ASSESSMENT OF FLEECE COMPONENTS USING MULTIVARIATE STATISTICS

T.W. HANCOCK*

Summary

This paper discusses the previous approaches to the resolution of the relationships among the components of clean fleece weight during selection experiments. In view of the statistical difficulties encountered it is proposed that multivariate statistical techniques should be applied. This is demonstrated on data from the Roseworthy Agricultural College fleece weight selection experiment which have previously been analysed by Mayo *et al.* (1969), using univariate statistics.

I. INTRODUCTION

Turner (1958) considered the following three possible methods of assessing the influence of each component on clean fleece weight:-

(i) Gross correlation of each component on clean fleece weight.
(ii) Apportioning of variance.
(iii) Percentage deviation technique.

For the first method the author clearly demonstrated thatcorrelations between the fleece components can greatly affect the observed correlation between each component and clean fleece weight. Similarly the second method was shown to have limitations especially with respect to' sampling errors. She concludes that while the third technique was ". . ..by no means perfect, [it] has proved to be a powerful tool in analysing the source of differences in clean wool weight between groups of sheep". In particular this technique suggested that fibre number and staple length were most closely associated with clean fleece weight. Dun (1958) using this third approach found fibre density and cross-sectional area were most important. Recently this technique was applied by Barlow (1974) but followed up with the calculation of realized correlated responses and realized genetic correlations. Barlow calculated his realized correlated responses by two methods:-

(i) regression of cumulative correlated response on cumulative selection differential;
(ii) regression of cumulative correlated response on cumulative response.

These two responses were subsequently used to calculate two realized genetic correlations. By these methods Barlow concluded that the response in clean fleece weight for the fleece plus flock was due to fibre density, fibre diameter and staple length, while for the fleece minus flock, staple length was the major contributor.

Similarly Robards *et al.* '(1974) used correlations when reporting that crimp frequency was related positively to live weight and negatively to clean fleece weight.

---

* Biometry Section, Waite Agricultural Research Institute, University of Adelaide, Glen Osmond, South Australia, 5064,

Mayo et al. (1969) also encountered difficulties when using repeated t-tests to compare two bases of selection for increased wool production.

It can be seen that repeated t statistics, correlations, regressions and Turner's percent deviation have been extensively used to resolve the responses to selection, especially regarding difference between breeding programs and the behaviour of the fleece components. Although informative, these approaches give no protection against both the effects of correlations among the subsets and the tendency for individual differences to be significant merely by chance as the number of variates increases.

Multivariate statistical techniques are proposed to overcome these difficulties.

II. MATERIALS AND METHODS

(a) Background

Complete details of the sheep used, the selection methods and characters recorded have been given by Mayo et al. (1969).

Briefly, the two flocks were raised at Roseworthy Agricultural College, between 1954 and 1966, either selecting rams by (i) visual appraisal alone (visual method) or (ii) clean fleece weight after initial visual appraisal (index method). The divergence in clean fleece weight of the index over the visual animals was previously established using t-tests. The following eight variates will be considered: clean fleece weight, clean scoured yield percentage, body weight, staple length, crimp, fibre diameter, primary follicle number and secondary follicle number. The other variates recorded were considered unsuitable for multivariate statistics. Only data from single born animals, for which all eight variates had been recorded, were used. All comparisons are made within sexes.

(b) Statistical analysis

The two flocks are compared using Hotelling's T^2 as described in Morrison (1967). This test is basically a multivariate analogue of the square of the univariate t-statistic. Thus two samples can be compared using

$$T^2 = \frac{N_1 N_2}{N_1 + N_2} (\bar{\underset{\sim}{y}}_1 - \bar{\underset{\sim}{y}}_2)' S^{-1} (\bar{\underset{\sim}{y}}_1 - \bar{\underset{\sim}{y}}_2)$$

where N_i, $\bar{\underset{\sim}{y}}_i$ are the number of observations and mean vector for sample i, (i = 1,2), and S is the pooled estimate of the covariance matrix (i.e. a p x p matrix where p is the number of variates measured).

The critical region is

$$T^2 . \frac{(N_1 + N_2 - 2)p}{(N_1 + N_2 - p - 1)} \geq F_{p,(N_1 + N_2 - p)}(\alpha)$$

with significance level α.

The mere significance of the T^2 statistic does not indicate which variates are likely to have led to the rejection of equality, of the two mean vectors. Further it would be erroneous to use univariate t-tests as the number of tests and the correlations among the variates would distort the critical value chosen for the t-statistic. However, use of

T^2 enables calculation of simultaneous confidence intervals for linear functions of the differences. That is, for any vector $\underset{\sim}{a}' = [a_1, \ldots, a_p]$ the probability that all intervals,

$$\underset{\sim}{a}'(\bar{\underset{\sim}{y}}_1 - \bar{\underset{\sim}{y}}_2) - \sqrt{\underset{\sim}{a}' \, S \underset{\sim}{a} \, \frac{N_1 + N_2}{N_1 \, N_2}} \; T\alpha{:}p, N_1 + N_2 - p - 1 \leq \underset{\sim}{a}'\delta$$

$$\leq \underset{\sim}{a}'(\bar{\underset{\sim}{y}}_1 - \bar{\underset{\sim}{y}}_2) + \sqrt{\underset{\sim}{a}' \, S \underset{\sim}{a} \, \frac{N_1 + N_2}{N_1 \, N_2}} \; T\alpha{:}p, N_1 + N_2 - p - 1$$

generated by different choices of the elements of $\underset{\sim}{a}$ are simultaneously true, is (1-a) (where δ is the vector of mean differences). By varying the form of $\underset{\sim}{a}$, a confidence interval can be calculated for each variate which indicates the-magnitude of the difference between flocks. If zero is outside the interval we conclude at the (l-a) ,100 percent joint significance level that the particular variate differs between the two flocks.

A generalised FORTRAN program has been written to apply the above . technique to large data sets.

III. RESULTS AND DISCUSSION

In Tables 1 and 2, the value of Hotelling's T^2, associated value of F, and significance, are presented along with the extremes of the 95% simultaneous confidence intervals.

Although T^2 for the 1960 rams and 1965 ewes are non-significant, the tables indicate the overall divergence of the two flocks. (The latter of these two anomalous results illustrates the well-known inadequacy of discrete cut-off probabilities as the observed value (2.01) is extremely close to the critical value (2.03).)

Assessment of the 95% simultaneous confidence intervals indicates the difference between flocks can seldom be associated with one character. However the position of zero in the interval gives good indication of the variates response to the selection. In particular, the increase in clean fleece weight, observed for the index flock over the visual flock, is seen to be positively associated with clean scoured yield percentage, secondary follicle number and staple length, but negatively associated with crimp number and body weight.

TABLE 1

Hotelling's T^2 and 95% simultaneous confidence interval for rams

Year	T^2	F	Sig	Body wt.	Yield	Clean fleece wt.	Staple length	Crimps per inch	Fibre diameter	Primary Foll. No.	Secondary Foll. No.
1954	8.05	0.95	NS	-3.30† 4.42†	-1.39 1.65	-.43 .61	-.83 .86	-1.19 .81	-2.17 1.42	-13.4 39.1	-274 872
1955	23.01	2.74	**	-4.59 2.22	- .72 1.70	-.32 .37	-.46 1.03	-1.18 .94	- .71 2.09	-13.1 45.1	-318 975
1956	37.16	4.40	***	-3.81 2.96	-1.76 1.65	-.14 .48	-.18 1.20	-1.41 2.08	-2.61 .22	-15.0 51.7	-407 940
1957	25.31	3.00	**	-3.99 3.13	-1.26 1.68	-.27 .41	-.17 1.23	-1.18 1.08	-1.10 2.26	-21.8 60.6	-282 1204
1958	48.71	5.75	***	-2.93 4.29	.04 2.70	-.06 .75	-.24 1.36	- .95 1.41	-1.44 1.58	-20.0 50.6	-604 665
1959	21.60	2.57	*	-3.90 2.35	-1.50 1.84	-.18 .60	-.15 1.24	-1.48 .39	- .58 2.07	-24.1 29.2	-446 535
1960	10.16	1.19	NS	-6.21 2.48	-1.18 2.02	-.29 .58	-.44 1.17	-1.04 .89	-1.28 1.65	-24.5 31.6	-455 709
1961	40.26	4.78	***	-3.30 5.65	- .56 2.74	.09 .89	-.18 1.35	-1.25 .14	-1.75 1.81	-42.6 21.3	-479 634
1962	59.76	7.15	***	-5.83 1.29	- .57 2.06	-.21 .55	-.49 1.07	- .70 .84	-2.51 -.22	-35.3 33.8	- 73 1104
1963	24.41	2.89	**	-4.52 3.29	-1.54 1.75	-.16 .63	-.60 1.02	-1.19 .79	-1.73 .55	-39.3 12.2	-404 712
1964	55.12	6.55	***	-6.06 1.79	- .80 2.18	.05 .78	-.31 1.24	-1.27 .36	-2.75 .44	-29.0 32.0	- 18 1076
1965	34.97	4.12	***	-4.80 3.50	-2.61 1.19	-.14 .65	-.30 1.16	-1.77 .16	-1.39 1.44	-21.3 44.7	-449 715

† Values shown are lower and upper limits of interval respectively.

TABLE 2

Hotelling's T^2 and 95% simultaneous confidence interval for ewes

Year	T^2	F	Sig	Body wt.		Yield		Clean fleece wt.		Staple length		Crimps per inch		Fibre diameter		Primary foll. no.		Secondary foll. no.	
1954	12.62	1.49	NS	-3.13†	3.62†	-1.86	.80	-.46	.30	-.84	.84	-1.21	.56	-2.14	.97	-30.9	33.1	-391	669
1955	29.50	3.47	**	-1.30	4.28	-1.53	1.20	-.04	.54	-.02	1.41	-1.90	.34	- .95	1.75	-30.5	38.4	-611	738
1956	31.84	3.75	***	-3.61	3.16	-1.08	1.72	-.01	.58	-.17	1.10	-2.40	.65	-2.30	1.03	-38.4	42.3	-405	1084
1957	39.96	4.72	***	-3.44	1.96	- .19	2.81	-.07	.47	-.31	.95	-1.66	.21	- .93	2.35	-31.1	52.1	-268	1169
1958	22.16	2.60	*	-2.10	4.02	- .61	2.90	-.05	.68	-.43	1.24	-1.50	.82	-2.49	.92	-35.2	39.1	-454	821
1959	57.72	6.87	***	-2.71	2.63	- .82	2.26	-.01	.59	-.27	1.09	-1.36	.27	.26	2.95	-17.7	57.4	-442	697
1960	28.22	3.37	**	-4.26	1.28	- .86	1.59	-.09	.57	-.44	.80	- .66	.81	-1.18	1.35	-25.4	33.4	-419	626
1961	26.38	3.15	**	-1.91	4.22	- .54	2.02	-.07	.51	-.33	1.00	- .63	.45	-1.48	1.12	-37.5	23.3	-270	804
1962	89.53	10.68	***	-6.43	-.77	- .20	2.36	-.10	.55	-.49	1.03	- .86	.49	-2.49	.11	-31.1	35.8	142	1403
1963	28.61	3.35	**	-3.56	3.20	-1.42	1.61	-.12	.52	-.47	1.01	-1.23	.82	-2.77	.45	-40.8	29.6	572	580
1964	48.73	5.78	***	-6.43	.85	- .78	1.90	-.04	.56	-.49	.85	-1.37	.22	-1.33	1.42	-39.5	49.9	-451	771
1965	17.26	2.01	NS	-6.25	1.79	-1.03	2.48	-.29	.49	-.64	.78	-1.72	.36	-1.45	1.26	-29.6	47.4	-613	916

† Values shown are lower and upper limits of interval respectively.

The trend for fibre diameter was unclear, considerable variation between seasons being observed. As fibre diameter is such an important factor in quality, and both Turner (1958) and Barlow (1974) have observed similar behaviour, further research on this variate could be rewarding.

III. ACKNOWLEDGEMENTS

Firstly, I would like to thank Dr. O. Mayo for his constructive , and patient interest in my work.

Secondly, I wish to thank the Director of Roseworthy Agricultural College for making the data available and acknowledge the splendid work of all those who maintained and recorded the Roseworthy flocks. (In particular, Messrs. R.E. Brady, C.W. Hooper, P.G. Schinckel, J.C. Hawker, K.J. Hutchinson, D. Heaton Harris, R.B. Porter, G. Ford, B. Schuff, C.S. Bungey and J. Wood, to mention just a few.)

IV. REFERENCES

BARLOW, R. (1974) Australian Journal of agricultural Research, 25: 973.

DUN, R.B. (1958) Australian Journal of agricultural Research, 9: 802.

MAYO, O. (1969) Australian Journal of agricultural Research, 20: 151.

MORRISON, D.F. (1967) "Multivariate statistical methods" (McGraw-Hill, New York).

ROBARDS, G.E., WILLIAMS, A.F. and HUNT, M.H.R. (1974) Australian Journal of experimental Agriculture and Animal Husbandry, 14: 441.

TURNER, H.N. (1958) Australian Journal of agricultural Research, 9: 521.