

Abstract No: PO-02

Heart disease prediction using multivariate data

I. M. L. N. Jayaweera^{1*}

¹Department of Mathematics and Statistics, Texas Tech University, USA
jayaweera.imiya-mohottige@ttu.edu*

This study addresses the problem of determining the important attributes of deciding heart disease. Heart disease can be considered one of the most common diseases in the world. Nowadays, diagnosing a disease in an early stage is a more crucial task. If it goes to later stages, it is hard to save human life and is a major problem for health care workers. In order to identify the heart condition as early as possible, it is essential to understand the main factors that may affect one's heart condition. With this motivation, we apply statistical methods to determine the significant attributes that cause heart disease. This study presents the application of various statistical methods to find the significant factors in heart disease for decision support during the diagnostic process. Heart Disease data from UCI Machine Learning Repository is considered for the analysis. The data set is divided into two parts, exploratory data and confirmatory data. We conducted a descriptive analysis for the classification of the heart disease dataset based on association, dimension reduction, and a confirmatory analysis based on principal components and hypothesis testing to discover the most important information in the heart disease dataset. In the explanatory analysis, we got a brief idea about the factors which need to be looked at when diagnosing heart disease and their relationship based on different classifications. Principal Component Analysis (PCA) and hypothesis testing were used in the confirmatory data analysis. As a result of a hypothesis test, it can be concluded that there is no significant difference between the mean of the variables on exploratory data and confirmatory data. This implies that we can use any part of the data for our analysis. According to the principal component analysis, the first four principal components explain 90.4% total variability of the data. PCA further reveals that the cholesterol level, maximum heart rate, and resting blood pressure have a major impact on heart disease.

Keywords: Dimension reduction, Heart disease, Multivariate, Principal component