Multivariate Statistical Analysis, Exercise 6, Fall 2011, Prof. S.K. Jeng

December 2, 2011 TA: H.C. Cheng

(Profile Analysis) Here we consider the problems of profile analysis. Before going on, recall the corollary:

Corollary Consider a linear transform of $X \sim N_p(\mu, \Sigma)$, $Y = AX$ where $A_{q\times p}$ with $(q \leq p)$. Let $\overline{x}$ and S_X be the sample mean and the covariance matrix, we have

$$\begin{aligned} \overline{y} = A\overline{x} &\sim N_q(A\mu, \frac{1}{n}A\Sigma A^T) \\ nS_Y = nAS_XA^T &\sim W_q(A\Sigma A^T, n-1) \\ (n-1)(A\overline{x} - A\mu)^T(AS_XA^T)^{-1}(A\overline{x} - A\mu) &\sim T^2(q, n-1), \end{aligned}$$

where T^2 is Hostelling's T^2-distribution and W is the Wishart distribution.

The repeated measurements problem arises in practice when we observe repeated measurements of characteristics (or measures of the same type under different experimental conditions) on the different groups which have to be compared. It is important that the p measures (the "profile") are comparable and in particular are reported in the same units. For instance, the observations may be the scores obtained from p different tests of two different experimental groups. One is then interested in comparing the profiles of each group: the profile being just the vectors of the means of the p responses.

We are thus in the same statistical situation as for the comparison of twomeans:

$$\begin{aligned} X_{i1} &\sim N_p(\mu_1, \Sigma) \quad i = 1, \ldots, n_1 \\ X_{i2} &\sim N_p(\mu_2, \Sigma) \quad i = 1, \ldots, n_2 \end{aligned}$$

where all variables are independent.

The following questions are of interest:

1. Are the profiles similar in the sense of being parallel (which means no interaction between the treatments and the groups)?
2. If the profiles are parallel, are they at the same level?
3. If the profiles are parallel, is there any treatment effect, *i.e.*, are the profiles horizontal (profiles keep the same no matter which treatment received)?

Now we consider these three scenarios as following.

Parallel Profiles Let C be a $(p-1)\times p$ matrix defined as

$$C = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & -1 \end{pmatrix}.$$

The hypothesis to be tested is

$$H_{01} : C(\mu_1 - \mu_2) = 0.$$

From the above corollary we know under H_0:

$$\frac{n_1 n_2}{(n_1+n_2)^2}(n_1+n_2-2)\{C(\overline{x}_1-\overline{x}_2)\}^T(CSC^T)^{-1}C(\overline{x}_1-\overline{x}_2) \sim T^2(p-1, n_1+n_2-2),$$

where $S = (n1S_1 + n_2S_2)/(n_1+n_2)$. The hypothesis is rejected if

$$\frac{n_1n_2(n_1+n_2-p)}{(n_1+n_2)^2(p-1)}(C\overline{x})^T(CSC^T)^{-1}C\overline{x} > F_{1-\alpha;p-1,n_1+n_2-p}.$$

Equality of Two Levels The question of equality of the two levels is meaningful only if the two profiles are parallel. In the case of interactions (rejection of H_{01}, the two populations react differently to the treatments and the question of the level has no meaning.

The equality of the two levels can be formalized as

$$H_{02} : 1_p^T(\mu_1 - \mu_2) = 0,$$

since

$$1_p^T(\overline{x}_1-\overline{x}_2) \sim N_1\left(1_p^T(\mu_1-\mu_2), \frac{n_1+n_2}{n_1n_2}1_p^T\Sigma 1_p\right)$$

and

$$(n_1+n_2)1_p^TS1_p \sim W_1(1_p^T\Sigma 1_p, n_1+n_2-2).$$

Using the above corollary we have

$$\begin{aligned} \frac{n_1n_2}{(n_1+n_2)^2}(n_1+n_2-2)\frac{\{1_p^T(\overline{x}_1-\overline{x}_2)\}^2}{1_p^TS1_p} &\sim T^2(1, n_1+n_2-2) \\ &= F_{1,n_1+n_2-2}. \end{aligned}$$

The rejection region is

$$\frac{n_1n_2(n_1+n_2-2)}{(n_1+n_2)^2}\frac{\{1_p^T(\overline{x}_1-\overline{x}_2)\}^2}{1_p^TS1_p} > F_{1-\alpha;1,n_1+n_2-2}.$$

Treatment Effect If it is rejected that the profiles are parallel, then two independent analyses should be done on the two groups using the repeated measurement approach. But if it is accepted that they are parallel, then we can exploit the information contained in both groups (eventually at different levels) to test a treatment effect, *i.e.*, if the two profiles are horizontal. This may be written as:

$$H_{03} : C(\mu_1 + \mu_2) = 0.$$

Consider the average profile $\overline{x}$:

$$\overline{x} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}.$$

Clearly,

$$\overline{x} \sim N_p\left(\frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}, \frac{\Sigma}{n_1 + n_2}\right).$$

With H_{03} with H_{01} implies that

$$C\left(\frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}\right) = 0.$$

So under parallel, horizontal profiles we have

$$\sqrt{n_1 + n_2}C\overline{x} \sim N_p(0, C\Sigma C^T).$$

From the above corollary, we again obtain

$$(n_1 + n_2 - 2)(C\overline{x})^T(CSC^T)^{-1}C\overline{x} \sim T^2(p - 1, n_1 + n_2 - 2).$$

This leads to the rejection region of H_{03} as

$$\frac{n_1 + n_2 - p}{p - 1}(C\overline{x})^T(CSC^T)^{-1}C\overline{x} > F_{1-\alpha;p-1,n_1+n_2-p}.$$

The data of this exercise contains two categories with group 1 contains 37 data, group 2 contains 12 data both with 4 sub-tests. Now using the data to calculate the test statistics for those three profiles analysis and justify which will be accepted or rejected.

- Solution:

(a) The test statistic for testing if the two profiles are parallel is $F = 0.4634$ which is not higher than 2.8115. Thus it is accepted

(b) The second test statistic (testing the equality of the levels of the 2 profiles) is $F = 17.2146$, which is higher than 4.0471. We reject the hypothesis.

(c) The final test (testing the horizontality of the average profile) has the test statistic $F = 53.3170$ which is higher than 2.8115. This implies that there are substantial differences among the means of the different sub-tests.