USE OF MULTIVARIATE TECHNIQUES FOR CLASSIFICATION OF RICE (*Oryza sativa* L.) GERMPLASM

A PROJECT REPORT PRESENTED BY

S.C.J. DISSANAYAKE

to the Board of Study in Statistics & Computer Science of the

POSTGRADUATE INSTITUTE OF SCIENCE

in partial fulfillment of the requirement
for the award of the degree of

MASTER OF SCIENCE IN APPLIED STATISTICS

of the

UNIVERSITY OF PERADENIYA
SRI LANKA
2009

USE OF MULTIVARIATE TECHNIQUES FOR CLASSIFICATION OF RICE (*Oryza sativa* L.) GERMPLASM

S.C.J. Dissanayake
Plant Genetic Resources Centre
Gannoruwa
Peradeniya
Sri Lanka

A total of 4153 rice (*Oryza sativa* L.) accessions have been conserved at the Plant Genetic Resources Centre in Sri Lanka. Some of the accessions have been characterized and morpho-agronomic traits were recorded. In the present work, the data on 13 quantitative traits of 347 rice accessions were analyzed to compare the effectiveness of different multivariate statistical techniques for classification of rice germplasm. It was observed that the sample variance could be summarized by seven principal components accounting for about 85.5 % of the total variance. The results of Principal Component Analysis (PCA) indicated that the first principal component (PC) explains 30 % of the total variance of rice accessions. The highest loading (41.7 %) on the first PC was given by number of days to heading. The other variables having positive loadings on the first PC were leaf blade length (39.3 %), culm length (37.9%), panicle length (34.5 %), days to maturity (31.3%), ligule length (29.8%), culm diameter (28.5%), leaf blade width (22.5%), and culm number (17.8%). Therefore the first principal component can be explained as the component for plant stature and age class. The second PC explains 19.0% of the total variance of rice accessions. The higher positive loadings on the second PC were given by grain weight (46.4 %), seedling height (39.6 %) and grain length (35.6%) and grain width (33.6%). The second principal component represents the grain characteristics of rice. The third principal component cannot be readily explained. The fourth PC explains slender rice grains. The 5^{th} , 6^{th} and 7^{th} PCs explain broader leaves, profuse tillering and slender culm.

Score plots created by plotting the scores for PC1 and PC2 showed several clusters of accessions. The accessions which were clustered on score plots were examined and it could be justified that the principal component analysis is a useful trchnique to visualize and to group rice germplasm.

Different factor solutions were applied to the data set and resulting outputs were used to explain major factors account for variance in the data. Based on these explanations and the general knowledge on the morphology of different rice varietal groups, it could be concluded that the varimax-rotated 7-factor solution of principal component factor analysis method is the most clearly explainable solution to interpret the data. Accordingly, the loadings in the first factor reflected the importance of plant stature in expressing the diversity of accessions and the second factor suggests the importance of age class. The grain characteristics were expressed by third and fourth factors. Broader leaves, profuse tillering and slender stem were related with 5^{th}, 6^{th} and 7^{th} factors respectively. This information will be useful to enhance the characterization procedure of rice germplasm and, thus to improve classification process of rice accessions. Identification of factors may be useful to understand the relationships among variables in order to optimize the selection criteria in rice varietal improvement programmes.

Cluster analyses were performed with four clustering methods: single linkage, average linkage, complete linkage and Ward's minimum variance method. The single, average and complete linkage clustering methods produced clusters having single accession or a few numbers of accessions during the formation of first few clusters. These clustering techniques are important in identifying accessions, which are different from the rest of the accessions. The Ward's minimum variance method produced distinct groups or clusters having more accessions. In addition, considering the grouping of accessions having similar accession names within a cluster, it can be predicted that the Ward's minimum variance clustering method is a useful technique in classification of rice germplasm collections.