Multivariate Statistical Analysis Final Exam
January 13, 2011
3 pages, 10 problems, 100 points

1. Let $\mathbf{X}_{1j}$ .and $\mathbf{X}_{2j}$ denote the responses to treatments 1 and 2, respectively, with $j = 1, 2, \cdots, n$ in a paired comparison experiment. Assume that the n independent observed differences $\mathbf{D}_j = \mathbf{X}_{1j} - \mathbf{X}_{2j}$ can be regarded as a random sample from a multivariate normal population. From a certain experiment of 30 observations, we found that the sample average and the sample covariance matrix of the observed differences are $\bar{\mathbf{d}} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ and $\mathbf{S}_d = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$, respectively.

   Test the hypothesis that there is no treatment difference at 5% significance level and compute the 95% Bonferroni simultaneous confidence interval for the individual mean difference. You may apply $\mathbf{S}_d^{-1} = \frac{1}{162}\begin{bmatrix} 14 & 4 & -2 \\ 4 & 14 & 2 \\ -2 & 2 & 17 \end{bmatrix}$ (10 %)

2. Two samples were taken from an experiment where two characteristics X_1 and X_2 were measured. Assume that samples 1 and 2 are taken from $N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, respectively. The summary statistics of the two samples for the observations are $\bar{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$, $n_1 = 16$ and $\bar{\mathbf{x}}_2 = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}$, $n_2 = 16$, respectively. Let the pooled sample covariance matrix be $\mathbf{S}_{pooled} = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$. Test the hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ .versus $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ at 5% significance level and find the 95% Bonferroni simultaneous confidence interval for the individual mean difference. (10%)

3. Twelve subjects are divided into 3 groups to rate the performances of three types of cell phones, A, B, and C, respectively, in a 1 to 10 scale. The results are given as the following table

Subject\System	A	B	C
1	7	6	5
2	5	4	5
3	6	4	5
4	4	5	6

Test the hypothesis that all three systems are with equal performance at 5% significance level, using ANOVA. (10%)

4. Consider the following independent samples. Population 1: $\begin{bmatrix}6\\7\end{bmatrix},\begin{bmatrix}5\\9\end{bmatrix},\begin{bmatrix}8\\6\end{bmatrix},\begin{bmatrix}4\\9\end{bmatrix},\begin{bmatrix}7\\9\end{bmatrix}$; Population 2: $\begin{bmatrix}3\\3\end{bmatrix},\begin{bmatrix}1\\6\end{bmatrix},\begin{bmatrix}2\\3\end{bmatrix}$; Population 3: $\begin{bmatrix}2\\3\end{bmatrix},\begin{bmatrix}5\\1\end{bmatrix},\begin{bmatrix}3\\1\end{bmatrix},\begin{bmatrix}2\\3\end{bmatrix}$. Assume that all three samples are from multivariate normal distributions with identical population covariance matrices. Are the mean vectors of those three populations the same at 1% significance level? (10%)

5. Calculate the least square estimates $\hat{\boldsymbol{\beta}}_{(1)}$, $\hat{\boldsymbol{\beta}}_{(2)}$ the residuals $\hat{\boldsymbol{\varepsilon}}_{(1)}$, $\hat{\boldsymbol{\varepsilon}}_{(2)}$ and the residual sum of squares for two straight line model $Y_1 = \beta_{01} + \beta_{11} z$, $Y_2 = \beta_{02} + \beta_{12} z$ fit to the data given in the table below. (10%)

z	-2	-1	0	1	2
y_1	4	2	3	1	-1
y_2	2	3	5	7	8

6. Use the results of Problem 5 to construct a 90% confidence region for $E([Y_1, Y_2] \mid z = -0.5)$ and a 90% prediction region for a new observation $[Y_1, Y_2]$ when $z = -0.5$. (10%)

7. Suppose 200 observations of random variables X_1, X_2, X_3.are with the sample covariance matrix $\mathbf{S} = \begin{bmatrix} 8 & 0 & -2 \\ 0 & 6 & 0 \\ -2 & 0 & 5 \end{bmatrix}$. Find the first sample principal component Y_1, its proportion of total variance, and the 95% confidence interval for the first

eigenvalue. (10%)

8. Assume an $m = 1$ orthogonal factor model. Estimate the loading matrix $\tilde{\mathbf{L}}$ and matrix of specific variance $\tilde{\mathbf{\Psi}}$ for the covariance matrix of Problem 7 using the principal component solution method. Compute also the residual matrix.. (10%)

9. The correlation matrix for chicken bone measurements (see Example 9.14) is

$$\begin{bmatrix} 1.000 & & & & & \\ .505 & 1.000 & & & & \\ .569 & .422 & 1.000 & & & \\ .602 & .467 & .926 & 1.000 & & \\ .621 & .482 & .877 & .874 & 1.000 & \\ .603 & .450 & .878 & .894 & .937 & 1.000 \end{bmatrix}$$

The following estimated factor loadings are extracted by the maximum likelihood procedure

Variable	Estimated factor loadings F_1	F_2	Varimax rotated estimated factor loadings F_1^*	F_2^*
1. Skull length	.602	.200	.484	.411
2. Skull breadth	.467	.154	.375	.319
3. Femur length	.926	.143	.603	.717
4. Tibia length	1.000	.000	.519	.855
5. Humerus length	.874	.476	.861	.499
6. Ulna length	.894	.327	.744	.594

Using the unrotated estimated factor loadings, obtain the maximum likelihood estimates of the communalities of the skull length, and the proportion of variance explained by the first factor. (10%)

10. Which part of this course is the most useful/interesting to you? Why? Give your suggestions to improve this course. (10%)