STUDIES ON THE DISCRIMINANT FUNCTION FOR CHARACTER GROUPS OF SOYBEAN VARIETIES[1]

HONG-PANG WU[2]

(Received December 3, 1965)

Soybean varieties may be roughly classified as early type or late type, according as whether or not they come to ripen early under a particular latitude or under given climatic conditions. The varieties grown in the United States have been divided into nine maturity groups (Morse and Cartter, 1949). Lu (1961) classified about 150 soybean varieties into eight ecotypes based on the days from sowing to flowering (flowering time) and the period from flowering to maturity. These two methods of classification are rather artificial and are of little help in the quantitative evaluation of the likeness of the various varieties.

The characters in which a plant breeder is interested are the metric characters. However, these characters can hardly be evaluated genetically, for their heritable variations are usually masked by large non-heritable or environmental variations. In the early days, breeders exercised selections on the basis of indirected characters of which, they believe, the favorable response may be extended to the characters of interest, but the actual effect attributable to each character is usually unknown. Especially, the metric character could be a super-character (compound character), which may not be measurable by itself but may be measured by several sub-characters. To approach this problem, R. A. Fisher offered the method of discriminant function in 1936. It is but a linear function of the sub-characters, by which the super-character can be adequately measured and discriminated.

In many cases varieties are so similar that they have no recognizable morphological differences; they can be distinguished only by their physiological or ecological characters. A number of statistical methods have been offered for overcoming these difficulties, but in this paper we are concerned

(1) Paper No. 47 of the Scientific Journal Series, Institute of Botany, Academia Sinica. This work was supported by the National Council on Science Development, Republic of China.

(2) Assistant Research Fellow of the Institute of Botany, Academia Sinica. The author wishes to thank Dr. H. W. Li, the Director of the Institute, for his encouragements and stimulations during this work. Thank are also due to Dr. W. Y. Tan and Mr. T. S. Weng for their valuable suggestion and critical reading of the manuscript.

only with the discriminant function method of Fisher (1936), i. e., constructing a linear discriminant function for metric characters of soybean varieties, and then discriminating those varieties as well as their F_2 hybrid progenies into several groups by means of their discriminant values, and finally, selecting the desirable cross-combinations and F_2 individual plants.

Materials

Nine soybean varieties and their F_2 progeny were used in this study, namely; Taichung Green (A, 臺中青皮豆), Tsao-ta-tou (B, 操大豆), Jackson (C), Shih-shih (D, 十石), Japanese Green (E, 日本青皮豆), Sheng-liang (F, 生娘), Wakashima (G, 和歌島), Palmetto (H, 百美豆), and Mi-yueh (I, 米岳). The nine cross-combinations were A×F, A×G, B×D, B×G, C×D, C×G, C×H, E×H and E×I etc.

Eleven metric characters were measured, i. e., A′: flowering time, B′: period from flowering to maturity, C′: maturity date, D′: plant height, E′: no. of branches, F′: no. of nodes on main stem, G′: no. of pods, H′: no. of seeds, I′: no. of seeds per pod, J′: wt. of 100 seeds and K′: seed yield. The data were recorded on an individual plant basis, forty plants were selected in a randomized manner from each variety, or from each cross-combination.

Statistical Methods

The statistical methods of computation used in this study were as follows:

A. The selection of single character for the construction of discriminant functions: The coefficient of discrimination (k) for single character between two varieties was estimated by using Lubischew's method (1962):

$$k=\frac{(\bar{X}_1-\bar{X}_2)^2}{2s_0^2}=\frac{(\bar{X}_1-\bar{X}_2)^2}{s_1^2+s_2^2}=\frac{F}{n}, \qquad (1)$$

if $n_1=n_2=n$

where $\bar{X}$ and s^2 are respectively the mean value and the mean square of a single character, while n is the number of individual plants of each variety. The estimated k can be used to estimate the probability of misclassification (p) when the character X is used alone. Nine best traits (i. e., those giving the largest value of k) were selected for the construction of discriminant functions by using Kendall's concordance coefficient (1955).

B. Construction of the discriminant function for metric characters of soybean varieties: The variances between and within varieties for the nine most promising characters were calculated and were used to construct two discriminant functions for the soybean varieties using the approach suggested in Mather (1949) and Rao (1952).

Suppose now that a linear discriminant function is express by

$$X=k_1C_1+k_2C_2+k_3C_3,$$

where k=coefficient of equation, C=mean value of sub-character.

Let,

$$S_v^2=k_1^2S_1^2+k_2^2S_2^2+k_3^2S_3^2+2k_1k_2S_{12}+2k_2k_3S_{23}+2k_3k_1S_{31}$$

=sum of squares of between varieties,

$$S_e^2=k_1^2E_1^2+k_2^2E_2^2+k_3^2E_3^2+2k_1k_2E_{12}+2k_2k_3E_{23}+2k_3k_1E_{31}$$

=sum of squares of within varieties,

where S_i^2 and E_i^2 are the sum of squares of between and within varieties, S_{ij} and E_{ij} are the sum of product of between and within varieties, respectively. To maximize the difference between varieties and minimize the difference within varieties, we set $\frac{\partial}{\partial k_i}\cdot\frac{S_v^2}{S_e^2}=0$. It follows that $\sum_i\sum_j k_i(S_{ij}-\delta E_{ij})=0$, where $\delta=S_v^2/S_e^2$.

We may solve $|S_{ij}-\delta E_{ij}|=0$ to obtain the three roote of δ, and then we may obtain the two best linear discriminant functions, if we used the two more efficient roots of δ (Bartlett, 1947) to construct these functions.

C. Discrimination of soybean varieties based on their discriminant values: A diagram of ellipses with equal probability deviation was drawn to substitute the region of acceptance of two-dimensional representation on the basis of the discriminant values of a variety.

Suppose that the critical region of the two-dimensional discriminant values is defined by

$$F_s=\frac{n(n-2)}{2(n+1)}\Big[\Sigma\Sigma\frac{x'_a x'_b}{\Delta}(X'_a-\bar{X}'_a)(X'_b-\bar{X}'_b)\Big]>F_\alpha\binom{n}{n-2}.$$

If $F_s=F_\alpha$, we may say that $(1-\alpha)\times100\%$ of the discriminant values are contained in this critical region. We may transform this equation into a standard equation of ellipses, thus

$$a^2X_a'^2+b^2X_b'^2=1, \tag{2}$$

where, if $\Sigma x'_a x'_b>0$, $\theta<0$, or $\Sigma x'_a x'_b<0$, $\theta>0$ (θ= rotation angle of axes), then,

$$a^2=\frac{\lambda}{2\Delta}\Big\{(\Sigma x_a'^2+\Sigma x_b'^2)+\Big[(\Sigma x_a'^2-\Sigma x_b'^2)^2+4(\Sigma x'_a x'_b)^2\Big]^{1/2}\Big\},$$
$$b^2=\frac{\lambda}{2\Delta}\Big\{(\Sigma x_a'^2+\Sigma x_b'^2)-\Big[(\Sigma x_a'^2-\Sigma x_b'^2)^2+4(\Sigma x'_a x'_b)^2\Big]^{1/2}\Big\}, \tag{3}$$

and, if $\Sigma x'_a x'_b>0$, $\theta>0$, or $\Sigma x'_a x'_b<0$, $\theta<0$, then,

$$a^2=\frac{\lambda}{2\Delta}\left\{(\Sigma x_a'^2+\Sigma x_b'^2)-\left[(\Sigma x_a'^2-\Sigma x_b'^2)^2+4(\Sigma x_a'x_b')^2\right]^{1/2}\right\},$$
$$b^2=\frac{\lambda}{2\Delta}\left\{(\Sigma x_a'^2+\Sigma x_b'^2)+\left[(\Sigma x_a'^2-\Sigma x_b'^2)^2+4(\Sigma x_a'x_b')^2\right]^{1/2}\right\}, \quad (4)$$

with $\Delta=(\Sigma x_a'^2)(\Sigma x_b'^2)-(\Sigma x_a'x_b')^2$

$$\lambda=\frac{n(n-2)}{2(n+1)F_\alpha}$$

$F_\alpha=F(^{n}_{n-2})$ ($\alpha=$ probability of risk$=0.05$ or 0.01)

$n=$number of individual plants of each variety.

$$\theta=\frac{1}{2}\tan^{-1}\frac{2\Sigma x_a'x_b'}{\Sigma x_a'^2-\Sigma x_b'^2}.$$

The areas of the ellipses may be rough estimates, which would contain 95% (if $\alpha=0.05$, or 99% if $\alpha=0.01$) of a large number of two-dimensional discriminant values.

D. Classification of the F_2 progeny on each cross-combination: Based on the two discriminant functions, the discriminant values of the F_2 individual plants of each cross-combination can be estimated, and the graphical method can be used to delineate the group of the F_2 progeny.

Results acd Discussion

A. The selection of single characters for discriminant function

Usually, the characters used in the construction of a linear discriminant function are selected unsystematically, and the function often gives very small or no improvement in discrimination over that given by any single character. To promote the efficiency of this method, we must select for the discriminant function those characters that are appropriate and discard those characters that are inappropriate, otherwise the calculations will be too cumbersome. This can be achieved by using the coefficient k.

The coefficient of discrimination (k) and the probability of misclassification (p) for each of the eleven single characters between two varieties were estimated, using equation (1), and the results are given in Table 1.

In Table I, we have for each character a total of $C_2^9=36$ combinations of varieties, taken two at a time. If we count the number of combinations under certain probability range of misclassification (p) for each of the single characters, and rank the results in according with the relative magnitude of the numbers counted, we may find a result as shown in Table 2.

In this table a total of eight sets of probability range of misclassification are given. In the set of $p=0-0.05$, the difference between any two varieties

Table 1. *Coefficient of discrimination* (k) *and probability of misclassification* (*p*) *for single characters*

Variety combination		A′ k	A′ *p*	B′ k	B′ *p*	C′ k	C′ *p*	D′ k	D′ *p*	E′ k	E′ *p*	F′ k	F′ *p*	G′ k	G′ *p*	H′ k	H′ *p*	I′ k	I′ *p*	J′ k	J′ *p*	K′ k	K′ *p*
A	B	0	.500	0.60	.290	0.62	.288	0	.500	0.19	.400	0.02	.490	0	.500	0.13	.425	0.61	.289	0.05	.468	0.12	.430
	C	1.04	.239	1.93	.167	0.45	.322	0.69	.276	0	.500	0.46	.320	1.10	.230	1.47	.195	0.44	.324	2.10	.155	0.46	.320
	D	1.37	.202	1.50	.195	0.06	.464	0.02	.490	0.03	.481	0.07	.458	0.08	.452	0.25	.378	0.21	.392	1.89	.167	0.01	.495
	E	0.02	.490	0	.500	0.02	.490	0.02	.490	0.06	.464	0.04	.473	0	.500	0	.500	0.01	.495	0.50	.309	0.01	.495
	F	0.04	.473	0.01	.495	0.01	.495	0.30	.362	0	.500	0.42	.330	0.70	.275	0.63	.286	0	.500	0.73	.270	0.19	.400
	G	0.03	.481	0.71	.273	1.05	.230	0.95	.244	0.01	.495	0.60	.290	0.01	.495	0.03	.481	0.63	.286	1.00	.239	0.01	.495
	H	0.61	.289	0.36	.345	1.63	.186	5.42	.051	0.68	.278	2.14	.155	0.21	.392	0.11	.435	0.41	.332	0	.500	0.09	.446
	I	0.01	.495	1.15	.221	1.69	.180	1.06	.230	0.15	.416	1.93	.167	0.63	.286	0.55	.300	0.01	.495	0.07	.458	0.58	.294
B	C	1.39	.202	4.00	.079	2.31	.143	0.77	.264	0.10	.440	0.65	.282	0.79	.261	0.62	.288	0.03	.481	2.00	.159	0.07	.458
	D	1.38	.202	4.19	.075	1.27	.210	0.02	.490	0.02	.490	0.02	.490	0.04	.473	0.01	.495	0.14	.420	1.91	.167	0.06	.464
	E	0.01	.495	0.54	.302	0.47	.317	0.02	.490	0.30	.362	0.10	.440	0	.500	0.06	.464	0.43	.330	0.35	.347	0.05	.468
	F	0.02	.490	0.35	.347	0.53	.300	0.34	.350	0.14	.420	0.59	.292	0.49	.312	0.14	.420	0.83	.257	0.58	.294	0	.500
	G	0.04	.473	2.26	.143	3.63	.092	0.97	.242	0.14	.420	0.51	.308	0.02	.490	0.01	.495	0	.500	0.87	.253	0.11	.435
	H	0.72	.272	1.70	.180	4.83	.061	5.53	.050	0.21	.392	1.96	.159	0.23	.385	0.25	.378	0.04	.473	0.02	.490	0.22	.388
	I	0.02	.490	3.08	.108	5.35	.051	1.07	.230	0.03	.481	1.78	.173	0.65	.282	0.70	.275	0.42	.330	0.19	.400	0.81	.259
C	D	0.04	.473	0.21	.392	0.21	.392	2.08	.155	0.24	.382	1.00	.239	0.73	.270	0.79	.261	0.05	.468	0.18	.404	0.29	.365
	E	0.99	.240	1.83	.173	0.70	.275	1.55	.186	0.04	.473	0.27	.371	0.84	.256	1.14	.230	0.28	.368	1.09	.230	0.26	.375
	F	1.91	.167	0.98	.241	0.22	.388	0.16	.412	0	.500	0	.500	0.06	.464	0.24	.382	0.60	.290	0.36	.345	0.05	.468
	G	2.84	.120	0.30	.362	0.13	.425	2.54	.134	0	.500	1.32	.210	0.57	.296	0.49	.312	0.10	.440	0.36	.345	0.29	.365
	H	5.48	.050	0.67	.279	0.36	.345	9.07	.017	0.58	.294	3.45	.095	0.70	.275	0.69	.276	0	.500	1.75	.173	0.37	.342
	I	1.83	.173	0.25	.378	0.34	.350	2.15	.147	0.15	.416	2.94	.116	1.09	.230	1.04	.239	0.28	.368	2.03	.159	1.06	.230
D	E	0.93	.246	1.38	.202	0.17	.408	0	.500	0.62	.288	0.23	.385	0.05	.468	0.14	.420	0.11	.435	0.73	.270	0	.5 (0
	F	2.31	.143	0.54	.302	0.01	.495	1.02	.239	0.33	.353	0.88	.252	0.38	.340	0.14	.420	0.27	.371	0.08	.452	0.10	.440
	G	3.34	.102	0.03	.481	0.70	.275	1.00	.239	0.32	.357	0.43	.327	0.07	.458	0.03	.481	0.16	.412	0.07	.458	0.02	.490
	H	6.52	.037	0.24	.382	1.21	.221	6.10	.041	0.15	.416	1.84	.173	0.32	.357	0.30	.362	0.04	.473	1.41	.202	0.11	.435
	I	1.89	.167	0	.500	1.25	.210	1.07	.230	0.01	.495	1.68	.180	0.75	.267	0.74	.269	0.11	.435	1.67	.180	0.62	.288
E	F	0	.500	0.01	.495	0.04	.473	0.77	.264	0.03	.481	0.25	.378	0.53	.300	0.44	.324	0.02	.490	0.13	.425	0.09	.446
	G	0.09	.446	0.67	.279	1.45	.195	0.93	.246	0.02	.490	0.79	.261	0.01	.495	0.01	.495	0.44	.324	0.22	.388	0.03	.481
	H	0.81	.259	0.34	.350	2.17	.147	5.74	.046	1.02	.239	2.54	.134	0.21	.392	0.15	.416	0.25	.378	0.34	.350	0.12	.430
	I	0.05	.468	1.08	.230	2.30	.143	1.02	.239	0.27	.371	2.23	.147	0.63	.286	0.60	.290	0	.500	0.62	.288	0.64	.284
F	G	0.22	.388	0.30	.362	0.59	.292	2.01	.159	0	.500	1.28	.210	0.39	.337	0.16	.412	0.86	.254	0	.500	0.15	.416
	H	1.56	.186	0.12	.430	0.94	.245	8.09	.023	0.71	.273	3.32	.102	0.58	.294	0.45	.322	0.56	.298	0.57	.296	0.26	.375
	I	0.10	.440	0.49	.312	0.93	.246	1.79	.173	0.18	.404	2.87	.116	1.00	.239	0.87	.253	0.02	.490	0.86	.254	0.89	.251
G	H	0.58	.294	0.07	.458	0.06	.464	1.38	.202	0.69	.276	0.24	.382	0.14	.420	0.17	.408	0.05	.468	0.77	.264	0.04	.473
	I	0	.500	0.03	.481	0.04	.473	0.07	.458	0.18	.404	0.34	.350	0.54	.302	0.62	.288	0.44	.324	1.06	.230	0.41	.332
H	I	0.50	.309	0.21	.392	0	.500	0.53	.300	0.02	.490	0.02	.490	0.17	.408	0.25	.378	0.25	.378	0.07	.458	0.17	.408

Table 2. *Numbers of variety combinations and ranks (in parenthesis) assigned to eleven single characters by eight probability ranges of misclassification*

Prob. range	A′	B′	C′	D′	E′	F′	G′	H′	I′	J′	K′
0-0.05	2(2)	0(4)	1(3)	5(1)	0(4)	0(4)	0(4)	0(4)	0(4)	0(4)	0(4)
0-0.10	2(3)	2(3)	4(2)	6(1)	0(5)	1(4)	0(5)	0(5)	0(5)	0(5)	0(5)
0-0.20	9(3)	8(4)	10(2)	12(1)	0(7)	12(1)	0(7)	1(6)	0(7)	7(5)	0(7)
0-0.25	14(3)	12(4)	16(2)	22(1)	1(7)	16(2)	3(6)	3(6)	0(8)	11(5)	1(7)
0-0.30	19(4)	16(5)	21(2)	26(1)	6(8)	20(3)	15(6)	13(7)	6(8)	19(4)	6(8)
0-0.33	20(4)	19(5)	23(3)	26(1)	6(10)	24(2)	17(6)	16(7)	10(8)	20(4)	7(9)
0-0.40	21(6)	28(2)	27(3)	28(2)	13(9)	29(1)	23(5)	21(6)	18(7)	26(4)	14(8)
0-0.50	36(1)	36(1)	36(1)	36(1)	36(1)	36(1)	36(1)	36(1)	36(1)	36(1)	36(1)

could be discriminated by any single character, in the set of $p=0-0.40$, about 2/3 of variety combinations can be discriminated, and in the set of $p=0-0.25$, 22 variety combinations are discriminated by plant height (D′).

For a datum consisting of m sets ($m=7$, excluded the set of $p=0-0.50$) of ranks, a descriptive measure of the agreement or concordance between these m sets is provided by Kendall's (1955) coefficient of concordance (w),

$$w=\frac{12S}{m^2(N^3-N)},$$

where, S = the sum of squares of mean diviations of the total ranks of the m sets. m = the number of sets of ranks, N = the number of ranks in each sets. Critical values of w depend on both m and N. For $N>7$, a χ^2-test may be applied by using, $\chi^2=m(N-1)w$ with N−1 degrees of freedom. For the data of Table 2, S=3072, $m=7$, N=11 and $w=+0.57$, $\chi^2=39.90^{**}$. When $df=11-1=10$, the value of Chi-square required for significance are 18.31 and 23.21 at the 5% and 1% levels respectively. Therefore, it may be concluded that there was concordance between the m sets. In other words, the relative degree of discrimination of a single character among the m sets remained unchanged.

The degree of discrimination of the various characters were found to be in the following descending order: plant height, maturity date, no. of nodes on main stem, flowering time, period from flowering to maturity, wt. of 100 seeds, no. of pods, no. of seeds, no. of seeds per pod, seed yield and no. of branches. From this result, it can be concluded that plant height is the best, while the no. of branches the worst, when each is used as a single character for discriminating the difference between two varieties. The difference in no. of branches between two varieties was too small, and, moreover, this single character is liable to be affected by the "planting spaces". The observed seed yield was no doubt an easy measure, but not quite reliable since it is affected by both the environment factors and the genotypes of it's sub-characters (i. e., no. of pods, no. of seeds per pod and wt. of 100 seeds). We, therefore, gave up the last two single characters, and selected the best nine single characters for constructing the discriminant function.

B. Construction of the discriminant function for metric characters of soybean varieties

The nine quantitative charactes, which were slected by using the method discussed in the previous section were divided into three groups:

Group 1, physiological characters; i. e., flowering time (C_1), period from flowering to maturity (C_2) and maturity date (C_3).

Group 2, characters of seed yield component; i. e., no. of pods (C_4), no. of seeds per pod (C_5) and wt. of 100 seeds (C_6).

Group 3, other characters; i. e., plant height (C_7), no. of nodes on main stem (C_8) and no. of seeds (C_9).

The analysis of variances and covariances of these character groups are given respectively in Table 3, 4, and 5.

The discriminant function was constructed for each character group based on the principle of maximizing the difference between varieties and minimizing the difference within varieties. This method has been described by Mather (1949) and Rao (1952). The standardized best linear discriminant function of the three character groups are given respectively as following:

$$\left.\begin{aligned} X_1' &= 0.23570C_1 - 0.03709C_2 + 0.20121C_3 \\ X_2' &= 0.03957C_4 + 0.00014C_5 - 0.00640C_6 \\ X_3' &= 0.06695C_7 + 0.00996C_8 + 0.00940C_9 \end{aligned}\right\} \quad (5)$$

The means of discriminant values of equation (5) for each variety are presented in Table 6.

Table 6. *Showing the mean of discriminant value for each variety*

Variety	X_1'	X_2'	X_3'
Taichung Green	25.0495	1.1973	2.6073
Tsao-ta-tou	24.4105	1.1535	2.4725
Jackson	24.2905	0.5945	1.8475
Shih-shih	24.4057	1.0098	2.5368
Japanese Green	24.7468	1.1703	2.9395
Sheng-liang	25.0638	0.7243	2.0913
Wakashima	26.0945	1.2480	3.4343
Palmetto	27.0273	1.9050	5.0800
Mi-yueh	26.1733	3.0423	4.7935

Table 3. *Analysis of variance and covariance of the characters of Group 1*

Source of variation	*df*	C_1	C_2	C_3	$C_1 \times C_2$	$C_2 \times C_3$	$C_3 \times C_1$	X'_1
between varieties	8	254.8750	537.7500	474.0695	456.6250	211.7500	235.3750	40.4419
within varieties	351	7.6695	17.1282	12.7600	16.3914	6.2143	7.4971	1.0905
F-value		33.2323**	31.3956**	37.1528**				37.0856**

Table 4. *Analysis of variance and covariance of the characters of Group 2 and seed yield*

Source of variation	*df*	C_4	C_5	C_6	$C_4 \times C_5$	$C_5 \times C_6$	$C_6 \times C_4$	X'_2	seed yield
between varieties	8	13,430.2500	0.9249	302.1250	0.9111	292.8750	9,788.3750	21.7664	667.5000
within varieties	351	637.3647	0.0900	9.7436	0.0902	9.7543	638.2229	0.9998	47.0541
F-value		21.0715**	10.2767**	31.0075**				21.7708**	14.1858**

Table 5. *Analysis of variance and covariance of the characters of Group 3*

Source of variation	*df*	C_7	C_8	C_9	$C_7 \times C_8$	$C_8 \times C_9$	$C_9 \times C_7$	X'_3
between varieties	8	6,587.2500	198.8750	58,245.6250	68.1250	15,534.8750	4,232.8750	53.1736
within varieties	351	78.5869	3.7094	2,835.6866	-3.1914	1,515.4857	50.9886	0.9687
F-value		83.8212**	53.6137**	20.5402**				54.8917**

**: Significant at 1% level

The test of a discriminant function of each character group by an analysis of variance is also shown in Table 3, 4, and 5, which shows that the F-value of the discriminant function of each character group is either greater than or approximately equal to the F-values of any single characters within the group, except for the characters such as wt. of 100 seeds (C_6) and plant height (C_7). In other words, the discriminant function of character group is effective in discriminating the difference between varieties, but gives no better discrimination power than the best single character in each character group. In Table 4, the F-value of Group 2 (the discriminant function of this character group can be regarded as a compound character relative to the respective yield components) is greater than the F-value of actual seed yield. In other words, it is more difficult to distinguish the difference between any two varieties by the character of actual seed yield than by the compound character corresponding to seed yield.

Using the derived X_1', X_2', and X_3' values, we may obtain the best two standardized linear discriminant functions of nine metric characters by the similar procedure described above;

$$X_a' = 0.48773X_1' + 0.34234X_2' + 0.59386X_3'$$
$$= 0.11496C_1 - 0.01809C_2 + 0.09814C_3 + 0.01355C_4 + 0.00005C_5$$
$$- 0.00219C_6 + 0.03976C_7 + 0.00591C_8 + 0.00558C_9 \quad (6)$$
$$X_b' = 0.95896X_1' + 0.51151X_2' - 0.05351X_3'$$
$$= 0.22603C_1 - 0.03557C_2 + 0.19295C_3 + 0.02024C_4 + 0.00007C_5$$
$$- 0.00327C_6 - 0.00358C_7 - 0.00053C_8 - 0.00050C_9 \quad (7)$$

The analysis of variance of these two discriminant functions were calculated and shown in Table 7, from which we find that the F-value of X_a' is greater than the F-value of X_b' and also greater than that of X_1', X_2' and X_3'.

Table 7. *Analysis of variance for two discriminant functions of nine metric characters*

Source of variation	*df*	X_a'	X_b'
between varieties	8	309.7627	61.6991
within varieties	351	4.2036	1.0874
F-value		73.6899**	56.7400**

**: Significant at 1% level

Incidently, both of the F-values of X_a' and X_b' are smaller than that of the plant height (C_7), but greater than those of other single characters. Therefore, the discriminant function (6) is the most efficient one among all these discriminant functions thus far constructed, but it still gives no better discrimination than the best single character.

C. Discrimination of soybean varieties based on their discriminant values

The mean values for the equations (6) and (7) for each variety are given in Table 8. These values can be used to estimate the parameters of elliptic equation given in equation (2) with equal probability deviation ($\alpha=0.05$ or 0.01) for each variety. The results are shown in Figure 1. The new origine, angle of rotation, and the coefficient of the elliptic equation are also given in Table 8.

Table 8. *Mean value, angle of rotation and the coefficient of the elliptic equation*

Variety	Mean value (new origine)		Angle of rotation	Coef. of elliptic equation			
				$\alpha=0.05$		$\alpha=0.01$	
	$\bar{X}'_a$	$\bar{X}'_b$	θ	$\pm a$	$\pm b$	$\pm a$	$\pm b$
Taichung Green	14.18	24.49	−35°23′	0.18	1.56	0.21	1.81
Tsao-ta-tou	13.77	23.87	−37°44′	0.16	1.22	0.19	1.41
Jackson	13.15	23.50	−38°38′	0.08	0.28	0.10	0.32
Shih-shih	13.58	23.45	−38°52′	0.09	0.30	0.11	0.44
Japanese Green	14.05	24.21	−32°50′	0.14	0.96	0.17	1.12
Sheng-liang	13.71	24.29	−36°10′	0.10	0.42	0.12	0.47
Wakashima	15.19	25.48	+43°10′	0.68	0.18	0.80	0.20
Palmetto	16.85	26.62	+34°42′	0.70	0.17	0.81	0.19
Mi-yueh	16.65	26.40	+22°30′	0.22	0.22	0.25	0.25

In Figure 1, the variety Jackson (an American variety which we call Group I here) occupies a position at one end of the scale, while the varieties, Palmetto and Mi-yueh, are spread out towards the other extreme. These latter varieties we call Group V, which show the same characteristics, when planted under the same environmental conditions. Thus they are similar in compound metric characters. Palmetto was an old Chinese variety, which was introduced to the United Statees from Naking in 1927, and to Taiwan in 1953. Mi-yueh is a Japanese variety. These two varieties were all native to Asia. The other four varieties, Taichung Green, Japanese Green, Tsao-ta-tou, and Shih-shih, all lying very close together in the figure, we call Group III. The first two varieties, Taichung Green and Japanese Green, have the same origin. Taichung Green was selected from Japanese Green by the College of Agric., N. T. U.. The other two varieties are also Japanese varieties. Shih-shih was introduced from Japanese in 1954. These later four varieties show the same characteristics in Taiwan. The other two groups, Sheng-liang (Group II) and Wakashima (Group IV), lie very close to Group III, but are clearly distinct from the latter. It may be concluded that the characteristics of the varieties within each group are similar.

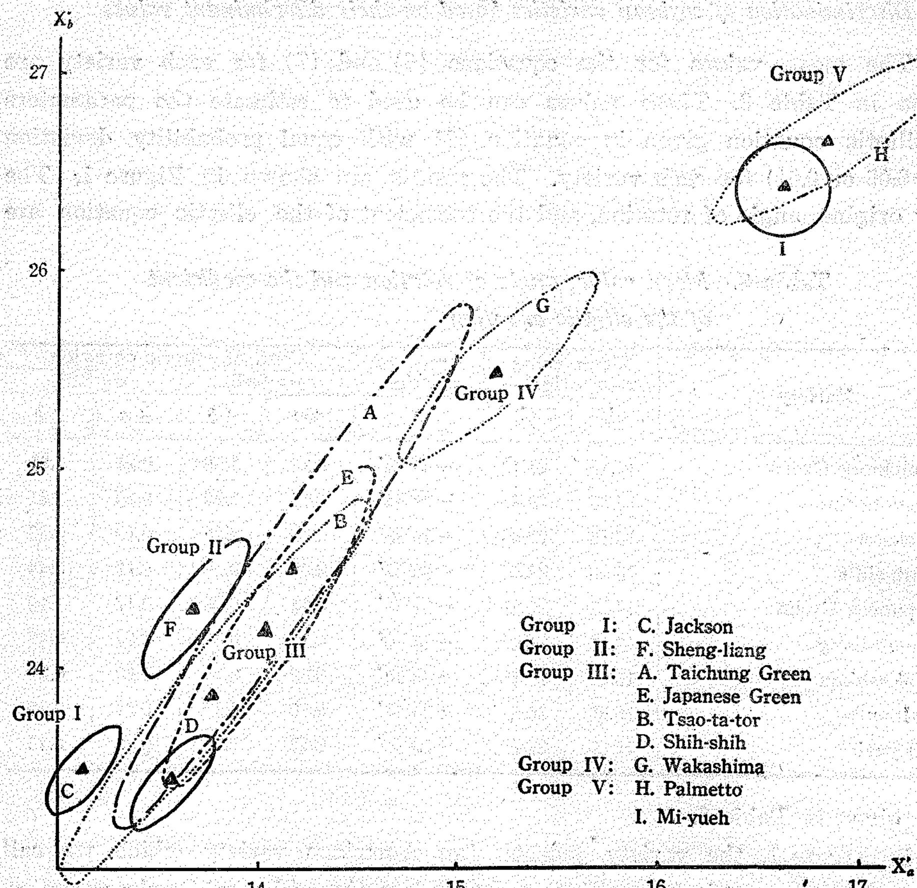

Figure 1. The configuration of the variety group of soybean varieties

D. Classification of F_2 hybrids in each cross-combination

For metric characters, it is considerably difficult to select genetically desirable type in breeding work, because heritable differences are to some extent masked by non-heritable variations. In this study, we used the discriminant function of the parents to estimate the discriminant value of F_2 individuals, and to determine the characteristic of F_2 plants.

The discriminant value of F_2 individual plants was estimated by the equations (6) and (7). The graphical method of a two-dimensional representation was used to denote the characteristic of the F_2 plants, the results being shown in Figure 2. From Fig. 2 we infer that most of the F_2 plants lie outside the 5% admissible range of the parents. However, it is not clear, which group they are likely to belong to. But a desirable cross-comination may be selected by examining the variation pattern of their discriminant values, and it is possible to obtain a desirable individual plant also from the combination which may show a large variation. In other words, large deviation of discriminant values can be obtained by inter-group crossing. In this study we have selected five desirable combinations, i. e., A×G, B×G, C×G, C×H, and E×H, from which the desired individual plants are to be chosen.

The above results suggest that this method may help in the quantitative evaluation of the likeness of different varieties. Thus we may apply this quantitative approach to the construction of a natural system of classification

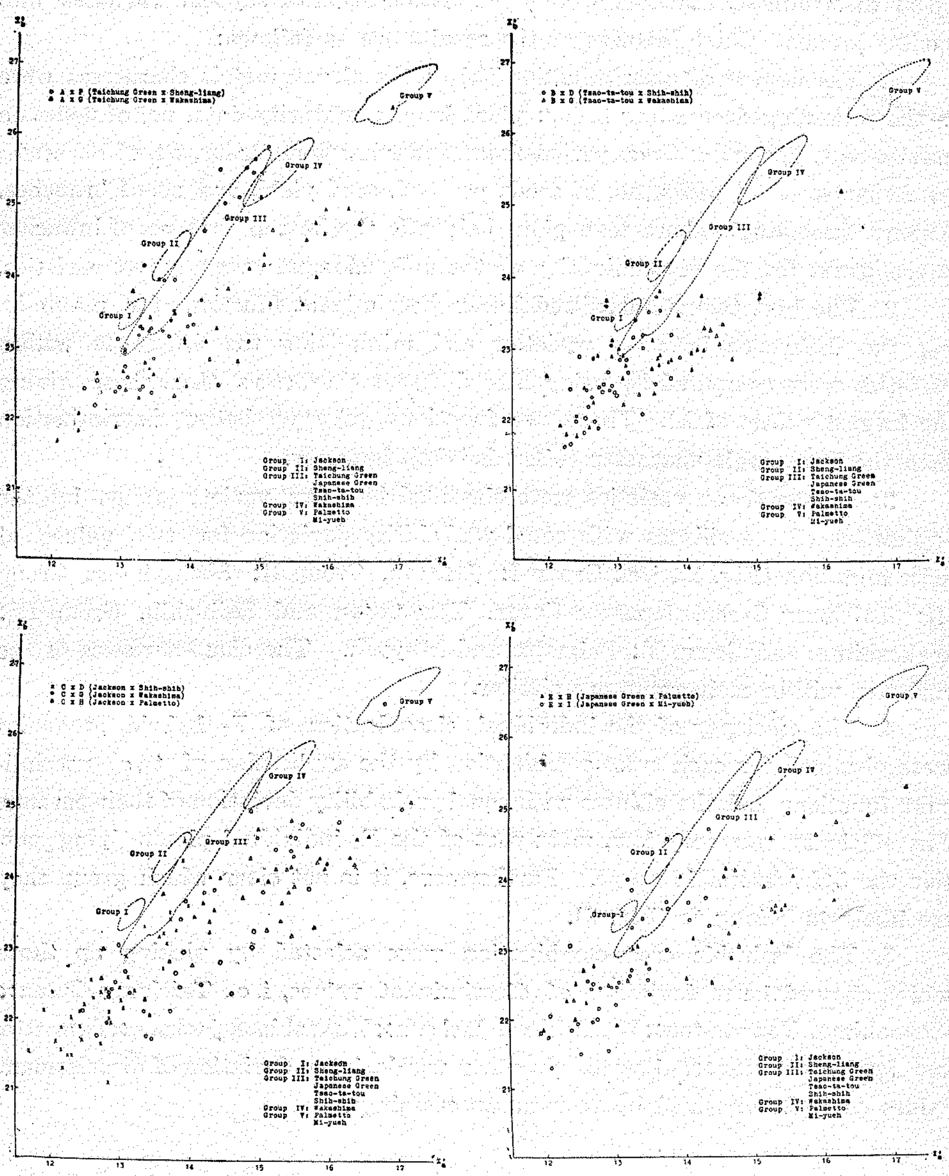

Figure 2. The distribution of discriminant value of F_2 hybrids

for various organisms, to the comparisons of taxons of the same level in different group, and to the earlier selection of desirable combinations from two or more crosses using records of the individual progeny.

Summary

Eleven metric characters of nine soybean varieties and the nine cross-combinations of their F_2 hybrids progeny were used in the construction of the linear discriminant function, and in the classification of soybean varieties into variety-groups. Chief features of the results are as follows:

1. The degree of discrimination (k) of the eleven metric characters were in the following descending order: plant height, maturity date, no. of nodes on main stems, flowering time, period from flowering to maturity, wt. of 100 seeds, no. of pods, no. of seeds, no. of seeds per pod, seed yield, and no. of branches. Among these single characters, plant height is the best and the no. of branches is the worst for the discrimination of the difference between any two varieties.

2. The best two standardized linear discriminant functions for maximizing the difference between varieties and minimizing the difference within varieties were computed by using nine different characters (i. e., those giving the largest values of k). These two functions gave no better discrimination than the best single character (plant height), however.

3. Nine soybean varieties were classified into five variety-groups by the application of the ellipses with equal probability deviation for two values of discrimination (Fig. 1), i. e., Group I: Jackson, Group II: Sheng-liang, Group III: Taichung Green, Japanese Green, Tsao-ta-tou and Shih-shih, Group IV: Wakashima, and Group V: Palmetto and Mi-yueh. The characteristics of the varieties within each group are similar.

4. The efficiency of the individual classification of F_2 hybrid progenies were obtained for each cross-combination by the application of two discriminant functions and the ellipses with equal probability deviation of their parents. The efficiency was very low, most part of the F_2 individual plants lying outside the 5% admissible range. Furthermore, it is not clear which group they are likely to belong to (Fig. 2).

5. The desirable cross-combination were selected by picking up those which showed large deviation of discriminant values, i. e., Taichung Green×Wakashima, Tsao-ta-tou×Wakashima, Jackson×Wakashima, Jackson×Palmetto, and Japanese Green×Palmetto. In other words, large deviation of discriminant values could be obtained by inter-group crossings.

大豆品種特性之判別函數研究

鄔 宏 潘

本研究係以九種大豆品種及其九種雜交組合之 F_2 後裔爲材料，研究大豆品種之數量性狀判別函數，及品種特性羣之分類，主要結果有：

1. 十一種數量性狀之判別能力不等，如依次序排之爲：株高、成熟期、主莖節數、開花日數、結莢日數、百粒重、總莢數、總粒數、每莢種子數、產量及枝數。株高爲判別任何二品種間差異最有效之性狀，而分枝數則反之。

2. 以品種間之差異爲最大而品種內之差異爲最小而估計之九種數量性狀之品種特性判別函數的判別效果不如單性狀之株高爲大。

3. 利用品種判別函數及其等機率偏差橢圓，將九種大豆品種分成五種品種特性羣，如第Ⅰ羣：Jackson，第Ⅱ羣：生娘，第Ⅲ羣：臺中青皮豆，日本青皮豆，操大豆，及十石，第Ⅳ羣：和歌島，第Ⅴ羣：百美豆及米岳。同一羣內各品種間之特性極爲相似。

4. 以親本之判別函數及其等機率偏差橢圓推測各什交組合 F_2 後裔之特性，其結果不甚理想，大部份 F_2 個體之判別值都落於親本之等機率偏差橢圓外，不能確實指認 F_2 單株之特性。

5. 由各什交組合 F_2 後裔所表現之判別值分佈趨勢，可推測變異性大之優良什交組合，如臺中青皮豆×和歌島，操大豆×和歌島，Jackson×和歌島，Jackson×百美豆及日本青皮豆×百美豆。這些什交組合都是不同特性羣間之品種什交。而同一特性羣內之品種什交所得後裔之變異性一般較小。故欲得優良之什交組合時，應選擇不同特性羣間之品種什交之。

Literature Cited

BARTLETT, M. S. The standard errors of discriminant function coefficients. J. R. S. S., Suppl. **6**: 169, 1939.

BARTLETT, M. S. Multivariate analysis, J. R. S. S., B. **9**:176, 1947.

FISHER, R. A. The use of multiple measurements in taxanomic problems. Ann. Eugen. **7**: 179, 1936.

KENDALL, M. G. Rank correlation methods. Charles Greffin & Co., 1955.

KENDALL, M. G. A course in multivariate analysis. Charles Greffin & Co., 1957.

HOSOKAWA, S. K., K. TAKEDA and TAKEDA, T. Selection Studies in sugar beet breeding. II. Bivariate selection for root weight and sucrose by application of ellipses with equal probability deviation. Jap. J. Breed. **14** :98, 1964.

LU, Y. C. The correlations between the agronomic characters and the ecotypes of soybean varieties. J. Agri. Asso. of China. **34**: 23, 1961.

LUBISCHEW, A. A. On the use of discriminant functions in taxanomy. Biometrics, **18**: 455, 1962.

MATHER, K. Biometrical genetics. Methuen and Co., 1949.

MOOD, A. M., and F. A. GRAYBILL Introduction to the theory of statistics. McGraw-Hill Book Co., 1963.

MORSE, W. J., and J. L. CARTTER Soybeans: Culture and varieties. U. S. D. A., Farmer's Bull. 1520, 1949.

RAO, C. R. Advanced statistical methods in biometric research. John Wiley & Sons, Inc, 1952.