Indian Journal of Science and Technology, *Vol 10(20), DOI: 10.17485/ijst/2017/v10i20/91294, May 2017*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Application of Principal Component Analysis (PCA) to Medical Data

Naeem Ahmed Qureshi[1], Velo Suthar[1], Habibullah Magsi[2], Muhammad Javed Sheikh[3], Mubeena Pathan[4] and Barkatullah Qureshi[4]

[1]Department of Statistics, Sindh Agriculture University, Tandojam, Pakistan; qureshistat@gmail.com, vsutahar@yahoo.co.uk
[2]Department of Agricultural Economics, Sindh Agriculture University, Tandojam, Pakistan; hmagsi@sau.edu.pk
[3]Department of Rural Sociology, Sindh Agriculture University, Tandojam, Pakistan; ruralsociologyst@gmail.com
[4]Information Technology Centre, Sindh Agriculture University, Tandojam, Pakistan; mubeena2009@gmail.com, qureshi5939@gmail.com

Abstract

Objectives: To apply Principal Component Analysis to medical data to explore the factors thought to be very important in increasing the risk of Ischemic Heart Diseases. **Methods/Statistical Analysis**: PCA was performed in R-mode using correlation and covariance for medical data. Variables pertaining to chemical tests of blood namely cholesterol, high density lipoprotein, triglyceride, Apo protein A-1, Apo protein B, low density lipoprotein, phospholipids, total lipid, glucose and uric acid, are undertaken to know the relationship between them and membership of group variable. **Findings**: The results indicated that among these factors cholesterol, triglyceride, Apo protein B, low density lipoprotein, phospholipids, total lipid, and uric acid were recorded to be higher in IHD group compared to those of control group. High density lipoprotein and Apo protein A-1 were recorded to be lower in IHD group, whereas found higher in control group. Cholesterol was highly correlated with low density lipoprotein and moderately correlated with total lipid. Cholesterol, Apo protein B and low density lipoprotein belonged to component 1, Apo protein A-1, phospholipids and uric acid belonged to component 2, triglycerides and total lipid belonged to component 3 and high density lipoprotein and glucose belonged to component 4. The first four components have explained 60.67 percent component variability. **Improvements/Applications:** The end results showed that the average cholesterol level, which is considered as the main risk factor of Ischemic Heart Disease, was found higher even in control group.

Keywords: Application, Ischemic Heart Diseases, PCA

1. Introduction

Principal component analysis (PCA) is statistical technique is applicable in the situation when the researcher is dealing with single set of variables and wants to discover that what are the main variable(s) which are important in the formation of a coherent factors in such a way that there exists no correlation between these newly formed factors. Since PCA works in a highly correlated environment so it chooses a set of variables that are highly dependent with each other but, at the same time, they are quite uncorrelated with other subset of variables which are combined together to form a factor. The basic idea is that these newly formed factors drive the underlying process due to which the variables in the data set are supposed to correlate with each other. The specific goals of the component analysis are to summarize patterns of correlations among observed variables, to reduce a large number of observed variables to a smaller number of factors, and to provide an operational definition (a regression equation) for an underlying process by using observed variables. Since the number of factors are usually far fewer than the number

*Author for correspondence