

Trabajo practico 1

Ejercicio 2

Grupo

12 mayo, 2021

A. Cuantifique (en forma manual) la similaridad entre las variedades correspondientes a la primera y segunda fila en función del porcentaje de caracteres comunes respecto al número de caracteres totales. Idem las variedades asociadas a las filas 12 y 13 (incluir la variable TIPO)

Obtenemos dos vectores que representan a cada una de las dos primeras variedades y luego calculamos el la proporción de variables donde ambas variedades coinciden.

```
variedad1 <- datos[1, -1] %>% unlist() %>% unname()
variedad2 <- datos[2, -1] %>% unlist() %>% unname()
similaridad_1_2 <- mean(variedad1 == variedad2, na.rm = TRUE) * 100
```

La similaridad entre las variedades de la fila 1 y 2 es del 88.89%. Luego, de manera analoga par de variedades 12 y 13

```
variedad12 <- datos[12, -1] %>% unlist() %>% unname()
variedad13 <- datos[13, -1] %>% unlist() %>% unname()
similaridad_12_13 <- mean(variedad12 == variedad13, na.rm = TRUE) * 100
```

La similaridad entre las variedades de la fila 12 y 13 es del 71.43%.

B. Halle una matriz de similaridad entre variedades en función del coeficiente SM generalizado.

Para esta tarea utilizamos la función `daisy()` del paquete `cluster`, donde especificamos que la métrica a utilizar es "gower", y nos devuelve la matriz de distancia entre las diferentes variedades. En la siguiente línea, convertimos la matriz de distancia a matriz de similaridad.

```
matriz_distancia <- daisy(datos_fct_df, metric = "gower")
matriz_similaridad <- 1 - matriz_distancia
```

Comprobemos, por ejemplo, si la similaridad entre las unidades 1 y 2 computada mediante `daisy()` es igual a la que computamos a mano.

```
matriz_similaridad[1] == (similaridad_1_2 / 100)
```

```
## [1] TRUE
```

Por lo que ambos resultados son iguales.

C. Aplique Análisis de Coordenadas principales para representar en un espacio bidimensional la semejanza entre las variedades.

Aplicamos el análisis de coordenadas principales de la siguiente manera:

Luego, graficamos a la caracterización cualitativa de las variedades de pepino en el plano principal.

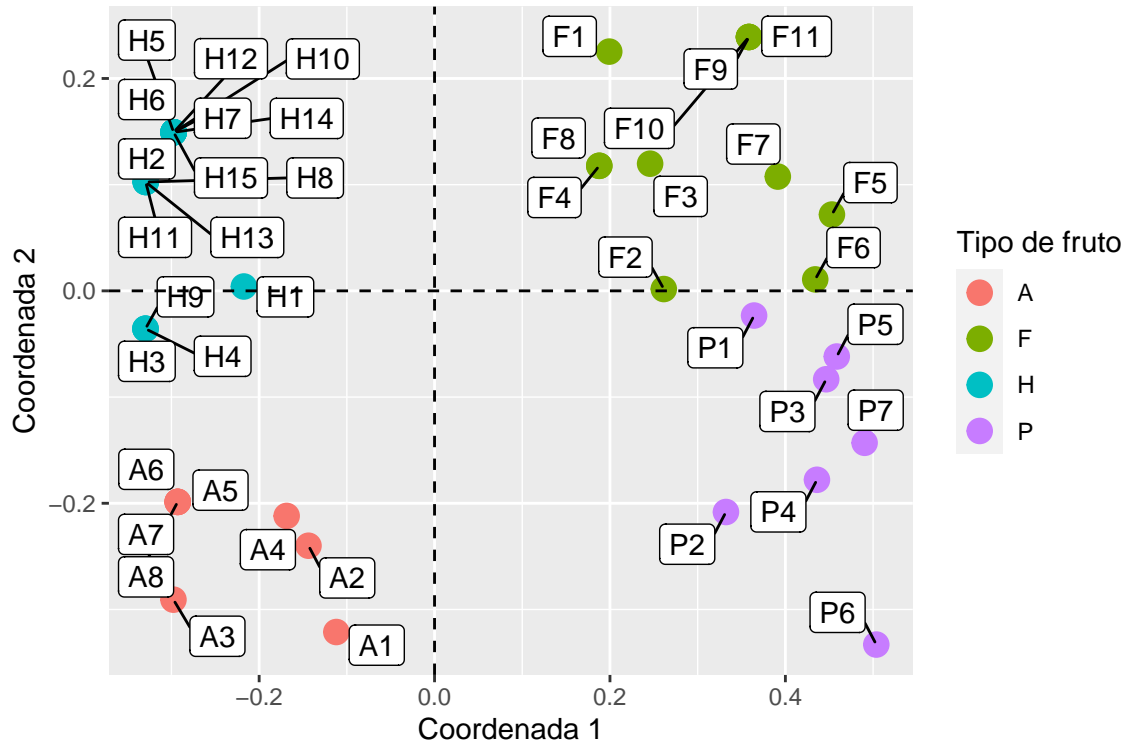


Figure 1: Caracterización cualitativa de las variedades de pepino.

En aquellos casos que mas de una etiqueta apunta hacia el mismo punto sucede que las variedades coinciden en terminos de las variables analizadas y en consecuencia los puntos que los representan estan encimados. Por ejemplo, las variedades A5, A6, A7, A8 tienen idénticos valores para todas las variables.

VARI...	TIPO	SEXO	CORN...	TORN...	PEDUNC	VERR...	MOTEADO	CLADOSP	CMV
A5	Alf	Fem	Negra	Espin	Cue	No	No	No	No
A6	Alf	Fem	Negra	Espin	Cue	No	No	No	No
A7	Alf	Fem	Negra	Espin	Cue	No	No	No	No
A8	Alf	Fem	Negra	Espin	Cue	No	No	No	No

D. Conforme grupos de variedades según su homogeneidad en la caracterización agronómica cualitativa.

En la Figura 1 se puede ver que las variedades de los tipos de fruto **A** y **H** se agrupan de manera que respetan al tipo de fruto y se diferencian del resto. También existe una agrupación, ya no tan clara, para las variedades de los grupos **F** y **P**. En este caso, si no estuviera el color que diferencie a los tipos de frutos, no podríamos diferenciar a estos dos grupos claramente. Por ejemplo, las variedades **F2**, **F6** y **P1** están muy cercanas en el plano y las podríamos haber tomado como parte de un mismo grupo.

También podemos ver que la variabilidad de las variedades dentro de cada tipo de fruto difiere. Por ejemplo, para el tipo de fruto **H**, se tiene que casi todas las variedades se corresponden con dos categorizaciones particulares (por eso vemos tantos puntos encimados). Por otro lado, todas las variedades del tipo de fruto **P** se corresponden con una configuración única de las variables cualitativas.

E. Encuentre el dendrograma ultramétrico con ligamiento UPGMA correspondiente

Obtenemos el dendrograma utilizando la función `hclust()`, a la que le pasamos la matriz de distancia previamente obtenida.

Luego graficamos el dendrograma y mostramos con diferentes colores a los 4 clusters obtenidos.

COMENTAR

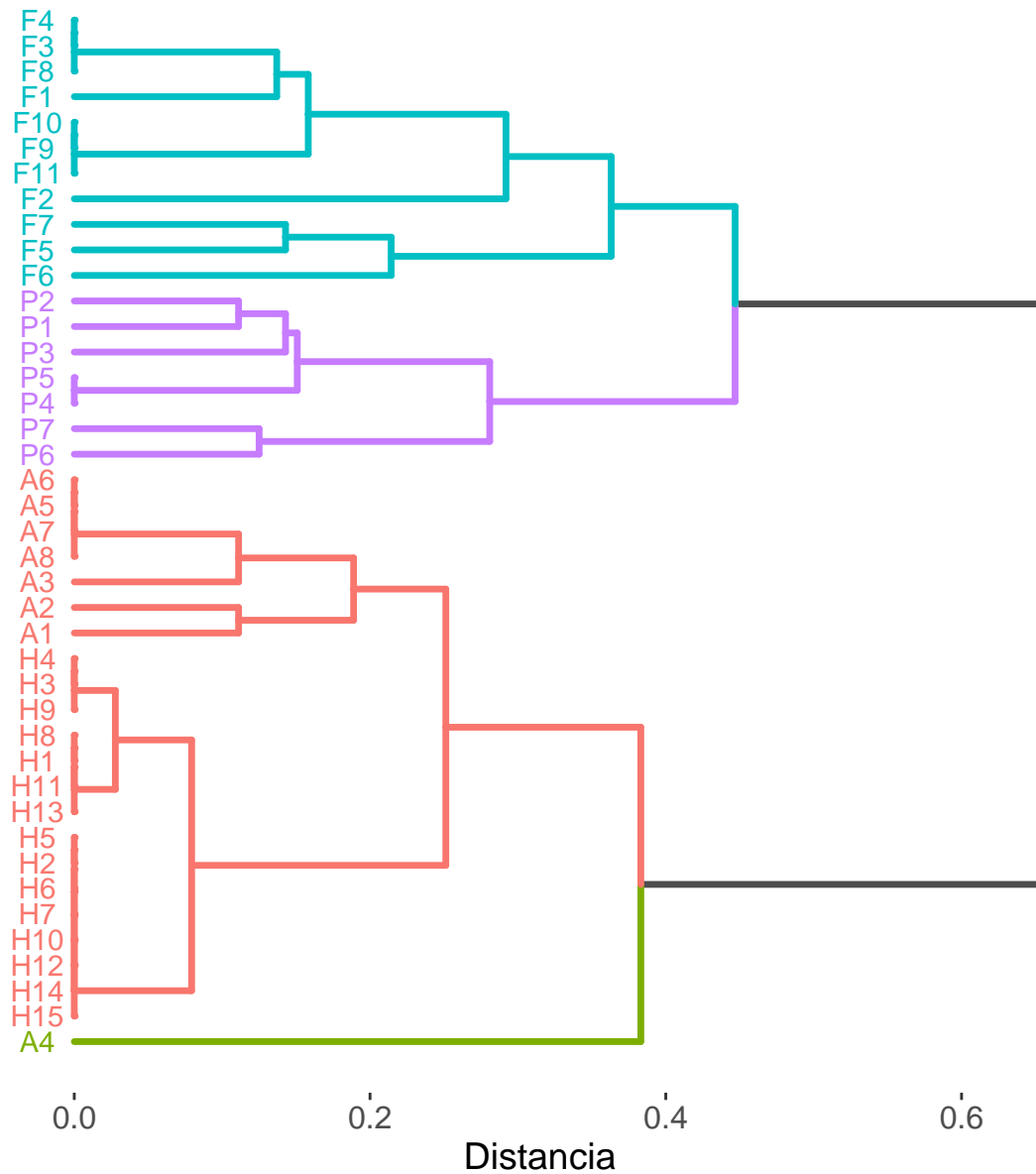


Figure 2: Dendrograma Ultrametrico con ligamiento UPGMA.

F. Mida a través de su matriz cofenética la concordancia con la matriz de distancias que le dio origen

Primero vamos a calcular la concordancia y luego obtenemos un grafico de dispersion donde se muestra la distancia original y la distancia cofenética.

```
distancia_cofenetica <- cophenetic(cluster_cualitativas)
concordancia <- cor(distancia_cofenetica, matriz_distancia)
```

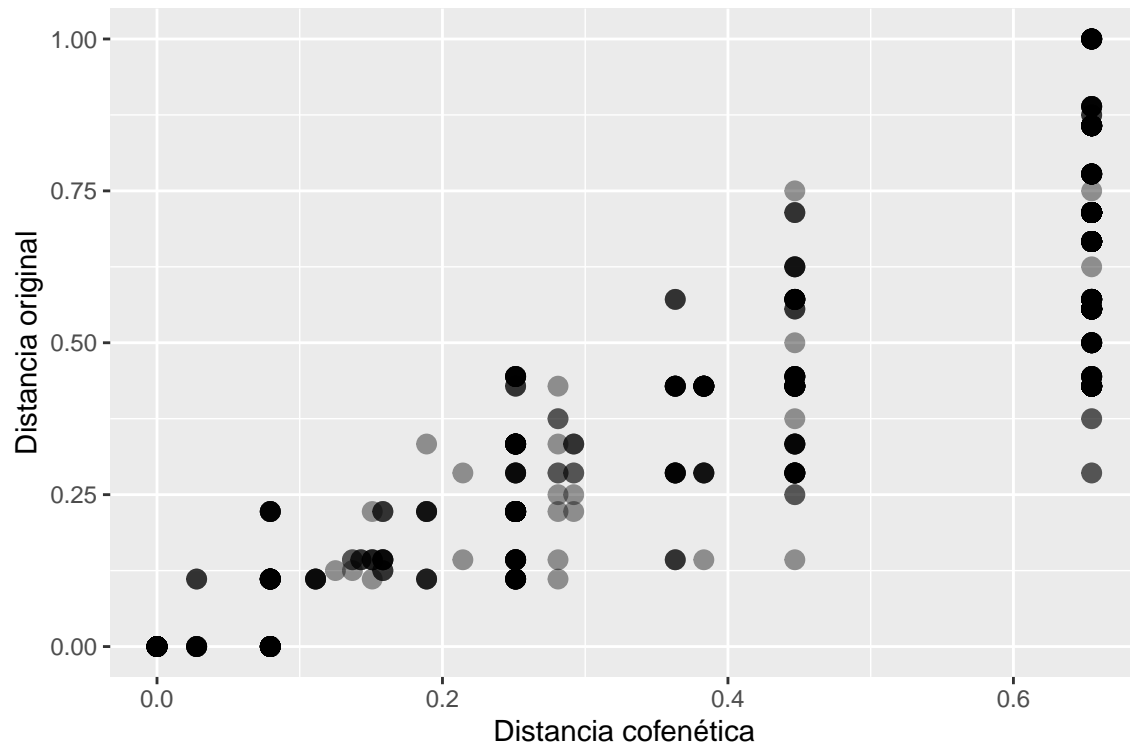


Figure 3: Grafico de dispersion entre distancia original y distancia cofenética.

La concordancia entre la matriz de distancias cofenética y la matriz de distancia original es igual a 0.883.

G. Cuantifique concordancia entre plano principal de ACoordP y Cluster

La concordancia concordancia entre plano principal de ACoordP y Cluster es 0.891.