Differentiating two popular cultivars of large cardamom (*Amomum subulatum* Roxb.) using a statistical method

PRIYA P. MENON[1]*, V.V. RADHAKRISHNAN[2]

[1]*Department of Statistics, Maharaja's College, Ernakulam, India
[2]Department of Botany, University of Calicut, Kerala, 673635, India

Research Article

ABSTRACT

The variability in large cardamom (*Amomum subulatum*) germplasm was studied by the application of multivariate methods such as Hotelling's T^2 statistic and discriminant function analysis. Data on variables such as girth of the clump, number of bearing tillers, number of spikes per clump, number of capsules per spike, capsule length, capsule diameter and capsule weight were recorded in 40 accessions belonging to two cultivars of large cardamom namely *Ramsey* and *Sawney.* Hotelling's T^2 test revealed that there is significant difference between the cultivars with respect to the characters studied. Discriminant function analysis brought about a discriminant function to distinguish between the two cultivars of large cardamom.

INTRODUCTION

Large cardamom *Amomum Subulatum* Roxb. is one of the main cash crops cultivated in the sub-Himalayan States of Sikkim and Darjeeling District of West Bengal. The dried fruit is used as a spice in several Ayurvedic preparations. Its cultivation is mainly confined to an elevation ranging from 900 to 2000 m above mean sea level at a temperature varying between 10° C and 30° C and rain fall of 2000 to 3000 mm per annum. In India, the crop is cultivated in an area of 44701.ha with an annual production of 8803 MT [1]. Apart from domestic consumption, a sizable quantity of the produce is exported to various countries. Though a large number of cultivars in large cardamom have been identified, very limited studies have been carried out on documentation of capsule characters in the crop [2]. Hence, the present study was undertaken to discriminate between the two popular cultivars of large cardamom namely *Ramsay* and *Sawney* using certain morphometric characters. The variables that contribute more towards discrimination were also identified.

Material and Methods

The study was undertaken in a well-managed plantation at Gangtok in Sikkim (5000',27^0.00'N and 88^0. 00'E). Forty accessions of yielding large cardamom (4 years old) belonging to cultivars *Ramsey* and *Sawney* were used for the study. Plant growth characters such as girth of the clump(X_1), number of the bearing tillers(X_2), number of spikes per clump(X_3), number of capsules per spike(X_4), capsule length(X_5), capsule diameter(X_6) and capsule weight (X_7) were recorded. The data were subjected to analysis such as Hotelling's T^2 and linear discriminant function analysis [3]. Correlation analysis of the discriminant function with the variables, identified the traits which contributed to the cultivar differentiation. (Absolute value of correlation coefficients > 0.5) as suggested by Jayaraman [4]. Moreover, similar work for differentiating Vanilla *planifolia* from Vanilla *tahitensis* using discriminant function analysis has also been done [5].

Result and discussion

The genuineness of the data was tested by using Q-Q plot and the data were found to be in order. Hotelling's T^2 statistic [6] was employed to check the significant difference between the germplasm at cultivar levels. Hotelling's T^2 can be used to test equality of means of two multivariate normal populations with the same but unknown dispersion matrix.

The hypothesis to be tested is $H_o : \mu^{(1)} = \mu^{(2)}$

Consider the problem of testing $H_0 : \mu^{(1)} = \mu^{(2)}$ where $\mu^{(1)}$ and $\mu^{(2)}$ are the means of two multivariate normal populations $N_p(\mu^{(1)}, \Sigma)$ and $N_p(\mu^{(2)}, \Sigma)$. Here the multivariate normal populations are the two cultivars *Ramsay* and *Sawney*.

The test statistic is given by

$$F = \left(\frac{T^2}{N1+N2-2}\right)\left(\frac{N1+N2-P-1}{P}\right) \sim F_{(p,N1+N2-p-1)}$$

Where,

$$T^2 = \left(\sqrt{\frac{N1N2}{N1+N2}}(\bar{X}^{(1)} - \bar{X}^{(2)})\right)' \left[\frac{A}{N1+N2-2}\right]^{-1} \left(\sqrt{\frac{N1N2}{N1+N2}}(\bar{X}^{(1)} - \bar{X}^{(2)})\right)$$

p = number of variables

N_1 = Number of observations in the first population.

N_2 = Number of observations in the second population.

We reject the null hypothesis if $F > F_\alpha$, where F_α is the tabled value at 5% level of significance with (p, N_1+N_2-p-1) degrees of freedom

T^2 = 44.585, F = 5.3636 (Calculated value)

The table value, F_α following the F distribution with (7, 32) degree of freedom and significance level, α = 0.05 is given by F_α = 2.3127 (Table value)

As F > F_α, the hypothesis is rejected, that is the populations differ significantly with respect to the characters studied.

R.A Fisher constructed a linear discriminant function Z= X'l such that it affords maximum discrimination between two multivariate populations with mean vectors $\mu^{(1)}$ and $\mu^{(2)}$ and common dispersion matrix $\sum$.

$$l = \sum{}^{-1}(\mu^{(1)} - \mu^{(2)})$$

Thus, the discriminant function is given by,

$$Z = X'\sum{}^{-1}(\mu^{(1)} - \mu^{(2)})$$

The classification rule is

Assign X to first population if $\quad X'l \geq \dfrac{(\mu(1) + \mu(2))'}{2} l$

Assign X to second population if $\quad X'l < \dfrac{(\mu(1) + \mu(2))'}{2} l$

If the population parameters $\mu^{(1)}$, $\mu^{(2)}$ and $\sum$ are not known, they are replaced by their sample counter parts. i.e $\overline{X}^{(1)}$, $\overline{X}^{(2)}$ and

$$\frac{(N_1 - 1) S_1 + (N_2 - 1) S_2}{N_1 + N_2 - 2}$$ respectively.

If $X'l \geq \frac{1}{2} (\overline{X1} + \overline{X2})'l$, then allocate X to first population and

If $X'l < \frac{1}{2} (\overline{X1} + \overline{X2})'l$, then allocate X to second population

Fisher's Discriminant function analysis brought about a discriminant function to distinguish between the two cultivars.

$$Z = X'l = -0.039 X_1 - 0.015 X_2 + 0.020 X_3 - 0.058 X_4 + 24.449 X_5 - 8.378 X_6 + 10.108 X_7$$

The classification rule is that if the discriminant score of a genotype X'l is **above 86.21**

Then it can be can be grouped under cultivar *Sawney* and that with a score **below 86.21** can be placed under *Ramsay*.

Correlation analysis of the discriminant function with the variables (Table-1) revealed that among the variables, capsule length and capsule weight contributed maximum discrimination between the cultivars *Ramsay and Sawney*.

Conclusion

Application of Hotelling's T^2 statistic for highlighting the significant difference between accessions belonging to two large cardamom cultivars based on growth characters revealed that they differ significantly. Discriminant function analysis could be employed for distinguishing between the two cultivars and also for identifying the variables that contribute towards discrimination.

References

1. Anonymous (2020). Spices Statistics, Spices Board, Cochin, Kerala, India, 281s
2. Gupta, U., Documentation of spike and capsule characters in large cardamom. Journal of Hill Res. 2000;13:122-124
3. Fisher, R.A., The use of multiple measurements in taxonomic problems. Ann. Eugen. 1938;7:240-250.
4. Jayaraman, K., A Hand Book on Statistical Analysis in Forestry Research. Kerala Forest Research Institute, Peechi, Thrissur, Kerala. 2001.
5. Radhakrishnan, V.V., Differentiating Vanilla planifolia from Vanilla tahitensis using a statistical method. Journal of plantation crops. 2005;33(1):75-76
6. Anderson, T. W., An introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc New York. 1984.