Arunachalam, V. and Bandyopadhyay, A.
A Method to Make Decisions Jointly on a Number of Dependent Characters.
Indian Journal of Genetics and Plant Breeding, 44 (1984): 419-424.

Indian J. Genet., **44(3)** : 419 - 424 (1984)

A METHOD TO MAKE DECISIONS JOINTLY ON A NUMBER OF DEPENDENT CHARACTERS

V. ARUNACHALAM AND A. BANDYOPADHYAY

National Research Centre, Indian Agricultural Research Institute, Regional Station, Hyderabad-500030

(Accepted : December 5, 1983)

ABSTRACT

A method is proposed to make decisions jointly on a number of dependent character variables. A score was allotted to each entry for each character. The scores were added across characters to provide a final score for each entry. Based on the final scores, the entries were ranked on their performance over a set of characters.

Experiments in diverse research disciplines, in general and agriculture, in particular deal inevitably with dependent characters and warrant, quite often, decisions taken jointly on them. While methods of analysis would be available to arrive at conclusions on each of the characters individually, it remains a problem to collate them over the characters and arrive at a final decision. When made, such decisions become subjective and would reflect the variation in the decision making processes adopted by individual investigators. The need is increasingly felt in this context to provide a method adequate to draw practical decisions. This paper is an attempt in this direction.

MATERIALS AND METHODS

A number of F_2 plants from selected single crosses was advanced to F_3 on a plant-to-progeny row basis in groundnut. Selection of upto 4 grades was made in each F_3 family based on the improvement of pod number over a national check. The unselected pods were pooled to form a bulk in each F_3 family. Based on kernel weight, kernel size and kernel colour, selections from different F_3 families were pooled to form populations. Each population thus consisted of selections from various F_3 families derived from one or more single crosses. The term 'population' is used here to define such a pool of F_3 selections. The populations were advanced to F_4 in plots of 50 m^2 containing 1500 plants and their kernel yield measured.

The ANOVA of F_5 populations was based on a completely randomised design. For illustration, the four direct yield components, pod weight (PW), kernel weight (KW), 100-kernel weight (TW) all measured in g/plant and shelling per cent (SP) were only considered.

The differences in the mean values of populations were tested by t-test. For example, the difference in the means of populations i and j was tested by

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{e\,[(1/n_i)+(1/n_j)]}}$$

where $\bar{x}_i$ was the mean of the population i,

$\bar{x}_j$ that of j,
n_i was the number of components (selections) making up population i,
n_j that in j,
e was the error m.s. in the ANOVA

The 't' would follow a t-distribution with error d.f.

The population means were arranged in groups based on t-test (and l.s.d.). The topmost group containing population with the highest mean was given a score 1, the next best a score 2 and so on. If 'k' was the number of groups for a particular character, the populations in group 1 were given a score=1/k, those in group 2, a score=2/k and so on to obtain standardised scores across the characters later on. When overlapping of groups occur, it is possible that a population was found in group 1 and also in group 2. The score for that population was taken to be the average which would thus be equal to $(1 + 2)/2k = 3/2k$. Populations occurring in more than 2 groups would be treated in a like manner for allotment of scores. The above points would become clear by a study of Tables 2 and 3.

The individual scores for each character were added up to provide a total score for each population. The populations were then ranked in descending order of the numerical values of total scores.

RESULTS

The differences among the 8 F_3 populations were highly significant (Table 1) indicating it would be useful to assess their relative order of importance. While only 2 groups, which were also non-overlapping, were obtained for PW and TW, there were 3 and 5 groups for KW and SP, some of which were overlapping (Table 2). The individual scores obtained by each population for

TABLE 1

ANOVA for 4 yield components in 8 F_3 populations of groundnut

Source	d.f.	Mean squares			
		PW	KW	TW	SP
Between pops	7	7983.98*	4103.18*	464.74*	731.66*
Within pops	147	428.13	165.91	69.94	46.95

PW=Pod weight; KW=Kernel weight; TW=100-Kernel weight; SP=Shelling per cent.
*Significant at 5% level.

TABLE 2

t-test of significance between means of populations for 4 yield components

Pop	n	PW	r	Pop	KW	r	Pop	TW	r	Pop	SP	r
III	9	82.87		I	59.36	1	I	53.00		I	73.03	1
I	5	81.10		III	54.72		III	52.35	1	III	67.26	2
V	9	78.41	1	V	50.69		VI	45.59		V	64.38	
IV	7	77.92		IV	49.72	2	X	40.09		IV	64.16	3
VI	26	74.17		VI	45.83		XII	38.84		VI	62.09	
IX	7	40.57		IX	23.91		IV	38.13	2	IX	60.06	4
X	36	37.62	2	X	21.79	3	IX	37.72		X	58.37	5
XII	15	32.60		XII	17.56		V	36.89		XII	55.02	

n = no. of component lines in the population; r = rank each component character and their total scores across the characters are presented in Table 3. It was found that populations I and III were at the top while X and XII were at the bottom. Further the populations IV and V received identical scores for individual characters and hence their total scores were also equal, implying they were of equal merit (See also 'Discussion').

The relative order of importance of F_3 populations was compared with the ranking of these populations based on their kernel yield in F_4 (Table 4). Populations I and VI had identical ranks in F_3 and F_4. Populations III, IV, IX and XII had adjacent ranks in F_3 and F_4. Populations V and X had quite different ranks in F_3 and F_4.

TABLE 3

Scores allotted to F_3 populations for 4 yield components

Pop	PW	KW	TW	SP	Total score
I	1/2	1/3	1/2	1/5	1.533
III	1/2	3/6	1/2	3/10	1.800
VI	1/2	2/3	1/2	5/10	2.167
IV	1/2	3/6	2/2	5/10	2.500
V	1/2	3/6	2/2	5/10	2.500
IX	2/2	3/3	2/2	7/10	3.700
X	2/2	3/3	2/2	9/10	3.900
XII	2/2	3/3	2/2	5/5	4.000

TABLE 4

Comparison of the ranked performance of F_3 populations with their performance in F_4 ranked on kernel yield

Pop	Rank in	
	F_3	F_4
I	1	1
III	2	4
VI	3	3
IV	4	6
V	4	8
IX	6	5
X	7	2
XII	8	7

DISCUSSION

Various approaches are adopted to tackle the complex problem considered in this paper. For example, breeders assign a floor value to each character and allot a population a score +1 or —1 when it exceeds or falls short of the floor value. The aggregate scores across the characters are used for ranking the populations. Alternatively, an arbitrary weight is associated with each character and a discriminant function is set up. Using the values of discriminant scores, the populations are arranged in their order of merit.

The method suggested in this paper has its own merits. For example, the differences are tested statistically for its significance. The number of groups that is obtained by t-test is taken into account in assigning the score for each population and each character. If there are no significant differences among populations for a character (which, incidentally, will also be reflected in the F-test in the ANOVA), all the populations will occur in one group, consequently each getting a score 1. It can be easily seen that such a character does not add to the differentiation between populations (as uniform addition of 1 to the score of every population will not alter the order of merit). But the scoring process takes care of the varying potential of characters in differentiation. For example, if 10 groups are obtained for a character X, the population in group 1 gets a score 1/10; if only 3 groups are obtained for a character y, the population in group 1 gets a score 1/3. Since lower total score will mean higher rank (as the topmost group is allotted score 1), this implies that character X is weighted more compared to Y for population 1 and similarly for others. Further, the t-test of significance being based on the error m.s. of ANOVA, if the character Y has a high error variance associated with it, its differentiating potential is proportionately lower compared to another character X with low error variance. For, the critical difference in the former would be high resulting in a comparatively fewer significant differences for Y compared to X. Following earlier arguments this would imply that X is automatically weighted higher than Y.

When a number of characters are considered, there could be a sequential relationship between them, in general. For example, poor germination affects seedling vigour that affects in turn initial leaf area and hence photosynthesis. Hence decisions based on several important characters spanning the entire growth phase will be fair and precise due to an automatic weighting in the expression of the various characters measured sequentially over the growth period.

Thus the three mechanisms—scoring process, t-test of differences and decision based on a large number of characters—ensure the superiority of the proposed method over the traditional ones. In the latter, only the numerical superiority over the floor value of a character is considered. Such superiority may not be upheld by statistical test leading to erroneous decisions. Though apparently no weight is associated with each character, it is evident the method proposed here takes into adequate account the relative importance of the characters.

On the other hand, discriminant function technique suffers essentially from arbitrary assignment of weights to characters. The problem of character weighting has been dealt with in great detail by Sneath and Sokal (1973) who defended equal weighting on several independent grounds. Moreover, exact rules for estimating weights cannot be formulated. 'When many characters are employed, the statistical analysis of similarity is only slightly affected by weighting (unless this weighting is extreme).'

The efficiency of the method is also borne out by the comparative performance of populations in F_4 (Table 4). Though there is no *a priori* reason to expect identical ranking of F_3 populations in F_4, it is also difficult to conceive of top-or bottom-ranking F_3 populations to perform otherwise in F_4. The results have upheld this point in bringing out the superior performance of populations I. III and VI and the inferiority of XII. The performance of F_3 population X suggests that the constitutent lines of the populations in F_3 must have been affected to a great extent by environment, diseases and pests while those of V could have had a higher contribution by environment. In addition, genetic differences in the constituent lines should also have been pronounced in those two populations to explain their differential performance in F_3 and F_4. But the results are, by and large, consistent in F_3 and F_4.

Sometimes two or more populations may receive the same total score as has happened for populations IV and V in F_3. The scores may be equal for every component character as in these populations in which case it can be concluded that they were of equal potential. Alternatively, there may be internal compensation in characters leading to such a result. In that case, those 'otherwise equal' populations can be differentiated on the scores computed on a subset of characters important for the objectives of the study.

Acknowledgement

The authors are grateful to Dr. S. Ramanujam for critical comments that helped to improve an earlier version.

Reference

Sneath, P.H.A. and R.R. Sokal. (1973). *Numerical Taxonomy.* W.H. Freeman and Co. San Francisco., pp. 96-113