# National Testing Agency

| | |
|---|---|
| **Question Paper Name :** | Applied Multivariate Analysis 25th March 2021 Shift 2 |
| **Subject Name :** | Applied Multivariate Analysis |
| **Creation Date :** | 2021-03-25 22:30:46 |
| **Duration :** | 180 |
| **Number of Questions :** | 37 |
| **Total Marks :** | 100 |
| **Display Marks:** | Yes |

## Applied Multivariate Analysis

| | |
|---|---|
| **Group Number :** | 1 |
| **Group Id :** | 512452129 |
| **Group Maximum Duration :** | 0 |
| **Group Minimum Duration :** | 120 |
| **Show Attended Group? :** | No |
| **Edit Attended Group? :** | No |
| **Break time :** | 0 |
| **Group Marks :** | 100 |
| **Is this Group for Examiner? :** | No |

## Applied Multivariate Analysis-1

| | |
|---|---|
| **Section Id :** | 512452761 |
| **Section Number :** | 1 |
| **Section type :** | Online |
| **Mandatory or Optional :** | Mandatory |
| **Number of Questions :** | 20 |
| **Number of Questions to be attempted :** | 20 |
| **Section Marks :** | 20 |
| **Mark As Answered Required? :** | Yes |
| **Sub-Section Number :** | 1 |
| **Sub-Section Id :** | 512452763 |
| **Question Shuffling Allowed :** | Yes |

**Question Number : 1 Question Id : 51245211479 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

In the Single Linkage method, the distance between two clusters is defined as

1. the distance between the closest two members of the two clusters
2. the distance between the farthest two members of the two clusters
3. the distance between the centroids of the two clusters
4. the average distance between two members of the two clusters

**Options :**

51245234501. 1
51245234502. 2
51245234503. 3
51245234504. 4

**Question Number : 2 Question Id : 51245211480 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

In a 2 × 2 table with *a* Yes-Yes, *b* Yes-No, *c* No-Yes and *d* No-No, the *matching proportions* similarity measure is given by

1. (a+c)/(a+b+c+d)
2. (a+b)/(a+b+c+d)
3. (a+d)/(a+b+c+d)
4. (b+c)/(a+b+c+d)

**Options :**

51245234505. 1
51245234506. 2
51245234507. 3
51245234508. 4

**Question Number : 3 Question Id : 51245211481 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

The discriminating function for two groups with underlying multivariate normal densities is *linear* if

1. their means are equal
2. their dispersion matrices are equal
3. their means are unequal
4. their dispersion matrices are unequal

**Options :**

51245234509. 1
51245234510. 2
51245234511. 3
51245234512. 4

**Question Number : 4 Question Id : 51245211482 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

For a m-dimensional random variable, the sum of the variances of its first p (< m) principal components is

1. greater than its own total variability
2. equal to the total variability of any of its p components
3. less than its own total variability
4. equal to its own total variability

**Options :**

51245234513. 1
51245234514. 2
51245234515. 3
51245234516. 4

**Question Number : 5 Question Id : 51245211483 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

Principal component analysis is primarily used

1. as a data reduction technique
2. to study cause-effect relationships
3. as a grouping technique
4. to study inter variable relationships

**Options :**

51245234517. 1
51245234518. 2
51245234519. 3
51245234520. 4

**Question Number : 6 Question Id : 51245211484 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

In the divisive hierarchical clustering method with n units, we

1. start with a single cluster of n units and split it into n/2 clusters of 2 units each
2. start with n clusters of 1 unit each and merge them into 2 clusters of n/2 units
3. start with a single cluster of n units and split it into n clusters of 1 unit each
4. start with n clusters of 1 unit each and merge them into 1 cluster of n units

**Options :**

51245234521. 1
51245234522. 2
51245234523. 3
51245234524. 4

**Question Number : 7 Question Id : 51245211485 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

The part of the variance of a variable explained by the factors is known as

1. communality
2. uniqueness
3. factor loading
4. factor score

**Options :**

51245234525. 1
51245234526. 2
51245234527. 3
51245234528. 4

**Question Number : 8 Question Id : 51245211486 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

The dispersion matrix $\Sigma$ of a factor model can be split as (where L is the matrix of loadings, $\Psi$ is the error dispersion matrix and I is the identity matrix) :

1. $\Sigma = LL' + I$
2. $\Sigma = L + \Psi$
3. $\Sigma = I + L\Psi L'$
4. $\Sigma = LL' + \Psi$

**Options :**

51245234529. 1
51245234530. 2
51245234531. 3
51245234532. 4

**Question Number : 9 Question Id : 51245211487 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

For the pair of canonical variables $(U_j, V_j)$, j= 1, ..., p, which of the following is true :

1. $Cov(U_j, V_j) = 0$ for all j and $Cov(U_j, V_k) = 0$ for all $j \neq k$
2. $Cov(U_j, V_j) \neq 0$ for all j and $Cov(U_j, V_k) = 0$ for all $j \neq k$
3. $Cov(U_j, V_j) = 0$ for all j and $Cov(U_j, V_k) \neq 0$ for all $j \neq k$
4. $Cov(U_j, V_j) \neq 0$ for all j and $Cov(U_j, V_k) \neq 0$ for all $j \neq k$

**Options :**

51245234533. 1
51245234534. 2
51245234535. 3
51245234536. 4

**Question Number : 10 Question Id : 51245211488 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

If $X \sim N_p(\mu, \Sigma)$, then for a $q \times p$ matrix L, $Y = LX$ follows

1. $N_q(\mu, \Sigma)$
2. $N_q(L\mu, L\Sigma)$
3. $N_q(L\mu L', L\Sigma L')$
4. $N_q(L\mu, L\Sigma L')$

**Options :**

51245234537. 1

51245234538. 2

51245234539. 3

51245234540. 4

**Question Number : 11 Question Id : 51245211489 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

If for $\mathbf{X} = (X_1, ..., X_m)$, $E(X_j) = \mu_j$, $Var(X_j) = \sigma^2$ for all j and $Cov(X_j, X_k) = \rho$ for all $j \neq k$, $j,k = 1, ..., m$, then the first principal component is

1. $m^{-1}(X_1 + ... + X_m)$
2. $m(X_1 + ... + X_m)$
3. $m^{-½}(X_1^2 + ... + X_m^2)$
4. $m^{-½}(X_1 + ... + X_m)$

**Options :**

51245234541. 1

51245234542. 2

51245234543. 3

51245234544. 4

**Question Number : 12 Question Id : 51245211490 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

If $\mathbf{l} = (l_1, ..., l_m)$ be the eigen-values of the dispersion matrix $\sum$ and $\mathbf{L}$ the diagonal matrix of the eigen-values, i.e. $\mathbf{L} = diag((l_1, ..., l_m))$, then the asymptotic distribution of the sample eigen-values is

1. $N(\mathbf{l}, n^{-1/2}\mathbf{L}^2)$
2. $N(\mathbf{l}, n^{-1}\mathbf{L}^2)$
3. $N(\mathbf{l}, 2n^{-1}\mathbf{L}^2)$
4. $N(\mathbf{l}, 2n^{-1/2}\mathbf{L}^2)$

**Options :**

51245234545. 1

51245234546. 2
51245234547. 3
51245234548. 4

**Question Number : 13 Question Id : 51245211491 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**
**Correct Marks : 1 Wrong Marks : 0**

If

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_{p+q}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix}\right)$$

then the conditional distribution of $\mathbf{X}_2^{qx1}$ given $\mathbf{X}_1^{px1} = \mathbf{x}_1$ is

1. $N_q\ (\mu_2 + \sum_{21}\sum^{-}{}_{11}{}^{1}\ (\mathbf{x}_1 - \mu_1), \sum_{11} - \sum_{12}\sum_{22}{}^{-1}\sum_{21})$
2. $N_p\ (\mu_2 + \sum_{21}\sum_{11}{}^{-1}\ (\mathbf{x}_1 - \mu_1), \sum_{22}{}^{-1} - \sum_{21}\sum_{11}{}^{-1}\sum_{12})$
3. $N_p\ (\mu_2 + \sum_{21}\sum_{11}{}^{-1}\ (\mathbf{x}_1 - \mu_1), \sum_{22}{}^{-1} - \sum_{21}\sum_{11}{}^{-1}\sum_{12})$
4. $N_q\ (\mu_2 + \sum_{21}\sum_{11}{}^{-1}\ (\mathbf{x}_1 - \mu_1), \sum_{22} - \sum_{21}\sum_{11}{}^{-1}\sum_{12})$

**Options :**
51245234549. 1
51245234550. 2
51245234551. 3
51245234552. 4

**Question Number : 14 Question Id : 51245211492 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**
**Correct Marks : 1 Wrong Marks : 0**

Multidimensional scaling is the technique of displaying

1. Low dimension data in high dimensions
2. Low dimension data in low dimensions
3. High dimension data in low dimensions
4. High dimension data in high dimensions

**Options :**
51245234553. 1
51245234554. 2
51245234555. 3
51245234556. 4

**Question Number : 15 Question Id : 51245211493 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**
**Correct Marks : 1 Wrong Marks : 0**

In a one-way MANOVA setting, Roy's test statistic, based on the measures of Within group variation (W) and Between group variation (B), is given by

1. $tr(BW^{-1})$
2. $|W| / (|W + B|)$
3. $tr[\ B\ (B + W)^{-1}]$
4. $W\ (B+W)^{-1}$

**Options :**
51245234557. 1
51245234558. 2
51245234559. 3
51245234560. 4

**Question Number : 16 Question Id : 51245211494 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**
**Correct Marks : 1 Wrong Marks : 0**

If $\ddot{\mathbf{X}}$ and S are respectively the sample mean and dispersion matrix based on a sample of size n from a normal distribution, the Hotelling $T^2$-statistics for testing the population mean $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is

1. $T^2 = n(\ddot{\mathbf{X}} - \boldsymbol{\mu}_0)S^{-1}(\ddot{\mathbf{X}} - \boldsymbol{\mu}_0)$
2. $T^2 = n(\ddot{\mathbf{X}} - \boldsymbol{\mu}_0)S(\ddot{\mathbf{X}} - \boldsymbol{\mu}_0)$
3. $T^2 = n^{-1}(\ddot{\mathbf{X}} - \boldsymbol{\mu}_0)S(\ddot{\mathbf{X}} - \boldsymbol{\mu}_0)$
4. $T^2 = n^{-1}(\ddot{\mathbf{X}} - \boldsymbol{\mu}_0)S^{-1}\ (\ddot{\mathbf{X}} - \boldsymbol{\mu}_0)$

**Options :**
51245234561. 1
51245234562. 2
51245234563. 3
51245234564. 4

**Question Number : 17 Question Id : 51245211495 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**
**Correct Marks : 1 Wrong Marks : 0**

In factor analysis, the *factor scores* give

1. the estimated values of the *common factors*
2. the estimated values of the *specific factors*
3. the estimated values of the *factor loadings*
4. the estimated values of the *elements of the covariance matrix*

**Options :**
51245234565. 1
51245234566. 2
51245234567. 3

51245234568. 4

**Question Number : 18 Question Id : 51245211496 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

In the partitioning method of clustering, with B as the between cluster variability and W as the within cluster variability, which of the following criterion is *not used* to reallocate the individuals optimally among the clustes.

1. Minimum $tr$W
2. Minimum |W|
3. Maximum $tr$BW$^{-1}$
4. Minimum $tr$BW$^{-1}$

**Options :**

51245234569. 1
51245234570. 2
51245234571. 3
51245234572. 4

**Question Number : 19 Question Id : 51245211497 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

For two sets of variables, the first canonical correlation is

1. Larger than all multiple correlations and simple correlations
2. Larger than the simple correlations, but smaller than the multiple correlations
3. Larger than the multiple correlations, but smaller than the simple correlations
4. Smaller than all multiple correlations and simple correlations

**Options :**

51245234573. 1
51245234574. 2
51245234575. 3
51245234576. 4

**Question Number : 20 Question Id : 51245211498 Question Type : MCQ Option Shuffling : No Is Question Mandatory : No**

**Correct Marks : 1 Wrong Marks : 0**

In a classification problem, the smallest value of the Total Probability of Miscalssification is referred to as

1. the Minimum Error Rate
2. the Optimum Error Rate
3. the Maximum Error Rate
4. the Average Error Rate

**Options :**
51245234577. 1
51245234578. 2
51245234579. 3
51245234580. 4

# Applied Multivariate Analysis-2

| | |
|---|---|
| **Section Id :** | 512452762 |
| **Section Number :** | 2 |
| **Section type :** | Offline |
| **Mandatory or Optional :** | Mandatory |
| **Number of Questions :** | 10 |
| **Number of Questions to be attempted :** | 10 |
| **Section Marks :** | 30 |
| **Mark As Answered Required? :** | Yes |
| **Sub-Section Number :** | 1 |
| **Sub-Section Id :** | 512452764 |
| **Question Shuffling Allowed :** | No |

**Question Number : 21 Question Id : 51245211499 Question Type : SUBJECTIVE**
**Correct Marks : 3**

Distinguish between hierarchical clustering and partitioning.

**Question Number : 22 Question Id : 51245211500 Question Type : SUBJECTIVE**
**Correct Marks : 3**

Distinguish between Euclidean and City-block Distances.

**Question Number : 23 Question Id : 51245211501 Question Type : SUBJECTIVE**
**Correct Marks : 3**

What is the apparent error rate in a classification problem ? (No derivation/expression required)

**Question Number : 24 Question Id : 51245211502 Question Type : SUBJECTIVE**
**Correct Marks : 3**

Discuss Fisher's separator for classifying two groups.

**Question Number : 25 Question Id : 51245211503 Question Type : SUBJECTIVE**
**Correct Marks : 3**

Discuss how you will choose the number of principal components to be used.

**Question Number : 26 Question Id : 51245211504 Question Type : SUBJECTIVE**
**Correct Marks : 3**

What is the contribution of the first factor to the total variability of all the variables ?

**Question Number : 27 Question Id : 51245211505 Question Type : SUBJECTIVE**
**Correct Marks : 3**

Discuss the weighted least squares method of finding factor score.

**Question Number : 28 Question Id : 51245211506 Question Type : SUBJECTIVE**
**Correct Marks : 3**

Write down the characteristic function of a multivariate normal distribution.

**Question Number : 29 Question Id : 51245211507 Question Type : SUBJECTIVE**
**Correct Marks : 3**

What will be the effect of standardization on the canonical correlation ?

**Question Number : 30 Question Id : 51245211508 Question Type : SUBJECTIVE**
**Correct Marks : 3**

Find the confidence region for the population mean vector of a multivariate normal distribution.

# Applied Multivariate Analysis-3

| | |
|---|---|
| **Section Id :** | 512452763 |
| **Section Number :** | 3 |
| **Section type :** | Offline |
| **Mandatory or Optional :** | Mandatory |
| **Number of Questions :** | 7 |

| | |
|---|---|
| **Number of Questions to be attempted :** | 5 |
| **Section Marks :** | 50 |
| **Mark As Answered Required? :** | Yes |
| **Sub-Section Number :** | 1 |
| **Sub-Section Id :** | 512452765 |
| **Question Shuffling Allowed :** | No |

**Question Number : 31 Question Id : 51245211509 Question Type : SUBJECTIVE**
**Correct Marks : 10**

For 4 units A, B, C and D with distances $d_{AB} = 5$, $d_{AC} = 8$, $d_{AD} = 4$, $d_{BC} = 3$, $d_{BD} = 5$ and $d_{CD} = 6$, form the dendogram using agglomerative hierarchical clustering method.

**Question Number : 32 Question Id : 51245211510 Question Type : SUBJECTIVE**
**Correct Marks : 10**

Given the apriori probabilities, joint densities and the misclassification costs for two groups, obtain a classification rule based on the minimum expected misclassification cost.

**Question Number : 33 Question Id : 51245211511 Question Type : SUBJECTIVE**
**Correct Marks : 10**

Discuss how you will derive the first principal component.

**Question Number : 34 Question Id : 51245211512 Question Type : SUBJECTIVE**
**Correct Marks : 10**

Describe how, in a factor model, you will estimate the factor loadings using the principal component technique.

**Question Number : 35 Question Id : 51245211513 Question Type : SUBJECTIVE**
**Correct Marks : 10**

Find the first canonical correlation in terms of the eigen-values.

**Question Number : 36 Question Id : 51245211514 Question Type : SUBJECTIVE**
**Correct Marks : 10**

If $X \sim N_p(\mu, \Sigma)$, where $\Sigma$ unknown, describe how you will test $H_0 : \mu = \mu_0$ (known) against $H_1 : \mu \neq \mu_0$.

**Question Number : 37 Question Id : 51245211515 Question Type : SUBJECTIVE**
**Correct Marks : 10**

Describe the least square method of estimating the parameters of a multivariate linear model.