample.

Table 1

Shows test of wilk by usingapproximation F test (Rao's F)

	statistics	approximation	df_1	df_2	P-value
1 to 2:	0.9050499	1.619677	6	190	0.1436474
2 to 2:	0.9612434	1.935325	2	96	0.1499689

Table 2

Shows Canonical Loadings for set 1 and set 2

Set 1			Set 2		
Variable	1	2	Variable	1	2
occupation	.585	-.476	Type	-.497	-.868
address	-.118	.809	weeks	-.896	.443
age	.834	.379			

The (CCA) estimates the grade of linear association between two groups of variables. The correlation coefficients measured in CCA on a par with the number of variables in small group: m = minimum (p, q). The coefficients are in go down order of size: $r_1 > r_2 > r_3 > r_4 > ... > r_m$. exclude for $t_{stat}= r$, the p.asym function calculates the m p values for everyone test; first p-value was computed included all canonical correlation coefficients, second p-value was computed by excluded r_1, in same way the third p-value was computed by excluded r_1 & r_2, and so on.(the full program coding is in Appendix No. one).
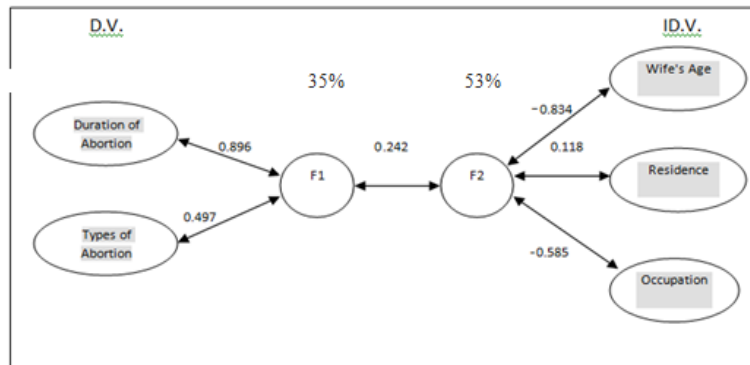

Fig. 6

Canonical correlation and loading of Spontaneous abortion problem

The correlations in Fig (6) are among independent variables and dependent variables, which recognized it as canonical loadings. The canonical variate is a type of latent variable; generally, the number of canonical size is equivalent to the number of variables in the more modest group; while the number of significant variables maybe evens less. Canonical dimensions, also known as canonical variates, are latent variables analogous to factors obtained in factor analysis (PCA).

Here we verify the canonical correlation between the first and second group of variables, its equal to (0.242), the canonical correlation are explained in a same manner to explanation of regression modeling, that

means when (wife's Age) in above figure decreasing leads to (duration & type of abortion) increasing because wife's age has negative sign, when all other variables are caught constant, in same way when residence variable tends to rural level more than downtown leads to (duration & type of abortion) increasing together because residence variable has positive sign, Being Occupation leads to (duration & type of abortion) decreases by (0.585), with the other predictors held constant.

10. Conclusions

CCA give us a simple and effective technique of prescribing the correspondence between the two variable groups (Type of abortion, duration of abortion in weeks), and the second dataset contains (wife's age, Occupation, and residence) Can be useful in the description the complex and nested relationship in Spontaneous abortion problem groups, It is appear from figure (6) Low correlation (r = 0.242) between independent set one set two insides tables (1&2), The demand of Canonical correlation analysis is probable to increase by details the simultaneously multi-modal in abortion problem data groups. in many uses, Canonical correlation analysis help as the central of the analysis flow. Give us variations and capacity become a building block for more detailed lines of data analysis, rather than the subject matter of the analysis itself, Our interest is in determining the number of dimensions (canonical variables in Abortion problem) that are significant or not to explain the relationship between the two sets of variables of the abortion problem.

11. Recommendations

There are some alternative techniques and methods made important extensions of the canonical correlation analysis modeling; each one has been improved performance subject to specific analytical cases with the following recommendations.

1. We recommend taking other risk factors that were not considered in this research, for example (chronic diseases, problems of the uterus or cervix, smoking, alcohol consumption and drug abuse ... etc.).

2. We recommend the statistics departments in Iraqi hospitals in general and the Kurdistan Region in particular to improve the data that is taken from the patient in order to provide data that has more relevance to the patient's condition in order to conduct useful studies and researches.

3. Principal components analysis has some likeness with canonical correlation analysis, in spite of the fact that Principal components

analysis implements matrix analysis of a single variable set. The common properties of Principal components analysis and canonical correlation analysis are the orthogonal restriction forced during the detection of the structure, assumed there are no covered sources of difference (the patterns) do not correlate with each other in both directions.

4. Analogous to Principal components analysis and canonical correlation analysis, there is independent component analysis (ICA) also excerpt the hidden changes in sets of potential variables. however canonical correlation analysis is interested with discovering the sources of variance depend upon the structure of linear variance, ICA determine more complex nonlinear associations in data that can hold statistical analysis further on differences in means and variances.

5. Partial least squares (PLS) regression is further comparable to canonical correlation analysis more than Principal components analysis or independent component analysis. because Partial least squares and canonical correlation analysis can distinguish latent dimensions of variance over two variable groups, and the main characteristic advantage of Partial least squares is the aim of optimal mode to reduce the covariance replaced than linear correlation.

6. Further calculations of CCA solutions can be done using under libraries in R there some packages (PMA , lme4), The package (canocorr) built in MATLAB and in python programming there is scikit in machine learning library (sklearn cross decomposition.CCA) can be used.

Reference

1. Cohain JS, et al, (2017), "Spontaneous first trimester miscarriage rates per woman among parous women with 1 or more pregnancies of 24 weeks or more". DOI: 10.1186/s12884-017-1620-1.

2. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J., 2009, "Modeling wine preferences by data mining from physicochemical properties", Decision Support Systems, 47(4), 547–553.

3. D. P. Foster, S. M. Kakade, & T. Zhang, 2008, "Multi-view dimensionality reduction via canonical correlation analysis". Penn labraries-University of Pennsylvania.

4. González, I., Déjean, S., Martin, P. G., & Baccini, A. , 2008, "CCA: An R package to extend canonical correlation analysis. Journal of Statistical Software", 23(12), 1–14.

5. H. Hotelling, 1936, "Relations between two sets of variates", Biometrika.

6. Hao-Ting Wang et al, 2020, "Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists" , Elsevier.

7. Idowu Ayodeji & Titi Obilade, 2016, "A Predictive Model for Canonical Correlation Analysis with Implications for the Simple and Multiple Correlations", researchgate.net/publication /318877866.

8. Ignacio González et al, 2020, "An R package to extend canonical correlation analysis", Journal of Statistical Software.

9. Takoua Jendoubi & Korbinian Strimmer, 2019, "A whitening approach to probabilistic canonical correlation analysis for omics data integration", Jendoubi and Strimmer BMC Bioinformatics 20:15 https://doi.org/10.1186/s12859-018-2572-9.

10. The Royal Women's Hospital (Miscarriage), 2019, "NSW Health (Early Pregnancy: When Things Go Wrong)", development and quality assurance of health direct content.

11. Tomasz Górecki et al, 2020, "Generalized canonical correlation analysis for functional data", sciendo.

12. Wenqi Fan & et al, 2020, "Deep Adversarial Canonical Correlation Analysis", Society for Industrial and Applied Mathematics.

Appendix No.1

```
#loading the required packages
library(CCA) #facilitates canonical correlation analysis
library(CCP) #facilitates checking the significance of the canonical variates
library(foreign)
library(CC)
library(CCP)
```

```
library(heplots)
rm(list = ls())
abodata1<- read.spss("C:/Users/asa/Desktop/abortion.sav",use.value.label=TRUE, to.data.frame=T)
abodata1
head(abodata1)
dim(abodata1)
summary(abodata1)
y1 <- abodata1[,2]
y1
y2 <- abodata1[,3]
y2
x1 <- abodata1[,7]
x1
x2 <- abodata1[,5]
x2
x3 <- abodata1[,6]
x3
DV<-cbind(y1,y2)
dim(DV)
DV
IDV<-cbind(x1,x2,x3)
IDV
cormat<-matcor(DV,IDV)
cormat
y.stand<-sweep(DV,2,sqrt(apply(DV,2,var)),FUN="/")
y.stand
x.stand<-sweep(IDV,2,sqrt(apply(IDV,2,var)),FUN="/")
x.stand
R11<- cor(y.stand)
R11
R22<- cor(x.stand)
R22
R12<- cor(x.stand,y.stand)
R12
R21<- t(R12)
R21
E1<-solve(R11)%*%(R21)%*%solve(R22)%*%(R12)
E1
E2<-solve(R22)%*%(R12)%*%solve(R11)%*%(R21)
E2
eigen(E1)
eigen(E2)
# display the canonical correlations
can.cor<-sqrt(eigen(E1)$values)
can.cor
# display the canonical correlations
cc1<-cancor(DV,IDV)
cc1
# display the canonical correlations
cc2<-cc(DV,IDV)
cc2
correl=matcor(DV,IDV)
img.matcor(correl)
img.matcor(correl,type=2)
res.cc=cc(DV,IDV)
plt.cc(res.cc)
rho <- cancor(IDV,DV)$cor
```