

Chennai Mathematical Institute
Multivariate Statistical Analysis - End Semester Examination

Time: 3 hours

Maximum Marks: 40

Part A - Answer TWO questions (2 x 10 = 20 Marks)

1. The following artificial data refers to a sample of six approximately two-year-old boys from a high-altitude region in Asia. MU AC is the mid-upper-arm circumference. For lowland, children of the same age in the same country, the height, chest, and MUAC means are considered to be 90, 58, and 16 cm respectively. (i) Test the hypothesis that the highland children have the same means. In addition, if similar data on nine two-year-old girls are provided, (ii) examine, if there exists any difference in mean vectors between boys and girls.

Individual	Height (cm)	Chest circumference (cm)	MUAC (cm)
1	78	60.6	16.5
2	76	58.1	12.5
3	92	63.2	14.5
4	81	59.0	14.0
5	81	60.8	15.5
6	84	59.5	14.0

Individual	Height (cm)	Chest circumference (cm)	MUAC (cm)
1	80	58.4	14.0
2	75	59.2	15.0
3	78	60.3	15.0
4	75	57.4	13.0
5	79	59.5	14.0
6	78	58.1	14.5
7	75	58.0	12.5
8	64	55.5	11.0
9	80	59.2	12.5

2. The admission officer of a business school has used an "index" of undergraduate grade point average (GPA) and graduate management aptitude test (GMAT), scores to help decide which applicants should be admitted to the school's graduate programs. The pairs of $X_1 = \text{GPA}$, $X_2 = \text{GMAT}$ values for groups of recent applicants who have been categorized as Π_1 : admit; Π_2 : do not admit and Π_3 borderline. Classify the student with GMAT and GPA scores as 497 and 3.21 respectively with equal and unequal probabilities.

$$\begin{aligned}
 n_1 &= 31 & n_2 &= 28 & n_3 &= 26 \\
 \bar{\mathbf{x}}_1 &= \begin{bmatrix} 3.40 \\ 561.23 \end{bmatrix} & \bar{\mathbf{x}}_2 &= \begin{bmatrix} 2.48 \\ 447.07 \end{bmatrix} & \bar{\mathbf{x}}_3 &= \begin{bmatrix} 2.99 \\ 446.23 \end{bmatrix} \\
 \bar{\mathbf{x}} &= \begin{bmatrix} 2.97 \\ 488.45 \end{bmatrix} & S_{\text{pooled}} &= \begin{bmatrix} .0361 & -2.0188 \\ -2.0188 & 3655.9011 \end{bmatrix}
 \end{aligned}$$

3. Let $Z = [Z1 \ Z2]$ with $Z1 = [Z1^1 \ Z2^1]$ and $Z2 = [Z1^2 \ Z2^2]$ are all standardized variables, then find the canonical variates and canonical correlation, given

$$\text{Cov}(Z) = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} = \begin{bmatrix} 1.0 & .4 & .5 & .6 \\ .4 & 1.0 & .3 & .4 \\ .5 & .3 & 1.0 & .2 \\ .6 & .4 & .2 & 1.0 \end{bmatrix}$$

PART-B (20 marks)

4. Public Utilities Data Table 1 below gives corporate data on 22 US public utilities. The objective of interest is forming groups of similar utilities. There are 8 measurements on each utility described in Table 2. Clustering would be useful to predict the cost impact of deregulation. The objects to be clustered are the utilities and there are 8 measurements on each utility.
- Carry out the process of clustering using hierarchical algorithms and offer your comments.
 - Also, carry out non-hierarchical k-means clustering with $k = 4$ and 5 and offer your comments. Identify the hierarchical algorithm that matches well with the k-means clustering.
 - Adopt PCA and draw a comparison with the hierarchical algorithm

No.	Company	X1	X2	X3	X4	X5	X6	X7	X8
1	Arizona Public Service	1.06	9.2	151	54.4	1.6	9077	0	0.628
2	Boston Edison Company	0.89	10.3	202	57.9	2.2	5088	25.3	1.555
3	Central Louisiana Electric Co.	1.43	15.4	113	53	3.4	9212	0	1.058
4	Commonwealth Edison Co.	1.02	11.2	168	56	0.3	6423	34.3	0.7
5	Consolidated Edison Co. (NY)	1.49	8.8	142	51.2	1	3300	15.6	2.044
6	Florida Power and Light	1.32	13.5	111	60	-2.2	11127	22.5	1.241
7	Hawaiian Electric Co.	1.22	12.2	175	67.6	2.2	7642	0	1.652
8	Idaho Power Co.	1.1	9.2	245	57	3.3	13082	0	0.309
9	Kentucky Utilities Co.	1.34	13	168	60.4	7.2	8406	0	0.862
10	Madison Gas & Electric Co.	1.12	12.4	197	53	2.7	6455	39.2	0.623
11	Nevada Power Co.	0.75	7.5	173	51.5	6.5	17441	0	0.768
12	New England Electric Co.	1.13	10.9	178	62	3.7	6154	0	1.897
13	Northern States Power Co.	1.15	12.7	199	53.7	6.4	7179	50.2	0.527
14	Oklahoma Gas and Electric Co.	1.09	12	96	49.8	1.4	9673	0	0.588
15	Pacific Gas & Electric Co.	0.96	7.6	164	62.2	-0.1	6468	0.9	1.4
16	Puget Sound Power & Light Co.	1.16	9.9	252	56	9.2	15991	0	0.62
17	San Diego Gas & Electric Co.	0.76	6.4	136	61.9	9	5714	8.3	1.92
18	The Southern Co.	1.05	12.6	150	56.7	2.7	10140	0	1.108
19	Texas Utilities Co.	1.16	11.7	104	54	-2.1	13507	0	0.636
20	Wisconsin Electric Power Co.	1.2	11.8	148	59.9	3.5	7297	41.1	0.702
21	United Illuminating Co.	1.04	8.6	204	61	3.5	6650	0	2.116
22	Virginia Electric & Power Co.	1.07	9.3	1784	54.3	5.9	10093	26.6	1.306

X1:	Fixed-charge covering ratio (income/debt)
X2:	Rate of return on capital
X3:	Cost per KW capacity in place
X4:	Annual Load Factor
X5:	Peak KWH demand growth from 1974 to 1975
X6:	Sales (KWH use per year)
X7:	Percent Nuclear
X8:	Total fuel costs (cents per KWH)