

# 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 homotéticas<sup>1</sup>. 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.

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.

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

El valor de los coeficiente  $N_g$ , es decir, la raiz 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.

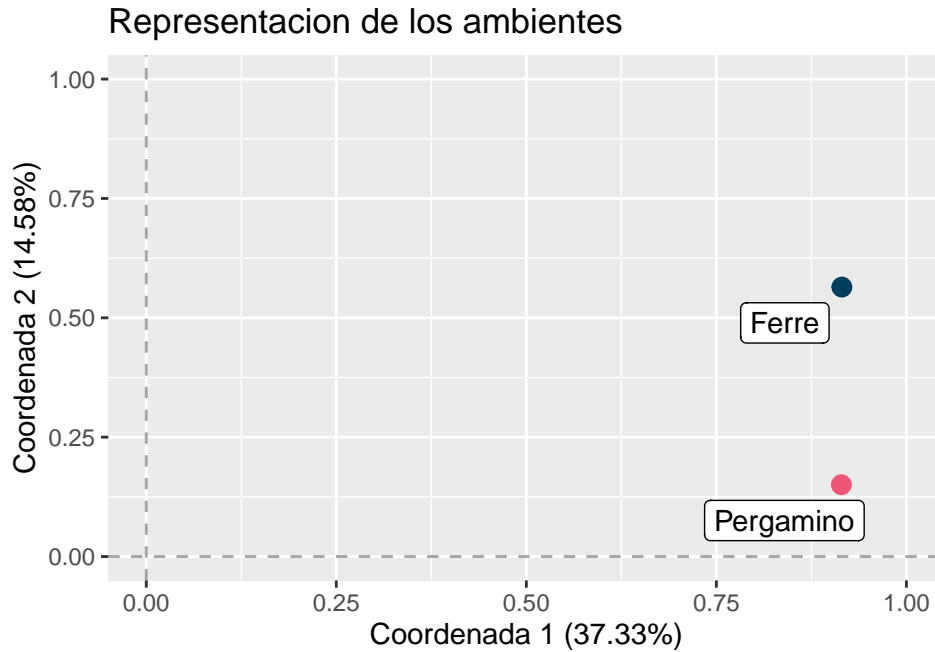


Figure 1: Asociacion de los ambientes con los ejes del plano principal AFM.

En el grafico de los ambientes se puede ver que ambos ambientes tienen relacion muy alta con la primer 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. Alli 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. para estas variedades la caracterizacion promedio que se obtiene mediante AFM no es representativa de su comportamiento

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

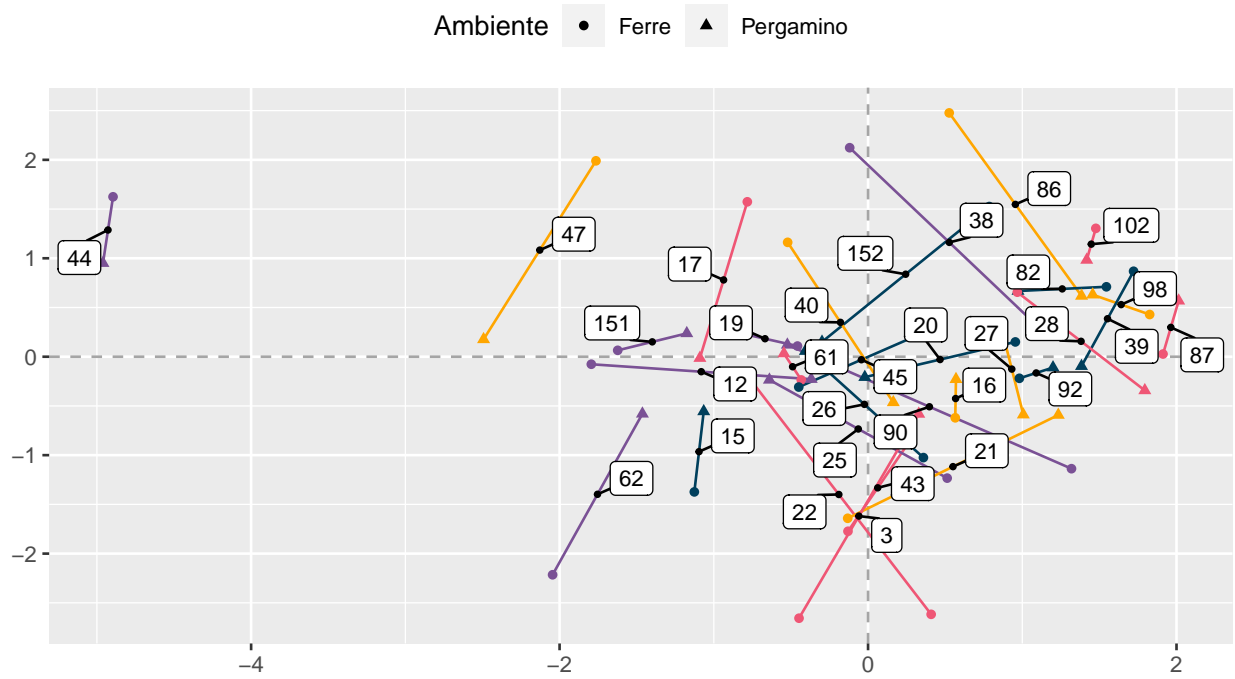


Figure 2: Caracterizaciones parciales y de consenso mediante AFM.

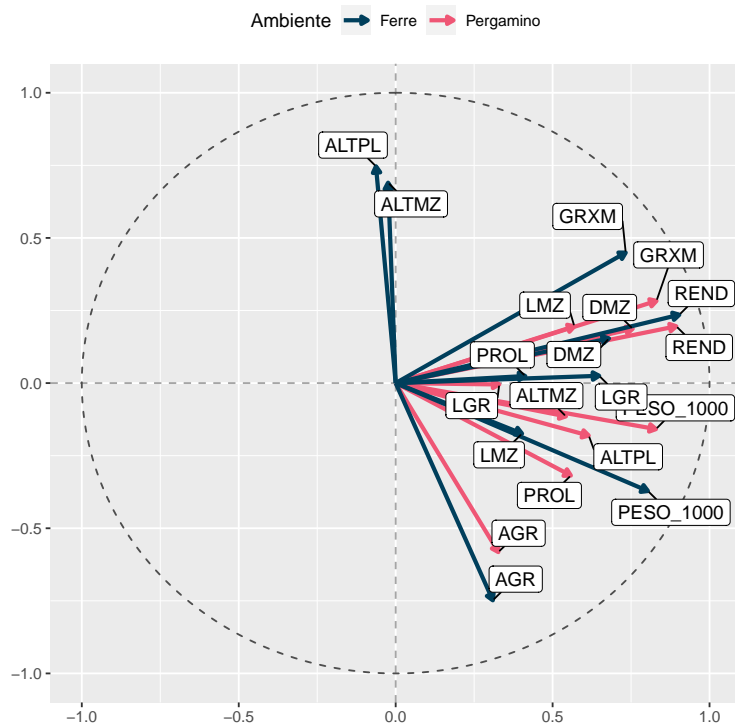


Figure 3: Correlacion de las variables con los ejes del plano principal AFM.

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 dimensión).

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 limpio, 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 maiz. Esta variedad resalta en ambos ambientes por su bajísimo rinde y otras características no deseables para una variedad de maiz como pueden ser los granos pequeños y livianos.

El grupo 2 presenta variedades de maiz cuyo rendimiento es mas bajo que el promedio. Estas variedades ademas presentan granos livianos y mediciones morfologicas mas tirando a bajas que medias. Excepcionalmente, alguna de las variedades de este grupo, como puede ser la 47 y 17, estan 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). Tambien 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 estan asociadas a granos anchos, por ejemplo.

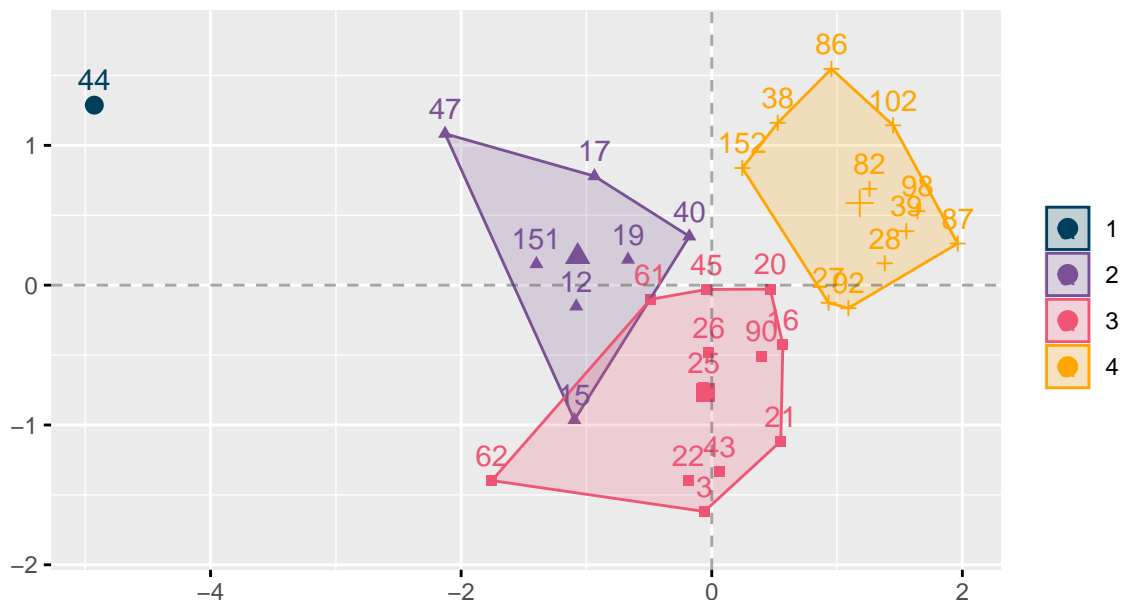


Figure 4: Agrupamiento basado en HCPC.

- 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.

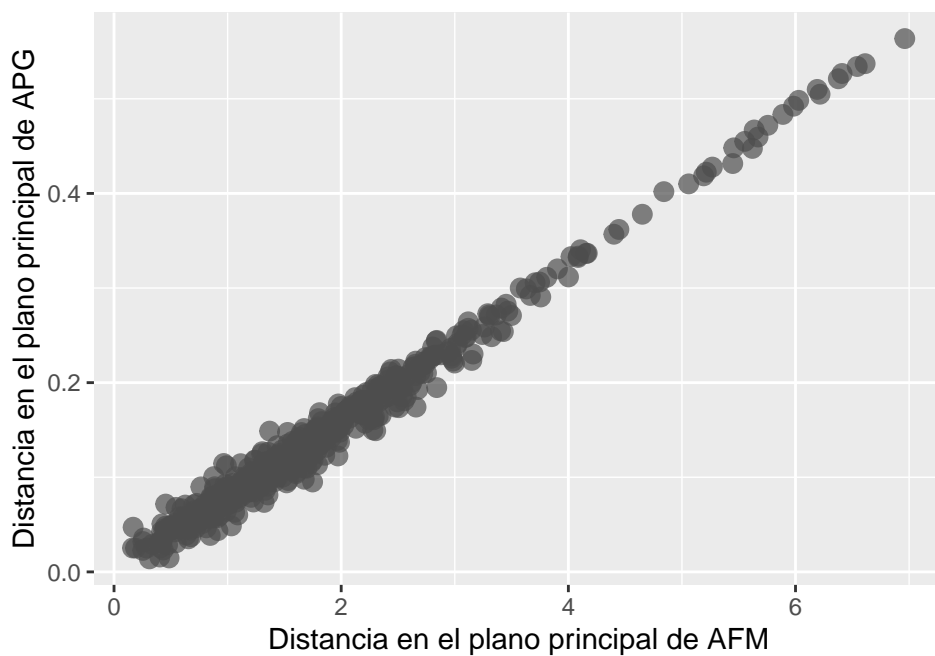


Figure 5: Comparacion de distancias en el plano principal para AFM y APG.

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 similitud entre ambas representaciones.

## Ejercicio 2

Los datos del archivo QUINOA corresponden a 24 accesiones de quinoa nativa del Noroeste Argentino conservadas en el Banco de Germoplasma, caracterizadas a través de 10 variables cuantitativas y 8 variables cualitativas. En el identificador de cada accesión se encuentra la indicación de la procedencia del mismo: AL (altiplano), VA (Valles de altura), VS (Valles secos), VH (Valles húmedos orientales).

A) Realice un ACP con las variables cuantitativas.

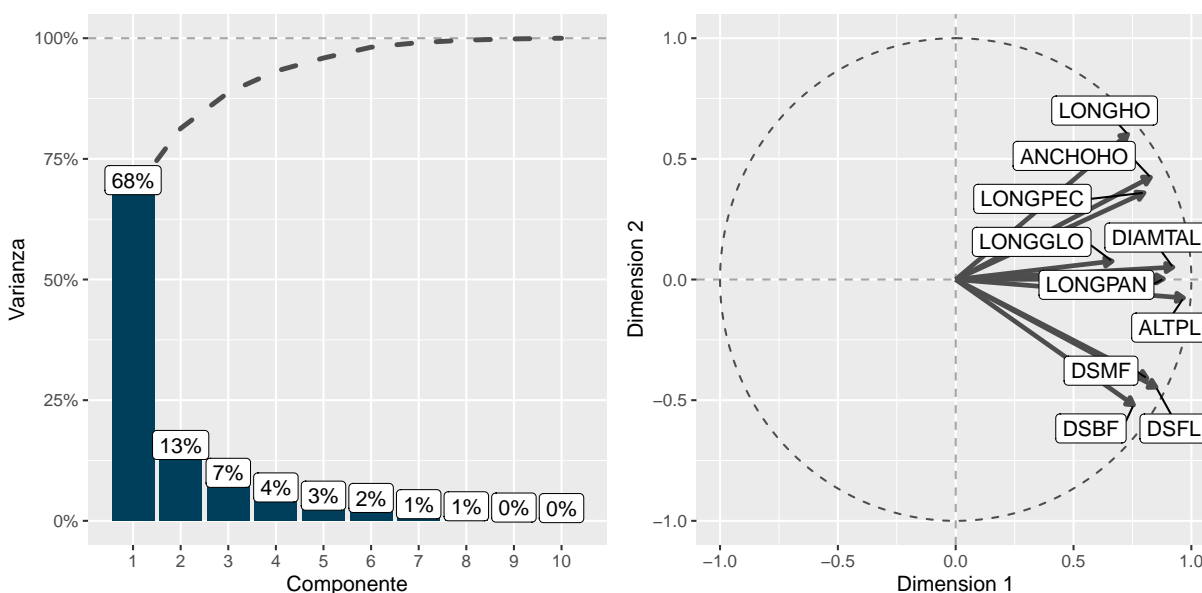


Figure 1: Izquierda: porcentaje de varianza retenido por cada componente. Derecha: Correlacion de las variables con los componentes del plano principal.

Luego de realizar el analisis de componentes principales, podemos observar que el porcentaje de variabilidad total explicada por las dos primeras componentes es de 81.31%. Esto indica que la configuracion en el plano principal conserva en muy buena manera la configuracion de los datos en el espacio original.

En relacion a las variables que mas aportan a la formacion de las componentes, podemos decir que todas las variables contribuyen de forma positiva a la primer componente, y que longitud del glomérulo es quien lo hace en menor medida. Con respecto a la segunda componente, no se cuenta con variables con correlaciones muy altas, pero si se puede notar que hay dos grupos que se asocian a este componente de manera moderada. Por un lado, longitud y ancho de hoja junto a longitud del peciolo (de forma positiva) y días desde siembra hasta la aparicion del botón floral, días hasta la floración, y días hasta la madurez fisiológica (de forma negativa).

El panel derecho de la **Figura 1** tambien nos permite identificar grupos de variables segun su correlacion. Podemos identificar a los siguientes tres grupos que presentan correlacion muy alta entre las variables que los componen:

- Grupo 1: Longitud del glomérulo, diámetro del tallo, longitud de la panoja y altura de planta (medidas morfológicas de la planta y su fruto).

- Grupo 2: Longitud de la hoja, ancho de la hoja y longitud del peciolo (medidas morfológicas asociadas a la hoja).
- Grupo 3: Días desde siembra a botón floral, días desde siembra a floración y días desde siembra a madurez fisiológica (medidas temporales).

Se puede notar que el Grupo 1 presenta una correlación moderada con los grupos 2 y 3, mientras que estos dos últimos están prácticamente no correlacionados entre sí ya que forman un ángulo aproximado de 90 grados.

A partir de la **Figura 2** podemos hacer una caracterización de accesiones de quinoa según la procedencia.

En primer lugar, se puede notar que las accesiones del Altiplano poseen valores bajos en la primera componente, y en torno al promedio o más bien altos en la segunda. Esto nos permite decir que estas accesiones presentan valores pequeños en cuanto a las características morfológicas de la hoja y de la planta, pero que sus variables temporales también son menor que el promedio. Es decir, la aparición de botón floral, la floración y la madurez fisiológica se dan más tempranamente que en otras regiones.

Por otro lado, se tiene que las accesiones de Valles Húmedo y Seco presentan similares valores en cuanto a la primera dimensión, pero difieren en cuanto a la segunda. Las accesiones del Valle Húmedo suelen presentar valores menor al promedio y mientras que las del Valle Seco son mayores. Así, podemos decir que en ambas procedencias se dan similares características en cuanto a las medidas morfológicas de la planta (de gran tamaño en ambos casos), pero que las accesiones del Valle Seco suelen asociarse a mayor tamaño de hoja y peciolo, mientras que las del Valle Húmedo se asocian a mayor tiempo hasta la aparición del botón floral, floración y madurez fisiológica.

Con respecto a Valles de Altura, podemos decir que presentan un comportamiento cercano al promedio en términos de la primera componente y están en torno al promedio, o más bien bajo el promedio, en cuanto a la segunda. En otras palabras, se asocian a accesiones de tamaño de planta promedio, tamaño de hoja tirando a bajo, y tiempos de floración y madurez tardíos.

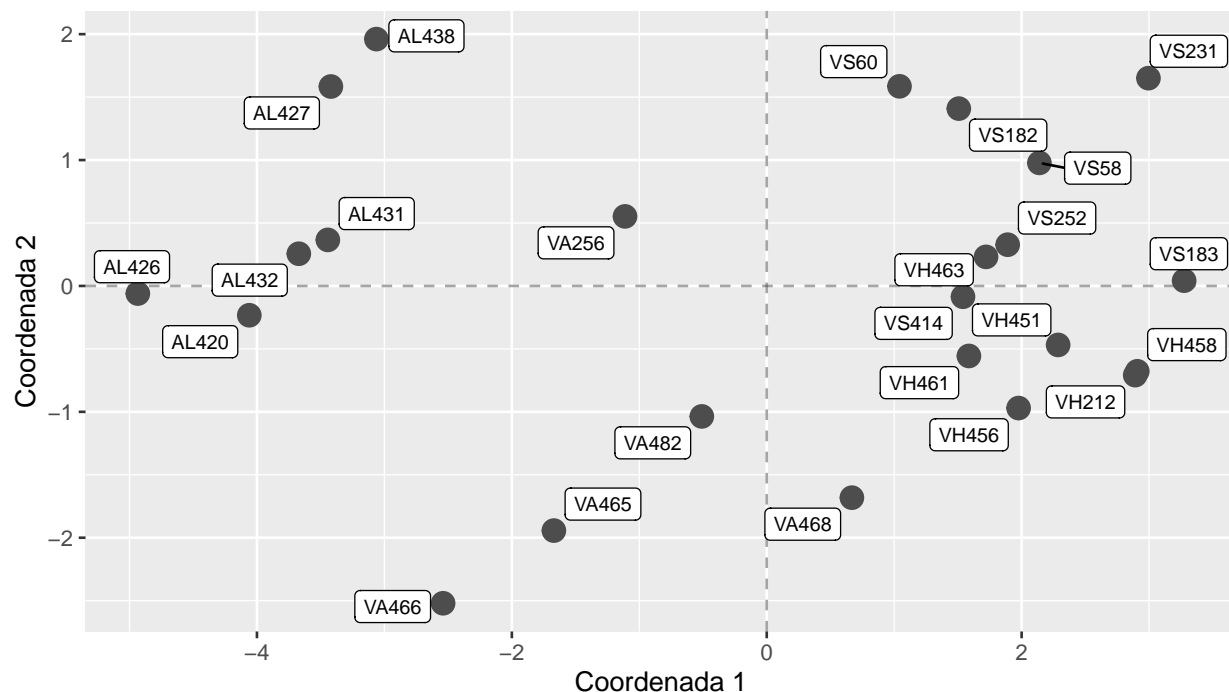
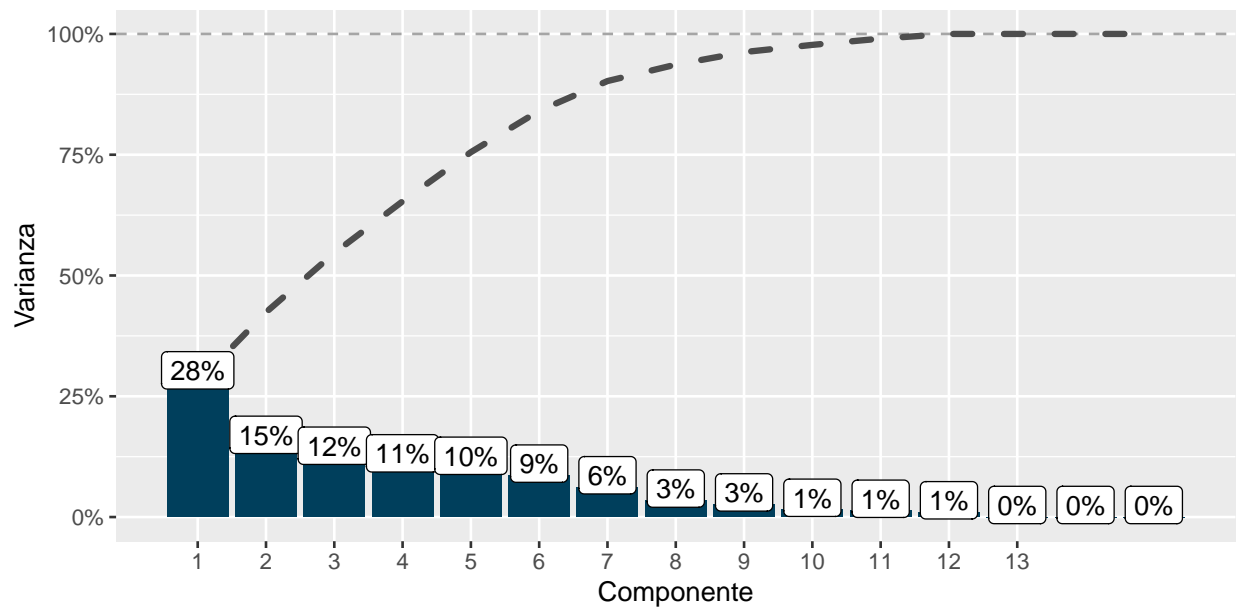


Figure 2: Accesiones de Quinoa en el plano principal del ACP.

- B) Realice un ACM con las variables cualitativas y compare las configuraciones de individuos provistas por ambas técnicas en el plano principal.





El porcentaje de inercia explicado es por las primeras dos componentes es de 42.39%, bastante menor que lo obtenido con ACP. Esto indica que la representacion que obtenemos en el plano principal del analisis obtenido con ACM no conserva la configuracion original de los datos de la misma manera que lo hace el analisis ACP con las variables cuantitativas.

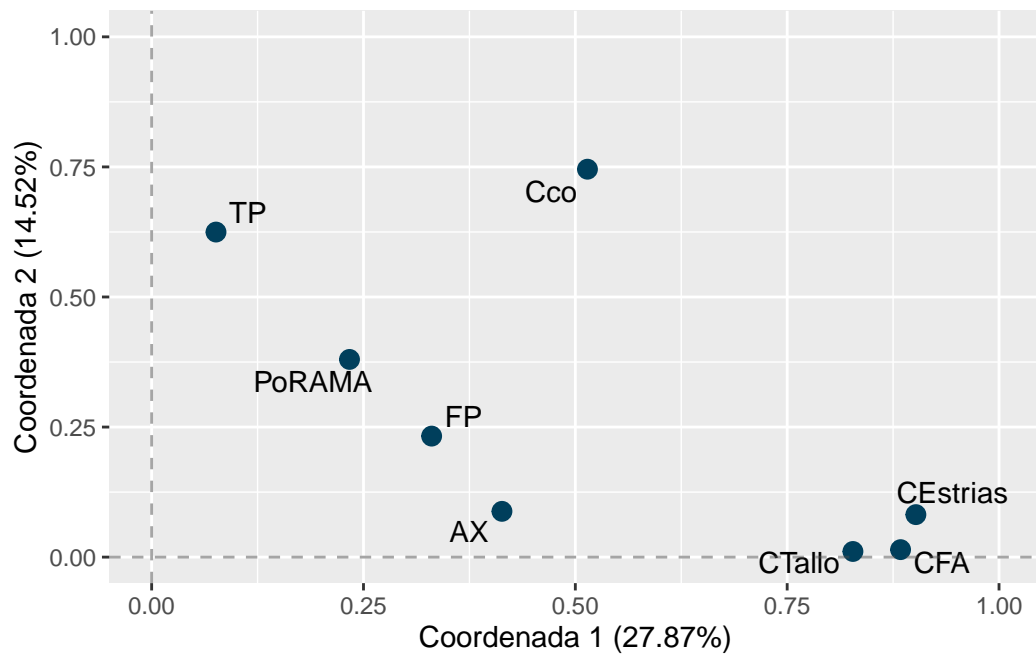


Figure 3: Contribucion de las variables a los dos primeros ejes.

Las variables que mas aportan a la formacion de la Dimension 1 son Color de las estrias, Color de panoja a fin de antesis y Color de tallo. Esta contribucion se da de manera practicamente exclusiva, ya que no se observa contribucion en la segunda dimension. Por otro lado, tipo de panoja y el color de la panoja a la cosecha son

las que mas contribuyen a la Dimension 2. Esta ultima variable tambien contribuye moderadamente a la primera dimension.

Finalmente notamos que la presencia de ramas, la forma de la panoja, y la presencia de axilas tienen muy poca contribucion a la formacion de cualquiera de los dos ejes.

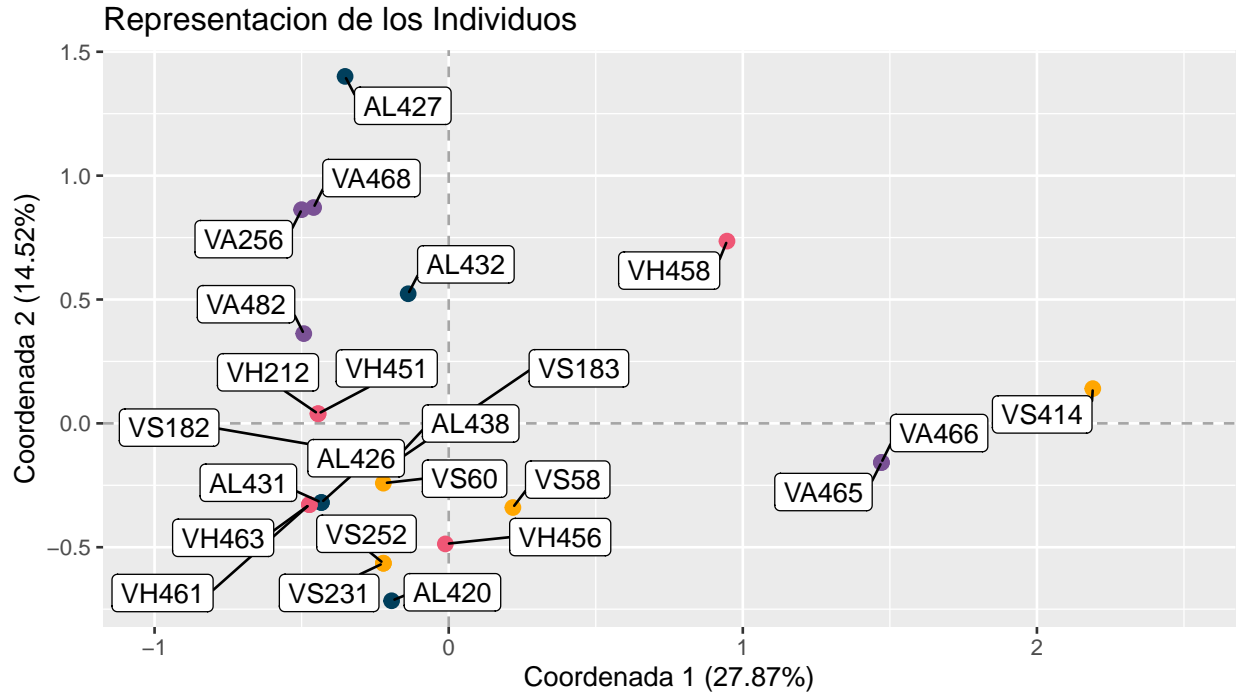


Figure 4: Representacion de individuos en el plano principal de ACM

Se nota en la **Figura 4** que en el analisis ACM basado en las variables cualitativas no permite formar agrupamientos en terminos del ambiente en el plano principal, a diferencia de lo que sucedio con el analisis ACP basado en variables cuantitativas.

Luego, si comparamos las configuraciones de individuos provistas por las dos tecnicas realizadas, podemos concluir que las mismas parecen no ser homogeneas, ya que, como se pudo observar, una tecnica logra una diferenciacion por ambientes mientras que la otra no.

En la **Figura 5** podemos ver que las representaciones sobre el plano principal no son para nada concordantes. La correlacion entre las distancias en los planos principales es de -0.006, lo que indica, consistentemente con el grafico, la discrepancia entre ambas representaciones.

Realice un AFM considerando como grupos a ambos tipos de variables.

C) Realice el análisis de la interestructura. ¿Qué puede decir de la relación existe entre los grupos?

Table 1: Coeficiente  $L_g$

	Cuantitativas	Cualitativas	MFA
Cuantitativas	1.06	0.23	1.11
Cualitativas	0.23	1.92	1.85
MFA	1.11	1.85	2.55

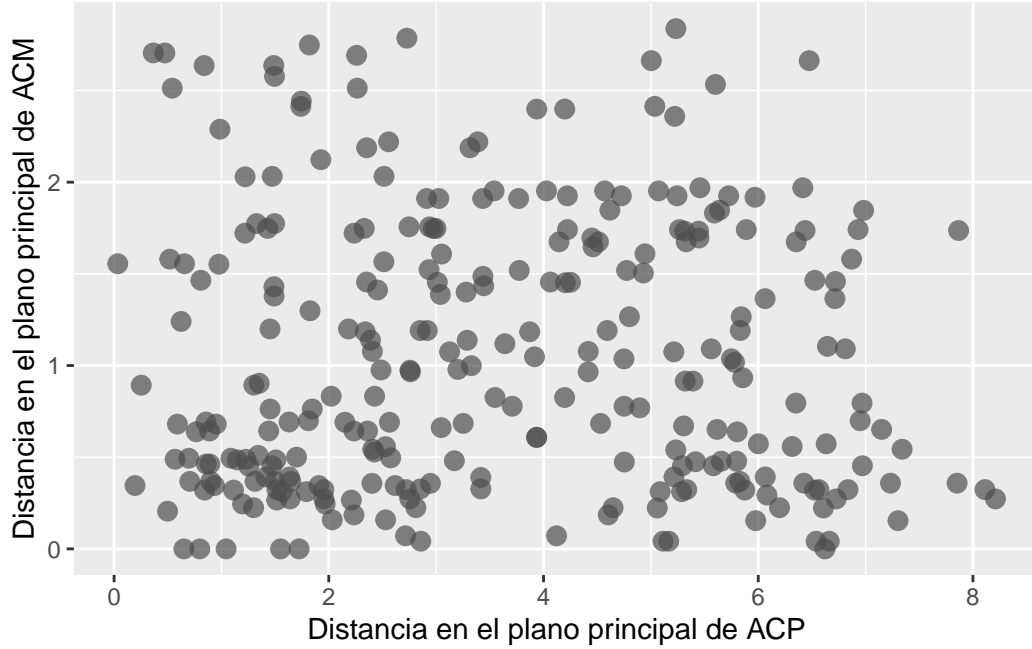


Figure 5: Comparacion de distancias en el plano principal para ACP y ACM.

Al obtener el Coeficiente  $L_g$ , el cual nos brinda una medida de estructura comun entre las variables Cautitativas y Cuantitativas, observamos en la **Tabla 1** que el mismo es de 0.23. Dado que se obtuvo un valor bajo, podemos decir que las variables del grupo “Cuantitativas” no estan correlacionadas con las del grupo “Cualitativas”. Es decir, las configuraciones no tienen estructura en comun.

Table 2: Coeficiente  $N_g$

	Cuantitativas	Cualitativas	MFA
Cuantitativas	1.03	0.48	1.05
Cualitativas	0.48	1.39	1.36
MFA	1.05	1.36	1.60

Al obtener una medida de dimensionalidad de cada grupo (coeficiente  $N_g$ ), observamos que el grupo “Cualitativas” tiene una dimensionalidad mayor que “Cuantitativas”

Table 3: Coeficiente RV

	Cuantitativas	Cualitativas	MFA
Cuantitativas	1.00	0.16	0.68
Cualitativas	0.16	1.00	0.84
MFA	0.68	0.84	1.00

Tal como se intuyo al realizar el analisis por separado, las configuraciones cualitativas y cuantitativas de las quinoas no son “homoteticas” entre si. Luego, al ser un valor bajo indica que los grupos de variables brindan informacion complementaria.

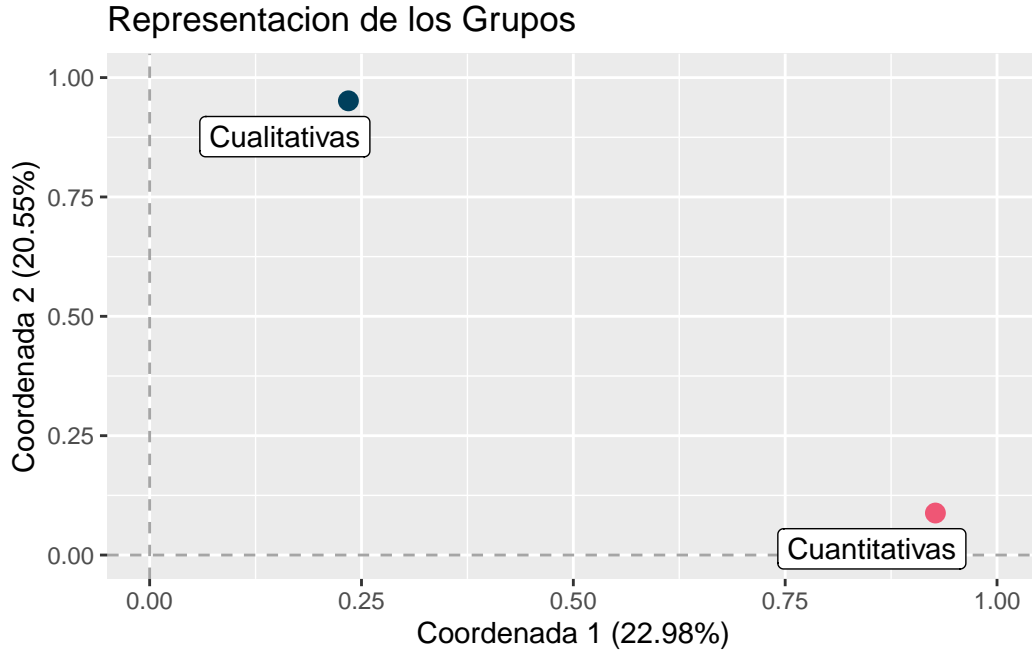


Figure 6: Caracterizacion de los grupos de variables en el plano principal AFM.

Si observamos la **Figura 6**, podemos ver que las variables Cuantitativas contribuyen unicamente a la formacion de la Coordenada 1, mientras que las Cualitativas son las que contribuyen a la formacion de la Coordenada 2 del Analisis Factorial Multiple. Estas ultimas, presentan una leve contribucion en la primer coordenada, que se corresponde con el coeficiente  $N_g$  un poco mayor a 1 como es el 1.39 antes visto.

En la **Tabla 4** vemos que as variables Cualitativas que más aportan en la formación del eje 2 son Color de Estrías y de Panoja a fin de antesis Púrpura

En la **Tabla 5** vemos que todas las variables cuantitativas contribuyen a la construccion del primer eje. El grupo de variables que mas contribuye es el de medidas morfologicas de la planta (Altura de planta, diámetro del tallo), longitud de la panoja y las características temporales (Días de siembra a boton floral y dias desde siembra a floracion). Las medidas morfologicas de la hoja contribuyen de menor manera. Esto nos permite decir que individuos que presenten valores altos en la primer componente van a ser de plantas y hojas grandes, pero que demoran mas en presentar el boton floral, en florecer y madurar fisiologicamente.

En cambio, solo las variables morfologicas de la hoja contribuyen a la segunda componente (LONGHO, ANCHOHO y LONGPEC).

Table 4: Contribuciones a los Ejes - Cualitativas

	Dim.1	Dim.2
AxNo	0.07	3.73
AxSi	0.12	6.22
CEstAm	0.14	0.50
CEstNo	1.26	2.64
CEstPu	0.74	16.54
CEstRo	2.95	0.85
CTaPu	0.38	10.47
CTaRo	1.44	6.63
CTaVe	0.01	2.43
PoRaBa	0.65	0.64
PoRaNo	0.39	1.56
PoRaOb	0.00	3.61
CFabl	0.15	3.31
CFapu	0.74	16.54
CcoAm	0.00	0.28
CcoBl	0.26	0.54
CcoGr	2.73	0.05
CcoMar	1.44	6.63
CcoPu	1.95	0.97
TPDif	0.86	0.14
TPNoDif	3.27	0.52
FPAma	0.42	4.51
FPGlom	0.21	2.26

Table 5: Contribuciones a los Ejes - Cuantitativas

	Dim.1	Dim.2
DIAMTAL	9.41	0.77
ALTPL	10.98	0.59
LONGHO	5.48	1.79
ANCHOHO	7.47	2.85
LONGPEC	6.04	1.61
LONGPAN	8.59	0.32
LONGGLO	6.09	0.02
DSBF	8.47	0.43
DSFL	9.97	0.04
DSMF	7.31	0.04

D) ¿Existe un agrupamiento de individuos por zona?

Al observar la **Figura 7** podemos referenciar algunos agrupamientos según la zona de procedencia. El primer grupo que podemos mencionar es el conformado por las variedades VA466 y VA465, ambas procedentes de Valles de Altura; el resto de las variedades de Valles de Altura pueden asociarse en un segundo grupo. El tercer grupo que puede identificarse está conformado por las variedades procedentes de Altiplano; y por último, el cuarto agrupamiento se compone por las que son procedentes de Valles Húmedos y Secos, no pudiendo distinguir entre estas dos zonas.

Cabe descartar que hay dos accesiones, VS414 y VH458, que se alejan del cuarto agrupamiento; probablemente se deba a que presenten valores más altos en la configuración Cualitativa que el resto de las Quinoas procedentes de su misma zona.

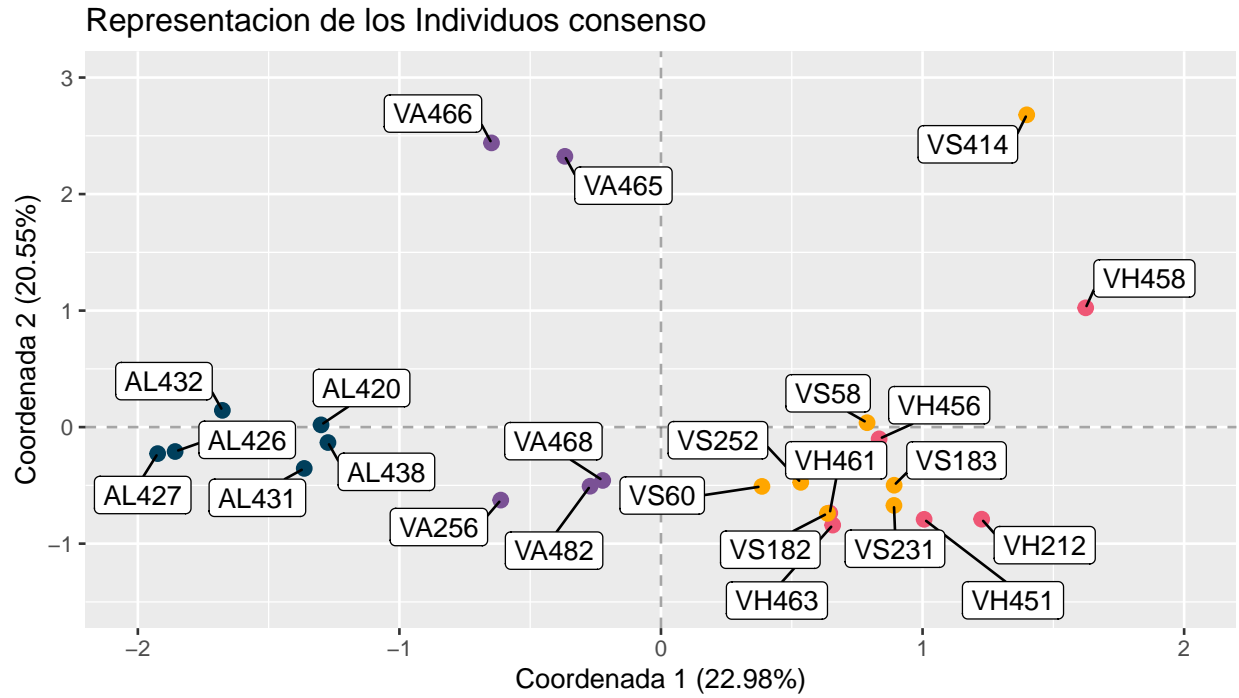


Figure 7: Representacion de individuos consenso en el plano principal de AFM

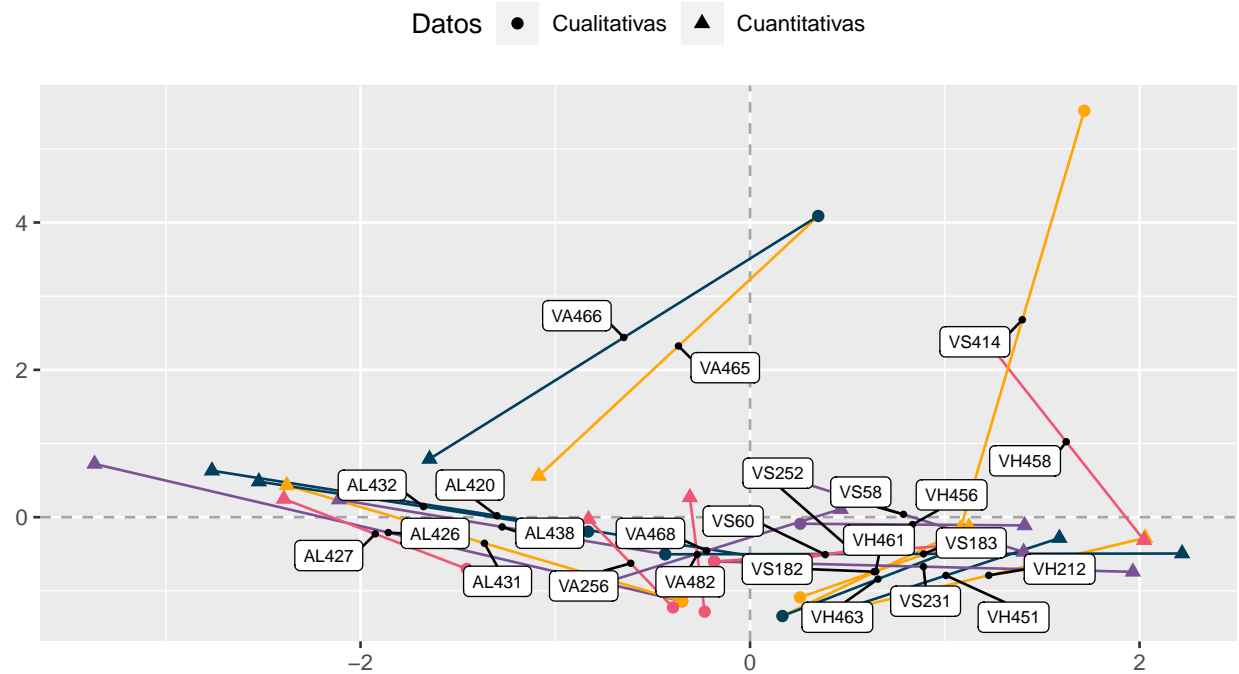


Figure 8: Caracterizaciones parciales y de consenso mediante AFM.

A través de la **Figura 8** se puede confirmar lo comentado anteriormente; los individuos VS414 y VH458 se apartan del grupo conformado por Valles Húmedos y Secos ya que poseen valores mas altos de la configuración

cualitativa.

E) Caracterice los grupos de individuos a través de todas las variables.

Las accesiones VA466 y VA465 que conforman el primer grupo, se caracterizan por tener el color de las estrias y de la panoja a fin de anthesis de color púrpura. Además, estas accesiones tienen un comportamiento en torno al promedio con respecto a las variables cuantitativas, es decir, son de plantas y hojas moderadas, y el tiempo que demoran en presentar el botón floral, en florecer y madurar fisiológicamente también es moderado.

El segundo grupo, conformado por las restantes accesiones procedentes de Valles de Altura, comparten las mismas características en cuanto a la morfología y medidas temporales, pero se diferencian de las procedentes de su misma zona por tener características cualitativas distintas, principalmente por no poseer estrías y panoja a fin de anthesis de color púrpura.

El tercer grupo, ubicado en la esquina inferior izquierda se caracteriza por poseer plantas y hojas pequeñas, pero de rápida floración y maduración fisiológica. Al igual que el grupo 2, no poseen estrías y panoja a fin de anthesis de color púrpura.

El cuarto grupo, Valles Húmedos y Secos, comparten las mismas características cualitativas que el segundo y tercer grupo, pero en cuanto a sus características morfológicas y temporales podemos decir que este grupo presenta valores más bien altos, es decir tienden a ser plantas grandes, y la floración y maduración fisiológica tiende a ser lenta.

## Ejercicio 3

Los datos del archivo banana16 corresponden a 40 poblaciones de banana (*Musa spp.*) de las cuales 20 son clones de origen genético desconocido (recolectados en campos de productores) y 20 son variedades de uso internacional Jaffa, Gal, Gran Enanao y Williams, respectivamente (testigos). Los datos son un extracto de la base original perteneciente al Centro Regional Chaco – Formosa de INTA.

A) Analice las correlaciones entre variables en ambos grupos.

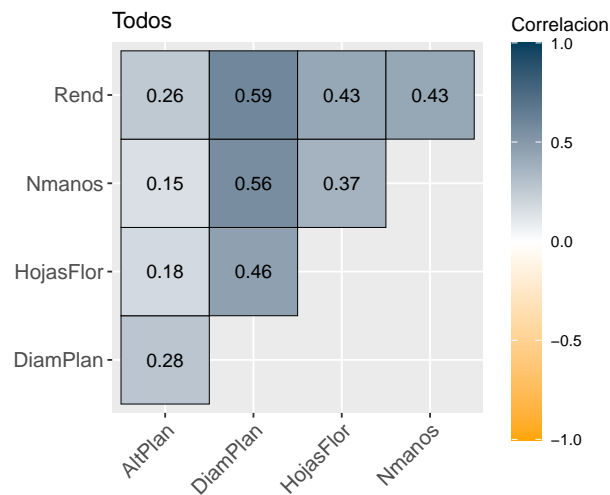


Figure 1: Correlacion entre las variables para el conjunto de datos completo

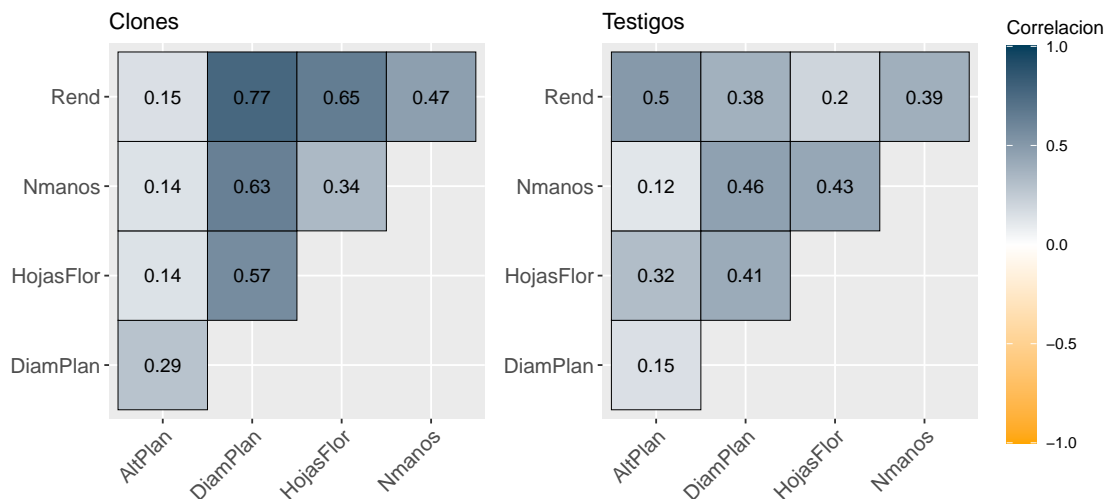


Figure 2: Correlacion entre las variables segun origen genetico.



Al obtener las matrices de correlaciones entre las variables podemos ver que las mismas son todas positivas, tanto en el grupo “Clones” como en el grupo “Testigos”. Cabe destacar que es notable la diferencia en las magnitudes de las correlaciones en los grupos.

Si observamos la matriz correspondiente al grupo “Clones” podemos ver que la variable Rendimiento presenta una correlacion alta con Diametro de Planta y, en menor medida con Hojas a Floracion. Ademas se observan correlaciones, de moderadas a altas, para las variables Diamtro de Planta con Cantidad de Manos y Hojas a Floracion.

Por otro lado, observando la matriz del grupo “Testigo” vemos que la variable Rendimiento, a diferencia de lo observado en la matriz anterior, esta mas correlacionada con Altura de planta.

Otra diferencia a destacar, es que la variable Diamtro de planta presenta correlaciones mas altas en el grupo “Clones” que en el grupo “Testigos”.

Realice un Análisis Factorial Múltiple Dual y responda las siguientes cuestiones

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

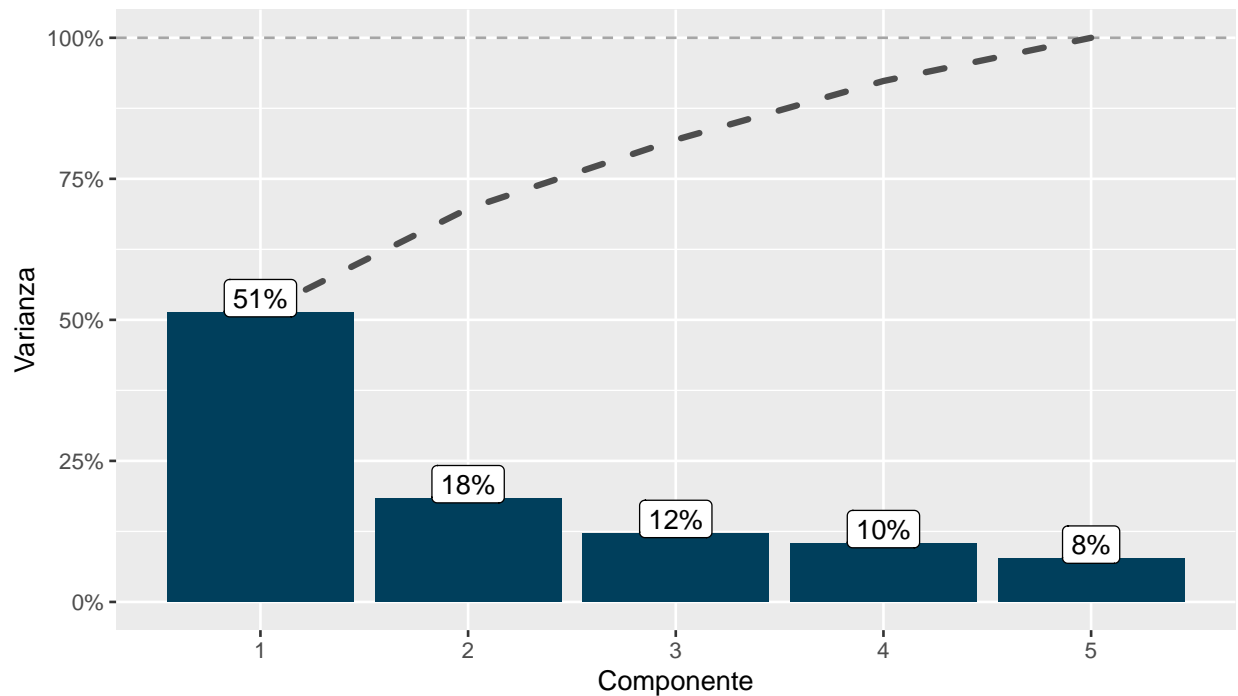


Figure 3: Variabilidad explicada por componente.

El porcentaje de variabilidad explicada sobre el plano princial es del 69.72%. Como puede observarse en la **Figura 3**, el primer autovalor explica un 51.35% mientras que el segundo un 18.37%

C) ¿Qué puede decir a partir del gráfico de las condiciones?

Al representar los grupos sobre el plano principal podemos observar que ambos tienen una alta relacion con la primera dimension, aportando en igual medida a la misma; pero si se proyectan los puntos sobre la segunda dimension, puede observarse que el grupo “Testigo” tiene un aporte levemente mayor en la formacion de esta.

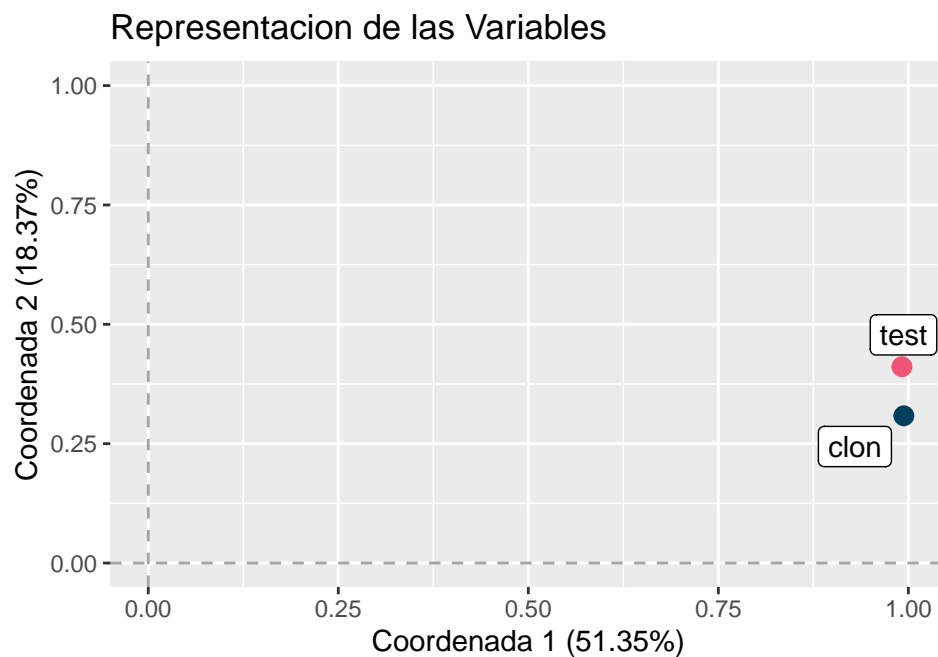


Figure 4: Caracterizacion de los grupos en el plano principal AFM.

D) ¿Cómo se ven representadas en los gráficos del DMFA las conclusiones del inciso a)?

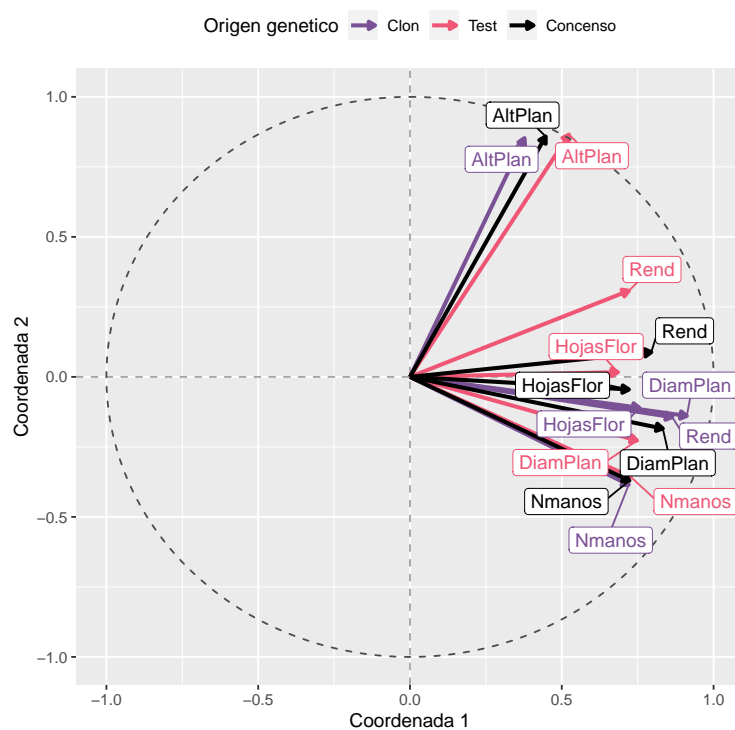


Figure 5: Variables conseso vs Variables parciales

En la **Figura 5** se representa el comportamiento de las variables conseso y parciales correspondientes a cada

grupo. Esta representación nos va a permitir observar los principales cambios que introdujo el agrupamiento de individuos.

Al considerar el grupo “Clones”, vemos que, luego del agrupamiento de individuos, Altura de planta y Cantidad de manos se mantienen prácticamente igual; pero observando la variable Rendimiento vemos que la misma tiene ahora una coordenada negativa sobre el Eje 2. Como vimos en la matriz de correlación correspondiente a este grupo (Figura 2), la variable Rendimiento presenta mayores correlaciones con Diámetro de Planta, Cantidad de manos y Hojas a Floración que con la variable Altura de Planta. Esto explica la transición de la variable Rendimiento sobre el eje 2.

Luego, si consideramos el grupo “Testigos”, podemos concluir que, a diferencia de lo observado en el grupo “Clones”, luego del agrupamiento de individuos la variable Rendimiento tiene una coordenada mayor sobre el eje 2. Este cambio se debe a que, como vimos en la matriz de correlacion correspondiente a este grupo (Figura 2), la variable Rendimiento presenta una correlacion positiva alta con la variable Altura de Planta.

## Ejercicio 4

Los datos del archivo hongos.csv se refieren a 8 sitios (primera vía) que corresponden a chacras identificadas por zona (centro/este del Alto Valle de Río Negro) y por sistema de cultivo (orgánico/tradicional), caracterizadas a través del recuento de hongos por género (en unidades formadoras de colonias por metro de aire) (segunda vía) y a lo largo de las cuatro estaciones (tercer vía).

A) Interprete los coeficientes  $L_g$ ,  $R_v$  y  $N_g$ . Analice el gráfico de las condiciones y el de ejes parciales.

Table 1: Coeficiente  $L_g$

	invierno	verano	primavera	otoño	MFA
invierno	1.654	0.770	0.683	0.850	1.265
verano	0.770	1.137	0.652	0.483	0.972
primavera	0.683	0.652	1.038	0.792	1.011
otoño	0.850	0.483	0.792	1.345	1.109
MFA	1.265	0.972	1.011	1.109	1.392

El coeficiente  $L_g$  demuestra que otoño e invierno, comparten mucha inercia. Por el contrario entre otoño y verano se observa el menor valor, por lo cual habrían de compartir menor estructura.

Table 2: Coeficiente  $N_g$

invierno	1.29
verano	1.07
primavera	1.02
otoño	1.16
MFA	1.18

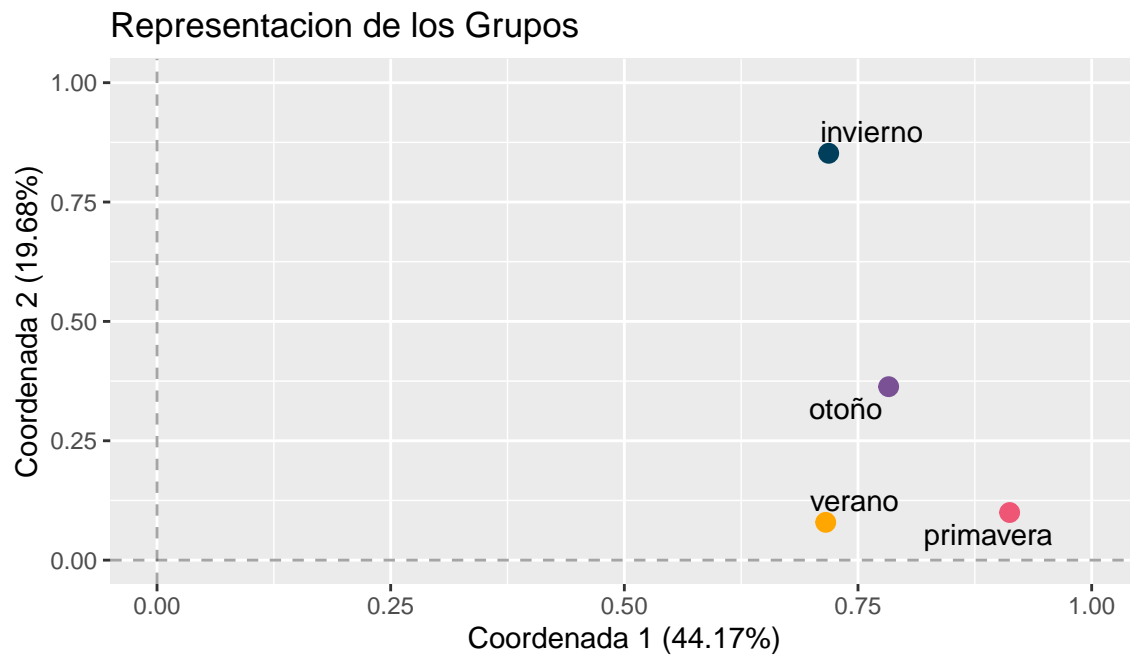
El coeficiente  $N_g$  demuestra que a los grupos de verano y primavera se los puede pensar como unidimensional, mientras que invierno y otoño requieren una mayor dimensión en su representación.

Table 3: Coeficiente  $R_v$

	invierno	verano	primavera	otoño	MFA
invierno	1.000	0.561	0.521	0.57	0.833
verano	0.561	1.000	0.600	0.39	0.772
primavera	0.521	0.600	1.000	0.67	0.841
otoño	0.570	0.390	0.670	1.00	0.810
MFA	0.833	0.772	0.841	0.81	1.000

Al analizar la tabla  $R_v$ , la cual estandariza los coeficientes, se observa que ningún grupo es tan similar al otro. Aunque existe un grado de similaridad no despreciable, esta no es tan fuerte entre las variables. La relación entre verano y otoño se destaca ya que es la de menor similaridad, al igual que la de otoño primavera por ser la más intensa.

En el gráfico de las condiciones se observa aquello que se intuía en las tablas anteriores: primavera y verano tienen un componente casi exclusivo de la dimensión 1, mientras que invierno necesita de ambas dimensiones. Otoño posee un comportamiento un poco más intermedio, aunque cercano a primavera-verano.



En el gráfico de ejes parciales se observa que otoño y primavera comparten su representación, y en cambio invierno y verano tienden a tener un comportamiento propio.

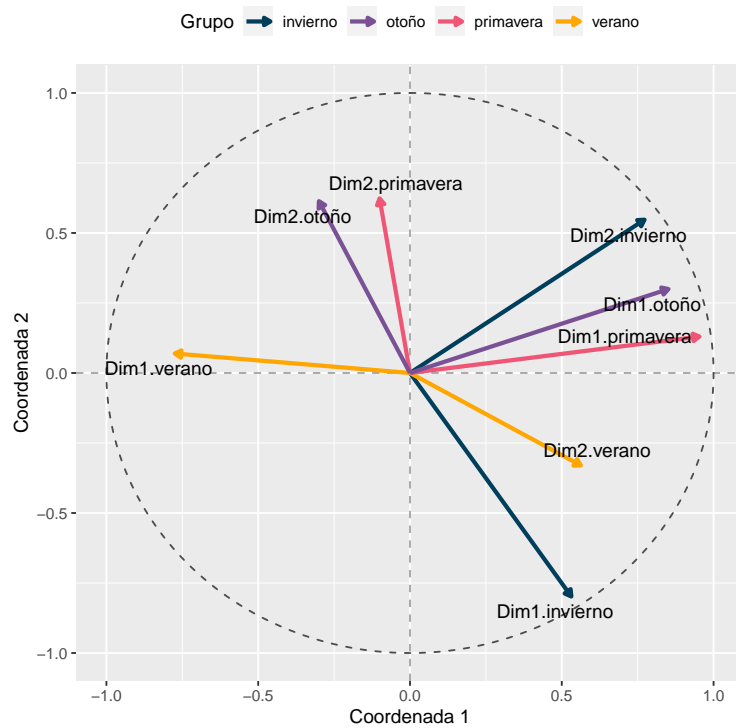


Figure 1: Ejes Parciales

B) ¿Hay un agrupamiento de los sitios por zona o sistema de cultivo?

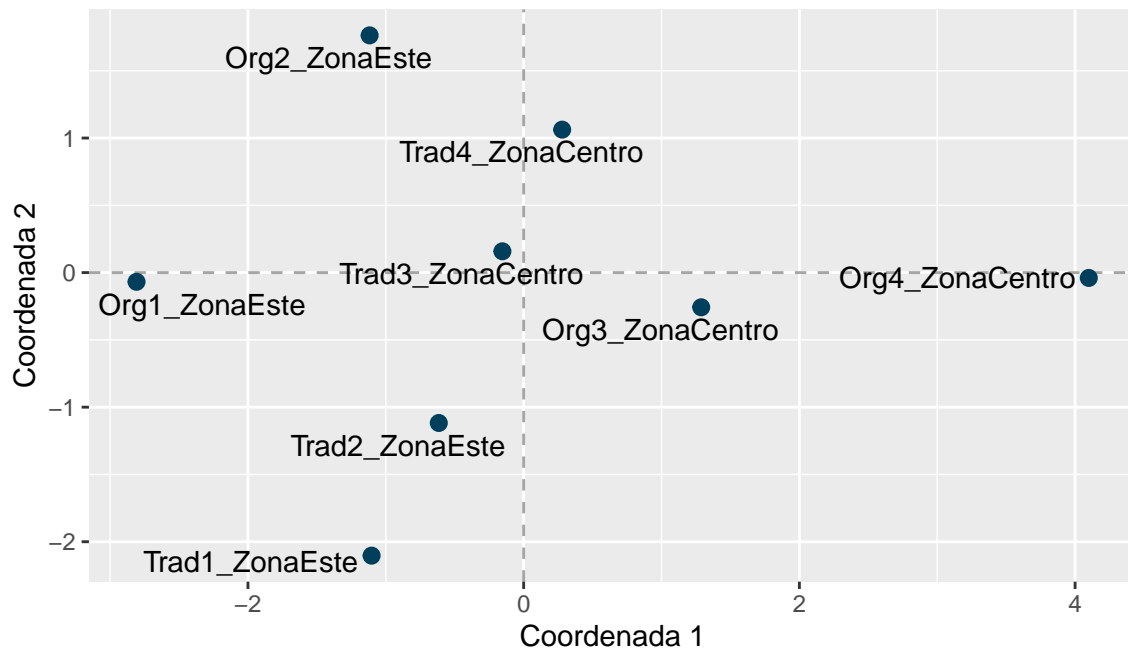


Figure 2: Configuración de Concenso

Al analizar a los individuos en función de su configuración de concenso se observa que la primera dimensión permitiría separar aquellas zonas centro de las zonas este. Un análisis más atento denota que el primer eje ordena de izquierda a derecha el sistema de cultivo orgánico de 1 a 4 y el segundo componente ordena de abajo a arriba el sistema de cultivo tradicional de 1 a 4. Respecto a los agrupamientos, no hay una tendencia clara en las configuraciones de concenso.

C) ¿Qué puede decir sobre la similitud de los sitios en las distintas estaciones?

Visualmente, se observa que existe mucha variabilidad de las estaciones alrededor de la “configuración de concenso” o coordenadas globales. En otoño y primavera, tiende a existir más similitud entre los sitios cercanos al origen de coordenadas (**Figura 3**).

D) Analice los cocientes de inercia por eje (1 y 2) y por individuo.

Table 4: Ratio de Inercia

	x
Dim.1	0.851
Dim.2	0.468

Como puede verse en la tabla del cociente de inercia, los puntos parciales están más concentrados sobre el primer eje que sobre el segundo eje.

### Grafico de individuos: Consenso y Parciales

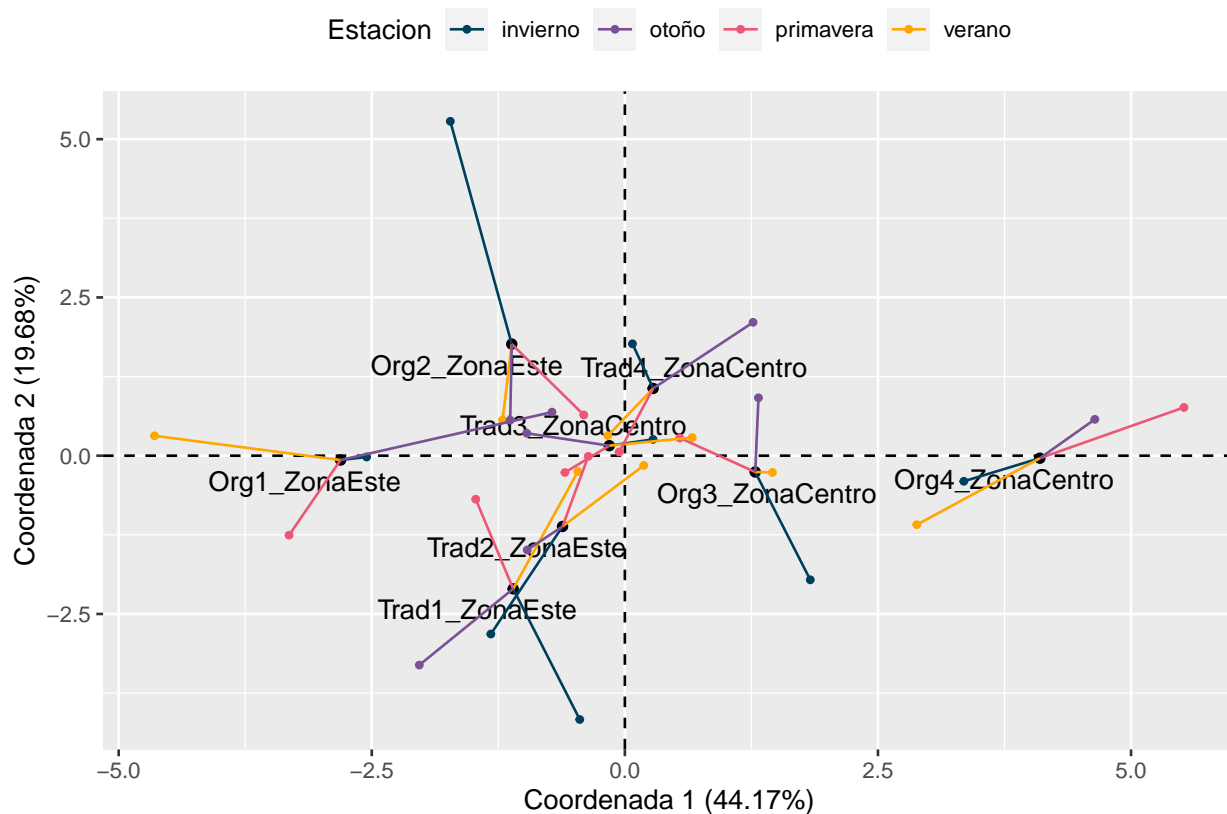


Figure 3: Grafico de individuos: Consenso y Parciales

Table 5: Inercia por Individuo

	Dim.1	Dim.2
Org1_ZonaEste	28.830	2.627
Org2_ZonaEste	7.270	46.843
Org3_ZonaCentro	5.716	10.204
Org4_ZonaCentro	21.860	3.897
Trad1_ZonaEste	11.793	24.661
Trad2_ZonaEste	5.134	6.879
Trad3_ZonaCentro	14.271	0.701
Trad4_ZonaCentro	5.126	4.189

Respecto de la inercia de los individuos, Trad2\_ZonaEste y Trad4\_ZonaCentro poseen una representación equilibrada de consenso. Con respecto a la primer componente, podemos observar que Org1\_ZonaEste, Org4\_ZonaCentro y Trad3\_ZonaCentro poseen mayor variabilidad a lo largo de la misma. Luego, puede comentarse que Org2\_ZonaEste, Org3\_ZonaCentro y Trad1\_ZonaEste poseen mayor variabilidad a lo largo del segundo componente.

E) Analice el comportamiento de los géneros de hongos CL, BO, AL y EP en las distintas estaciones.

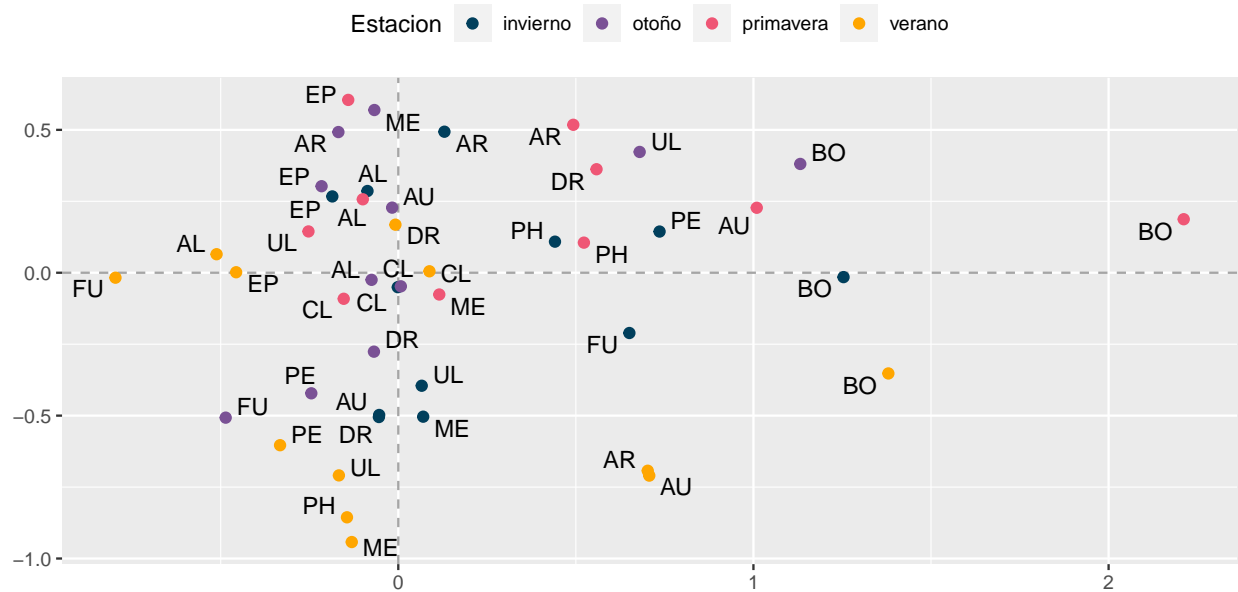
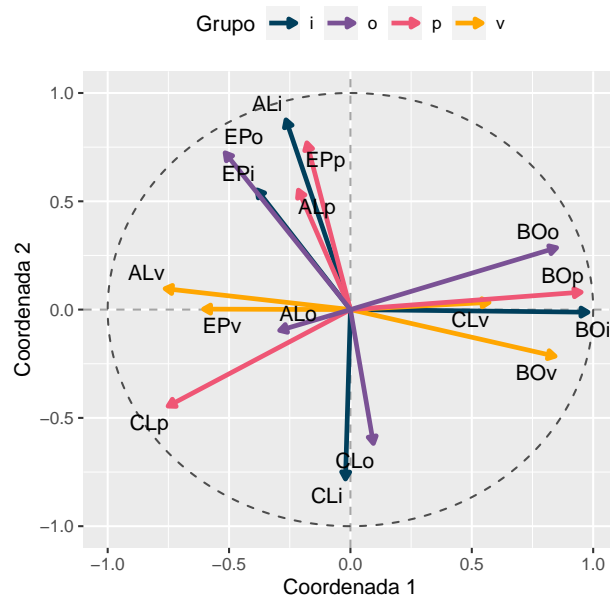


Figure 4: Comportamiento de los Géneros

El género de hongos BO posee notable variabilidad a lo largo de las estaciones, aunque su emplazamiento en las coordenadas es marcado ya que se localiza en el extremo derecho del gráfico, sobre los valores positivos de la dimensión 1. Por otra parte, puede observarse que el género CL posee poca variación a lo largo de las estaciones y, además, se concentra cerca del origen. Con respecto al género AL, puede decirse que el mismo posee considerable variación y se encuentra en el cuadrante superior izquierdo. Finalmente, el género EP posee un comportamiento similar a AL.

Nótese que si se grafican las correlaciones, se obtienen los mismos resultados:



F) Caracterice las principales asociaciones entre las distintas condiciones consideradas (zona, sistema de cultivo, estación y géneros).



En función de los resultados anteriores, se puede establecer que invierno presenta una estructura tanto más rica en términos de frecuencias resultantes que el resto de las estaciones para con los hongos, siendo primavera y verano aquellas mas sencillas de caracterizar. En cuanto a las locaciones, es posible utilizar la primer coordenada para diferenciar la zona centro de la este. A su vez, si se proyectaran los datos de los establecimientos de método tradicional de cultivo, las frecuencias de los establecimientos establecerían un gradiente que coincide con la numeración del mismo. Lo mismo sucede que fuera a proyectarse los datos de los cultivos con métodos orgánicos sobre el eje dos, con lo cual se insta a estudiar con mayor profundidad el criterio de selección de los establecimientos productivos. Respecto a las variedades de hongos, se nota que existe una gran variabilidad de resultados en las frecuencias a medida que cambian las estaciones. Cabe destacar los resultados de la variedad BO, ya que presentó una fuerte concordancia en sus resultados a lo largo de las estaciones, y se espera que la misma se adapte bien a aquellos establecimientos con métodos de cultivos orgánicos. Por último, se espera que los establecimientos de la zona centro puedan aprovechar específicamente dicha variedad.

## Ejercicio 5

Retome los datos del archivo MAIZ.csv. Realice un análisis STATIS y obtenga la matriz de distancias euclídeas entre individuos en el plano principal. Compare dicha matriz con las obtenidas a partir de AFM y APG.

### Estimación

A continuación, se detalla el procedimiento

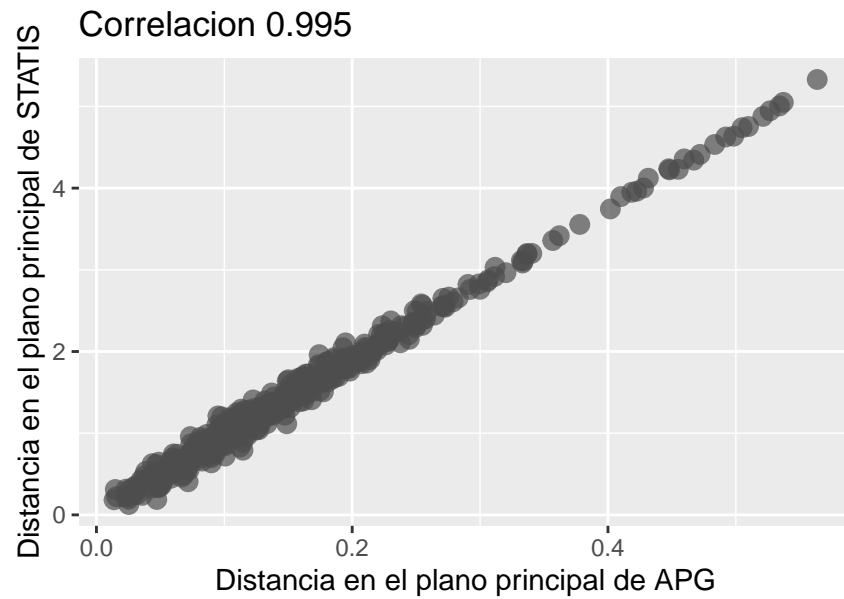
```
# Datos
datos_m1 <- read.csv2(
  here("data", "raw", "MAIZ1.csv"), row.names = 1, stringsAsFactors = FALSE
)
datos_m2 <- read.csv2(
  here("data", "raw", "MAIZ2.csv"), row.names = 1, stringsAsFactors = FALSE
)
datos <- cbind(datos_m1, datos_m2)

# STATIS
bloques <- as.integer(c(10, 10))
nombres <- c("Pergamino", "Ferre")
names(bloques) <- nombres
k <- ktab.data.frame(data.frame(scale(datos)), bloques)
s <- statisc(k, scannf = FALSE, nf = 2, tol = 1e-07)
dist_statisc <- dist(s$C.li, method = "euclidean")

# AFM
mfa <- MFA(
  base = datos,
  group = c(10, 10),
  type = c("s", "s"),
  name.group = c("Pergamino", "Ferre"),
  graph = FALSE
)
dist_mfa <- dist(mfa$ind$coord[, 1:2])

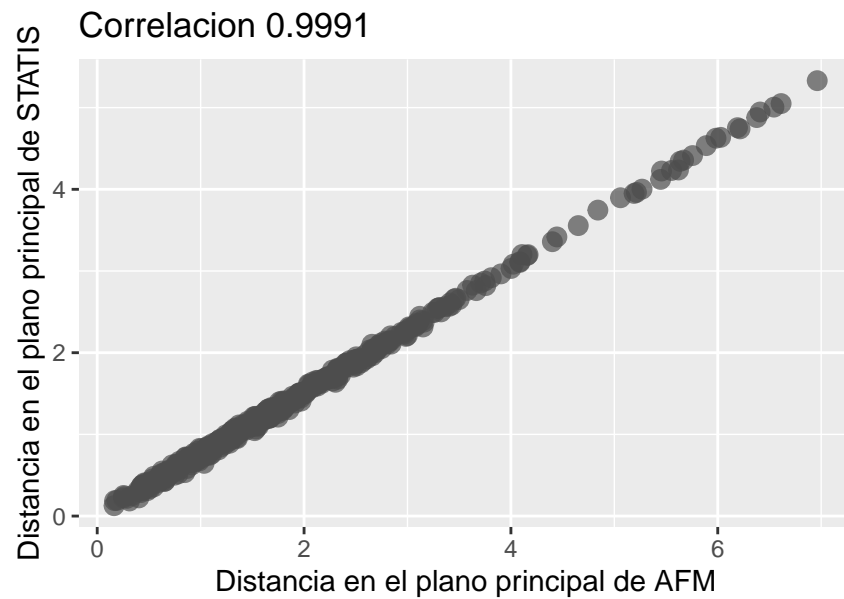
# APG.
acp_m1 <- PCA(datos_m1, ncp = 10, graph = FALSE)
acp_m2 <- PCA(datos_m2, ncp = 10, graph = FALSE)
df <- data.frame(cbind(acp_m1$ind$coord, acp_m2$ind$coord))
gpa <- GPA(df, group = c(10, 10), name.group = c("Pergamino", "Ferre"), graph = FALSE)
dist_gpa <- dist(gpa$consensus[, 1:2])
```

### Comparación de AFM - STATIS



La correlación entre ambas matrices es de 0.9950

### Comparación de AFM - STATIS



La correlación entre ambas matrices es de 0.9991

### Conclusión

STATIS y AFM coinciden mejor en término de sus configuraciones. Se observa que hay mayor variabilidad para aquellas distancias con menor valor.