Medical Sciences
ISSN: 1308-7312 (NWSAMS)
ID: 2021.16.1.1B0105

Status : Research Article
Received: 02.11.2020
Accepted: 15.01.2021

Sadi Elasan
Van Yüzüncü Yıl University, sadielasan@yyu.edu.tr, Van-Turkey

DOI	http://dx.doi.org/10.12739/NWSA.2021.16.1.1B0105
ORCID ID	0000-0002-3149-6462
CORRESPONDING AUTHOR	Sadi Elasan

DETERMINATION OF SOME HEART DISEASE INDICATORS WITH MULTIPLE CORRESPONDENCE ANALYSIS

ABSTRACT

Various statistical techniques are used by researchers to help diagnose diseases. One of these, the detection of the presence of heart disease, it is important to develop rapid and effective techniques. Multiple Correspondence analysis can also be used to determine variables associated with some diseases. In this study, it is aimed to determine some variables that may cause heart diseases by multiple correspondence analysis. In this study, multiple correspondence analysis was applied to the data set of 303 patients presenting with heart disease. Multiple correspondence analysis is an analysis method that presents the relationships between categorical variables in two-dimensional space. The statistical study was conducted in June-September 2019 in Van. The application material for this study was obtained from the free access data site Kaggle.[1,2] This is a retrospective study. In this study; the relationship of the variables between the "presence of Heart Disease" and "some heart disease indicators" were investigated. According to "the transformed correlation coefficients for the presence of heart disease"; The variables associated with the presence of heart disease are "exercise-related angina, gender, heart rate, age, electrocardiography, systolic blood pressure, fasting blood sugar", respectively. In the study, some variables that may have an impact on heart diseases were determined by multiple correspondence analysis. It is hoped that the development of rapid and effective techniques for the detection of heart diseases will be important in terms of providing new perspectives to statistical decision-making processes.

Keywords: Multiple Correspondence Analysis, Heart Disease, Inertia, Variance, Dimension

1. INTRODUCTION

Multiple correspondence analysis is an analysis method that presents the relationships between categorical variables in two-dimensional space. Various statistical techniques are used by researchers to help diagnose diseases. It is important to develop rapid and effective techniques for the detection of heart diseases from these diseases [3, 4 and 5]. Multiple correspondence analysis can also be used to determine variables associated with some diseases. Multiple correspondence analysis methods; It allows the evaluation of categorical variables, explaining the relationships between row and column variables in cross tables and showing these relationships graphically in less dimensional space. Multiple correspondence analysis is the general case of simple correspondence analysis and examines the relationship between two or more variables by dividing the data matrix into row and column regions. The Chi-square test, which is used in many studies to determine the relationships between categorical variables, only determines whether there is a relationship between categorical variables and cannot determine the relationship

How to Cite:
Elasan, S., (2021). Determination of Some Heart Disease Indicators with Multiple Correspondence Analysis, Medical Sciences (NWSAMS), 16(1):36-40, DOI: 10.12739/NWSA.2021.16.1.1B0105.

between categories. However, multiple correspondence analysis, by dealing with more than two categorical variables together, can present the relationships between them graphically in two-dimensional space and provides easily interpretable results to researchers [6, 7, 8 and 9]. Correspondence analysis is examined in the two different headings as simple correspondence analysis and multiple correspondence analysis. In the simple correspondence analysis, only the relationship structure between two categorical variables is examined. In other words, simple correspondence analysis is used only in two-sided cross tables. In the multiple- correspondence analysis, the relationship structure between three or more categorical variables is examined. It is possible to consider simple correspondence analysis as a special case of multiple correspondence analysis for the case where the number of categorical variables is two [6 and 7]. Correspondence analysis is needed especially in researches in areas where categorical data are common such as medicine, health sciences, biometrics, economics, marketing and social sciences. In this study, an application of multiple correspondence analysis on heart disturbance data was performed. It can be said that there are many studies in the literature about multiple correspondence analysis. However, unlike previous studies, in this study; this study aimed to investigate the relationship between personal characteristics and heart disease in patients presenting with heart disease by multiple correspondence analysis.

2. RESEARCH SIGNIFICANCE

Various statistical techniques are used by researchers that can help diagnose diseases. It is important to develop fast and effective techniques for detecting the presence of heart diseases, which is one of them. Multiple Correspondence analysis can also be used to determine variables associated with some diseases. In this study, it is aimed to determine some variables that may cause heart diseases by multiple coherence analysis. This study; It is hoped that it will be important due to the development of fast and effective techniques in detecting the presence of heart diseases, providing applied information on multiple compliance analysis, providing a new perspective to the statistical decision-making processes and thus contributing to such studies in the future.

3. MATERIALS AND METHODS

In this study, the data set given descriptive statistics in Table 1 was used to make an application of multiple correspondence analysis. The application material for this study was obtained from the free access data site Kaggle [1 and 2]. Ethical approval for this study was obtained concerning the Ethical Principles for Medical Research Involving Human Subjects (the Helsinki Declaration) from the local ethics committee [1 and 2]. In this study, multiple correspondence analysis was performed on the data set of 303 patients who presented with heart disease. "The presence of heart disease" was used as an independent variable. SPSS (IBM SPSS for Windows, ver.23) statistical package program was used for data analysis. In this study, the relationship between "the presence of heart disease" and "gender, fasting blood glucose, electrocardiography desirable, exercise-induced angina, age, systolic blood pressure and heart rate" variables were examined by multiple correspondence analysis. The relationship between 8 categorical variables was presented graphically in two-dimensional space by using multiple correspondence analysis. Thus, it was tried to obtain easily interpretable results. The statistical study of this study was conducted in September 2019 in Van. Since the data used in this study is free-access data, the ethics committee report is not required again. This is a retrospective study.

Table 1. Descriptive statistics of patients presenting with heart disease

		N	%
Gender	Male	207	68.3%
	Female	96	31.7%
Fasting Blood Sugar (mg/dL)	FBS>120	45	14.9%
	FBS<120	258	85.1%
Resting ECG Results	ECG (+)	152	50.2%
	ECG (-)	151	49.8%
Exercise-Induced Angina	Angina(+)	99	32.7%
	Angina(-)	204	67.3%
Age (Years)	25-45 Years	64	21.1%
	46-65 Years	206	68.0%
	66-80 Years	33	10.9%
Systolic Blood Pressure (mmHg)	SBP 90-140	238	78.5%
	SBP 141-200	65	21.5%
Max Heart Rate	HR 70-120	37	12.2%
	HR 121-200	266	87.8%
Heart Disease	HD (+)	165	54.5%
	HD (-)	138	45.5%

4. RESULTS

4.1. Results of Multiple Adaptation Analysis in Determining Some Heart Disease Indicators

Table 1 shows the results of the initial matrix obtained based on dimensions in the multiple correspondence analysis applied to determine the relationship between "the presence of Heart Disease" and "Gender, Fasting Blood Sugar, Electrocardiography Desired, Exercise-Induced Angina, Age, Systolic Blood Pressure and Heart Rate" variables.

Table 2. Model Summary

	Variance Accounted For		
Dimension	Total (Eigenvalue)	Inertia	% of Variance
1	1.946	0.243	24.325
2	1.309	0.164	16.368

Table 2 shows the variance amounts of each dimension in the total change (inertia). While Size-1 explains 24.3% of the total change, Size-2 explains 16.4% of the total change. Accordingly, considering the percentages of variance explanation of dimensions; in patients presenting with heart disorders; the multiple correspondence analysis diagrams for the relationship between "the presence of Heart Disease" and "Gender, Fasting Blood Glucose, Electrocardiography Desired, Exercise-Induced Angina, Age, Systolic Blood Pressure and Heart Rate" variables were given in Figure 1.

4.2. Multiple Correspondence Analysis Diagram

In this Study; "the relationship between the presence of heart disease" and "gender, fasting blood glucose, electrocardiography, exercise-induced angina, age, systolic blood pressure and heart rate" variables were investigated. According to the obtained multiple correspondence analysis diagrams (Figure 1); "Systolic blood pressure between 90-140, Fasting blood sugar <120(mg/dL), Heart rate from 121-200, 46-65 years old, Requesting ECG, Exercise-related Angina and Male" variables were associated with the "presence of Heart Disease (+)". However; In this study, there was angina in exercise, heart rate between 70-120 and no ECG demand were associated with "no heart disease (-)". Also; It was not found to be in a strong relationship "having an AKŞ>120, an SBP of 141-200, and an age range of 66-80 and 25-45" variables with the "presence or absence of Heart Disorder".

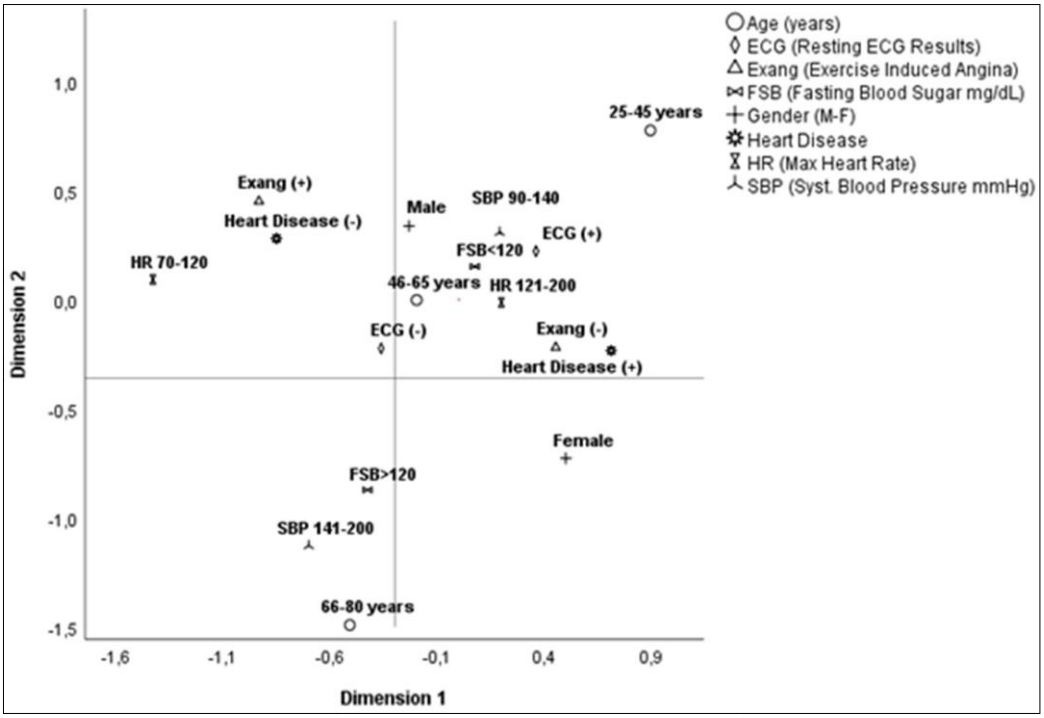

Figure 1. Diagram of multiple correspondence analysis between the presence of heart disease and other independent variables

Table 3. Correlation coefficients according to Heart Disease Presence

	Heart Disease
Exercise-Induced Angina	0.437
Gender	0.281
Max Heart Rate	0.266
Age (years)	0.205
Resting ECG Results	0.175
Systolic Blood Pressure	0.136
Fasting Blood Sugar	0.028

Table 3 shows the correlation coefficients of "the presence of heart disease" and other variables. According to this; The variables associated with "the presence of Heart Disease" are "Exercise-related Angina, Gender, Heart Rate, Age, Electrocardiography, Systolic Blood Pressure, Fasting Blood Sugar".

5. DISCUSSION

Various statistical techniques are used by researchers to help diagnose diseases. One of these, the detection of the presence of heart disease, it is important to develop rapid and effective techniques. Multiple correspondence analysis can also be used to determine variables associated with some diseases. In this study, it is aimed to determine some variables that may cause heart diseases by multiple correspondence analysis. According to the transformed correlation coefficients for" the presence of heart disease"; The variables associated with "the presence of heart disease" are "exercise-related angina, gender, heart rate, age, electrocardiography, systolic blood pressure, fasting blood sugar". In this study, an application of multiple correspondence analysis on heart disturbance data was performed. Some variables that may effect heart diseases were determined by this analysis method. It can be said that multiple correspondence analysis can be used to examine the relationships between variables in the data set of patients presenting with heart disease in detail and to present them visually in an easily interpretable manner.

In this study; It is hoped that the development of rapid and effective techniques for the detection of heart disease, providing practical information on multiple compliance analysis, providing a new perspective to statistical decision-making processes and thus contributing to such studies in the future.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Aha, D. and Kibler, D., (1988). Instance-based Prediction of Heart-disease Presence with the Cleveland Database. The University of California, 3(1):3-2.
[2] Kaggle Datasets, (2019). (internet) (Access:10.07.2019). https://www.kaggle.com/ronitf/heart-disease-uci.
[3] Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S., and Froelicher, V., (1989). Int. Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease. Am. J. Cardiology, 64:304-310.
[4] Gennari, J.H., Langley, P., and Fisher, D., (1989). Models of Incremental Concept Formation. Artificial Intelligence, 40: 11-61.
[5] Patel, J., Tejalpadhyay, D., and Patel, S., (2015). Heart Disease Prediction Using Machine Learning and Data Mining Technique. Heart Disease, 7(1):129-137.
[6] Greenacre, M.J., (1981). Practical Correspondence Analysis. Looking at Multivariate Data, 81-107.
[7] Clausen, S.E., (1988). Applied Correspondence Analysis, An Introduction. California, 1. Ed, Sage Publications.
[8] Suner, A. ve Çelikoğlu, C.C., (2008). Uygunluk Analizinin Benzer Çok Değişkenli Analiz Yöntemleri Ile Karşılaştırılması. İstatistikçiler Dergisi: İstatistik ve Aktüerya, 1(1):9-15.
[9] Onat, A., Şurdumavci, G., Şenocak, M., Örnek, E., Gözükara, Y., Karaaslan, Y. ve Özcan, R., (1991). Türkiye'de Erişkinlerde Kalp Hastalığı Ve Risk Faktörleri Sıklığı Taraması: 3. Kalp Hastalıkları Prevalansı. Türk Kardiyoloji Derneği Arşivi, 19(1):26-33.