

## 4. Análise classificatória

4.1. Considere que foram observadas as variáveis X, Y e Z em 5 indivíduos, tendo-se obtido o seguinte conjunto de dados:

Indivíduo	X	Y	Z
A	7	6	3
B	6	8	4
C	3	2	4
D	5	6	2
E	4	1	2

- Calcule a matriz de distâncias euclidianas entre os indivíduos.
- Proceda à junção dos indivíduos com base nos métodos:
  - Ligação simples (i.e., menor distância).
  - Ligação completa (i.e., maior distância).
  - Distância média (i.e., maior distância).
- Construa os respetivos dendogramas.
- Confirme os resultados com auxílio do R,
- No R explore outros métodos de agrupamento, e compare os resultados.

4.2. (🖨) A escala WAIS III (Wechsler Adult Intelligence Scale – Third Edition) de inteligência para adultos é um instrumento muito utilizado na avaliação de habilidades cognitivas. Esta escala é dividida em quatro índices principais:

- Compreensão verbal (ICV): avalia habilidades de compreensão verbal e raciocínio.
- Organização perceptiva (IOP): mede habilidades não verbais e de resolução de problemas utilizando informações visuais.
- Memória de trabalho (IMT): avalia a capacidade de manter e manipular informações na memória a curto prazo.
- Velocidade de processamento (IVP): mede a rapidez com que uma pessoa pode processar informações e executar tarefas simples.

Quanto maior for o valor maior será a habilidade cognitiva.

Considere que esta escala foi aplicada a um grupo de 10 indivíduos com predisposição familiar para a síndrome de Alzheimer's e em idade de risco, tendo-se obtidos os seguintes resultados disponíveis em *Wescelher.xlsx*.

- Utilize a análise classificatória hierárquica para tentar identificar grupos de indivíduos aos quais deva ser recomendado um acompanhamento médico mais frequente.
  - Considere a distância euclidiana como medida de distância, e utilize diferentes métodos de aglomeração (menor distância, maior distância, distância média e Ward).
  - Obtenha os dendogramas associados a cada um dos métodos de aglomeração que usou e compare-os.
  - Confirme se obteria os mesmos agrupamentos se tivesse usado outra medida de distância, por ex., a distância de Manhattan ou a distância de Minkowski com  $m = 3$ , ou outra à sua escolha.

4.3. Foram analisadas algumas profissões segundo os níveis de prestígio, taxa de suicídio, rendimento e níveis de instrução, havendo a suspeita de que existem grupos homogêneos.

- a) Calcule a distância euclidiana entre os cozinheiros e os canalizadores, com base nas variáveis originais e as variáveis estandardizadas.
- b) Aplicar o método da maior distância com base nas variáveis originais (use a distância euclidiana).
  - i) Obtenha o dendograma.
  - ii) Determine o número de grupos com base na análise do dendograma e no gráfico das distâncias (scree plot).
  - iii) Caracterize os grupos.
- c) Repita a alínea b com base nas variáveis estandardizadas.
- d) Compare os resultados obtidos em b) e c).

4.4. Considere que foram observadas as variáveis X e Y em 5 indivíduos, tendo-se obtido o seguinte conjunto de dados:

Indivíduo	X	Y
A	7	6
B	6	8
C	3	2
D	5	6
E	4	1

Divida os indivíduos em 2 grupos, usando o algoritmo *K-means*, para que os indivíduos dentro do mesmo grupo estejam mais próximos entre si do que dos indivíduos nos outros grupos.

- a) Considere como grupos iniciais: (A, B) e (C, D, E).
- b) Considere como grupos iniciais: (E, A) e (B, C, D).
- c) Compare os resultados obtidos em a) e b).

4.5. Retome o exercício 4.2.

- a) Aplique o método *K-means*, considerando o número de grupos que obteve nesse exercício.
- b) Compare os grupos obtidos com o método não hierárquico e os métodos hierárquicos. Na comparação, considere comparar o valor de  $R^2$  produzido pelos diferentes métodos.

4.6. Retome o exercício 4.3. Em ambas as alíneas que se seguem considere os cenários com os dados originais e os dados estandardizados.

- a) Aplique o método *K-means*, considerando o número de grupos que obteve nesse exercício.
- b) Compare os grupos obtidos com o método não hierárquico e os métodos hierárquicos, nos casos em que considerou a distância euclidiana.