STRATHMORE UNIVERSITY
STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES (SIMS)
MASTER OF SCIENCE IN STATISTICS
STA 8301 MULTIVARIATE STATISTICAL ANALYSIS

Date: Monday, 19th August 2019 TIME: 3 Hours

INSTRUCTION: Answer Question **1** and any other **Two** Questions.

Question 1 (20 Marks)

(a) Suppose $\underline{X} = x_1, x_2, ..., x_n$ are n independent observations and that $X \sim N(\mu, \sigma^2)$. Show that the joint distribution of $x_1, x_2, ..., x_n$ is the n-variate normal distribution. (4 Marks)

(b) Let $\underline{X} = (x_1, x_2, .., x_p)'$ be a p -dimensional column vector of random variables each which is distributed by $N_p \sim (\underline{\mu}, \Sigma)$, with $\Sigma > 0$, as shown below

$$f(\underline{X}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}Q\right] \quad -\infty < \underline{X} < \infty$$

Where, $Q = (\underline{X} - \underline{\mu})' \Sigma^{-1} (\underline{X} - \underline{\mu})$

Show that $f(\underline{X})$ is a probability density function (4 Marks)

(c) Consider a random vector $\underline{Y}$ that are independently distributed with a distribution defined by $N_3 \sim (\underline{\mu}, \Sigma)$

$$\text{where } \underline{\mu} = \begin{pmatrix} 2 \\ 1 \\ 4 \end{pmatrix}, \Sigma = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 4 \end{pmatrix}$$

(i) Find the distribution of Z= $y_1 - y_2 + 3y_3$ (4 Marks)

(ii) Determine the Marginal distribution of y_3 (2 Marks)

(d) Suppose $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ where and Σ is the dispersion matrix and that the density of $\underline{X}$ is defined by

$f(\underline{X}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[\frac{1}{2}\left(\underline{X}-\underline{\mu}\right)^{\prime}\Sigma^{-1}\left(\underline{X}-\underline{\mu}\right)\right] \quad -\infty < \underline{X} < \infty$. Suppose a random sample of size n is taken from this population. Determine the Maximum likelihood estimator (M.L.E) of the population mean vector $\underline{\mu}$. (6 Marks)

Question 2 (20 Marks)

(a) Explain two properties of covariance matrix for a multivariate distribution. (2 Marks)

(b) Suppose the sample covariance matrix is given by

$$S=\begin{pmatrix} 10 & 14 & 1 \\ 14 & 12 & 3 \\ 1 & 3 & 25 \end{pmatrix}$$

Determine the correlation matrix, R (6 Marks)

(c)A random sample of 12 households were investigated on three variables; income X_1, expenditure X_2, and savings X_3 in thousand Kenya shillings. The following is a summary of the results:

$$\bar{X}_1 = 28,\ \bar{X}_2 = 14 \text{ and } \bar{X}_3 = 3,\ S=\begin{pmatrix} 10 & 2 & 4 \\ 2 & 6 & 3 \\ 4 & 3 & 5 \end{pmatrix}$$

test the hypothesis $H_0:\underline{\mu}=\begin{pmatrix} 12 \\ 4 \\ 3 \end{pmatrix}$ against $H_a:\underline{\mu}\neq\begin{pmatrix} 12 \\ 4 \\ 3 \end{pmatrix}$ at $\alpha = 0.05$ (6 Marks)

(d) Two indices measuring students' behavior were considered by a researcher. The values of these indices for n=20 students at a university produced the following summary statistics

$$\bar{y} = \begin{pmatrix} 71.45 \\ 164.7 \end{pmatrix}$$

covariance matrix $S = \begin{pmatrix} 20 & 100 \\ 100 & 1000 \end{pmatrix}$, $\mu = \begin{pmatrix} 70 & 170 \end{pmatrix}^{\prime}$

All three indices are evaluated for each patient. Test for the equality of mean indices (6 Marks)

Question 3 (20 Marks)

Suppose we have the following population of four observations made on three random variables X_1, X_2, and X_3:

X_1	X_2	X_3
1.0	6.0	9.0
4.0	12.0	10.0
3.0	12.0	15.0
4.0	10.0	12.0

Suppose the derived correlation matrix is given by;

$$\rho = \begin{pmatrix} 1 & .83 & .35 \\ .83 & 1 & .62 \\ .35 & .62 & 1 \end{pmatrix}$$

and the corresponding eigen value-eigen vector pairs:

$$\lambda_1 = 2.2 \; e_1 = (0.58 \quad 0.63 \quad 0.51)'$$
$$\lambda_2 = 0.62 \; e_2 = (0.54 \quad -0.16 \quad 0.82)'$$
$$\lambda_3 = 0.16 \, , e_3 = (0.60 \quad -0.75 \quad 0.26)'$$

(a) Find the three population principal components variables Y_1, Y_2, and Y_3 for the standardized random variables Z_1, Z_2, and Z_3 (3 Marks)

(b) Determine the proportion of total population variance due to each principal component (8 Marks)

(c) Obtain the correlations between the original random variables X_i and the principal components Y_i. Hence interpret the correlation matrix. (9 Marks)

Question 4 (20 Marks)

(a) Explain the objective of canonical correlation analysis (2 Marks)

(b) The (2x1) random vectors $X^{(1)}$ and $X^{(2)}$ have the joint mean vector and joint covariance matrix given by

$$\underline{\mu} = \begin{pmatrix} \mu^{(1)} \\ ..... \\ \mu^{(2)} \end{pmatrix} = \begin{bmatrix} -3 \\ 2 \\ .... \\ 0 \\ 1 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & . & \Sigma_{12} \\ . & . & . \\ \Sigma_{21} & . & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 8 & 2 & . & 3 & 1 \\ 2 & 5 & . & -1 & -3 \\ . & . & . & . & . \\ 3 & -1 & . & 6 & -2 \\ 1 & 3 & . & -2 & 7 \end{bmatrix}$$

(i) Determine the first and second canonical correlations. Hence comment on your results (7 Marks)

(ii) Determine the Canonical variate pairs (U_1, V_1) and (U_2, V_2) (7 Marks)

(iii) Find the eigen vectors of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ (4 Marks)