

Lista 12 – Análise de Agrupamento – Estatística 2024

1. Comente cada uma das questões abaixo:

- a) Técnicas de agrupamento podem ser utilizadas independentemente do tamanho da amostra utilizada;
- b) Técnicas de agrupamento hierárquico são mais versáteis que as de agrupamento não hierárquico, pois no agrupamento hierárquico, a definição do número de grupos pode ser definido depois que o dendrograma for obtido, ao passo que, no agrupamento não hierárquico esta definição é feita *a priori*.
- c) Quando se trabalha com muitas variáveis de diferentes grandezas é imprescindível normalizá-las antes do cálculo da matriz de distância.
- d) Técnicas de agrupamento hierárquico produzem dendrogramas semelhantes independentemente do método de ligação utilizado.

2. A partir da tabela abaixo, utilizando as técnicas de agrupamento hierárquico e não-hierárquico, tente descobrir quantos grupos aparentemente distintos existem nestes dados. Normalize os dados entre 0 e 1 (função *scale* do R) antes de proceder o cálculo da matriz de distância euclidiana entre os pontos (função *dist* do R). Apresente o dendrograma baseado no método de ligação completo (função *hclust* do R) e defina o número de grupos desejado (função *cutree* do R). Plote o diagrama de dispersão identificando os elementos de cada grupo. Repita o procedimento utilizando o método k-médias definindo-se o mesmo número de grupos encontrado anteriormente. Plote o diagrama de dispersão com o novo resultado identificando os elementos de cada grupo. Comente os resultados. Ambas as técnicas de agrupamento parecem concordar entre si?

Dica: para iniciar a análise, copie a tabela abaixo para a memória e execute no R o seguinte comando:

```
data <- read.table("clipboard", header=T, dec=",")
```

X1	X2
5,6	200,1
5,6	206,5
6,1	219,9
6,2	213,3
6,6	203,3
6,7	202,3
6,7	205,6
7,2	211,2
7,3	204,6
7,5	213,1
7,5	209,4
8,1	209,7
8,4	203,4
8,7	158,5
8,8	162,5
9,4	189,3
9,6	163,2
9,8	151
9,8	187,2
10	163,3
10,4	175
10,8	169,5
11,2	171,7
11,2	181,4
11,3	186,9
11,4	166,7
11,8	161,1
12,1	160
12,6	181,3
13,8	187,2
13,8	151,5
13,9	188,9
14,1	163,5
14,1	189,1
14,4	170,8
14,4	151,7
14,5	164,4
14,7	175,2
14,7	186,5
15	166,2
22,6	213,3
23,6	202,6
25,2	210,1
25,6	216,5
26,5	216,9
27	208,4
27,1	205,4
28,4	204,7
29,8	216,2
29,9	202,8