## Chennai Mathematical Institute Multivariate Statistical Analysis Mid Term Test

Time: 60 Minutes

Max marks: 20

## Choose the Best possible answer

- 1. An advantage of using an experimental multivariate design over separate univariate designs is that using the multivariate analysis is:
  - A. Allows you to look at more complex relationships than does univariate strategy
  - B. Provides a more powerful test of hypotheses
  - a) Both A and B holds
  - b) A holds but not B
  - c) B holds but not A
  - d) Both A and B do not hold
- A 2. If X has a multivariate Normal distribution, then
  - A. Linear combinations of X are normally distributed
  - B. All subsets of the components of X have a MVN distribution
  - C. All the conditional distributions of the components are not MVN
  - a) A and B are true
  - b) A and C are true
  - c) B and C are true
  - d) A,B, and C are true
- C 3. Let X and Y be independent standard normal random variables. Define the random variable Z by Z = X if XY > 0; Z = -X if XY < 0
  - A. Z is a standard normal random variable
  - B. Random Vector (Y, Z) is not jointly normal
  - a) A is true but not B
  - b) B is true but not A
  - c) Both A and B are true
  - d) Both A and B are not true
- 4. Examine the following statements
  - A. If the components of MVN are not independent, then the vector may not be jointly Normal
  - B. Uncorrelated jointly normal random variables are independent
  - a) A holds but not B
  - b) B holds but not A
  - c) A and B do not hold
  - d) A and B holds

- 5. If Y follows MVN with  $(\mu, \Sigma)$ , then Y'AY  $\sim \chi^2(r, \gamma)$  where  $\gamma = \mu'A\mu$  and r(A) = r iff
- A.  $A\Sigma A = \Sigma$  and  $A\Sigma$  is idempotent B.  $A\Sigma A = \Sigma$  and  $A^{-1}\Sigma$  is idempotent
  - a) A holds
  - b) B holds
  - c) A and B holds
  - d) A and B does not hold
- 6. Examine the following properties on p variate MVN
- A. Invariant under linear transformation
  - B. Uncorrelated sub-vectors are independent
  - a) A and B holds
  - b) A holds but not B
  - c) B holds but not A
  - d) A and B hold under certain other conditions
- 7. The MD between Y and  $\mu$  in the metric of  $\Sigma$  is  $D(Y,\mu) = [(Y \mu) \Sigma^{-1} (Y \mu)]^{1/2}$
- A. Invariant under appropriate linear transformations
   B. Adjusted Euclidean distance in the metric of reciprocal variance
  - a) A holds and not B
  - b) B holds and not A
  - c) A and B holds together
  - d) A and B do not hold
- 8. Q-Q plot can be used to
- A. order Mahalanobis distances versus estimated quantiles for a sample of size n from a chi-squared distribution with p degrees of freedom for evaluating multivariate normality
  - B. points on the right side of the plot for which the Mahalanobis distance is notably greater than the chi-square quantile value exhibit the presence of outliers.
    - a) A is true
    - b) B is true
    - c) A and B are true
    - d) A and B are not true
- 9. Examine the following statements in PCA
  - A. The construction of principal components does not require that the variables in Y have a normal multivariate distribution.
  - B. Construction of PCA can be carried out through SVD of the data matrix
  - PCA minimizes the perpendicular distance between a data point and the principal component
  - a) A and B are true
  - b) A and C are true
  - c) B and C are true
  - d) A, B, and C are true

## 10. Examine the following statements in PC



- A. PC's are invariant to linear transformation to the original variables
- B. PC's are not sensitive to the presence of outliers F
- a) Both A and B are true
- b) A alone is true
- c) B alone is true
- d) A and B are untrue

## 11. Examine the following statements on Mahalanobis distance

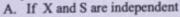


- A. MD is effectively a multivariate equivalent of Euclidean distance
- B. MD could be used for classification and Outlier detection
- C. MD is not possible for highly correlated variables F
- a) A,B and C are true
- b) A,B is true but not C
- c) A,C is true but not B
- d) B,C is true but not A
- 12. Suppose X1, ..., Xn are i.i.d. Np( $\mu$ ,  $\Sigma$ ); then the sample mean and sample variance are independent, with



A. 
$$\sqrt{n(\bar{X}-\mu)} \sim \text{Np}(0, \Sigma)$$
,  
B.  $(n-1)S \sim \text{Wp}(n-1, \Sigma)$ .

- a) A is true
- b) B is true
- c) A and B are true
- d) A and B are not true
- 13. let  $X \sim Np(\mu, \Sigma)$ ,  $mS \sim Wp(m, \Sigma)$ . then  $T^2_p(m) = (\overline{X} \mu)$  'S · 1  $(\overline{X} \mu)$  and  $[(m - p + 1)/(mp)] T^{2}_{p}(m)$  follows Fp,m-p+1:

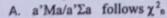




B. If X and S are independent



- a) A is true
- b) B is true
- c) A and B are true
- d) A and B are not true
- 14. If  $M \sim Wp(n; \Sigma)$  and a  $\in \mathbb{R}^p$  is such that a  $\Sigma \neq 0$  and n > p-1 then





- B. a', Σ-1a/a'M-1 a follows χ 2 n-p+1
  - a) A is true b) B is true
  - c) A and B are true

  - d) A and B are not true

- 15. If A is idempotent of rank r then
  - A. If y is Np(O, I), then y'Ay is  $\chi 2(r)$
  - B. y is Np( $\mu$ ,  $\sigma^2$ 1), then y'Ay/ $\sigma^2$  is  $\chi^2$ (r,  $\mu'A\mu/\sigma^2$ )



- a) A is true
- b) B is true
- c) A and B are true
- d) A and B are not true
- 16. Examine the following
- A. Let B is a  $k \times p$  matrix of constants, A is a  $p \times p$  symmetric matrix of constants, and y is distributed as Np( $\mu$ ,  $\Sigma$ ). Then By and y'Ay are independent if and only if B $\Sigma$ A = O
  - B. Let A and B be symmetric matrices of constants. If y is Np (  $\mu$ ,  $\Sigma$ ), then y 'Ay and y 'B y are independent if and only if A  $\Sigma$  B = O.



- a) A is true
- b) B is true
- c) A and B are true
- d) A and B are not true
- 17. If y us  $Np(\mu, \Sigma)$ , then
  - A.  $E(Y'AY) = tr(A\Sigma) + \mu'A\mu$
  - B.  $var(y'Ay) = 2tr[(A\Sigma)^2] + 4\mu'A\Sigma A\mu$ .
  - C.  $cov(y, y'Ay) = 2\Sigma A\mu$ .



- a) A and B are true
- b) A and C are true
- c) B and C are true
- d) A, B and C are true
- 18. Examine the following for Mahalanobis distance:
  - A. a measure that quantifies the dissimilarity between two data points in a multidimensional space, considering the covariance structure of the data. T
  - B. It takes into account the correlations between variables, making it suitable for datasets where variables are interrelated. T
  - C. It is not scale-invariant, meaning it is affected by the scaling of variables F



- a) A and B are true
- b) A and C are true
- c) B and C are true
- d) A, B, and C are true
- 19. Let X follows Np( $\mu$ ,  $\Sigma$ ), and Y =  $\Sigma^{-1/2}$  (X  $\mu$ ) then
  - A. Y follows Np(0, I p) T
  - B. Y follows Np(μ, Σ) F
    - a) A is true
- b) B is true
  - c) A and B are true
  - d) A and B are not true

20. If  $X \sim N_3(\mu, \Sigma)$  with  $\mu = [-2 \ 3 \ 1]'$  and

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

Then marginal distribution of  $X_1$  and  $X_3$  and  $[X_1 \quad X_3]$ ' are

A. 
$$X_1 \sim N(2, 2)$$

B. 
$$X_2 \sim N(1.4) T$$

A. 
$$X_1 \sim N(2, 2) T$$
  
B.  $X_2 \sim N(1,4) T$   
C.  $X_2 \sim [X_1, X_3] \sim$ 

$$N_2\left(\begin{pmatrix} -2\\1 \end{pmatrix}, \begin{pmatrix} 2&0\\0&4 \end{pmatrix}\right)$$
.  $\top$ 

- a) A and B are but not C
- b) B and C are true but not A
- c) A and C are true but not B
- d) A,B, and C are true