

# Ejercicio 1

Para obtener una única caracterización agronómica incluyendo toda la información brindada en los archivos CUALITATIVAS y CUANTITATIVAS vistos en el TP Parte I, una alternativa válida es trabajar con el Coeficiente General de Similitud de Gower que permite considerar variables cuantitativas y cualitativas en forma simultánea.

- A) Calcule a mano este coeficiente para medir el grado de asociación entre las variedades A1 y A2, y F4 y F5 (filas 1, 2, 12 y 13 respectivamente).

Definimos la siguiente función para calcular el coeficiente de Gower. El primer argumento, `individuos`, recibe un vector de longitud 2 con los dos individuos para los cuales se quiere computar el coeficiente de Gower. El segundo argumento, `datos`, es un data frame con las variables a utilizar.

```
coeficiente_gower <- function(individuos, datos) {
  numerador <- 0
  denominador <- 0

  # Contribucion categoricos
  datos_c <- datos[sapply(datos, function(x) is.character(x) | is.factor(x))]
  x <- datos_c[individuos[1], ]
  y <- datos_c[individuos[2], ]
  numerador <- numerador + sum(x == y, na.rm = TRUE)
  denominador <- denominador + sum(!is.na(x == y))

  # Contribucion numericos
  datos_n <- datos[sapply(datos, is.numeric)]
  rangos <- sapply(datos_n, function(x) {
    max(x, na.rm = TRUE) - min(x, na.rm = TRUE)
  })
  x <- datos_n[individuos[1], ]
  y <- datos_n[individuos[2], ]
  numerador <- numerador + sum(1 - abs(x - y) / rangos)
  denominador <- denominador + ncol(datos_n)

  # Resultado
  numerador / denominador
}
```

Luego, simplemente hacemos

```
gwr1 <- coeficiente_gower(c("A1", "A2"), datos)
gwr2 <- coeficiente_gower(c("F4", "F5"), datos)
```

El coeficiente de Gower para los individuos "A1" y "A2" es 0.9271. Esto quiere decir que de todas aquellas variables medidas, ya sean cuantitativas o cualitativas, los individuos "A1" y "A2" comparten el 92.71% de los caracteres observados. Por otro lado, para los individuos "F4" y "F5" este coeficiente es del 0.7827. El mismo indica que los individuos "F4" y "F5" se parecen bastante, aunque no tanto como "A1" y "A2".

B) Recorra al software para encontrar la matriz de similitud de Gower entre todos los individuos.

**Nota:** `daisy()` no funciona con variables de tipo "character", por lo que las convertimos a tipo "factor".

```
datos_fct <- as.data.frame(unclass(datos), stringsAsFactors = TRUE)
distancia <- daisy(datos_fct, metric = "gower")
similitud <- 1 - distancia
```

C) Aplique Análisis de Coordenadas Principales para obtener la configuración conjunta cuali-cuantitativa de las variedades de pepino

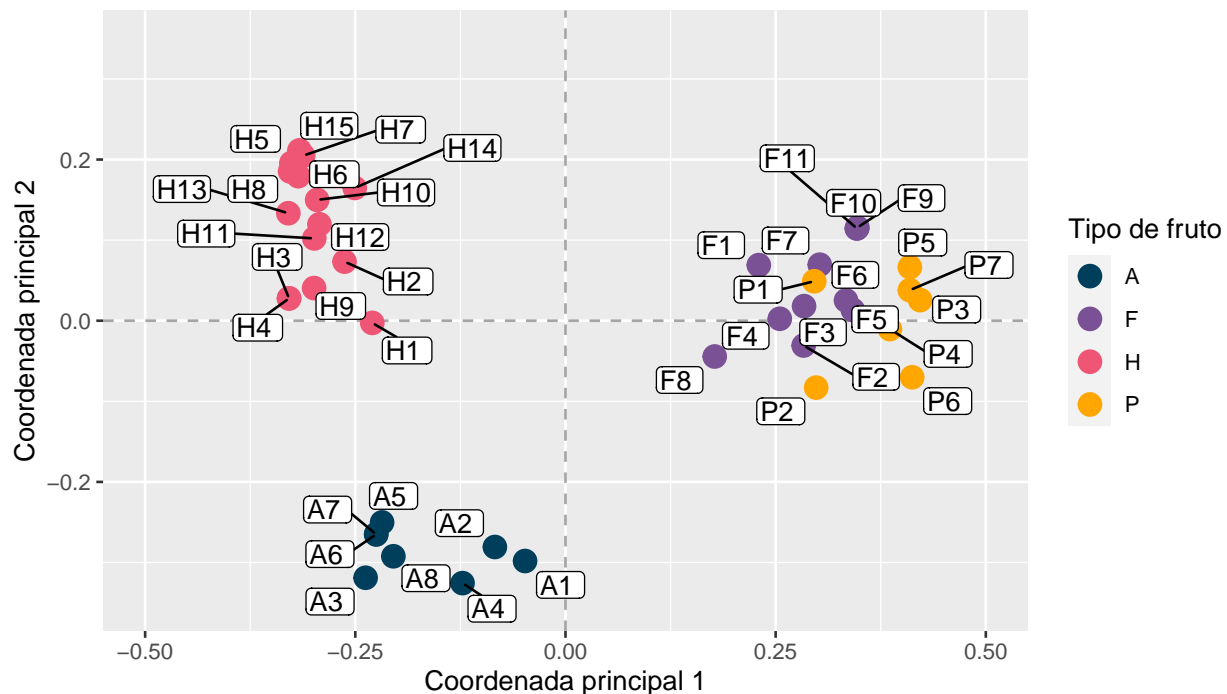


Figure 1: Caracterización de las variedades de pepino en el plano principal.

En la **Figura 1** se puede ver que el análisis de correspondencias principales conserva de muy buena manera el agrupamiento para los tipos de pepino alpha-beta (A) y holandes (H) en el plano principal. Por otro lado, no se puede obtener un agrupamiento tan claro para los pepinos de tipos franceses (F) y pepinillo (P). Es posible ver que los pepinillos tienden a tener mayor valor en la primera coordenada que los franceses. Sin embargo, sino fuera por los colores que distinguen a los tipos de pepino, uno agruparía a los pepinos franceses y pepinillos en una misma clase.

Ahora comparamos este resultado con el análisis de procrustes del Ejercicio 6 del primer trabajo práctico. Podemos concluir que esta configuración en el plano principal y la configuración de consenso en el plano principal de APG logran ordenamientos similares. Sin embargo, el ordenamiento de consenso del APG funciona un poco mejor ya que logra diferenciar más claramente a los pepinillos de los franceses.

D) Halle un cluster con encadenamiento UPGMA en función de la distancia de Gower

En la **Figura 2** se observan los agrupamientos resultantes del cluster con encadenamiento UPGMA. Este método logra discriminar de manera perfecta a los individuos por tipo de fruto. Es esperable que este agrupamiento discrimine mejor que el agrupamiento obtenido en el plano principal de ACP ya que se enriquece de las distancias en el espacio original, mientras que el plano principal en la **Figura 1** es una proyección de una configuración en un espacio de mayor dimensionalidad.

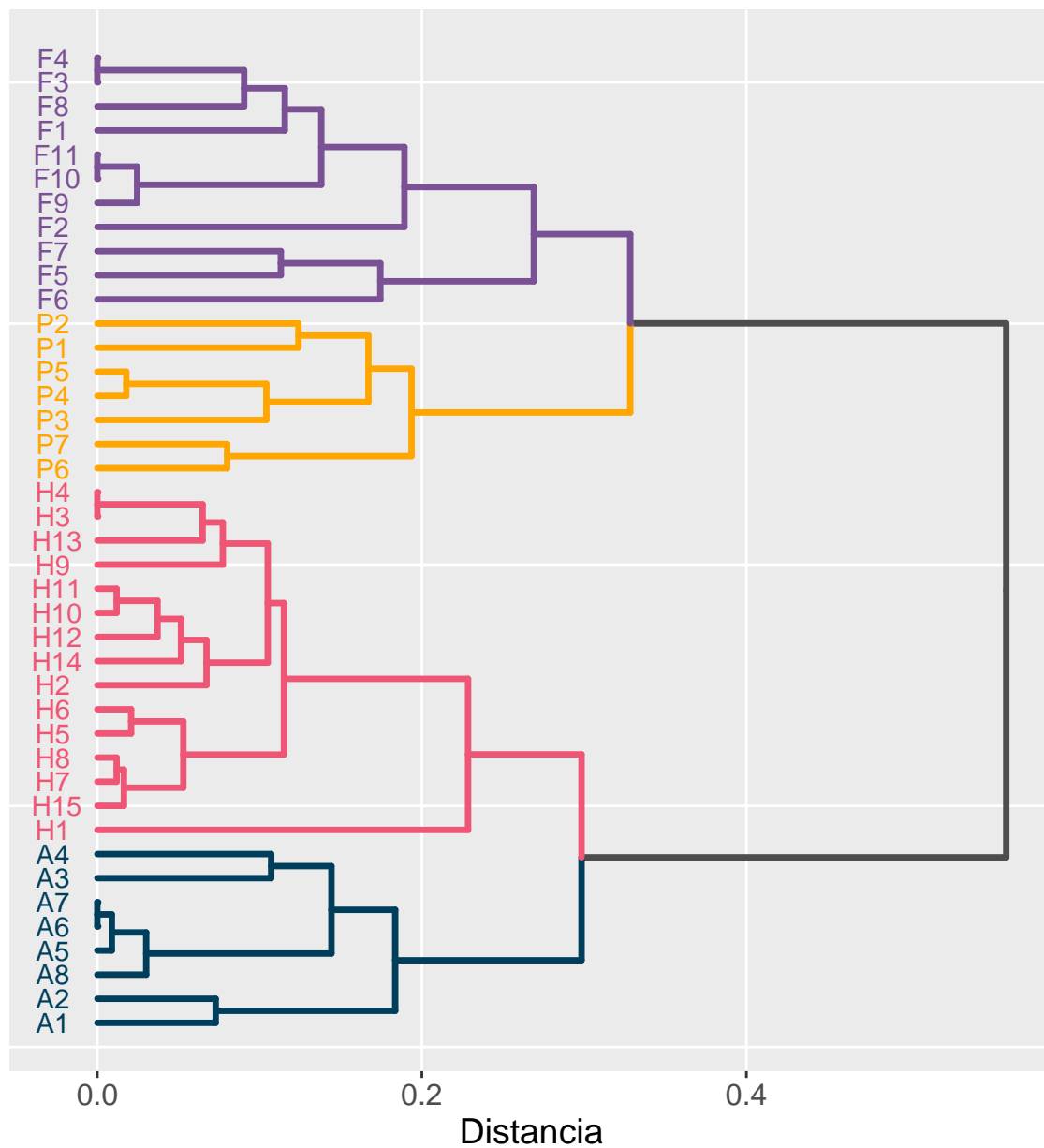


Figure 2: Dendrograma ultrametrico con ligamiento UPGMA en función de la distancia de Gower .

E) Compare resultados de (c) con los hallados en ejercicio 4 de Parte I. Cuantifique la relación con la configuración de consenso (bidimensional) encontrada con APG

En la **Figura 3** se observa la alta relación entre las configuraciones de consenso bidimensional encontrada con APG y la configuración de consenso obtenida mediante ACP, que se observa en la **Figura 1**. La correlación entre ambas configuraciones es 0.9835. Es decir que ambos análisis son altísimamente consistentes.

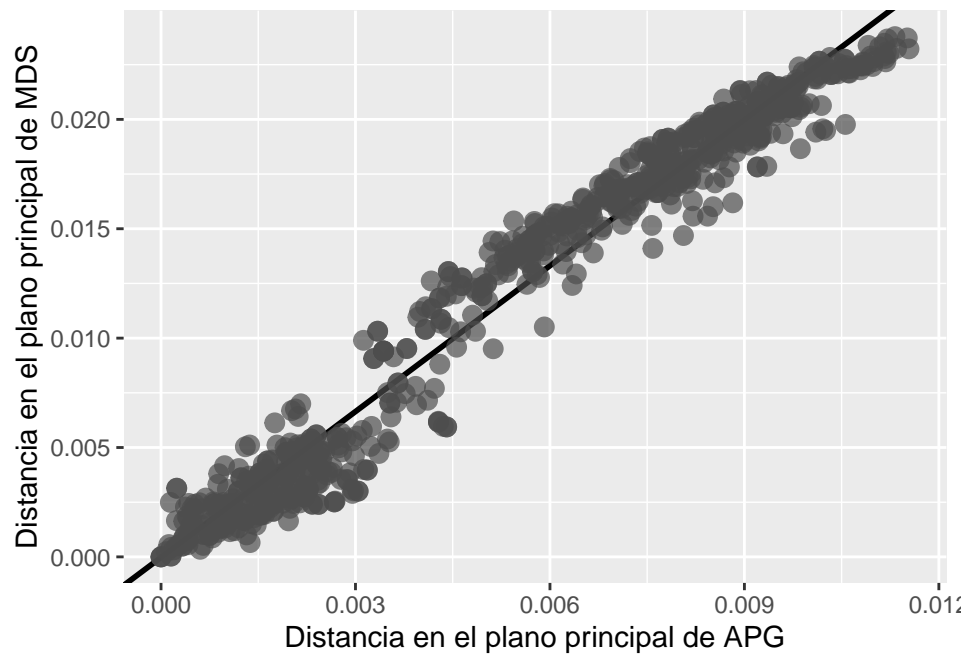


Figure 3: Concordancia entre las distancias segun APG y ACP. La linea negra representa a la regresion lineal entre ambas distancias.

**NOTA:** Que les parece si borramos este grafico para ahorrar espacio?