

Ejercicio 3.1 *Demostrar que la matriz de correlaciones de X es la matriz de la transformación $XD^{-1/2}$, con $D = \text{diag}(s_1, \dots, s_p)$ y s_i la desviación típica de la variable x_i .*

Para esta demostración partimos de la matriz centrada $Y = XD^{-1/2}$ que es la matriz de transformación con lo que tenemos que:

$$S_y = \frac{1}{n} Y' H Y$$

Sustituyendo la Y obtenemos:

$$\frac{1}{n} Y' H Y = D^{-1/2} \frac{1}{n} X' H X D^{-1/2}$$

Siendo $X' H X = S_x$:

$$D^{-1/2} \frac{1}{n} X' H X D^{-1/2} = D^{-1/2} S_x D^{-1/2} = R$$

Siendo R la matriz de correlación con lo que queda demostrado que la matriz de correlación es la transformación que estábamos buscando.

Ejercicio 3.2 *Ejercicio 3.3: Realizar un ACP sobre los datos del fichero.
http://campusvirtual.ugr.es/moodle/file.php/1934/Datos_nutricionales_de_alimentos.csv.*

Lo primero que hacemos es definir la matriz como datos:

```
datos<-read.csv('Alimentos.csv',
  header=TRUE)
```

Lo que queremos ver a continuación es lo correlacionados que están los datos por lo que primero abriremos con el comando `attach` para trabajar con los diferentes datos de la tabla, una vez hecho esto haremos la correlación de los datos:

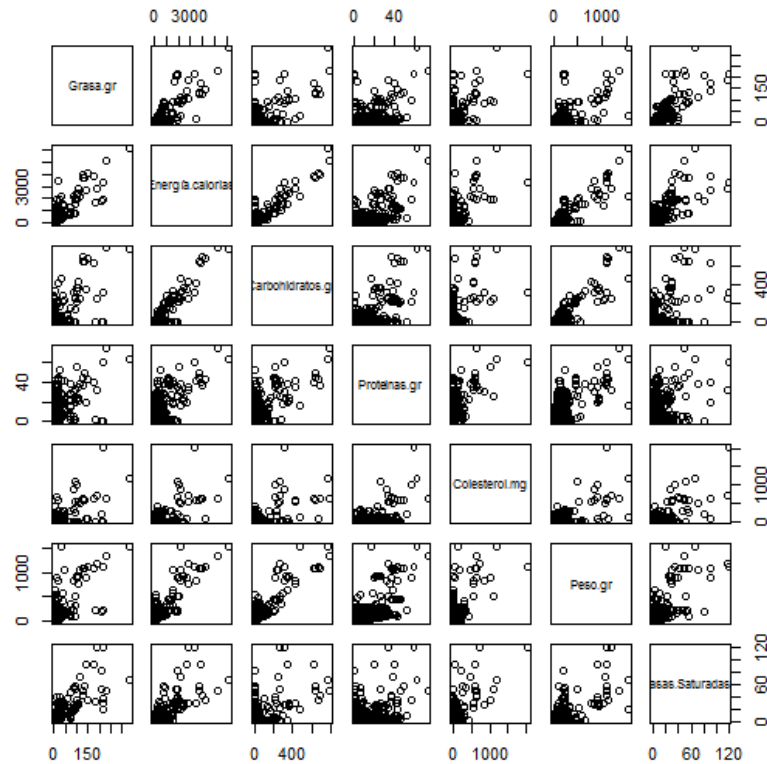
```
attach(datos)
cor(datos)
```

Una vez hecho esto lo vamos a hacer para que salgan solo con dos decimales para que se vea mejor.

```
round(cor(datos), 2)
```

A continuación veremos la representación gráfica de los datos y cerraremos el attach con el detach:

```
plot(datos)
detach(datos)
```



En la gráfica podemos observar que efectivamente la mayor parte de los datos están correlacionados. Ahora cogeremos las componentes principales:

```
datos_pca=prcomp(datos, scale=T)
```

Para ver la cantidad de variabilidad explicada usaremos el comando `summary`:

```
summary(datos_pca)
```

```
Importance of components:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7
Standard deviation  2.2047 0.8793 0.77381 0.62790 0.45693 0.35853 0.1889
Proportion of Variance 0.6944 0.1105 0.08554 0.05632 0.02983 0.01836 0.0051
Cumulative Proportion 0.6944 0.8048 0.89039 0.94671 0.97654 0.99490 1.0000
```

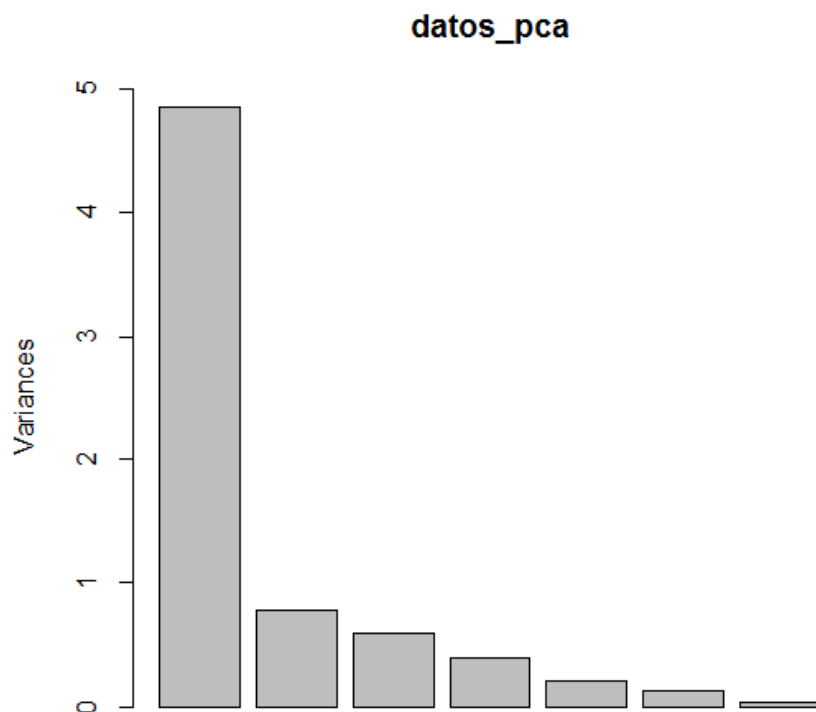
En el `summary` podemos observar por ejemplo que la desviación estandar va disminuyendo con forme va avanzando el número de componentes, la varianza de la proporción también mientras que en la última fila nos dicen que por ejemplo la primera componente solo explica el 69.44 % de la distribución y claramente la última explica el 100 %.

Para extraer la primera componente de la matriz usamos la función `predict()[1]`, el uno es para que lo haga solo de la primera variable, de la forma:

```
predict(datos_pca)[,1]
```

Y para terminar obtendremos la gráfica de barras con el comando:

```
plot(datos_pca, ylim=c(0,5))
```



Las barras explican la cantidad de variabilidad que explican las componentes principales.