NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Faculty of Health, Natural Resources and Applied Sciences

School of Natural and Applied Sciences

Department of Mathematics, Statistics and Actuarial Science

13 Jackson Kaujeua Street
Private Bag 13388
Windhoek
NAMIBIA

T: +264 61 207 2913
E: msas@nust.na
W: www.nust.na

QUALIFICATION: Bachelor of Science Honours in Applied Statistics	
QUALIFICATION CODE: 08BSHS	**LEVEL:** 8
COURSE CODE: MVA802S	**COURSE NAME:** MULTIVARIATE ANALYSIS
SESSION: NOVEMBER 2023	**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 6 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. Nonprogrammable scientific calculators with no cover.

THIS QUESTION PAPER CONSISTS OF 5 PAGES (Including this front page)

ATTACHMENTS

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

Question 1 [12 Marks]

1.1. Briefly discuss multivariate statistical analysis and compare the multivariate techniques: discriminant analysis and cluster analysis. **[3]**

1.2. Briefly discuss a two-way MANOVA additive model. Your answer should include (the model, three assumptions, hypothesis to be tested under two-way MANOVA and two of the most common test statistics used to test the hypothesis). **[2+3+2+2]**

Question 2 [10 Marks]

2. In assessment of an experiment of assessing plant growth, a researcher collected data on height (y_1), root length (y_2), and width (y_3) of three new species of plants in mm. The results obtained were listed below:

Plant	Height	length	Width
1	320	260	120
2	260	250	100
3	350	300	140

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

2.1. mean vector. **[2]**

2.2. variance-covariance matrix. **[6]**

2.3. the total sample variance. **[2]**

Question 3 [34 Marks]

3.1. If $y \sim N_p(\mu_y, \Sigma_y)$ and $x \sim N_p(\mu_x, \Sigma_x)$ are independent, then show that $y + x \sim N_p(\mu_y + \mu_x, \Sigma_y + \Sigma_x)$. Hint use the uniqueness property of joint moment generating function. **[9]**

3.2. Each Delicious Candy Company sore makes 4 sizes candy bars: mini (x_1), regular (x_2), fun (x_3) and big size (x_4). Assume the weights (in ounces) of the candy bars (x_1, x_2, x_3, x_4) follow multivariate normal distribution with parameters:

$$\mu = \begin{pmatrix} 3 \\ 4 \\ 2 \\ 6 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 5 & -2 & 1 & 4 \\ -2 & 4 & -1 & 0 \\ 1 & -1 & 4 & 2 \\ 4 & 0 & 2 & 9 \end{pmatrix}.$$

3.2.1. Find the conditional distribution of x_3 given (x_2, x_4). **[12]**

3.2.2. If $y = 2x_2 - 3x_3 + x_4$, then find $P(y > 7)$ **[5]**

3.3. Suppose that x_1, x_2, x_3 and x_4 are independent variables with the $N(0, 2)$ distribution. Define the following random variables:

$$\begin{aligned} z_1 &= x_1 + \frac{y_1}{2} \\ z_2 &= x_1 + x_2 + \frac{y_2}{2} \\ z_3 &= x_1 + x_2 + x_3 + \frac{y_3}{2} \\ z_4 &= x_1 + x_2 + x_3 + x_4 + \frac{y_4}{2} \end{aligned}$$

where y_1, y_2, y_3 and y_4 have the $N(0, 1)$ distribution and are independent of each other and independent of x_1, x_2, x_3 and x_4. Find the variance-covariance matrix for the vector $z' = (z_1\ z_2\ z_3\ z_4)$ **[8]**

NAMIBIA UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Faculty of Health, Natural Resources and Applied Sciences

School of Natural and Applied Sciences

Department of Mathematics, Statistics and Actuarial Science

13 Jackson Kaujeua Street
Private Bag 13388
Windhoek
NAMIBIA

T: +264 61 207 2913
E: msas@nust.na
W: www.nust.na

QUALIFICATION: Bachelor of Science Honours in Applied Statistics	
QUALIFICATION CODE: 08BSHS	**LEVEL:** 8
COURSE CODE: MVA802S	**COURSE NAME:** MULTIVARIATE ANALYSIS
SESSION: NOVEMBER 2023	**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 6 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. Nonprogrammable scientific calculators with no cover.

THIS QUESTION PAPER CONSISTS OF 5 PAGES (Including this front page)

ATTACHMENTS

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

Question 1 [12 Marks]

1.1. Briefly discuss multivariate statistical analysis and compare the multivariate techniques: discriminant analysis and cluster analysis. **[3]**

1.2. Briefly discuss a two-way MANOVA additive model. Your answer should include (the model, three assumptions, hypothesis to be tested under two-way MANOVA and two of the most common test statistics used to test the hypothesis). **[2+3+2+2]**

Question 2 [10 Marks]

2. In assessment of an experiment of assessing plant growth, a researcher collected data on height (y_1), root length (y_2), and width (y_3) of three new species of plants in mm. The results obtained were listed below:

Plant	Height	length	Width
1	320	260	120
2	260	250	100
3	350	300	140

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 population:

2.1. mean vector. **[2]**

2.2. variance-covariance matrix. **[6]**

2.3. the total sample variance. **[2]**

Question 3 [34 Marks]

3.1. If $\boldsymbol{y} \sim N_p(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$ and $\boldsymbol{x} \sim N_p(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ are independent, then show that $\boldsymbol{y} + \boldsymbol{x} \sim N_p(\boldsymbol{\mu}_y + \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_y + \boldsymbol{\Sigma}_x)$. Hint use the uniqueness property of joint moment generating function. **[9]**

3.2. Each Delicious Candy Company sore makes 4 sizes candy bars: mini (x_1), regular (x_2), fun (x_3) and big size (x_4). Assume the weights (in ounces) of the candy bars (x_1, x_2, x_3, x_4) follow multivariate normal distribution with parameters:

$$\boldsymbol{\mu} = \begin{pmatrix} 3 \\ 4 \\ 2 \\ 6 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 5 & -2 & 1 & 4 \\ -2 & 4 & -1 & 0 \\ 1 & -1 & 4 & 2 \\ 4 & 0 & 2 & 9 \end{pmatrix}.$$

3.2.1. Find the conditional distribution of x_3 given (x_2, x_4). **[12]**

3.2.2. If $y = 2x_2 - 3x_3 + x_4$, then find $P(y > 7)$ **[5]**

3.3. Suppose that x_1, x_2, x_3 and x_4 are independent variables with the $N(0, 2)$ distribution. Define the following random variables:

$$\begin{aligned} z_1 &= x_1 + \frac{y_1}{2} \\ z_2 &= x_1 + x_2 + \frac{y_2}{2} \\ z_3 &= x_1 + x_2 + x_3 + \frac{y_3}{2} \\ z_4 &= x_1 + x_2 + x_3 + x_4 + \frac{y_4}{2} \end{aligned}$$

where y_1, y_2, y_3 and y_4 have the $N(0, 1)$ distribution and are independent of each other and independent of x_1, x_2, x_3 and x_4. Find the variance-covariance matrix for the vector $\boldsymbol{z}' = (z_1\ z_2\ z_3\ z_4)$ **[8]**

Question 4 [16 Marks]

4. Bars of soap manufactured in each of two ways. The characteristics y_1 =lather and y_2 = mildness is measured. The summary statistics for 9 bars produced by each method (methods 1 and 2) are:

$$\bar{y}_1 = \begin{pmatrix} 8 \\ 4 \end{pmatrix}, \bar{y}_2 = \begin{pmatrix} 10 \\ 3 \end{pmatrix}, S_1 = \begin{pmatrix} 5 & 4 \\ 4 & 7 \end{pmatrix} \text{ and } S_2 = \begin{pmatrix} 7 & 2 \\ 2 & 3 \end{pmatrix}.$$

Assume that the observations are bivariate and follow multivariate normal distributions $N(\mu_i, \Sigma)$, for $i = 1 \; and \; 2$.

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

4.2. Conduct a test if there is any significant difference between the vector of expected mean measurements of the two methods at 5% level of significance. Your answer should include the following:

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

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

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

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

Question 5 [7 Marks]

5. As part of the study of love and marriage, a sample of husbands and wives were asked to respond to these questions:

Question 1: What is the level of passionate love you feel for your partner?

Question 2: What is the level of passionate love that your partner feels for you?

Question 3: What is the level of companionate love that you feel for your partner?

Question 4: What is the level of companionate love that your partner feels for you?

The responses were recorded on the following 5-point scale. Thirty husbands and 30 wives gave the responses in Table 3, where y_1, = a 5-point- scale response to Question 1, y_2 = a 5-point-scale response to Question 2, y_3 = a 5-point-scale response to Question 3, and y_4 = a 5-point-scale response to Question 4.

The R package output for the analysis of this data is given Table 1. Based on the output provided Answer the following questions based on the output provided (use $\alpha = 0.05$ level of significancy). Your answer should include the hypothesis to be tested (H_o), the test statistics, the p-value and your decision and conclusions.

5.1. Is the husband rating wife profile parallel to the wife rating husband profile? **[3]**

5.2. Test for coincident profiles at the same level of significance. Thus, are the groups at the same level? **[2]**

5.3. Test for level profiles. **[2]**

Table 1: Profile analysis software result

```
>library(profileR)
>Y<-read.csv(file=mydata.csv", header = T)
>profres <- pbg(Y[,1:4], Y[,5], original.names = TRUE, profile.plot = TRUE)
>summary(profres)
Call:
pbg(data = Y[, 1:4], group = Y[, 5], original.names = TRUE, profile.plot = TRUE)

Hypothesis Tests:
$` Profiles are parallel`
```

	Multivariate.Test	Statistic	Approx.F	num.df	den.df	p.value
1	Wilks	0.878573	2.579917	3	56	0.062559
2	Pillai	0.121427	2.579917	3	56	0.062559
3	Hotelling-Lawley	0.13821	2.579917	3	56	0.062559
4	Roy	0.13821	2.579917	3	56	0.062559

$` Profiles have equal levels`

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	1	0.234	0.2344	1.533	0.221
Residuals	58	8.869	0.1529		

$` Profiles are flat`

Hotteling T2	F	df1	df2	p-value
25.4415	8.18807	3	56	0.000131

Question 6 [21]

6.1. Let $\boldsymbol{X}' = [X_1, X_2, \ldots, X_p]$ have covariance matrix $\boldsymbol{\Sigma}$ with eigenvalue-eigenvector pairs $(\lambda_1, \boldsymbol{e}_1)$, $(\lambda_2, \boldsymbol{e}_2)$, ..., $(\lambda_p, \boldsymbol{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let $Y_i = \boldsymbol{e}_i'X$, $Y_2 = \boldsymbol{e}_2'\boldsymbol{X}$, ..., $Y_p = \boldsymbol{e}_p'\boldsymbol{X}$ be the principal components. Then show that $\sum_{i=1}^{p} Var(Y_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} Var(X_i)$ **[9]**

6.2. At the beginning of the 20th century, one researcher obtained measurements on seven physical characteristics for each of 3000 convicted male criminals. The characteristics he measured are: X_1 Length of head from front to back (in cm); X_2 Head breadth (in cm); X_3 Face breadth (in cm); X_4 Length of left forefinger (in cm); X_5 Length of left forearm (in cm); X_6 Length of left foot (in cm); X_7 Height (in inches).

6.2.1. Based on the correlation matrix provided, it the data suitable for PCA? **[2]**

6.2.2. Show that the value of $\lambda_7 = 0.11$ **[2]**

6.2.3. What is the percentage of the total standardized variation attributed to the second principal component? **[2]**

6.2.4. Using the results of the principal components analysis, draw a scree plot. How many principal components do you recommend to retain based on the scree plot? **[4]**

6.2.5. Give a formula for computing the scores of the first principal component. **[2]**

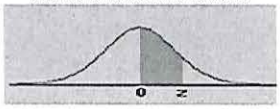
Table 2: The sample correlation matrix is:

x1	x2	x3	x4	x5	x6	x7
1	0.402	0.395	0.301	0.305	0.399	0.34
	1	0.618	0.15	0.135	0.206	0.183
		1	0.321	0.289	0.363	0.345
			1	0.846	0.759	0.661
				1	0.797	0.8
					1	0.736
						1

Table 3: The eigenvalues and eigenvectors for the sample correlation matrix are:

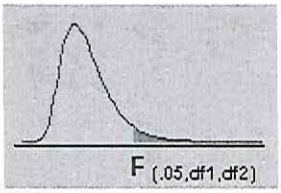
	1	2	3	4	5	6	7
Eigenvectors	0.285	-0.351	877	-0.088	-0.076	0.112	-0.023
	0.211	-0.643	-0.246	0.686	-0.098	-0.01	0.02
	0.294	-0.515	-0.387	-0.693	-0.112	0.029	-0.074
	0.435	0.24	-0.113	0.126	-0.604	0.33	0.5
	0.453	0.282	-0.079	0.127	-0.024	0.27	-0.787
	0.453	0.167	0.028	0.023	-0.065	-0.873	0.024
	0.434	0.182	-0.027	-0.09	0.776	0.208	0.352
Eigenvalues	3.82	1.49	0.65	0.36	0.34	0.23	λ_7

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



df1

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

df2