

- **4.** a) Briefly discuss how each of the following techniques can help in data reduction:
 - i) Clustering
 - ii) Box Plots

(4 Marks)

- b) Explain the difference between single linkage and complete linkage in cluster analysis. Under what circumstances would you use one of these techniques in preference to the other? (4 Marks)
- c) (7 Marks)

Data were collected on seven soybean plants. Five characteristics were measured on each plant. The Euclidean distances between the vectors of measurements are displayed in the following table for all pairs of plants.

Plant							
	1	2	3	4	5	6	7
Plant 1		33	37	24	31	36	39
Plant 2			42	22	39	42	35
Plant 3				41	45	30	42
Plant 4					41	32	40
Plant 5						46	48
Plant 6							34
Plant 7							

Use the complete linkage clustering procedure to make three clusters.

5. The doctor of a school has measured the height of pupils in a 5th grade class. The result (in cm) is as follows:

130 132 138 136 131 153 131 133 129 133 110 132 129 134 135 132 135 134 133 132 130 131 134 135 135 134 136 133 133 130

a- Which ones are outliers and why?

(2 Marks)

b- The weight of those pupils was measured in kg and the results is as follows:

37 40 39 40.5 42 51 41.5 39 41 30 40 42 40.5 39.5 41

- 40.5 37 39.5 40 41 38.5 39.5 40 41 39 40.5 40 38.5 39.5 41.5 Draw the box-plot for weight. (4 Marks)
- c- Draw the scatter-plot for both variables height and weight.

(4 Marks)

END of QUESTIONS