AGRIALLIS
Science for Agriculture and Allied Sector: A Monthly Magazine
Volume 6, Issue 3
www.agriallis.com

Article Id
AL04317

MULTIVARIATE TECHNIQUES FOR RESEARCH ANALYSIS IN PLANT BREEDING

Email
dimplesharma17071992@gmail.com

[1]Sarita Devi*, [2]Shikha Bhagta and [3]Neena Kumari

[1]Statistics Department, Dr YS Parmar College of Horticulture & Forestry, Thunag, H.P, India

[2]Tree Improvement and Genetic Resources Department, Dr YS Parmar College of Horticulture & Forestry, Thunag, H.P, India

[3]Forest Product Department, Dr YS Parmar College of Horticulture & Forestry, Thunag, H.P, India

Plant breeding is a continuously evolving field. From the discovery of Mendel's laws of heredity to the development of gene-based markers, advances on analytical tools and emerging use of big data from trials, plant breeders have constantly utilized scientific breakthroughs to increase the rate of genetic gain and optimize breeding processes. Thus, multivariate analysis helps plant breeders in finding the pattern between variables, analysing the effects that different factors have on each other and the relationships between them. Multivariate data arise when a researcher takes measurements for two or more variables. Multivariate analysis offers a more complete examination of data by looking at all possible independent variables and their relationships with each other. In design and analysis, this technique is used to perform trade studies across multiple dimensions as multivariate techniques consider the effects of all variables on the responses of interest. The development of multivariate methods emerged to analyse large databases and increasingly complex datasets. Multivariate methods are designed to simultaneously analyse data sets. Always keeping in mind that all variables must be treated accurately reflect the reality of the problem addressed. There are different types of multivariate analysis and each one should be employed according to the type of variables to analyse: dependent, interdependence and structural methods.

Plant breeding is the purposeful manipulation of qualities in plants to create new varieties with a new set of desired characteristics. Multivariate techniques are the most common methods used for data analysis in plant breeding. Multivariate analysis is preferred over univariate analysis in plant breeding research studies because it can exploit correlated traits and environments. Genetic study based on the multivariate analysis is a powerful tool for determining the degree of divergence between populations, the relative contribution of different components to the total divergence and the nature of forces operating at different levels (Sanwal *et al.*, 2015). In this article three main multivariate techniques used in plant breeding i.e. Principal Component Analysis (PCA), Factor Analysis (FA) and Discriminant Analysis will be discussed.

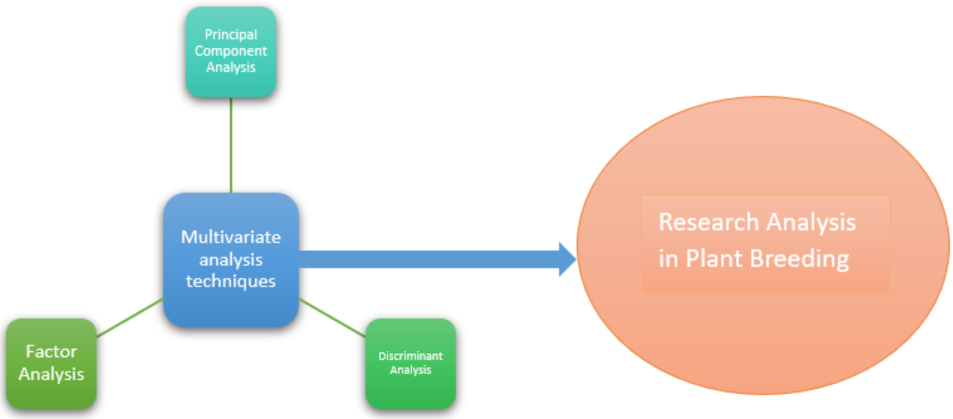

Figure 1

Discussion

I. Principal Component Analysis

A principal component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. Its general objectives are data reduction and interpretation. Usually, p components are required to reproduce the total system variability, often much of this variability can be accounted for by a small number k of the principal components. If this is the case, k components give as much information as there is in the original p variables. The k principal components can then replace the initial p variables, and the original data set, consisting of n measurements on p variables, is finally reduced to a data set consisting of n measurements on k principal components. There is always a question of how many components to retain. There is no definite answer to this question. Things to consider while retaining principal components include the amount of total sample variance explained., the relative sizes of the eigen values and the subject-matter interpretations of the components. A useful aid to determine an appropriate number of principal components is a scree plot. With the eigen values ordered from largest to smallest, a scree plot is a plot of λ_i versus i- the magnitude of an eigen value versus its number. The number of principal components to be retained depends on the elbow (bend) in the scree plot.

In plant breeding, our objective is to construct uncorrelated linear combinations of the measured characteristics that account for much of the variation in the sample. The uncorrelated combinations with the largest variances will be called the sample principal components. In plant breeding, PCA is used to estimate the contribution of each trait for the total observed variations in the genotypes, identify the major traits accounting for the greater share in observed variations, focus on specific traits of interest for crop improvement, predict the breeding value of hybrids from the genetic divergence of their parents. Magudeeswari *et al.* (2019) conducted a study to evaluate the plant nutrient traits in 12 baby corn genotypes by using Principal Component Analysis and revealed that the first three principal components together accounted for 87.49 % of variability. Anandhinatchiar *et al.* (2023) studied the genetic diversity and genetic relationship among seed traits in ricebean and using PCA revealed that the traits viz., length breadth ratio, bulk density, hundred seed weight, seed volume, seed length and seed thickness contributed to the maximum genetic variability.

II. Factor Analysis

The purpose of factor analysis is to describe, if possible, the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called factors. Suppose all variables within a particular group are highly correlated among themselves, but have relatively small correlations with variables in a different group. Then it is conceivable that each group of variables represents a single underlying construct, or factor, that is responsible for the observed correlations. Factor analysis can be considered as an extension of principal component analysis. Both can be viewed as attempts to approximate the covariance matrix Σ.

In plant breeding, factor analysis helps to reduce large number of variables into fewer number of factors and is a way to find hidden patterns and show what characteristics are seen in multiple patterns. It provides a way of explaining the observed variability in behaviour in terms of these traits. Filipovic *et al.* observed interrelationships of yield and yield components of 15 commercial maize hybrids using factor analysis and pointed out significant effect of two factors on grain yield.

III. Discriminant Analysis

Discriminate analysis is a multivariate technique concerned with separating distinct sets of objects and with allocating new objects to previously defined groups. Discriminant analysis is exploratory in nature, it is employed on a one-time basis to investigate observed differences when causal relationships are not well understood. To describe, either graphically or algebraically, the differential features of objects from several known collections, find "discriminants" whose numerical values are such that the collections are separated as much as possible.

In plant breeding, discriminant analysis measures the efficiency of various traits combinations in selection and provides information on yield components and thus aids indirect selection for genetic improvement of yield. Discriminant Analysis provides information on weight coefficient, general selection index and restricted selection index. Kanbar *et al* (2010) conducted a study to determine the effectiveness of discriminant analysis in recognizing deep rooted types of rice based on a few plant measurements.

Conclusion

Multivariate methods are ideal for the analysis of large data sets and to find the cause-and-effect relationships between variables. Multivariate statistical technique is a form of statistics encompassing the simultaneous observations and analysis of more than one statistical variable at a time. In this article we tried to clarify how multivariate statistical methods such as principal component analysis (PCA), factor analysis (FA), and discriminant can be used as methods to explain relationships among different variables and making decisions for future works with examples relating to the agriculture and plant science.

References

Anandhinatchiar S, Jayamani P, Kumaresan D, Bhuvaneswari K and Sudha M. (2023) Principal Component Analysis and Genetic Association of Seed Related Traits in an Underutilized Pulse Crop, Ricebean (*Vigna umbellata*). *Agricultural Science Digest* 10: 1-8.

Filipovic M, Babic M, Delic N, Bekavac G and Babic V. (2014) Determination relevant breeding criteria by the path and factor analysis in maize. *Genetika* 46(1): 49-58

Kanbar A, Motohashi T, Kondo K and Shashidhar G. (2010) Evaluation of discriminant analysis in identification of deep and shallow rooted plants in early segregating generation of rice (*Oryza sativa* L.) using single tiller approach. *Australian Journal of Basic and Applied Sciences* 4(8): 3909-3916.

Magudeeswari P, Sastry Divakara EV and Devi Renuka Th. (2019) Principal Component (PCA) and cluster analyses for plant nutrient traits in baby corn (*Zea mays* L.). *Indian Journal of Agricultural Research* 53(3): 353-357.

Sanwal S, Singh B, Singh V and Mann A. (2015) Multivariate analysis and its implication in breeding of desired plant type in garden pea (*Pisum sativum*). *The Indian Journal of Agricultural Sciences* 85(10): 1298-1302.