ROHAN KARTHIKEYAN
MDS202226

Chennai Mathematical Institute

Multivariate statistical Analysis

Mid Term Test

Time : 60 Minutes | Max marks: 20

**Choose the Best possible answer**

1. (A) 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 ✗

2. (C) If X = (X1, . . . , Xn) has a multivariate Normal distribution, then
   A. Each component Xi has a Normal distribution.
   B. Every subvector of X has a multivariate Normal distribution.

   a) A is true but not B
   b) B is true but not A ✗
   c) Both A and B are true ✓
   d) Both of them are not true ✗

3. (C) Let X and Y be independent standard normal random variables. Define the random variable Z by

$$Z = X \text{ if } XY > 0 \;\&\; Z = -X \text{ if } XY < 0$$

   A. Z is a standard normal random variable
   B. Random Vector (Z,Y) 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. (A) Let $X \sim N(0, \sigma^2 I)$ and let $A_1$ and $A_2$ be symmetric idempotent matrices, then $X'A_1X$ and $X'A_2X$ are independent, if $\boxed{\text{Cochran's theorem}}$?
   A. $A_1A_2 = A_2A_1 = 0$.
   B. $A_1A_2 = A_2A_1 = I$

   a) A is true
   b) B is true
   c) A and B are true
   d) A and B are not true ? (best)

$\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$

$a^2+b^2=a$
$b(a+c)=b$
$b^2+c^2=c$

$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

$aa'+bb'$
$ab'+bc'=0$
$ba'+cb'=0$
$bb'+cc'$

$\mathbb{E}(X^TA_1X - tr(\sigma^2A_1))(X^TA_2X - tr(\sigma^2A_2)) = 0$

5. Let X ~ N(0, In). Then X′AX ~ χ 2 (k), and A has rank k, if and only if .
C
   A. A is symmetric
   B. A is idempotent

   a) B holds but not A
   b) A holds but not B
   c) Both A and B holds ✓
   d) A and B do not hold

6. Examine the following statements
D
   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 ⟹ For eg. consider the bivariate density:

   Ind ⇒ Uncorrelated

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp\{-q/2\}; \quad -\infty < x < \infty,\ -\infty < y < \infty$$

with $q = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$

→ If $\rho = 0$, we can write the above density as a product of two normal densities
⇒ RVs 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

7. If Y follows MVN with (μ,Σ) then the quadratic form Y'AY follows $\chi^2$ (r,γ) where γ = μ'Aμ and r(A) = r if and only if
C
   A. AΣA = Σ
   B. AΣ is idempotent

   a) A holds
   b) B holds
   c) A and B holds ✓
   d) A and B does not hold

8. Examine the following properties on p variate MVN
A
   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 holds under certain other conditions

9. For **Y** ~ *N(μ,Σ_)* the following result holds.
C

   a) E(Y'AY) = tr(AΣ)
   b) E(Y'AY) = μ'_Aμ ?
   c) E(Y'AY) = tr(AΣ) + μ'Aμ (best)
   d) E(Y'AY) = tr(AΣ + μ'Aμ)

$$E[(Y-\mu)^T(Y-\mu)] = \Sigma$$
$$E[Y^TY - Y^T\mu - \mu^TY + \mu^T\mu] = \Sigma$$
$$E[Y^TY - 2Y^T\mu] = \Sigma - \mu^T\mu$$
$$E[Y^TY] - 2\mu^T\mu = \Sigma - \mu^T\mu$$
$$\therefore E[Y^TY] = \Sigma + \mu^T\mu$$

10. The Mahalanobis distance between Y and $\mu$ in the metric of $\Sigma$ is $D(Y,\mu) = [(Y - \mu)\ \Sigma^{-1}$
C $(Y - \mu)]^{1/2}$

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

11. Let **Y** and **Q** be independent random variables where **Y** $\sim Np(\mu,\Sigma)$ and **Q** $\sim Wp(n,\Sigma)$, and n >
D p. Then Hotelling's $T^2$ statistic $T^2 = n\mathbf{Y}'\mathbf{Q}^{-1}\mathbf{Y}$ has a distribution proportional to

a) Hotelling Distribution
b) Multivariate Normal Distribution
c) Noncentral $\chi^2$ distribution
d) Noncentral F distribution ✓

12. Let $X = [X1,X2]'$ be a multivariate normal distribution with mean $\mu = [1\ 2]'$ and $\Sigma =$
B

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

then the random variable $Y = X1 + X2$ is MVN with mean and variance respectively as

a) 2 and 5
b) 3 and 7 ✓
c) 4 and 3
d) 5 and 2

13. Examine the following statements
C
A. Independent => uncorrelated ( always holds)
B. Uncorrelated => Independent (for MVN)

a) A is true but not B ✓
b) B is true but not A
c) A and B are true
d) A and B are not true

14. Let X1,X2,.... Xn be a random sample from $N_p\ [\mu,\Sigma]$ then
C
A. Sample mean has a normal distribution and sample covariance matrix has Wishart distribution
B. Both the sample distributions are independent

a) A holds but not B
b) B holds but not A
c) A and B holds always
d) A and B holds conditionally ✓

15. Q-Q plot can be used to picture the Mahalanobis distances for the sample and is useful in
A
a) evaluating multivariate normality and outliers ✓
b) evaluating chi-square distribution and outliers
c) evaluating Wishart distribution and outliers
d) detecting outliers only

16. Examine the following statements in PCA

A

A. Construction of principal components do not require that the variables in **Y** have a multivariate normal distribution.

B. Statistical procedure for transforming a MVN distribution into a set of independent univariate normal distributions Simply $\Sigma^{-1/2}(Y-\mu)$ suffices

a) Both A and B are true
b) A alone is true ? ✓
c) B alone is true
d) A and B are untrue

17. Examine the following statements in PC

C

A. PC's are invariant to linear transformation to the original variables If we standardize the data, PCs change!

B. PC's are not sensitive to the presence of outliers

a) Both A and B are true ✗
b) A alone is true ✗
c) B alone is true (best)
d) A and B are untrue

$\boxed{\text{Actually, PCA is sensitive to outliers...}}$

Ans D

$Y_{ij} \sim N(\mu_i, \dots)$

$Y_1 \quad \dots \quad Y_m$

$\begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1p} \end{bmatrix} \quad \begin{bmatrix} Y_{m1} \\ Y_{m2} \\ \vdots \\ Y_{mp} \end{bmatrix}$

18. Let Yi be m independent MVN random p-vectors with mean **μ** and covariance Matrix Σ, Yi ~ INp (μ, Σ). Define X j = $\sum^{m}_{i=1}$ $Y^2_{ij}$ for j = 1, 2, . . . , p. Then the joint distribution of **X** = [X1, X2, . . . , X p] is a

C

a) Central chi square distribution with m df
b) Noncentral chi square distribution with m df → if $\mu = 0$, we can't choose this option
c) Central or Noncentral chi square distribution with m df (Best)
d) Central or Noncentral chi square distribution with m-1 df

19. Examine the following statements on Mahalanobis distance

A

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 As $\rho \to 1$, actually, $\Sigma^{-1} \to$ singular $(X-Y)^T \Sigma^{-1} (X-Y) \to$ problematic

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

$\boxed{\text{Actually, MD works well even when two or more variables are highly correlated.}}$

Ans B

20. MD can be calculated using PCs obtained after PC analysis and need not be corrected for covariance as

C

a) number of PCs are smaller
b) as PCs are easy to calculate
c) as PCs are orthogonal ✓
d) as PCs use the maximum variability