

Practica de componentes principales.

Ejercicio número 1:

Consideraciones:

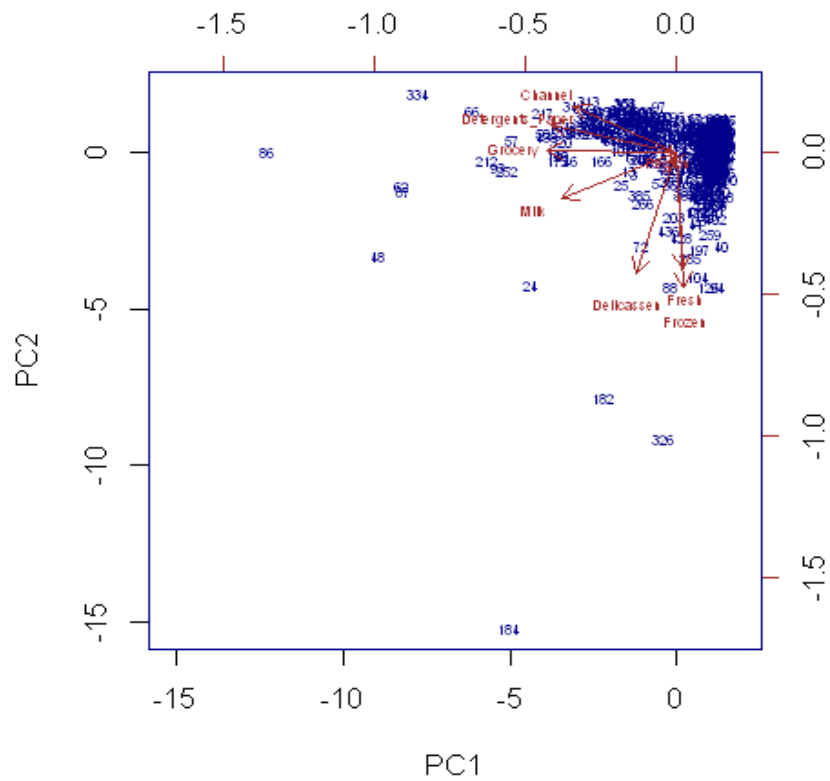
Al cargar el data set se visualizan todas las variables, se dejan todas puesto que tienen relación, se obtiene la varianza, la media y la desviación estándar de todo el data set y se guarda en un data frame, posteriormente se utiliza la función `prcomp()` para realizar el cálculo de los componentes principales.

Después de calcularlos se grafican y posteriormente se vuelve a graficar después de cambiarles el signo a dichos componentes.

Código:

```
#Practica de analisis de componentes principales ejercicio número 1  
#Se carga el data set correspondiente  
file <- read.csv(file.choose())  
  
#Descripcion de los datos  
media <- apply(file, MARGIN = 2, FUN = mean)  
varianza <- apply(file, MARGIN = 2, FUN = var)  
desviacion <- apply(file, MARGIN = 2, FUN = sd)  
  
descr_file <- data.frame(media, varianza, desviacion)  
descr_file  
  
#Se estandarizan las variables  
pca <- prcomp(file, scale = TRUE)  
names(pca)  
pca$center #media  
pca$scale #Sd  
pca$rotation #componentes principales  
  
biplot(x = pca, scale = 0.0, cex = 0.5, col = c("blue4", "brown"))  
  
#se invierten los signos de las variables  
pca$rotation <- -pca$rotation  
pca$x <- -pca$x  
biplot(x = pca, scale = 0, cex = 0.6, col = c("blue4", "brown3"))  
|
```

Primera grafica que se obtiene:

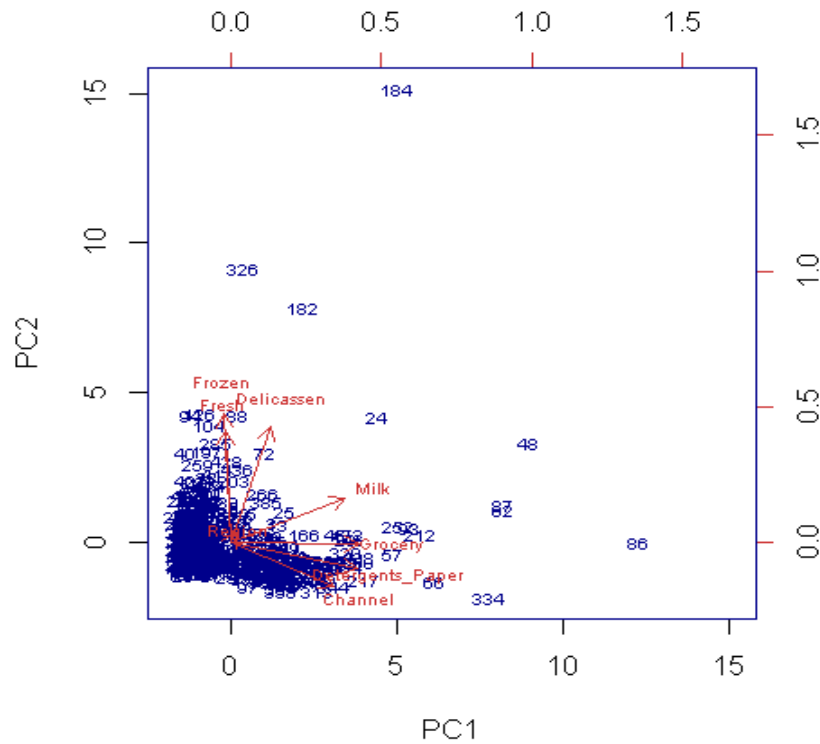


Como puede observarse salen todos los valores de manera negativa, por lo que se puede interpretar que todas las variables están alejadas excepto las variables

- Dellacase
- Fresk
- Froze

Pero como son negativos no se puede obtener algo concluyente por lo que se vuelve a graficar cambiando los signos

Grafica 2:



De esta grafica ya podemos tener algo más concluyente puesto que los valores ya son positivos y se mantienen las variables iguales, las más cercanas son las variables de:

- Frozen
- Delicassen
- Fresh

Mientras que las más alejadas son la de:

- Milk
- Channel
- Grocery
- Detergents_Paper

Ejercicio número 2:

Consideraciones:

Al cargar el data set se puede observar que una columna es de tipo texto o string por lo que se elimina de la data set para poder hacer el calculo de los componentes principales. Restando esto se procede a obtener la media la varianza y la desviación del conjunto de datos.

Código:

```
#Practica de analisis de componentes principales ejercicio número 2
#Se carga el data set correspondiente
vino <- read.csv(file.choose())
vino <- data.frame(vino$fixed.acidity, vino$volatile.acidity,
                  vino$citric.acid, vino$residual.sugar,
                  vino$chlorides, vino$free.sulfur.dioxide,
                  vino$total.sulfur.dioxide, vino$density,
                  vino$pH, vino$sulphates,
                  vino$alcohol, vino$quality)

View(vino)

#Descripcion de los datos
media <- apply(vino, MARGIN = 2, FUN = mean)
varianza <- apply(vino, MARGIN = 2, FUN = var)
desviacion <- apply(vino, MARGIN = 2, FUN = sd)

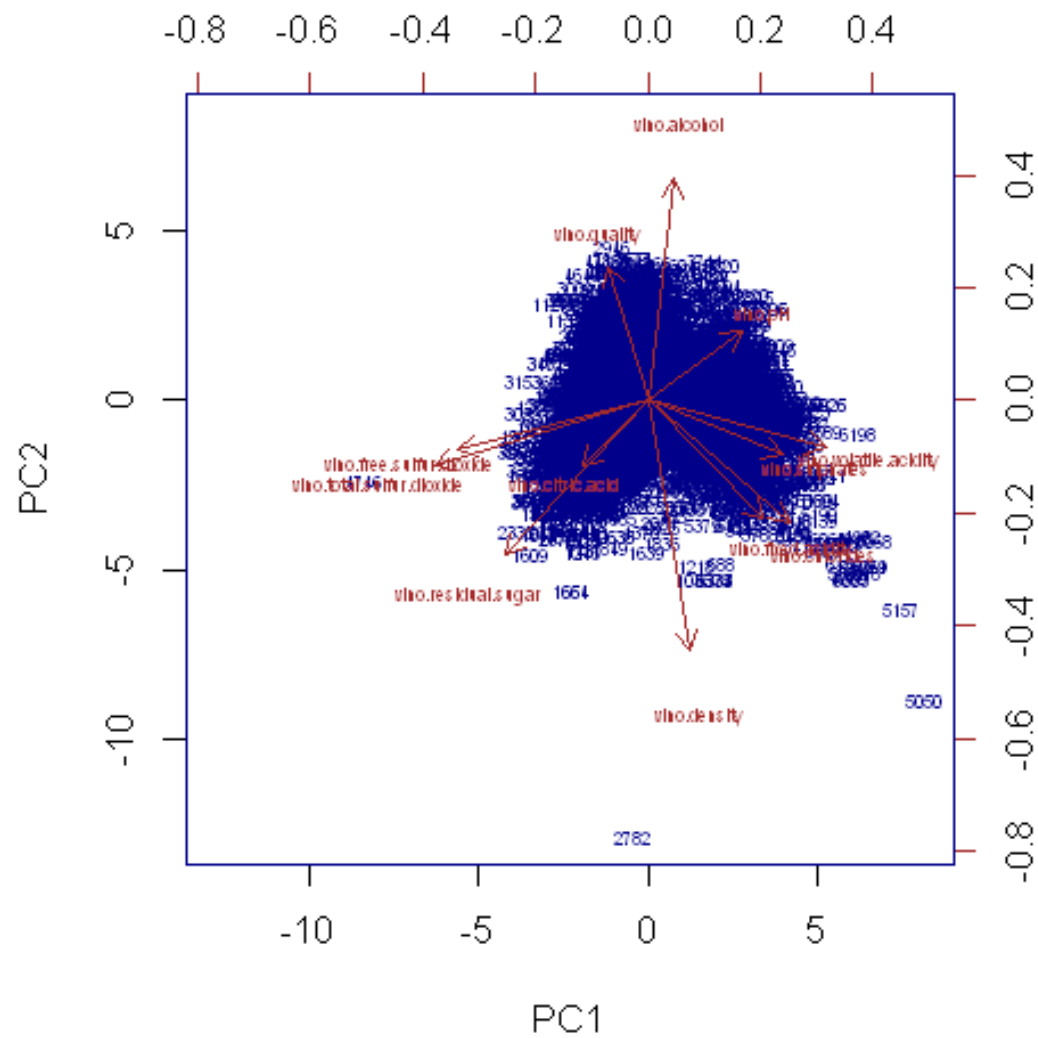
descr_file <- data.frame(media, varianza, desviacion)
descr_file

#Se estandarizan las variables
acp <- prcomp(vino, scale = TRUE)
names(acp)
acp$center #media
acp$scale #Sd
acp$rotation #componentes principales

biplot(x = acp, scale = 0.0, cex = 0.5, col = c("blue4", "brown"))

#se invierten los signos de las variables
pca$rotation <- -pca$rotation
pca$x <- -pca$x
biplot(x = pca, scale = 0, cex = 0.6, col = c("blue4", "brown3"))
```

Con lo siguiente obtenemos el siguiente gráfico:



Dado la grafica anterior podemos concluir que la mayoría de los datos tiene están con las variables:

- Alcohol
- Quality

Y otra variable que los puntos tapan, es más podríamos decir que esta un medio distribuido las variables que están relacionada con al menos tres variables de la data set.

Conclusiones.

En esta practica hemos plasmado los temas vistos en las ultimas 2 sesiones de clases en la que utilizamos el análisis de los componentes principales, quizá no haciéndolo de forma manual (paso a paso) si no con las funciones incorporadas de RStudio.

El análisis de los datos de esta forma nos dimensiona de mejor manera como tiene relación dos o más variables entre ellas usando la desviación y la varianza al igual que la media para normalizar dichos datos y que así sea más fácil nuestra labor en el análisis de estos datos.