

1-Práctica de Análisis de Cluster Jerárquico en R

Los datos de la siguiente tabla corresponden a las medidas de pecho, cintura y cadera en pulgadas de una muestra de 20 hombres y mujeres. La pregunta que nos hacemos es: ¿existen subgrupos de la forma del cuerpo entre hombres y mujeres, de forma que dentro de cada subgrupo los individuos sean similares y entre ambos subgrupos se diferencien?

Vamos a cargar el paquete **MVA** (Análisis multivariante aplicado), el paquete **mvtnorm** (Distribución t y normal multivariante), el paquete **mclust** (Cluster para modelos basados en mezclas de distribuciones), **flexclust** (Algoritmos flexibles de cluster)

pecho	cintura	cadera	sexo
34	30	32	m
37	32	37	m
38	30	36	m
36	33	39	m
38	29	33	m
43	32	38	m
40	33	42	m
38	30	40	m
40	30	37	m
41	32	39	m
36	24	35	f
36	25	37	f
34	24	37	f
33	22	34	f
36	26	38	f
37	26	37	f
34	25	38	f
36	26	37	f
38	28	40	f
35	23	35	f

1ª parte

Vamos a utilizar la función **hclust** para ejecutar un Análisis cluster jerárquico con distancia mínima, distancia máxima y promedio ponderado. La distancia que vamos a utilizar es la distancia euclídea.

hclust: realiza un análisis cluster jerárquico sobre un conjunto de distancias y métodos diferentes.

hclust (d, method, members)

d: una matriz de distancias

method: la estrategia del método cluster a usar

- ward.D método de Ward (la única opción para versiones <=3.03)
- ward.D2 método 2 de Ward (las diferencias se elevan al cuadrado antes de actualizar el cluster)
- single distancia mínima
- complete distancia máxima
- average promedio
- mcquitty promedio ponderado
- median mediana
- centroid centroide

plot(x, labels = NULL, hang = 0.1, axes = TRUE, frame.plot = FALSE, ann = TRUE, main = "Cluster Dendrogram", sub = NULL, xlab = NULL, ylab = "Height", ...)

Vamos a introducir la tabla de datos de las tallas en R

```
> medidas<-read.table("medidas.txt", h=T) Leemos los datos de la tabla
> medidas
```

	pecho	cintura	cadera	sexo
1	34	30	32	m
2	37	32	37	m
3	38	30	36	m
4	36	33	39	m
5	38	29	33	m
6	43	32	38	m
7	40	33	42	m
8	38	30	40	m
9	40	30	37	m
10	41	32	39	m
11	36	24	35	f
12	36	25	37	f
13	34	24	37	f
14	33	22	34	f
15	36	26	38	f
16	37	26	37	f
17	34	25	38	f
18	36	26	37	f
19	38	28	40	f
20	35	23	35	f

```
> dm1 <- dist(medidas[, c("pecho", "cintura", "cadera")]) Calculamos la matriz de
distancias euclídeas
```

```
> dm1
```

	1	2	3	4	5	6	7
2	6.164414						
3	5.656854	2.449490					
4	7.874008	2.449490	4.690416				
5	4.242641	5.099020	3.162278	7.483315			
6	11.000000	6.082763	5.744563	7.141428	7.681146		
7	12.041595	5.916080	7.000000	5.000000	10.049876	5.099020	
8	8.944272	3.741657	4.000000	3.741657	7.071068	5.744563	4.123106
9	7.810250	3.605551	2.236068	5.385165	4.582576	3.741657	5.830952
10	10.099505	4.472136	4.690416	5.099020	7.348469	2.236068	3.316625
11	7.000000	8.306624	6.403124	9.848858	5.744563	11.045361	12.083046
12	7.348469	7.071068	5.477226	8.246211	6.000000	9.949874	10.246951
13	7.810250	8.544004	7.280110	9.433981	7.549834	12.083046	11.916375
14	8.306624	11.180340	9.643651	12.449900	8.660254	14.696938	15.297059
15	7.483315	6.164414	4.898979	7.071068	6.164414	9.219544	9.000000
16	7.071068	6.000000	4.242641	7.348469	5.099020	8.544004	9.110434
17	7.810250	7.681146	6.708204	8.306624	7.549834	11.401754	10.770330
18	6.708204	6.082763	4.582576	7.280110	5.385165	9.273618	9.486833
19	9.165151	5.099020	4.472136	5.477226	7.071068	6.708204	5.744563
20	7.681146	9.433981	7.681146	10.816654	7.000000	12.409674	13.190906
	8	9	10	11	12	13	14
2							
3							
4							
5							
6							
7							
8							
9	3.605551						
10	3.741657	3.000000					
11	8.062258	7.483315	10.246951				
12	6.164414	6.403124	8.831761	2.236068			
13	7.810250	8.485281	10.816654	2.828427	2.236068		
14	11.180340	11.045361	13.747727	3.741657	5.196152	3.741657	
15	4.898979	5.744563	7.874008	3.605551	1.414214	3.000000	6.403124
16	5.099020	5.000000	7.483315	3.000000	1.414214	3.605551	6.403124
17	6.708204	7.874008	9.949874	3.741657	2.236068	1.414214	5.099020
18	5.385165	5.656854	8.062258	2.828427	1.000000	2.828427	5.830952
19	2.000000	4.123106	5.099020	6.708204	4.690416	6.403124	9.848858
20	9.110434	8.831761	11.532563	1.414214	3.000000	2.449490	2.449490
	15	16	17	18	19		
2							
3							
4							
5							
6							
7							
8							
9							

```

10
11
12
13
14
15
16 1.414214
17 2.236068 3.316625
18 1.000000 1.000000 2.449490
19 3.464102 3.741657 5.385165 4.123106
20 4.358899 4.123106 3.741657 3.741657 7.681146

```

```

> cs <- hclust(dm1, method = "single") Método del distancia mínima
> cs

```

```

Call:
hclust(d = dm1, method = "single")

```

```

Cluster method : single
Distance       : euclidean
Number of objects: 20

```

```

> cc <- hclust(dm1, method = "complete") Método del distancia máxima
> cc

```

```

Call:
hclust(d = dm1, method = "complete")

```

```

Cluster method : complete
Distance       : euclidean
Number of objects: 20

```

```

> ca <- hclust(dm1, method = "average") Método del promedio
> ca

```

```

Call:
hclust(d = dm1, method = "average")

```

```

Cluster method : average
Distance       : euclidean
Number of objects: 20

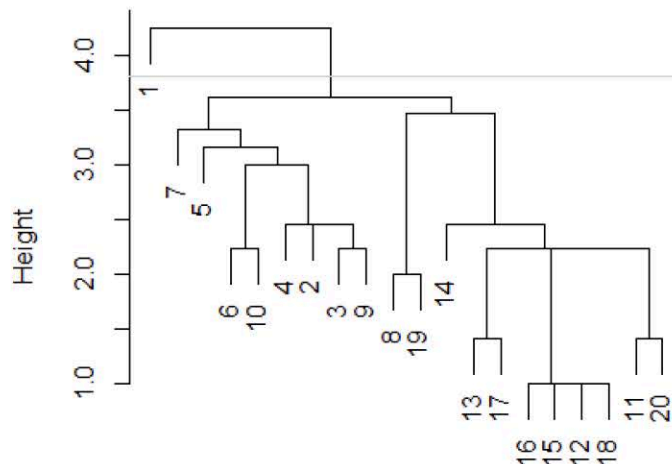
```

```

> plot(cs <- hclust(dm1, method = "single")) Dendrograma con el corte a una altura de
3.8
abline(h = 3.8, col = "lightgrey")

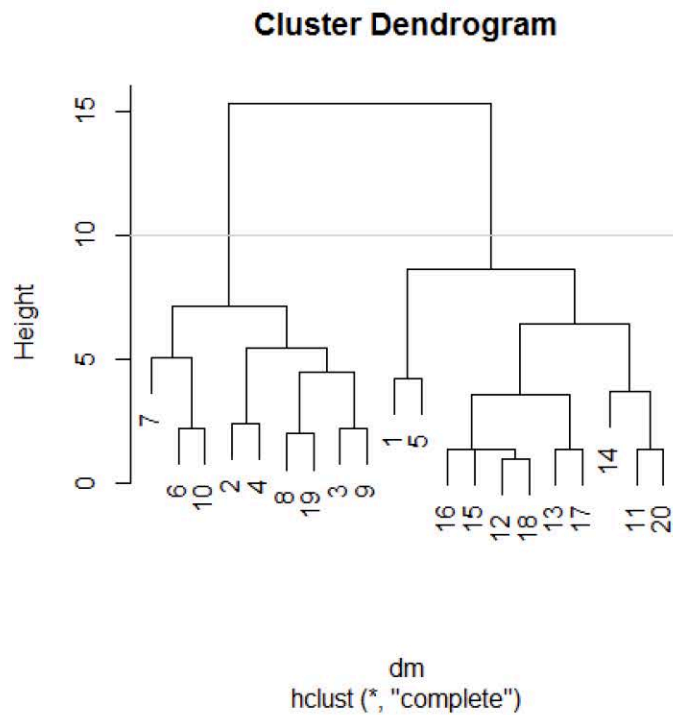
```

Cluster Dendrogram

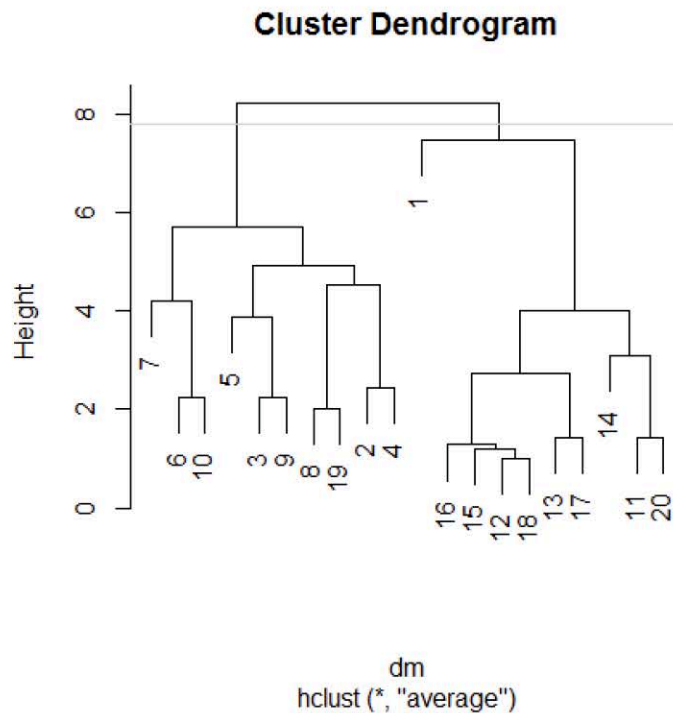


dm
hclust(*, "single")

```
> plot(cc <- hclust(dm1, method = "complete")) Dendrograma con corte a la altura 10  
abline(h = 10, col = "lightgrey")
```



```
> plot(ca <- hclust(dm1, method = "average")) Dendrograma  
abline(h = 7.8, col = "lightgrey")
```



Veamos ahora cómo seleccionar particiones de los datos de los dendrogramas. Como vimos en el desarrollo teórico de esta técnica, tenemos que cortar el dendrograma por algún punto, que nos dará un número particular de grupos. Ahora bien ¿cómo elegir el punto donde cortar, o dicho de otra forma, cómo decidir que un número particular de grupos es el óptimo para los datos? Esta cuestión es difícil de resolver, es en principio el investigador que realiza el

análisis, buen conocedor de los datos que maneja, el que debe decidir, en qué punto los grupos son más homogéneos y compactos.

En nuestro ejemplo como sabemos que los datos consisten en medidas sobre 10 hombres y 10 mujeres, buscaremos una solución de 2 grupos, cortando el dendrograma por las alturas adecuadas.

Podemos representar y comparar las tres soluciones gráficamente, dibujando los valores de las dos primeras componentes principales de los datos, etiquetando los puntos para identificar la solución cluster de uno de los métodos:

Realizamos por lo tanto un Análisis de Componentes Principales con la función **princomp** a la matriz de distancias

```
> cuerpo_pc <- princomp(dm1, cor = TRUE) Análisis de componentes principales
> cuerpo_pc
Call:
princomp(x = dm1, cor = TRUE)
```

```
Standard deviations:
  Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6    Comp.7
3.46007015 2.10009281 1.49848592 0.80761233 0.44083266 0.40303615 0.30290549
  Comp.8    Comp.9    Comp.10   Comp.11   Comp.12   Comp.13   Comp.14
0.26905442 0.21377621 0.21001515 0.16384224 0.14673317 0.13322578 0.11386412
  Comp.15   Comp.16   Comp.17   Comp.18   Comp.19   Comp.20
0.09829363 0.08313989 0.07558329 0.06295785 0.06137847 0.00000000
```

```
20 variables and 20 observations.
```

```
xlim <- range(cuerpo_pc$scores[,1]) Rango de valores del eje x
> xlim
[1] -5.193501 4.657388
```

```
> plot(cuerpo_pc$scores[,1:2], type = "n", Representa el marco del dibujo
+ xlim = xlim, ylim = ylim)
```

```
> cuerpo_pc$scores[,1:2] Valores de las 2 primeras componentes principales
  Comp.1    Comp.2
1  0.5558846  2.70150224
2 -3.2394203 -1.06062400
3 -1.9803969 -2.31862771
4 -3.7376752  0.55148284
5 -0.3238166 -0.03305419
6 -4.9570240  2.66979088
7 -5.1935010  2.97399611
8 -2.9702560 -1.67301445
9 -3.1051525 -1.56680353
10 -4.9210859  0.83988425
11  3.9033661  0.42587938
12  3.3464781 -1.57828199
13  4.1614115  0.71634381
14  4.6573878  4.87325857
15  2.3510749 -2.30862321
16  2.1670693 -2.52244439
17  3.4374768 -0.14256479
18  2.7003195 -2.36323102
19 -1.3144762 -2.08277614
20  4.4623359  1.89790735
```

```
> lab <- cutree(cs, h = 3.8) Establece el corte del dendrograma a la altura 3.8 en el
método de la distancia mínima
```

```
> lab
[1] 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
```

```
> text(cuerpo_pc$scores[,1:2], labels = lab, cex = 0.6) Representa las etiquetas de los
puntos
```

```
> plot(cuerpo_pc$scores[,1:2], type = "n", Igual para la distancia máxima
+ xlim = xlim, ylim = ylim)
```

```
> lab <- cutree(cc, h = 10)
```

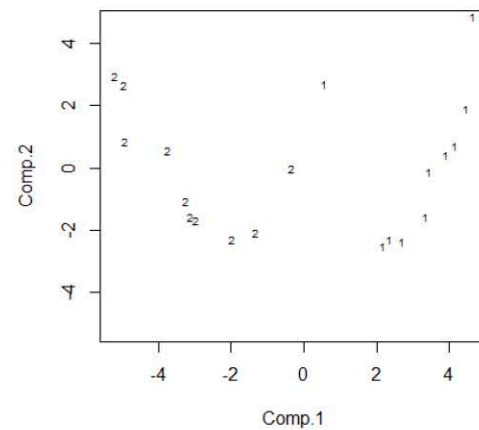
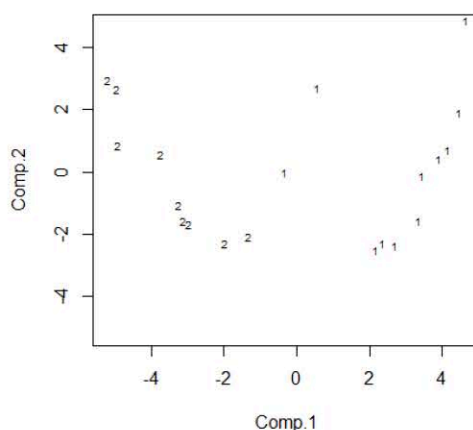
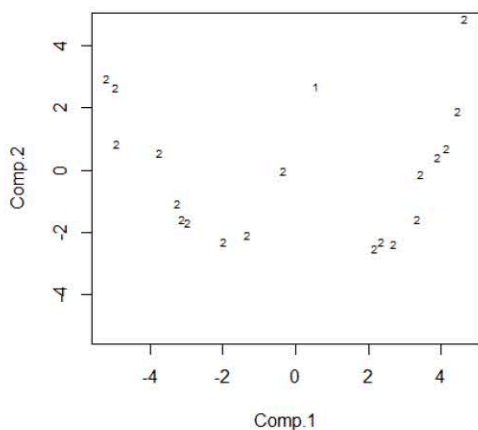
```
> lab
[1] 1 2 2 2 1 2 2 2 2 2 1 1 1 1 1 1 1 1 2 1
```

```
> text(cuerpo_pc$scores[,1:2], labels = lab, cex = 0.6)
```

```

> plot(cuerpo_pc$scores[,1:2], type = "n", Igual para la distancia promedio
+ xlim = xlim, ylim = xlim)
> lab <- cutree(ca, h = 7.8)
> lab
[1] 1 2 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 2 1
> text(cuerpo_pc$scores[,1:2], labels = lab, cex = 0.6)

```



Estas gráficas representan los cluster correspondientes a las tres estrategias elegidas en el espacio de las dos primeras componentes principales. Si hubiésemos querido poner todos los gráficos juntos para establecer una mejor comparación podemos utilizar la función **layout**.

layout divides the device up into as many rows and columns as there are in matrix mat, with the column-widths and the row-heights specified in the respective arguments.

```

> layout(matrix(1:6, nr = 2), height = c(2, 1))

> plot(cs <- hclust(dm1, method = "single"), main = "Simple")
> abline(h = 3.8, col = "lightgrey")
> xlim <- range(cuerpo_pc$scores[,1])
> plot(cuerpo_pc$scores[,1:2], type = "n", xlim = xlim, ylim = xlim,
+ xlab = "PC1", ylab = "PC2")
> lab <- cutree(cs, h = 3.8)
> text(cuerpo_pc$scores[,1:2], labels = lab, cex=0.6)

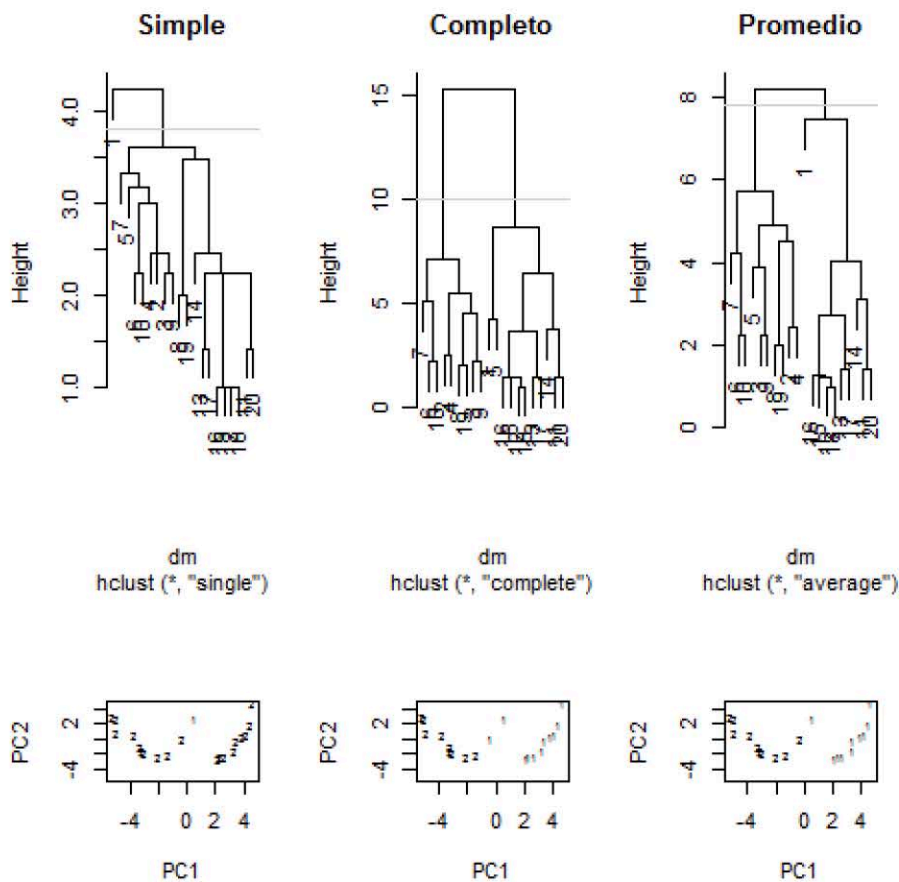
> plot(cc <- hclust(dm1, method = "complete"), main = "Complete")
> abline(h = 10, col = "lightgrey")
> plot(cuerpo_pc$scores[,1:2], type = "n", xlim = xlim, ylim = xlim,
+ xlab = "PC1", ylab = "PC2")
> lab <- cutree(cc, h = 10)
> text(cuerpo_pc$scores[,1:2], labels = lab, cex=0.6)

> plot(ca <- hclust(dm1, method = "average"), main = "Average")
> abline(h = 7.8, col = "lightgrey")
> plot(cuerpo_pc$scores[,1:2], type = "n", xlim = xlim, ylim = xlim,
+ xlab = "PC1", ylab = "PC2")
> lab <- cutree(ca, h = 7.8)
> text(cuerpo_pc$scores[,1:2], labels = lab, cex=0.6)

```

La solución que nos muestra el gráfico asociado con el linkage simple es uno de los problemas asociados a este método en la práctica, conocido como encadenamiento, y que se refiere a la tendencia a incorporar puntos intermedios entre clusters, en un cluster existente, antes de iniciar uno nuevo.

Como resultado de esta situación la solución de la distancia mínima, a menudo contiene grupos muy dispersos que no permiten una descripción útil de los datos. Las soluciones de los otros dos métodos, distancia máxima y distancia promedio, sitúan en esencia a los hombres en un grupo y a las mujeres en otro.



2ª parte

Con los mismos datos vamos a utilizar la distancia manhattan y los otros cuatro métodos: promedio ponderado, centroide, mediana y ward. Debes cortar cada dendrograma para obtener 2 grupos y analiza la coherencia de los resultados obtenidos.

```
> dm2 <- dist(medidas[, c("pecho", "cintura", "cadera")], method="manhattan")
```

```
> dm2
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19
2 10
3  8  4
4 12  4  8
5  6  8  4 12
6 17  7  9  9 13
7 19  9 11  7 15  8
8 12  6  4  6  8  9  7
9 11  5  3  9  7  6  8  5
10 16  6  8  6 12  3  5  6  5
11 11 11  9 13  9 18 20 13 12 17
12 12  8  8 10 10 15 17 10  9 14  3
13 11 11 11 13 13 18 20 13 12 17  4  3
14 11 17 15 19 13 24 26 19 18 23  6  9  6
15 12  8  8  8 10 13 15  8  9 12  5  2  5 11
16 12  6  6 10  8 13 15  8  7 12  5  2  5 11  2
17 11 11 11 11 13 16 18 11 12 15  6  3  2  8  3  5
18 11  7  7  9  9 14 16  9  8 13  4  1  4 10  1  1  4
19 14  8  6  8  8 11  9  2  7  8 11  8 11 17  6  6  9  7
20 11 13 11 15 11 20 22 15 14 19  2  5  4  4  7  7  6  6 13
```

```
> pp <- hclust(dm2, method = "mcquitty")
> pp
```

```
Call:
hclust(d = dm2, method = "mcquitty")
```

```
Cluster method : mcquitty
Distance       : manhattan
Number of objects: 20
```

```
> ce <- hclust(dm2, method = "centroid")
> ce
```

```
Call:
hclust(d = dm2, method = "centroid")
```

```
Cluster method : centroid
Distance       : manhattan
Number of objects: 20
```

```
> me <- hclust(dm2, method = "median")
> me
```

```
Call:
hclust(d = dm2, method = "median")
```

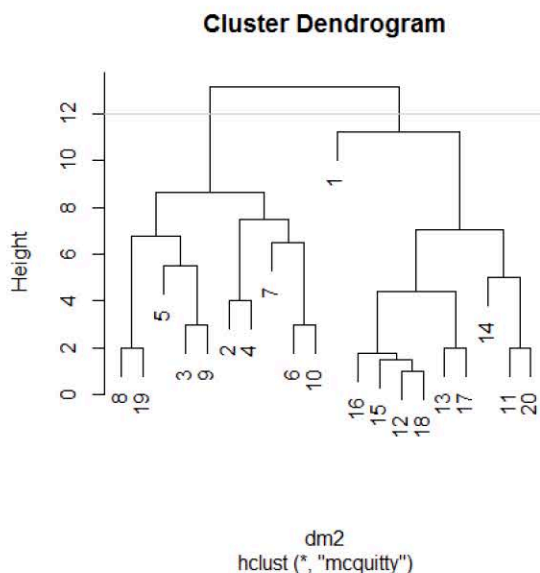
```
Cluster method : median
Distance       : manhattan
Number of objects: 20
```

```
> ward <- hclust(dm2, method = "ward.D2")
> ward
```

```
Call:
hclust(d = dm2, method = "ward.D2")
```

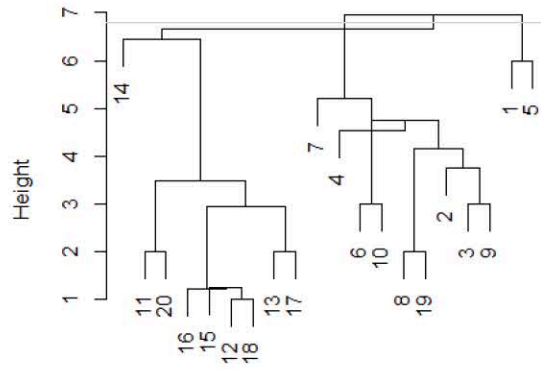
```
Cluster method : ward.D2
Distance       : manhattan
Number of objects: 20
```

```
> plot(pp <- hclust(dm2, method = "mcquitty"))
> abline(h = 12, col = "lightgrey")
```



```
> plot(ce <- hclust(dm2, method = "centroid"))
> abline(h = 6.8, col = "lightgrey")
```

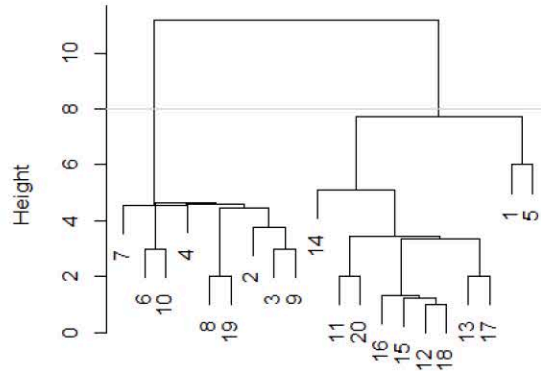

Cluster Dendrogram



dm2
hclust(*, "centroid")

```
> plot(me <- hclust(dm2, method = "median"))  
> abline(h = 8, col = "lightgrey")
```

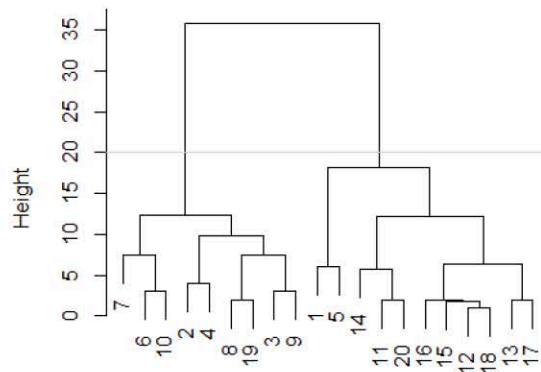
Cluster Dendrogram



dm2
hclust(*, "median")

```
> plot(ward <- hclust(dm2, method = "ward.D2"))  
> abline(h = 20, col = "lightgrey")
```

Cluster Dendrogram



dm2
hclust(*, "ward.D2")

Los 3 métodos excepto el centroide, proporcionan dos grupos coherentes de hombres y mujeres

2-Práctica de Análisis de Cluster Jerárquico en R

Los datos de la siguiente tabla debidos a Stanley y Miller corresponden a los valores de 6 variables sobre 22 aviones de combate de EEUU, las variables son:

FFD: primera fecha de vuelo, en meses después de 1940

SPR: potencia específica, proporcional a la potencia por unidad de peso

RGF: factor del rango de vuelo

PLF: carga útil como una fracción del peso bruto del avión

SLF: factor de carga sostenido

CAR: una variable binaria que toma el valor 1 si puede aterrizar en un portaaviones o 0 en caso contrario

	FFD	SPR	RGF	PLF	SLF	CAR
FH-1	82	1.468	3.30	0.166	0.10	0
FJ-1	89	1.605	3.64	0.154	0.10	0
F-86A	101	2.168	4.87	0.177	2.90	1
F9F-2	107	2.054	4.72	0.275	1.10	0
F-94A	115	2.467	4.11	0.298	1.00	1
F3D-1	122	1.294	3.75	0.150	0.90	0
F-89A	127	2.183	3.97	0.000	2.40	1
XF10F-1	137	2.426	4.65	0.117	1.80	0
F9F-6	147	2.607	3.84	0.155	2.30	0
F-100A	166	4.567	4.92	0.138	3.20	1
F4D-1	174	4.588	3.82	0.249	3.50	0
F11F-1	175	3.618	4.32	0.143	2.80	0
F-101A	177	5.855	4.53	0.172	2.50	1
F3H-2	184	2.898	4.48	0.178	3.00	0
F-102A	187	3.880	5.39	0.101	3.00	1
F-8A	189	0.455	4.99	0.008	2.64	0
F-104B	194	8.088	4.50	0.251	2.70	1
F-105B	197	6.502	5.20	0.366	2.90	1
YF-107A	201	6.081	5.65	0.106	2.90	1
F-106A	204	7.105	5.40	0.089	3.20	1
F-4B	255	8.548	4.20	0.222	2.90	0
F-111A	328	6.321	6.45	0.187	2.00	1

Vamos a aplicar el método de la distancia máxima pero usando solo las variables 2 a 5, utilizando la distancia euclídea, y dado que las variables están dadas en diferentes escalas, las tipificaremos a varianza unidad previamente. También utilizaremos la proyección sobre las dos primeras componentes principales como en la práctica 1ª.

Introducimos los datos de la tabla en R

```
> aviones
  FFD  SPR  RGF  PLF  SLF  CAR
FH-1   82 1.468 3.30 0.166 0.10  0
FJ-1   89 1.605 3.64 0.154 0.10  0
F-86A  101 2.168 4.87 0.177 2.90  1
F9F-2  107 2.054 4.72 0.275 1.10  0
F-94A  115 2.467 4.11 0.298 1.00  1
F3D-1  122 1.294 3.75 0.150 0.90  0
F-89A  127 2.183 3.97 0.000 2.40  1
XF10F-1 137 2.426 4.65 0.117 1.80  0
F9F-6  147 2.607 3.84 0.155 2.30  0
F-100A 166 4.567 4.92 0.138 3.20  1
F4D-1  174 4.588 3.82 0.249 3.50  0
F11F-1 175 3.618 4.32 0.143 2.80  0
F-101A 177 5.855 4.53 0.172 2.50  1
```

F3H-2	184	2.898	4.48	0.178	3.00	0
F-102A	187	3.880	5.39	0.101	3.00	1
F-8A	189	0.455	4.99	0.008	2.64	0
F-104B	194	8.088	4.50	0.251	2.70	1
F-105B	197	6.502	5.20	0.366	2.90	1
YF-107A	201	6.081	5.65	0.106	2.90	1
F-106A	204	7.105	5.40	0.089	3.20	1
F-4B	255	8.548	4.20	0.222	2.90	0
F-111A	328	6.321	6.45	0.187	2.00	1

> X Se tipifican los datos

	SPR	RGF	PLF	SLF
FH-1	0.31366825	0.6954522	0.8609797	0.0395801
FJ-1	0.34294111	0.7671049	0.7987402	0.0395801
F-86A	0.46323758	1.0263189	0.9180325	1.1478229
F9F-2	0.43887915	0.9947074	1.4263217	0.4353811
F-94A	0.52712505	0.8661542	1.5456141	0.3958010
F3D-1	0.27648959	0.7902866	0.7779937	0.3562209
F-89A	0.46644264	0.8366501	0.0000000	0.9499224
XF10F-1	0.51836456	0.9799554	0.6068351	0.7124418
F9F-6	0.55703892	0.8092535	0.8039268	0.9103423
F-100A	0.97583304	1.0368561	0.7157542	1.2665632
F4D-1	0.98032012	0.8050387	1.2914695	1.3853035
F11F-1	0.77305976	0.9104102	0.7416873	1.1082428
F-101A	1.25104061	0.9546663	0.8920994	0.9895025
F3H-2	0.61921703	0.9441291	0.9232192	1.1874030
F-102A	0.82904143	1.1359053	0.5238491	1.1874030
F-8A	0.09722006	1.0516081	0.0414930	1.0449146
F-104B	1.72816677	0.9483440	1.3018427	1.0686627
F-105B	1.38928540	1.0958641	1.8983046	1.1478229
YF-107A	1.29933013	1.1906985	0.5497822	1.1478229
F-106A	1.51812870	1.1380128	0.4616096	1.2665632
F-4B	1.82645519	0.8851210	1.1514306	1.1478229
F-111A	1.35061105	1.3592930	0.9698988	0.7916020

attr("scaled:scale")

SPR	RGF	PLF	SLF
4.6801039	4.7451138	0.1928036	2.5265222

> dj Se calcula la matriz de distancias

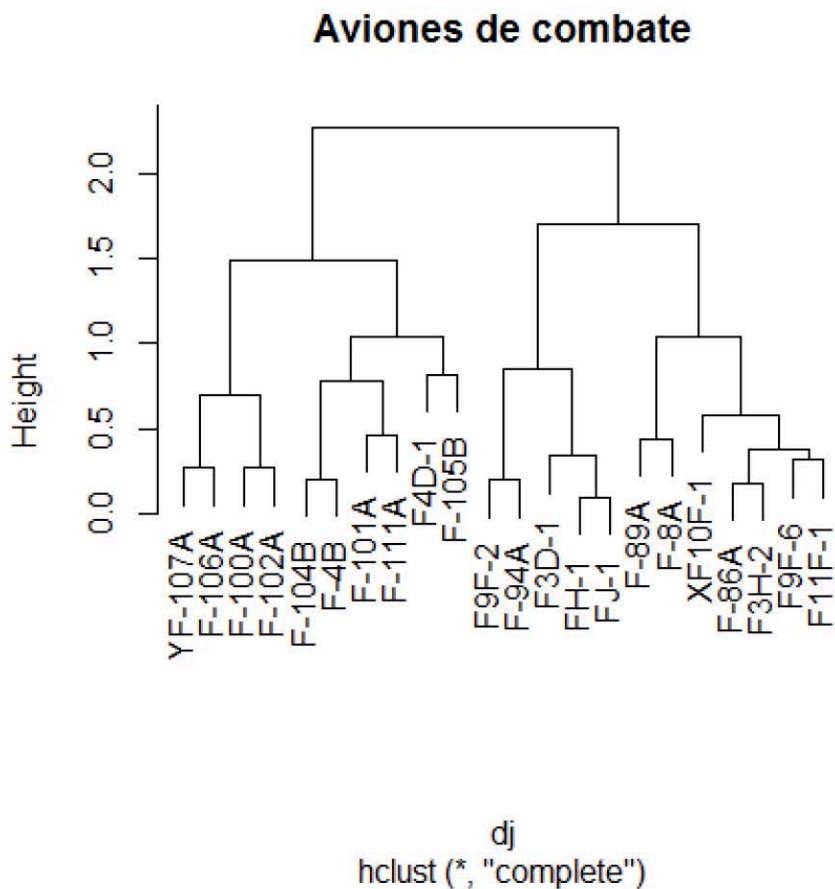
	FH-1	FJ-1	F-86A	F9F-2	F-94A	F3D-1
FJ-1	0.09932149					
F-86A	1.16760476	1.15069367				
F9F-2	0.76256248	0.78200005	0.87608438			
F-94A	0.81873108	0.85349192	0.99455010	0.20027616		
F3D-1	0.34281757	0.32503078	0.85838840	0.70338741	0.81202322	
F-89A	1.27015193	1.21934331	0.95808812	1.52475843	1.64332780	0.99799166
XF10F-1	0.80010936	0.75209652	0.53998993	0.86882482	0.99729380	0.50066591
F9F-6	0.91305085	0.89770156	0.35402434	0.81321491	0.90497808	0.62192501
F-100A	1.44277398	1.40914587	0.56380930	1.21896100	1.29513408	1.17578680
F4D-1	1.56611840	1.56889918	0.71567556	1.11788652	1.11929690	1.34842999
F11F-1	1.18891468	1.16225361	0.37694673	1.01992641	1.10286630	0.90987180
F-101A	1.35984552	1.33074884	0.80715896	1.11966671	1.14518603	1.17933712
F3H-2	1.21514254	1.20027429	0.18077123	0.92397460	1.01418450	0.92362531
F-102A	1.37505013	1.32867219	0.55024433	1.24585604	1.35445807	1.08637883
F-8A	1.36233356	1.31356070	0.95578242	1.55217601	1.70380170	1.05695732
F-104B	1.82157392	1.80660481	1.32653781	1.44255078	1.40050838	1.70714379
F-105B	1.90303100	1.90792111	1.35030898	1.28212551	1.21902200	1.79260211
YF-107A	1.59431622	1.54411584	0.92826706	1.43341664	1.50294639	1.37304285
F-106A	1.81977162	1.77137886	1.16090135	1.67536540	1.72895573	1.60978356
F-4B	1.90710839	1.88873886	1.39024234	1.58761295	1.55226804	1.78254756
F-111A	1.44683629	1.40032771	1.01384755	1.13990918	1.18719066	1.30533116
	F-89A	XF10F-1	F9F-6	F-100A	F4D-1	F11F-1
FJ-1	0.66923697					
F-86A	0.81044625	0.32961281				
F9F-2	0.95505311	0.72899176	0.60154732			
F-100A	1.45688692	1.07956750	0.80154274	0.63190729		
F4D-1	1.82135255	0.49451971	0.31612663	0.28782535	0.65809050	
F11F-1	1.19454164	0.83402165	0.71890469	0.43629651	0.64174208	0.51686299
F-101A	0.97140228	0.58063911	0.33623034	0.43021079	0.56966176	0.25303103
F3H-2	0.74286368	0.59440331	0.57956271	0.27285974	0.87221770	0.32817956
F-102A	0.43963329	0.78271747	0.93251494	1.12957275	1.58712571	0.98538249
F-8A	1.82025193	1.44032971	1.28991172	0.97801066	0.82473027	1.10860681
F-104B	2.13577982	1.62153880	1.42437340	1.25974182	0.82248276	1.32418173
F-105B	1.07725435	0.92039753	0.90414770	0.41226670	0.92574187	0.62762744
YF-107A	1.22890512	1.16303543	1.12953557	0.60737693	1.05017929	0.84287574
F-106A	1.79358319	1.48534026	1.33952585	0.97493322	0.89351310	1.13125490
F-4B	1.42150566	0.98722480	0.98688709	0.73116466	0.94879000	0.82909778
	F-101A	F3H-2	F-102A	F-8A	F-104B	F-105B
FJ-1						
F-86A						

```

F9F-2
F-94A
F3D-1
F-89A
XF10F-1
F9F-6
F-100A
F4D-1
F11F-1
F-101A
F3H-2 0.66290657
F-102A 0.62105058 0.49020492
F-8A 1.43781130 1.04008486 0.89198590
F-104B 0.63391268 1.17781254 1.20953671 2.06390312
F-105B 1.03757535 1.25235256 1.48531850 2.26489088 0.70614150
YF-107A 0.44753719 0.81509022 0.47582862 1.31657236 0.90249543 1.35484245
F-106A 0.60583505 1.03198004 0.69640922 1.50069089 0.90843085 1.44795332
F-4B 0.65441310 1.23067177 1.20546761 2.06410361 0.20627140 0.89070238
F-111A 0.46781880 0.93066412 0.82313771 1.60989551 0.70595812 1.02942756
      YF-107A      F-106A      F-4B
FJ-1
F-86A
F9F-2
F-94A
F3D-1
F-89A
XF10F-1
F9F-6
F-100A
F4D-1
F11F-1
F-101A
F3H-2
F-102A
F-8A
F-104B
F-105B
YF-107A
F-106A 0.26929959
F-4B 0.85628224 0.80558785
F-111A 0.57831229 0.74898142 0.78173772

```

```
> plot(cc <- hclust(dj), main = "Aviones de combate")
```



El resultado del dendrograma sugiere la presencia de dos grupos de aviones.

```
> cc
```

```
Call:
hclust(d = dj)
```

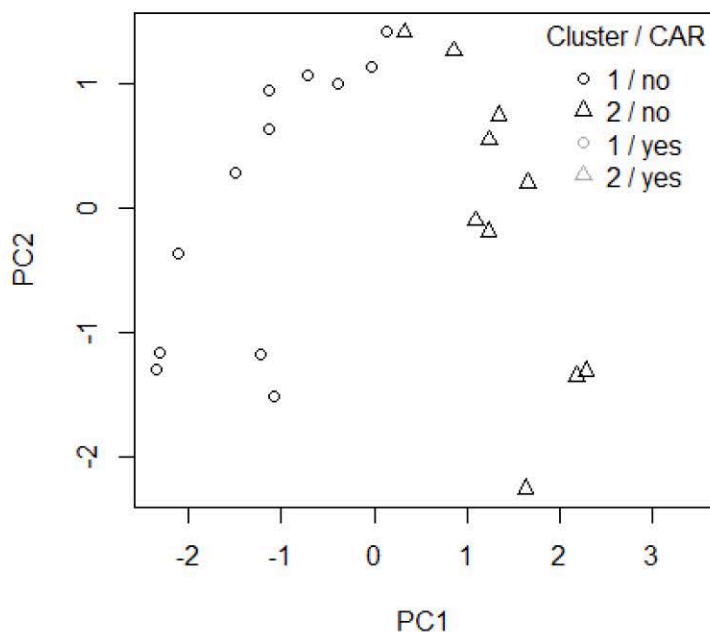
```
Cluster method : complete
Distance       : euclidean
Number of objects: 22
```

```
> pr Se calculan las dos primeras componentes principales
```

	PC1	PC2
FH-1	-2.35123899	-1.29064386
FJ-1	-2.32149092	-1.16139966
F-86A	-0.39637806	1.00518658
F9F-2	-1.23131537	-1.16859977
F-94A	-1.08288401	-1.51471688
F3D-1	-2.11065550	-0.35845008
F-89A	-1.13156882	0.64121890
XF10F-1	-1.13169763	0.94341402
F9F-6	-0.71430742	1.06216939
F-100A	0.86375794	1.26728237
F4D-1	1.09967995	-0.09616507
F11F-1	0.13485591	1.41874755
F-101A	1.23932013	0.54939976
F3H-2	-0.02046055	1.14196098
F-102A	0.32693900	1.41512764
F-8A	-1.49655884	0.28687196
F-104B	2.18332330	-1.34662343
F-105B	1.63009130	-2.25203455
YF-107A	1.34113820	0.74735701
F-106A	1.65320183	0.20724272
F-4B	2.28272152	-1.30954671
F-111A	1.23352703	-0.18779885

```
> plot(pr, pch = (1:2)[cutree(cc, k = 2)],
+ col = c("black", "darkgrey")[aviones$CAR],
+ xlim = range(pr) * c(1, 1.5))
```

```
> legend("topright", col = c("black", "black",
+ "darkgrey", "darkgrey"),
+ legend = c("1 / no", "2 / no", "1 / yes", "2 / yes"),
+ pch = c(1:2, 1:2), title = "Cluster / CAR", bty = "n")
```



Los datos están representados sobre las dos primeras componentes principales, representados por el cluster al que pertenecen y en función de si pueden o no pueden aterrizar.