NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

FACULTY OF HEALTH, APPLIED SCIENCES AND NATURAL RESOURCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

QUALIFICATION: Bachelor of Science Honours in Applied Statistics	
QUALIFICATION CODE: 08BSHS	LEVEL: 8
COURSE CODE: MVA802S	COURSE NAME: MULTIVARIATE ANALYSIS
SESSION: JANUARY 2023	PAPER: THEORY
DURATION: 3 HOURS	MARKS: 100

SUPPLEMENTARY / SECOND OPPORTUNITY EXAMINATION QUESTION PAPER	
EXAMINER	Dr D. B. GEMECHU
MODERATOR:	Prof L. PAZVAKAWAMBWA

INSTRUCTIONS
1. There are 8 questions, answer ALL the questions by showing all the necessary steps.
2. Write clearly and neatly.
3. Number the answers clearly.
4. Round your answers to at least four decimal places, if applicable.

PERMISSIBLE MATERIALS

1. Non-programmable scientific calculator

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

ATTACHMENTS
Two statistical distribution tables (z-and F-distribution tables)

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 [9 marks]

2. A number of patients with bronchus cancer were treated with ascorbate and compared with matched patients who received no ascorbate. The data are given in Table 1. The variables measured were

y_1, x_1 = survival time (days) from date of first hospital admission,
y_2, x_2 = survival time from date of untreatability.

Table 1: Survival Times for Bronchus Cancer Patients and Matched Controls

No Ascorbate Patients		Ascorbate Patients		Difference	
x_1	x_2	y_1	y_2	d_1	d_2
72	33	81	74	9	41
34	18	61	23	27	5
84	20	20	16	-64	-4
98	58	50	50	-48	-8
48	13	46	87	-2	74
42	49	66	75	24	26
70	38	63	50	-7	12
90	24	64	50	-26	26

2.1. calculate the sample mean vector for the difference, $\bar{d}$. **[3]**

2.2. If $S_d = \begin{pmatrix} 1081.269 & 390.0714 \\ 390.0714 & 722.8571 \end{pmatrix}$, test the hypothesis $H_o: \mu_x - \mu_y = 0$ using a paired comparison test at 5% of significancy **[6]**

Question 3 [16 marks]

3. Let x_1, x_2, x_3 and x_4 be independent variables that each have the normal distribution with mean 2 and standard deviation one. Define the random variables $y_1, y_2,$ and y_3 as follows:

$$\begin{aligned} y_1 &= (x_1 + x_2)/2 \\ y_2 &= x_3 + x_4 \\ y_3 &= x_1 + x_4 \end{aligned}$$

If we let the random vector $y = (y_1 \; y_2 \; y_3)'$, then

3.1. derive the distribution of y_2 and compute $P(y_2 < 0.3)$. **[8]**

3.2. derive the distribution of y. **[8]**

Question 4 [10 marks]

4. Let $z = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \sim N_3 \left(\begin{pmatrix} 2 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \right)$, derive the conditional distribution of

(z_1, z_2) given z_3 **[10]**

Question 5 [11 marks]

5. Perspiration from 19 healthy females was analyzed. Two components, y_1 = sweat rate, and y_2 = sodium, were measured. Assume that the data is from a multivariate normal distribution $N_2(\mu, \Sigma)$ with unknown μ and unknown Σ. The mean score and covariance matrix of the score are:

$$\bar{y} = \begin{pmatrix} 4.640 \\ 45.400 \end{pmatrix}$$

$$S = \begin{pmatrix} 2.879 & 10.010 \\ 10.010 & 199.788 \end{pmatrix}$$

Test the hypothesis $H_0: \mu = (4, 50)'$ vs $H_1: \mu \neq (4, 50)'$ at 5% level of significance. Your solution should include the following:

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

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

5.3. Compute the test statistics and write up your decision and conclusion [7]

Question 6 [16 marks]

6. Independent samples of size $n_1 = 11$ and $n_2 = 11$ were taken from populations of Katatura homeowners with and without air conditioning, respectively. Two measurements of electrical usage (in kilowatt hours) were made on each home: y_1, a measure of total on-peak consumption during September 2018 and, y_2, a measure of total off-peak consumption during September 2018. The resulting summary statistics are:

	Sample mean vector	Sample covariance matrices
Home with air conditioning	$\begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$	$\begin{pmatrix} 13.8 & 23.4 \\ 23.4 & 73.1 \end{pmatrix}$
Home without air conditioning	$\begin{pmatrix} 130.0 \\ 355.0 \end{pmatrix}$	$\begin{pmatrix} 86.3 & 19.6 \\ 19.6 & 55.9 \end{pmatrix}$

Assuming equality of population covariance matrices and bivariate normal distributions of observations within each group,

6.1. Compute the pooled covariance matrix [3]

6.2. Test the hypothesis that homes with air conditioning have the same vector of means for on-peak and off-peak consumption as home without air conditioning at 5% level of significance. Your answer should include the following:

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

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

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

6.2.4. Compute the test statistics and write up your decision and conclusion [7]

Question 7 [20 Marks]

7. Observations on two responses are collected for three treatments. The observation vectors $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ are

Treatment 1: $\begin{bmatrix} 10 \\ 4 \end{bmatrix}$, $\begin{bmatrix} 7 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 10 \\ 8 \end{bmatrix}$

Treatment 2: $\begin{bmatrix} 1 \\ 5 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$

Treatment 3: $\begin{bmatrix} 4 \\ 9 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 10 \end{bmatrix}$, $\begin{bmatrix} 3 \\ 8 \end{bmatrix}$

7.1.	Construct the one-way MANOVA table	[14]
7.2.	Evaluate Wilks' lambda, Λ_{wilks}	[3]
7.3.	Test for vector of treatment effects at 5% level of significance. Your answer should include specification of the null and alternative hypothesis.	[3]

Hint: Use the test statistics: $\left(\frac{N-g-1}{g-1}\right)\left(\frac{1-\sqrt{\Lambda_{wilks}}}{\sqrt{\Lambda_{wilks}}}\right) \sim F_{2(g-1),2(N-g-1)}$

Question 8 [10 Marks]

8. A principal component analysis (PCA) was performed on a dataset involving six variables representing scores from six tests of different aspects of educational ability (Visual perception "visperc", Cube "cubes" and lozenge "lozenges" identification, Word meanings "wordmean", sentence structure "sentence" and paragraph understanding "paragraph") on 73 girls from seventh-eighth grade students. The resulting software output is presented in Tables 1 and 2 below. Use this output and answer the following questions.

|		
8.1.	Discuss and interpret Bartlett's Test of Sphericity.	[4]
8.2.	Using the results of the principal components analysis, draw a scree plot.	[3]
8.3.	What percentage of total variation is explained by the first principal component? What percentage of total variation is explained by the first two principal component?	[3]

Table 1: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.763
Bartlett's Test of Sphericity	Approx. Chi-Square	180.331
	df	15
	Sig.	<0.001

Table 2: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.099	51.648		3.099		
2	1.349	22.478		1.349		
3	.549	9.153				
4	.488	8.127				
5	.282	4.698				
6	.234	3.895				
Extraction Method: Principal Component Analysis.						

=== END OF PAPER===
TOTAL MARKS: 100

Area between 0 and z
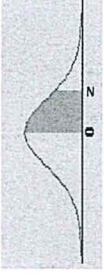
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Table for α=.05

$F_{(.05, df1, df2)}$
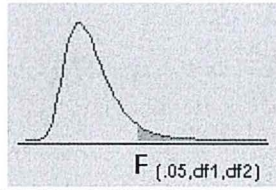
df2/df1	1	2	3	4	5	6	7	8	9	10	12
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882	243.906
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.384	19.396	19.413
3	10.128	9.552	9.277	9.117	9.014	8.941	8.887	8.845	8.812	8.786	8.745
4	7.709	6.944	6.591	6.388	6.256	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