# DISCRIMINATORY TECHNIQUES IN CLASSING OF FLEECES

By U.G. Nadkarni
*Institute of Agricultural Research Statistics, New Delhi*
(Received in November, 1965)

## 1. Introduction

In an investigation sponsored by Indian Council of Agricultural Research for evaluating suitable grades of wool of important sheep breeds in Rajasthan, a large and representative sample of fleeces from Magra tract was collected. These fleeces were first visually graded by the laboratory graders at Bikaner, into four classes. They were then analysed for quality characters such as fibre diameter, crimps per centimetre, staple length and percentage of meduallated fibres. The mean values of the different characters in these grades differed, defining each grade in terms of means and their standard errors. When a large number of fleeces are to be sensorily graded on the basis of the quality characters, it is necessary to have an objective procedure for discriminating the fleeces into grades. Such an objective procedure evolved on the basis of data on fleeces graded by expert graders may be useful in testing the ability of grading of a new grader.

The classifactory techniques for classifying given observations into one of two or more populations have been discussed in number of papers. To quote some of these techniques for two populations we have (*i*) method of discriminant function, Fisher (1936), (*ii*) method of likelihood ratio test, Anderson (1958), (*iii*) Wald's, Anderson's, Rao's and John's classificatory statistics, John (1960). For classification into more than two classes, the procedures suggested are (*i*) Fisher's general method of discrimination, Fisher (1937), (*ii*) canonical analysis, Bartlett (1947), (*iii*) method of minimum expected loss, Anderson (1952), (*iv*) Fairfield Smith's technique of discrimination, Fairfield Smith (1936). Where a single discriminant function is not adequate, one can use a discriminant function for every pair generated by dichotomising classes. In the present paper, the discriminant function for classifying the fleeces into grades are estimated

through canonical analysis, and it is shown by multivariate tests of linearity and direction of population means, that a single discriminant function is in general not adequate for classifying the fleeces into four grades, and discrimination through dichotomy is to be adopted.

## 2. SOURCE DATA AND ASSUMPTIONS

In the investigation referred to earlier, *2610* ewe fleeces, *1741* lamb fleeces of Magra breed were collected more or less equally from the three shearings, in autumn, spring and monsoon, through a multi-stage stratified sampling design, with ten geographically contiguous areas having about 30,000 sheep as strata and ten villages per strata, three flocks per village, three ewes and two lambs per flock as successive sampling units. Except for a sub-sample of fleeces kept apart for clean wool yield determination all were visually graded into fine, medium 1, medium 2 and coarse classes and were also analysed in the laboratory for the four characters, fibre diameter, crimps per centimetre, staple length and percentage of medullated fibres. Out of the data on these visually graded ewe fleeces, that on a representative sample of 399 ewe fleeces in autumn, 447 in spring, 294 in monsoon, equally spread over all circles in each season, have been used for fitting theoretical models and the rest kept for testing such models. The averages and their standard errors of all the four characters for the fleeces falling in each grade in every season were calculated. For brevity the data for autumn clip are discussed in greater detail in the following, and comments on the results of the analysis of the other two clips are made.

The means and standard errors of the four characters of fleeces of autumn clip were as given in Table 1.

TAELE 1

| | *Fine* | *Medium 1* | *Medium 2* | *Coarse* |
|---|---|---|---|---|
| Numbers of fleeces | 28 | 85 | 140 | 146 |
| Characters<br>Diameter ($x_1$)<br>(in microns) | 33·83<br>( ·82) | 35·39<br>( ·47) | 36·53<br>( ·37) | 39 02<br>( ·36) |
| Crimps per ($x_2$)<br>centimetre | 1·14<br>( ·05) | ·74<br>( ·03) | ·58<br>( ·0 ) | ·48<br>( ·02) |
| Percentage of ($x_3$)<br>medullated fibres | 47·75<br>( 3·83) | 56·11<br>( 2·20) | 59·98<br>( 1·71) | 66·46<br>( 1·68) |
| Staple length ($x_4$)<br>(in cms.) | 4·78<br>( ·31) | 5·42<br>( ·18) | 5·52<br>( ·14) | 5·80<br>( ·13) |

(Numbers in brackets indicate standard errors)

The means of the four laboratory characters differed significantly over the grades. A significant increasing trend from fine to coarse in the case of all characters, except crimps per centimetre which decreased with decrease in fineness, was observed. The same trends were also observed for the averages of characters of fleeces collected in spring and monsoon clips. In order to obtain discriminating models for classification following basic assumptions were made.

(*i*) The grades into which the sample of fleeces have been classified by the graders are taken as defining classes into which all the fleeces are to be classified, (*ii*) The data are assumed to conform to multivariate normal law, (*iii*) The population variance-covariance matrix for each of the four classes is assumed to be identical.

## 3. Selection of characters

In fitting a suitable linear combination of variables for discrimination the variables selected should be such as would determine a best discriminant function. The analysis of variance for each of the four characters showed that the ratios of between grade to within grade variations for diameter, ($x_1$) crimps per centimetre, ($x_2$) percentage of medullated fibres ($x_3$) and staple length ($x_4$) were 20·1, 60·8, 3·4, and 7·9 respectively. This shows that, taken separately, fibre diameter and crimps per centimetre are better discriminators than percentage of medullated fibres and staple length.

In order to test whether the characters, percentage of medullated fibres ($x_3$) and staple length ($x_4$) show significant variation in the four grades independently of the variation due to characters fibre diameter ($x_1$) and crimps per centimetre ($x_2$) the estimate of the statistic (6),

$$\Lambda=\frac{E(x_3x_4/x_1x_2)}{S(x_3x_4/x_1x_2)}$$

worked out for the autumn data was ·9967. (Appendix I) It was tested by considering

$$V=-m\log\Lambda=1{\cdot}3107 \text{ with } m=\left[\, n-\frac{(p+q+1)}{m}\,\right]=396$$

as a $x^2$ variate with $pq=6$ *d.f.*

The estimate is not significant at 5% level of significance showing that the inclusion of $x_3$ and $x_4$ would not increase the joint discriminating ability of $x_1$ and $x_2$. This held true in the data for spring and monsoon clips also. In the subsequent analysis only fibre diameter and crimps per centimetre have been used.

## 4. Discriminant function

The best discriminant functions are estimated from the equations $(B-r^2S)\,.\,a=0$ in coefficient vector $a$ for the largest root $r^2$, of the determinental equation in $r^2$ (5). These equations for the coefficients corresponding to $r^2=\cdot3679$ (with $r_2^2=\cdot0264$) are

$$2005{\cdot}67\ a_1+63{\cdot}36\ b_1=0$$
$$63{\cdot}36\ a_1+2\ 00\ b_1=0$$

which give $a_1/b_1=-\cdot0316$. The best discriminant function is, therefore, $x=x_2-\cdot0316\ x_1$.

Substituting the mean values for each grade into this expression, we obtain :

$$\bar{X}_C=-1{\cdot}0630,\ \bar{X}_{M_2}=-\cdot5743,\ \bar{X}_{M_1}=-\cdot3786,\ \bar{X}_F=\cdot0710$$

## 5. Conditions for linearity

The necessary condition that the above best discriminant function may be valid one for efficient classification, is that the means of the characters in the grades are collinear. In the exact test (5, 9) for linearity of means for a discriminant function, the residual likelihood criterion $\frac{V}{1-\phi}$ where $\Lambda=(1-r_1^2)(1-r_2^2)$ and $\phi=\frac{\text{Between } S.S.}{\text{Total } S.S.}$ for the proposed discriminant function is factorised into two factors one of which is $\left(1-\frac{r_1^2\,r_2^2}{\phi}\right)$ and corresponds to the component for deviation from linearity. The $x^2$ value for the component for deviation from linearity for the estimated discriminant function is

$$-\{(n-1)-\tfrac{1}{2}[(p-1)+(q-1)+1]\}\log\left[1-\frac{r_1^2\,r_2^2}{\phi}\right]$$
$$=-\{(396)(\cdot02676)\}=10{\cdot}99$$

which is highly significant at 5% level with $(p-1)\,(q-1)=2$ d.f. This shows that the population means are not collinear and a single discriminant function of diameter and crimps/centimetre is not suitable for discriminating between populations. This was found to be true in the case of data for spring and monsoon clips also.

## 6. Discrimination through dichotomy

The alternative procedure of repeated discrimination taking two populations at a time has therefore to be adopted for the data under study. For classifying the fleeces through dichotomy into four grades, *viz.*, fine, medium 1, medium 2 and coarse, three discriminant functions, *viz.*, one for discriminating fine and medium 1 fleeces (class

$G_1$) from medium 2 and coarse (class $G_2$) and the other two, for discriminating fine ($G_{11}$) from medium 1 ($G_{12}$) and medium 2 ($G_{21}$) from coarse ($G_{22}$) respectively will have to be fitted.

The procedure of classification consists in classifying a fleece using a discriminant function $G_1$ or $G_2$ and subsequently to use the discriminant function for $G_{11}$ and $G_{12}$ or $G_{21}$ and $G_{22}$ according as a fleece is put into $G_1$ or $G_2$. The three discriminant functions for autumn data were found.

These turned out to be

$$X_1=10^{-2}[\cdot 2320\ X_1-\cdot 7442\ x_2] \qquad \bar{X}_{1G_1}=10^{-2}\ (\cdot 1819)$$

$$\bar{X}_{1G_1}=10^{-2}\ (\cdot 4836)$$

$$X_2=10^{-2}[\cdot 0275\ x_1-2\cdot 5559\ x_2] \qquad \bar{X}_{2G_{11}}=10^{-2}\ (-1\cdot 9986)$$

$$\bar{X}_{2G_{12}}=10^{-2}\ (-\cdot 9196)$$

$$X_3=10^{-2}[\cdot 0412\ X_1-\cdot 7081\ x_2] \qquad \bar{X}_{3G_{21}}=10^{-2}\ (109\cdot 1200)$$

$$\bar{X}_{3G_{22}}=10^{-2}\ (127\ 08)$$

In all, 143 ewe fleeces were classified through repeated dichotomy using discriminant function. Frequency of correct classification by repeated dichotomy was 48%.

Though in the above all the three discriminant functions have been fitted, it will reduce the routine computation, if the discriminant function for $G_{11}$ and $G_{12}$ do not differ significantly from that corresponding to $G_{21}$ and $G_{22}$. If $d_1$, $d_2$ denote differences in mean values of diameter and crimps per centimetre for $G_{11}$, $G_{12}$ and $d_1$, $d_2$ are those for $G_{21}$ and $G_{22}$, then the hypotheses of equality of discriminant function corresponds to testing $E(di)=E(di')$. The corresponding variance ratio with $p=2$,

$$n'=n_1+n_2+n_3+n_4-5. \quad d.f.$$

where $n_i$ is number of observations in $i$th class is

$$\frac{n'\,f(n)}{p(n'-p-1)}\ \Sigma\ s^{ij}(d_i-d_i')(d_j-d')$$

where $$f(n)=\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}$$

In the case of autumn data, the value comes to be ·004297 and is not significant. It is concluded from the test that only two discriminant functions may be used for classification into four classes.

In general, if a single discriminant function be used at every stage of dichotomy, then for $n$ classes, such that $2^{r-1} < n < 2^r$ we need only $r$ discriminant functions.

## CONCLUSIONS

Analysis of the data on fleeces classified by trained graders showed that fibre diameter and crimps per centimetre conveyed most of the information for discrimination. The inclusion of percentage of medullated fibres and staple length did not add to the discriminating ability of the function. It was found that a single discriminant function corresponding to the maximum canonical root did not satisfy the condition of linearity in all the three clips. The alternative procedure of repeated dichotomy was therefore used for classing of fleeces.

## ACKNOWLEDGEMENTS

I am grateful to Dr. V.G. Panse, Statistical Adviser, I.C.A.R. for suggesting this study to me and to Shri V.N. Amble, Deputy Statistical Adviser for the constant encouragement in the preparation of this paper. I am also grateful to Dr. G.R. Seth, Additional Statistical Adviser, for the critical discussion and the referee for his comments which led to the refinement of the analysis. My thanks are also due to Shri L.R. Kad for the computational work.

## REFERENCES

(1) Fisher, R A. : 'The use of multiple measurements in taxonomic problems.' Ann. Eugen. London 7. 179.

(2) Fairfied Smith : 'A discriminant function for plant selection Ann. Eugen.' London 7, 240.

(3) Fisher, R.A. : 'Statistical Utilisation of multiple measurements Ann. Eugen.' London 8, 376.

(4) Bartlett, M.S. : 'Multivariate Analysis Jour. Roy. Stat. Society.' Supplement 9. 1947.

(5) Bartlett, M.S. : 'The goodness of fit of a single hypothetical discriminant function in the case of several groups. Ann. of Eugen' London 16, 1951.

(6) Rao, C.R. : 'Advanced Statistical Methods in Biometric Research Wiley Publ. 1952.

(7) Anderson, T.W. : 'Introduction of Multivariate Statistical Analysis.' Wiley Publ. 1958.

(8) John, S. : 'On some classification statistics Sankhya' 52, 1960.

(9) William, E.J. ; 'Some exact tests in multivariate analysis.' Biometrika 39.

# APPENDIX

*Analysis of Dispersion*

| S. S. and S. P. of Variates | D.F. | *Between Grades* | *Within Grades* | *Total* |
|---|---|---|---|---|
| | | 3 | 395 | 398 |
| $x_1^2$ | | 1127·15 | 7386·43 | 8513·58 |
| $x_2^2$ | | 12 16 | 26·33 | 38·49 |
| $x_3^2$ | | 27·04 | 1046·23 | 1073·27 |
| $x_4^2$ | | 9801·87 | 162291·78 | 172093·65 |
| $x_1\ x_2$ | | —103·15 | —4·95 | —108·10 |
| $x_1\ x_3$ | | 160·77 | 388·58 | 549·35 |
| $x_1\ x_4$ | | 3228·22 | 18738·98 | 21967·20 |
| $x_2\ x_3$ | | —17·73 | —31·08 | —48·81 |
| $x_2\ x_4$ | | —334·30 | —126·31 | —460·61 |
| $x_3\ x_4$ | | 504·32 | 397·37 | 901·69 |

$x_1$=fibre diameter in microns, $x_2$=crimp per centimetre,
$x_3$=percentage of medullated fibres, $x_4$=staple length in cm.

$$E\ (x_3\ x_4/x_1\ x_2)=\begin{pmatrix}1046{\cdot}23 & 397{\cdot}37\\ 397{\cdot}37 & 162291{\cdot}78\end{pmatrix}-\begin{pmatrix}388{\cdot}58 & -31{\cdot}08\\ 18738{\cdot}78 & -126{\cdot}51\end{pmatrix}\begin{pmatrix}{\cdot}000135 & {\cdot}000025\\ {\cdot}000025 & {\cdot}037984\end{pmatrix}\times\begin{pmatrix}388{\cdot}58 & 18738{\cdot}88\\ -31{\cdot}08 & 126{\cdot}51\end{pmatrix}$$

$$=\begin{pmatrix}989{\cdot}71 & -721{\cdot}28\\ 721{\cdot}81 & 114259{\cdot}49\end{pmatrix}=112562750{\cdot}17$$

$$S\ (x_3\ x_4/x_1\ x_2)=\begin{pmatrix}1073{\cdot}27 & 901{\cdot}69\\ 901{\cdot}69 & 172093{\cdot}65\end{pmatrix}-\begin{pmatrix}549{\cdot}35 & -48{\cdot}81\\ 21967{\cdot}20 & -460{\cdot}81\end{pmatrix}\begin{pmatrix}{\cdot}000122 & {\cdot}000342\\ {\cdot}000342 & {\cdot}026941\end{pmatrix}\times\begin{pmatrix}549{\cdot}35 & 21967{\cdot}20\\ -48{\cdot}81 & 460{\cdot}81\end{pmatrix}$$

$$=\begin{pmatrix}990{\cdot}67 & -720{\cdot}73\\ -720{\cdot}73 & 114522{\cdot}94\end{pmatrix}=112934989{\cdot}24$$

$$\frac{E(x_3\ x_4/x_1\ x_2)}{S\ (x_3\ x_4/x_1\ x_2)}={\cdot}9967$$