KENYATTA UNIVERSITY

UNIVERSITY EXAMINATIONS 2011/2012

SECOND SEMESTER EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE

SMA 563: MULTIVARIATE ANALSYSI II

DATE: Wednesday 28th March 2012 TIME: 9.00a.m – 12.00p.m

INSTRUCTIONS

Answer question one and any other two questions

Question 1 (30marks)

a) Suppose $\underline{x}_2' = (x_2\ x_3 \ldots x_p)$ has the *p*-variate normal distribution. Then the conditional distribution of x_1 given $\underline{x}^1_2 = x_2, x_3, \ldots, x_p$ is normal with mean $\mu + \Sigma_{12}\ \ \Sigma_{22}^{-1}\ (\underline{x}_2 - \mu_{-2})$ and variance $\sigma_{1-}^{2}\ \ \Sigma_{22}^{-1}\ \Sigma_{21}$ where $\Sigma_{12} = cov\left(x_{1,}x_{-1}\right) = \Sigma_{21}^{1}$ and $\Sigma_{22} = var(\underline{x}_2)$.

i) Define the multiple correlation $R_{1\,v\,2\,3\ldots p}$ between x_1 and $\underline{x}_2$

[give and expression/formula] [2marks]

ii) Show that $R^2_{1,2\,3--p} = \Sigma_{12}\ \ \varepsilon_{22}^{-1}\ \ \Sigma_{21}\ \div \sigma_{11}^2$

[5marks]

iii) Given a triavariate normal density with variance-covariance matrix

$$\Sigma = \tfrac{1}{6}\begin{bmatrix} 4 & 2 & 1 \\ 2 & 7 & ½ \\ 1 & ½ & 7/4 \end{bmatrix}$$

Compute $R^2_{1.\ 2\,3}$ and $P^2_{13.2}$ [8marks]

b) Consider a p_{x1}random vector $\underline{x}$ such that $\underline{x} \sim N_p\ (\ \underline{\mu}\ ,\ \Sigma)$ where Σ is a symmetric positive definite matrix.

i) Define the principle components of $\underline{x}$ and show if $\underline{u}$ is a vector of principal components, then $u_{1,}\ u_{2,\ldots}\ u_p$ are independently distributed as normal random variables with var $(\underline{u}) = \ diag\ (\lambda_1\ , \lambda_2\ --\lambda_p\ )$ where $\lambda_1\ \geq \lambda_2\ \geq \cdots \geq \lambda_p$ are the characteristics roots of Σ. [6marks]

ii) Show that the generalized variance of $\underline{u}$ is equal to the generalized variance of $\underline{x}$ [3marks]

c) Consider the linear model $X\underline{\beta}\ + \underline{e}$ where $\underline{e}\ \sim \mu_n\ (\underline{o}\ ,\ \underline{\mathrm{I}_n})$ and rank $(X) = r\ < p$

i) When is a linear function $\underline{a}^1\ \underline{\beta}$ said to be estimable? [2marks]

ii) Define a "best' linear unbiased estimator (BLUE) of an estimable linear function $\underline{a}^1\ \underline{\beta}$ [2marks]

iii) For the following linear model $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \beta \\ 1 \\ \beta_1 \\ \ \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$

Determine X^1X [2marks]

Question 2 **(20marks)**

a) Let $f_1\ (\underline{x}\ )$ and $f_2\ (\underline{x}\ )$ be two p-variate normal probability density functions corresponding to two populations II_1 and II_2 respectively. The distributions have mean vectors $\underline{\mu}_1$ and $\underline{\mu}_2$ respectively and a common variance-covariance matrix Σ which is symmetric positive definite. Show that log $\left\{ f_1, (\bar{\underline{x}}) / f_2\ (\underline{x}) \right\} = \underline{a}^1\ \underline{x} - \frac{1}{2}\ \underline{a}^1\ (\underline{\mu}_1 + \underline{\mu}_2)$

Where $\underline{a} = \Sigma^{-1}\underline{\sigma}$ and $\underline{\sigma}\ \underline{\mu}_1 - \underline{\mu}_2$ [5marks]

b) A random sample of 49 men participating in study of aging where classified by psychiatric examination into one of two categories, <u>senile</u> or <u>non-senile</u>. An independently administered adult intelligence test revealed large differences between the two groups in certain standardized subsets of the test. The results for these sections of the text are given below

Subject	Senile (n_1 = 37)	Non-Senile (n_2 = 12)
x_1 Information	12.57	8.75
x_2 Similarities	9.57	3.35
x_3 Arithmetic	11.49	8.50
x_4 Picture completion	7.97	4.75

The within group covariance matrix inverse is 5^{-1} =

$$\begin{bmatrix} 02591 & -1358 & 0.588 & -0.0647 \\ & 0.1865 & 0.0383 & -0.0144 \\ & & 0.1510 & -0170 \\ & & & 0.2112 \end{bmatrix}$$

Determine the maximum likelihood discriminant function between the two groups and find the estimated errors of misclassification. [10marks]

c) Four examinations in reading speed, reading power, arithmetic speed arithmetic power were given to n = 148 children. The question of interest is whether reading ability is correlated with arithmetic ability. The correlations are given by.

$R_{11} = \begin{bmatrix} 1.0 & 0.6328 \\ & 1.0 \end{bmatrix}$, $R_{22} = \begin{bmatrix} 1.0 & 0.4248 \\ & 1.0 \end{bmatrix}$

$R_{12} = \begin{bmatrix} 0.2412 & 0.0586 \\ -0.0553 & 0.0655 \end{bmatrix}$

Find the estimated canonical correlations λ_1 and λ_2 [5marks]

Question 3 **(20 marks)**

A 2-way classification Multivariate Analysis of Variance (MANOVA) model is given by

$$\underline{y}_{ij} = \underline{\mu} + \underline{\alpha}_i + \underline{\beta}_j + \underline{e}_{ij} \;:\; i = 1,2,--,v$$
$$j = 1,2,--b$$

Where $\underline{\mu} = (\mu_1, \mu_{2,}, .., \mu_p)$ '

$\underline{\alpha}_j = (\alpha_{l1}, \alpha_{i2} ... \alpha_{i3})'$

$\underline{\beta}_j = (\underline{\beta}_{j1}, \underline{\beta}_{j2}, \underline{\beta}_{jp})'$

And $\underline{e}_{ij}$ is a *p*-variate normal vector with mean vector $\underline{o}$ and a dispersion matrix $\sigma^2 \ \mathrm{I}_p$

a) Let $\underline{a}$ be any non zero *p*-component vector of constants. Make unvariable transformation of the above model using the vector $\underline{a}$. Hence derive the least squares estimators of the parameters in the resulting univariate mode. [3marks]

b) Derive in terms of $\underline{a}$ and $\underline{y}_{lj}$ the sum of squares due to $\underline{\alpha}\ \ j\ s$, $\beta \underline{j\, s}$ ', and $\underline{e}_{ij}$'s. Hence from an Analysis of variance (ANOVA) tale for transformed model for testing equality of parameters. Indicate the test for hypothesis. [8marks]

c) By applying Roy's union-intersection principle on the critical region of the ANOVA test, form a MANOVA table for testing the null hypothesis of equality of the $\underline{e}_{ij}$'s. and equality of the β_{ij}'s. [9marks]

Question 4 **(20 marks)**

a) Let a *p*-dimensional random vector $\underline{x}$ be distributed as $Np\ (\underline{\mu},\ \Sigma)$, where Σ axis a symmetric positive definite (*spd)* matrix. Explain briefly how you would transform the vector $\underline{x}$ of correlated variables to a new set of uncorrelated variables $\underline{u}$ using principal axis rotation. [5marks]

b) Explain how principal component analysis can be used to reduce the dimensionality of the original vector. What are the possiblems in the interpretation of the principle components? [5marks]

c) Let p_{ij} denote the coefficient of correlation between the *i-th* component of $\underline{x}$ and the *j-th* principal component. Determine the expression factor p_{ij} and give its maximum likelihood estimator r_{ij}. Explain how r_{ij}'s can be used to assess the effect of discarding some components of $\underline{x}$ [10marks