

PROVA 3

ANÁLISE MULTIVARIADA 1

Tailine J. S. Nonato (190038144)

2023-12-14

Lista 7

Exercício 12.11

Dada a matriz, os centroides dos clusters (AB) e (CD) são dados por:

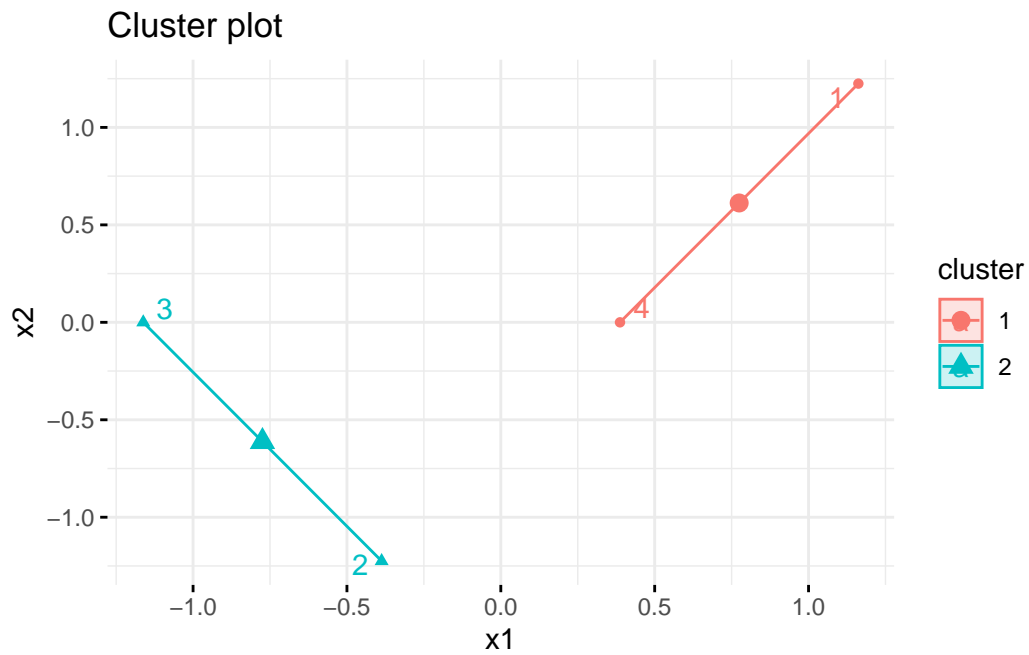
	x1	x2
[1,]	3	1
[2,]	1	1

Utilizando a função `kmeans` tem-se que o cluster final é:

```
[1] 1 2 2 1
```

Ou seja, 1:(AD) e 2:(BC). E os novos centroides são dados por:

	x1	x2
1	4	2.5
2	0	-0.5



Exercício 12.12

Dada a matriz, os centroides dos clusters (AC) e (BD) são dados por:

```

      x1 x2
[1,]  5  4
[2,]  3  1

```

O cluster final é:

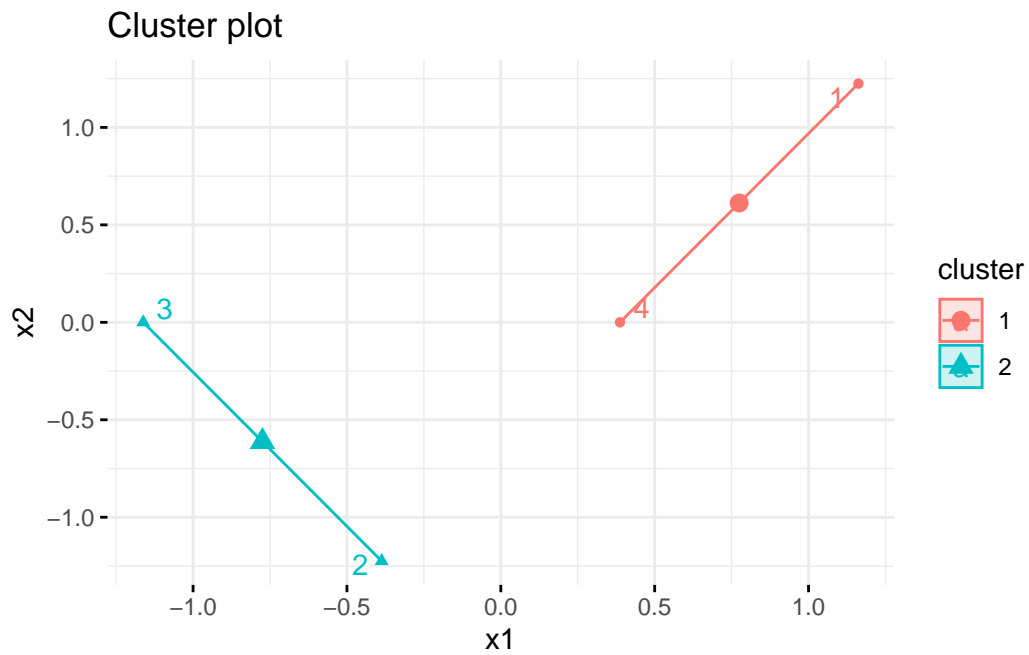
```
[1] 1 2 2 1
```

Ou seja, 1:(AD) e 2:(BC). E os centroides são dados por:

```

      x1  x2
1  4  2.5
2  0 -0.5

```



Resultado similar ao do item anterior.

Exercício 12.13

Dada a matriz, os centroides dos clusters (AB) e (CD) são dados por:

```

      x11 x21
[1,]  -1  -2
[2,]   2   2

```

O cluster final é:

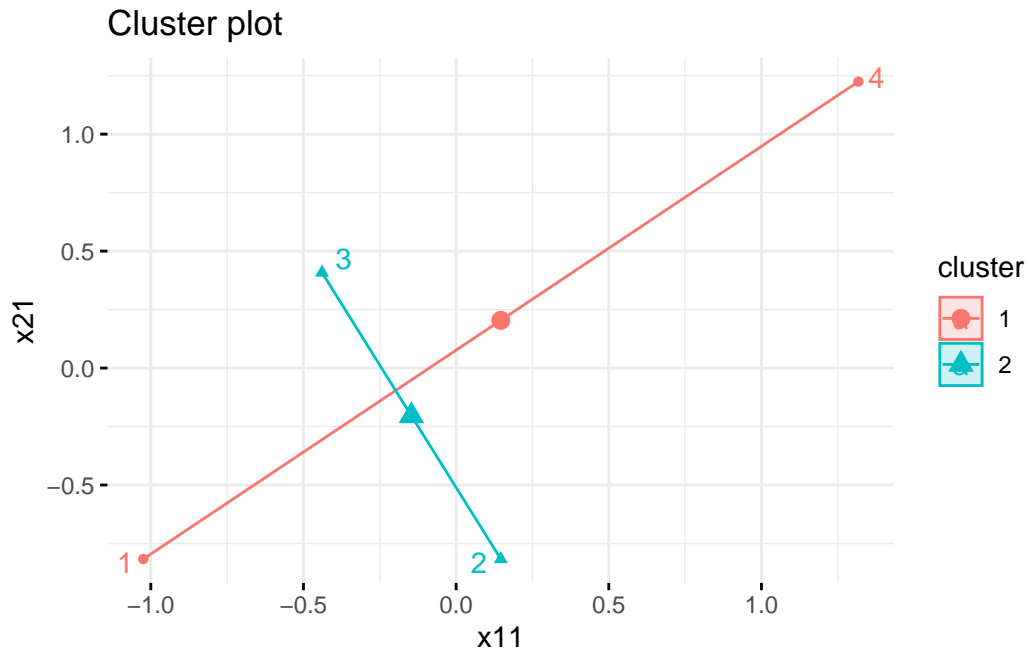
```
[1] 1 1 1 2
```

Ou seja, 1:(ABC) e 2:(D). E os centroides são dados por:

```

      x11 x21
1  -1  -1
2   5   3

```



Exercício 12.14

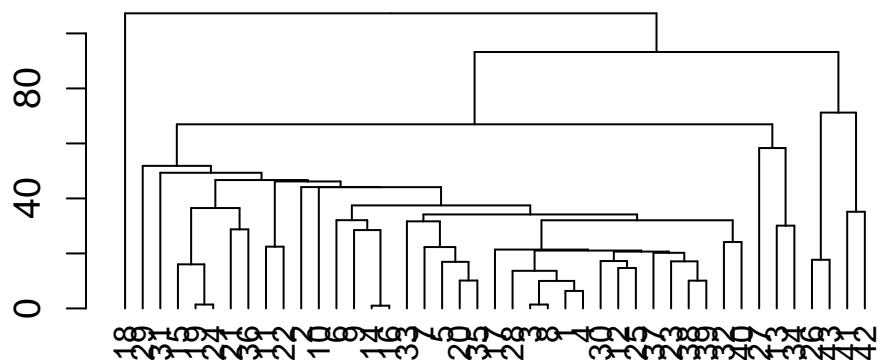
Item A

Verificando as 5 primeiras linhas e 5 primeiras colunas da matriz de distâncias é dada por:

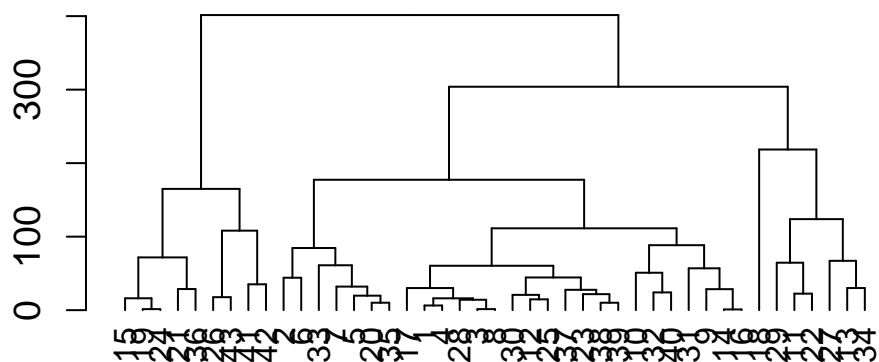
	1	2	3	4	5
1	0.000000	116.03663	15.50806	6.363961	103.20126
2	116.036632	0.00000	121.65114	117.894020	61.62792
3	15.508062	121.65114	0.00000	10.000000	100.62306
4	6.363961	117.89402	10.00000	0.000000	102.10289
5	103.201260	61.62792	100.62306	102.102889	0.00000

Por questões de tamanho, o output completo pode ser verificado no arquivo em R.

Item B



Ligação Simples



Ligação Completa

Comparando os dendogramas, pode-se dizer que os resultados são similares.

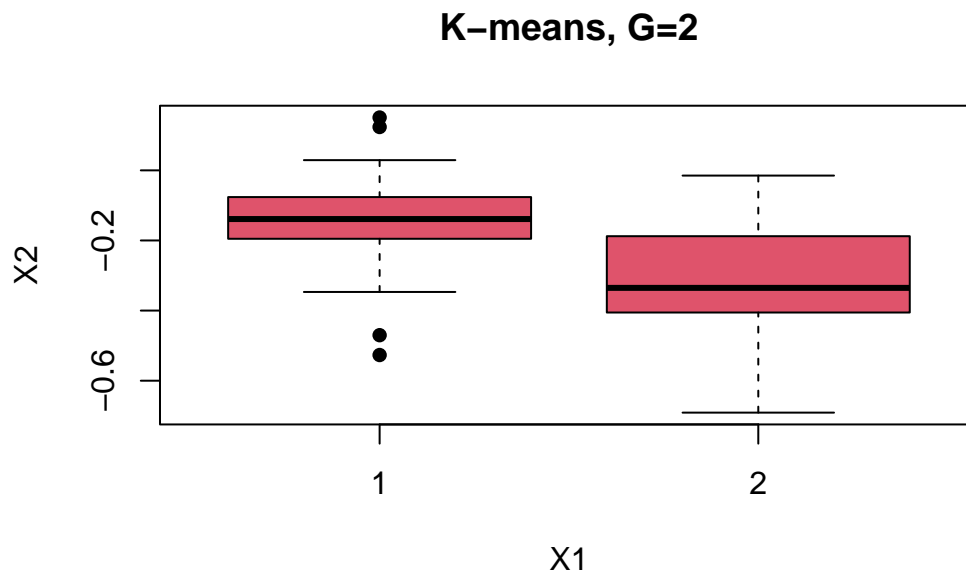
Exercício 12.15

```

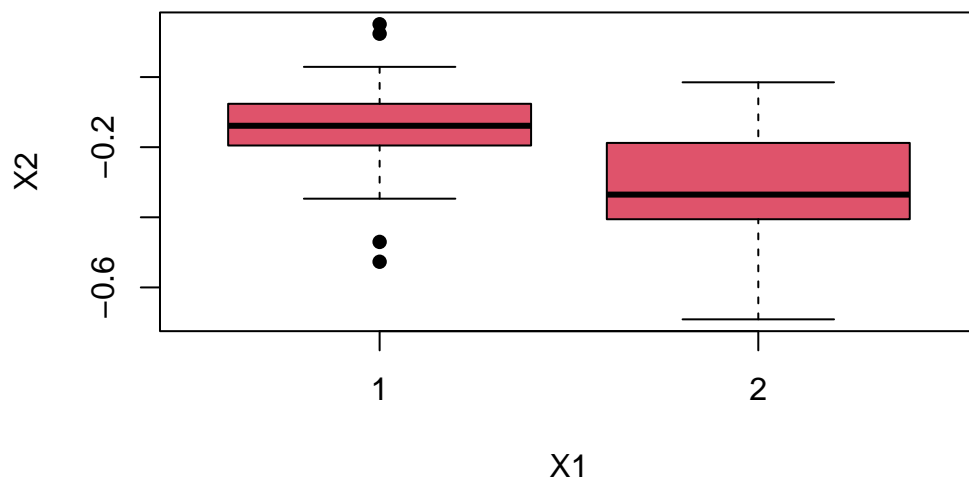
      V2      V3
[1,] -0.1348700 -0.077856667
[2,] -0.3079467 -0.005991111

```

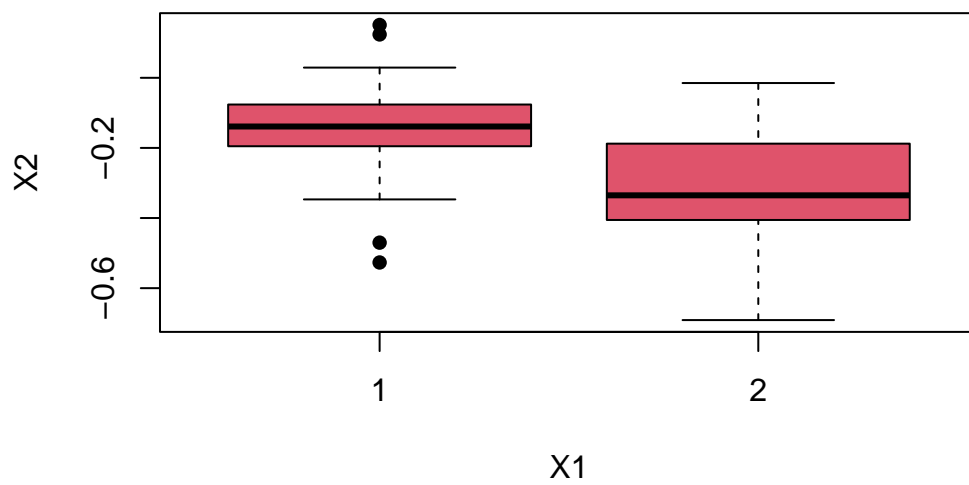
	V1	V2	V3	km15_2\$cluster	km15_3\$cluster	km15_4\$cluster
1	1	-0.0056	-0.1657	2	2	2
2	1	-0.1698	-0.1585	2	2	2
3	1	-0.3469	-0.1879	1	3	1
4	1	-0.0894	0.0064	2	2	4
5	1	-0.1679	0.0713	2	1	3
6	1	-0.0836	0.0106	2	2	4



K-means, G=3



K-means, G=4



Lista 8

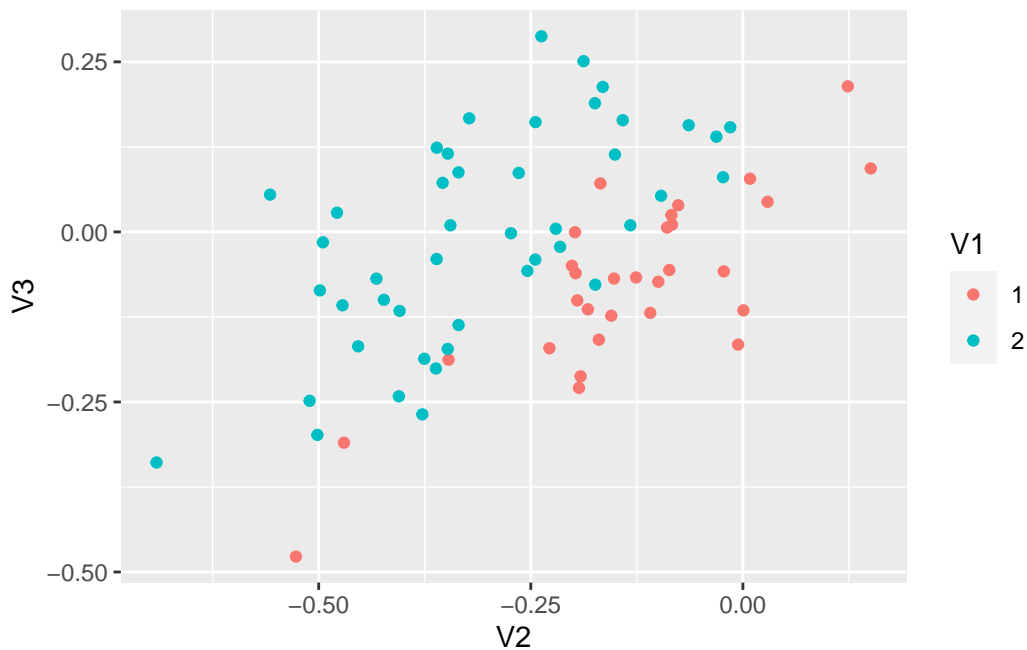
Exercício 11.32

Item A

Para investigar a suposição de Normal Bivariada, tem-se como hipóteses:

$$\begin{cases} H_0 : \text{Os dados seguem distribuição Normal} \\ H_1 : \text{Os dados não seguem distribuição Normal} \end{cases}$$

Realiza-se análise gráfica:



Pela análise gráfica, observa-se que os grupos têm formatos que lembram formatos elípticos, tornando a suposição de Normal Bivariada algo razoável a se dizer. Além disso, não há evidências para rejeitar H_0 .

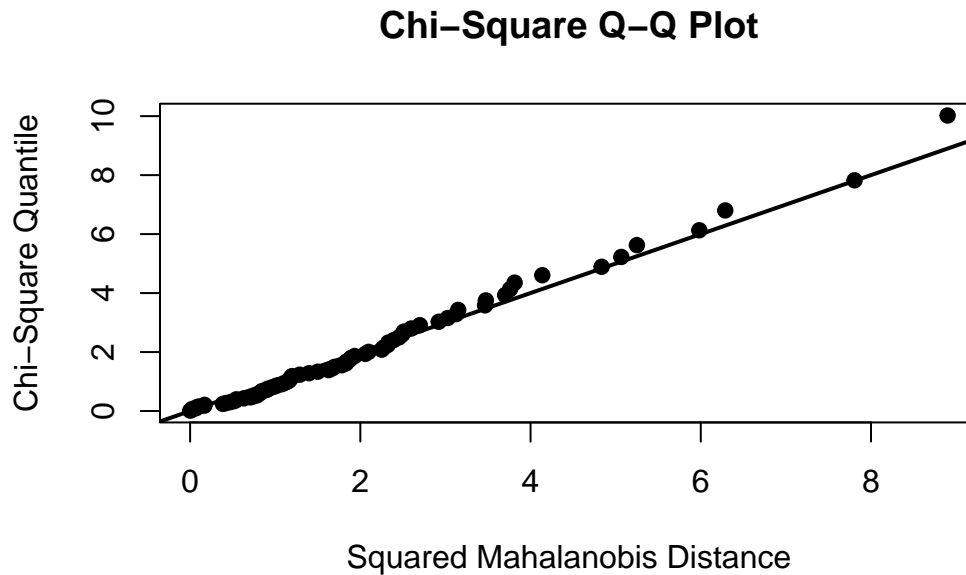
Shapiro-Wilk normality test

```
data: df32$V2
W = 0.98496, p-value = 0.5185
```


Shapiro-Wilk normality test

```
data: df32$V3  
W = 0.99255, p-value = 0.9428
```

Realizando o teste de Shapiro-Wilk para normalidade nas marginais, também não há evidências para rejeitar H_0 .



Anderson-Darling test for Multivariate Normality

```
data : df32[, 2:3]
```

```
AD          : 0.7583493  
p-value     : 0.2261774
```

```
Result  : Data are multivariate normal (sig.level = 0.05)
```

O teste de Anderson-Darling para normalidade multivariada também não rejeitou H_0 . Observa-se também pelo QQPlot mais uma indicação de normalidade. Assim, pode-se dizer que os dados seguem distribuição Normal Multivariada, especificamente para o caso, Bivariada.

Item B ao D

Uma tentativa de solucionar a questão está no arquivo R. Tive problemas ao gerar a função `cov_pooled` e não fui capaz de gerar os outputs que gostaria, mas deixei a estrutura.