MACHAKOS UNIVERSITY

University Examinations for 2020/2021

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR FIRST SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE (MATHEMATICS AND COMPUTER SCIENCE)

BACHELOR OF SCIENCE (STATISTICS AND PROGRAMMING)

BACHELOR OF SCIENCE (MATHEMATICS)

BACHELOR OF EDUCATION (SCIENCE)

SMA 360 MULTIVARIATE STATISTICAL METHODS I

DATE: 16/8/2021 **TIME: 2:00 – 4:00 PM**

INSTRUCTIONS

Answer Question ONE and ANY OTHER TWO

QUESTION ONE (COMPULSORY) (30 MARKS)

a) Define the multivariate normal density. (2 marks)

b) State three properties of the multivariate normal distribution. (3 marks)

c) Suppose $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and given that $\boldsymbol{Y_1} = \boldsymbol{X_1} - \boldsymbol{X_2}$ and $\boldsymbol{Y_2} = \boldsymbol{X_1} - \boldsymbol{X_3}$. Determine

i. $E(\boldsymbol{Y})$ (2 marks)

ii. $Var(\boldsymbol{Y})$ (4 marks)

d) Given two random variables X and Y with joint and marginal probabilities as below, determine the variance covariance matrix. (8 marks)

	0	1	P(x)
-1	0.24	0.06	0.3
0	0.16	0.14	0.3
1	0.40	0.00	0.4
P(y)	0.8	0.2	1

e) Given the covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{4} & \mathbf{1} & \mathbf{2} \\ \mathbf{1} & \mathbf{9} & \mathbf{-3} \\ \mathbf{2} & \mathbf{-3} & \mathbf{25} \end{bmatrix}$, determine the correlation matrix ρ. (5 marks)

f) Given $\boldsymbol{S} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$ and $\boldsymbol{S} = \begin{bmatrix} \mathbf{1} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \mathbf{1} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \mathbf{1} \end{bmatrix}$.

i. Calculate the total sample variance for each **S.** Compare the results. (2 marks)

ii. Calculate the generalised sample variance for each **S.** Compare the results. Comment on the discrepancies, if any, between i and ii. (4 marks)

QUESTION TWO (20 MARKS)

a) Consider the data matrix $\boldsymbol{X} = \begin{bmatrix} 9 & \mathbf{1} \\ \mathbf{5} & \mathbf{3} \\ \mathbf{1} & \mathbf{2} \end{bmatrix}$ with $n = 3$ observations on $p = 2$ variables X_1 and X_2. Form the linear combinations $\boldsymbol{c}'\boldsymbol{X} = [-\mathbf{1} \quad \mathbf{2}] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and $\boldsymbol{b}'\boldsymbol{X} = [\mathbf{2} \quad \mathbf{3}] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$**.** Evaluate the sample means, variances and covariance of $\boldsymbol{c}'\boldsymbol{X}$ and $\boldsymbol{b}'\boldsymbol{X}$ from first principles. (8 marks)

b) Let $\boldsymbol{X} \sim N_{\mathbf{3}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\mu' = [-3 \quad 1 \quad 4]$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{1} & \mathbf{-2} & \mathbf{0} \\ \mathbf{-2} & \mathbf{5} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{2} \end{bmatrix}$**.** Determine whether the following random variables are independent.

i. X_2 and X_3

ii. (X_1, X_2) and X_3 (4 marks)

c) State the multivariate version of the central limit theorem. (3 marks)

d) State the law of large numbers and briefly explain its practical importance. (5 marks)

QUESTION THREE (20 MARKS)

a) Suppose $X \sim N(\mu, \sigma^2)$**.** Find the characteristic function of a standard normal variable $z = \frac{x-\mu}{\sigma}$. (5 marks)

b) Let $\boldsymbol{X} \sim \boldsymbol{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Suppose $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$ where **A** is a qxp matrix of constants and **b** is a non-zero $qx1$ vector of constants. Determine the distribution of **Y**. (5 marks)

c) Given that $f(x_1, x_2)$ is the bivariate normal density, show that $f(x_1|x_2)$ is $N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{{\sigma_{12}}^2}{\sigma_{22}}\right)$. (10 marks)

QUESTION FOUR (20 MARKS)

a) Consider a bivariate normal distribution with $\mu_1 = 1$, $\mu_2 = 3$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ and $\rho_{12} = -0.8$.

i. Write out the bivariate normal density. (3 marks)

ii. Write out the squared statistical distance expression as a function of x_1 and x_2. (7 marks)

b) Let $\boldsymbol{X_1}, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X_4}, \boldsymbol{X_5}$ be independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

i. Determine the mean vector and covariance matrices for each of the two linear combinations of random vectors

$\frac{1}{5}\boldsymbol{X_1} + \frac{1}{5}\boldsymbol{X}_2 + \frac{1}{5}\boldsymbol{X}_3 + \frac{1}{5}\boldsymbol{X_4} + \frac{1}{5}\boldsymbol{X_5}$ and $\boldsymbol{X_1} - \boldsymbol{X}_2 + \boldsymbol{X}_3 - \boldsymbol{X_4} + \boldsymbol{X_5}$

in terms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. (8 marks)

ii. Determine the covariance between the two linear combinations of random vectors. (2 marks)

QUESTION FIVE (20 MARKS)

a) Given the data

z_1	10	5	7	19	11	8
y	15	9	3	25	7	13

Fit the linear regression model $Y_j = \beta_0 + \beta_1 z_{jt} + \varepsilon_j$, $j = 1,2,\ldots,6$. Specifically, calculate the least squares estimates $\widehat{\boldsymbol{\beta}}$, the fitted values $\widehat{\boldsymbol{y}}$, the residuals $\widehat{\boldsymbol{\varepsilon}}$ and the residual sum of squares $\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}$. (10 marks)

b) Let $\boldsymbol{X_1}, \boldsymbol{X}_2, \ldots, \boldsymbol{X_n}$ be a random sample from a joint distribution that has mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Prove that $\overline{\boldsymbol{X}}$ is an unbiased estimator of $\boldsymbol{\mu}$ and its covariance matrix is $\frac{1}{n}\boldsymbol{\Sigma}$. (10 marks)