INDIAN STATISTICAL INSTITUTE
CHENNAI CENTRE
M.Stat First Year
2015-16 Semester II

Multivariate Analysis
End-Sem Examination

Total Marks 100. Date: 4 May 2016 Duration: 3 hours

Comments : This paper carries 108 *marks. Answer as much as you can. Maximum you can score is* 100

1. (a) Consider the problem of classification into two multivariate normal populations with unequal mean vectors μ_1 and μ_2 and common positive definite dispersion matrix Σ. Assuming μ_1, μ_2 and Σ to be unknown, express the Bayes error i.e. the misclassification probability for the Bayes rule in terms of the standard normal distribution function, the prior probabilities (q_1 and q_2), the loss due to misclassification $C(1|2)$ and $C(2|1)$ and Δ the Mahalanobis distance between the two populations.

   (b) Suppose $q_1 = 0.6, \quad q_2 = 0.4$ and $\mu_1 = (2,4)', \quad \mu_2 = (6,8)'$, and $\Sigma = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$
   Compute the Bayes error for $C(1|2)$ and $C(2|1) = c$ [15+10=25]

2. (a) Describe the $k-$factor model and how the related parameters are estimated in Principal Factor Analysis.

   (b) If the $k-$factor model holds, show that it is scale invariant but the factor loadings may not be unique.

   (c) Suppose that $X' = (X_1, X_2)$ has a bivariate multinomial distribution with $n = 1$, so that $X_1 = 1$ with probability p and $X_2 = 1 - X_1$. Find the principal components of X and their variances. [10+8+7=25]

3. (a) Let $\mathbf{y}_{ij}$ be independently distributed as $N_p(\mu_i, \Sigma)$, where $i = 1, \ldots, g$ and $j = 1, \ldots, n_i$. Define $\mathbf{z}_{ij} = A\mathbf{y}_{ij} + \mathbf{b}$, where $\mathbf{b}$ is a fixed p-vector and A is $p \times p$ non-singular matrix. Show that the test criterion to test equality of several multivariate means is invariant under the above transformation, alternatively $\Lambda_y^* = \Lambda_z^*$ where Wilk's Λ, is given by $\Lambda^* = \frac{|SSE|}{|SSE+SSTr|}$

   (b) Consider the observations on two variables $\mathbf{x}_1$ and $\mathbf{x}_2$ displayed in the form of a two way table in 12 factor-level combinations of two factors, factor 1 with three levels and factor 2 with four levels with no replications. Obtain the matrices $SSP_{Corrected}$, $SSP_{Factor\ 1}$, $SSP_{Factor\ 2}$ and $SSP_{Residual}$ with appropriate degrees of freedom and summarize the calculations in a two way MANOVA table to test for factor effects.

		Factor 2				
		Level 1	Level 2	Level 3	Level 4	Average
Factor 1	Level 1	$\begin{bmatrix} 6 \\ 8 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 6 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 6 \end{bmatrix}$	$\begin{bmatrix} 5 \\ 8 \end{bmatrix}$
	Level 2	$\begin{bmatrix} 3 \\ 8 \end{bmatrix}$	$\begin{bmatrix} -3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 3 \end{bmatrix}$	$\begin{bmatrix} -4 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 4 \end{bmatrix}$
	Level 3	$\begin{bmatrix} -3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} -4 \\ -5 \end{bmatrix}$	$\begin{bmatrix} 3 \\ -3 \end{bmatrix}$	$\begin{bmatrix} -4 \\ -6 \end{bmatrix}$	$\begin{bmatrix} -2 \\ -3 \end{bmatrix}$
	Average	$\begin{bmatrix} 2 \\ 6 \end{bmatrix}$	$\begin{bmatrix} -1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 5 \\ 4 \end{bmatrix}$	$\begin{bmatrix} -2 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 3 \end{bmatrix}$

for $p = 2$ and $g \geq 2$, $\left(\frac{\sum n_l - g - 1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1),2(\sum n_l - g - 1)}$. Where p is the number of components in $\mathbf{x}$, g is the number of groups. $F_{4,16,0.95} = 3.007$, $F_{4,16,0.99} = 4.773$, $F_{6,16,0.95} = 2.741$, $F_{6,16,0.99} = 4.202$.

[15+20=35]

4. Let $\mathbf{x} \sim N_p(\mathbf{0}, \Sigma)$. Find the distribution of $\frac{\mathbf{1}'\Sigma^{-1}\mathbf{x}}{(\mathbf{1}'\Sigma^{-1}\mathbf{1})^{\frac{1}{2}}}$ [8]

5. Write short notes on

   (a) Clustering Methods and needs

   (b) Multi Dimensional Scaling [15]