THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS

By R. A. FISHER, Sc.D., F.R.S.

I. Discriminant functions

When two or more populations have been measured in several characters, $x_1, \ldots, x_s$, special interest attaches to certain linear functions of the measurements by which the populations are best discriminated. At the author's suggestion use has already been made of this fact in craniometry (*a*) by Mr E. S. Martin, who has applied the principle to the sex differences in measurements of the mandible, and (*b*) by Miss Mildred Barnard, who showed how to obtain from a series of dated series the particular compound of cranial measurements showing most distinctly a progressive or secular trend. In the present paper the application of the same principle will be illustrated on a taxonomic problem; some questions connected with the precision of the processes employed will also be discussed.

II. Arithmetical procedure

Table I shows measurements of the flowers of fifty plants each of the two species *Iris setosa* and *I. versicolor*, found growing together in the same colony and measured by Dr E. Anderson, to whom I am indebted for the use of the data. Four flower measurements are given. We shall first consider the question: What linear function of the four measurements

$$X = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \lambda_4 x_4$$

will maximize the ratio of the difference between the specific means to the standard deviations within species? The observed means and their differences are shown in Table II. We may represent the differences by d_p, where $p = 1, 2, 3$ or 4 for the four measurements.

The sums of squares and products of deviations from the specific means are shown in Table III. Since fifty plants of each species were used these sums contain 98 degrees of freedom. We may represent these sums of squares or products by S_{pq}, where p and q take independently the values 1, 2, 3 and 4.

Then for any linear function, X, of the measurements, as defined above, the difference between the means of X in the two species is

$$D = \lambda_1 d_1 + \lambda_2 d_2 + \lambda_3 d_3 + \lambda_4 d_4,$$

while the variance of X within species is proportional to

$$S = \sum_{p=1}^{4} \sum_{q=1}^{4} \lambda_p \lambda_q S_{pq}.$$

The particular linear function which best discriminates the two species will be one for

Table I

Iris setosa				*Iris versicolor*				*Iris virginica*			
Sepal length	Sepal width	Petal length	Petal width	Sepal length	Sepal width	Petal length	Petal width	Sepal length	Sepal width	Petal length	Petal width
5·1	3·5	1·4	0·2	7·0	3·2	4·7	1·4	6·3	3·3	6·0	2·5
4·9	3·0	1·4	0·2	6·4	3·2	4·5	1·5	5·8	2·7	5·1	1·9
4·7	3·2	1·3	0·2	6·9	3·1	4·9	1·5	7·1	3·0	5·9	2·1
4·6	3·1	1·5	0·2	5·5	2·3	4·0	1·3	6·3	2·9	5·6	1·8
5·0	3·6	1·4	0·2	6·5	2·8	4·6	1·5	6·5	3·0	5·8	2·2
5·4	3·9	1·7	0·4	5·7	2·8	4·5	1·3	7·6	3·0	6·6	2·1
4·6	3·4	1·4	0·3	6·3	3·3	4·7	1·6	4·9	2·5	4·5	1·7
5·0	3·4	1·5	0·2	4·9	2·4	3·3	1·0	7·3	2·9	6·3	1·8
4·4	2·9	1·4	0·2	6·6	2·9	4·6	1·3	6·7	2·5	5·8	1·8
4·9	3·1	1·5	0·1	5·2	2·7	3·9	1·4	7·2	3·6	6·1	2·5
5·4	3·7	1·5	0·2	5·0	2·0	3·5	1·0	6·5	3·2	5·1	2·0
4·8	3·4	1·6	0·2	5·9	3·0	4·2	1·5	6·4	2·7	5·3	1·9
4·8	3·0	1·4	0·1	6·0	2·2	4·0	1·0	6·8	3·0	5·5	2·1
4·3	3·0	1·1	0·1	6·1	2·9	4·7	1·4	5·7	2·5	5·0	2·0
5·8	4·0	1·2	0·2	5·6	2·9	3·6	1·3	5·8	2·8	5·1	2·4
5·7	4·4	1·5	0·4	6·7	3·1	4·4	1·4	6·4	3·2	5·3	2·3
5·4	3·9	1·3	0·4	5·6	3·0	4·5	1·5	6·5	3·0	5·5	1·8
5·1	3·5	1·4	0·3	5·8	2·7	4·1	1·0	7·7	3·8	6·7	2·2
5·7	3·8	1·7	0·3	6·2	2·2	4·5	1·5	7·7	2·6	6·9	2·3
5·1	3·8	1·5	0·3	5·6	2·5	3·9	1·1	6·0	2·2	5·0	1·5
5·4	3·4	1·7	0·2	5·9	3·2	4·8	1·8	6·9	3·2	5·7	2·3
5·1	3·7	1·5	0·4	6·1	2·8	4·0	1·3	5·6	2·8	4·9	2·0
4·6	3·6	1·0	0·2	6·3	2·5	4·9	1·5	7·7	2·8	6·7	2·0
5·1	3·3	1·7	0·5	6·1	2·8	4·7	1·2	6·3	2·7	4·9	1·8
4·8	3·4	1·9	0·2	6·4	2·9	4·3	1·3	6·7	3·3	5·7	2·1
5·0	3·0	1·6	0·2	6·6	3·0	4·4	1·4	7·2	3·2	6·0	1·8
5·0	3·4	1·6	0·4	6·8	2·8	4·8	1·4	6·2	2·8	4·8	1·8
5·2	3·5	1·5	0·2	6·7	3·0	5·0	1·7	6·1	3·0	4·9	1·8
5·2	3·4	1·4	0·2	6·0	2·9	4·5	1·5	6·4	2·8	5·6	2·1
4·7	3·2	1·6	0·2	5·7	2·6	3·5	1·0	7·2	3·0	5·8	1·6
4·8	3·1	1·6	0·2	5·5	2·4	3·8	1·1	7·4	2·8	6·1	1·9
5·4	3·4	1·5	0·4	5·5	2·4	3·7	1·0	7·9	3·8	6·4	2·0
5·2	4·1	1·5	0·1	5·8	2·7	3·9	1·2	6·4	2·8	5·6	2·2
5·5	4·2	1·4	0·2	6·0	2·7	5·1	1·6	6·3	2·8	5·1	1·5
4·9	3·1	1·5	0·2	5·4	3·0	4·5	1·5	6·1	2·6	5·6	1·4
5·0	3·2	1·2	0·2	6·0	3·4	4·5	1·6	7·7	3·0	6·1	2·3
5·5	3·5	1·3	0·2	6·7	3·1	4·7	1·5	6·3	3·4	5·6	2·4
4·9	3·6	1·4	0·1	6·3	2·3	4·4	1·3	6·4	3·1	5·5	1·8
4·4	3·0	1·3	0·2	5·6	3·0	4·1	1·3	6·0	3·0	4·8	1·8
5·1	3·4	1·5	0·2	5·5	2·5	4·0	1·3	6·9	3·1	5·4	2·1
5·0	3·5	1·3	0·3	5·5	2·6	4·4	1·2	6·7	3·1	5·6	2·4
4·5	2·3	1·3	0·3	6·1	3·0	4·6	1·4	6·9	3·1	5·1	2·3
4·4	3·2	1·3	0·2	5·8	2·6	4·0	1·2	5·8	2·7	5·1	1·9
5·0	3·5	1·6	0·6	5·0	2·3	3·3	1·0	6·8	3·2	5·9	2·3
5·1	3·8	1·9	0·4	5·6	2·7	4·2	1·3	6·7	3·3	5·7	2·5
4·8	3·0	1·4	0·3	5·7	3·0	4·2	1·2	6·7	3·0	5·2	2·3
5·1	3·8	1·6	0·2	5·7	2·9	4·2	1·3	6·3	2·5	5·0	1·9
4·6	3·2	1·4	0·2	6·2	2·9	4·3	1·3	6·5	3·0	5·2	2·0
5·3	3·7	1·5	0·2	5·1	2·5	3·0	1·1	6·2	3·4	5·4	2·3
5·0	3·3	1·4	0·2	5·7	2·8	4·1	1·3	5·9	3·0	5·1	1·8

Table II. *Observed means for two species and their difference (cm.)*

	Versicolor	*Setosa*	Difference ($V-S$)
Sepal length (x_1)	5·936	5·006	0·930
Sepal width (x_2)	2·770	3·428	−0·658
Petal length (x_3)	4·260	1·462	2·798
Petal width (x_4)	1·326	0·246	1·080

Table III. *Sums of squares and products of four measurements, within species (cm.^{2})*

	Sepal length	Sepal width	Petal length	Petal width
Sepal length	19·1434	9·0356	9·7634	3·2394
Sepal width	9·0356	11·8658	4·6232	2·4746
Petal length	9·7634	4·6232	12·2978	3·8794
Petal width	3·2394	2·4746	3·8794	2·4604

which the ratio D^2/S is greatest, by variation of the four coefficients λ_1, λ_2, λ_3 and λ_4 independently. This gives for each λ

$$\frac{D}{S^2}\left\{2S\frac{\partial D}{\partial \lambda}-D\frac{\partial S}{\partial \lambda}\right\}=0,$$

or

$$\frac{1}{2}\cdot\frac{\partial S}{\partial \lambda}=\frac{S}{D}\frac{\partial D}{\partial \lambda},$$

where it may be noticed that S/D is a factor constant for the four unknown coefficients. Consequently, the coefficients required are proportional to the solutions of the equations

$$\left.\begin{aligned}S_{11}\lambda_1+S_{12}\lambda_2+S_{13}\lambda_3+S_{14}\lambda_4=d_1,\\ S_{12}\lambda_1+S_{22}\lambda_2+S_{23}\lambda_3+S_{24}\lambda_4=d_2,\\ S_{13}\lambda_1+S_{23}\lambda_2+S_{33}\lambda_3+S_{34}\lambda_4=d_3,\\ S_{14}\lambda_1+S_{24}\lambda_2+S_{34}\lambda_3+S_{44}\lambda_4=d_4.\end{aligned}\right\} \qquad \text{......(1)}$$

If, in turn, unity is substituted for each of the differences and zero for the others, the solutions obtained constitute the matrix of multipliers reciprocal to the matrix of S; numerically we find:

Table IV. *Matrix of multipliers reciprocal to the sums of squares and products within species (cm.$^{-2}$)*

	Sepal length	Sepal width	Petal length	Petal width
Sepal length	0·1187161	−0·0668666	−0·0816158	0·0396350
Sepal width	−0·0668666	0·1452736	0·0334101	−0·1107529
Petal length	−0·0816158	0·0334101	0·2193614	−0·2720206
Petal width	0·0396350	−0·1107529	−0·2720206	0·8945506

These values may be denoted by s_{pq} for values of p and q from 1 to 4.

Multiplying the columns of the matrix in Table IV by the observed differences, we have the solutions of the equation (1) in the form

$$\lambda = -0{\cdot}0311511, \quad \lambda_2 = -0{\cdot}1839075, \quad \lambda_3 = +0{\cdot}2221044, \quad \lambda_4 = +0{\cdot}3147370,$$

so that, if we choose to take the coefficient of sepal length to be unity, the compound measurement required is

$$X = x_1 + 5{\cdot}9037x_2 - 7{\cdot}1299x_3 - 10{\cdot}1036x_4.$$

If in this expression we substitute the values observed in *setosa* plants, the mean, as found from the values in Table I, is

$$5{\cdot}006 + (3{\cdot}428)\,(5{\cdot}9037) - (1{\cdot}462)\,(7{\cdot}1299) - (0{\cdot}246)\,(10{\cdot}1036) = 12{\cdot}3345 \text{ cm.};$$

for *versicolor*, on the contrary, we have

$$5{\cdot}936 + (2{\cdot}770)\,(5{\cdot}9037) - (4{\cdot}260)\,(7{\cdot}1299) - (1{\cdot}326)\,(10{\cdot}1036) = -21{\cdot}4815 \text{ cm.}$$

The difference between the average values of the compound measurements being thus 33·816 cm.

The distinctness of the metrical characters of the two species may now be gauged by comparing this difference between the average values with its standard error. Using the values of Table III, with the coefficients of our compound, we have

$$\begin{aligned}
19{\cdot}1434 + (9{\cdot}0356)\,(5{\cdot}9037) - (9{\cdot}7634)\,(7{\cdot}1299) - (3{\cdot}2394)\,(10{\cdot}1036) &= -29{\cdot}8508,\\
9{\cdot}0356 + (11{\cdot}8658)\,(5{\cdot}9037) - (4{\cdot}6232)\,(7{\cdot}1299) - (2{\cdot}4746)\,(10{\cdot}1036) &= 21{\cdot}1224,\\
9{\cdot}7634 + (4{\cdot}6232)\,(5{\cdot}9037) - (12{\cdot}2978)\,(7{\cdot}1299) - (3{\cdot}8794)\,(10{\cdot}1036) &= -89{\cdot}8206,\\
3{\cdot}2394 + (2{\cdot}4746)\,(5{\cdot}9037) - (3{\cdot}8794)\,(7{\cdot}1299) - (2{\cdot}4604)\,(10{\cdot}1036) &= -34{\cdot}6699,
\end{aligned}$$

and finally,

$$-29{\cdot}8508 + (21{\cdot}1224)\,(5{\cdot}9037) + (89{\cdot}8206)\,(7{\cdot}1299) + (34{\cdot}6699)\,(10{\cdot}1036) = 1085{\cdot}5522.$$

The average variance of the two species in respect of the compound measurements may be estimated by dividing this value (1085·5522) by 95; the variance of the difference between two means of fifty plants each, by dividing again by 25. For single plants the variance is 11·4269, so that the mean difference, 33·816 cm., between a pair of plants of different species has a standard deviation of 4·781 cm. For means of fifty the same average difference has the standard error 0·6761 cm., or only about one-fiftieth of its value.

III. Interpretation

The ratio of the difference between the means of the chosen compound measurement to its standard error in individual plants is of interest also in relation to the probability of misclassification, if the specific nature were judged wholly from the measurements. For reasons to be discussed later we shall estimate the variance of a single plant by dividing 1085·5522 by 95, giving 11·4269 cm.^{2} for the variance, and 3·3804 cm. for the standard deviation. Supposing that a plant is misclassified, if its deviation in the right direction exceeds half the difference, 33·816 cm., between the species, the ratio to the standard as estimated is 5·0018.

The table of the normal distribution (*Statistical Methods*, Table II) shows that a ratio 4·89164 is exceeded five times in a million, and 5·32672 only once in two million trials. By logarithmic interpolation the frequency appropriate to a ratio 5·0018 is about 2·79 per million. If the variances of the two species are unequal, this frequency is somewhat overestimated by this method, since we ought to divide the specific difference in proportion to the two standard deviations, and for constant sum of variances the sum of the standard deviations is greatest when they are equal. We may, therefore, at once conclude that if the measurements are nearly normally distributed the probability of misclassification, using the compound movement only is less than three per million.

The same ratio is of interest from another aspect. If the chosen compound X is analysed in respect to its variation within and between species, the sum of squares between species must be $25D^2$. Numerically we have, therefore,

Table V. *Analysis of variance of the chosen compound* X, *between and within species*

	Degrees of freedom	Sum of squares
Between species	4	28588·05
Within species	95	1085·55
Total	99	29673·60

Of the total only 3·6583 per cent. is within species, and 96·3417 per cent. between species. The compound has been chosen to maximize the latter percentage. Since, in addition to the specific means, we have used three adjustable ratios, the variation within species must contain only 95 degrees of freedom.

In making up the variate X, we have multiplied the original values of λ by $-32·1018$ in order to give to the measurement sepal length the coefficient unity. Had we used the original values, the analysis of Table V would have appeared as:

Table VI. *Analysis of variance of the crude compound* X, *between and within species*

	Degrees of freedom	Sum of squares	
Between species	4	27·74160	$=25D^2$
Within species	95	1·05341	$= \quad D = S$
Total	99	28·79501	$D(1+25D)$

On multiplying equations (1) by λ_1, λ_2, λ_3 and λ_4 and adding, it appears that $S = \Sigma\lambda d = D$, the specific difference in the crude compound X. The proportion (3·6 per cent.) of the sum of squares within species could therefore have been found simply as $1/(1+25D)$.

IV. The analogy of partial regression

The analysis of Table VI suggests an analogy of some interest. If to each plant were assigned a value of a variate y, the same for all members of each species, the analysis of variance of y, between the portions accountable by linear regression on the measurements $x_1, \ldots, x_4$, and the residual variation after fitting such a regression, would be identical with Table VI, if y were given appropriate equal and opposite values for the two species.

In general, with different numbers of representatives of the two species, n_1 and n_2, if the values of y assigned were

$$\frac{n_2}{n_1+n_2} \text{ and } \frac{-n_1}{n_1+n_2},$$

differing by unity, the right-hand sides of the equations for the regression coefficients, corresponding to equation (1), would have been

$$\frac{n_1 n_2}{n_1+n_2} d_p,$$

where d_p is the difference between the means of the two species in any one of the measurements. The typical coefficient of the left-hand side would be

$$S_{pq}+\frac{n_1 n_2}{n_1+n_2} d_p d_q.$$

Transferring the additional fractions to the right-hand side, we should have equations identical with (1), save that the right-hand sides are now

$$\frac{n_1 n_2}{n_1+n_2} d_p (1-\Sigma\lambda' d),$$

where λ' stands for a solution of the new equations; hence

$$\lambda'=\frac{n_1 n_2}{n_1+n_2} (1-\Sigma\lambda' d)\,\lambda,$$

multiply these equations by d and add, so that

$$\Sigma\lambda' d=\frac{n_1 n_2}{n_1+n_2} \Sigma\lambda d\,(1-\Sigma\lambda' d),$$

or
$$(1-\Sigma\lambda' d)\left(1+\frac{n_1 n_2}{n_1+n_2} \Sigma\lambda d\right)=1,$$

and so in our example
$$1-\Sigma\lambda' d=\frac{1}{1+25D}.$$

The analysis of variance of y is, therefore,

Table VII. *Analysis of variance of a variate y determined exclusively by the species*

	Degrees of freedom	Sum of squares	
Regression	4	24·0854	$25^2D/1+25D$
Remainder	95	0·9146	$25/1+25D$
Total	99	25·0000	

The total $S\ (y^2)$ is clearly in general $\frac{n_1 n_2}{n_1+n_2}$; the portion ascribable to regression is

$$\frac{n_1 n_2}{n_1+n_2}\,\Sigma\lambda' d=\frac{25^2 D}{1+25D}.$$

In this method of presentation the appropriate allocation of the degrees of freedom is evident.

The multiple correlation of y with the measurements $x_1, \ldots, x_4$ is given by

$$R^2=25D/1+25D.$$

V. Test of significance

It is now clear in what manner the specific difference may be tested fcr significance, so as to allow for the fact that a variate has been chosen so as to maximise the distinctness of the species. The regression of y on the four measurements is given 4 degrees of freedom, and the residual variation 95; the value of z calculated from the sums of squares in any one of Tables V, VI or VII is 3·2183 or

$$\tfrac{1}{2}\,(\log 95-\log 4+\log 25+\log D),$$

a very significant value for the number of degrees of freedom used.

VI. Applications to the theory of allopolyploidy

We may now consider one of the extensions of this procedure which are available when samples have been taken from more than two populations. The sample of the third species given in Table I, *Iris virginica*, differs from the two other samples in not being taken from the same natural colony as they were—a circumstance which might considerably disturb both the mean values and their variabilities. It is of interest in association with *I. setosa* and *I. versicolor* in that Randoph (1934) has ascertained and Anderson has confirmed that, whereas *I. setosa* is a "diploid" species with 38 chromosomes, *I. virginica* is "tetraploid", with 70, and *I. versicolor*, which is intermediate in three measurements, though not in sepal breadth, is hexaploid. He has suggested the interesting possibility that *I. versicolor* is a polyploid hybrid of the two other species. We shall, therefore, consider whether, when we use the linear compound of the four measurements most appropriate for discriminating three such species, the mean value for *I. versicolor* takes an intermediate value, and, if so, whether it differs twice as much from *I. setosa* as from *I. virginica*, as might be expected, if the effects of genes are simply additive, in a hybrid between a diploid and a tetraploid species.

If a third value lies two-thirds of the way from one value to another, the three deviations from their common mean must be in the ratio 4 : 1 : − 5. To obtain values corresponding with the differences between the two species we may, therefore, form linear compounds of their mean measurements, using these numerical coefficients. The results are shown in Table VIII where, for example, the value 7·258 cm. for sepal length is four times the mean sepal length for *I. virginica* plus once the mean sepal length for *I. versicolor* minus five times the value for *I. setosa*.

Table VIII

Means	S_{pq}			
Iris virginica. Fifty plants				
6·588	19·8128	4·5944	14·8612	2·4056
2·974	4·5944	5·0962	3·4976	2·3338
5·552	14·8612	3·4976	14·9248	2·3924
2·026	2·4056	2·3338	2·3924	3·6962
Iris versicolor. Fifty plants				
5·936	13·0552	4·1740	8·9620	2·7332
2·770	4·1740	4·8250	4·0500	2·0190
4·260	8·9620	4·0500	10·8200	3·5820
1·326	2·7332	2·0190	3·5820	1·9162
Iris setosa. Fifty plants				
5·006	6·0882	4·8616	0·8014	0·5062
3·428	4·8616	7·0408	0·5732	0·4556
1·462	0·8014	0·5732	1·4778	0·2974
0·246	0·5062	0·4556	0·2974	0·5442
$4vi+ve-5se$				
7·258	482·2650	199·2244	266·7762	53·8778
−2·474	199·2244	262·3842	74·3416	50·7498
19·158	266·7762	74·3416	286·5618	49·2954
8·200	53·8778	50·7498	49·2954	74·6604

Since the values for the sums of squares and products of deviations from the means within each of the three species are somewhat different, we may make an appropriate matrix corresponding with our chosen linear compound by multiplying the values for *I. virginica* by 16, those for *I. versicolor* by one and those for *I. setosa* by 25, and adding the values for the three species, as shown in Table VIII. The values so obtained will correspond with the matrix of sums of squares and products within species when only two populations have been sampled.

Using the rows of the matrix as the coefficients of four unknowns in an equation with our chosen compound of the mean measurements, e.g.

$$482{\cdot}2650\lambda_1+199{\cdot}2244\lambda_2+266{\cdot}7762\lambda_3+53{\cdot}8778\lambda_4=7{\cdot}258,$$

we find solutions which, when multiplied by 100, are

Coefficient of sepal length	−3·308998
sepal breadth	−2·759132
petal length	8·866048
petal breadth	9·392551

defining the compound measurement required.

It is now easy to find the means and variances of this compound measurement in the three species. These are shown in the table below (Table IX):

Table IX

	Mean	Sum of squares	Mean square	Standard deviation
I. virginica	38·24827	923·7958	18·8530	4·342
I. versicolor	22·93888	873·5119	17·8268	4·222
I. setosa	−10·75042	292·8958	5·9775	2·444

From this table it can be seen that, whereas the difference between *I. setosa* and *I. versicolor*, 33·69 of our units, is so great compared with the standard deviations that no appreciable overlapping of values can occur, the difference between *I. virginica* and *I. versicolor*, 15·31 units, is less than four times the standard deviation of each species.

The differences do seem, however, to be remarkably closely in the ratio 2 : 1. Compared with this standard, *I. virginica* would appear to have exerted a slightly preponderant influence. The departure from expectation is, however, small, and we have the material for making at least an approximate test of significance.

If the differences between the means were exactly in the ratio 2 : 1, then the linear function formed by adding the means with coefficients in the ratio 2 :−3 : 1 would be zero. Actually it has the value 3·07052. The sampling variance of this compound is found by multiplying the variances of the three species by 4, 9 and 1, adding them together and dividing by 50, since each mean is based on fifty plants. This gives 4·8365 for the variance and 2·199 for the standard error. Thus on this test the discrepancy, 3·071, is certainly not significant, though it somewhat exceeds its standard error.

In theory the test of significance is not wholly exact, since in estimating the sampling variance of each species we have divided the sum of squares of deviations from the mean by 49, as though these deviations had in all 147 degrees of freedom. Actually three degrees of freedom have been absorbed in adjusting the coefficients of the linear compound so as to discriminate the species as distinctly as possible. Had we divided by 48 instead of by 49 the standard error would have been raised by a trifle to the value 2·231, which would not have affected the interpretation of the data. This change, however, would certainly have been an over-correction, since it is the variances of the extreme species *I. virginica* and *I. setosa* which are most reduced in the choice of the compound measurement, while that of *I. versicolor* contributes the greater part of the sampling error in the test of significance.

The diagram, Fig. 1, shows the actual distributions of the compound measurement adopted in the individuals of the three species measured. It will be noticed, as was anticipated above, that there is some overlap of the distributions of *I. virginica* and *I. versicolor*, so that a certain diagnosis of these two species could not be based solely on these four measurements of a single flower taken on a plant growing wild. It is not, however, impossible that in culture the measurements alone should afford a more complete discrimination.

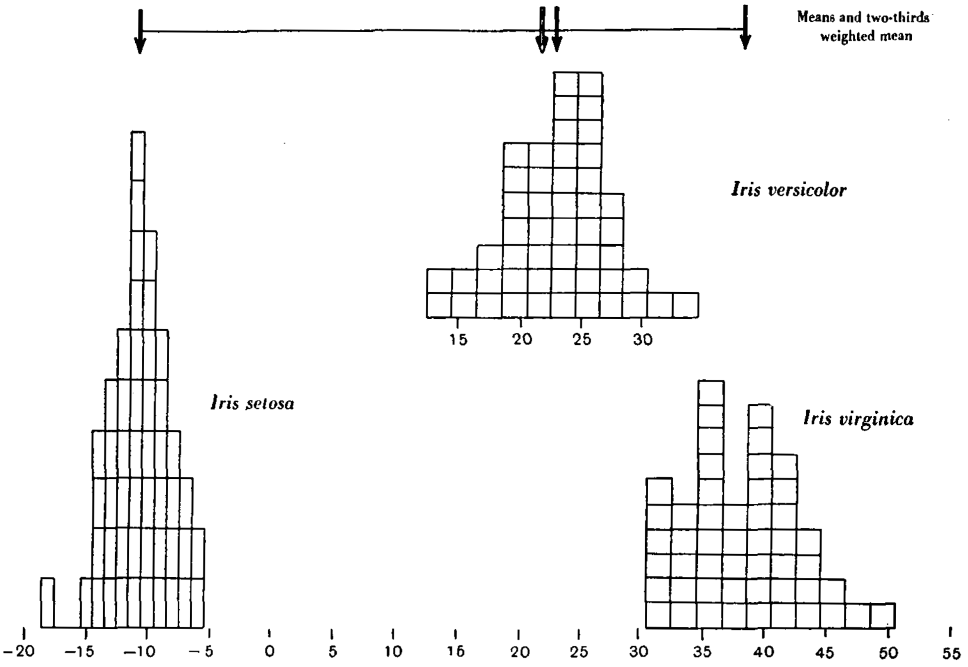

Fig. 1. Frequency histograms of the discriminating linear function, for three species of *Iris*.

REFERENCES

RANDOLPH, L. F. (1934). "Chromosome numbers in native American and introduced species and cultivated varieties of Iris." *Bull. Amer. Iris Soc.* **52**, 61–66.

ANDERSON, EDGAR (1935). "The irises of the Gaspe Peninsula." *Bull. Amer. Iris Soc.* **59**, 2–5.

—— (1936). "The species problem in *Iris*." *Ann. Mo. bot. Gdn.* (in the Press).