Evaluation of Statistical Classificatory Techniques in Chickpea (*Cicer arietinum* L.) Genotypes

Tanveer Ahmed Khan[1], A. R. S. Bhat[2], Tanveer Ahmed*[3], Sudheesh Kulkarni[4] and Viresh Hiremath[5]

ABSTRACT: *Wealth variability is the prime requirement for improvement of any crop. In breeding programme diversity of parents is always emphasized, and classification helps in identifying the groups (genotype), which on crossing would produce greater variability in the progenies, which is needed by the breeders to operate their selection efficiently (Seber, 1984). Such classification could be done by using the techniques like, a) Scaling technique and b) Linear Discriminant Analysis. In scaling technique were employed to assess the performance of 175 chickpea genotypes, results revealed that 67 Chickpea genotypes were classified under best performing. In medium and poor performing genotypes 51 and 57 have identified respectively, in contrast 74.63% of these cases are true into the group of high yielding, 78.43 % of correct classification was achieved for Medium performing genotypes, and true classification of Poor performing genotypes estimated is about 85.96% from Linear Discriminant analysis. An overall 79.4 % of original grouped genotypes correct classification was best achieved using discriminant analysis compared with Scaling technique classification.*

Key word: *Linear Discriminant Analysis (LDA), Scaling technique, Chickpea, genotypes*

INTRODUCTION

Scaling technique is very much vogue in behavioral sciences like psychology and management (Bhat *et. al.,*). In discriminant analysis, a linear combination of the independent variables is formed and serves as the basis of assigning cases to groups. Discriminant analysis, also known as 'statistical pattern recognition', is a statistical technique (Seber, 1984) used popularly in the classification of biological materials. Application of linear discriminant analysis could, therefore, provide new insights regarding the evaluation of Chickpea genotypes. The objective of this Study is to classify the genotypes into different homogeneous groups by using scaling technique and discriminant analysis. Frank *et al.,* (1965) discuss in detail the upward biases that can occur in classification using LDA. Morrison. (1969) Discriminant analysis was used in order to classify the cocoon. Bhat *et al.,* (1991) obtained three groups

of agriculture services cooperative societies as poor , satisfactory and good using scaling technique. Petrovic *et al.,* (1998) used the Discriminant Analysis in Identification of Low- and High-Water Use Kentucky Bluegrass Cultivars (KBG) . Pedro *et al.,* (2001) studied the metal content of 46 tea samples, including green, black, and instant teas, was analyzed. Varadachari. and Mukherjee. (2004) studied discriminant analysis of Clay Mineral Compositions Compositional data for 464 clay minerals (2:1 type) were analyzed by statistical techniques.

MATERILS AND METHODS

For the purpose classification of 175 chick pea genotypes under study observations from Eight mean vectors have recorded viz., Plant height , Number of branches per plant , Days to 50% flowering , Number of pods per plant , Wilt index at 1st ,2nd and 3rd at the stage of plant growth and Total yield of crop . The

[1]   Research Associate, NIVEDI, Hebbal, Bangalore,

[2]   Professor Department of Agricultural Statistics, UAS, Dharwad,

[3]   Assistance Professor, Department of Agricultural Economics, UHS, Bagalkot, Karnataka,

[4]   Assistant Professor, Dept of Spices and Plantation Crops, UHS, Bagalkot.

[5]   Assistant Professor, Department of Post Harvest Technology, UHS, Bagalkot.

Correspondence Author: tanveerecon@gmail.com

Classification of chick pea genotypes for efficient yielders can be archived using i) Scaling technique and ii) Linear Discriminant Analysis, discussed below

Scaling technique

Let n be the total number of genotypes and k variables which indicates the performance of genotype. Let X_{ij} denote the observed value of ith genotype at j^{th} plot. The procedure of Scaling technique is as follows:

(i) Take the mean of all genotype for all k variables.

(ii) Transform the Xij into Standard normal variate Z_{ij} by the relation

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j}$$

Where;

X_{ij} – observed value of ith genotype of jth plot for k^{th} variable

μ_j – Mean of jth plot for k^{th} variable

σ_j – Standard deviation for j^{th} plot of kth variable

(iii) Read the area below the Zij by referring to the table of area under normal curve using *Biometrica tables for Statistics (1966)*

(iv) Take the sum of all the variables for each genotype.

(v) The genotypes can be classified into number of categories ranging from Best performing genotype to Poor yielding using fractiles of normal distribution.

For instance if it is intended to categorize the genotype into three groups then the cutoff point is , the Genotypes above the total area of X + 0.425σ were classified as Best performing , Genotypes below the total area of X - 0.425σ were classified as Poor performing and the Genotypes between the total area of X - 0.425σ to X + 0.425σ were classified as Medium performing .

Linear Discriminant Analysis (LDA)

Discriminant analysis is a Multivariate technique concerned with classifying set of objects (or set of observations) and with allocating new objects or observations to the previously defined groups. It involves deriving variates, which are combination of two or more independent variables that will discriminate best between a priori defined groups.If the population covariance matrices are equal then linear discriminate function for classification is used,

otherwise quadratic discriminant function is used for this purpose. The maximum number of discriminant functions that can be computed is equal to minimum of G-1 and p, where G is the number of groups and p is the number of variables. Suppose the first discriminant function is

$$Z_1 = W_{11}X_1 + W_{12}X_2 + \ldots, + W_{1p}X_p$$

Where the W_{ij} is the weight of jth variable for the 1st discriminant function. The weights of the discriminant function are such that the ratio

$$\lambda_1 = \frac{Between\ group\ SS\ of\ Z_1}{Within\ group\ SS\ of\ Z_1}\ is\ maximized.$$

Suppose the second discriminant function is given by,

$$Z_2 = W_{21}X_1 + W_{22}X_2 + \ldots, + W_{2p}X_p$$

The weights of above discriminant function are estimated such that the ratio

$$\lambda_2 = \frac{Between\ group\ SS\ of\ Z_2}{Within\ group\ SS\ of\ Z_2}$$

is maximized subject to the constraint that the discriminant scores Z_1 and Z_2 are uncorrelated. The procedure is repeated until all possible discriminant functions identified. Once the discriminant functions are identified, the next step is to determine a rule for classifying the future observations. Classification procedure involves the division of discriminant space in g mutually exclusive and collectively exhaustive regions. {Chatfield and Collins (1990), Jhonson and Wichern (1996), Sharma (1996)}.

RESULT AND DISCUSSION

Case wise group membership of Chick pea genotype is computed and the actual group membership from scaling technique is compared to its membership as predicted by the discriminant analysis. Estimates of the classification function coefficient are used to compute a score for each group and then classify into a particular group.

The table 1. shows the three groups of genotypes Best, Medium, and Poor performing of which about 67 genotypes were classified under High yielder. In medium and Low yielder genotypes 51 and 57 have identified respectively .The table:2. shows the actual and misclassification of genotypes. It is observed that Fifty genotypes from original group of rank 1 is correctly classified estimated to 74.63% of these cases are true into the group of high performing, sixteen genotypes have predicted to be miss classified into

rank1 from rank 2 and one genotype predicted to be misclassified under rank 1 from rank 3, which accounts to 23.88% and 1.49% of true classification of Medium and Low performing genotypes have been misclassified into High performing group. Forty Medium performing genotypes having rank 2 is correctly classified estimated to 78.43% of these cases true into Medium performing genotypes , and six genotypes and is predicted under rank 1 and rank 3 from actual group of rank 2 respectively, which accounts to 11.76% and 9.80% misclassification in Medium performing genotypes. Similarly, about Forty nine genotypes of Low performing genotypes having rank 3 are correctly classified , which estimates to 85.96% of true classification under rank 3, and 8 genotypes of low performing would have been classified under medium performing genotypes,

which accounts to 14.04% of misclassification in Low performing genotypes.

An overall 79.4 % of original grouped genotypes correct classification was best achieved using discriminant analysis into High performing, Medium performing and Low performing groups that obtained with all 8 original variables in the analysis simultaneously.

CONCLUSION

The classification results from Linear Discriminant analysis and Scaling technique for different Category of Chickpea Genotypes are found similar to each other without many variations, as shown in Fig. 1, thus researchers interpretations were true by using either of the techniques to ascertain the performance of Chickpea Genotypes.

Table 1
Classification of genotypes using Scaling technique

Classified Group	Genotypes of Chick pea (Genotype Number)											
Best	4	5	7	9	12	14	17	19	21	23	30	36
Performing	37	40	43	46	47	48	50	51	52	53	54	55
Genotypes	57	64	67	69	71	77	79	81	90	94	100	102
(Rank 1)	103	114	120	122	125	128	129	131	132	140	142	143
	148	149	150	151	152	153	154	155	157	159	161	162
	164	165	172	174	175	176	179					
Medium	2	6	10	15	18	20	22	25	35	38	39	42
Performing	45	58	60	62	63	66	70	72	73	75	78	82
Genotypes	85	86	87	89	91	92	93	95	97	98	99	104
(Rank 2)	107	108	110	117	119	127	130	134	138	144	156	158
	166	169	178									
Poor	1	3	11	13	16	24	26	27	28	29	31	32
Performing	33	34	41	49	56	59	61	65	68	74	76	80
Genotypes	83	84	88	96	105	106	109	111	112	113	115	116
(Rank 3)	118	121	123	124	126	133	135	136	137	139	145	146
	147	160	163	167	168	170	171	173	177			

Table 2
Discriminant analysis results for classification Chick pea genotypes

		Rank	Predicted Group Membership			Total
			1	2	3	
Original	Count	1	50	16	1	67
		2	6	40	5	51
		3	0	8	49	57
	%	1	74.63	23.88	1.49	100
		2	11.76	78.43	9.80	100
		3	0.00	14.04	85.96	100

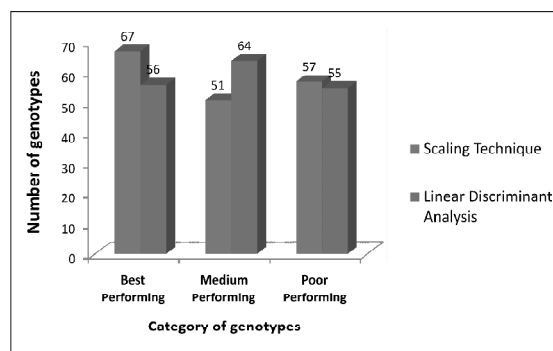79.4% of original grouped cases correctly classified

Figure 1: The classification results from Linear Discriminant analysis and Scaling technique for different Category of Chickpea Genotypes.

REFERENCES

BHATT, A.R.S., ARUNA RAO, K., and RAMACHANDRA BHATTA., (1991), A Methodology to Measure the Efficiency of Business Organizations ,*Indian Cooperative Review, Jr.*, PP. 237-244.

CHANDRIKA VARADACHARI and GARGI MUKHERJEE (2004), Discriminant Analysis of Clay Mineral Composition. *Clays and Clay Minerals*, 2004; **52**(3), 311–320.

CHATFIELD, C. and COLLINS, A.J, (1990), Introduction to Multivariate Analysis. Chapman and Hall publications.

FRANK. R.E, MASSY F.W and MORRISON G.D, (1965), Bias in Multiple Discriminant Analysis, *Journal of Marketing Research,* II: 250-258.

JHONSON, R.A. and WICHERN, W. (1996), Applied Multivariate Statistical Analysis. Prentice-Hall of India Private Limited.

MORRISON G. DONALD, (1969), On Interpretation of Discriminant Analysis, *Journal of Marketing Research,* VI: 156-163.

PEDRO L. FERNÁNDEZ-CÁCERES, MARÍA J., and MARTÍN., (2001), Differentiation of Tea (Camellia sinensis) Varieties and Their Geographical Origin According to their Metal Content J. Agric. Food Chem., **49** (10): 4775–4779.

PETROVIC A. M., EBDON J. S., and SCHWAGER S. J., (1998), Evaluation of Discriminant Analysis in Identification of Low- and High-Water Use Kentucky Bluegrass Cultivars ,Crop Sci 38: 152-157.

SHARMA, S . (1996), Applied Multivariate Techniques , Jhon Wiely & Sons, New York.

SEBER, G.A.F., (1984), Multivariate Observations. John Wiley, New York,: pp 279–46.