NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FACULTY OF HEALTH AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

QUALIFICATION: Bachelor of Science Honours in Applied Statistics	
QUALIFICATION CODE: 08BSSH	**LEVEL:** 8
COURSE CODE: MVA802S	**COURSE NAME:** MULTIVARIATE ANALYSIS
SESSION: NOVEMBER 2019	**PAPER:** THEORY
DURATION: 3 HOURS	**MARKS:** 100

FIRST OPPORTUNITY EXAMINATION QUESTION PAPER	
EXAMINER	Dr D. B. GEMECHU
MODERATOR:	PROF. S. SUSUMAN

INSTRUCTIONS

1. Answer ALL the questions in the booklet provided.
2. Show clearly all the steps used in the calculations.
3. All written work must be done in blue or black ink and sketches must be done in pencil.

PERMISSIBLE MATERIALS

1. Non-programmable calculator without a cover.

ATTACHMENTS

1. Statistical tables (F-table).

THIS QUESTION PAPER CONSISTS OF 4 PAGES (Including this front page)

Question 1 [8 Marks]

1.1. State three features (properties) of Multivariate normal distribution. **[3]**

1.2. Briefly explain Principal Components Analysis (PCA) and state three assumptions of PCA. **[2+3]**

Question 2 [13 Marks]

2. In a psychological experiment, three subjects are tested on how long they take to perform f tasks. The variables y_1, y_2 and y_3 are the times in minutes for the three tasks. The results obtained are listed below:

Individual	Task 1	Task 2	Task 3
1	26	12	32
2	25	10	26
3	30	14	35

Then, using the matrices, calculate:

2.1. the sample mean vector $\bar{\boldsymbol{y}}$. **[3]**

2.2. the sample variance-covariance matrix, $\boldsymbol{S}$. **[5]**

2.3. the sample correlation matrix, $\boldsymbol{R}$, in terms of $\mathbf{DSD}$, clearly defining your matrix $\mathbf{D}$ and interpret your result. **[5]**

Question 3 [10 Marks]

3. If $\boldsymbol{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{b}$ is a $p \times 1$ vector of constant and $\boldsymbol{A}$ is a constant $q \times p$ matrix of rank q, where $q \leq p$, then show that $\boldsymbol{Ay} - \boldsymbol{b} \sim N_p(\boldsymbol{A\mu} - \boldsymbol{b}, \mathbf{A\Sigma A'})$. **Hint**: Use the uniqueness property of joint moment generating function.

Question 4 [9 Marks]

4. In an investigation of adult intelligence, scores were obtained on two tests "verbal" and "performance" for **20** randomly selected subjects aged 60 to 64. Assume that the scores follow a multivariate normal distribution $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unknown $\boldsymbol{\mu}$ and unknown $\boldsymbol{\Sigma}$. The mean score and covariance matrix of the score are:

$$\bar{y} = \begin{pmatrix} 55.24 \\ 34.97 \end{pmatrix}$$

$$\boldsymbol{S} = \begin{pmatrix} 210.54 & 126.99 \\ 126.99 & 119.68 \end{pmatrix}$$

Test the hypothesis $H_0: \boldsymbol{\mu} = (60,\ 50)'$ vs $H_1: \boldsymbol{\mu} \neq (60,\ 50)'$ at 5% level of significance. Your solution should include the following:

4.1. State the test statistics to be used and its corresponding distribution **[2]**

4.2. State the decision (rejection) rule and compute the tabulated value using an appropriate statistical table **[2]**

4.3. Compute the test statistics and write up your decision and conclusion **[5]**

Question 5 [14 Marks]

5. Two psychological tests were given to 11 men and 10 women. The variables are y_1 = tool recognition and y_2 = vocabulary. The mean vectors and covariance matrices of the two samples are $\bar{y}_1 = \begin{pmatrix} 12 \\ 13 \end{pmatrix}$, $\bar{y}_2 = \begin{pmatrix} 16 \\ 17 \end{pmatrix}$, $\boldsymbol{S}_1 = \begin{pmatrix} 5 & 4 \\ 4 & 13 \end{pmatrix}$ and $\boldsymbol{S}_2 = \begin{pmatrix} 9 & 7 \\ 7 & 18 \end{pmatrix}$.
Assume that the observations are bivariate and follow multivariate normal distributions $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, for $i = 1 \; and \; 2$.

5.1. Compute the pooled covariance matrix **[3]**

5.2. Conduct a test if there is any significant difference between the vector of expected mean scores of men and women at 5% level of significance. Your answer should include the following:

5.2.1. State the null and alternative hypothesis to be tested **[1]**

5.2.2. State the test statistics to be used and its corresponding distribution **[2]**

5.2.3. State the decision (rejection) rule and compute the tabulated value using an appropriate statistical table **[3]**

5.2.4. Compute the test statistics and write up your decision and conclusion **[5]**

Question 6 [20 Marks]

6. Let $\boldsymbol{x} \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} 5 \\ 6 \\ 7 \\ 8 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 3 & 2 & 0 \\ 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix}$.

If we define a new random variable $y = 2x_2 - 3x_3 + x_4$, then

6.1. drive the distribution of y. **[4]**

6.2. derive the joint distribution of y and x_3. Are they independently distributed? Provide explanation for your answer. **[6]**

6.3. drive the conditional distribution of x_2, given $\begin{pmatrix} x_1 \\ x_3 \end{pmatrix}$ **[10]**

Question 7 [11 Marks]

7. Suppose that x_1, x_2, x_3 and x_4 are independent variables with the $N(0,4)$ distribution. Define the following random variables:

$$\begin{aligned} z_1 &= x_1 + y_1 \\ z_2 &= x_1 + x_2 + y_2 \\ z_3 &= x_1 + x_2 + x_3 + y_3 \\ z_4 &= x_1 + x_2 + x_3 + x_4 + y_4 \end{aligned}$$

where y_1, y_2, y_3 and y_4 have the $N(0,1)$ distribution, and are independent of each other and also independent of x_1, x_2, x_3 and x_4 .

7.1. Find the covariance matrix for the vector $\boldsymbol{z}' = (z_1 \; z_2 \; z_3 \; z_4)$ **[8]**

7.2. Find the variance of $\bar{z} = \frac{1}{4}(z_1 + z_2 + z_3 + z_4)$ **[3]**

Question 8 [15 Marks]

8. Technical and artistic scores of figure skaters are correlated, and therefore, they can probably be represented by one principal component. A sample of skaters yields covariance matrix $\boldsymbol{S} = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}$.

8.1. Compute coefficients for the first sample principal component y_1. **[9]**

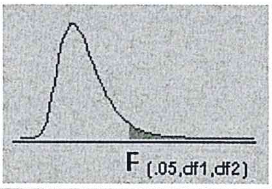
8.2. What percent of the total variance is attributed to y_1? **[2]**

8.3. Two competing figure skaters receive the scores $x_1 = (5.5, 5.9)$ and $x_2 = (5.7, 5.7)$. Which of them should be considered the winner, according to the first principal component? **[4]**

=== END OF PAPER===

TOTAL MARKS: 100

Table for α=.05



df2/df1	1	2	3	4	5	6	7	8	9	10	12
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882	243.906
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.384	19.396	19.413
3	10.128	9.552	9.277	9.117	9.014	8.941	8.887	8.845	8.812	8.786	8.745
4	7.709	6.944	6.591	6.388	6.256	6.163	6.0942	6.041	5.998	5.964	5.912
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678
6	5.987	5.143	4.757	4.533	4.387	4.284	4.207	4.147	4.099	4.060	3.999
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.676	3.637	3.575
8	5.318	4.459	4.066	3.838	3.688	3.581	3.501	3.438	3.388	3.347	3.284
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.229	3.178	3.137	3.073
10	4.965	4.103	3.708	3.478	3.326	3.217	3.136	3.072	3.020	2.978	2.913
11	4.844	3.982	3.587	3.358	3.204	3.095	3.012	2.948	2.896	2.854	2.788
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.645	2.602	2.534
15	4.543	3.682	3.287	3.056	2.901	2.791	2.707	2.641	2.587	2.544	2.475
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.537	2.494	2.425
17	4.451	3.591	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.441	2.393	2.348	2.278