METHODS AND THEIR APPLICATIONS FOR COMPARISON OF
TWO HIGHLY MULTIVARIATE SMALL SAMPLES

Shunichi TANI, Mitsugu KANBA and Masashi GOTO

SHIONOGI Kaiseki Center

SHIONOGI & CO.,LTD.

22-41,1-chome,Izumicho,Suita City, Osaka, Japan

SUMMARY

In this paper, we consider the problem of comparison of two highly multivariate small samples, where the number of variables is large relative to size of these samples. We are sometimes faced with such a problem practically in medical, biochemical or biological fields, for instance on the comparison of blood test results, or on the comparison of weight of organs observed in the experiment of toxicity about some drugs, and so on. In clinical trial to compare the efficacy of two drugs, one of the two drugs is given to either of the two homogeneous disease groups and many characteristics of each case are observed. If there are many cases, we can apply usual multivariate methods, but if the number of cases is extremely smaller than the number of the characteristics, as in clinical trial of incurable disease, it is difficult to compare them such as to make the best use of multivariate characteristics.

For such a data, we cannot apply classical Hotelling's T^2 test. Ordinarily, we compare the two samples under selection of some variables or reduction on all variables. However, in such a small characteristics, relation between the samples is often obscure. Here, we propose a modified test procedure based on Dempster's method and examine its

relative features to other test procedures, applying these procedures to three kinds of data of multivariate small samples.

KEYWORDS: *Highly multivariate small sample; Hotelling's T^2 ; Stepwise MANOVA*

1. Introduction

As for some methods to compare two highly multivariate small samples, we can give approximate F test and its more precise test based on similar theory to Hotelling's T^2 test (cf. Dempster (1958,1960), Goto (1973), Kanba, et al. (1984), Tani, et al. (1984)). These test procedures assume homogeneity of within population variance-covariance matrix and multivariate normality. The degrees of freedom of F-statistic are estimated based on the vectors to orthogonalize the data matrix. In this paper we propose use of a convenient fixed orthogonal matrix. We evaluate and compare the feature of this procedure with those of alternatives, through applying these procedures to a example in published literature and two practical examples. As for alternatives, we consider randomization test based on permutation to assign F-statistic an unique significance probability. Since the F-statistic is constructed based on decomposition of variation due to the sample, scaling of variables affects the F-value. Incidentally in carrying out above procedures, we usually standardize observations on each variable by its range or standard deviation. We can also consider nonparametric randomization test where the data is replaced with their ranks on each variable (cf. Chung & Fraser (1958)). However, this randomization procedure has difficulty in numerical computation as sample size is getting large. Further, we consider the application of stepwise MANOVA (Multivariate ANalysis Of VAriance) based on principal variables (cf.

Dempster(1963a,b)). This procedure uniquely determines the significance probability and has little difficulty in numerical computation. However, it has disadvantage in difficulty of interpretation of the principal variables.

All test procedures mentioned above other than the stepwise MANOVA give overall tests on all variables. They do not evaluate which variables contribute to the separation of groups. The stepwise MANOVA is the procedure which tests separately several criterions based on the principal variables. Therefore, we do not require previously an arbitrary ordering of the variables. It has been recommended to use the stepwise MANOVA as a way to get a priori information which produces more sensitive test result, particularly when variable selection is our concern. Further, we examine influence of each variable for the test procedures mentioned above on the separation of samples.

2. Derivation of test procedures based on F-statistic

Let $X = (x_{ij})$ represent $n \times p$ data matrix, and x_i represent the i -th row of X , and the size of the first sample $\{x_i, i=1, \dots, n_1\}$ and the second sample $\{x_i, i=n_1+1, \dots, n_1+n_2\}$ are n_1 and n_2 , respectively, where $n=n_1+n_2$. We assume that each population has multivariate normal distribution with mean vector μ_k ($k=1,2$) and common variance-covariance matrix Σ . All of these parameters are unknown. We are interested in the hypothesis

$$H_0 : \mu_1 = \mu_2.$$

Now, n vectors of observations are transformed into $Y=AX$ by an orthogonal matrix A to yield n new orthogonal vectors with single d.f.. One vector corresponds to the overall mean, another to the difference between sample means, and the remainder to within-sample variation. A_1 ,

the first row of A corresponding to the overall mean is given by

$$A_1/\sqrt{n}=(1/n, 1/n, \dots, 1/n)$$

and A_2 , the second row of A corresponding to difference between sample means is given by

$$A_2/\sqrt{1/n_1+1/n_2}=(1/n_1, \dots, 1/n_1, -1/n_2, \dots, -1/n_2).$$

The remaining $(n-2)$ rows are arbitrarily constructed to satisfy the orthogonality of A (usually constructed with the aid of random numbers from normal distribution). Let $Y_1, Y_2, \dots, Y_n$ be the rows of $Y=AX$. Their means and variances are

$$E(Y_1)=(n_1\mu_1+n_2\mu_2)/\sqrt{n}$$

$$E(Y_2)=(\mu_1-\mu_2)/\sqrt{1/n_1+1/n_2}$$

$$E(Y_i)=0 \quad (3 \leq i \leq n)$$

$$V(Y_i)=\Sigma \quad (1 \leq i \leq n).$$

Y_2 which represents the difference between sample means has the same variance with that of Y_i ($3 \leq i \leq n$) and the mean of Y_2 is shifted only a multiple of $\mu_1-\mu_2$ from the mean 0 of Y_i ($3 \leq i \leq n$). Consequently, taking use of the statistic

$$F=(n-2)Q_2/(Q_3+\dots+Q_n),$$

we can see whether Y_2 be larger than $Y_3, \dots, Y_n$ in the sense of each squared length, where Q_i ($i=2, \dots, n$) is the squared length of Y_i ,

$$Q_i=Y_i Y_i^T.$$

Then under the hypothesis H_0 , F has apporoximately F -distribution on r and $(n-2)r$ degrees of freedom, where r may be thought of as a reduced dimensionality from an ideal dimensionality p when Σ is a unit matrix. Hence, we need to estimate r . Rough estimation procedure and more precise procedure of r are proposed.

(1) Approximate F test (M_1)

Setting $\hat{r}_1$ to be an estimate of r , we solve

$$t = \{ 1/\hat{r}_1 + (n-1)/3(n-2)\hat{r}_1^2 \} (n-3)$$

about $\hat{r}_1$, where

$$t = (n-2) \ln \{ 1/(n-2)(Q_3 + \dots + Q_n) \} - \sum_{i=3}^n \ln Q_i.$$

(2) More precise F test (M_2)

We make use of the angles among vectors $Y_3, \dots, Y_n$. Let θ_{ij} be the angle between vectors Y_i and Y_j . Then $\hat{r}_2$, more precise estimate of r than $\hat{r}_1$ can be obtained as the solution of equation

$$t+u = \{ (1/\hat{r}_2) + (n-1)/3(n-2)\hat{r}_2^2 \} (n-3) + \{ (1/\hat{r}_2) + (3/2)\hat{r}_2^2 \} \binom{n-2}{2},$$

where t is given in a same manner as t defined in (1) and

$$u = -\sum_{3 \leq i < j}^n \ln(\sin^2 \theta_{ij}).$$

(3) Randomization test (M_3)

We compute the F-statistic for each of $\binom{n}{n_1}$ possible ways of dividing the n individuals into subsamples of size n_1 and n_2 , and rank these F-values in increasing order. Suppose that S is the rank corresponding to the true division of samples. Then, under the hypothesis H_0 , S is uniformly distributed over the integers 1 to $\binom{n}{n_1}$. If s is the rank of the F-value actually computed from observations, the significance probability is given by

$$1 - s / \binom{n}{n_1}.$$

(4) Nonparametric randomization test (M_4)

This is a randomization test which uses the ranks of observations on each variable.

(5) Modifications which use a fixed orthogonal matrix (M_1^*, M_2^*)

The orthogonal matrix used in the above mentioned transformation affects the estimators of degrees of freedom of F-statistic in M_1 and M_2 . Therefore, we will apply M_1 and M_2 with use of a convenient fixed orthogonal matrix. Denote these corresponding methods by M_1^* and M_2^* . This orthogonal matrix is given as follows except for normalization of every row vector to unit length.

1	1	1	...	1	1	1	1	...	1
n_2	n_2	n_2	...	n_2	$-n_1$	$-n_1$	$-n_1$	...	$-n_1$
-1	1	0	...	0	0	0	0	...	0
-1	-1	2	...	0	0	0	0	...	0
:	:	:	...	:	:	:	:	...	:
-1	-1	-1	...	$n_1 - 1$	0	0	0	...	0
0	0	0	...	0	-1	1	0	...	0
:	:	:	...	:	:	:	:	...	:
0	0	0	...	0	-1	-1	-1	...	$n_2 - 1$

3. Stepwise MANOVA based on principal variables

The stepwise MANOVA is a stepwise test procedure which performs analysis of variance of scores on principal variables obtained from principal component analysis (cf. Dempster(1963a,b)). An important feature of this method is that it is applicable when the number of variables is larger than the sample size. Suppose that there are g samples, and each sample has size n_k ($k=1, \dots, g$). Denote the between group dispersion matrix by S_B , and the within group dispersion matrix by S_W obtained from multivariate analysis of variance. Then, the test statistic is given by Wilks' Λ criterion

$$\Lambda = |S_Y| / |S_T|,$$

where $S_T = S_B + S_Y$ is the total dispersion matrix. This method is not applicable when $n - g < p$, where $n = n_1 + \dots + n_g$. So we can apply the following stepwise procedure to principal variables. Principal component analysis based on S_T (usually, correlation matrix of S_T is used) is carried out. Let the eigenvalues of S_T to be

$$d_{11t} \geq \dots \geq d_{rrt} > 0,$$

where $r \leq \min(p, n - g)$. We denote the matrix of eigenvectors by

$$V = (v_1, \dots, v_r),$$

where v_i is the eigenvector corresponding to the eigenvalue d_{iit} .

Further, we set

$$D_T = \text{diag}(d_{11T}, \dots, d_{rrT})$$

$$D_Y = \text{Diag}(V^T S_Y V)$$

$$= \text{diag}(d_{11Y}, \dots, d_{rrY}).$$

and call

$$P_i = d_{i1Y} / d_{i1T} \quad (i = 1, 2, \dots, r)$$

stepwise statistic. Under the null hypothesis that g mean vectors are equivalent, statistic

$$F_i = \{ (n - g - i + 1) / (g - 1) \} \{ (1 - P_i) / P_i \}$$

has F-distribution with degrees of freedom $(g - 1, n - g - i + 1)$.

4. Illustration

We apply above mentioned six test procedures to three examples with small samples from a published literature and practical fields. Further, we examine the effect of each variable on the separation of two groups by deletion of the corresponding variable. Incidentally we perform t-test on each variable.

4.1 Metabolic characteristic data on alcoholics

This is an example cited from Beerstecher et al. (1950). The subject of this data is normal controls ($n_1=8$) and alcoholics ($n_2=4$). As for observed characteristics, 62 biochemical variables are measured. These are blood serum (8 items, 1 ~ 8), urine (30 items, 9 ~ 38), saliva (16 items, 39 ~ 54), taste thresholds (5 items, 55 ~ 59), phagocytic indices (3 items, 60 ~ 62), where the number of items and item no. are presented in the parentheses. We did log transformation on each item of taste thresholds, and then standardized every item by its range. Following is the result of the significance tests for the separation of normal controls and alcoholics by 62 items.

$$M_1; F=1.381, \text{ estimated d.f.}=(27.8, 278.0), \quad P=0.102$$

$$M_1^*; F=1.381, \text{ estimated d.f.}=(39.7, 396.7), \quad P=0.068$$

$$M_2; F=1.381, \text{ estimated d.f.}=(29.0, 290.1), \quad P=0.097$$

$$M_2^*; F=1.381, \text{ estimated d.f.}=(24.6, 245.7), \quad P=0.114$$

$$M_3; P=0.111, \quad M_4; P=0.081$$

Though this result indicated that there were not so large differences of significance probabilities among test procedures, significance probability of M_1^* was minimum. It seems that this is resulted from the unstability of approximate F-test. Degrees of freedom of F-statistic in M_2 and M_2^* were estimated at 29.0 and 24.6, respectively. In other words, dimensionality of this data decreased from 62 to about 24. We examined the effect of each item on the separation of two groups. In table 1, for each test procedure except for the t-test we give ten numbers of items which show great increase of significance probabilities by deletion of them. With respect to the t-test, each number corresponds to the item with which small significance probability is obtained.

Table 1. Ten items having large effect on separation of two groups

rank	M_1	M_1^*	M_2	M_2^*	M_3	M_4	t	rank	M_1	M_1^*	M_2	M_2^*	M_3	M_4	t
1	31	31	31	31	31	31	31	6	8	16	16	16	16	37	41
2	39	39	39	39	39	38	39	7	35	8	8	8	35	41	35
3	38	38	38	38	38	35	38	8	27	4	4	35	27	27	4
4	41	41	41	37	41	39	37	9	16	60	60	29	29	16	32
5	37	37	37	41	8	4	16	10	4	35	35	27	60	8	27

Item 31 had shown the largest effect in all methods. The three items with the highest ranks are agreed in M_1^*, M_2, M_2^*, M_3 . This result is similar to the result of the t-test only except for the exchange of order of items in M_1 and M_4 . However, as for comparison of M_2 and M_2^* , 7 items with the lowest ranks are perfectly agreed with each other except for rank 4 and 5. While item 37 has rank 4 in the t-test, it cannot be seen among ten items with the highest ranks in M_3 and M_4 . Though item 8 is not appeared in the t-test, it seems to be influential in multivariate consideration except for M_1 .

Next, we applied the stepwise MANOVA to this data. Significance probability for each principal component is given in table 2.

Table 2. Result of the stepwise MANOVA

component	1	2	3	4	5
probability	0.162	0.162	0.198	0.342	0.516
component	6	7	8	9	10
probability	0.383	0.034	0.224	0.836	0.559

The 7-th and the 1-st principal components play the important roles in the separation of two groups. Items which have large factor loadings for the 7-th principal component are item 28, 38, 2. For the 1-st

principal component, items 39, 33, 25 have large factor loadings. Note that items 38 and 39 have also shown large effect in previous investigation. We could not discriminate clearly two groups by the usual plot of scores on the axes of the 1-st and the 2-nd principal components. If we use the 1-st and the 7-th principal components, normal controls are located in the direction where the sum of scores of both principal components is large, and alcoholics is located in the direction where the sum of scores of both principal components is small (see figure 1).

Fig.1 Plot of scores on axes of the 1-st and the 7-th principal components (example (1)).

4.2 Experimental data on subacute toxicity of antifungal drug A

For six female rats on each dosage of control (zero dosage), 25, 60, 160 and 400 (mg/kg), sixteen biochemical ingredients of blood are measured (Tasaki & Goto (1984)). They are 1.TP, 2.ALB, 3.A/G, 4.T-CHO, 5.GOT, 6.GPT, 7.ALP, 8.GLV, 9.BUN, 10.CRE, 11.T-BIL, 12.LDH, 13.Ca,

14.Na, 15.K, 16.Cl. We performed the same analysis as 4.1 using the test procedures mentioned above about two dosage groups which are control and moderate 60 mg/kg dosage groups. Result of each test procedure is following.

$$M_1; P=0.029 (F=2.254, \hat{r}_1=8.2) \quad M_1^*; P=0.059 (F=2.254, \hat{r}_1=5.3)$$

$$M_2; P=0.035 (F=2.254, \hat{r}_2=7.7) \quad M_2^*; P=0.046 (F=2.254, \hat{r}_2=6.4)$$

$$M_3; P=0.012$$

$$M_4; P=0.026$$

In figure 2, we show the plot of significance probabilities obtained from deletion of one variable for each test procedure. As for the t-test, we plot values of $-\log_{10}P$ for convenience. From this figure, we find that item 5 (GOT), item 8 (GLU), item 10 (CRE) contribute to the separation of control and 60 mg/kg dosage groups. M_4 shows more extreme change at items 5 and 8 than other methods.

Significance probabilities for ten principal components are given in table 3 with the result of the stepwise MANOVA on this data.

Table 3. Result of the stepwise MANOVA

component	1	2	3	4	5
probability	0.179	0.163	0.137	0.006	0.484
component	6	7	8	9	10
probability	0.088	0.401	0.705	0.488	0.555

From the result, we can observe that the 4-th principal component contributes to the separation of two groups. Cumulative contribution ratio up to the 4-th principal component was 82 percent. The largest factor loading for the 4-th principal component was item 8 (GLU). Incidentally, relatively (secondly and thirdly) large factor loadings for the 6-th principal component were items 5 and 8. While item 10 was the largest factor loading for the 3-rd principal component. We plotted

scores on the axes of the 4-th and the 6-th principal components. The result was that 60 mg/kg dosage group was located in the 4-th quadrant and control group was located in the 2-nd quadrant.

Fig.2 Effect of each variable on the separation of two groups (example (2)).

4.3 Data on concentration of drug H and drug P in blood

For normal adult males, drugs H and P are injected. Then, concentration in blood at 14 points of time after injection are observed (Uragari et al. 1982). Sample size is four for each group. Usually, a certain compartment analysis is carried out for each individual.

Here, taking into consideration that this is a comparison of drugs H and P with small samples, we applied the test procedures above mentioned to this data. Results of the significance tests for the separation of two drugs are as follows.

$$M_1; P=0.057 (F=2.292, \hat{r}_1=6.0) \quad M_1^*; P=0.102 (F=2.292, \hat{r}_1=3.4)$$

$$M_2; P=0.048 (F=2.292, \hat{r}_2=6.8) \quad M_2^*; P=0.058 (F=2.292, \hat{r}_2=5.9)$$

$$M_3; P=0.029 \quad M_4; P=0.057$$

All of these significance were near about 0.05 except for M_1^* (0.102). Examining the effect of each variable on separation of two groups, large changes were recognized at the 6-th and the 10-th time points. The result of a certain compartment analysis of each individual shows that the 6-th time point is corresponding to the peak (after one hour) of concentration in blood where drug H shows high concentration. While the 10-th time point is corresponding to half-life period (after three hours) of concentration in blood and again concentration of drug H is high at this period. Therefore, it is suggested that concentration of drug H is higher than drug P in average (particularly at peak and half-life period).

In the result of the stepwise MANOVA, significance probability for the 1-st principal component was 0.051. Its largest factor loading was given at the 8-th time point which corresponded to the middle time point between peak and half-life period. In the plot of scores on axes of the 1-st and the 2-nd principal components, drug H was located in the 1-st and the 4-th quadrants, and drug P was located in the 2-nd and the 3-rd quadrants (see figure 3). After all, two drugs were well separated by the 1-st principal components only.

Fig.3 Plot of scores on axes of the 1-st and the 2-nd principal components (example (3)).

5. Concluding remarks

In this paper, we considered the test procedures based on F-statistic and the stepwise MANOVA to compare two highly multivariate small samples. We evaluated these test procedures through applying them to examples from a published literature and two practical fields. We compared M_1 and M_2 based on decomposition of variation due to sample and then compared M_1^* and M_2^* which used a fixed orthogonal matrix. It was indicated that M_2 gave more stable result than M_1 . However, as for M_1 and M_2 , estimates of degrees of freedom of F-statistic depend on the orthogonal matrix. For this reason, significance probabilities are not determined uniquely. While as for M_3 and M_4 based on permutation of F-statistic, significance probabilities are obtained uniquely, their numerical computation are getting difficult according with the increase

of the sample size. Hence, it is desirable to use test procedures which give unique significance probabilities and have little difficulty in numerical computation. As for M_2^* , a problem is kept open to check whether it has deviation of test. As for randomization test, it needs to investigate how to use the subgroup of permutation even in intermediate sample size. In contrast, the stepwise MONOVA based on principal variables gives unique significance probability for each principal variable and has little difficulty in numerical computation. A substantial difficulty in applying this method is interpretation of the principal variables. If they are interpretable, we can verify the relative configuration of two groups by the graphical representation with the plot of scores on the meaningful principal axes. Further, we can evaluate the contribution of each variable to the separation of two groups from the significance probability of each principal component. When it is difficult to interpret the principal components, it will be recommended to use this procedure as the supplement of overall test procedures. It can be used as the way to incorporate a priori information for variable selection.

REFERENCES

- Beerstecher, E., Sutton, H.E., Berry, H.K., Brown, W.D., Reed, J., Rich, G.B., Berry, L.J. & Williams, R.J. (1950). Biochemical individuality. V. Explorations with respect to the metabolic pattern of compulsive drinkers. *Archives of Biochemistry*, 29, 27-39.
- Chung, J.H. & Fraser, D.A.S. (1958). Randomization test for a multivariate two-sample problem. *J. Amer. Statist. Assoc.*, 53, 729-735.
- Dempster, A.P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.*, 29(4), 995-1010.
- Dempster, A.P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, 16, 41-50.
- Dempster, A.P. (1963a). Multivariate theory for general stepwise methods. *Ann. Math. Statist.*, 34, 873-883.

- Dempster, A.P. (1963b). Stepwise multivariate analysis of variance based on principal variables. *Biometrics*, 19,478-490.
- Goto, M. (1973). *Analysis of Multivariate Data*. Kagakujoho-sya, 173-176(in japanese).
- Kanba, M., Tani, S. & Goto, M. (1984). On comparison of two highly multivariate small samples. Abstract of the 14th conference of Japanese Society of Quality Control, 49-52 (in Japanese).
- Tani, S., Kanba, M. & Goto, M. (1984). On comparison of two highly multivariate small samples. Presentation paper at the Symposium of Informatics & Cybernetics in Fukuoka(1984.12.7).