Application of Principal Component Analysis in Annual Health Survey Data-Factors Affecting Infant Mortality

Feba Anna Varghese
Department of Biostatistics, St. Thomas College, Pala

There has been an increase in the number of composite indicators developed by various national and international agencies. Unfortunately, individual indicators are sometimes selected in an arbitrary manner with little attention paid to the interrelationships between them. The underlying nature of the data needs to be carefully analysed before the construction of a composite indicator. The goal of principal components analysis (PCA) is to reveal how different variables change in relation to each other and how they are associated. This is achieved by transforming correlated variables into a new set of uncorrelated variables using a covariance matrix or its standardised form. Cluster analysis is also used in the development of composite indicators to group information on countries based on their similarity on different individual indicators (OECD (Organisation for Economic Co-operation and Development) 2008). The aim of this study is to use Principal Component Analysis to identify factors associated with infant and under five mortality in the AHS Data (Annual Health Survey Data 2012-2013). In the study we have explored the use of two dimension reduction methods for determining the factors affecting the Infant mortality rate and under five mortality rate. The problem of multicollinearity between the independent variables was explored. The approaches of principal component analysis, factor analysis were examined to address this issue and compared against the stepwise linear regression with original variables. A substantial benefit in terms of model fit could not be observed with the dimension reduction technique. And some common factors were identified. All the immunization and breast feeding exclusively for six months can decreases the mortality rate and under five mortality rate. The diet factors like feeding vegetable and semisolid food during six months can increase the infant mortality rate. The childhood diseases like diarrhoea and acute respiratory disease can increase infant mortality and under five mortality rate. Two
clusters of Indian districts were identified with the variables that were considered for the regression of IMR and UMR.