Ann. Inst. Statist. Math.
Vol. 50, No. 2, 241–252 (1998)

THE DECOMPOSITION OF THE BEHRENS-FISHER STATISTIC IN q-DIMENSIONAL COMMON PRINCIPAL COMPONENT SUBMODELS

D. G. NEL[1] AND I. PIENAAR[2]

[1] *Department of Mathematical Statistics, University of the OFS, P.O. Box 339, Bloemfontein, 9300, South Africa*
[2] *Department of Statistics, Vista University, Bloemfontein, 9300, South Africa*

(Received December 11, 1995; revised May 27, 1996)

Abstract. Takemura (1985, *Multivariate Analysis VI*, ed. P. R. Krishnaiah, 583 597, Elsevier, Amsterdam) presented a decomposition of Hotelling's T^2-statistic into analogues of univariate Student-t variates along the principal component axes of the (pooled) sample covariance matrix. In this paper the idea is extended to the heteroscedastic situation where an analogous decomposition of the Behrens-Fisher statistic is considered when the nature of the heteroscedasticity between the two samples can be described by a common principal component (CPC) model, or more particularly a q-dimensional CPC subspace model.

Key words and phrases: Covariance matrices, heteroscedasticity, iterative procedure, model fitting, randomization testing, Takemura decomposition.

1. Introduction

Consider the following setup from multivariate quality control. A machine produces items and p variates are measured on each item. In an initial phase, N_1 "good" observations $\underline{x}_{11}, \dots, \underline{x}_{1N_1}$ are collected and at a later stage N_2 observations $\underline{x}_{21}, \dots, \underline{x}_{2N_2}$ are obtained which the quality controller compares with the first set of observations, which then serves as a standard for comparison.

Flury *et al.* (1995) introduced a q-dimensional Common Principal Components (CPC) subspace model in the sense that the two populations are allowed to differ in both mean and variance in q directions ($0 \le q \le p$) corresponding to q CPC's-more formally: The p-variate random vector $\underline{X}$, distributed as $N_p(\underline{\mu}_j, \Sigma_j)$ ($j = 1, 2$) in two populations, follows a q-dimensional CPC subspace model if there exists an orthogonal matrix $\boldsymbol{\beta} = [\underline{\beta}_1 \cdots \underline{\beta}_p]$ such that:

(i) $\boldsymbol{\beta}'\Sigma_j\boldsymbol{\beta}$ is diagonal, $j = 1, 2$
(ii) $\underline{\beta}_i'(\underline{\mu}_1 - \underline{\mu}_2) = 0$ for $i = q + 1, \dots, p$ and
(iii) $\underline{\beta}_i'\Sigma_1\underline{\beta}_i = \underline{\beta}_i'\Sigma_2\underline{\beta}_i$ for $i = q + 1, \dots, p$.

So in a q-dimensional CPC subspace model all differences between the two groups can be summarized in a subspace of dimension q; the remaining $p-q$ directions being redundant. For $q=p$, the model corresponds to Flury's CPC model with no constraints imposed on the mean vectors (Flury (1988)). For $q=1$, the populations differ in exactly one CPC direction and for $q=0$ the two distributions are identical. By varying q between 0 and p and choosing q directions from the p directions for each value of q, a total of 2^p such models can be considered. The question of identifying q, which is the selection of the best fitting model, is of particular interest. Flury *et al.* (1995) used a likelihood ratio statistic and a model selection procedure based on the Akaike Information Criterion (AIC). In this paper we propose an adaptation of a method used by Takemura (1985) to identify and confirm the q directions of change in a CPC(q) subspace model. The method of Takemura itself is used to identify the principal components (PC's) which contribute the most to the rejection of the null hypothesis of equal mean vectors in the homoscedastic situation.

In Section 2 we give a brief overview of the estimation and selection procedure used by Flury *et al.* (1995) to identify q, since it is essential to the rest of this paper. In Section 3 we describe the Takemura decomposition as applied to the heteroscedastic case and particularly for a CPC(q) subspace model. An example is presented in Section 4.

2. Estimation, testing and selection procedures for a CPC(q) subspace model

Suppose independent samples $\underline{x}_{j1},\ldots,\underline{x}_{jN_j}$ are obtained from two p-variate normal populations, $N_j > p$, $j=1,2$. The joint log-likehood function is proportional to

$$L(\underline{\mu}_1,\underline{\mu}_2,\Sigma_1,\Sigma_2)=\sum_{j=1}^{2}[N_j\log|\Sigma_j|+\operatorname{tr}(\Sigma_j^{-1}A_j)],$$

where

$$A_j=\sum_{i=1}^{N_j}(\underline{x}_{ji}-\underline{\mu}_j)(\underline{x}_{ji}-\underline{\mu}_j)',\qquad j=1,2.$$

Assuming that a CPC(q) subspace model holds, then

$$\boldsymbol{\beta}'\Sigma_j\boldsymbol{\beta}=\Lambda_j=\operatorname{diag}(\lambda_1^{(j)},\ldots,\lambda_q^{(j)},\lambda_{q+1},\ldots,\lambda_p),$$

where $\boldsymbol{\beta}=[\underline{\beta}_1,\ldots,\underline{\beta}_p]$ is orthogonal and

$$\boldsymbol{\beta}'\underline{\mu}_j=\underline{\nu}_j=[\nu_1^{(j)},\ldots,\nu_q^{(j)},\nu_{q+1},\ldots,\nu_p]';\qquad j=1,2.$$

Using standard techniques it was shown that for given $\underline{\beta}_1,\ldots,\underline{\beta}_p$ the maximum likelihood estimates are given by:

$$\hat{\nu}_\ell^{(j)}=\underline{\beta}_\ell'\underline{\bar{x}}_\ell,\qquad \hat{\lambda}_\ell^{(j)}=\underline{\beta}_\ell'S_j\underline{\beta}_\ell\qquad j=1,2;\qquad \ell=1,\ldots,q,\tag{2.1}$$

where

$$\underline{\overline{x}}_j = \frac{1}{N_j}\sum_{i=1}^{N_j} \underline{x}_{ji} \quad \text{and} \quad S_j = \frac{1}{N_j}\sum_{i=1}^{N_j}(\underline{x}_{ji} - \underline{\overline{x}}_j)(\underline{x}_{ji} - \underline{\overline{x}}_j)'$$

and

$$\hat{\nu}_\ell = \underline{\beta}_\ell' \underline{\overline{x}}_T, \qquad \hat{\lambda}_\ell = \underline{\beta}_\ell' S_T \underline{\beta}_\ell \qquad \ell = q+1, \ldots, p, \tag{2.2}$$

where

$$\underline{\overline{x}}_T = \frac{1}{(N_1+N_2)}(N_1\underline{\overline{x}}_1 + N_2\underline{\overline{x}}_2)$$

and

$$S_T = \frac{1}{(N_1+N_2)}\left(N_1S_1 + N_2S_2 + \frac{N_1N_2}{(N_1+N_2)}(\underline{\overline{x}}_1 - \underline{\overline{x}}_2)(\underline{\overline{x}}_1 - \underline{\overline{x}}_2)'\right). \tag{2.3}$$

The maximum likelihood estimates of the orthonormal vectors $\underline{\beta}_1, \ldots, \underline{\beta}_p$ follow from:

$$\hat{\underline{\beta}}_m' T_{m\ell} \hat{\underline{\beta}}_\ell = 0 \qquad 1 \le m < \ell \le p \tag{2.4}$$

where

$$\begin{aligned} T_{m\ell} &= \sum_{j=1}^{2}\left(\frac{\hat{\lambda}_{jm} - \hat{\lambda}_{j\ell}}{\hat{\lambda}_{jm}\hat{\lambda}_{j\ell}}\right)A_j \qquad \text{if} \quad 1 \le m < \ell \le q \\ &= \frac{1}{\hat{\lambda}_\ell}(N_1+N_2)S_T - \sum_{j=1}^{2}\frac{1}{\hat{\lambda}_m^{(j)}}N_jS_j \qquad \text{if} \quad 1 \le m \le q < \ell \le p \\ &= S_T \qquad \text{if} \quad q < m < \ell \le p \end{aligned} \tag{2.5}$$

and

$$\hat{\lambda}_{jh} = \begin{cases} \hat{\lambda}_h^{(j)} & \text{for} \quad h = 1, \ldots, q \\ \hat{\lambda}_h & \text{for} \quad h = q+1, \ldots, p \end{cases} \qquad j = 1, 2. \tag{2.6}$$

Thus the likelihood equations consist of (2.1), (2.2) and (2.4), to be solved under the orthogonality constraints $B'B = I_p$ by an iterative procedure described by Flury *et al.* (1995). The paper of Schott (1988) also gives valuable information regarding procedures to obtain the estimate $B = \hat{\boldsymbol{\beta}} = [\hat{\underline{\beta}}_1 \ldots \hat{\underline{\beta}}_p]$.

To test the hypothesis that a CPC(q) model resulted from a CPC(r) model where $r > q$ i.e.

$$H(r,q) : \nu_h^{(1)} = \nu_h^{(2)} \quad \text{and} \quad \lambda_h^{(1)} = \lambda_h^{(2)} \qquad \text{for} \quad h = q+1, \ldots, r,$$

the log likelihood ratio statistic is given by

$$\text{LRS}(r,q) = (N_1+N_2)\sum_{\ell=q+1}^{r}\log\hat{\lambda}_\ell - \sum_{j=1}^{2}N_j\sum_{\ell=q+1}^{r}\log\tilde{\lambda}_{j\ell}, \tag{2.7}$$

where "$\wedge$" denotes the maximum likelihood estimates under the CPC(q) model and "$\sim$" denotes these estimates under the CPC(r) model. The hypothesis $H(r,q)$ is rejected at approximate significance level α under the usual chi-square approximation, if $\text{LRS}(r,q) > \chi^2_{2(r-q);1-\alpha}$, where $\chi^2_{v;1-\alpha}$ denotes the $100(1-\alpha)$ the percentile of the chi-square distribution with v degrees of freedom.

The statistic LRS(1, 0) can be computed for each of the p directions in order to identify directions with large differences. The statistic

$$M_1 = \max\{\text{LRS}(1,0)\} \tag{2.8}$$

then asymptotically follows the distribution of the maximum of p independent chi-square random variables with two degrees of freedom each. The independence follows from the orthogonality of the latent variables.

When the use of (2.8) has lead to the identification of the directions of change, the Akaike information criterion (Akaike (1973))

$$\text{AIC}(q) = g_q - g_{nc} + 4q \tag{2.9}$$

can be used to select the best fitting model, where $g_q = \sum_{j=1}^{2} N_j \sum_{i=1}^{p} \log(\hat{\lambda}_{j\ell})$ denotes the log likelihood function for group specific change in q directions and $g_{nc} = \sum_{j=1}^{2} N_j \log|S_j|$ denotes the corresponding function under a no-constraints model. The AIC for the no-constraints model is then defined by: $\text{AIC}(nc) = p(p+3)$.

The model with the smallest AIC will then be chosen as the best fitting model.

In the next section we present the Takemura decomposition of the Behrens-Fisher statistic as a confirmatory procedure for the choice of the best fitting model by the AIC method.

3. The Takemura decomposition of the Behrens-Fisher statistic under a CPC(q) model

Takemura (1985) considered the following decomposition of Hotelling's T^2 statistic, which in our notation can be written as:

$$T^2 = \frac{N_1 N_2}{N_1 + N_2}(\underline{\overline{X}}_1 - \underline{\overline{X}}_2)' S_e^{-1} (\underline{\overline{X}}_1 - \underline{\overline{X}}_2), \tag{3.1}$$

where S_e is the unbiased estimator of the common covariance matrix $\Sigma = \Sigma_1 = \Sigma_2$ under the assumption of homoscedasticity. Using the principal component transformation $\underline{Y} = B'\underline{X}$, where B is an orthogonal matrix such that $B'S_eB = L = \text{diag}(\ell_1, \ldots, \ell_p)$ and $\underline{\overline{Y}}_i = B'\underline{\overline{X}}_i$, $i = 1, 2$, Takemura (1985) decomposed Hotelling's T^2 statistic as the sum of p t_h^2-components

$$T^2 = \sum_{h=1}^{p} t_h^2, \tag{3.2}$$

where the t_h's are analogues of Student-t variates

$$t_h = \frac{(\overline{Y}_{1h} - \overline{Y}_{2h})}{\sqrt{\ell_h\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}} \qquad h = 1, \ldots, p. \tag{3.3}$$

Comparing the relative contribution of each t_h^2 to Hotelling's T^2 makes the T^2-test much more interesting from an interpretation point of view.

When S_i, $i = 1, 2$ is defined as in (2.1), the multivariate Behrens-Fisher statistic, (Behrens (1929), Fisher (1939)), which is the union-intersection test statistic for the test of equal mean vectors $H_0 : \underline{\mu}_1 = \underline{\mu}_2$ under the assumption that Σ_1 and Σ_2 are not the same, is given by

$$\text{BF} = (\underline{\overline{X}}_1 - \underline{\overline{X}}_2)'\left(\frac{1}{N_1 - 1}S_1 + \frac{1}{N_2 - 1}S_2\right)^{-1}(\underline{\overline{X}}_1 - \underline{\overline{X}}_2). \tag{3.4}$$

One of the approximate degrees of freedom tests like those of Nel and van der Merwe (1986) or Kim (1992) can be used to test for the significance of the BF-statistic.

This statistic may also be considered as the estimator of the corresponding heteroscedastic Mahalanobis distance:

$$\Delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)'\left(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2\right)^{-1}(\underline{\mu}_1 - \underline{\mu}_2). \tag{3.5}$$

Under a CPC(q) model, this Mahalanobis distance may be written as:

$$\begin{aligned}\Delta^2 &= \sum_{h=1}^{q} \frac{(\nu_h^{(1)} - \nu_h^{(2)})^2}{\frac{1}{N_1}\lambda_h^{(1)} + \frac{1}{N_2}\lambda_h^{(2)}} + \sum_{h=q+1}^{p} \frac{(\nu_h - \nu_h)^2}{\lambda_h\left(\frac{1}{N_1} + \frac{1}{N_2}\right)} \\ &= \sum_{h=1}^{q} \Delta_h^2,\end{aligned} \tag{3.6}$$

where

$$\Delta_h^2 = \frac{(\nu_h^{(1)} - \nu_h^{(2)})^2}{\frac{1}{N_1}\lambda_h^{(1)} + \frac{1}{N_2}\lambda_h^{(2)}}.$$

The Behrens-Fisher statistic (3.4) can likewise be decomposed by using the estimated orthogonal matrix B in Section 2 for a CPC(q) model namely

$$\begin{aligned}\text{BF} &= (B(\underline{\overline{X}}_1 - \underline{\overline{X}}_2))'\left(\frac{1}{N_1 - 1}B'S_1B + \frac{1}{N_2 - 1}B'S_2B\right)^{-1}(B(\underline{\overline{X}}_1 - \underline{\overline{X}}_2)) \\ &= \text{TAK} + \sum_{\substack{h=1 \\ h\neq k}}^{p}\sum_{k=1}^{p} \frac{\underline{\hat{\beta}}_h'(\underline{\overline{X}}_1 - \underline{\overline{X}}_2)(\underline{\overline{X}}_1 - \underline{\overline{X}}_2)'\hat{\beta}_k}{\frac{1}{N_1 - 1}\underline{\hat{\beta}}_h'S_1\underline{\hat{\beta}}_k + \frac{1}{N_2 - 1}\underline{\hat{\beta}}_h'S_2\underline{\hat{\beta}}_k},\end{aligned} \tag{3.7}$$

where

$$(3.8)\qquad \begin{aligned} \mathrm{TAK} &= \sum_{h=1}^{p} \frac{(\hat{\underline{\beta}}_h'(\overline{X}_1-\overline{X}_2))^2}{\frac{1}{N_1-1}\hat{\underline{\beta}}_h' S_1 \hat{\underline{\beta}}_h + \frac{1}{N_2-1}\hat{\underline{\beta}}_h' S_2 \hat{\underline{\beta}}_h} \\ &= \sum_{h=1}^{q} \frac{(\hat{\nu}_h^{(1)}-\nu_h^{(2)})^2}{\frac{1}{N_1-1}\hat{\lambda}_h^{(1)} + \frac{1}{N_2-1}\hat{\lambda}_h^{(2)}} \\ &\quad + \sum_{h=q+1}^{p} \frac{(\hat{\underline{\beta}}_h'(\overline{X}_1-\overline{X}_2))^2}{\frac{1}{N_1-1}\hat{\underline{\beta}}_h' S_1 \hat{\underline{\beta}}_h + \frac{1}{N_2-1}\hat{\underline{\beta}}_h' S_2 \hat{\underline{\beta}}_h} \\ &= \sum_{h=1}^{q} t_h^2 + \sum_{h=q+1}^{p} t_h^2 \end{aligned}$$

and where

$$(3.9)\qquad t_h^2 = \frac{(\hat{\underline{\beta}}_h'\overline{X}_1 - \hat{\underline{\beta}}_h'\overline{X}_2)^2}{\frac{1}{N_1-1}\hat{\underline{\beta}}_h' S_1 \hat{\underline{\beta}}_h + \frac{1}{N_2-1}\hat{\underline{\beta}}_h' S_2 \hat{\underline{\beta}}_h} \qquad \text{for} \quad h = 1,\ldots,p.$$

If a CPC(q) model fits the data exactly, then TAK equals BF and the relative contributions of the t_h^2 to BF (or TAK) in the first term of (3.8) will indicate which directions contribute mostly to the BF statistic. To test for significant contributions of the individual t_h^2 is more difficult, since even in the homoscedastic situation as in (3.3), the distributions of the t_h^2 are unknown (Takemura (1985)).

Since no CPC(q) model can fit the data perfectly, we are always looking at a situation where the matrix B diagonalizes S_1 and S_2 as closely as possible under the fitted CPC(q) model for a particular value of q. Consequently $B'S_1B$ and $B'S_2B$ will be close to diagonal matrices under a good fit but far from diagonal matrices under a bad fit and TAK will be close to BF under a good fit and far from BF under a bad fit.

The decomposition method itself however, is useful to illustrate and confirm the directions of change discovered by using (2.7) and (2.8) under a CPC(q) subspace model, particularly when the directions are to be interpreted in terms of a multivariate quality control setup.

It is interesting to note that if $q = p$, when the CPC model of Flury is valid, the decomposition (3.8) may still be used to explore the nature of the differences between mean vectors when no constraints are applied to them (Richards (1993)).

4. Example

To illustrate the method we use the head measurements of football players data in Rencher ((1995), p. 305) where six head measurements were made on two groups of football players (we consider groups 1 and 2 of Rencher's three groups) namely high school football players (group 1) of size $N_1 = 30$ and college football players (group 2) of size $N_2 = 30$. The six variables are:

X_1 = WDIM = head width at widest dimension,
X_2 = CIRCUM = head circumference,
X_3 = FBEYE = front-to-back measurement at eye level,
X_4 = EYEHD = eye-to-top-of-head measurement,
X_5 = EARHD = ear-to-top-of-head measurement,
X_6 = JAW = jaw width.

The sample mean vectors are

$$\underline{\overline{x}}_1' = [15.20 \quad 58.95 \quad 20.11 \quad 13.08 \quad 14.73 \quad 12.27]$$
$$\underline{\overline{x}}_2' = [15.42 \quad 57.39 \quad 19.80 \quad 10.08 \quad 13.45 \quad 11.94]$$

and the sample covariance matrices (see (2.1)) are

$$S_1 = \begin{bmatrix} 0.526 & 0.522 & 0.166 & 0.225 & 0.170 & 0.238 \\ & 4.061 & 1.382 & 0.753 & 0.823 & 0.693 \\ & & 0.688 & 0.206 & 0.403 & 0.227 \\ & & & 1.051 & 0.522 & 0.169 \\ & & & & 0.862 & 0.079 \\ & & & & & 0.462 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 0.393 & 0.598 & 0.189 & -0.224 & 0.109 & 0.247 \\ & 2.834 & 0.912 & 0.191 & 0.090 & 0.303 \\ & & 0.534 & -0.061 & -0.001 & 0.124 \\ & & & 1.114 & 0.084 & -0.152 \\ & & & & 0.551 & -0.008 \\ & & & & & 0.365 \end{bmatrix}$$

First consider the model for $q = 0$, which corresponds to equality of mean vectors and covariance matrices. The maximum likelihood estimates of $\boldsymbol{\beta}$ and $\Lambda = \Lambda_1 = \Lambda_2$ are given by the eigenvectors and eigenvalues of the total sample covariance matrix S_T as in (2.2). The column vectors of the estimated matrix $B = \hat{\boldsymbol{\beta}}$ are given as row vectors in Table 1(b).

Table 1(a) gives the estimates $\hat{\nu}_1 = \hat{\underline{\nu}}_2$ $(= \hat{\underline{\nu}})$ and $\hat{\Lambda} = \mathrm{diag}(B'S_TB)$ together with the groupwise sample mean vectors $B'\underline{\overline{x}}_1$, $B'\underline{\overline{x}}_2$ and diagonal elements of $B'S_1B$ and $B'S_2B$.

Note that the largest differences among $\hat{\nu}_h$, $\underline{\hat{\beta}}_h'\underline{\overline{x}}_1$ and $\underline{\hat{\beta}}_h'\underline{\overline{x}}_2$ and between $\underline{\hat{\beta}}_h'S_1\underline{\hat{\beta}}_h$ and $\underline{\hat{\beta}}_h'S_2\underline{\hat{\beta}}_h$ occur at component $h = 2$. The Takemura decomposition of the Behrens-Fisher statistic (3.7) is given in Table 1(b). The value of the Behrens-Fisher statistic for the test of $H_0 : \underline{\mu}_1 = \underline{\mu}_2$ is given by BF $= 139.64$. Using the approximate degrees of freedom test of Nel and van der Merwe (1986), the approximate degrees of freedom is $\hat{f} = 55.92$ and the approximate value of the F-statistic $\mathrm{F} = \frac{\mathrm{BF}}{\hat{f}} \cdot \frac{\hat{f}-p+1}{p}$ is $\mathrm{F} = 21.19$, yielding a p-value of $p = 1.68 \times 10^{-8}$ from

Table 1(a).

h	$\hat{\nu}_h$	$\hat{\lambda}_h$	$\hat{\underline{\beta}}'_h\bar{x}_1$	$\hat{\underline{\beta}}'_h\bar{x}_2$	$\hat{\underline{\beta}}'_h S_1\hat{\underline{\beta}}_h$	$\hat{\underline{\beta}}'_h S_2\hat{\underline{\beta}}_h$
1	0.88	0.163	0.85	0.91	0.186	0.137
2	59.60	6.241	61.25	57.95	4.509	2.513
3	−1.67	0.214	−1.63	−1.72	0.225	0.198
4	−29.99	2.401	−29.22	−30.77	1.648	1.954
5	5.61	0.572	5.59	5.63	0.503	0.640
6	2.95	0.464	2.98	2.93	0.580	0.347

Table 1(b).

h	$\hat{\underline{\beta}}'_h S_1\hat{\underline{\beta}}_h$	$\hat{\underline{\beta}}'_h S_2\hat{\underline{\beta}}_h$	t_h^2			$\hat{\underline{\beta}}'_h$			
1	0.186	0.137	0.42	0.47	−0.26	0.69	0.21	−021	−0.39
2	4.509	2.513	45.12	0.06	0.73	0.22	0.57	0.29	0.12
3	0.225	0.198	0.60	−0.49	−0.23	0.59	−0.02	0.01	0.60
4	1.648	1.954	19.31	−0.22	−0.54	−0.26	0.72	0.22	−0.12
5	0.503	0.640	0.04	0.41	−0.22	0.06	−0.25	0.83	0.16
6	0.580	0.347	0.07	0.57	−0.11	−0.23	0.22	−0.36	0.66

TAK = 65.55

Table 2(a).

h	$\hat{\nu}_{1h}$	$\hat{\nu}_{2h}$	$\hat{\lambda}_{1h}$	$\hat{\lambda}_{2h}$	$\hat{\underline{\beta}}'_h\bar{x}_1$	$\hat{\underline{\beta}}'_h\bar{x}_2$	$\hat{\underline{\beta}}'_h S_1\hat{\underline{\beta}}_h$	$\hat{\underline{\beta}}'_h S_2\hat{\underline{\beta}}_h$
1	0.31	0.31	0.163	0.163	0.28	0.35	0.188	0.136
2	67.68	65.60	5.099	3.286	67.68	65.60	5.099	3.286
3	0.91	0.91	0.214	0.214	−0.86	−0.95	0.229	0.194
4	4.00	4.00	3.371	3.371	5.50	2.50	0.055	1.183
5	5.13	5.13	0.572	0.572	5.11	5.15	0.503	0.641
6	3.43	3.43	0.464	0.464	3.45	3.40	0.578	0.349

the F-distribution with p and $\hat{f} - p + 1$ degrees of freedom. The null-hypothesis is therefore rejected and we wish to determine with our procedure which common principal directions contributed mainly to this change between high school and college football players' head measurements.

Note that the largest contributing t_h^2 is at $h = 2$ and $h = 4$. The fact that TAK is so different from the BF statistic's value indicates that the model fit for $q = 0$ is rather poor.

For $q = 1$, a CPC(1) model is fitted to the data to model change in direction

Table 2(b).

h	$\underline{\hat\beta}_h' S_1 \underline{\hat\beta}_h$	$\underline{\hat\beta}_h' S_2 \underline{\hat\beta}_h$	t_h^2	$\underline{\hat\beta}_h'$					
1	0.188	0.136	0.46	0.46	−0.26	0.69	0.21	−0.21	−0.38
2	5.099	3.286	14.91	0.17	0.90	0.32	0.13	0.15	0.15
3	0.229	0.194	0.68	−0.49	−0.22	0.59	−0.02	0.02	0.61
4	1.055	1.183	116.72	−0.17	−0.10	−0.12	0.91	0.34	−0.03
5	0.503	0.641	0.04	0.41	−0.23	0.06	−0.25	0.83	0.16
6	0.578	0.349	0.07	0.57	−0.11	−0.23	0.22	−0.35	0.66

TAK = 132.90

Table 3(a).

h	$\hat\nu_{1h}$	$\hat\nu_{2h}$	$\hat\lambda_{1h}$	$\hat\lambda_{2h}$	$\underline{\hat\beta}_h' \overline{\underline{x}}_1$	$\underline{\hat\beta}_h' \overline{\underline{x}}_2$	$\underline{\hat\beta}_h' S_1 \underline{\hat\beta}_h$	$\underline{\hat\beta}_h' S_2 \underline{\hat\beta}_h$
1	0.93	0.93	0.236	0.236	0.75	1.11	0.246	0.161
2	67.88	65.64	5.142	3.247	67.88	65.64	5.142	3.247
3	0.10	0.10	0.169	0.169	0.02	0.18	0.168	0.159
4	1.50	−1.34	1.027	1.231	1.50	−1.34	1.027	1.231
5	4.95	4.95	0.574	0.574	4.88	5.00	0.494	0.647
6	3.66	3.66	0.476	0.476	3.79	3.54	0.574	0.345

Table 3(b).

h	$\underline{\hat\beta}_h' S_1 \underline{\hat\beta}_h$	$\underline{\hat\beta}_h' S_2 \underline{\hat\beta}_h$	t_h^2	$\underline{\hat\beta}_h'$					
1	0.246	0.161	9.09	0.61	0.17	−0.45	0.00	−0.12	−0.62
2	5.142	3.247	17.45	0.16	0.89	0.31	0.18	0.17	0.14
3	0.168	0.159	2.12	0.37	−0.30	0.79	0.17	−0.23	−0.24
4	1.027	1.231	103.56	−0.14	−0.15	−0.13	0.89	0.35	−0.15
5	0.494	0.647	0.33	0.40	−0.22	0.07	−0.27	0.84	0.12
6	0.574	0.345	2.06	0.53	−0.13	−0.22	0.26	−0.29	0.71

TAK = 134.61

2. For this model the numerical results for $\hat\Lambda_1$, $\hat\lambda_2$, $\underline{\hat\nu}_1$, $\underline{\hat\nu}_2$ along with the group-wise sample means $B'\overline{\underline{x}}_1$, $B'\overline{\underline{x}}_2$ and variances in $B'S_1B$ and $B'S_2B$ are given in Table 2(a).

Note that the largest differences between $\underline{\hat\beta}_h'\overline{\underline{x}}_1$ and $\underline{\hat\beta}_h'\overline{\underline{x}}_2$ and between $\underline{\hat\beta}_h' S_1 \underline{\hat\beta}_h$ and $\underline{\hat\beta}_h' S_2 \underline{\hat\beta}_h$ occur at $h = 4$. Thus direction 4 may be the next possible direction

Table 4.

h	$\underline{\hat{\beta}}'_h S_1 \underline{\hat{\beta}}_h$	$\underline{\hat{\beta}}'_h S_2 \underline{\hat{\beta}}_h$	t_h^2	$\underline{\hat{\beta}}'_h$					
1	0.290	0.079	8.34	0.78	−0.15	0.19	0.14	−0.18	−0.54
2	5.130	3.253	17.39	0.18	0.89	0.32	0.18	0.16	0.13
3	0.174	0.219	1.25	−0.16	−0.33	0.89	0.12	−0.09	0.24
4	1.020	1.236	104.10	−0.17	−0.15	−0.13	0.89	0.36	−0.13
5	0.485	0.652	0.16	0.31	−0.21	0.06	−0.29	0.86	0.16
6	0.554	0.352	2.76	0.46	−0.13	−0.23	0.24	−0.25	0.77

TAK = 134.00

Table 5.

q	Components	AIC	TAK
0		105.43	65.55
1	1	98.61	72.94
1	2	104.81	132.90
1	3	98.61	72.93
1	4	62.14	130.66
1	5	107.35	70.65
1	6	106.71	67.16
2	1 , 2	97.92	139.63
2	1 , 4	50.92	133.65
2	2 , 3	97.92	139.63
2	2 , 4	52.35	134.61
2	2 , 5	106.24	135.82
3	1 , 2 , 4	39.07	134.00
3	1 , 3 , 4	39.07	133.99
3	2 , 4 , 5	55.06	146.71
4	1 , 2 , 4 , 6	38.83	135.35
4	2 , 3 , 4 , 6	38.83	135.34
5	2 , 3 , 4 , 5 , 6	40.8 8	145.76
6	1 , 2 , 3 , 4 , 5 , 6	42.85	145.54

to include.

The Takemura decomposition of the Behrens-Fisher statistic is now given by Table 2(b). The column vectors of the corresponding orthogonal matrix B are given as the row vectors in Table 2(b).

The largest contributing t_h^2 are at $h = 4$ and $h = 2$ and TAK is much closer to the value of the BF-statistic indicating that we have made some progress in choosing the correct directions.

For $q = 2$, a CPC(2) model is fitted to model change in directions 2 and 4. The numerical results for this fit are given in Table 3(a) and the Takemura decomposition in Table 3(b).

Note that TAK has moved even closer to the value of the BF-statistic and that from Table 3(a) direction 1 now emerges as a possible direction where change has occurred. If we construct another Takemura decomposition for change in directions 1, 2 and 4 we get the results as displayed in Table 4.

Note that TAK became slightly smaller, an indication that we may have reached the optimum model for change in directions 2 and 4 as given in Table 3(b).

In Table 5 we report the smallest AIC's computed from (2.8) together with the largest TAK's for different q and corresponding components for change.

From these values we notice that small AIC-values are usually accompanied by large TAK-values. For $q = 1$, component 4 again seems to be the best choice for change. For $q = 2$, components 2 and 4 seem reasonable while 1 and 4 seem like a possibility too. However for $q = 1$, component 1 seems to be a bad choice. For $q = 3$ either components 1, 2 and 4 or components 2, 3 and 4 can be used, although for $q = 3$, the values of TAK are moving away in general from the BF statistic's value. On the basis of these values, it seems reasonable to choose components 2 and 4 for $q = 2$ as those yielding small AIC and large TAK-values.

If we investigate the loadings in these directions i.e. the components of the vectors $\hat{\underline{\beta}}_2'$ and $\hat{\underline{\beta}}_4'$, we note that the variables X_2, (head circumference) and X_4, (eye-to-top-of-head measurement) are mainly responsible for this change.

5. Discussion

The Takemura decomposition of the Behrens-Fisher statistic is a useful method to confirm the choice of a particular CPC(q) subspace model in describing the change which has occurred when a multivariate quality control process gives an out-of-control signal.

Acknowledgements

Part of this research was done while the first author was visiting the Statistics Department at Stanford University during the summer of 1995. We wish to thank professors T. W. Anderson and Richard Olshen for valuable discussions.

We also wish to thank the referees for their comments regarding more consistent notation and clarification.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csáki), 267–281, Akadémiai Kiadó, Budapest.

Behrens, W. U. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen, *Landwirtsch Jahrbucher*, **68**, 607–837.

Fisher, R. A. (1939). The comparison of samples with possibly unequal variances, *Annals of Eugenics*, **9**, 174–180.

Flury, B. (1988). *Common Principal Components and Related Multivariate Models*, Wiley, New York.

Flury, B., Nel, D. G. and Pienaar, I. (1995). Simultaneous detection of shift in means and variances., *J. Amer. Statist. Assoc.*, **90** (432), 1474–1481.

Kim, S-J. (1992). A practical solution to the multivariate Behrens-Fisher problem, *Biometrika*, **79**, 171–176.

Nel, D. G. and van der Merwe, C. A. (1986). A solution to the multivariate Behrens-Fisher problem, *Comm. Statist. Theory Methods*, **19**, 279–298.

Rencher, A. C. (1995). *Methods of Multivariate Analysis*, Wiley, New York.

Richards, W. D. (1993). Modelling multivariate heteroscedasticity with a common principal component decomposition of the Behrens-Fisher statistic, Tech. Report, No. 193, Department of Mathematical Statistics, University of the Orange Free State, Bloemfontein, South Africa.

Schott, J. R. (1988). Common principal component subspaces in two groups, *Biometrika*, **75** (2), 229–236.

Takemura, A. (1985). A principal decomposition of Hotelling's T^2 statistic, *Multivariate Analysis VI* (ed. P. R. Krishnaiah), 583–597, Elsevier Science Publishers, Amsterdam.