P

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:** JANUARY 2025 | **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 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 [15 Marks]

1.1. State at least two techniques of multivariate analysis and describe their objectives **[2]**

1.2. If a random variable $z$ is defined as a linear combination of $y_1, y_2, \ldots, y_p$ as $z_i = a_1 y_{i1} + a_2 y_{i2} + \cdot \cdot \cdot + a_p y_{ip}$, for $i = 1, 2, \ldots, n$, then show that $S_z^2 = \boldsymbol{a}'\boldsymbol{S}\boldsymbol{a}$, where $\boldsymbol{a}' = (a_1\ a_2\ \cdots\ a_p)$ and $\boldsymbol{S}$ is the sample variance covariance. **[6]**

1.3. Let $T_x^2$ be the Hotelling's $T^2$ –statistic for $\boldsymbol{x}$ and $T_y^2$ be the Hotelling's $T^2$ –statistic for $\boldsymbol{y}$, where $\boldsymbol{y}$ is a linear combination of $\boldsymbol{x}$ variables given by:

$$\boldsymbol{y}_{p\times1} = \boldsymbol{C}_{p\times p}\boldsymbol{x}_{p\times1} + \boldsymbol{d}_{p\times1},$$

where $\boldsymbol{C}$ is non-singular and $\boldsymbol{d} \in \mathbb{R}^p$. Show that $T_x^2 = T_y^2$, where

$$T_x^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\boldsymbol{S}_x^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)$$ **[7]**

## Question 2 [13 Marks]

2. Table below contains data on concentrations of heavy metals The researcher collected data on concentrations of heavy metals [Cu ($y_1$), Ni ($y_2$) and Pb($y_3$)] in river's sediment in Namibia at three different streams. The first three measurements are presented in table below:

| Sample | Cu | Ni | Pd |
|---|---|---|---|
| 1 | 19.8 | 17.3 | 33.2 |
| 2 | 17.2 | 15.5 | 36.2 |
| 3 | 20.1 | 19.2 | 40.9 |

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. correlation matrix in terms of **DSD**, by defining your matrix **D** and interpret your result. **[5]**

## Question 3 [14 Marks]

3. An experiment was conducted to determine whether protein and fiber content for wheat grown with fertilizer 1 is different from that for wheat grown with fertilizer 2. Wheat was grown in 22 plots. On 11 of these plots, fertilizer A was used; on the other 11 plots, fertilizer B was used. The protein and fiber content (in percent) of the wheat from each plot was measured. Assume also that the observations are bivariate and follow multivariate normal distributions for $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, i = 1 and 2. The sample mean vectors and sample covariance matrices from these measurements are:

$$\overline{\boldsymbol{y}}_1 = (12.1\ \ 14.3)', \qquad \overline{\boldsymbol{y}}_2 = (10.1\ \ 13.3)'$$

$$\boldsymbol{S}_1 = \begin{pmatrix} 2.2 & -1.1 \\ -1.1 & 0.9 \end{pmatrix}, \qquad \boldsymbol{S}_2 = \begin{pmatrix} 2.3 & -1.0 \\ -1.0 & 1.1 \end{pmatrix}$$

3.1. Find the pooled estimate of the covariance matrix for this data. **[3]**

3.2. Test the null hypothesis that the mean protein and fiber content is the same for both fertilizers at 5% level of significance. **[11]**

**Question 4 [27 Marks]**

4. Suppose a vector of random variable $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ is from a multivariate normal population with mean vector $\boldsymbol{\mu} = \begin{pmatrix} -3 \\ 1 \\ 4 \end{pmatrix}$ and variance covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. If we define a new random variable $y = \frac{x_1+x_2}{2}$, then
   - 4.1. Derive the distribution of $y$ and compute $P(y < -2)$ **[8]**
   - 4.2. Derive the joint distribution of $x_3$ and $y$. Are they independently distributed? Provide explanation for your answer. **[8]**
   - 4.3. Drive the conditional distribution of $(x_1, x_3)$ given $x_2$. **[11]**

**Question 5 [9 Marks]**

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

**Question 6 [22 Marks]**

6.1. 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). **[9]**

6.2. Heavy metals in river sediments are a significant concern due to their toxicological impacts on aquatic ecosystems and human health. Understanding the distribution and concentration levels of these metals across different rivers streams is essential for effective environmental management. Victor et al. (2024) analysed heavy metals concentrations across three different river streams in Namibia- upper, middle and lower. The heavy metals are nickel (Ni), zinc (Zn), copper (Cu), manganese (Mn), and lead (Pd). One of the objectives of the study was to investigate whether there is a mean concentrations difference of heavy metals among the three streams. The statistical summary of portion of the data, modified for this question, are presented below. Answer the following questions based on these results. Your answer to each question below should include **the hypothesis to be tested, test statics** and $\boldsymbol{p - value}$ and **conclusion.**

6.2.1. Draw conclusion of the Box's M test for equality of covariance matrix using the 5% significance level. Your answer should include the hypothesis to be tested, test statistics and $p - value$ and conclusion. **[3]**

6.2.2. Are there mean concentrations differences of heavy metals among the three streams? If so, for which heavy metal? Your answer should include the hypothesis to be tested, test statics and $p - value$ and conclusion. **[6]**

6.2.3. Briefly discuss the results of pairwise comparison. **[4]**

########Software output############
#Box's M-test for Homogeneity of Covariance Matrices

data: heavym_data
Chi-Sq (approx.) = 74.448, df = 42, p-value = 0.001497
Multivariate linear model

#One-way MANOVA

| | Df | Wilks | approx F | num Df | den Df | Pr(>F) |
|---|---|---|---|---|---|---|
| stream | 2 | 0.28618 | 3.1874 | 12 | 44 | 0.002417 |
| Residuals | 27 | | | | | |

#Univaraite ANOVA

Response Fe :

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| stream | 2 | 12077198 | 6038599 | 5.9776 | 0.007092 |
| Residuals | 27 | 27275442 | 1010202 | | |

---

Response Mn :

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| stream | 2 | 30499 | 15249.7 | 13.517 | 8.559e-05 |
| Residuals | 27 | 30461 | 1128.2 | | |

---

Response Ni :

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| stream | 2 | 0.098 | 0.0490 | 0.02 | 0.9803 |
| Residuals | 27 | 66.290 | 2.4552 | | |

---

Response Pb :

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| stream | 2 | 668.19 | 334.09 | 4.3302 | 0.02338 |
| Residuals | 27 | 2083.16 | 77.15 | | |

---

Response U :

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| stream | 2 | 9.6656 | 4.8328 | 4.2551 | 0.02475 |
| Residuals | 27 | 30.6659 | 1.1358 | | |

---

Response Zn :

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| stream | 2 | 72.651 | 36.325 | 4.1044 | 0.02777 |
| Residuals | 27 | 238.959 | 8.850 | | |

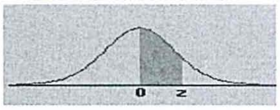
---

### Multiple Comparisons: Bonferroni method

| Dependent Variable | (I) stream | (J) stream | Mean Difference (I-J) | Std. Error | Sig. |
|---|---|---|---|---|---|
| Fe | lower | middle | 1063.6 | 449.48894 | 0.076 |
| | lower | upper | 1513.2000* | 449.48894 | 0.007 |
| | middle | upper | 449.6 | 449.48894 | 0.978 |
| Mn | lower | middle | 38.6000* | 15.02129 | 0.048 |
| | lower | upper | 78.1000* | 15.02129 | <0.001 |
| | middle | upper | 39.5000* | 15.02129 | 0.042 |
| Ni | lower | middle | 0.07 | 0.70074 | 1.000 |
| | lower | upper | -0.07 | 0.70074 | 1.000 |
| | middle | upper | -0.14 | 0.70074 | 1.000 |
| Pb | lower | middle | 0.651 | 3.92821 | 1.000 |
| | lower | upper | 10.3210* | 3.92821 | 0.042 |
| | middle | upper | 9.67 | 3.92821 | 0.062 |
| U | lower | middle | 0.923 | 0.47661 | 0.190 |
| | lower | upper | 1.3620* | 0.47661 | 0.024 |
| | middle | upper | 0.439 | 0.47661 | 1.000 |
| Zn | lower | middle | 2.16 | 1.33044 | 0.348 |
| | lower | upper | 3.8000* | 1.33044 | 0.024 |
| | middle | upper | 1.64 | 1.33044 | 0.685 |

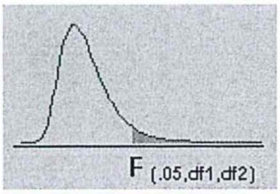
**===========END OF QUESTION PAPER===========**

**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

| 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 |