CLASSIFICATION OF IPNOPIDAE

BY MEANS OF PRINCIPAL COMPONENTS AND DISCRIMINANT FUNCTIONS

By KNUD P. ANDERSEN

Danmarks Fiskeri- og Havundersøgelser, Charlottenlund Slot

As HOTELLING's technique of principal components and FISCHER's (RAO 1952) of discriminatory analysis are little known and used by zoologists the author feels it would be a good idea to use these techniques on the data discussed by JØRGEN NIELSEN (1966). The author (and many others) thinks of it as a very powerful tool for supporting an introduction of new species as well as other classification problems that often arise in zoology. As the mathematical and statistical theory is well known, the author's aim is to show how the calculations are performed and what they mean.

The calculations needed can be performed by means of a desk calculator, and it is the author's hope that the reader will go through some of the tables using a desk calculator.

Principal components

This method is a mathematical tool for splitting up data consisting of several components (e.g. meristic characters) into groups. The aim is to find combinations of the components which give rise to a clustering of the data. The clustering is demonstrated by plotting the found combinations against each other in a co-ordinate system. The method uses all data as a whole and is in this way independent of classifications done by other means.

The data used is shown in Table 1. In the original data some values of the characters are missing. These gaps have been filled by means of a random number technique.

The original distributions of the 5 characters where values are missing are given in Table 2, and if we take Dorsal as an example three values are missing. The column acc. n is the accumulated frequences and we want to produce three D's by choosing randomly between 8, 9, 10, and 11 in accordance with the given frequency distribution. From Table XIX Random sampling numbers in A. HALD: Statistical Tables and Formulas I have taken the first column on page 97 where the first three figures are 05, 78, 24. Division by 44 (actual numbers of characters) gives the following residues 05, 34, 24. The accumulated series now provides us with the following D's: 8, 10, and 9, and these are the figures in Table 1 in brackets. The procedure has been used for all missing values.

This procedure for the filling up of gaps has the advantage that it uses the material as a whole and does not use the splitting up into species already done. As a matter of fact it diffuses the species a bit, but this is weighted up by being able to use all fish in the calculations. In Table 3 the means, variances and the correlation matrix L for all characters (augmented values included) are given. We can think of the data as points distributed in a six dimensional space with coordinates (l, D, A, G, V, S). If the data are distributed as a six dimensional normal distribution the points should fill up a hyper-elipsoid and by finding the main axes of this elipsoid and ordering them after magnitude, we should get combinations of the characters which give all the spread of the points. The greatest axes would be responsible for most of the spread, number two will be responsible for most of the spread left, etc. If now the data include different species one would expect this to give rise to a spread of the data and in this way the main axes of the hyper-elipsoid should give a combined character which groups the different species in well-defined clusters.

To do this properly one should use dimensionfree characters with unit standard deviations. This is done by dividing each character with its standard deviation, and these new characters are given in Table 3.

Table 1.

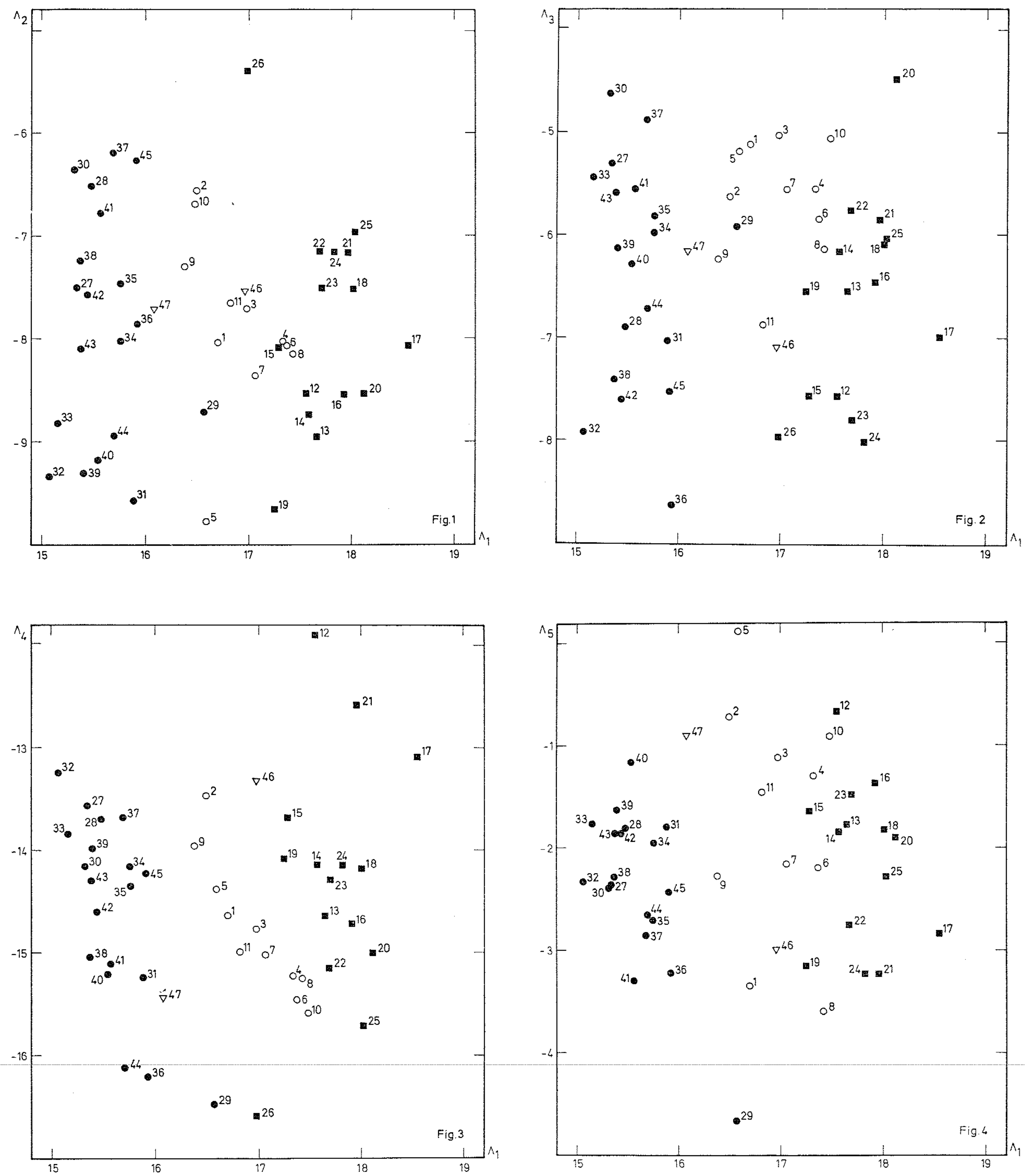
		Standard length	Dorsal fin rays	Anal fin rays	Gill rakers	Vertebrae	Lateral line scales
Ipnops murrayi ...	1	106	10	14	20	54	54
○	2	116	9	14	21	55	(51)
	3	107	10	14	22	55	53
	4	106	10	14	22	57	55
	5	60	10	14	21.5	55	53
	6	109	10	14	21.5	57	56
	7	100	10	14	21	56	55
	8	113	10	15	21	56	57
	9	107	9	14	20	54	53
	10	131	10	13	22.5	58	54
	11	103	9	14	21.5	55	55
Ipnops agassizi ...	12	90	9	19	22	58	55
■	13	94	10	15	21	60	57
	14	97	10	16	21.5	58	56
	15	100	9	16	21	58	56
	16	103	10	16	22.5	59	57
	17	123	10	19	22	61	58
	18	124	10	16	22	60	56
	19	81	10	16	20	57	57
	20	112	11	16	23	58	56
	21	134	10	18	21	59	55
	22	131	10	15	22	57	56
	23	115	9	16	22	59	57
	24	129	9	17	21.5	58	58
	25	137	10	15	23	58	57
	26	135	(8)	(11)	23	(58)	(55)
Ipnops meadi	27	92	9	11	17	54	50
●	28	105	8	12	18.5	53	51
	29	98	(10)	12	19	54	(57)
	30	113	9	11	18.5	51	(49)
	31	59	9	12	19	54	(55)
	32	50	8	(13)	17	52	(53)
	33	60	9	11	(18)	53	(50)
	34	86	9	12	18.5	54	52
	35	99	9	12	18.5	53	52
	36	92	8	12	19.5	52	(57)
	37	121	9	13	18	53	49
	38	93	8	11	18.5	52	53
	39	57	9	12	18	53	(52)
	40	60	9	11	19	53	(53)
	41	113	9	11	18.5	52	52
	42	86	8	11	18	54	53
	43	80	9	11	18	53	(51)
	44	72	9	11	19	52	(55)
	45	117	8	12	18.5	55	53
Ipnops specimens .	46	111	9	16	20	56	55
▼	47	92	(9)	11	20	55	(53)

The problem of finding the main axes is an eigenvalue problem and this means in matrix notation that we have to solve the equation

$$L V = k V$$

where L is the correlation matrix of the data, V an unknown vector and k an unknown constant.

The V's are the wanted main axes and k / 6 (number of characters) is the proportion of spread this axes is responsible for.



Figs. 1-4. Principal components plotted. For symbols and numbers see Table 1.

HOTELLING (1933 and 1936) has given a method for solving this problem and it gives the axes in the right order.

The process is illustrated in Table 4 for the first main axes.

We first take an arbitrary vector ($a_1\ a_2\ a_3\ a_4\ a_5\ a_6$) and calculate the new vector

$$(\sum_i a_i r_{1i},\ \sum a_i r_{2i},\ \sum a_i r_{3i},\ \sum a_i r_{4i},\ \sum a_i r_{5i},\ \sum a_i r_{6i})$$

where the r's are the elements of L.

Then we find the greatest of the product sums, say $\Sigma\, a_i r_{2i}$ and calculate

$$(b_1, b_2, b_3, b_4, b_5, b_6) = \left(\frac{\Sigma\, a_i r_{1i}}{\Sigma\, a_i r_{2i}},\ 1\ ,\ \frac{\Sigma\, a_i r_{3i}}{\Sigma\, a_i r_{2i}},\ \ldots.\right)$$

Table 2.

Dorsal fin rays (D)

	n	acc. n
11	1	44
10	17	43
9	20	26
8	6	6
Total	44	

Anal fin rays (A)

	n	acc. n
19	2	45
18	1	43
17	1	42
16	8	41
15	4	33
14	9	29
13	2	20
12	8	18
11	10	10
Total	45	

Gill Rakers (G)

	n	acc. n
23.0	3	46
22.5	2	43
22.0	7	41
21.5	5	34
21.0	6	29
20.5	0	23
20.0	5	23
19.5	1	18
19.0	4	17
18.5	7	13
18.0	4	6
17.5	0	2
17.0	2	2
Total	46	

Vertebrae (V)

	n	acc. n
61	1	46
60	2	45
59	3	43
58	7	40
57	4	33
56	3	29
55	6	26
54	7	20
53	7	13
52	5	6
51	1	1
Total	46	

Lateral line scales (S)

	n	acc. n
58	2	34
57	6	32
56	6	26
55	6	20
54	2	14
53	6	12
52	3	6
51	1	3
50	1	2
49	1	1
Total	34	

Now the vector $(\Sigma\, b_i\, r_{1i},\ \Sigma\, b_i\, r_{2i},\ \ldots)$ is found, and if the greatest product sum is, say $\Sigma\, b_i\, r_{4i}$ the vector

$$(c_1, c_2 \ldots) = \left(\frac{\Sigma\, b_i\, r_{1i}}{\Sigma\, b_i\, r_{4i}}, \ldots\right)$$

is calculated.

This process is stopped when the figures do not change any more.

In Table 4 I have started with (1, 1, 1, 1, 1, 1) and in the end the calculations have been carried out with 6 decimals (6 decimals were used in L too!). In most cases one must use 6 decimals in the calculations, but in most tables in this paper I have given substantially fewer, as the many decimals have been used for checking purposes only. k is the greatest $\Sigma\, d_i\, b_{ji}$ in the last row of this kind. From Table 4 is thus found $k_1 = 3.8882$.

After k_1 has been found $SS = \sum_i f_i^2$ is calculated and in order k_1 / SS and $\sqrt{k_1 / SS}$. $V_1 = (a_{11}, a_{21}, \ldots)$ is finally determined as $V_1 =$

$$(\sqrt{k_1 / SS}\, f_1, \ldots)$$

and this is the wanted main axis.

After k_1 and V_1 have been found the covariance Matrix L is reduced for the spread caused by the combined characters that corresponds to the main axis. This is done in the following way: The matrix $B_1 = \{b_{ij}\} = \{a_{i1}\, a_{j1}\}$ is calculated and subtracted from L. The matrix equation

$$L_2\, V = (L - B_1)\, V = k\, V$$

is now solved, and the process is continued until all 6 k's and V's are found. As a control the matrix $R = L - \sum_1^6 B_n$ is calculated and should be very close to the 0 matrix. B_1, L_2 are shown in Table 5. In Table 6 the k's and V's are given together with R. The linear components of the standard characters which gives the main axes are the principal components. From C (the V's) the standard characters are easily found as functions of the principal components which are called $\Lambda_1, \Lambda_2, \ldots \Lambda_6$, and from C^{-1} (also given in Table 6) the Λ's are found as functions of the standard and actual characters. In Table 7 are given: 1) Standard characters as functions of the Λ's, 2) Principal components (Λ's) as functions of the standard characters, 3) The Λ's as functions of the actual characters.

The main properties of the principal components are that they are uncorrelated. The principal components for each single fish as found from the formulas in Table 7 are given in Table 8. Some controls of calculation are given in Table 9. The means of the Λ's are determined from Table 8 and from Tables 3 and 7. The variances are also calculated and should theoretically be 1. All combinations of Λ's plotted against each other are shown in Figs. 1-15. (Λ_1, Λ_4) gives a definite splitting up into three clusters whereas (Λ_1, Λ_2), (Λ_1, Λ_3), (Λ_1, Λ_5), (Λ_1, Λ_6) mainly give two groups

Table 3.

	Mean	Variance	Standard deviation
Standard length (l)	100.40	500.99	22.38
Dorsal fin rays (D)..........	9.28	.55	.74
Anal fin rays (A)	13.68	5.22	2.29
Gill rakers (G)..........	20.22	3.05	1.74
Vertebrae (V)..........	55.49	6.73	2.59
Lat. line scales (S)	54.19	5.73	2.39

$$L = \{ r_{ij} \} = \begin{Bmatrix} 1.000000 & .254519 & .399122 & .567466 & .519401 & .279540 \\ .254519 & 1.000000 & .539655 & .596012 & .559679 & .429345 \\ .399122 & .539655 & 1.000000 & .674491 & .800450 & .640398 \\ .567466 & .596012 & .674491 & 1.000000 & .816921 & .668526 \\ .519401 & .559679 & .800450 & .816921 & 1.000000 & .701502 \\ .279540 & .429345 & .640398 & .668526 & .701502 & 1.000000 \end{Bmatrix} \begin{matrix} l \\ D \\ A \\ G \\ V \\ S \end{matrix}$$

$$\begin{matrix} l & D & A & G & V & S \end{matrix}$$

Total number of specimens: 47

Standard characters

$l^* = .045\ l$
$D^* = 1.346\ D$
$A^* = .438\ A$
$G^* = .572\ G$
$V^* = .385\ V$
$S^* = .418\ S$

Table 4.

(a_i)	1	1	1	1	1	1
$(\Sigma\ a_i\ r_{ji})$	3.02	3.38	4.05	4.33	*4.40*	3.72
$(\Sigma\ a_i\ r_{ji}\ /\ 4.40)$	.686	.768	.920	.984	1.000	.845
$(\Sigma\ b_i\ r_{ji})$	2.56	2.95	3.61	3.83	*3.92*	3.31
$(\Sigma\ b_i\ r_{ji}\ /\ 3.92)$	.653	.753	.921	.977	1.000	.844
(Σ)	2.522	2.920	3.588	3.799	*3.888*	3.294
$(\Sigma\ /\ 3.888)$	.6487	.7510	.9228	.9771	1.000	.8472
(Σ)	2.5189	2.9197	3.5889	3.7984	*3.8884*	3.2967
$(\Sigma\ /\ 3.8884)$	.64780	.75087	.92298	.97685	1.00000	.84783
(Σ)	2.51802	2.91955	3.58884	3.79815	*3.88820*	3.29692
$(\Sigma\ /\ 3.88820)$	.647606	.750874	.923008	.976840	1.000000	.847930
(Σ)	2.51787	2.91956	3.58885	3.79812	*3.88819*	3.29698
$(\Sigma\ /\ 3.88819)$	.647569	.750879	.923013	.976835	1.000000	.847947
(Σ)	2.51783	2.91956	3.58885	3.79811	*3.88818*	3.29699
$(\Sigma\ /\ 3.88818)$	.647560	.750881	.923015	.976835	1.000000	.847952
$(\Sigma\ d_i\ r_{ji})$	2.51783	2.91956	3.58886	3.79811	*3.88818*	3.29700
$(\Sigma\ /\ 3.88818) = f_i)$	.647560	.750881	.923018	.976835	1.000000	.847955

$k_1 = 3.88818$ $\quad SS = 4.50835$ $\quad \frac{k_1}{SS} = .862440$ $\quad \sqrt{\frac{k_1}{SS}} = .928676$ $\quad V = \left(f_i \sqrt{\frac{k_1}{SS}} \right)$

$a_{11} = .601373$
$a_{21} = .697325$
$a_{31} = .857185$
$a_{41} = .907163$
$a_{51} = .928676$
$a_{61} = .787475$

Control: (Length of V)$^2 = \Sigma\ a_{i1}^2 = 3.88818$

Table 5.

$B_1 = \{a_{n1}\}\{a_{n1}\}^* = \{a_{i1}\,a_{j1}\} = \{b_{ij}\} =$

.361649	.419352	.515488	.545543	.558481	.473566
.419352	.486262	.597737	.632587	.647589	.549126
.515488	.597737	.734766	.777607	.796047	.675012
.545543	.632587	.777607	.822945	.842461	.714368
.558481	.647589	.796047	.842461	.862439	.731309
.473566	.549126	.675012	.714368	.731309	.620117

$L_2 =$ (= $L - B_1$)

.638351	–.164833	–.116366	.021923	–.039080	–.194026
–.164833	.513738	–.058182	–.036575	–.088010	–.119781
–.116366	–.058182	.265234	–.103216	.004403	–.034614
.021923	–.036575	–.103216	.177055	–.025540	–.045842
–.039080	–.088010	.004403	–.025540	.137561	–.029807
–.194026	–.119781	–.034614	–.045842	–.029807	.379883

Table 6.

	k_1	k_2	k_3	k_4	k_5	k_6	Σk
	3.8882	.7965	.5915	.3543	.2280	.1416	6.0001
% of spread	64.8	13.3	9.9	5.9	3.8	2.4	100.1

$C =$

V_1	V_2	V_3	V_4	V_5	V_6
.6014	.7720	.0958	–.0411	–.1774	.0072
.6973	–.3230	.6220	–.0852	–.1195	.0307
.8572	–.1334	–.1122	.4520	–.0963	–.1457
.9072	.0881	.0303	–.1889	.3122	–.1878
.9287	.0051	–.1045	.1152	.1710	.2900
.7875	–.2656	–.4135	–.3034	–.2151	.0003

$R =$

.000001	.000005	.000000	.000003	.000009	.000000
.000000	.000000	–.000002	–.000001	.000000	–.000001
.000004	.000002	.000004	.000003	.000005	.000005
–.000003	–.000002	–.000002	–.000002	–.000001	–.000001
.000001	.000000	.000002	.000001	.000003	.000002
.000003	.000020	.000015	.000016	.000011	.000012

$C^{-1} =$

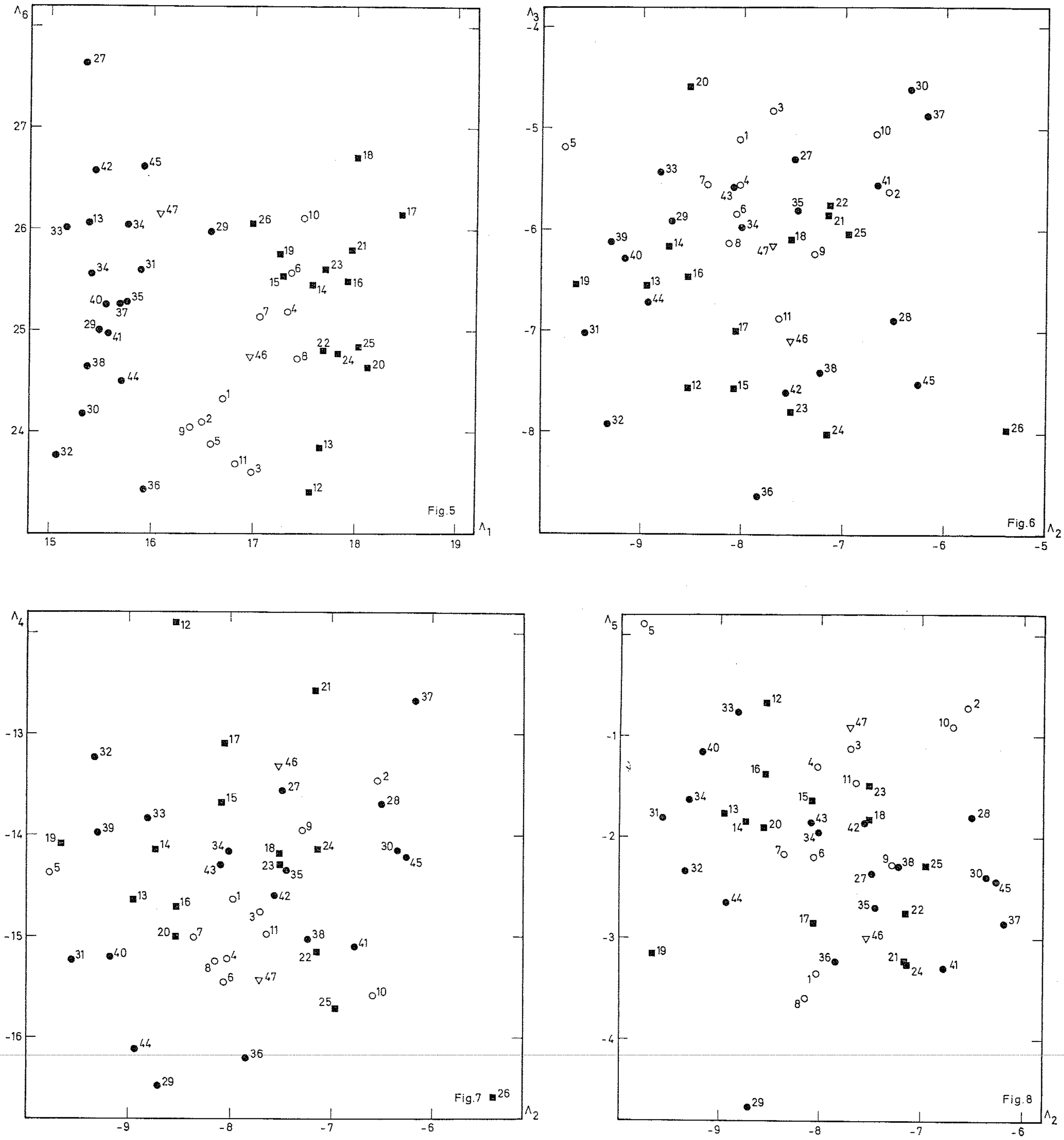
.1547	.1793	.2205	.2333	.2388	.2025	$= \frac{V_1}{k_1}$
.9693	–.4056	–.1676	.1106	.0064	–.3335	$\frac{V_2}{k_2}$
.1620	1.0516	–.1898	.0512	–.1767	–.6991	$\frac{V_3}{k_3}$
–.1161	–.2406	1.2759	–.5331	.3252	–.8565	$\frac{V_4}{k_4}$
–.7781	–.5242	–.4221	1.3690	.7499	–.9435	$\frac{V_5}{k_5}$
.0510	.2165	–1.0293	–1.3260	2.0482	.0020	$\frac{V_6}{k_6}$

(one consisting of *I. meadi*). The rest of the figures do not give clear clusterings.

In this way the figures show that Λ_1 gives a clustering that is in agreement with the classification into three species done by NIELSEN (1966).

Discriminant functions

The idea behind this method is to find combinations of the components which maximize the difference between already known groups.



Figs. 5-8. Principal components plotted. For symbols and numbers see Table 1.

To determine the discriminant functions we start with determining the means and mean variance for each Λ for each of the three species determined by NIELSEN. These values are given in Table 10. For three species we get three discriminant functions, viz. three linear combinations of the principal components that give the best distinction between any two of the three species. To be able to find these functions properly we have to suppose that the variances and correlations between the principal components are the same for all three species or in other words: The covariance matrixes for the three species are supposed to be identical.

In our case we will suppose that all variances are the variances given in Table 10 and that all correlations are zero. An inspection of Figs. 1-15 shows

Table 7.

1* =	$.6014\,\Lambda_1$	$+\ .7720\,\Lambda_2$	$+\ .0958\,\Lambda_3$	$-\ .0413\,\Lambda_4$	$-\ .1774\,\Lambda_5$	$+\ .0072\,\Lambda_6$
D* =	$.6973\,\Lambda_1$	$-\ .3230\,\Lambda_2$	$+\ .6220\,\Lambda_3$	$-\ .0852\,\Lambda_4$	$-\ .1195\,\Lambda_5$	$+\ .0307\,\Lambda_6$
A* =	$.8572\,\Lambda_1$	$-\ .1335\,\Lambda_2$	$-\ .1122\,\Lambda_3$	$+\ .4520\,\Lambda_4$	$-\ .0963\,\Lambda_5$	$-\ .1457\,\Lambda_6$
G* =	$.9072\,\Lambda_1$	$+\ .0881\,\Lambda_2$	$+\ .0303\,\Lambda_3$	$-\ .1889\,\Lambda_4$	$+\ .3122\,\Lambda_5$	$-\ .1878\,\Lambda_6$
V* =	$.9287\,\Lambda_1$	$+\ .0051\,\Lambda_2$	$-\ .1045\,\Lambda_3$	$+\ .1152\,\Lambda_4$	$+\ .1710\,\Lambda_5$	$+\ .2900\,\Lambda_6$
S* =	$.7875\,\Lambda_1$	$-\ .2656\,\Lambda_2$	$-\ .4135\,\Lambda_3$	$-\ .3034\,\Lambda_4$	$-\ .2151\,\Lambda_5$	$+\ .0003\,\Lambda_6$
Λ_1 =	.1547 1*	+ .1793 D*	+ .2205 A*	+ .2333 G*	+ .2388 V*	+ .2025 S*
Λ_2 =	.9693 1*	− .4056 D*	− .1676 A*	+ .1106 G*	+ .0064 V*	− .3335 S*
Λ_3 =	.1620 1*	+ 1.0516 D*	− .1898 A*	+ .0512 G*	− .1767 V*	− .6991 S*
Λ_4 =	−.1161 1*	− .2406 D*	+ 1.2759 A*	− .5331 G*	+ .3252 V*	− .8565 S*
Λ_5 =	−.7781 1*	− .5242 D*	− .4221 A*	+ 1.3690 G*	+ .7499 V*	− .9435 S*
Λ_6 =	.0510 1*	+ .2165 D*	− 1.0293 A*	− 1.3260 G*	+ 2.0482 V*	+ .0020 S*
Λ_1 =	.0069 1	+ .2413 D	+ .0965 A	+ .1335 G	+ .0920 V	+ .0846 S
Λ_2 =	.0433 1	− .5458 D	− .0733 A	+ .0633 G	+ .0025 V	− .1393 S
Λ_3 =	.0072 1	+ 1.4151 D	− .0830 A	+ .0293 G	− .0681 V	− .2921 S
Λ_4 =	−.0052 1	− .3238 D	+ .5583 A	− .3051 G	+ .1253 V	− .3578 S
Λ_5 =	−.0348 1	− .7054 D	− .1847 A	+ .7836 G	+ .2890 V	− .3942 S
Λ_6 =	.0023 1	+ .2914 D	− .4504 A	− .7590 G	+ .7893 V	+ .0008 S

that even if this is not true it is probably a good approximation to the truth. We start again by standardising our principal components by dividing by the standard deviations. Each of the new variables can be used as a discriminant function, and every one contributes somewhat to the discrimination. What we want is the following combinations: 1) X the combination that discriminates best between *I. murrayi* and *I. agassizi,* 2) Y the combination that discriminates best between *I. agassizi* and *I. meadi,* and 3) Z the combination that discriminates best between *I. murrayi* and *I. meadi.* R.A. FISHER has shown that such linear combination exists under the mentioned assumptions. In our case we have for X

$$X = (\text{Mean } \lambda_1 \text{ (I. mur.)} - \text{Mean } \lambda_1 \text{ (I. agas.)})\,\lambda_1 + \ldots + (\text{Mean } \lambda_6 \text{ (I. mur.)} - \text{Mean } \lambda_6 \text{ (I. agas.)})\,\lambda_6$$

and similar for Y and Z. This means by the way that Z = X + Y.

The calculations are shown in Table 11.

Table 12 gives the values of X and Y.

A test of the reliability of the grouping can now be performed in the following way. The variances and correlations of X, Y and Z = X + Y can be found the formulas

$$V(X) = \text{Mean } X_I - \text{Mean } X_{II}$$
$$V(Y) = \text{Mean } Y_{II} - \text{Mean } Y_{III}$$
$$V(Z) = V(X + Y) = \text{Mean } Z_I - \text{Mean } Z_{II} = V(X) + V(Y) + 2 \text{ cov. } (X,Y)$$
$$\text{cov. } (X,Y) = Y_I - Y_{II} = X_{II} - X_{III}$$
$$\text{cov. } (X,Z) = \text{cov. } (X, X + Y) = Z_I - Z_{II} = X_I - X_{III} = V(X) + \text{cov. } (X,Y)$$
$$\text{cov. } (Y,Z) = \text{cov. } (Y, X + Z) - Z_{II} - Z_{III} = Y_I - Y_{III} = V(Y) + \text{cov. } (X,Y)$$

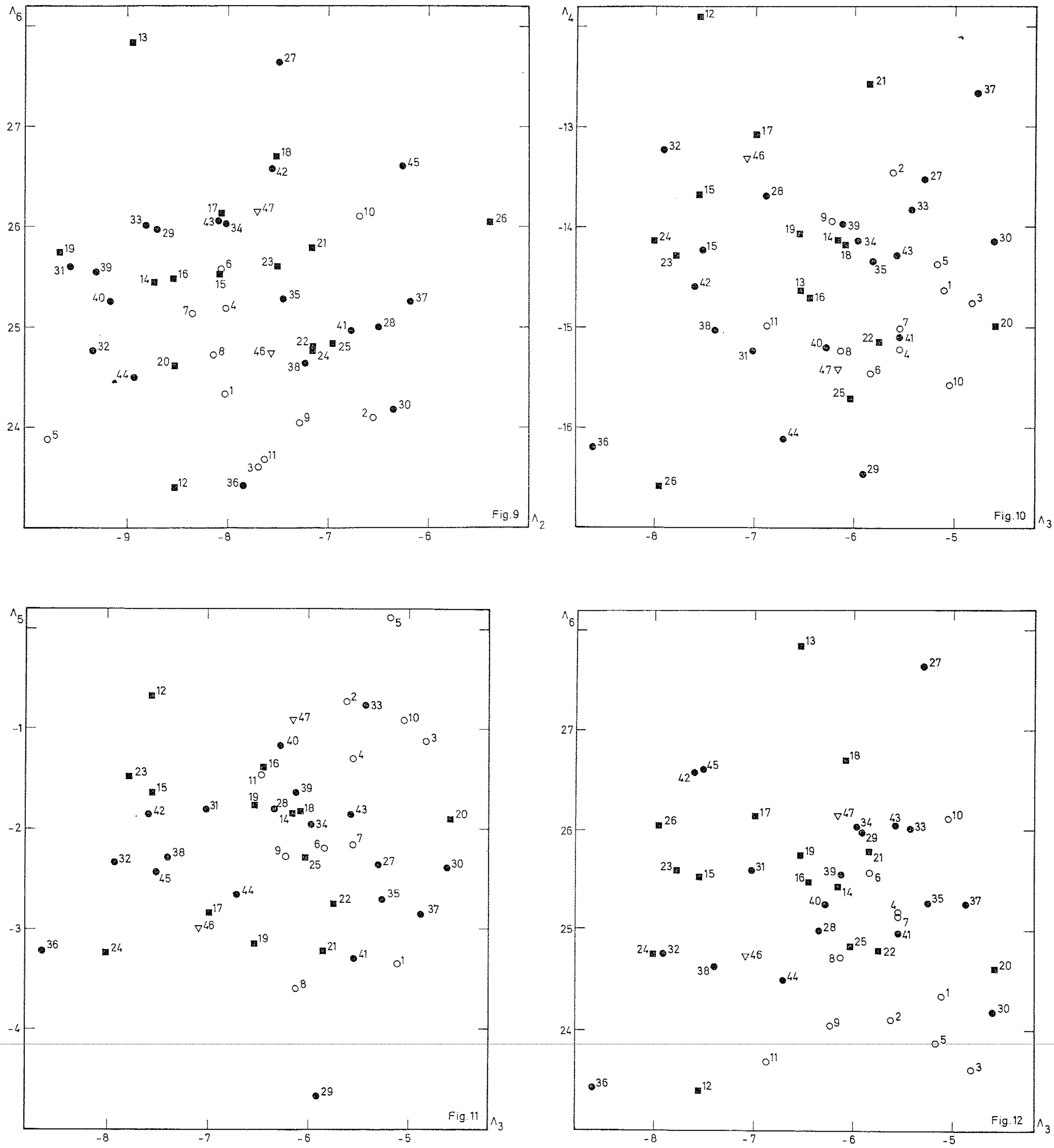
The calculations are shown in Table 13. The means are determined from the Mean Λ's of Table 10 and the Λ formulas of Table 11.

The variances and correlations can also be determined from Table 12, and even if the estimates are not independent of the values in Table 13, they give a good support to the reliability of the splitting up into species. Variances and correlations for X and Y determined from Table 12 are given in Table 14. The mean X's and Y's give a splitting up of the (X,Y) plane into three areas

$$X \geqq \frac{\text{Mean } X_I + \text{Mean } X_{II}}{2} \qquad X + Y \geqq \frac{\text{Mean } Z_I + Z_{III}}{2}$$

$$X \leqq \frac{\text{Mean } X_I + \text{Mean } X_{II}}{2} \qquad Y \geqq \frac{\text{Mean } Y_{II} + \text{Mean } Y_{III}}{2}$$

$$Y \leqq \frac{\text{Mean } Y_I + \text{Mean } Y_{III}}{2} \qquad X + Y \leqq \frac{\text{Mean } Z_I + \text{Mean } Z_{III}}{2}$$



Figs. 9-12. Principal components plotted. For symbols and numbers see Table 1.

This division of the (X,Y) plane is in a certain sense the best division of the plane in accordance with the three species. The divided (X,Y) plane is shown in Fig. 16 together with all X and Y values and the mean points for the three species. The figure shows that the separation is well effectuated by the discriminatioant functions X and Y. The equations of the division lines are given in the bottom of Table 14.

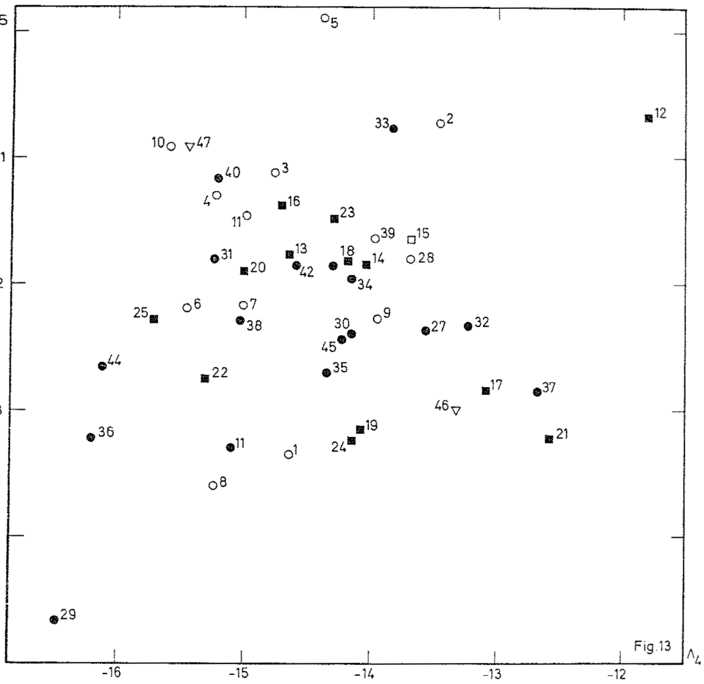
A measure of the efficiency of the separation can be found by finding the errors of misclassification, and even if this can be done for the three species under one, this will take us too far and we shall content ourselves by determining the pair misclassification probabilities.

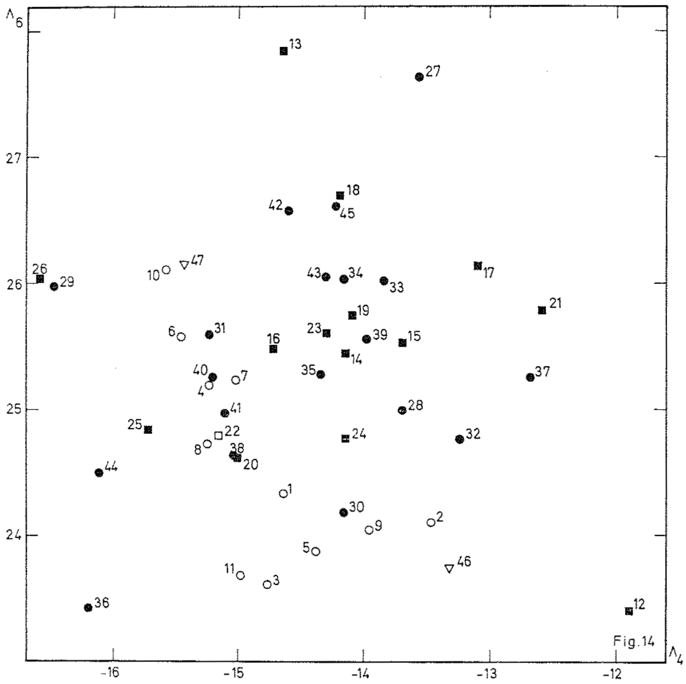
Table 8.

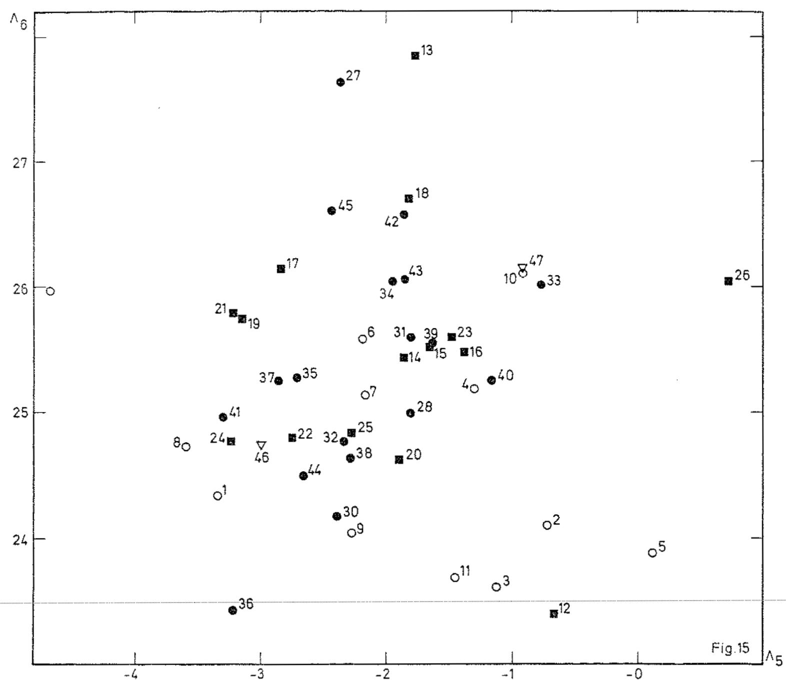
	A_1	A_2	A_3	A_4	A_5	A_6
1	16.70	-8.03	-5.11	-14.63	-3.34	24.34
2	16.50	-6.55	-5.62	-13.46	- .72	24.10
3	16.98	-7.70	-4.82	-14.76	-1.12	23.61
4	17.33	-8.02	-5.55	-15.22	-1.30	25.19
5	16.59	-9.77	-5.18	-14.37	+.12	23.88
6	17.37	-8.06	-5.84	-15.45	-2.19	25.58
7	17.06	-8.35	-5.55	-15.01	-2.16	25.14
8	17.42	-8.14	-6.13	-15.24	-3.59	24.73
9	16.38	-7.29	-6.23	-13.95	-2.27	24.05
10	17.48	-6.69	-5.05	-15.58	- .91	26.11
11	16.82	-7.64	-6.87	-14.98	-1.46	23.69
12	17.55	-8.53	-7.56	-11.90	- .67	23.40
13	17.65	-8.95	-6.54	-14.64	-1.77	27.84
14	17.57	-8.73	-6.16	-14.14	-1.85	25.44
15	17.28	-8.08	-7.56	-13.68	-1.64	25.53
16	17.92	-8.54	-6.45	-14.71	-1.38	25.48
17	18.55	-8.06	-6.99	-13.09	-2.84	26.14
18	18.01	-7.52	-6.08	-14.18	-1.82	26.70
19	17.25	-9.66	-6.54	-14.08	-3.15	25.75
20	18.12	-8.53	-4.59	-15.00	-1.90	24.62
21	17.96	-7.16	-5.85	-12.58	-3.22	25.79
22	17.68	-7.15	-5.75	-15.15	-2.75	24.80
23	17.70	-7.51	-7.79	-14.29	-1.48	25.60
24	17.82	-7.15	-8.01	-14.14	-3.23	24.77
25	18.03	-6.96	-6.03	-15.71	-2.28	24.84
26	16.98	-5.39	-7.96	-16.58	+.73	26.05
27	15.34	-7.49	-5.30	-13.56	-2.36	27.64
28	15.48	-6.50	-6.89	-13.69	-1.80	25.00
29	16.57	-8.70	-5.91	-16.47	-4.66	25.98
30	15.32	-6.35	-4.61	-14.15	-2.39	24.18
31	15.89	-9.56	-7.02	-15.23	-1.80	25.60
32	15.07	-9.33	-7.92	-13.23	-2.33	24.77
33	15.16	-8.81	-5.43	-13.83	- .76	26.02
34	15.76	-8.01	-5.97	-14.15	-1.95	26.04
35	15.76	-7.45	-5.81	-14.34	-2.70	25.28
36	15.93	-7.84	-8.63	-16.20	-3.22	23.43
37	15.69	-6.18	-4.87	-12.67	-2.85	25.26
38	15.37	-7.23	-7.40	-15.03	-2.28	24.64
39	15.40	-9.30	-6.12	-13.97	-1.63	25.56
40	15.54	-9.17	-6.28	-15.20	-1.16	25.26
41	15.57	-6.77	-5.55	-15.10	-3.29	24.97
42	15.44	-7.56	-7.60	-14.59	-1.85	26.58
43	15.38	-8.09	-5.58	-14.29	-1.85	26.06
44	15.70	-8.93	-6.71	-16.11	-2.65	24.50
45	15.91	-6.26	-7.52	-14.22	-2.43	26.61
46	16.96	-7.53	-7.09	-13.32	-2.99	24.74
47	16.08	-7.71	-6.16	-15.43	- .91	26.15

For each pair of species the corresponding discriminant function has a misclassification probability for each of the species that equals the probability that a normal distributed variate with zero mean and unit variance exceeds $^1/_2$ of the standard deviation of discriminant function. The misclassification probabilities are shown in Table 15.

It is possible using this method to start with the original components. The basic assumption would be that the covariance matrices of the three species for the original components were identical. The first step in the calculation would be from the best estimate of this common covariance matrix to find uncorrelated combinations of the original components. One would use the following combinations:







Figs. 13-15. Principal components plotted. For symbols and numbers see Table 1.

$$\nu_1 = 1$$
$$\nu_2 = D - a_{21}\,1$$
$$\nu_3 = A - a_{32}\,D - a_{31}\,1$$
$$\nu_4 = G - a_{43}\,A - a_{42}\,D - a_{41}\,1$$
$$\nu_5 = V - a_{54}\,G - a_{53}\,A - a_{52}\,D - a_{51}\,1$$
$$\nu_6 = S - a_{65}\,V - a_{64}\,G - a_{63}\,A - a_{62}\,D - a_{61}\,1$$

where the a's are determined by the conditions

$$\text{cov.}\ (\nu_i, \nu_j) = 0;\ (i \neq j).$$

In this case we have used the Λ's and the point is that we know that they are uncorrelated. Principally the two approaches should give the same answers, but as we have used two different approximations (1. The correlation matrices of original components are identical and 2. The correlation matrices of the Λ's are identical) slightly different answers will come out of the calculations, but the differences are very small.

Table 9.

	Means (Table 3 and 7)	Means (Table 8)	Variance
$\bar{\Lambda}_1$	16.641	16.639	1.006
$\bar{\Lambda}_2$	− 7.848	− 7.850	1.002
$\bar{\Lambda}_3$	− 6.301	− 6.302	1.005
$\bar{\Lambda}_4$	−14.495	−14.496	1.005
$\bar{\Lambda}_5$	− 2.043	− 2.044	1.005
$\bar{\Lambda}_6$	25.264	25.265	1.000

Table 10.

Mean	$\bar{\Lambda}_1$	$\bar{\Lambda}_2$	$\bar{\Lambda}_3$	$\bar{\Lambda}_4$	$\bar{\Lambda}_5$	$\bar{\Lambda}_6$
I. murr.	16.966	−7.840	−5.632	−14.786	−1.722	24.584
I. agas.	17.738	−7.861	−6.657	−14.258	−1.950	25.517
I. meadi	15.594	−7.870	−6.375	−14.528	−2.314	25.441
Mean variance	.1387	1.0948	.9204	1.0045	.9856	.9118
Stand. dev. δ	.3724	1.0463	.9594	1.0022	.9928	.9549

Table 11.

$$\lambda_1 = \Lambda_1 / \delta_1 = \Lambda_1 / .3724 = 2.6853\, \Lambda_1$$
$$\lambda_2 = \Lambda_2 / \delta_2 = \Lambda_2 / 1.0463 = .9557\, \Lambda_2$$
$$\lambda_3 = \Lambda_3 / \delta_3 = \Lambda_3 / .9594 = 1.0423\, \Lambda_3$$
$$\lambda_4 = \Lambda_4 / \delta_4 = \Lambda_4 / 1.0022 = .9978\, \Lambda_4$$
$$\lambda_5 = \Lambda_5 / \delta_5 = \Lambda_5 / .9928 = 1.0073\, \Lambda_5$$
$$\lambda_6 = \Lambda_6 / \delta_6 = \Lambda_6 / .9549 = 1.0472\, \Lambda_6$$

Mean	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	
I. murr.	45.5588	−7.4927	−5.8702	−14.7535	−1.7346	25.7444	I
I. agas.	47.6319	−7.5128	−6.9386	−14.2266	−1.9642	26.7214	II
I. meadi	41.8746	−7.5214	−6.6447	−14.4960	−2.3309	26.6418	III

	I-II	II-III	I-III
λ_1	−2.0731	5.7573	3.6842
λ_2	.0201	.0086	.0287
λ_3	1.0684	− .2939	.7745
λ_4	− .5269	.2694	− .2575
λ_5	.2296	.3667	.5963
λ_6	− .9770	.0796	− .8974

$$X = -2.0731\,\lambda_1 + .0201\,\lambda_2 + 1.0684\,\lambda_3 - .5269\,\lambda_4 + .2296\,\lambda_5 - .9770\,\lambda_6$$
$$= -5.5669\,\Lambda_1 + .0192\,\Lambda_2 + 1.1136\,\Lambda_3 - .5257\,\Lambda_4 + .2313\,\Lambda_5 - 1.0231\,\Lambda_6$$
$$= -.0372\,l - .0690\,D - .5065\,A + .4088\,G - 1.3945\,V - .7028\,S$$

$$Y = 5.7573\,\lambda_1 + .0086\,\lambda_2 - .2939\,\lambda_3 + .2694\,\lambda_4 + .3667\,\lambda_5 + .0796\,\lambda_6$$
$$= 15.4601\,\Lambda_1 + .0082\,\Lambda_2 - .3063\,\Lambda_3 + .2688\,\Lambda_4 + .3694\,\Lambda_5 + .0834\,\Lambda_6$$
$$= .0908\,l + 2.9693\,D + 1.5610\,A + 2.1996\,G + 1.6495\,V + 1.1545\,S$$

$$Z = X + Y = 3.6842\,\lambda_1 + .0287\,\lambda_2 + .7745\,\lambda_3 - .2575\,\lambda_4 + .5963\,\lambda_5 - .8974\,\lambda_6$$
$$= 9.8932\,\Lambda_1 + .0274\,\Lambda_2 + .8073\,\Lambda_3 - .2569\,\Lambda_4 + .6007\,\Lambda_5 - .9398\,\Lambda_6$$
$$= .0535\,l + 2.9003\,D + 1.0546\,A + 2.6085\,G + .2549\,V + .4517\,S$$

Fig. 16. The plane of the discriminant functions. For symbols and numbers see Table 1.

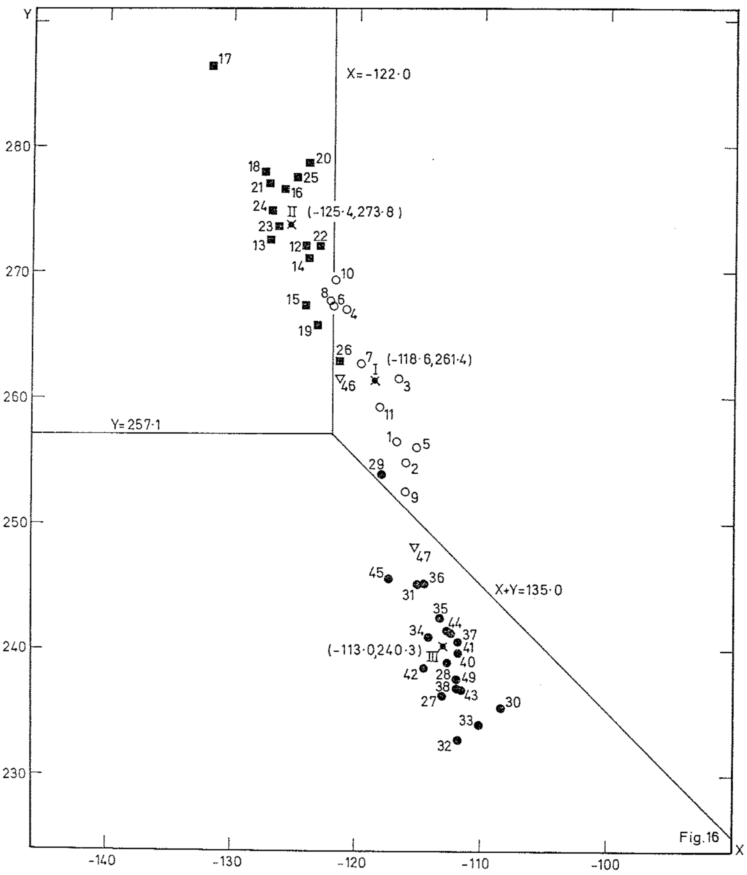


Table 12.

	X	Y		X	Y
1	–116.80	256.58	25	–124.92	277.62
2	–115.98	254.90	26	–121.28	262.94
3	–116.71	261.56			
4	–120.87	267.08	27	–113.11	236.44
5	–115.17	256.20	28	–112.72	239.02
6	–121.89	267.41	29	–118.01	254.00
7	–119.66	262.69	30	–108.39	235.54
8	–122.06	267.74	31	–115.08	245.18
9	–116.07	252.55	32	–111.81	232.94
10	–121.78	269.39	33	–110.11	234.08
11	–118.11	259.44	34	–114.18	243.06
			35	–113.27	242.60
12	–124.13	272.11	36	–114.65	245.31
13	–126.93	272.61	37	–112.69	241.59
14	–123.85	271.09	38	–111.78	237.02
15	–124.10	267.29	39	–111.92	237.68
16	–125.76	276.64	40	–111.81	239.75
17	–131.72	286.49	41	–111.89	240.66
18	–127.44	277.94	42	–114.52	238.59
19	–123.18	265.84	43	–111.56	237.05
20	–123.86	278.72	44	–112.27	241.50
21	–127.14	276.97	45	–117.37	245.71
22	–123.01	272.07			
23	–126.34	273.66	46	–121.42	261.64
24	–126.88	274.90	47	–115.38	248.15

Table 13.

Species	I	II	III
Mean	I. murrayi	I. agassizi	I. meadi
X	–118.648	–125.372	–112.987
Y	261.397	273.781	240.335
Z = X + Y	142.747	148.407	127.345

$V(X) = 6.724$ $\delta_X = 2.593$
$V(Y) = 33.446$ $\delta_Y = 5.783$
$V(Z) = 15.402$ $(= 15.402)$ $\delta_Z = 3.924$

cov. (X,Y) = –12.384 (= –12.385)
cov. (X,Z) = – 5.660 (= 5.661) (= 5.660)
cov. (Y,Z) = 21.062 (= 21.062) (= 21.062)

Correlation matrix

	X	Y	Z
X	—	–.826	–.556
Y	–.826	—	.928
Z	–.556	.928	—

Table 14.

I. murrayi

Mean X = –118.64	V(X) = 7.12	$r_{X,Y} = -.94$
Mean Y = 261.41	V(Y) = 34.88	

I. agassizi

Mean X = –125.37	V(X) = 6.33	$r_{X,Y} = -.81$
Mean Y = 273.79	V(Y) = 33.90	

I. meadi

Mean X = –113.00	V(X) = 5.23	$r_{X,Y} = -.83$
Mean Y = 240.41	V(Y) = 25.19	

Division lines

X = –122.010
Y = 257.058
X+Y = 135.046

(For a more detailed description of the discriminant fuctions see Rao, p. 273 foll. pp.).

We have now come to an end of our calculations,

Table 15.

Misclassification probabilities.

$$P\,(I.\ mur.,\ I.\ agas.) = P\left(u \geqq \frac{2.593}{2}\right) = 9.7\,\%$$

$$P\,(I.\ agas.,\ I.\ meadi) = P\left(u \geqq \frac{5.783}{2}\right) = .2\,\%$$

$$P\,(I.\ mur.,\ I.\ meadi) = P\left(u \geqq \frac{3.924}{2}\right) = 2.5\,\%$$

and it should be borne in mind that the calculations do not prove that the three species are real. The calculations only show that some consequences of the hypothesis that the 47 specimens belong to three different groups are in agreement with the data, and the calculations support in this way the hypothesis.

For one specimen (No. 26) only two characters are known and 4 are constructed. The figures show that this fish totter about in the diagrams, but that is of course only to be expected. Fish Nos. 46 and 47 are plotted on all figures in attempt to allocate them to the species. From Fig. 16 one would call No. 46 and *I. murrayi* and No. 47 for *I. meadi,* but as NIELSEN (1966) shows there are other characters (not used in this paper) that contradicts this classification.

REFERENCES

HALD, A., 1952: Statistical Tables and Formulas. – New York – John Wiley & Sons, Inc. London – Chapman & Hall, Ltd.

HOTELLING, H., 1933: Analysis of a complex of statistical variables into principal components. – J. Educ. Psych. 24: 417.

– 1936: Simplified calculation of principal components. – Psychometrika **1**: 27.

NIELSEN, J. G., 1966: Synopsis of the family Ipnopidae (Pisces, Iniomi) with description of two new abyssal species. Galathea Rep. **8**.

RAO, C. R., 1952: Advanced Methods in Biometric Research. – New York – John Wiley & Sons, Inc. London – Chapman & Hall, Ltd.

25. 10. 1966. 600. 125