Anekant Education Society's

Tuljaram Chaturchand College of

Arts, Commerce and Science, Baramati

(Autonomous)

QUESTION BANK

FOR

M.Sc SEM-II

STATISTICS

PAPER: STAT- 4203

Multivariate Analysis - 4 Credit

(With effect from June 2019)

Unit 1 No.

Question for <u>One</u> marks:

A) Choose the correct alternative of the following

1) Canonical correlation is correlation between

a) two sets of variables

b) linear combination of two sets of variables

c) a linear combination of one set of variable and a linear combination of another set of variable

d) all of the above

2) Dendogram plot is used for _________

a) Making clusters b) Analyze clusters

c) Both of them d) Discriminating observations

3) Consider the two statements.

1) Principal components are orthogonal to each other.

2) The variation explained by first principal component is maximum.

a) 1 true 2 false b) 1 false 2 true c) both true d) both false

4) Cluster analysis is used for

a) classify observations b) group observations

c) both a) and b) d) reducing variables

5) Dendogram plot is used for _________

a) Making cluster

b) Analyze clusters

c) Both of them

d) Discriminating observations

6) All of the following types of techniques is useful for data reduction except ______.

a) factor analysis b) principle component analysis

c) cluster analysis d) multivariate regression

7) A----------------graph is a graphical device for displaying clustering results..

a) dendrogram. b) Scattergram

c) Scree plot. d) Icicle diagram

8) A factor loading is:

a) empirically based hypothetical variable consisting of items which are strongly associated with each other

b) the correlation between a binomial variable and a variable which has a continuous distribution of scores.

c) the correlation of a variable with a whole score.

d) correlation coefficient between a variable and a factor (cluster of variables).

9) The goal of performing Principle Component Analysis (PCA) is

a) To explain variance – covariance structure of a large set of variables through a few linear combinations of these variables.

b) To explain only variance structure of a large set of variables through a few linear combination of these variables.

c) To explain only covariance structure of a large set of variables through a few linear combination of these variables.

d) To explain variance – covariance structure of a large set of variables through a few nonlinear combinations of these variables.

10) Which of the followings can be used to determine how many factors to take from a factor analysis:

a) Eigen values c) % of variance

b) Scree plot d) All of the above

B) State whether following statements are TRUE or FALSE.

1) Scree plot is used in factor analysis.
2) Factor analysis is extension of principle component analysis.
3) The first Principal component is the linear combination with maximum variance.
4) In factor analysis, the factors are always orthogonal.
5) Principal Component analysis is invariant of scale.
6) The scree plot is plot of ordered eigen values.
7) The principal components are always uncorrelated.
8) Factor analysis is invariant of scale.
9) Cluster analysis uses prior knowledge to group the cases or individuals.

10) Cluster analysis does not classify variables dependent and independent.
11) In factor analysis the factors are always orthogonal.
12) Canonical correlation is a particular case of multiple linear regression.
13) Non-hierarchical clustering is faster than hierarchical method.
14) The average linkage method of hierarchical clustering is preferred to the single and complete linkage methods
15) A factor loading of 0.80 means that the variable is poorly related with the factor
16) Rotation technique is used to provide a simpler and interpretable picture of the relationships between factors and variables
17) Bartlett's test is used to extract the principal components to be retained

C) Define the following terms

1) Sample mean vector
2) Dispersion matrix
3) Complete linkage
4) Orthogonal factor model.
5) Principal Component Analysis
6) Communalities
7) Factor Loadings
8) Correlation Matrix
9) Average linkage.
10) Specific Variance.
11) Factor scores.

Question for Three marks:

1) Write note on scree plot
2) Write note on dendogram plot.
3) Orthogonal factor model.
4) Principal Component Analysis

Question for <u>Five</u> marks:

1) Define Hierarchical cluster analysis. Also gives it's characteristics.
2) Write a note on Hierarchical aggregative method.
3) What is factor rotation? Show that $\Sigma = LL' + \Psi$
4) Define non-hierarchical cluster analysis. What is a feature of non hierarchical cluster analysis. Explain k-means clustering method in brief.
5) Define canonical correlation and variates. Explain relation between consecutive canonical correlations.
6) Define canonical correlation and write it's application

Question for <u>Seven</u> marks:

1) Write a note on cluster analysis. Describe a method of forming clusters from given observations by using a distance function
2) Describe canonical correlation analysis. Give one example. State the result to get the canonical variables from $\underline{X} = [X_1, X_2, \ldots, X_p]'$.
3) Show that the principal components are uncorrelated and have variances equal to the eigen values of Σ.

Question for <u>Ten</u> marks:

1) State and prove that any two properties of Principal Component Analysis.

2) Determine the principal components Y_1, Y_2 and Y_3 for the covariance matrix

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

   Also, calculate the proportion of total population variance explained by the first principal component.

3) The distance between pairs of five item are as follows.

$$
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}
\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{bmatrix} 0 & & & & \\ 4 & 0 & & & \\ 6 & 9 & 0 & & \\ 1 & 7 & 10 & 0 & \\ 6 & 3 & 5 & 8 & 0 \end{bmatrix} \end{array}
$$

Cluster the five items using the single linkage and complete linkage hierarchical method.

Draw the dendrograms.

4) The following data matrix contains data on test scores, with x_1=score on first test, x_2=score on second test, and x_3= Total score on the two tests:

$$
X = \begin{pmatrix} 12 & 17 & 29 \\ 18 & 20 & 38 \\ 14 & 16 & 30 \\ 20 & 18 & 38 \\ 16 & 19 & 35 \end{pmatrix}
$$

Obtain mean and sample covariance matrix

5) What are Principal Components? What purpose do they serve? Determine the first principal component and its variance when

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

6) Define orthogonal factor model with an assumptions. Also explain the terms factor loading and specific variance.

7) Let $\underline{X}$ of order P×1 has mean $\underline{\mu}$ and covariance matrix Σ. Obtain K-principal components of standardized vector $\underline{Z}$ of $\underline{X}$.

8) Define canonical correlations and canonical variables. Derive the first canonical correlation coefficient and the corresponding canonical variables. Also show that first canonical correlation explains maximum correlation.

9) Consider the matrix of distances.

$$
\begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 10 & 3 & 0 & \\ 6 & 5 & 4 & 0 \end{bmatrix}
$$

do the cluster analysis by using single and complete linkage method. Also draw the dendograms and commend on it.

10) Suppose the random variables X1, X2 and X3 have the covariance matrix

$$\Sigma = \begin{bmatrix} 5 & -4 & 0 \\ -4 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Obtain principle components. Prepare scree plot and determine the number of components to be retained. What are the correlation coefficients between the principle components and the X variables? What conclusion can you draw from this?

11) Let X ~ Np(μ,Σ). A random sample of size n = 5 is collected and shown below:

$$X = \begin{bmatrix} 9 & 12 & 3 \\ 2 & 8 & 4 \\ 6 & 6 & 0 \\ 5 & 4 & 2 \\ 8 & 10 & 1 \end{bmatrix}$$

Compute:

i) Mean vector for the data set

ii) Covariance matrix of X

iii) Correlation matrix of X

12) Suppose the random variable X1, X2 and X3 have the covariance matrix.

$$\Sigma = \begin{bmatrix} 5 & -4 & 0 \\ -4 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Assume m=1 factor model calculate the loading matrix L and matrix of specific variances Ψ using principle component solution method. Calculate communalities and interpret these equations. What proportion of the total variance is explained by the common factor.

13) Determine the population principal components Y_1 and Y_2 for the covariance matrix given below and also calculate the proportion of the total population variance explained by the first principal component. $\Sigma = \begin{bmatrix} 5 & 2 \\ 3 & 2 \end{bmatrix}$

14) Differentiate hierarchical and non-hierarchical cluster analysis. Explain k-means clustering method in brief.

Unit No. 2

Question for <u>one</u> marks:

A) Choose the correct alternative of the following

1) Let $X_1, \dots X_n$ a random sample of size n from a p – variate normal distribution with mean μ and positive definite covariance matrix ∑. Choose the correct statement

   a) $(X_1 - \mu)' \Sigma^{-1} (X_1 - \mu)$ has χ_1^2 b) $\underline{\overline{X}}\,\underline{\overline{X}}'$ has Wishart distribution with P d.f.

   c) $\sum_{i=1}^{n}(xi-\mu)(xi-\mu)'$ has Wishart distribution with n d.f.

   d) $X_1 + X_2$ & $X_1 - X_2$ are independently distribution

2) Let X be a P × 1 random vector such that X ~ Np (0, ∑) where rank (∑) = p which of the following is truer.

   a) $E(X'\Sigma^{-1}X) = 2P,\ V(X'\Sigma^{-1}X) = 2P$ b) $E(X'\Sigma^{-1}X) = 2P,\ V(X'\Sigma^{-1}X) = P$

   c) $E(X'\Sigma^{-1}X) = P,\ V(X'\Sigma^{-1}X) = P$ d) $E(X'\Sigma^{-1}X) = P,\ V(X'\Sigma^{-1}X) = 2P$

3) If $\underline{X}$ ~ Np (μ, ∑) then $\underline{X}$ +$\underline{d}$ follows

   a) Np (μ, ∑) b) Np ($\underline{\mu}$ + $\underline{d}$, ∑) c) Np ($\underline{\mu}$,∑+$\underline{d}$) d) Np ($\underline{\mu}$+$\underline{d}$, ∑+$\underline{d}$)

   i) Let A be the matrix of constants of order m×p and given that rank(A)=m then the probability distribution of A$\underline{X}$ is.

      a) $N_m(A\mu, A\Sigma A')$ b) $N_p(A\mu, A\Sigma A')$ c) $N_m(A\mu, \Sigma')$ d) $N_m(\mu, A\Sigma A')$

4) If $\underline{X}$~$N_p(\underline{\mu},\Sigma)$ then $\underline{X}$+$\underline{d}$ follows

   a) $N_p(\underline{\mu},\Sigma)$ b) $N_p(\underline{\mu}+\underline{d}, \Sigma)$ c) $N_p(\underline{\mu}, \Sigma + \underline{d})$ d) $N_p(\underline{\mu}+\underline{d}, \Sigma + \underline{d})$

5) Let A be the matrix of constants of order m×p and given that rank(A)=m then the probability distribution of A$\underline{X}$ is.

   a) $N_m(A\mu, A\Sigma A')$ b) $N_p(A\mu, A\Sigma A')$ c) $N_m(A\mu, \Sigma')$ d) $N_m($

6)The exponent of multivariate normal density function follows-

a) F-distribution b) t- distribution
c) Chi – square distribution d) None of these

7) If (X_1, X_2 , X_3)' follows $N_3(\underline{\mu}, \Sigma)$ where $\underline{\mu}' = [1 \quad 2 \quad 3]$ and $\Sigma = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ then X_1 – X_2 + X_3 follows

a) $N\left(2 \quad \frac{5}{2}\right)$ b) $N(0 \quad 2)$ c) $N(2 \quad 2)$ d) $N(2 \quad 3)$

8) When referring to a multivariate random variable X, then matrix Σ refers to the ____________ matrix.

a) Correlation b) Coefficients
c) Variance-covariance d) Variance-correlation

9) Let h be any fixed row vector of order n×1 and independently distributed with matrix D of order n×n then the what is the conditional distribution of $\frac{\underline{h}'D\underline{h}}{\underline{h}'\Sigma\underline{h}}$?

a) χ^2_n b) χ^2_{n-1}
c) χ^2_{n-p} d) χ^2_{n-p-1}

B) State whether following statements are TRUE or FALSE.

1) If $\underline{X} \rightarrow N_p(\underline{\mu}, \Sigma)$ then maximum likelihood estimator of $\underline{\mu}$ *and* Σ is $\overline{X}$ and S
2) If $\underline{X}$ follows multivariate normal distribution then it's marginal distribution also follows multivariate normal distribution
3) The two vectors X_1 and X_2 are independent iff cov(X_1,X_2)=0.
4) The variance-covariance matrix of a bivariate random variable is always symmetric.
5) The matrix S/(n-1) is a biased estimate of the variance-covariance matrix of the multivariate population.
6) The probability distribution is said to be multivariate normal iff every linear combination of $\underline{X}$ is univariate Normal.
7) If $X_i \sim N(\mu_i, \sigma_i^2)$ then $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$

8) The probability distribution is said to be multivariate normal iff every linear combination of $\underline{X}$ is univariate Normal.
9) If $X_i \sim N(\mu_i, \sigma_i^2)$ then $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$

10) If $\underline{X} \sim N_p(\underline{\mu},\Sigma)$ then the moment generating function of

11) The probability distribution is said to be multivariate normal iff every linear combination of $\underline{X}$ is univariate Normal.

12) If $\underline{X} \sim N_p(\underline{\mu},\Sigma)$ then the moment generating function of

$$Y=\left(\underline{X}-\underline{\mu}\right)\Sigma^{-1}\left(\underline{X}-\underline{\mu}\right) \text{ is } e^{\underline{\mu}'\underline{t}-\frac{1}{2}\underline{t}\Sigma\underline{t}}$$

13) If $X_i \sim N(\mu_i,\sigma_i^2)$ then $\underline{X} \sim N_p(\underline{\mu},\Sigma)$

14) If $\underline{X} \to N_p\left(\underline{\mu},\Sigma\right)$ then $\underline{X}+d \to N_p\left(\mu+d,\Sigma\right)$

Question for <u>Two</u> marks:

Define the following terms.

1) Non-singular multivariate normal distribution.

2) Sample generalized variance.

3) Singular multivariate normal

4) Moment generating function of multivariate normal distribution.

Question for <u>Five</u> marks:

1) State and prove necessary and sufficient condition for the two multivariate normal vectors to be independent.
2) Show that a random vector $\underline{X}$ follows multivariate normal distribution iff every linear combination of $\underline{X}$ is unvariate normal distribution.
3) Let X have covariance matrix

$$\Sigma=\begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$$

   a) Determine ρ and $V^{1/2}$
   b) Multiply your matrices to check the relation $V^{1/2}\rho V^{1/2}=\Sigma$
4) Derive the first two moments of a multivariate normal distribution.
5) Define non-singular multivariate normal distribution. write down the density function for the following characteristic function

$$\phi = \exp\{i(t_1 + 2t_2 + 5t_3) - 1/2(t_1^2 - 4t_1t_2 + 5t_2^2 + t_3^2)\} \;;\; \underline{t} \in R^3$$

6) Let $\underline{X}$ have covariance matrix $\Sigma = \begin{matrix} 25 & -2 & 4 \\ & 4 & 1 \\ & & 9 \end{matrix}$

Find the correlation between X_1 and $\frac{X_2}{2} + \frac{X_3}{2}$

7) Define orthogonal factor model with an assumptions.

8) Show that every linear function of multivariate normal vector is normally distributed.

9) Let $\underline{X} \sim N_3(\underline{\mu}, \Sigma)$ derive the characteristic function of X. Hence find E(X) and Var(X).

10) Let $X \to N_2(\mu, \Sigma)$ with covariance matrix $\Sigma = \begin{bmatrix} 9 & 6 \\ 6 & 25 \end{bmatrix}$ Obtain it's correlation matrix.

11) Suppose $\underline{X} \to N_2((\underline{\mu}, \Sigma)$ with $\underline{\mu} = (2,2)'$ and $\Sigma = I_2$ consider $A = (1,1)$ and $B = (1,-1)$.

Verify whether $A\underline{X}$ and $B\underline{X}$ are independent.

12) Let $\underline{X} \sim N_3(\mu, \Sigma)$ where $\underline{\mu}$ = (-3 1 4)' and $\Sigma = \begin{matrix} 1 & -2 & 0 \\ & 5 & 0 \\ & & 2 \end{matrix}$

i) Check whether X_2 and $X_2 - \frac{5}{2} X_2 - X_3$ are independent or not.

ii) Find the conditional probability distribution of $X_1 + 3X_3$ given $X_2 = 5$.

13) Let $\underline{X} \sim N_p(\mu, \Sigma)$ Obtain sufficient statistic for μ and Σ where (μ, Σ) are unknown.

Question for <u>Eight</u> marks:

1) State and prove relation between multivariate normal and χ^2 distribution.

2) Identify the sampling distribution of maximum likelihood estimators of multivariate normal distribution.Show that these maximum likelihood estimators are independently distributed.

3) Derive moment generating function and characteristic function for multivariate normal distribution.

Question for <u>Ten</u> marks:

4) Let $\underline{X} \sim Np(\mu, \Sigma)$

   i) Obtain characteristics function of $\underline{X}$

   ii) Hence find the distribution of ($A\,\underline{x} + \underline{b}$) where $A_{k \times p}$ matrix of constants with $k \leq p$ and $\underline{b}_{k \times 1}$ vector of constant.

5) Let $\underline{X} \sim N_p(\mu, \Sigma)$ with $\mu' = [2, -1, 1]$ and $\Sigma = \begin{bmatrix} 2 & 2 & 2 \\ & 4 & 3 \\ & & 3 \end{bmatrix}$

   i) Find the distribution of X_1

   ii) Find the distribution of $[X_1+3X_2, 3X_3]'$

   iii) Find the conditional distribution of X_1+3X_2 given $3X_3=2$

   iv) Find a 2×1 vector $\underline{b}$ such that X_2 and $X_2 - \underline{b}' [X_1, X_3]'$ are independent.

6) Define multivariate normal distribution. Write probability density function of multivariate normal distribution. Derive the characteristic function of multivariate normal distribution from its probability density function.

7) Derive the sampling distribution of the maximum likelihood estimators of the parameter of the multivariate normal distribution.

8) If $\underline{X} \to N_3(\underline{\mu}, \Sigma)$ where $\mu = (-3 \quad 1 \quad 4)'$ and

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Check whether

i) X_1 and X_2 are independent.
ii) X_2 and X_3 are independent.
iii) (X_1, X_2) and X_3 are independent.
iv) $\frac{X_1 + X_2}{2}$ *and* X_3 are independent.
v) X_2 *and* $X_2 - \frac{5}{2}X_1 - X_3$ are independent.

9) Let $\underline{X}$ is distributed as $N_p(\underline{\mu},\Sigma)$ where $\underline{\mu}$ = (2, -1, 1) and have covariance matrix

$$\Sigma = \begin{pmatrix} 5 & -2 & 0 \\ & 4 & 0 \\ & & 9 \end{pmatrix}$$

i) Check whether X_2 and $X_1 + 2X_2$ are independent.
ii) Find the conditional probability distribution of $X_1 + 3X_3$ given $X_2 = 5$.

8) Give that X ~ $N_4(\underline{\mu}, \Sigma)$ distribution with $\underline{\mu}$ = (4, 3, 2, 1)' and

$$\Sigma = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 3 & 2 & 0 \\ 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$$

i) Obtain distribution of $A\underline{X} + \underline{b}$ where $A = \begin{bmatrix} -1 & 2 & -4 & 3 \\ 1 & 2 & 1 & 3 \end{bmatrix}$ $\underline{b} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

ii) Conditional distribution of $(X_1 + X_2 \mid X_3 = 14, X_4 = 12)$

10) Let X1, X2, …. Xn be a random sample of size n from Np (μ, Σ). Find maximum likelihood estimator of $\underline{\mu}$ and Σ based on these n observations.

11) Let X_1, X_2, X_3, X_4 and X_5 be independent and identically distributed random vectors with mean vector μ and covariance matrix Σ. Find the mean vector and covariance matrices for each of the two linear combinations of random vectors.

$$\frac{1}{5}X_1 + \frac{1}{5}X_2 + \frac{1}{5}X_3 + \frac{1}{5}X_4 + \frac{1}{5}X_5$$

$$X_1 - X_2 + X_3 - X_4 + X_5$$

In terms of μ and ∑. Also obtain the covariance between the two linear combinations of random vectors.

Unit No.3

Question for <u>One</u> marks:

A) Choose the correct alternative of the following

1) Let Y_1, Y_2, Y_3, Y_4 be i.i.d. standard normal variables. Which of the following has Wishart distribution with 2 d.f. ?

   a) $\begin{bmatrix} Y_1^2+Y_2^2 & Y_2^2+Y_3^2 \\ Y_2^2+Y_3^2 & Y_3^2+Y_4^2 \end{bmatrix}$ b) $\begin{bmatrix} Y_1^2 & Y_2^2 \\ Y_3^2 & Y_4^2 \end{bmatrix}$

   c) $\begin{bmatrix} Y_1^2+Y_2^2 & 0 \\ 0 & Y_3^2+Y_4^2 \end{bmatrix}$ d) $\begin{bmatrix} Y_1^2+Y_2^2 & Y_1Y_3+Y_2Y_4 \\ Y_1Y_3+Y_2Y_4 & Y_3^2+Y_4^2 \end{bmatrix}$

2) Let X_{pxn} is a random sample from Np (μ, ∑). The distribution of $s=\left(x-\bar{x}E_{1n}\right)\left(x-\bar{x}E_{1n}\right)'$ is

   a) w_p (n, H ∑ H') b) w_p(n, ∑) c) a) w_m (n, H ∑ H') d) a) w_m (n, ∑)

3) Let $X_{p\times n}$ is a random sample from $N_p(\underline{\mu}, \sum)$. The distribution of $S = (X - \bar{X}\,E_{1n})(X - \bar{X}\,E_{1n})'$ is

   a) $w_p(n-1,\Sigma)$ b) $w_p(n,\Sigma)$ c) $w_p(n+1,\Sigma)$ d) χ_n^2

4) In Hotelling T2 statistics distribution of <u>u</u> is

   a) Singular multivariate normal b) Non-singular multivariate normal

   c) Wishart d) Uniform

5) Let $X_{p\times n}$ is a random sample from $N_p(\underline{\mu}, \sum)$. The distribution of

   a) $w_p(n-1,\Sigma)$ b) $w_p(n,\Sigma)$ c) $w_p(n+1,\Sigma)$ d) χ_n^2

6) In Hotelling T2 statistics distribution of <u>u</u> is

   a) Singular multivariate normal b) Non-singular multivariate

   c) Wishart

7) If A~ $W_p(n-1, \Sigma)$ then the distribution of $\underline{l}' A \underline{l}$ where $\underline{l}$ is a known vector would be

a) $(\underline{l}'\Sigma\underline{l})\chi^2_{n-1}$ b) Wishart c) $n\chi^2_{n-1}$ d) χ^2_{n-1}

8) Hotelling T^2 test can be used for testing:

a) a general linear hypothesis in the mean vector

b) an arbitrary hypothesis about the mean vector

c) simultaneous hypothesis about the mean vector

d) two-sided hypothesis in the mean vector

9) Let $X_{p\times n}$ is random sample from $N_p(\underline{\mu}, \Sigma)$. The distribution of $S = (X - \underline{X}E_{1n})(X - \underline{X}E_{1n})'$ is

a) $W_P(n-1, \Sigma)$ b) $W_P(n, \Sigma)$

c) $W_P(n+1, \Sigma)$ d) χ^2_n

10) The profile described by p dimensional vectors $\underline{X}$ and $\underline{Y}$ will parallel if

a) H_0: $E_{1p}\, \underline{\mu} = E_{1p}\, \underline{\gamma}$ b) $H_0 : \mu_i - \mu_{i-1} = \gamma_i - \gamma_{i-1}$

c) H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_p$ d) all of the above

B) State whether following statements are TRUE or FALSE.

1) Chi – square distribution is particular case of Wishart distribution

2) If A ~ $W_p(n-1, \Sigma)$ with n > p and then $\underline{X}$ ~$N_p(\underline{\mu},\Sigma)$ $|\Sigma| > 0$ and let $\underline{U}$ be distributed independently of A.Then the $\frac{u\Sigma^{-1}\underline{U}}{\underline{u}' A^{-1} \underline{u}}$ follows χ^2_{n-1}

3) Hotelling T^2 test is generalization of F-test.

Question for <u>Five</u> marks:

1) If A ~ $W_p(n, \Sigma)$ then $|A| / |\Sigma|$ is distributed as product of p independent chi-square variate with d.f.n, n-1,…. N – p + 1.
2) Define Hotelling T^2 and show that it is invariant under non-singular linear transformation.
3) Explain how Hotelling's T^2 statistic is used to test the problem of symmetry
4) Write note on Rao's union-intersection principal for obtaining Hotelling's T^2 statistic.
5) Let D be the symmetric positive definite matrix with $w_p(n, \Sigma)$ distribution and $\underline{h}$ be any random vector then find the distribution of $\frac{h' D h}{h' \Sigma h}$

6) Let $\underline{Z} \sim N_p(\underline{0}, \Sigma)$ and C has the Wishart distribution then $\frac{n-p+1}{p} X' C^{-1} X \sim F_{p,n-p+1}$

derive the distribution of T^2 under the null hypothesis $H_0: \mu = \mu_0$

7) Describe Characteristic function of Wishart distribution.

8) Write a short note on distribution of sample generalized variance.

9) Define Hotelling T^2.State with two examples the uses of Hotelling's T^2

10) Explain how Hotelling's T^2 stastistic is used in profile analysis.

Question for <u>Eight</u> marks:

1) Define Hotelling's T^2 statistic and state it's null distribution. Discuss the use T^2 as a test statistic for testing $H_0: \underline{\mu} = \underline{\mu}_0$ where $\underline{\mu}$ is the mean vector of a multivariate normal population.
2) Define Wishart Matrix. State and prove any two important properties of Wishart distribution.
3) Evaluate Hotelling's T^2 for testing $H_o: \underline{\mu}' = [9 \;\; 5]$ using the data

$$\mathbf{X} = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}$$

Specify the distribution of T^2 for this situation. Test H_o at α=0.05 level. What conclusion do you reach?

Question for <u>Ten</u> marks:

4) Define Wishart distribution. Obtain the characteristics function of canonical wishart distribution.
5) Define Wishart distribution. State and prove the additive property of canonical wishart distribution.
6) Define Bartlett's decomposition of Wishart matrix. Also find distribution of B.
7) Define the Wishart distribution and describe how it generalizes the chi-square distribution. How is the Wishart distribution related to the multivariate normal distribution? State the density function of the Wishart distribution and identify its parameter.
8) Define Bartlet's decomposition of Wishart matrix in B canonical case. Find distribution of matrix B.

Unit No.4

Question for <u>One</u> marks:

B) State whether following statements are TRUE or FALSE.

i) The number of correctly classified cases in discriminant analysis is given by the hit ratio.
ii) The discriminant function allows us to determine group membership.
iii) The discriminant function allows us to determine group membership.

Question for Two marks:

Define the following terms.

1) Mahalnobis D^2 statistic
2) Misclassification error
3) Fisher discriminnt function

Question for Five marks:

1) Derive a test for the problem of deciding a specified linear function is a good discriminate function.
2) Roy's union – intersection principal to test H_0: $\underline{\mu} = \underline{0}$ against H_1: $\underline{\mu} \neq \underline{0}$.
3) Derive relationship between Hotelling T^2 and Mahalonobis distance.
4) Derive test of equality of mean vectors of full multivariate normal population with same but unknown variance covariance matrix.
5) Derive test of equality of dispersion matrix of full multivariate normal population
6) Derive Fisher's Discriminant function to discriminate between two populations. State the assumptions clearly.
7) Rsearcher has enough data to illustrate density function f1(x) and f2(x) Given C(1|2)=5 units, C(2|1)=10 units.It is known that about 20% of all objects belong to π_2 i.e. prior probability are P1=0.8, P2=0.2.
   Using the above information derie the classification regions for two populations.

Question for Eight marks:

1) Define the problem of classifying an observation into one of the two populations. Define linear discriminant function and probabilities of misclassification.
2) Derive Fisher's Discriminant function to discriminate between two populations. State the assumptions clearly.

3) Suppose we have $\pi_1 = \pi_2 = \frac{1}{2}$ and two bivaraite normal groups with common covariance matrix $\Sigma = \begin{bmatrix} 6 & 2 \\ 2 & 1 \end{bmatrix}$ and respective means $\mu_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ Derive the discriminant function for assigning a new observation X to one of these two groups. Find an expression for the probability of misclassifying an individual using this rule when C(1|2)=C(2|1)=1.If X=$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ to which group would you assign this individual?

4) Consider the two data sets

$$X_1 = \begin{bmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{bmatrix} \quad and \quad X_2 = \begin{bmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{bmatrix}$$

$$for\ which\ \overline{X_1} = \begin{bmatrix} 3 \\ 6 \end{bmatrix},\ \overline{X_2} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}\ and\ S_{pooled} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

i) Calculate the linear discriminant function.

ii) Classify the observation $x_0' = [2 \quad 7]$ as population π_1 or population π_2.