Multivariate Statistical Analysis Final Exam
January 14, 2010
2 pages, 10 problems, 100 points

1. Bars of soap are manufactured in each of two ways. The characteristics X_1 = lather and X_2 = mildness are measured. The summary statistics for 61 and 62 bars produced by methods 1 and 2, respectively, are $\bar{\mathbf{x}}_1 = [8 \quad 4]'$, $\bar{\mathbf{x}}_2 = [10 \quad 3]'$, $\mathbf{S}_1 = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix}$, . $\mathbf{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$. Test the hypothesis that the population means $\boldsymbol{\mu}_1$ = $\boldsymbol{\mu}_2$ at 10% significance level, using the spooled covariance matrix to estimate the population covariance matrix $\boldsymbol{\Sigma}$ and assuming that both populations are with multivariate normal distribution. (10 %)

2. Use Box's test to Test the hypothesis $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ .versus $H_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ at 10% significance level for Problem 1. (10%)

3. Four subjects rate the performances of three types of cell phones, A, B, and C, 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 for repeated measures. (10%)

4. Consider the following independent samples. Population 1: $\begin{bmatrix} 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \end{bmatrix}$; Population 2: $\begin{bmatrix} 7 \\ 3 \end{bmatrix}, \begin{bmatrix} 8 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 3 \end{bmatrix}$; Population 3: $\begin{bmatrix} 7 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 4 \end{bmatrix}$. 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 10% 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
y_1	4	2	3	1
y_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 all the variables $Y, Z_1, Z_2, \cdots, Z_r$ are random and have a joint distribution, not necessarily normal, with a $(r+1)\times 1$ mean vector $\boldsymbol{\mu}$ and a $(r+1)\times(r+1)$ covariance matrix $\boldsymbol{\Sigma}$. Partition $\boldsymbol{\mu}' = \left[\mu_Y \quad | \quad \boldsymbol{\mu}_{\mathbf{Z}}'\right]$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{YY} & | & \boldsymbol{\sigma}_{\mathbf{Z}Y}' \\ -- & + & -- \\ \boldsymbol{\sigma}_{\mathbf{Z}Y} & | & \boldsymbol{\Sigma}_{\mathbf{ZZ}} \end{bmatrix}$ with $\boldsymbol{\sigma}_{\mathbf{Z}Y}' = \left[\sigma_{YZ_1} \quad \sigma_{YZ_2} \quad \cdots \quad \sigma_{YZ_r}\right]$. Consider the problem of predicting Y using the linear predictor $b_0 + b_1 Z_1 + \cdots + b_r Z = b_0 + \mathbf{b}'\mathbf{Z}$. Find b_0 and $\mathbf{b}$ to minimize the mean square error $E(Y - b_0 - \mathbf{b}'\mathbf{Z})^2$. (10%)

8. Suppose 200 observations of random variables X_1, X_2, X_3.are with the sample covariance matrix $\mathbf{S} = \begin{bmatrix} 13 & 0 & 5 \\ 0 & 7 & 0 \\ 5 & 0 & 13 \end{bmatrix}$. Find the first sample principal component Y_1, its proportion of total variance, and the 95% confidence interval for the first eigenvalue. (10%)

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

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