Study of the Relationship between Dependent and Independent Variable Groups by Using Canonical Correlation Analysis with Application

Thanoon Y. Thanoon[1,3], Robiah Adnan[1] & Seyed Ehsan Saffari[2]

[1] Department of Mathematical Science, Faculty of Science, Universiti Teknologi Malaysia, Malaysia

[2] Education Development Centre, Sabzevar University of Medical Sciences, Sabzevar, Iran

[3] Department of Operation Management Technique, Technical College of Management, Foundation of Technical Education, Mosul, Iraq

Correspondence: Thanoon Y. Thanoon, Department of Mathematical Science, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia. E-mail: Thanoon.younis80@gmail.com/ robiaha@utm.my/ehsanreiki@yahoo.com/

Abstract

Canonical correlation analysis is used to study the relationship between two groups of variables (dependent and independent). Since each group represents the linear combination to a number of variables, canonical correlation analysis measures the relationship between these variables that maximally correlate with linear combinations of another subset of variables. Statistical analysis involves canonical correlation between two groups of variables, canonical variates, standard canonical variates, canonical factor loadings, canonical cross factor loadings for both groups. Test of significance of canonical correlation using Wilk's Lambda showed that the first and second canonical correlation was significant and the third and fourth canonical correlation were insignificant. This method is illustrated by using a real data set. Results obtained by using SPSS program.

Keywords: canonical correlation analysis, dependent variables, independent variables, canonical variates

1. Introduction

Canonical correlation analysis (see Weenink, 2003; Fan & Konold, 2010; Vía et al., 2007) is one of the multivariate statistical analysis methods which measures the strength of the overall relationship between the linear structures (Canonical variables) of the dependent and the independent variables. It is a bivariate correlation between two Canonical variables, for example: a group of personal variables and a group of potential measures, group of price indices and a group of production indices, a group of psychological characteristics and a group of physiological characteristics, a group of academic achievement variables and a group of measures for business success (Rencher, 2002).

Canonical correlation analysis is a generalization of the concept of regression analysis, but rather than being a relationship between one variable Y and a group of variables $X_1, X_2, \ldots, X_q$, the Canonical correlation measures (with respect to) the relationship between a group of independent variables $X_1, X_2, \ldots, X_q$, and another group of dependent variables $Y_1, Y_2, \ldots, Y_p$ (Hair, 2009).

More statistically sound methods in the field are based on canonical correlation analysis and involve linear and nonlinear relationships between the groups of variables proposed by (Böckenholt and Böcknholt, 1990; Cook et al., 1996; Thorndike, 2000; Hardoon et al., 2004; Thompson, 2005).

Canonical correlation depends on finding a linear function (linear fitting) in the $X_1, X_2, \ldots, X_q$ (variable U) and a linear function (linear fitting) in the $Y_1, Y_2, \ldots, Y_p$ (variable V). The selected function which represents the correlation between the two largest (correlation between U and V is the greatest), in that case, there will be r correlation relations, which is equal to the smallest value between p and q (Gittins, 1985).

The aim of the canonical correlation analysis is to get a simple description of the structure of the relationship between subgroups of variables.

The paper is organized as follows. Canonical correlation analysis is described in section 2. Some important definitions are explained in Section 3. Test of Significance for Canonical Correlation is discussed in section 4. Statistical analysis which involves data collection and empirical results are presented in Section 5. Some concluding remarks are given in section 6.

2. Canonical Correlation Analysis

Suppose that there are two groups of variables $X' = [X_1, X_2,...,X_q]$, $Y' = [Y_1, Y_2,...,Y_P]$

Each of them has a variance matrix $\sum xx$, $\sum yy$ respectively, where $s = \min(p,q)$. The basic objective of the Canonical correlation is to find the canonical variables $U = \alpha'Y$ and $V = \beta'X$, where it should be a correlation between the U and V whichever is greater.

Suppose that the joint variance-covariance matrix of the vector

$$(Y_1, Y_2,...,Y_P, X_1, X_2,...,X_q) = (Y', X')$$

known as:

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} \tag{1}$$

Also, the joint variance-covariance matrix of the sample

$$S = \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix} \tag{2}$$

Therefore, the correlation between U and V are as follows: -

$$\rho_{UV} = \frac{\alpha' \Sigma_{yx} \beta}{\left[(\alpha' \Sigma_{yy} \alpha)(\beta' \Sigma_{xx} \beta) \right]^{1/2}} \tag{3}$$

because ρ_{UV} includes canonical variables U and V so-called canonical correlation.

$$R = [diag(S_{yy}, S_{xx})]^{-1/2} [diag(S_{yy}, S_{xx})]^{-1/2} \tag{4}$$

From the previous two equations the correlation matrix can be expressed as follows:

$$R = \begin{bmatrix} R_{yy} & R_{yx} \\ R_{xy} & R_{xx} \end{bmatrix} \tag{5}$$

and we can prove it by second method as follows: -

$$\left| S_{yx} S_{xx}^{-1} S_{xy} - \rho^2 S_{yy} \right| = 0 \tag{6}$$

$$\left| R_{yx} R_{xx}^{-1} R_{xy} - \rho^2 R_{yy} \right| = 0 \tag{7}$$

where (a_i, b_i) is the Eigen vector of matrix S, (c_i, d_i) is the Eigen vector of matrix R, and can be expressed in canonical variables as a vector of variables

$$U_i = a_i' y \tag{8}$$
$$V_i = b_i' x \tag{9}$$

Standard parameters can also calculated according to the following formula:

$$U_i = c_i' z_y \tag{10}$$

$$V_i = d_i' z_x \tag{11}$$

where they can express on (Z_x, Z_y) as standardized variables (Timm, 2002).

3. Some Important Definitions (Black *et al.*, 1998)

3.1 Canonical Function

Represents the relationship (correlation) between two structures (Canonical variables). Each Canonical function has two variables, one of the Canonical group of independent variables and the other to the group of dependent variables. The strength of this relationship is given by the canonical correlation.

3.2 Canonical Loadings

A measure of simple linear correlation between the independent variables and canonical variables. Interpretation of the canonical loadings are similar to the interpretation of factor loadings in factor analysis.

3.3 Canonical Cross-Loadings

Represents the correlation between the independent or dependent corresponding canonical variables, for example: the independent variables associated with the canonical dependent variables, the dependent variables associated with the canonical independent variables.

3.4 Canonical Variates

Represents the linear structure of the total weighted sum of two variables or more and can be defined as either independent or dependent variables.

3.5 Canonical Roots

Represent the square of canonical correlation which is used to estimate the amount of the variance between the weighted optimal canonical variables for independent variables and dependent and can be named Eigen values.

4. Test of Significance for Canonical Correlation (Härdle and Simar, 2007)

We can test the following alternative hypothesis

$$H_o : \rho_1 = \rho_2 = ... = \rho_m = 0$$
$$H_1 : \rho_1 \neq \rho_2 \neq .... \neq \rho_m \neq 0 \tag{12}$$

for canonical correlation coefficient and by using the most common measures used a Wilk's lambda described as follows:

$$\lambda = \prod_{i=1}^{s} (1 - \rho^2) \tag{13}$$

where ρ^2 represent the eigenvalues of the sample:

$$\left| S_{yx} S_{xx}^{-1} S_{xy} - \rho^2 S_{yy} \right| = 0 \tag{14}$$

which is a square canonical correlation of the sample.

Equation (13) proves that the two variables X, Y are uncorrelated linearity and mathematical application (Wilk's λ) will be described almost as a variable distribution following a Chi-square with degrees of freedom v = (p-k) (q-k), as well statistically significant application (Wilk's λ) needs to account for the following statistics:

$$\chi^2 = -\left[(n-1) - (p+q+1)/2 \right] \log \lambda \tag{15}$$

where:

  n: the number of cases.

  Log: the natural logarithm function.

  q: the number of variables in the first group.

  p: the number of variables in the second group.

5. Statistical Analysis

Statistical program SPSS was used to find a canonical correlation analysis through finding the correlation matrix between all independent variables and the correlation matrix between all dependent matrix; the correlation

matrix between independent and dependent variables in both groups; finding Wilks' Lambda test to see significance of the canonical correlation; finding a canonical correlation between the two groups; finding standard canonical coefficients in the first and second groups; creating factor loadings matrix in the first and second groups; finding cross loadings matrix in first and second groups; and finding the canonical scores of the first and second groups.

5.1 Data Collection

The data has been taken from Ibn Sina Hospital (surgery and fractures ward) for (80) patients with infection of the urinary tract were selected as a group of variables which could be influential on the disease. These variables were divided into two groups: the first group was a group of personal variables and the second group was a group of pathological variables as follows:

5.2 Results and Discussion

By Equation (5) will be displayed a correlation matrix for personal variables in addition to the correlation matrix of pathological variables, the joint correlation matrix between personal and pathological variables as shown in tables (1, 2, 3) respectively.

Table 1. Correlation Matrix between the Independent Variables

Var.	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1	-0.0832	-0.4823	-0.0185	-0.0885	-0.3640
X_2	-0.0832	1	0.0448	0.0618	-0.6364	0.0051
X_3	-0.4823	0.0448	1	-0.2834	-0.0348	0.4984
X_4	-0.0185	0.0618	-0.2834	1	-0.0745	-0.1673
X_5	-0.0885	-0.6364	-0.0348	-0.0745	1	0.0951
X_6	-0.3640	0.0051	0.4984	-0.1673	0.0951	1

Table 2. Correlation Matrix between the Dependent Variables

Var.	Y_1	Y_2	Y_3	Y_4
Y_1	1.0000	0.1701	-0.0771	0.1283
Y_2	0.1701	1.0000	0.2090	0.1541
Y_3	-0.0771	0.2090	1.0000	0.1014
Y_4	0.1283	0.1541	0.1014	1.0000

Table 3. Correlation Matrix between the Dependent and the Independent Variables

Var.	Y_1	Y_2	Y_3	Y_4
X_1	0.4121	0.1711	0.2281	0.1273
X_2	-0.1183	0.5063	0.1838	0.0064
X_3	-0.1820	-0.1325	-0.3085	-0.0391
X_4	0.1230	0.1691	0.0685	-0.1768
X_5	-0.0305	-0.3617	-0.1165	-0.0797
X_6	-0.1205	0.0490	-0.1743	-0.0623

Table 4. Wilks' Lambda Test

Wilk's Lambda	Chi-SQ.	DF	Sig.
0.415	64.571	24	0.000
0.683	28.033	15	0.021
0.893	8.280	8	0.407
0.961	2.893	3	0.408

Table 4 represents the first and second canonical correlations were significant but the rest of the canonical correlations were not significant, based on the test results (Wilk's lambda).

Table 5. Canonical Correlations

F_1	F_2	F_3	F_4
0.626	0.485	0.266	0.196

Table 5, which contains the first, second, third and fourth canonical correlation, the canonical correlations are strong between the first canonical variable which is extracted from the canonical correlation function. The rest of the canonical correlation was weak and it refers to the weakness of the relationship between the canonical variables extracted from the functions and canonical correlation.

Table 6. Standardized Canonical Coefficients for Group-1

	F_1	F_2	F_3	F_4
X_1	-0.460	0.762	-0.359	0.457
X_2	-0.796	-0.535	0.166	-0.222
X_3	0.260	-0.020	-0.710	0.536
X_4	-0.282	0.131	-0.881	-0.434
X_5	0.021	-0.101	-0.086	-0.621
X_6	-0.315	-0.114	-0.360	0.225

Table 7. Canonical Coefficients for Group-1

	F_1	F_2	F_3	F_4
X_1	-0.922	1.525	-0.719	0.916
X_2	-0.045	-0.030	0.009	-0.013
X_3	0.191	-0.015	-0.523	0.395
X_4	-0.658	0.307	-2.057	-1.013
X_5	0.050	-0.245	-0.210	-1.510
X_6	-0.351	-0.127	-0.402	0.251

Table 8. Standardized Canonical Coefficients for Group-2

	F_1	F_2	F_3	F_4
Y_1	-0.122	0.947	-0.372	0.088
Y_2	-0.877	-0.434	-0.240	0.292
Y_3	-0.320	0.407	0.633	-0.631
Y_4	0.173	0.082	0.653	0.762

Table 9. Canonical Coefficients for Group-2

	F_1	F_2	F_3	F_4
Y_1	-0.089	0.690	-0.271	0.064
Y_2	-0.563	-0.279	-0.154	0.188
Y_3	-0.255	0.324	0.504	-0.502
Y_4	0.044	0.021	0.167	0.195

Table 10. Canonical Loadings for Group-1

	F_1	F_2	F_3	F_4
X_1	-0.401	0.864	0.125	0.199
X_2	-0.778	-0.528	0.163	0.134
X_3	0.368	-0.502	-0.457	0.562
X_4	-0.345	0.117	-0.596	-0.599
X_5	0.550	0.153	-0.104	-0.486
X_6	0.027	-0.435	-0.444	0.338

Table 10 represents the factor loadings for the first group (personal variables). Through the first factor we show that the independent variables (X_2, X_5) have a simple linear correlation with the corresponding canonical independent variables, while variables appeared (X_1, X_2, X_3) with a linear relationship with the corresponding independent variables canonical in Group II. However the rest of the factors (third and fourth) cannot be relied upon to describe the data, because the canonical correlation coefficients were not significant, according to the test (Wilk's lambda).

Table 11. Cross Loadings for Group-1

	F_1	F_2	F_3	F_4
X_1	-0.251	0.419	0.033	0.039
X_2	-0.51	-0.256	0.043	0.026
X_3	0.230	-0.244	-0.121	0.110
X_4	-0.216	0.057	-0.158	-0.118
X_5	0.344	0.074	-0.028	-0.095
X_6	0.017	-0.211	-0.118	0.066

Table 11 represents cross loadings of the first group (personal variables). Through the first factor it can be seen that the independent variable (X_2) has a linear relationship with the canonical dependent variables. However, the second, third and the fourth factors could not be relied upon in the description of the data, because the canonical correlation coefficients were not significant, according to the test (Wilk's lambda).

Table 12. Canonical Loadings for Group-2

	F_1	F_2	F_3	F_4
Y_1	-0.224	0.852	-0.378	0.284
Y_2	-0.938	-0.175	-0.070	0.292
Y_3	-0.476	0.251	0.678	-0.500
Y_4	-0.010	0.177	0.632	0.754

Table 12 above represents canonical loadings for the second group (pathological variables) and through the first factor it is shown that the variable (Y_2) has a simple linear correlation with the corresponding canonical dependent variables, while the variable (Y_1) has a linear relationship with the corresponding canonical dependent variables in the second group. The rest of the factors (third and fourth) cannot be relied upon to describe the data, because the canonical correlation coefficients were not significant, according to the test (Wilk's lambda).

Table 13. Cross Factor Loadings for Group-2

	F_1	F_2	F_3	F_4
Y_1	-0.140	0.414	-0.100	0.056
Y_2	-0.587	-0.085	-0.019	0.057
Y_3	-0.298	0.122	0.180	-0.098
Y_4	-0.006	0.086	0.168	0.148

Table 13 represents cross factor loadings for the second group (pathological variables). Through the first factor it is shown that the dependent variable (Y_2) has a linear relationship with the canonical independent variables. The second cross factor loading does not appear to have any significant effect variable with any of the supported canonical changes. The third and fourth factor could not be relied upon in the description of the data because the canonical correlation coefficients were not significant, according to the Wilk's lambda test. The canonical values of the first and second groups are shown in Table 14 in Appendix A**.**

6. Conclusions

Canonical correlation analysis method (CCA) is very useful in interpretation of data by discovering the structures and similar relationships between two sets of multi-dimensional variables and categories of those variables are often used in medical data. The significant values (sig.) are selected when the value is less than or

equal to 0.05. Wilk's lambda test showed significant first and the second canonical correlation and the rest of the canonical correlations were not significant.

There is a strong relationship between the first group (personal variables) and the second group (pathological variables), because the correlation function has worked to maximize the correlation between two groups, and through factor loading matrix have been identified canonical variables that have a relationship with the original values. There are significant and non signifigant relationships in the results of factor loadings and cross factor loadings. Theses results are are very important to interpretation the correlation relationships between the dependant and the independent variable groups.

Reference

Black, W., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). Multivariate data analysis. *New Jersey*.

Böckenholt, U. and Böcknholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika, 55*(4), 633-639.

Cook, J. A., Razzano, L., & Cappelleri, J. C. (1996). Canonical correlation analysis of residential and vocational outcomes following psychiatric rehabilitation. *Evaluation and Program Planning, 19*(4), 351-363.

Fan, X., & Konold, T. R. (2010). Canonical correlation analysis. *The reviewer's guide to quantitative methods in the social sciences*, 29-40.

Gittins, R. (1985). *Canonical analysis: a review with applications in ecology*: Springer Berlin.

Hair, J. F. (2009). Multivariate data analysis.

Härdle, W., & Simar, L. (2007). *Applied multivariate statistical analysis* (Vol. 22007): Springer.

Hardoon, D. R., Szedmak, S., & Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural computation, 16*(12), 2639-2664.

Rencher, A. C. (2002). *Methods of multivariate analysis*. New York: J. Wiley.

Thompson, B. (2005). Canonical correlation analysis. *Encyclopedia of statistics in behavioral science*.

Thorndike, R. M. (2000). Canonical correlation analysis. *Applied multivariate statistics and mathematical modeling*, 237-263.

Timm, N. H. (2002). *Applied multivariate analysis*: Springer.

Vía, J., Santamaría, I., & Pérez, J. (2007). A learning algorithm for adaptive canonical correlation analysis of several data groups. *Neural Networks, 20*(1), 139-152.

Weenink, D. (2003). Canonical correlation analysis. Proceedings of the 2003 *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 81-99.

Appendix A

Table 14. Canonical scores for first and second Groups

	Canonical scores for Group-1				Canonical scores for Group-2			
	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
Obs1	-3.81	-1.46	0.41	2.09	-8.55	0.98	-0.32	-1
Obs2	-6.41	-3.07	1.16	2.31	-6.03	-0.53	-1.93	0.2
Obs3	-2.57	-0.94	0.43	1.89	-4.41	1.48	-2.39	0.68
Obs4	-2.16	-0.94	0.57	1.89	-4.98	1.48	-1.93	0.68
Obs5	-4.61	-3.7	-0.48	1.43	-5.69	1.87	-0.05	0.42
Obs6	-4.44	-3.93	1.6	1.88	-4.73	0.17	-0.21	1.74
Obs7	-3.69	-1.38	-0.18	3.15	-4.46	-0.06	-0.84	-0.37
Obs8	-4.22	-1.41	1.75	0.93	-4.78	1.6	-0.15	0.8
Obs9	-3.11	-1.38	0.96	3.15	-6.01	-0.06	-3.71	-0.37
Obs10	-4.22	-1.95	1.75	2.87	-4.78	-0.22	-0.15	-0.18
Obs11	-4.43	-1.38	1.19	3.15	-5.42	-0.06	-1.39	-0.37
Obs12	-3.52	-1.64	1.93	3.47	-5.89	0.44	0.73	-0.87
Obs13	-3.85	-2.14	2.61	4.57	-6.77	2.95	-2.74	-1.21
Obs14	-3.77	-1.07	2.05	1.1	-4.87	0.92	-0.03	-0.36
Obs15	-4.22	-1.59	1.75	3.04	-4.78	1.12	-0.15	1.38
Obs16	-5.55	-1.68	1.6	3	-6.69	0.78	-1.35	0.99

Obs17	-3.51	-1.12	2.37	3.27	-4.44	0.94	-0.3	0.8
Obs18	-4.23	-2.24	-0.5	2.72	-6.74	0.63	0.74	1.18
Obs19	-4	-0.86	-0.4	2.95	-4.4	0.43	-0.92	1.31
Obs20	-4.61	-2.2	-0.52	0.54	-3.3	0.61	-1.78	0.01
Obs21	-5.29	-3.37	1.92	2.16	-6.26	0.32	-1.62	1.55
Obs22	-3.5	-1.03	0.83	3.32	-6.21	1.27	-1.96	1.19
Obs23	-2.94	-0.94	0.9	1.89	-7.49	1.48	-2.63	0.68
Obs24	-2.65	-0.73	0.51	3.01	-4.68	0.94	-0.55	1.89
Obs25	-4.1	-3.19	-0.32	2.24	-3.89	0.99	-1.29	2.33
Obs26	-3.74	-1.21	2.22	3.96	-4.4	0.67	-0.36	0.87
Obs27	-5.97	-3.64	1.47	2.03	-6.12	-0.68	-1.81	0.38
Obs28	-3.95	-1.64	0.53	3.47	-6.12	0.44	-2.08	-0.87
Obs29	-3.27	-1.37	2.11	3.6	-5.18	1.44	-1.41	0.3
Obs30	-2.46	-0.81	0.5	1.96	-5.2	1.98	-0.15	1.26
Obs31	-5.97	-3.64	1.47	2.03	-6.12	-0.68	-1.81	0.38
Obs32	-3.32	-1.46	-0.22	2.09	-4.74	0.98	-2.26	-1
Obs33	-5.31	-1.94	1.16	3.32	-4.07	1.29	-0.8	0.49
Obs34	-3.77	-1.37	2.05	3.6	-4.87	1.44	-0.03	0.3
Obs35	-2.2	-1.11	0.79	2.26	-4.96	2.32	-1.94	0.56
Obs36	-4.44	-1.89	1.6	3.8	-4.73	0.94	-0.21	-1.37
Obs37	-5.04	-1.94	-0.06	3.32	-5.41	1.29	-2.73	0.49
Obs38	-3.19	-2.93	2.01	3.56	-6.44	1.54	1.16	1.9
Obs39	-6.19	-4.46	1.31	2.08	-6.08	-0.33	-1.87	0.07
Obs40	-3.27	-1.37	2.11	3.6	-5.18	1.44	-1.41	0.3
Obs41	-3.32	-1.12	2.35	3.27	-4.97	0.94	0.1	0.8
Obs42	-4.22	-1.68	1.75	4.18	-4.78	0.34	-0.15	-0.61
Obs43	-4.9	-3.06	-0.96	2.76	-6.6	0.98	0.56	0.86
Obs44	-2.65	-1.37	0.51	3.6	-4.68	1.44	-0.55	0.3
Obs45	-3.35	-1.37	2.19	3.6	-5.44	1.44	0.43	0.3
Obs46	-3.14	-3.37	0.04	2.16	-5.06	0.32	-0.34	1.55
Obs47	-4.44	-0.42	1.6	1.7	-4.73	1.98	-0.21	2.35
Obs48	-4.63	-1.94	-0.45	1.12	-4.73	1.1	-0.74	-0.87
Obs49	-2.66	-0.82	0.16	0.78	-6.79	0.41	-0.69	0.14
Obs50	-3.55	-1.99	-0.09	0.2	-4.49	-0.06	-0.8	0.32
Obs51	-3.81	-1.33	-0.41	0.97	-4.92	1.93	-0.53	1.19
Obs52	-4.36	-1.68	-0.6	0.8	-6.71	0.6	0.71	-0.37
Obs53	-4.98	-2.09	1.23	2.39	-4.62	2.93	-0.36	-2.37
Obs54	-5.28	-0.8	-0.97	1.67	-3.16	3.42	-1.96	1.47
Obs55	-5.29	-1.55	1.92	3.06	-6.26	1.29	-1.62	1.58
Obs56	-3.66	-1.12	1.84	1.08	-5.86	0.75	0.69	-0.56
Obs57	-5.12	-4.21	1.14	3.54	-4.59	3.01	-0.4	0.32
Obs58	-3.7	-1.16	0.85	1.06	-5.69	0.58	-2.35	-0.75
Obs59	-3.62	-1.33	-0.43	1.7	-5.44	1.99	-0.13	1.64
Obs60	-3.79	-1.94	-0.11	1.12	-3.96	1.1	-1.21	-0.87
Obs61	-5.31	-1.95	1.16	2.87	-4.07	-0.22	-0.8	-0.18
Obs62	-2.18	-0.86	0.52	0.76	-5.93	0.25	-1.25	-0.06
Obs63	-3.62	-0.81	-0.43	1.23	-5.44	1.92	-0.13	0.81
Obs64	-5.12	-3.09	1.14	0.54	-4.59	2.64	-0.4	2.03
Obs65	-3.54	-1.37	0.67	1.4	-6.69	1.25	-1.62	-1.06
Obs66	-3.27	-1.37	-0.05	1.4	-5.03	1.25	-0.38	-1.06
Obs67	-4.81	-3.12	-0.9	1.83	-6.62	-0.18	0.58	2.06
Obs68	-2.29	-1.12	0.73	1.08	-4.94	0.75	-1.96	-0.56
Obs69	-4.29	-1.62	0.02	2	-7.01	2.33	-1.49	0.95
Obs70	-4.71	-1.33	-1.02	1.42	-4.73	1.42	-0.78	-0.87
Obs71	-3.58	-0.61	2.04	2.63	-5.4	-0.07	0.37	1.81
Obs72	-3.16	-0.61	2.15	2.63	-6.16	-0.07	-0.69	1.81

Obs73	-4.28	-2	-0.12	1.94	-7.5	-1.38	-1.15	1.01
Obs74	-2.72	-0.87	0.18	2.5	-5.63	-1.07	0.12	0.64
Obs75	-3.54	-0.7	0.67	1.12	-6.69	-0.53	-1.62	0.51
Obs76	-2.14	-0.31	0.55	2.78	-5.94	-0.92	-1.24	0.45
Obs77	-2.19	-0.31	0.65	2.78	-5.45	-0.92	-1.59	0.45
Obs78	-4.23	0.13	-0.5	1.53	-6.74	0.63	0.74	1.49
Obs79	-3.87	-0.27	-0.38	2.8	-6.34	-0.75	0.49	0.65
Obs80	-3.08	0.09	2.34	2.97	-5.69	0.59	-1.01	2.2

Appendix B

X: Group of Personal Variables

X_1: Gender (0: Male, 1: Female)

X_2: Age

X_3: Cultural level (1: Not able to read or write, 2: Primary, 3: Secondary1, 4: Secendary2, 5: University)

X_4: Residence (1: City, 2: Rural)

X_5: Marital Status (1: Married, 2: Single)

X_6: Occupation (1: earner or a housewife, 2: Student, 3: Employee, 4: Retired)

Y: Group of Pathological Variables

Y_1: current disease (1: skeletal system, 2: muscular system, 3: nervous system, 4: digestive system, 5: reproductive system 6: lymphatic system.

Y_2: Chronic disease (1: None, 2: D.M, 3: H.T, 4: I.H.D, 5: 2 + 3, 6: 3 + 4.

Y_3: Day incidence (1: Unknown, 2: the second day, 3: the third day, 4: the fourth day, 5: fifth day, 6: The Sixth Day, 7: The Seventh Day.

Y_4: drugs taken (1: cannot take an antibiotic, 2: cephalosporines, 3: penicillins, 4: aminoglycosides, 5:antiprotozoal, 6: quinolones + cephalosporines + antiprotozoal, 7: antituberculosis, 8: 5 +2,   9: 3 + 2, 10: 5 +4 +3, 11: 5 +3, 12: 5 +4 +2, 13: 4 +3, 14: 5 +3 +2,15: 5 +4.

Copyrights