**Taller 7: Clasificación de razas de perros**

El objetivo que se busca al realizar el ACM del ejemplo de Razas de Perros, presentado en la sección 6.9.1, se cumple mejor complementando el ACM con la clasificación automática.

Realice una clasificación de las razas de perros utilizando todas las coordenadas factoriales del ACM previo. Obtenga los resultados con el programa estadístico que desee.

Desarrolle los siguientes puntos:

[**1. Numere los nodos del árbol con los números de la descripción de los nodos (histograma de índices).**](#_q9escrdthnlc) **1**

[**2. Describa las tres primeras uniones en la clasificación jerárquica.**](#_1g8hih94yurz) **2**

[**3. Justifique la selección de 4 clases o cambie la decisión.**](#_xpeg6csqokpt) **3**

[**4. Para la partición en cuatro clases deduzca la inercia entre clases a partir de los Índices de nivel.**](#_xd7v6y20wz84) **3**

[**5. A partir del árbol determine las razas de cada clase.**](#_iyzq526ikeu6) **4**

[**7. Qué porcentaje de inercia explica la clasificación?**](#_9ah17k9xpi2m) **4**

[**8. Resuma las características de cada una de las clases.**](#_t7zcyt520l4q) **4**

[**9. Comente el primer plano factorial del ACM incluyendo las clases obtenidas (centros de gravedad y distinción de las razas de cada clase).**](#_ic8pas510ued) **7**

[**10. Haga un resumen del análisis que responda a los objetivos del ejercicio.**](#_5nb1djv64yi6) **8**

# 1. Numere los nodos del árbol con los números de la descripción de los nodos (histograma de índices).

library(FactoClass)

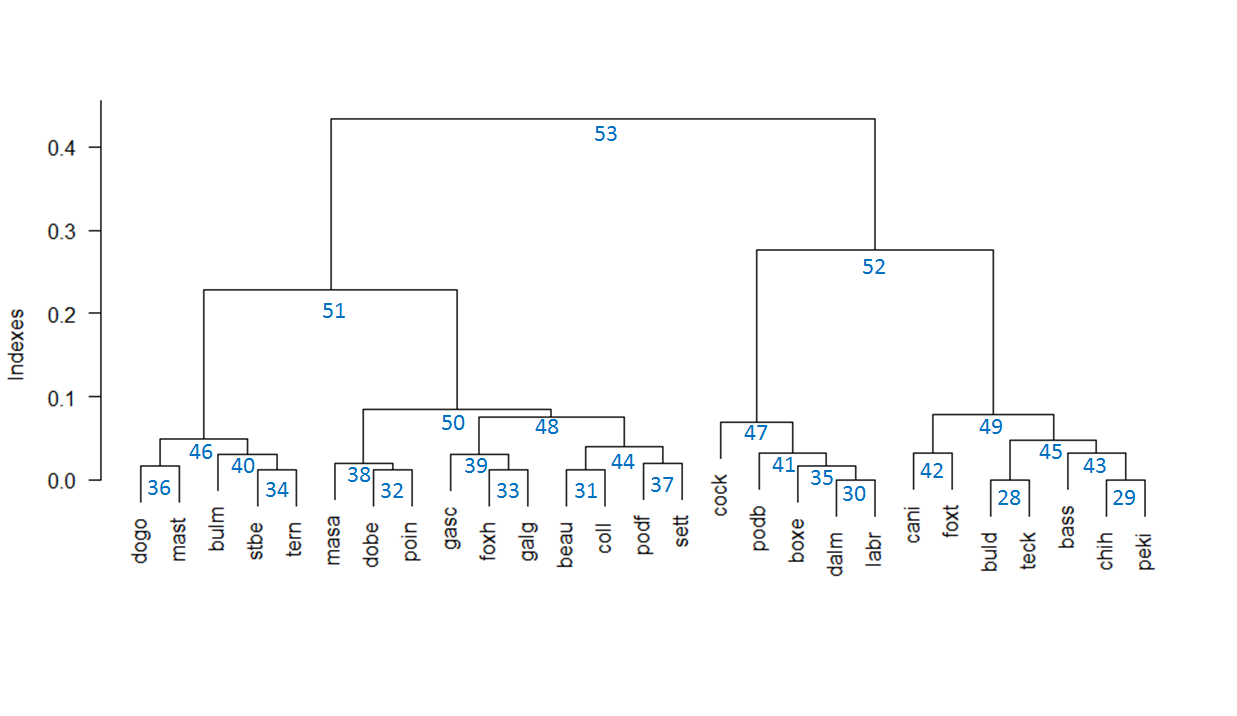
data(BreedsDogs)

Y <- BreedsDogs[-7] # active variables

Ys <- BreedsDogs[7] # ilustrative variables

fc <- FactoClass(Y,dudi.acm,k.clust=4,scanFC=FALSE,nf=2,nfcl=10,dfilu=Ys,conso=T)

grupos <- fc$cluster

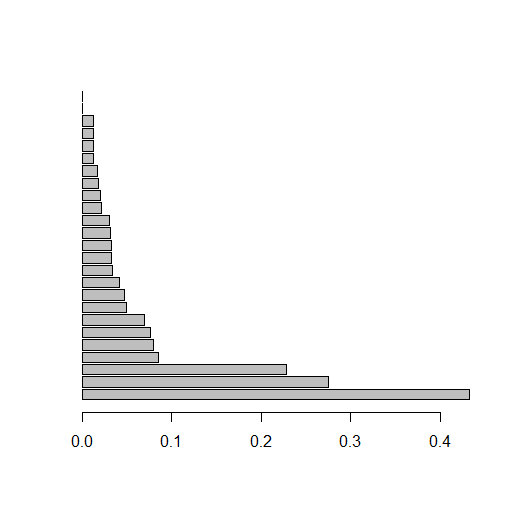


# 2. Describa las tres primeras uniones en la clasificación jerárquica.

Los primeros tres grupos presentan distancia 0 entre sus individuos; corresponden a los duplos de las razas Chihuahua-Pequinés, Buld-Tec y Dalmata-Labrador. Entre los 6 individuos comparten tener afectividad alta, agresividad baja y velocidad baja; y a nivel grupal son separados por peso, talla e inteligencia, de modo que Chihuahua-Pequinés y Buld-Teck son perros de tamaño pequeño, que difieren en su inteligencia baja y media respectivamente, mientras que los Labrador-Dálmata son de tamaño mediano y con inteligencia media.

# 3. Justifique la selección de 4 clases o cambie la decisión.

De acuerdo al histograma de índices de nivel para el ACM de las razas de perros (figura \ref{hist}), se puede observar que la mejor altura para el corte del árbol es cuando el valor del índice es 0.1. Es decir, el mayor salto en el histograma de los índices de nivel se observa cuando la inercia-entre está alrededor de 0.1, a esta altura del árbol (figura \ref{arb}) se pueden definir claramente 4 clases para agrupar las razas de perros, la cual sería la partición de mejor calidad. Lo anterior indica que en la agrupación anterior las clases estaban próximas, y luego de esta partición los grupos se encuentran necesariamente alejados (se incrementa la inercia-entre), y se pueden por tanto, diferenciar claramente las agrupaciones para las razas de perros.



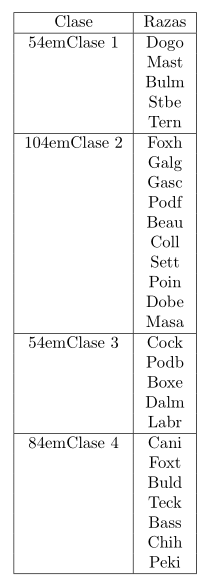
# 4. Para la partición en cuatro clases deduzca la inercia entre clases a partir de los Índices de nivel.

Inercia entre = Inercia total - inercia intra

Inercia entre = 1.6667 - 0.073

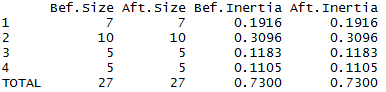
Inercia entre = 0.9366

# 5. A partir del árbol determine las razas de cada clase.



# 6. Describa el proceso de consolidación

fc$clus.summ



Ninguna de las clases tuvo cambios en el proceso de consolidación indicando que el agrupamiento jerárquico fue suficiente para encontrar grupos con una inercia intra baja.

# 7. Qué porcentaje de inercia explica la clasificación?

El porcentaje de inercia que explica la clasificación corresponde al 83% de la inercia total, el cual representa la inercia entre los cuatro grupos obtenidos.

# 8. Resuma las características de cada una de las clases.

# Basado en el análisis descriptivo desde la Pag. 23

library(FactoClass)

data(BreedsDogs)

Y <- BreedsDogs[-7] # active variables

Ys <- BreedsDogs[7] # ilustrative variables

fc <-FactoClass(Y,dudi.acm,k.clust=4,scanFC=FALSE,nf=2,nfcl=10,dfilu=Ys)

grupos <- fc$cluster

#Perfiles fila

dev.new()

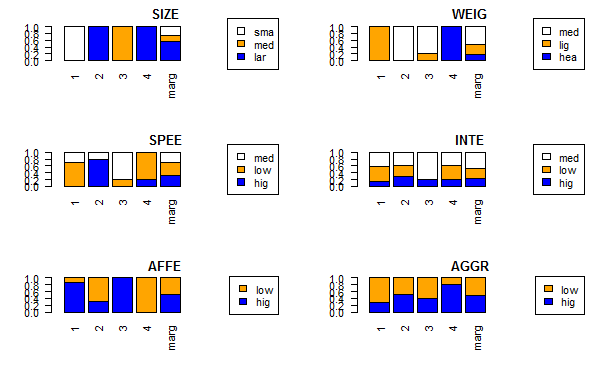
par(mfrow=c(3 ,2))

for (col in colnames(Y)) {

ct <- table(Y[,col], grupos)

plotct(ct,"col", col=c("blue","orange","white"), legend.text=T, main=col)

}



De acuerdo al gráfico de asociación entre clases y variables cualitativas de las razas (Fig. \_\_) así como al ordenamiento de valores test de las categorías de dichas variables (Tabla. \_\_) se obtiene:

La clase 1 contiene al 26% de las razas, los perros son de tamaño pequeño y peso ligero, con tendencia a poca velocidad.

La clase 2 representa al 37% de las razas de perros, se encuentra conformado por organismos de mayor tamaño y peso mediano, de gran agilidad y velocidad.

En la clase 3 se encuentran el 18% de las razas, las cuales son de tamaño mediano, velocidad media y afectividad alta.

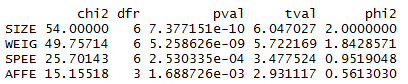
La clase 4 contiene el restante 18% de razas de perros los cuales se caracterizan por tener tamaño grande, peso alto y afectividad baja.

En general, la mayoría de las razas analizadas son de gran tamaño (55.6%) y peso medio (51.9%).

A continuación se presenta el índice de asociación entre la clasificación y las variables cualitativas:

fc <-FactoClass(Y,dudi.acm,k.clust=4,scanFC=FALSE,nf=2,nfcl=10)

chisq.carac(Y, fc$cluster)

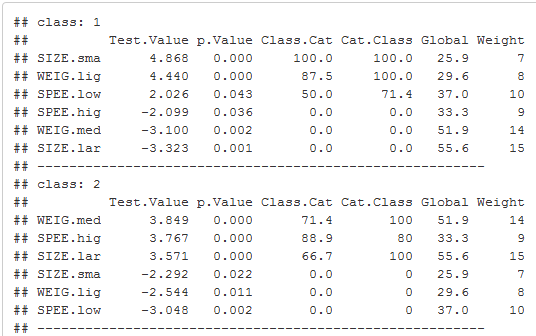


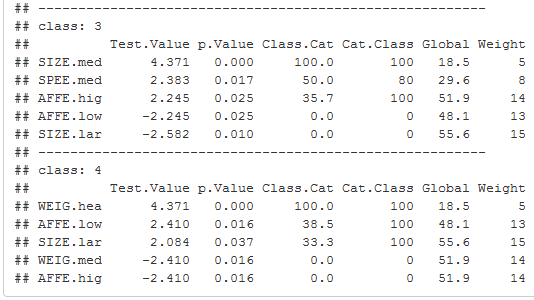
Las variables con mayor índice de asociación son tamaño y peso, indicando que son las mejores discriminatorias para clasificar las razas entre los grupos. Aunque la inteligencia y la agresividad a menudo se pueden asociar más con unas razas de perros que con otras, según este índice no son determinantes de la agrupación propuesta.

#pag. 21 ordienamiento de categorías mediante valores test

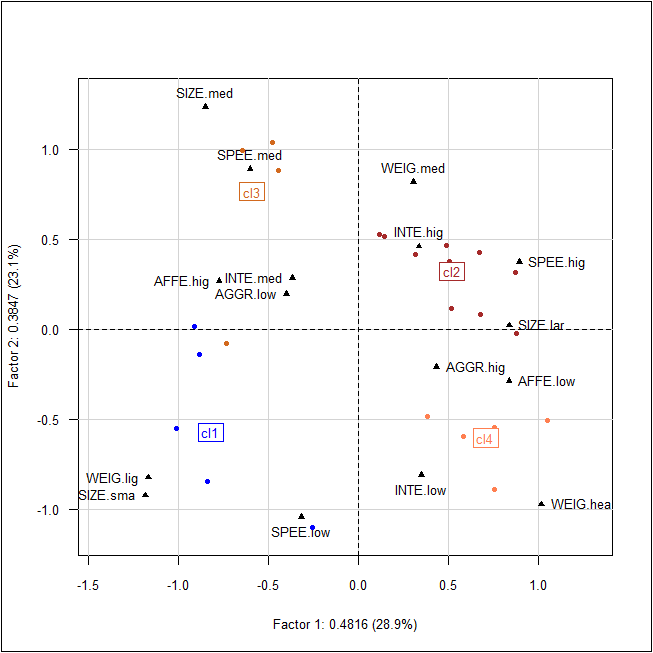
#cluster.carac(Y, fc$cluster)

fc$carac.cate





# 9. Comente el primer plano factorial del ACM incluyendo las clases obtenidas (centros de gravedad y distinción de las razas de cada clase).



El centro de gravedad de la clase 2 es el más cercano al primer eje en el primer plano factorial, comparte con la clase 4 talla, velocidad y agresividad altas y afecto bajo, difieren en la inteligencia y el peso; las características que comparten son totalmente contrastantes a las de la clase 1. Los razas asociadas a las clases 2 y 4 se observan fuertemente atraídas por sus respectivos centros de gravedad, lo cual también se refleja en el valor de su inercia (0.085 y 0.079 respectivamente), mientras que las razas asociadas a las clases 1 y 2 se observan más dispersas de su centro de gravedad y presentan los valores más bajos de inercia (0.0494 y 0.0693 respectivamente).

# 10. Haga un resumen del análisis que responda a los objetivos del ejercicio.

De entrada el ACM en su histograma de inercias permite conocer que existen tres puntos de divergencia entre las razas y posteriormente la clasificación obtenida presenta a los cuatro grupos que retienen al 83% de la variabilidad. Esto significa que las características evaluadas logran definir grupos de razas, principalmente con respecto a las variables de talla, peso y velocidad en una muy buena aproximación estadística. De acuerdo a estos rasgos, validados por tener una alta asociación con las razas de perros y su relación con la variable indicativa función, es posible seleccionar razas de perros de acuerdo con su función:

-De compañía: razas pequeñas y de baja velocidad.

-De utilidad: razas medianas, de gran peso y gran velocidad.

-De caza: razas grandes, de peso medio y gran velocidad.