NANYANG TECHNOLOGICAL UNIVERSITY

SEMESTER II EXAMINATION 2023–2024

MH4501 – MULTIVARIATE ANALYSIS

May 2024 Time Allowed: 2 hours

INSTRUCTIONS TO CANDIDATES

1. This examination paper contains **FIVE (5)** questions and comprises **SIX (6)** printed pages.
2. Answer each question beginning on a **FRESH** page of the answer book.
3. This is a **RESTRICTED OPEN BOOK** exam. You are only allowed to bring into the examination hall **ONE DOUBLE-SIDED A4-SIZE REFERENCE SHEET WITH TEXTS HANDWRITTEN OR TYPED ON THE A4 PAPER WITHOUT ANY ATTACHMENTS** (e.g. sticky notes, post-it notes, gluing or stapling of additional papers)
4. Calculators may be used. However, you should write down systematically the steps in the workings.

QUESTION 1 **(15 marks)**

A bank developed two indices: $\boldsymbol{x}_1$ = financial index and $\boldsymbol{x}_2$ = social status index to classify their average customers from prestige customers. The following are the sample mean and sample covariance matrix of the realisations $x_i = (x_{i1}, x_{i2})^\top$, $i = 1, 2$ for two groups of customers:

$$\text{Group 1 (average)} \quad n_1 = 50 \quad \bar{x}_1 = (14.6, 2.5)^\top \quad S_1 = \begin{pmatrix} 3 & 0.6 \\ 0.6 & 1.1 \end{pmatrix}$$

$$\text{Group 2 (prestige)} \quad n_2 = 40 \quad \bar{x}_2 = (43.2, 13.5)^\top \quad S_2 = \begin{pmatrix} 20.4 & 7.4 \\ 7.4 & 4.3 \end{pmatrix}$$

We use the sample proportion in the data as prior probabilities and the loss of misclassifying a prestige customer to average group is three times as much as the loss of misclassifying an average customer to prestige group.

(a) Assume that the covariance matrices of Groups 1 and 2 are the same. Give the classification rule to allocate a new observation (customer) $x_0 = (x_{01}, x_{02})^\top$ to either of these two groups.

(b) Assume that the covariance matrices of Groups 1 and 2 are different. Give the classification rule to allocate a new observation (customer) $x_0 = (x_{01}, x_{02})^\top$ to either of these two groups. Which group will you allocate the new observation $x_0 = (22, 8)^\top$?

QUESTION 2 **(15 marks)**

We have the following sample correlation matrix for four original variables

$$R = \begin{pmatrix} 1 & 0.632 & 0.511 & 0.115 \\ 0.632 & 1 & 0.574 & 0.322 \\ 0.511 & 0.574 & 1 & 0.183 \\ 0.115 & 0.322 & 0.183 & 1 \end{pmatrix}$$

The first two eigenvalues and the corresponding normalized eigenvectors of R are

$$\begin{aligned} \lambda_1 &= 2.251 \quad u_1 = (0.541, 0.589, 0.532, 0.279)^\top \\ \lambda_2 &= 0.929 \quad u_2 = (0.317, -0.007, 0.174, -0.932)^\top \end{aligned}$$

(a) A two-factor model is used on this R, i.e., $R = \hat{L}\hat{L}^\top + \widehat{\Psi}$. Calculate the factor loading matrix $\hat{L}$ and matrix of specific variances $\widehat{\Psi}$ using the principal component solution method.

(b) Using the factor loadings in part (a), obtain the estimates of the following.

1. The communalities.
2. The proportion of variance explained by each factor.

(c) Explain the differences between Principal Component Analysis and Factor Analysis.

QUESTION 3 **(20 marks)**

Consider a population covariance matrix Σ of the form

$$\Sigma = \gamma I_p + aa^\top,$$

with $\gamma > 0$, I_p the $p \times p$ identity matrix and $a \in \mathbb{R}^p$.

(a) Show that a is an eigenvector of Σ.

(b) Show that if b is orthogonal to a, then b is also an eigenvector of Σ and calculate the corresponding eigenvalue.

(c) Obtain all the eigenvalues of Σ.

(d) Determine the proportion of the total variance explained by

1. the largest principal component of Σ.
2. the r largest principal components of Σ, where $r \in \{2, \ldots, p\}$.

QUESTION 4 **(20 marks)**

Suppose that $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4)^\top$ is a random vector with the covariance matrix

$$\Sigma = \begin{pmatrix} 5 & 3 & 2 & 1 \\ 3 & 5 & 1 & 2 \\ 2 & 1 & 5 & 3 \\ 1 & 2 & 3 & 5 \end{pmatrix}$$

(a) Verify that the principal components for this covariance matrix are

$$\boldsymbol{y}_1 = \frac{1}{2}(\boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3 + \boldsymbol{x}_4), \quad \boldsymbol{y}_2 = \frac{1}{2}(\boldsymbol{x}_1 + \boldsymbol{x}_2 - \boldsymbol{x}_3 - \boldsymbol{x}_4),$$
$$\boldsymbol{y}_3 = \frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{x}_2 + \boldsymbol{x}_3 - \boldsymbol{x}_4), \quad \boldsymbol{y}_4 = \frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{x}_2 - \boldsymbol{x}_3 + \boldsymbol{x}_4).$$

(b) Determine the least number of principal components that can account for at least 85% of the total variance. Show your calculations (Answers of only a number without justification will be marked zero).

(c) Compute the correlation coefficients between $\boldsymbol{y}_2$ and $\boldsymbol{x}_k$, $k = 1, 3, 4$.

(d) Let $\boldsymbol{z}$ be the standardised version of $\boldsymbol{x}$. What is the relationship between the principal components for $\boldsymbol{z}$ and $\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4$? Briefly justify your answer.

QUESTION 5 **(30 marks)**

Suppose that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are n i.i.d. random vectors drawn from $N_p(\mu, \Sigma)$ with sample mean $\bar{\boldsymbol{x}}$ and sample covariance matrix $\boldsymbol{S}$. For any $a \in \mathbb{R}^p$, define $\tilde{\boldsymbol{S}}(a)$ as the $p \times p$ matrix

$$\tilde{\boldsymbol{S}}(a) = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - a)(\boldsymbol{x}_i - a)^\top.$$

(a) Give without proving the maximum likelihood estimate of $\log\det(\Sigma)$.

(b) Calculate $\mathbb{E}[\tilde{\boldsymbol{S}}(\mu)]$, simplifying as much as possible.

(c) Show that

$$\tilde{\boldsymbol{S}}(a) = \frac{n-1}{n}\boldsymbol{S} + (\bar{\boldsymbol{x}} - a)(\bar{\boldsymbol{x}} - a)^\top$$

(d) Using Part (c), show that $n\,\mathrm{tr}(\tilde{\boldsymbol{S}}(a)) \geq (n-1)\,\mathrm{tr}(\boldsymbol{S})$.

END OF PAPER

MH4501 MULTIVARIATE ANALYSIS

Please read the following instructions carefully:

1. **Please do not turn over the question paper until you are told to do so. Disciplinary action may be taken against you if you do so.**

2. You are not allowed to leave the examination hall unless accompanied by an invigilator. You may raise your hand if you need to communicate with the invigilator.

3. Please write your Matriculation Number on the front of the answer book.

4. Please indicate clearly in the answer book (at the appropriate place) if you are continuing the answer to a question elsewhere in the book.