

A two stage procedure for identification of molecular markers and stepwise testing for the partial contributions.

H. Sridhara¹ and Raghavendra S. G²

¹ Elizabeth & Emmanuel Kaye Biostatistics and Bioinformatics Unit, Thrombosis Research Institute, Narayana Hrudayalaya Campus, Hosur Road, Bangalore, INDIA..

² Department of Agricultural statistics, University of Agricultural Sciences, Bangalore, INDIA

Analysis of gene expression profiles involves a holistic approach, that must be suited for the typical data situation with small number of genotypes n (= number of observations) and a large number of molecular markers p (= variables), the so called “small n and large p ” paradigm. A two stage procedure is proposed for reduction in dimensionality and identifying contributing markers associate with several phenotypic characters. The first stage analysis involves a novel procedure of Quantitative Row-Column analysis technique for identifying the markers contributing to genetic variability. The second stage analysis involves maximizing the correlations between linear functions of phenotypic characters and the selected markers. Further under the assumption of random effect multivariate regression model a stepwise test for partial contribution of each genetic marker for the phenotypic expression is proposed. The methods are applied to a data set with 135 Rice lines of Double Haploid (DH) population involving 235 molecular markers spanning over all chromosomes and 11 phenotypic characters.