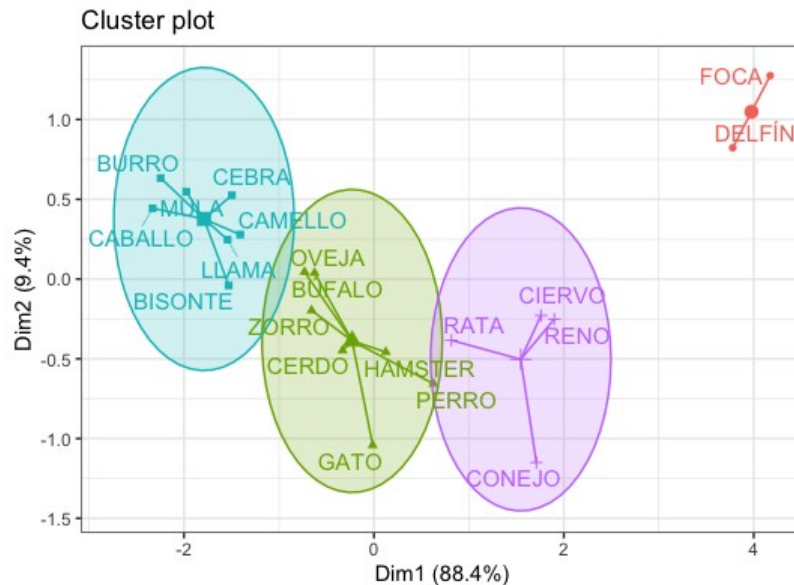


Ejercicio 1

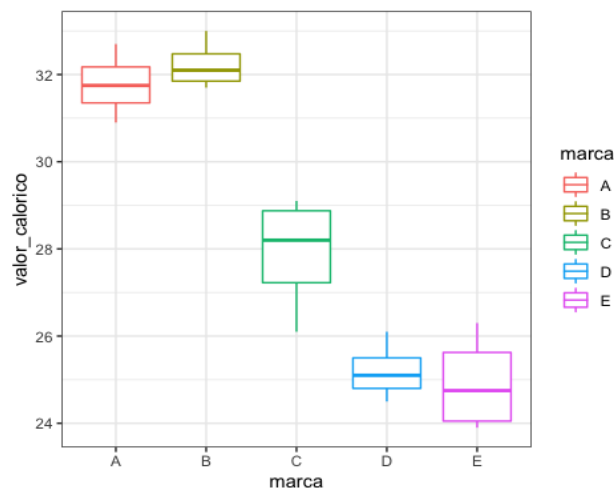
- a) Elijo como cantidad de clusters 4, con la utilización del método k-means y la distancia euclídea. Usando elbow method se pudo ver que entre la cantidad de clusters 3 y 4 hay diferencia pero ya entre 4 y 5 no. Los clusters formados son:



- a) Cluster 1: "FOCA", "DELFIN"
- b) Cluster 2: "HAMSTER", "OVEJA", "CERDO", "BÚFALO", "ZORRO", "GATO", "PERRO"
- c) Cluster 3: "RATA", "RENO", "CONEJO", "CIERVO"
- d) Cluster 4: "CABALLO", "BURRO", "CEBRA", "MULA", "CAMELLO", "LLAMA", "BISONTE"
- b) A diferencia del método jerárquico que con 4 cluster obtengo dos conjuntos de dos elementos y con método elegido obtengo una mejor separación. Al tratarse de valores continuos elegí la distancia euclídea para hacer el análisis de jerárquico.
- c) No estandarize las variables ya que la distancia usada en el método no lo requiere.
- d) N/A.
- e) Voy a elegir el cluster 2 y 4 para testear la igualdad de los vectores medios. Usando el test de Hotelling para testear la igualdad de los vectores medios obtengo un P-value: 0.00476, lo cual indica que hay evidencia estadística para rechazar la igualdad.

Ejercicio 2

- a) Se va a analizar si la media de los valores calóricos entre las distintas marcas de cerveza difiere. Los supuestos es que las muestras son independientes y las distribuciones de los residuos son normales y son homocedasticos.
- a) H_0 : para todo i, j $\mu_i = \mu_j$. Donde i, j son las distintas marcas.
- b) H_1 : existe algún i, j tal que $\mu_i \neq \mu_j$. Existen dos marcas tales que sus medias son diferentes.
- b) Hay evidencia a nivel de significación 0.95 de rechazar la hipótesis nula de que las medias de todas las marcas sean igual. El p-valor obtenido es $1.03e-08$. El test de Bartlett da un p-valor 0.5997 lo cual no rechazaría la hipótesis de normalidad y Levene dio un p-valor de 0.2831 lo cual es consistente estadísticamente entre ambos. Los test de normalidad de Shapiro y Anderson-Darling dieron un p-valor de 0.7 y 0.6087 respectivamente con lo cual no rechaza la normalidad de las distribuciones. Como los supuestos se cumplen y el test de ANOVA rechaza la hipótesis nula puedo decir que existe evidencia estadística para decir que las medias son diferentes. Adjunto box plot que confirma visualmente.



- c) N/A.
- d) Usando Tukey se puede ver que las marcas A y B tienen poca diferencia dentro de las que son blancas y entre las negras las marcas D y E son similares entre sí, mientras que ambas con respecto a C son muy diferentes. Además el grupo A, B y D, E contiene al 0 lo cual indica que no son tan diferentes.
- e) N/A.

Ejercicio 3

- a) El test de Hotelling dio un p-valor 0.0009905 lo que indica que tiene sentido seguir haciendo un análisis de discriminante.
- b) Los supuestos para el análisis son de muestras independientes, de distribución normal y matrices de varianzas y covarianzas iguales. Utilizando Shapiro tengo una evidencia estadística de que las distribuciones no son normales. Y realizando el test de Box-M con un p-valor de 0.001754, puedo rechazar la hipótesis de las matrices de varianzas iguales.
- c) Usaría un análisis de discriminante robusto ya que no se cumplieron los supuestos y no se observan outliers. O bien uno cuadrático ya que no son linealmente separables.