

Ejercicio 1

- A) En el ejercicio 7b TP Parte I, se realizó un ACP por cada ambiente. ¿Cuánto vale el primer autovalor de cada ACP? ¿Qué ocurriría si se hace un ACP de los ambientes concatenados pero sin utilizar la ponderación del AFM?

En el caso del ambiente Pergamino el primer autovalor es 4.67, y en el caso del ambiente Ferre es 3.62.

Si concateno las matrices sin realizar la ponderación del AFM, el ambiente Pergamino va a tener mayor peso que el ambiente Ferre en la construcción de un ACP de compromiso porque su primer autovalor, que representa la varianza en la primera dirección principal, es mayor.

En cambio, al ponderar a cada tabla por el inverso del primer autovalor, el primer autovalor de cada tabla ponderada es 1 y en consecuencia ambos ambientes tienen el mismo peso en el ACP global.

Realice un AFM y responda las siguientes cuestiones:

- B) ¿Cuál es el porcentaje de explicación del plano principal?

El porcentaje de variabilidad explicada en el plano principal es 51.92%.

- C) ¿Qué puede decir a partir de los coeficientes RV , L_g y Ng ? Analice el gráfico de los ambientes (condiciones).

Table 1: Coeficiente RV

	Pergamino	Ferre	MFA
Pergamino	1.000	0.567	0.867
Ferre	0.567	1.000	0.902
MFA	0.867	0.902	1.000

Table 2: Coeficiente L_g

	Pergamino	Ferre	MFA
Pergamino	1.268	0.830	1.146
Ferre	0.830	1.690	1.377
MFA	1.146	1.377	1.379

En la **Tabla 1**, vemos que el coeficiente RV entre los ambientes Pergamino y Ferre es de 0.57. Esto significa que las configuraciones de las variedades de maíz en ambos ambientes son medianamente similares. En otras palabras, si bien no hay un desentendimiento total entre estas dos configuraciones, tampoco presentan estructuras homotéticas¹. Por otro lado, los valores del coeficiente RV entre las representaciones parciales de cada ambiente y la representación promedio del AFM es de 0.87 para Pergamino y 0.90 para Ferre. Estos valores indican que la representación de consenso se parece bastante a las representaciones parciales de cada ambiente, siendo la configuración en el ambiente Ferre levemente más similar a la configuración promedio.

¹Si alguien nos pregunta: ¿Cuánto se parecen? Responderíamos en criollo *ni tan tan, ni muy muy*.

En cuanto a la **Tabla 2**, vemos que el mayor valor del coeficiente L_g lo presenta el ambiente Ferre. Esto significa que los datos en este ambiente tienen mayor riqueza, o menor redundancia, que para el ambiente Pergamino.

El valor de los coeficiente N_g , es decir, la raíz cuadrada de los coeficientes L_g de un grupo respecto de si mismo, son 1.12 y 1.3 para Pergamino y Ferre, respectivamente. Esto significa que el ambiente Pergamino aporta casi de manera exclusiva en una sola dimension, mientras que el ambiente Ferre tiene una dimensionalidad un poco mayor a 1, por lo que aporta un poco en una segunda dimension.

Naturalmente, la informacion obtenida mediante el coeficiente N_g es redundante con lo que concluimos mediante el coeficiente L_g . En otras palabras, en el ambiente Ferre se tiene una descripcion mas rica y menos redundante de las variedades de maiz que en el ambiente Pergamino.

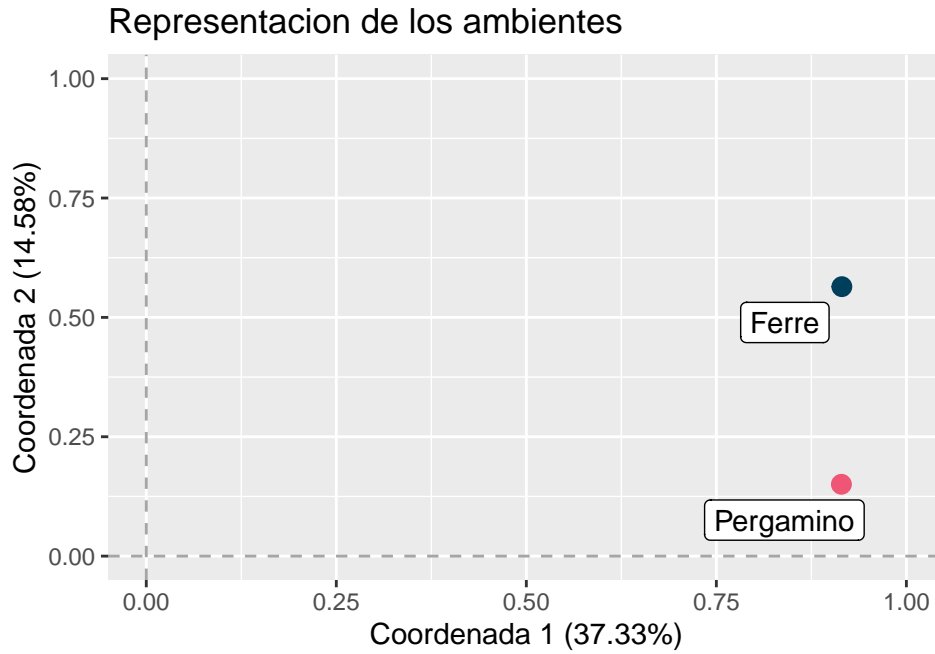


Figure 1: Caracterizacion de los ambientes en el plano principal AFM.

En el grafico de los ambientes se puede ver que ambos ambientes tienen relacion muy alta con la primera coordenada, y que solamente el ambiente Ferre tiene relacion con la segunda coordenada. Esto es consistente con lo que pudimos apreciar mediante el coeficiente N_g , donde vimos que la dimensionalidad del ambiente Pergamino era practicamente 1, mientras que la del ambiente Ferre era mayor.

D) Analice si encuentra indicios de interacción tanto genotipo-ambiente como variable-ambiente.

En la Figura 2 podemos analizar la interaccion genotipo-ambiente. Allí se puede ver que muchos genotipos de maiz presentan caracterizaciones muy distintas en ambos ambientes. A simple vista, uno puede resaltar a las variedades 38, 90, y 22 como algunas de las variedades con mayor interaccion genotipo ambiente, ya que las configuraciones parciales difieren mucho entre ellas en el plano principal. En otras palabras, para estas variedades, la caracterizacion promedio que se obtiene mediante AFM no es representativa de su comportamiento en ninguno de los ambientes estudiados.

Por otro lado, tambien hay poblaciones de maiz que presentan muy poca interaccion genotipo-ambiente, como por ejemplo las variedades 44, 102, y 92.

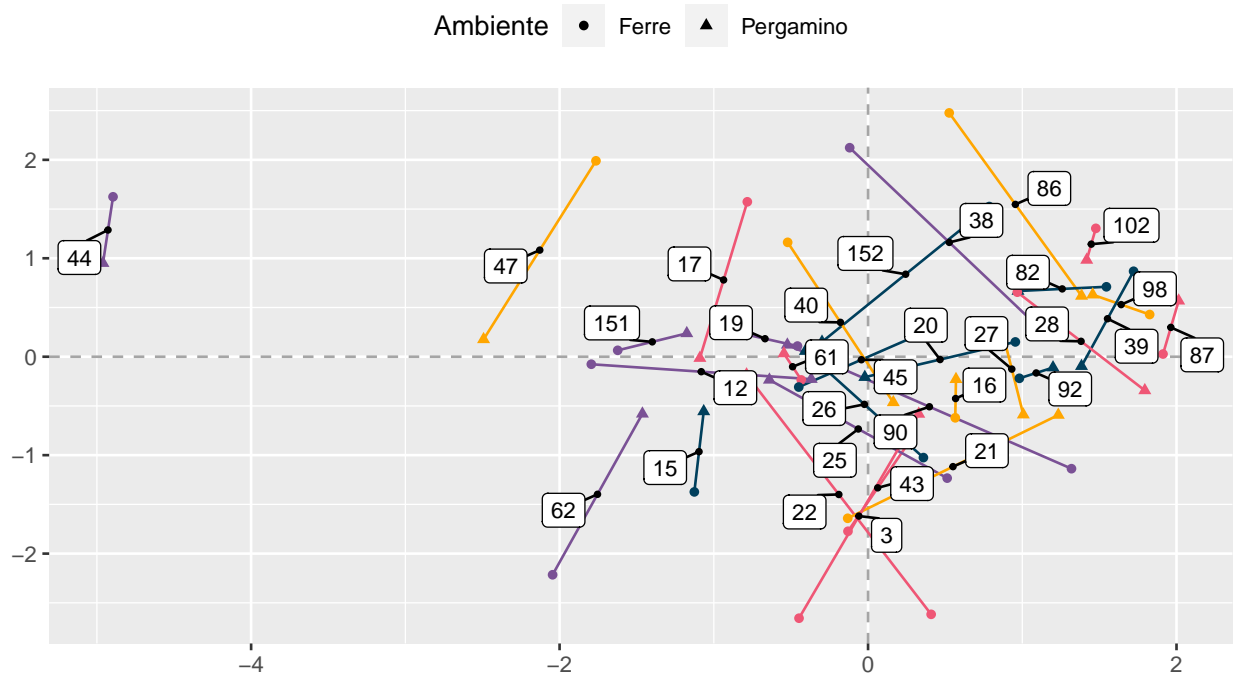


Figure 2: Caracterizaciones parciales y de consenso mediante AFM.

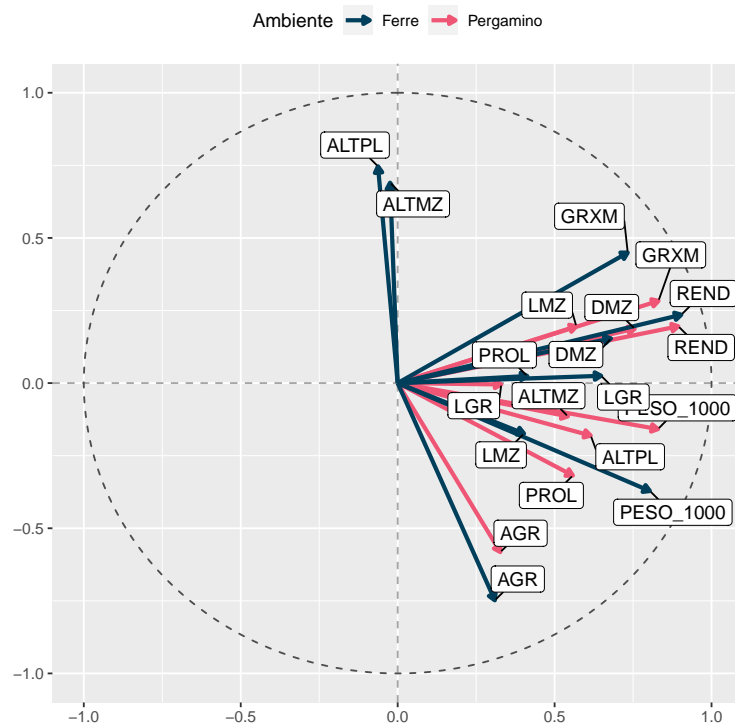


Figure 3: Caracterizacion de las variables en el plano principal AFM.

En el caso de la interaccion variable-ambiente, miramos el angulo que forman los dos vectores que

representan a la misma variable en ambos ambientes en la Figura 3. Nuevamente, podemos concluir que hay interaccion variable-ambiente. Por ejemplo, tenemos que en Pergamino se da que la altura de la planta y de la mazorca se relacionan positivamente con el rinde, indicando que plantas con mayor altura y mazorcas mas largas se asocian a rindes mayores. Sin embargo, esta asociacion no sucede en Ferre, donde vemos que el rendimiento de la planta no se asocia a estas variables de altura, ya que forman un angulo aproximado de 90 grados.

- E) ¿Cuáles son las poblaciones que tuvieron un comportamiento más diferenciado entre ambos ambientes en el plano principal? (nombrar 5 para cada dimension).

Para realizar este trabajo vamos a obtener los valores de la inercia *within* de cada individuo, para el plano principal.

Table 3: Poblaciones de maiz con mayor inercia en cada dimension del plano principal

	Inercia		Inercia
90	17.172	22	7.248
12	10.288	3	3.050
21	9.434	38	8.447
38	8.447	86	3.726
22	7.248	47	2.688

A la izquierda, tenemos los valores de inercia para la primer dimension, y al a derecha, los valores de inercia para la segunda dimension. Las poblaciones de maiz que se incluyen son aquellas 5 con mayor inercia en cada dimension.

En otras palabras, las 5 poblaciones con mayor diferencia en el comportamiento entre ambientes en terminos de la primera dimension son las 90, 12, 21, 38 y 22. Para la segunda dimension, son las poblaciones 22, 3, 38, 86 y 47.

- F) ¿Cuáles son las variables que más contribuyen a la determinación de los ejes?

Para determinar que variables presentan la mayor contribucion con los ejes vamos a tomar como referencia a la maxima correlacion y nos quedamos con aquellas variables cuya correlacion sea al menos dos tercios de esta cantidad.

En la **Tabla 4** podemos ver que las variables mas asociadas con la primer dimension son el diametro de mazorca, la cantidad de granos por metro, el peso de 1000 granos, y el rendimiento, para ambos ambientes.

El hecho de que para ambos ambientes tengamos las mismas variables altamente correlacionadas con la primera dimension nos indica que la relacion entre estas variables es consistente en ambos ambientes.

En cuanto al segundo eje, vemos que esta asociado de manera negativa con el ancho del grano en ambos ambientes, y de manera positiva con la altura de la planta y la altura de la mazorca en el ambiente Ferre.

Que esta dimension este principalmente asociada con variables del ambiente Ferre es consistente con lo que observamos en el **item C** de este ejercicio, donde vimos que la dimensionalidad para el ambiente Ferre era mayor que para el ambiente Pergamino. Esto se da, principalmente, porque en el ambiente Pergamino las variables altura de mazorca y altura de planta estan muy correlacionadas positivamente con variables tales como el rendimiento, mientras que en el ambiente Ferre son practicamente independientes, requiriendo mas de una dimension para reflejar la riqueza de los datos.

Table 4: Correlacion entre variables y el primer eje

Variable	Ambiente	Correlacion
DMZ	Pergamino	0.755
GRXM	Pergamino	0.831
PESO_1000	Pergamino	0.829
REND	Pergamino	0.894
DMZ	Ferre	0.680
GRXM	Ferre	0.733
PESO_1000	Ferre	0.804
REND	Ferre	0.904

Table 5: Correlacion entre variables y el segundo eje

Variable	Ambiente	Correlacion
AGR	Pergamino	-0.582
AGR	Ferre	-0.748
ALTPL	Ferre	0.747
ALTMZ	Ferre	0.690

- G) Busque en el Help de R el comando HCPC. Utilícelo con los datos de este ejercicio y comente el resultado (Usar UPGMA).

```
hcpc <- HCPC(mfa, nb.clust = 4, method = "average", graph = FALSE)
```

- H) Arme grupos de individuos y realice una caracterización.

Los cuatro grupos de individuos pueden verse en la **Figura 5**. El primero que se destaca, es el grupo numero 1, que contiene solamente a la variedad 44 de maíz. Esta variedad resalta en ambos ambientes por su bajísimo rinde y otras características no deseables para una variedad de maíz como pueden ser los granos pequeños y livianos.

El grupo 2 presenta variedades de maíz cuyo rendimiento es mas bajo que el promedio. Estas variedades además presentan granos livianos y mediciones morfológicas mas tirando a bajas que medias. Excepcionalmente, alguna de las variedades de este grupo, como puede ser la 47 y 17, están asociadas a plantas y mazorcas altas en el ambiente Ferre. Pero como mencionamos anteriormente, esto no se asocia a mejor rinde.

Por otro lado, el grupo numero 3 se compone de variedades que en ambos ambientes presentan granos anchos y rendimientos cercanos al promedio, o incluso mas bajos que el promedio (esto se da particularmente para la variedad 62). También notamos que las variedades de este grupo presentan plantas y mazorcas mas bajas que el promedio para el ambiente Ferre.

Finalmente, el grupo 4. Estas variedades se asocian principalmente con rindes altos, granos pesados, y muchos granos por metros. Esto es consistente con que estas variedades no están asociadas a granos anchos, por ejemplo.

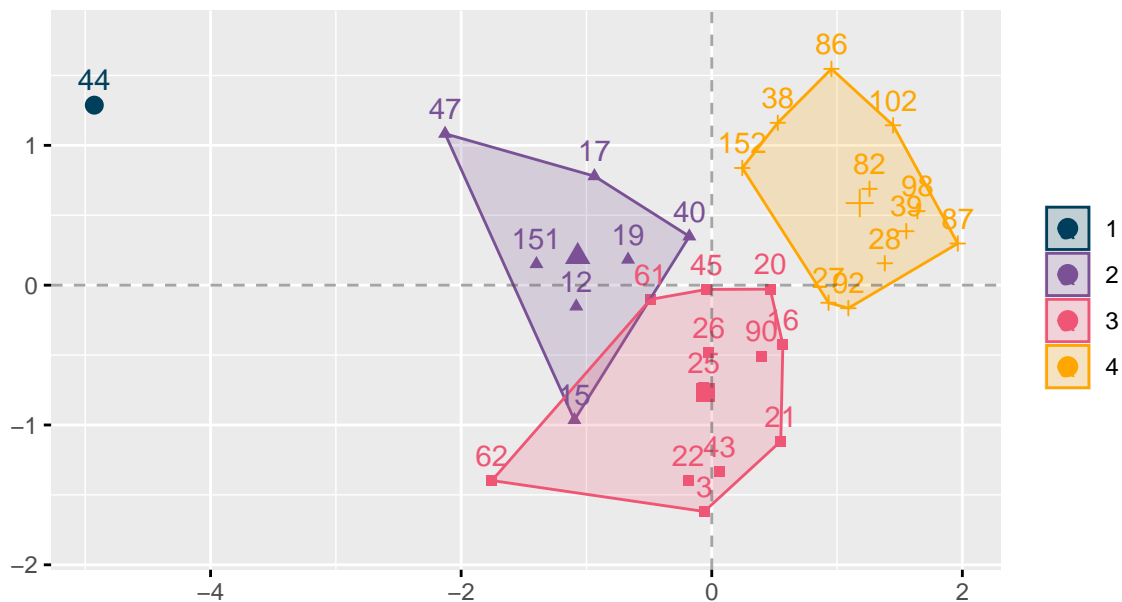


Figure 4: Agrupamiento basado en HCPC.

- I) Obtenga la matriz de distancias euclídeas entre individuos en el plano principal del AFM. Compare dicha matriz con la obtenida a partir de APG.

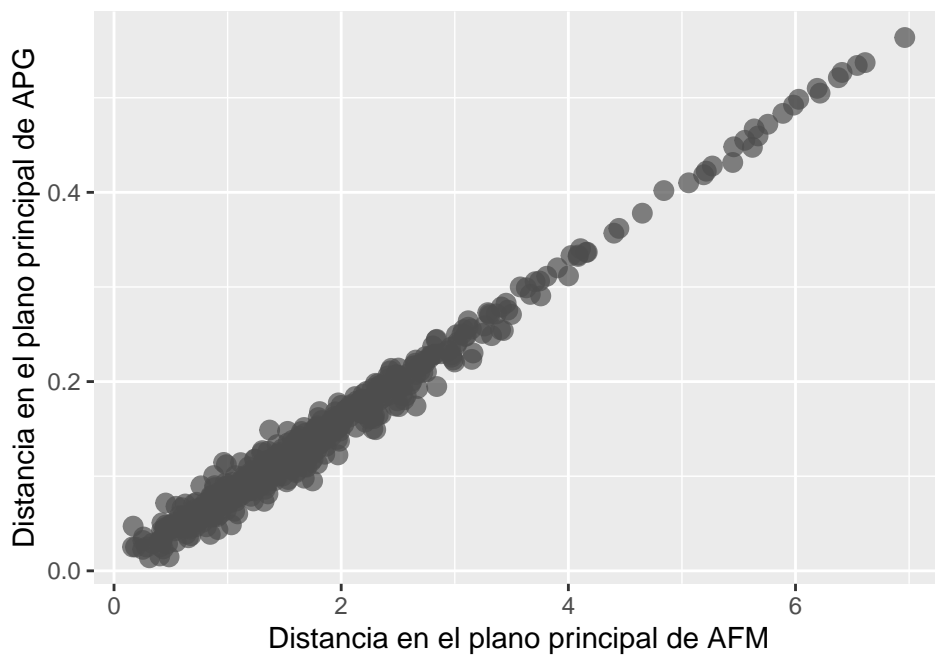


Figure 5: Comparacion de distancias en el plano principal para AFM y APG.

El grafico nos permite ver la alta concordancia que existe entre ambas representaciones en el plano principal. La correlacion entre las distancias en los planos principales es de 0.992, lo que indica, consistentemente con el grafico, la muy alta similaridad entre ambas representaciones.