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: NOVEMBER 2022	**PAPER:** THEORY
DURATION: 3 HOURS	**MARKS:** 100

FIRST 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 6 PAGES (Including this front page)

ATTACHMENTS

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

Question 1 [11 Marks]

1.1. Briefly discuss a one-way MANOVA. Your answer should include (Definition, three assumptions of one-way MANOVA, hypothesis to be tested under one-way MANOVA and two of the most common test statistics used to test the hypothesis). **[1+2+2]**

1.2. Briefly discuss two-sample profile analysis. Your answer should include the definition, the assumptions, and the possible hypothesis of interests that can be tested using this approach. **[1+2+3]**

Question 2 [9 Marks]

2. The data in table below are three measurements on air-pollution variables recorded on three different days.

Days	**Solar radiation,** y_1	**Nitrogen Dioxide (NO_2),** y_2	**Ozone (O_3),** y_3
1	72	18	9
2	70	11	7
3	80	13	11

Assume that $\boldsymbol{y} \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unknown $\boldsymbol{\mu}$ and unknown $\boldsymbol{\Sigma}$. Then, **using the matrices approach**, calculate the **maximum likelihood estimate** of the population:

2.1. mean vector. [3]

2.2. variance-covariance matrix. [6]

Question 3 [10 Marks]

3. If $\boldsymbol{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{z} = \left(\boldsymbol{\Sigma}^{1/2}\right)^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$, then show that $\boldsymbol{z} \sim \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{I})$. **Hint**: Use the uniqueness property of joint moment generating function. **[10]**

Question 4 [11 marks]

4. 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 $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unknown $\boldsymbol{\mu}$ and unknown $\boldsymbol{\Sigma}$. The mean score and covariance matrix of the score are:

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

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

Test the hypothesis $H_0: \boldsymbol{\mu} = (4,\ 50)'$ vs $H_1: \boldsymbol{\mu} \neq (4,\ 50)'$ 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 **[7]**

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 $\bar{y}_1 = \begin{pmatrix}12\\13\end{pmatrix}, \bar{y}_2 = \begin{pmatrix}16\\17\end{pmatrix}, S_1 = \begin{pmatrix}5 & 4\\4 & 13\end{pmatrix}$ and $S_2 = \begin{pmatrix}9 & 7\\7 & 18\end{pmatrix}$.
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 [23 Marks]

6. Let $\boldsymbol{x} \sim N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{x} = \begin{pmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix}5\\3\\7\\4\\9\end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix}4 & -1 & 0 & 0 & 2\\-1 & 4 & 2 & 0 & 4\\0 & 2 & 9 & 0 & 3\\0 & 0 & 0 & 9 & 7\\2 & 4 & 3 & 7 & 16\end{pmatrix}$.
Answer the following questions based on the above information.
   - 6.1. If $z_1 = \frac{x_1+x_3}{2}$ and $z_2 = x_1 - \frac{1}{2}x_2$ then, find the joint distribution of $z_1$and z_2. Are they independently distributed? Provide explanation for your answer. **[7]**
   - 6.2. Find the conditional distribution of x_2 given (x_1, x_3). **[11]**
   - 6.3. If $y = 2x_1 - 3x_2 + x_3$, then find $P(y > 7)$ **[5]**

Question 7 [9 Marks]

7. Let $\boldsymbol{X}' = [X_1, X_2, \ldots, X_p]$ have covariance matrix $\boldsymbol{\Sigma}$ with eigenvalue-eigenvector pairs $(\lambda_1, \boldsymbol{e}_1), (\lambda_2, \boldsymbol{e}_2), \ldots, (\lambda_p, \boldsymbol{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let $Y_i = \boldsymbol{e}_i'X$, $Y_2 = \boldsymbol{e}_2'\boldsymbol{X}, \ldots, Y_p = \boldsymbol{e}_p'\boldsymbol{X}$ be the principal components. Then show that
   - 7.1. $Var(Y_i) = \lambda_i$ **[4]**
   - 7.2. $tr(\boldsymbol{\Sigma}) = \sum_{i=1}^{p} Var(Y_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p$ **[5]**

Question 8 [13 marks]

8. A researcher compared judges' scores on fish prepared by three methods. Twelve fish were cooked by each method, and several judges tasted fish samples and rated each on four variables: y_1 = aroma, y_2 = flavor, y_3 = texture, and y_4 = moisture. The summary statistics of the data are given in the attached software output (**Tables 1-5 given below**).
   - 8.1. Draw conclusion of the Box test for equality of covariance matrix using the 5% significance level. Your answer should include the hypothesis to be tested, test statics and $p - value$ and conclusion. **[3]**
   - 8.2. Are there significant mean difference of judges' scores (as rated each on four variables) between three different methods? Your answer should include the hypothesis to be tested, test statics and $p - value$ and conclusion. **[4]**
   - **8.3.** Are there significant mean difference of judges' score on flavour of fish prepared by three methods? If so, which cooking methods differ? **[4]**
   - **8.4.** Are there significant mean difference judges' score on moisture of fish prepared by three methods? Explain in detail. **[2]**

Table 1: Box's Test of Equality of Covariance Matrices[a]

Box's M	16.292
F	.669
df1	20
df2	3909.028
Sig.	.860

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.
a. Design: Intercept + Method

Table 2: Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.993	1109.613[b]	4.000	30.000	.000	.993
	Wilks' Lambda	.007	1109.613[b]	4.000	30.000	.000	.993
	Hotelling's Trace	147.948	1109.613[b]	4.000	30.000	.000	.993
	Roy's Largest Root	147.948	1109.613[b]	4.000	30.000	.000	.993
Method	Pillai's Trace	.864	5.897	8.000	62.000	.000	.432
	Wilks' Lambda	.220	8.488[b]	8.000	60.000	.000	.531
	Hotelling's Trace	3.162	11.461	8.000	58.000	.000	.613
	Roy's Largest Root	3.036	23.526[c]	4.000	31.000	.000	.752

a. Design: Intercept + Method

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Table 3: Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
flavor	Based on Mean	.158	2	33	.855
	Based on Median	.245	2	33	.784
	Based on Median and with adjusted df	.245	2	32.566	.784
	Based on trimmed mean	.166	2	33	.848
texture	Based on Mean	.592	2	33	.559
	Based on Median	.547	2	33	.584
	Based on Median and with adjusted df	.547	2	32.090	.584
	Based on trimmed mean	.588	2	33	.561
moisture	Based on Mean	1.167	2	33	.324
	Based on Median	1.263	2	33	.296
	Based on Median and with adjusted df	1.263	2	32.455	.296

	Based on trimmed mean	1.195	2	33	.316
aroma	Based on Mean	.684	2	33	.512
	Based on Median	.680	2	33	.514
	Based on Median and with adjusted df	.680	2	31.390	.514
	Based on trimmed mean	.695	2	33	.506

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Method

Table 4: Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Flavour	4.605[a]	2	2.303	9.378	.001	.362
	Texture	2.382[b]	2	1.191	3.386	.046	.170
	Moisture	.811[c]	2	.405	1.266	.295	.071
	Aroma	1.051[d]	2	.525	1.293	.288	.073
Intercept	Flavour	995.402	1	995.402	4054.092	.000	.992
	Texture	1110.000	1	1110.000	3155.719	.000	.990
	Moisture	1309.234	1	1309.234	4089.096	.000	.992
	Aroma	975.521	1	975.521	2400.910	.000	.986
Method	Flavour	4.605	2	2.303	9.378	.001	.362
	Texture	2.382	2	1.191	3.386	.046	.170
	Moisture	.811	2	.405	1.266	.295	.071
	Aroma	1.051	2	.525	1.293	.288	.073
Error	Flavour	8.103	33	.246			
	Texture	11.607	33	.352			
	Moisture	10.566	33	.320			
	Aroma	13.408	33	.406			
Total	Flavour	1008.110	36				
	Texture	1123.990	36				
	Moisture	1320.610	36				
	Aroma	989.980	36				
Corrected Total	Flavour	12.708	35				
	Texture	13.990	35				
	Moisture	11.376	35				
	Aroma	14.459	35				

a. R Squared = .362 (Adjusted R Squared = .324)

b. R Squared = .170 (Adjusted R Squared = .120)

c. R Squared = .071 (Adjusted R Squared = .015)

d. R Squared = .073 (Adjusted R Squared = .016)

Table 5: Pairwise Comparisons

Dependent Variable	(I) Method	(J) Method	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
flavour	1	2	.475	.202	.075	-.035	.985
		3	.875*	.202	.000	.365	1.385
	2	1	-.475	.202	.075	-.985	.035
		3	.400	.202	.169	-.110	.910
	3	1	-.875*	.202	.000	-1.385	-.365
		2	-.400	.202	.169	-.910	.110
texture	1	2	.133	.242	1.000	-.477	.744
		3	-.467	.242	.188	-1.077	.144
	2	1	-.133	.242	1.000	-.744	.477
		3	-.600	.242	.055	-1.211	.011
	3	1	.467	.242	.188	-.144	1.077
		2	.600	.242	.055	-.011	1.211
moisture	1	2	.108	.231	1.000	-.474	.691
		3	-.250	.231	.861	-.833	.333
	2	1	-.108	.231	1.000	-.691	.474
		3	-.358	.231	.391	-.941	.224
	3	1	.250	.231	.861	-.333	.833
		2	.358	.231	.391	-.224	.941
aroma	1	2	.125	.260	1.000	-.531	.781
		3	.408	.260	.378	-.248	1.065
	2	1	-.125	.260	1.000	-.781	.531
		3	.283	.260	.852	-.373	.940
	3	1	-.408	.260	.378	-1.065	.248
		3	-.283	.260	.852	-.940	.373

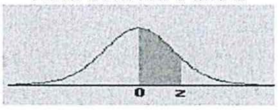
Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

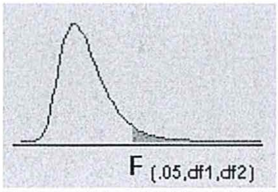
=== END OF PAPER===
TOTAL MARKS: 100

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



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