

## 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.

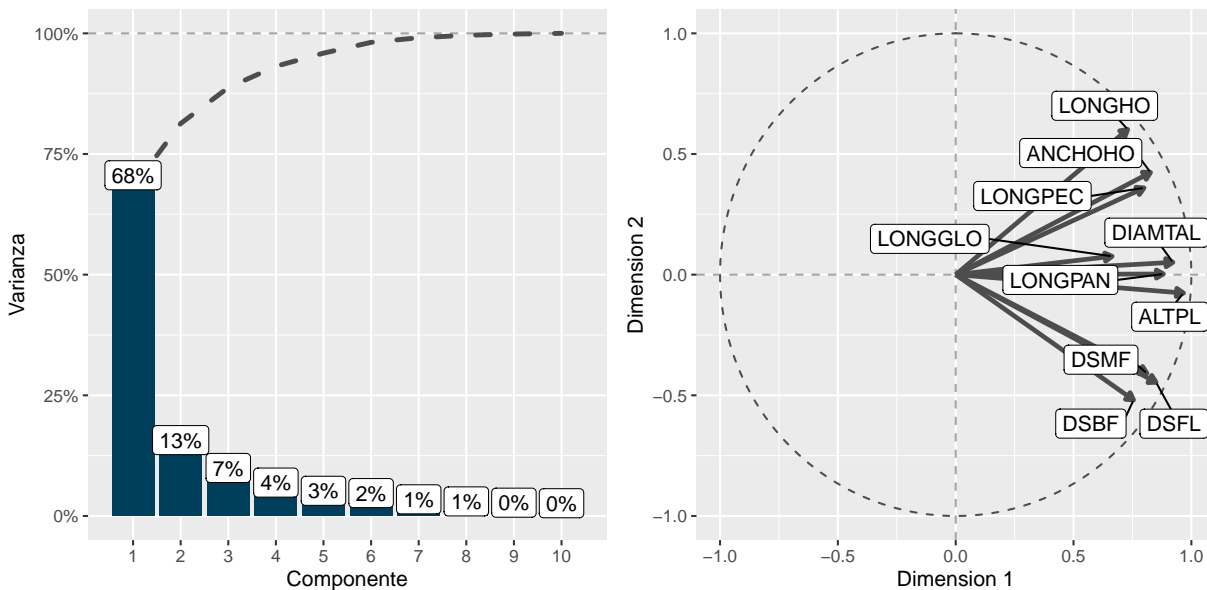


Figure 1: Izquierda: porcentaje de varianza retenido por cada componente. Derecha: Correlacion de las variables con los componentes.

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%; el 68.09% corresponde a la primera componente, mientras que el 13.22% restante corresponde a la segunda componente. 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, se puede notar que las variables que mas aportan son Longitud y Ancho de la hoja (de forma positiva) y días desde siembra a botón floral y a floración (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 sus variables:

- 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 menores 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 menores 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.

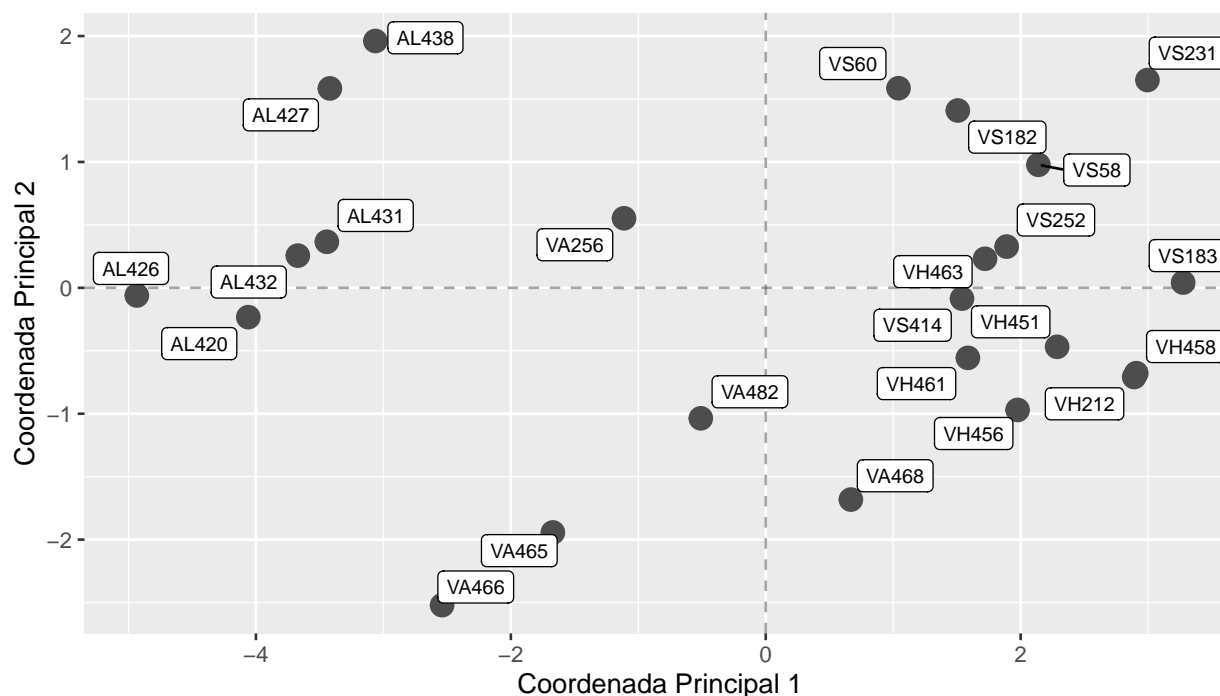
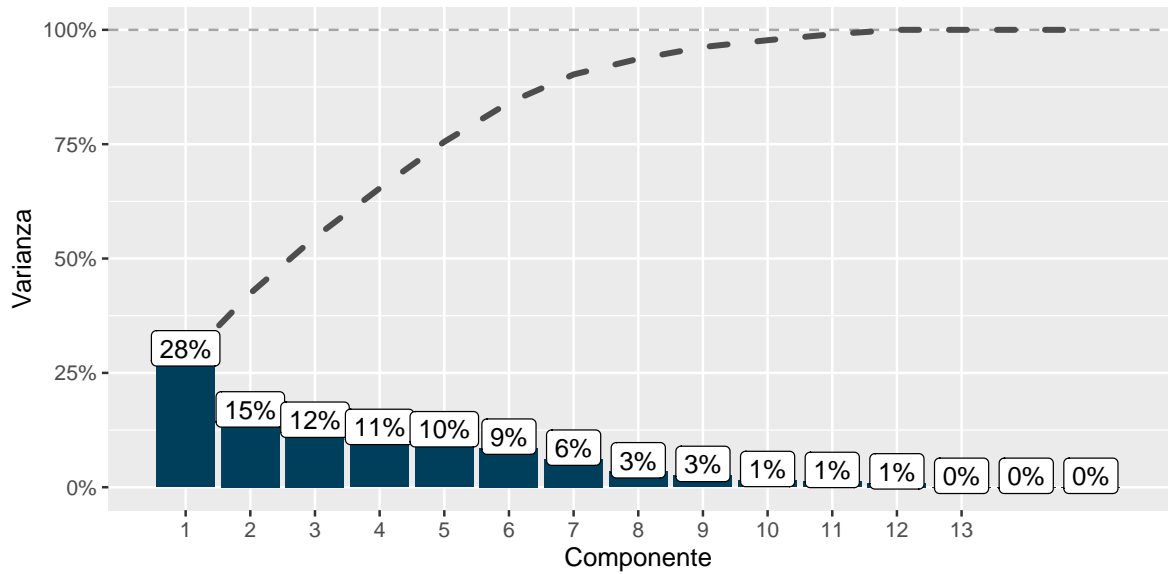


Figure 2: Accesiones de Quinoa en el plano principal del ACP.

- 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.

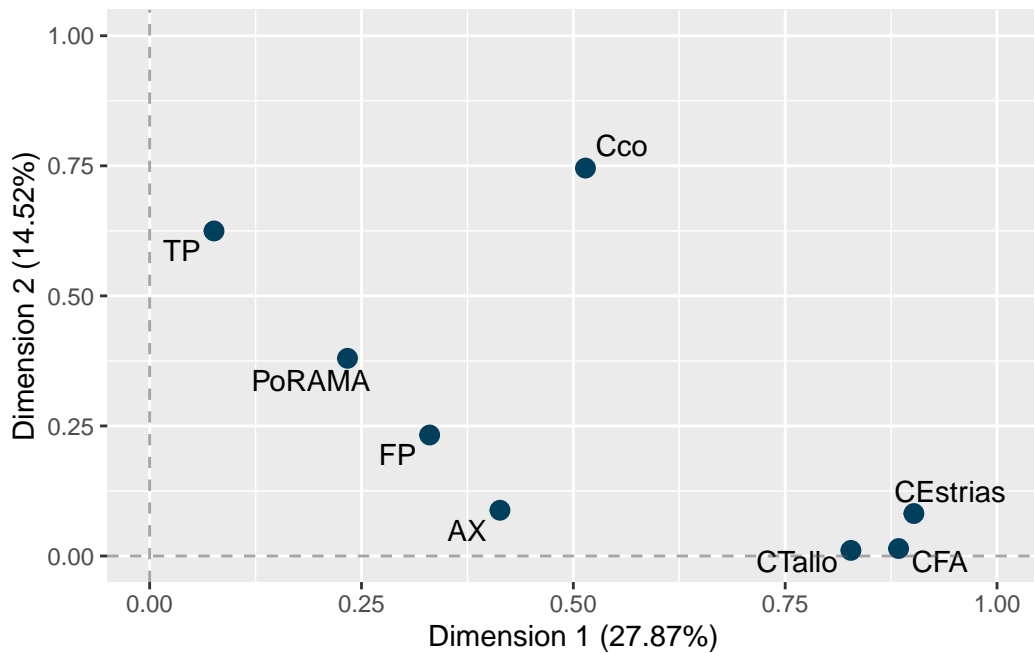


Figure 3: Contribucion de las variables a los dos primeros ejes.

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 panojy el color de la panoja a la cosecha y a 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.

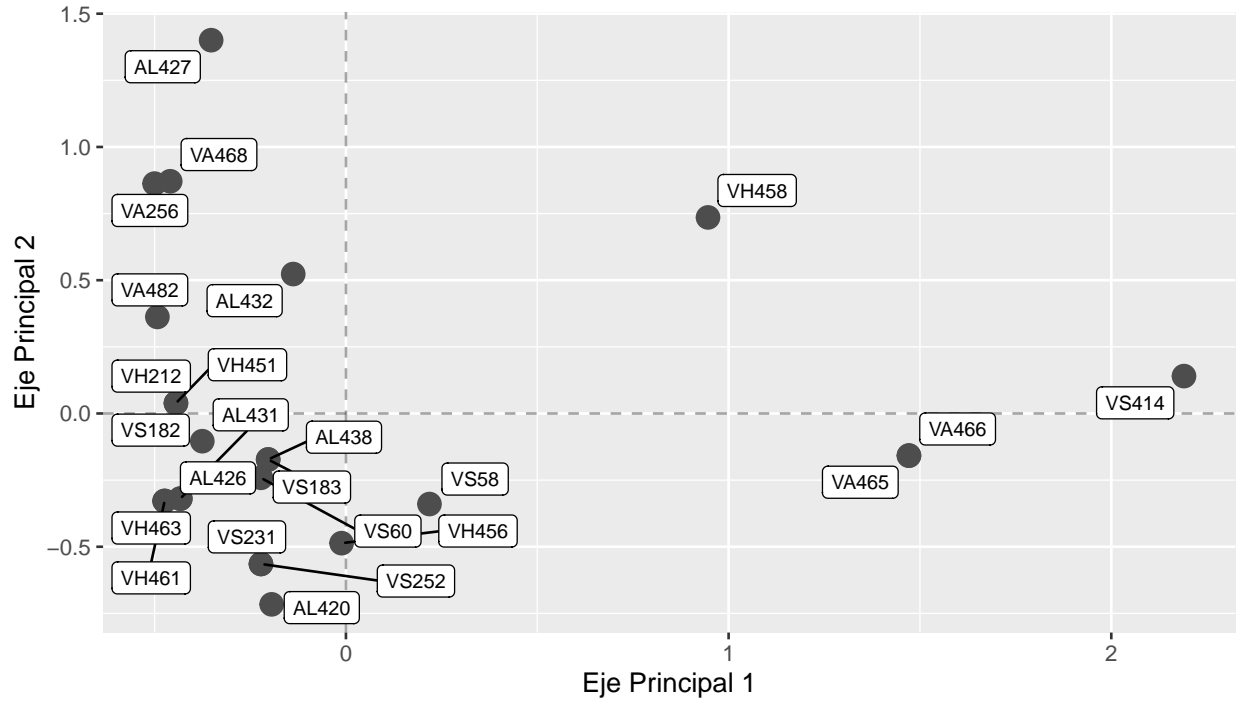


Figure 4: Configuración de los individuos en el plano principal de ACM.

Se nota en la **Figura 4** que en el análisis AFM basado en las variables cualitativas no permite formar agrupamientos en términos del ambiente en el plano principal, a diferencia de lo que sucedió con el análisis ACP basado en variables cuantitativas.

Luego, si comparamos las configuraciones de individuos provistas por las dos técnicas realizadas, podemos concluir que las mismas parecen no ser homogéneas, ya que, como se pudo observar, una técnica logra una diferenciación por ambientes mientras que la otra no.

Con alguna técnica numérica podríamos realizar esa comparación: Para poder comparar, siempre en el plano principal, en acp, se podría buscar una matriz de distancias euclídeas entre individuos, lo mismo puedo tmb hacer para acm. Luego, puedo comparar esas matrices de distancia, medir el “acuerdo”, a través de una corr de matrices de distancias (tiene que dar un valor bajo)

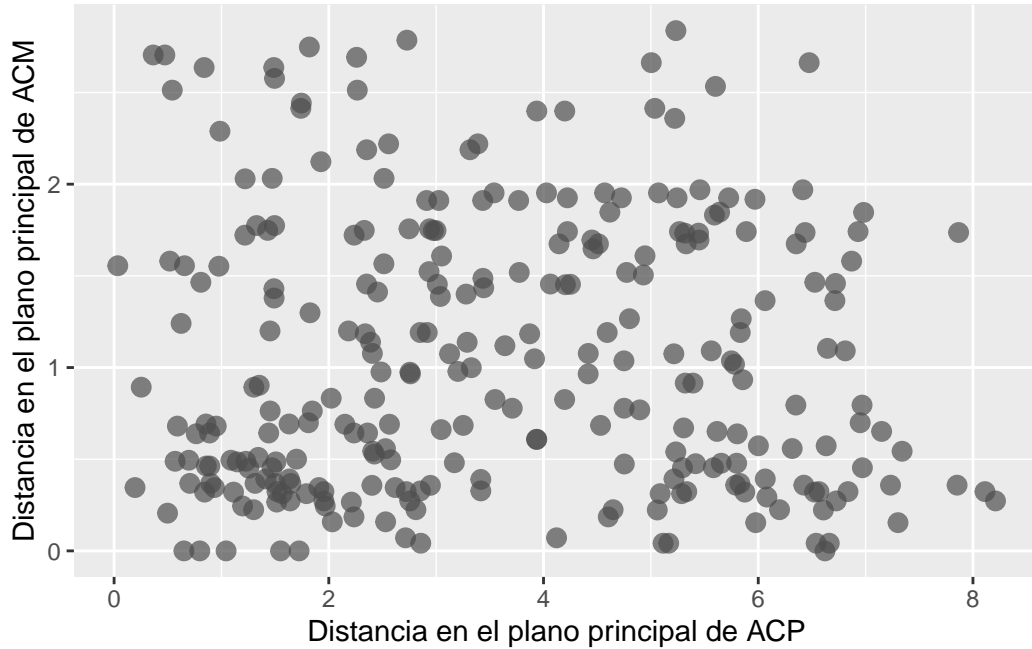


Figure 5: Comparacion de distancias en el plano principal para ACP y ACM.

El grafico nos permite ver que las representaciones sobre el plano principal no son 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:

A) 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

Al obtener el Coeficiente  $L_g$ , el cual nos brinda una medida de estructura comun entre las variables Cuantitativas y Cualitativas, observamos 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.

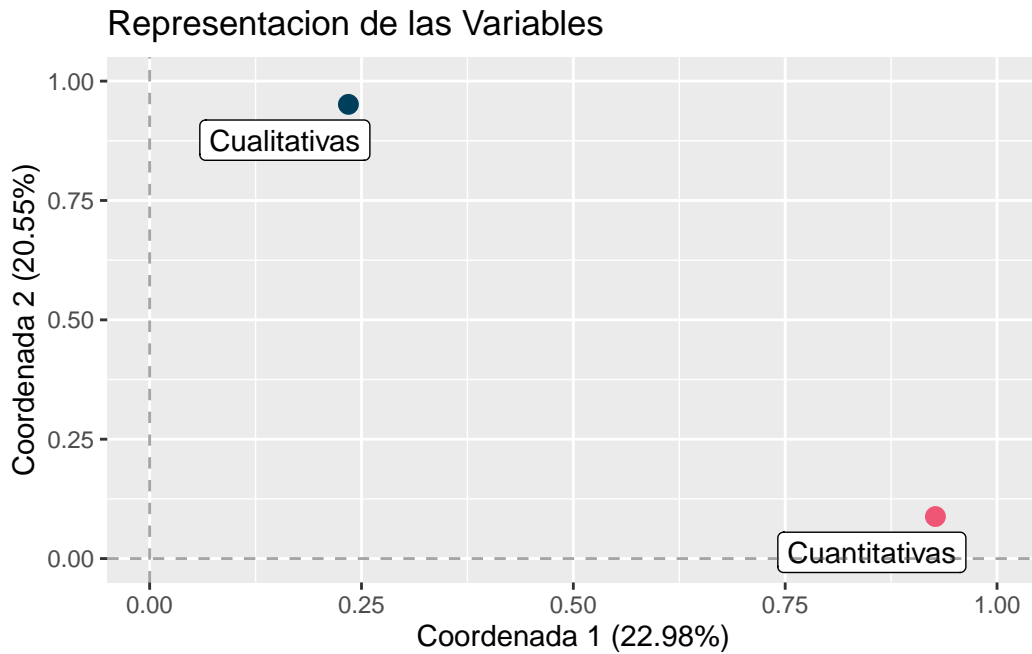


Figure 6: Caracterizacion de los grupos de variables en el plano principal AFM.

Si observamos el grafico anterior, 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.

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

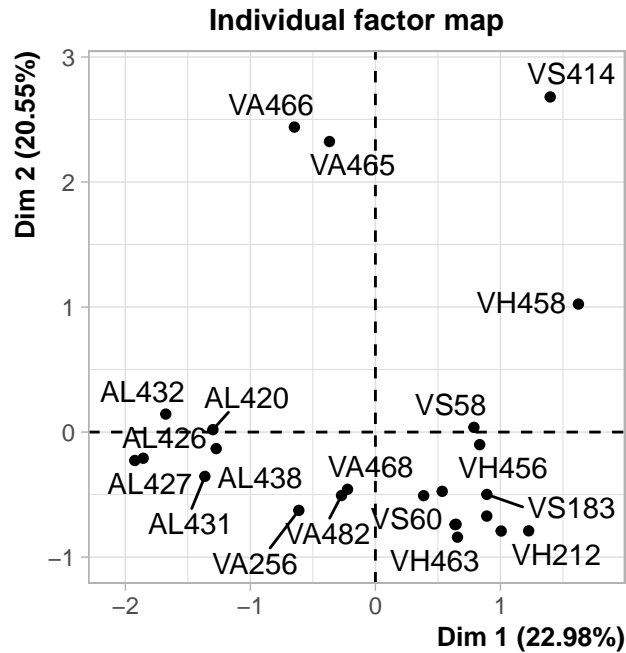
Las variables Cualitativas que más aportan en la formación del eje 2 son Color de Estrías y de Panoja a fin de antesis Púrpura

Table 5: Contribuciones a los Ejes - Cuantitativas

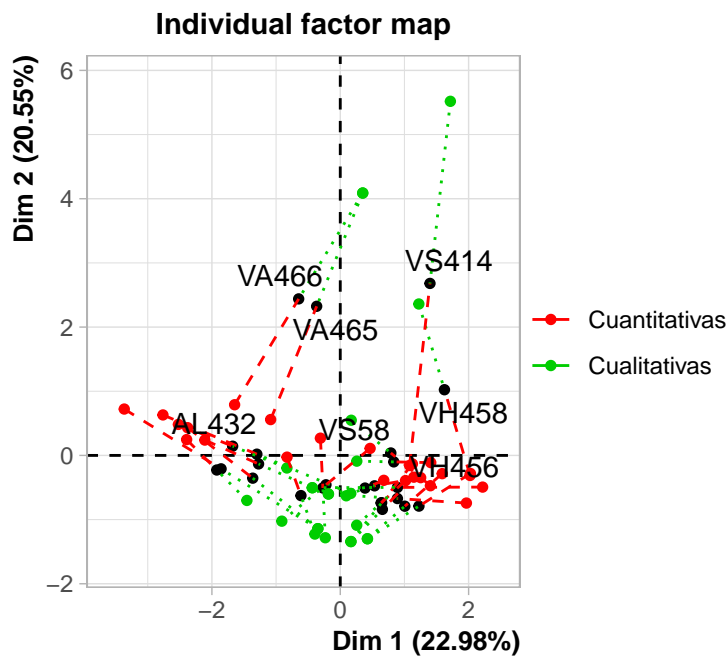
	Dim.1	Dim.2
DIAMTAL	9.41	0.77
ALTPL	10.98	0.59
LONGHO	5.48	1.79
ANCHOHO	7.47	2.85
LONGPEC	6.04	1.61
LONGPAN	8.59	0.32
LONGGLO	6.09	0.02
DSBF	8.47	0.43
DSFL	9.97	0.04
DSMF	7.31	0.04

Las variables cuantitativas que más aportan en la formación del eje 1 son: altura de planta, diámetro del tallo, longitud de la panoja, Días de siembra a botón floral y días desde siembra a floración. Es decir, aquellos individuos que posean valores altos de la primer componente van a tener... COMPLETAR

B) ¿Existe un agrupamiento de individuos por zona?



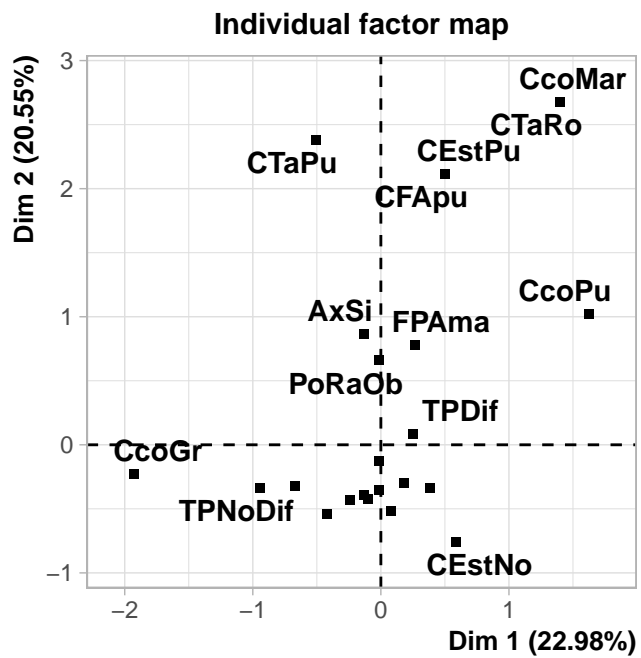
Al observar la figura VER NUMERO Q CORRESPONDE podemos referenciar algunos agrupamientos según la zona de procedencia. El primer grupo es el conformado por las variedades VA466 y VA465, procedentes de Valles de Altura; el resto de las variedades de Valles de Altura pueden asociarse en un segundo grupo. 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. Cabe descartar que hay dos “individuos”, VS414 y VH458, que se alejan del cuarto agrupamiento. Es muy probable que estos valores extremos se deban a que presenten valores más altos en la configuración Cualitativa.



Con este grafico confirmamos lo anterior, VS414 y VH458 poseen valores mas altos de cualitativas q



el resto, es por ello q se apartan del grupo. REDACTAR MEJOR!



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

Las variedades VA466 y VA465, las cuales conforman el primer grupo, se caracterizan por tener el color de las estrias y de la panoja a fin de antesis de color púrpura. Además, estas variedades tienen un comportamiento promedio con respecto a las variables cuantitativas