# INDIAN STATISTICAL INSTITUTE
## CHENNAI CENTRE
## M.STAT First Year
## 2014-15 Semester I

Multivariate Analysis
Final Examination

*Total Marks 100.* Date: 29 April 2015 Duration: 3 hours

*Comments : This paper carries* 115 *marks. Answer as much as you can. Maximum you can score is* 100

1. (a) Consider the classification problem of an observation $\mathbf{x}_0$ into either of normal populations $\Pi_1$ or $\Pi_2$ with common variance-covariance matrix $\Sigma$ and $\mu_1$ and $\mu_2$ as their mean vectors respectively. Let $\delta = \mu_1 - \mu_2$ and $K$ be defined by $\frac{1}{2}\delta'\Sigma^{-1}(\mu_1 + \mu_2)$. Show that $(\delta'\Sigma^{-1}\mathbf{x}_0 - K)$ can be written as $\frac{1}{2}(D_2^2 - D_1^2)$, where $D_i^2 = (\mathbf{x}_0 - \mu_i)'\Sigma^{-1}(\mathbf{x}_0 - \mu_i), \quad i = 1, 2$

   (b) Suppose prior probabilities of the two normal distributions $\Pi_1$ and $\Pi_2$ are $q_1 = 0.6$ and $q_2 = 0.4$. $\mu_1 = (2, 4)$ and $\mu_2 = (6, 8)$ and the common dispersion matrix for both populations $\Sigma = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$. Compute the Bayes error for $C(1|2) = C(2|1) = c$. [5+10=15]

2. Consider $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \sim N_3(\mu, \Sigma)$. Compute $\mu$ and $\Sigma$ knowing that

$$Y|Z \sim N_1(-Z, 1)$$

$$\mu_{Z|Y} = -\frac{1}{3} - \frac{1}{3}Y \quad \text{and}$$

$$X|(Y, Z) \sim N_1(2 + 2Y + 3Z, 1)$$

   Determine the distribution of $X|Y$ and of $X|(Y + Z)$ [10+5+5=20]

3. Let $Y \sim N_n(\mu, \sigma^2 I_n)$ be an $n$-vector with a spherical normal distribution and $A$ be an $n \times n$ symmetric matrix. Then prove that $\frac{Y'AY}{\sigma^2}$ will have a $\chi_r^2(\lambda^2)$ distribution with non-centrality parameter $\lambda^2 = \frac{\mu' A \mu}{\sigma^2}$ if and only if $A$ is idempotent with rank$(A) = r$. [10]

4. Let $X = \begin{bmatrix} X^{(1)} & q \times 1 \\ X^{(2)} & p - q \times 1 \end{bmatrix}$ follow a $p-$variate normal distribution with mean vector $\mathbf{0}$ and dispersion matrix $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ with usual break-up as in $X$. Let $x_i$ be the $i$th component of $X^{(1)}$. Then show that

   (a) $x_i - \beta^{(i)} X^{(2)}$ and $X^{(2)}$ are independently distributed where $\beta^{(i)}$ is the $i$th row vector of $\Sigma_{12}\Sigma_{22}^{-1}$

(b) Among all linear functions of $X^{(2)}$, $E[x_i - \alpha' X^{(2)}]^2$ is minimum when $\alpha'(X^{(2)}) = \beta^{(i)} X^{(2)} = E[x_i | X^{(2)}]$

(c) The correlation coefficient between $x_i$ and $\alpha'(X^{(2)})$ is maximum when $\alpha'(X^{(2)}) = \beta^{(i)} X^{(2)}$ and that maximum correlation coefficient is the multiple correlation coefficient of $x_i$ on $X^{(2)}$ and is given by

$$\rho_{i.q+1\ldots p} = \sqrt{\frac{\sigma_{(i)} \Sigma_{22}^{-1} \sigma'_{(i)}}{\sigma_{ii}}}$$

where $\sigma_{ii} = Var(x_i)$ and $\sigma_{(i)}$ is the $i$th row of $\Sigma_{12}$. [5+10+10=25]

5. Let $X = [X_1, X_2, X_3, X_4]'$ be a random variable with covariance matrix $\Sigma$ given by

$$\begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

and $Y$ be the vector of principal components of $X$. Show that the variance explained by the last three principal components are the same. Also find the first principal component $Y_1$ and compute the variance explained by it. What proportion of the total variation is accounted for by $Y_1$? [10]

6. (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 $\Lambda$, is given by $\Lambda^* = \frac{|SSE|}{|SSE+SSTr|}$

(b) Consider the observations on two variables $x_1$ and $x_2$ 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 |
| | Level 1 | $\begin{bmatrix}6\\8\end{bmatrix}$ | $\begin{bmatrix}4\\6\end{bmatrix}$ | $\begin{bmatrix}8\\12\end{bmatrix}$ | $\begin{bmatrix}2\\6\end{bmatrix}$ | $\begin{bmatrix}5\\8\end{bmatrix}$ |
| Factor 1 | 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 | $\begin{bmatrix}2\\6\end{bmatrix}$ | $\begin{bmatrix}-1\\1\end{bmatrix}$ | $\begin{bmatrix}5\\4\end{bmatrix}$ | $\begin{bmatrix}-3\\-1\end{bmatrix}$ | $\begin{bmatrix}-\frac{1}{3}\\-3\end{bmatrix}$ |

[15+20=35]