act_pca

Rogelio Lizárraga

2024-10-08

PARTE I

Realiza el análisis de los valores y vectores propios con la matriz de covarianzas y con la de correlación. Analiza la varianza explicada por cada componente en cada caso e interpreta dentro del contexto del problema.

```
library(FactoMineR)
library(factoextra)
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
datos <- read.csv("corporal.csv")  # Ejemplo si los datos están en un CSV
datos
```

```
##    edad peso altura   sexo muneca biceps
## 1    43 87.3  188.0 Hombre   12.2   35.8
## 2    65 80.0  174.0 Hombre   12.0   35.0
## 3    45 82.3  176.5 Hombre   11.2   38.5
## 4    37 73.6  180.3 Hombre   11.2   32.2
## 5    55 74.1  167.6 Hombre   11.8   32.9
## 6    33 85.9  188.0 Hombre   12.4   38.5
## 7    25 73.2  180.3 Hombre   10.6   38.3
## 8    35 76.3  167.6 Hombre   11.3   35.0
## 9    28 65.9  183.0 Hombre   10.2   32.1
## 10   26 90.9  183.0 Hombre   12.0   40.4
## 11   43 89.1  179.1 Hombre   11.3   36.5
## 12   30 62.3  170.2 Hombre   11.5   34.2
## 13   26 82.7  177.8 Hombre   11.5   35.2
## 14   51 79.1  179.1 Hombre   11.8   34.0
## 15   30 98.2  190.5 Hombre   10.7   34.8
## 16   24 84.1  177.8 Hombre   11.5   38.6
## 17   35 83.2  180.3 Hombre   11.1   36.4
## 18   37 83.2  180.3 Hombre   10.5   34.0
## 19   22 51.6  161.2  Mujer    9.2   24.3
## 20   20 59.0  167.5  Mujer    9.9   27.8
## 21   19 49.2  159.5  Mujer    8.9   24.0
```

```
## 22   25 63.0  157.0  Mujer    9.5   28.0
## 23   21 53.6  155.8  Mujer    9.1   26.9
## 24   23 59.0  170.0  Mujer   10.0   26.5
## 25   26 47.6  159.1  Mujer    9.4   24.1
## 26   22 69.8  166.0  Mujer   10.7   29.2
## 27   28 66.8  176.2  Mujer    9.8   29.0
## 28   40 75.2  160.2  Mujer   11.5   33.6
## 29   32 55.2  172.5  Mujer    8.6   24.8
## 30   25 54.2  170.9  Mujer    9.7   25.4
## 31   25 62.5  172.9  Mujer    9.2   25.9
## 32   29 42.0  153.4  Mujer    8.3   24.0
## 33   22 50.0  160.0  Mujer    8.6   25.6
## 34   25 49.8  147.2  Mujer    9.0   26.0
## 35   23 49.2  168.2  Mujer    9.6   23.5
## 36   37 73.2  175.0  Mujer   11.0   31.0
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
datos <- datos %>%select(-sexo)
summary(datos)
```

```
##       edad            peso           altura          muneca
##  Min.   :19.00   Min.   :42.00   Min.   :147.2   Min.   : 8.300
##  1st Qu.:24.75   1st Qu.:54.95   1st Qu.:164.8   1st Qu.: 9.475
##  Median :28.00   Median :71.50   Median :172.7   Median :10.650
##  Mean   :31.44   Mean   :68.95   Mean   :171.6   Mean   :10.467
##  3rd Qu.:37.00   3rd Qu.:82.40   3rd Qu.:179.4   3rd Qu.:11.500
##  Max.   :65.00   Max.   :98.20   Max.   :190.5   Max.   :12.400
##      biceps
##  Min.   :23.50
##  1st Qu.:25.98
##  Median :32.15
##  Mean   :31.17
##  3rd Qu.:35.05
##  Max.   :40.40
```

Eliminaremos sexo, pues es una variable categórica y para poder aplica análisis de componentes principales necesitamos trabajar con variables numéricas. Como podemos observar: Las edades en el conjunto de datos van desde los 19 hasta los 65 años. El promedio de edad es 31.44 años. Los pesos van desde 42 kg a 98.2 kg, con un promedio de 68.95 kg. Las alturas varían entre 147.2 cm y 190.5 cm, con una media de 171.6 cm. El perímetro de la muñeca va de 8.3 cm a 12.4 cm, con una media de 10.467 cm. Los tamaños de bíceps oscilan entre 23.5 cm y 40.4 cm, con un promedio de 31.17 cm. ### Calcule las matrices de varianza-covarianza S con cov(X) y la matriz de correlaciones R con cor(X) y realice los siguientes pasos con cada una: Calcule los valores y vectores propios de cada matriz.La función en R es: eigen(). Calcule la proporción de varianza explicada por cada componente en ambas matrices. Se sugiere dividir cada lambda entre la varianza total (las lambdas están en eigen(S)$values). La varianza total es la suma de las varianzas de la diagonal de S. Una forma es sum(diag(S)). La varianza total de los componentes es la suma de los valores propios (es decir, la suma de la varianza de cada componente), sin embargo, si sumas la diagonal de S (es decir, la varianza de cada x), te da el mismo valor (¡comprúebalo!). Recuerda que las combinaciones lineales buscan reproducir la varianza de X.

```
S <-cov(datos)
R<- cor(datos)
cat('Matriz de varianza-covarianza:\n')
```

```
## Matriz de varianza-covarianza:
```

```
S
```

```
##              edad      peso     altura    muneca    biceps
## edad   111.396825  80.88159  36.666032  7.698095 26.720952
## peso    80.881587 221.08713 124.728698 14.844667 70.738381
## altura  36.666032 124.72870 110.673968  8.156476 39.021048
## muneca   7.698095  14.84467   8.156476  1.381714  5.400571
## biceps  26.720952  70.73838  39.021048  5.400571 27.398857
```

```
cat('Matriz de correlaciones:\n')
```

```
## Matriz de correlaciones:
```

```
R
```

```
##             edad      peso    altura    muneca    biceps
## edad   1.0000000 0.5153847 0.3302211 0.6204942 0.4836702
## peso   0.5153847 1.0000000 0.7973737 0.8493361 0.9088813
## altura 0.3302211 0.7973737 1.0000000 0.6595849 0.7086144
## muneca 0.6204942 0.8493361 0.6595849 1.0000000 0.8777369
## biceps 0.4836702 0.9088813 0.7086144 0.8777369 1.0000000
```

```
# Valores y vectores propios de la matriz de varianza-covarianza
eigen_S <- eigen(S)
eigen_R <- eigen(R)
cat('Valores propios S:', eigen_S$values)
```

```
## Valores propios S: 359.398 80.37579 27.6229 4.307432 0.2343571
```

```
cat('\nVectores propios S\n:')
```

```
##
## Vectores propios S
## :
```

```r
eigen_S$vectors
```

```
##             [,1]       [,2]        [,3]         [,4]         [,5]
## [1,] -0.34871002  0.9075501 -0.23248825 -0.001589466  0.026473941
## [2,] -0.76617586 -0.1616581  0.52166894 -0.338508602  0.010707863
## [3,] -0.47632405 -0.3851755 -0.78905759  0.046160807  0.003543154
## [4,] -0.05386189  0.0155423  0.02785902  0.126103480 -0.990039959
## [5,] -0.24817367 -0.0402221  0.22455005  0.931330496  0.137814357
```

```r
cat('\nValores propios R:', eigen_R$values)
```

```
##
## Valores propios R: 3.757497 0.7258566 0.3203298 0.1246187 0.07169749
```

```r
cat('\nVectores propios R:\n')
```

```
##
## Vectores propios R:
```

```r
eigen_R$vectors
```

```
##            [,1]       [,2]        [,3]       [,4]       [,5]
## [1,] -0.3359310  0.8575601 -0.34913780 -0.1360111  0.1065123
## [2,] -0.4927066 -0.1647821  0.06924561 -0.5249533 -0.6706087
## [3,] -0.4222426 -0.4542223 -0.73394453  0.2070673  0.1839617
## [4,] -0.4821923  0.1082775  0.36690716  0.7551547 -0.2255818
## [5,] -0.4833139 -0.1392684  0.44722747 -0.3046138  0.6739511
```

En las matrices anteriores observamos lo siguiente: En la matriz de varianza-covarianza se cuenta con las varianzas:

Edad: 111.4 (alta varianza, lo que indica una variabilidad significativa en la edad). Peso: 221.1 (alta varianza, lo que indica un amplio rango de pesos). Altura: 110.7 Muñeca: 1.38 (baja varianza, lo que indica que el tamaño de la muñeca no varía mucho). Bíceps: 27.4

y las covarianzas:

Edad y peso: 80.88 (covarianza positiva, lo que significa que a medida que aumenta la edad, el peso tiende a aumentar también). Peso y altura: 124.7 (covarianza positiva, lo que indica que las personas más altas tienden a pesar más). Muñeca y bíceps: 5.40 (covarianza positiva, mostrando una relación entre las circunferencias de la muñeca y los bíceps).

En la matriz de correlaciones observamos las siguientes correlaciones: Edad y peso: Correlación = 0.515 (correlación positiva moderada, lo que significa que las personas mayores tienden a pesar más). Peso y altura: Correlación = 0.797 (correlación positiva fuerte, lo que indica que las personas más altas generalmente pesan más).
Muñeca y bíceps: Correlación = 0.878 (correlación positiva muy fuerte, lo que significa que las personas con mayor circunferencia de muñeca tienden a tener bíceps más grandes).

En los eigenvalores de S: Eigenvalor 1 (359.398): Explica la mayor parte de la varianza. Eigenvalor 2 (80.37579): Explica menos varianza que el primero. Eigenvalor 3 (27.6229), Eigenvalor 4 (4.307432), Eigenvalor 5(0.2343571): Estos explican porciones progresivamente menores de la varianza.

En la matriz de de los eigenvectores de S:

El primer componente tiene cargas negativas fuertes tanto para la edad (-0.3487), como para el peso (-0.766) y la altura (-0.4763), por lo que el primer componente principal está fuertemente influenciado por la edad, el peso y la altura. El segundo componente tiene una carga positiva fuerte para la edad (0.90755) y una carga negativa fuerte para la altura (-0.385), por lo que la edad y la altura contribuye mucho a este componente.

En los eigenvalores de R:

Estos valores propios representan la cantidad de varianza explicada cuando los datos están estandarizados. Eigenvalor 1 (3.757497): Explica una porción significativa de la varianza estandarizada. Eigenvalor 2 (0.7258566): Explica una porción relevante de la varianza. El eigenvalor 3, 4 y 5 (0.3203298 0.1246187 0.07169749) explican porciones de la varianza muy bajas.

Estos vectores propios representan las cargas de cada variable al utilizar los datos estandarizados (matriz de correlación).

En el primer compoentne,la edad tiene una carga de -0.3359310 y el peso de -0.4927066, y así sucesivamente.

Se hará un análisis más relevante posteriormente de la varianza explicada y los pesos de las variables sobre los componentes principales.

Acumule los resultados anteriores (cumsum() puede servirle) para obtener la varianza acumulada en cada componente.

Según los resultados anteriores, ¿qué componentes son los más importantes?

```
# Proporción de varianza explicada para la matriz de covarianza
proporcion_varianza_cov <- eigen_S$values / sum(eigen_S$values)
#Proporión de varianza explicada para la matriz de correlación
proporcion_varianza_cor <- eigen_R$values / sum(eigen_R$values)

# Mostrar la proporción de varianza explicada por componente (covarianza)
cat('% de varianza explicada por componente en matriz S:\n')
```

```
## % de varianza explicada por componente en matriz S:
```

```
proporcion_varianza_cov
```

```
## [1] 0.7615357176 0.1703098726 0.0585307219 0.0091271040 0.0004965839
```

```
# Mostrar la proporción de varianza explicada por componente (correlación)
cat('% de varianza explicada por componente en matriz R:\n')
```

```
## % de varianza explicada por componente en matriz R:
```

```
print(proporcion_varianza_cor)
```

```
## [1] 0.75149947 0.14517133 0.06406596 0.02492375 0.01433950
```

```
# Mostrar la proporción de varianza acumulada para la matriz de covarianza
cat('% de varianza acumulada en matriz S:\n')
```

```
## % de varianza acumulada en matriz S:
```

```r
varianza_acumulada_cov <- cumsum(proporcion_varianza_cov)
print(varianza_acumulada_cov)
```

```
## [1] 0.7615357 0.9318456 0.9903763 0.9995034 1.0000000
```

```r
# Mostrar la proporción de varianza acumulada para la matriz de correlación
cat('% de varianza acumulada en matriz R:\n')
```

```
## % de varianza acumulada en matriz R:
```

```r
varianza_acumulada_cor <- cumsum(proporcion_varianza_cor)
print(varianza_acumulada_cor)
```

```
## [1] 0.7514995 0.8966708 0.9607368 0.9856605 1.0000000
```

```r
# Similarmente, muestra la varianza acumulada usando la matriz de correlación
```

Los componentes más importantes son el PC1 y el PC2, pues el primer componente explica el 76.15% de la varianza total y el segundo componente explica el 17.03% de la varianza total en la matriz de varianza-covarianza, mientras que el primer componente explica el 75.14% de la varianza total y el segundo componente explica el 14.57% de la varianza total en la matriz de correlaciones. Además, con estos dos componentes se logra una varianza $\approx 90\%$ de ambas matrices, lo cual es bastante bueno.

Escriba la ecuación de la combinación lineal de los Componentes principales CP1 y CP2 (eiX, donde ei está en eigen(S)$vectors[1], e2X para obtener CP2, donde X = c(X1, X2, . . .)) ¿qué variables son las que más contribuyen a la primera y segunda componentes principales? (observe los coeficientes en valor absoluto de las combinaciones lineales). Justifique su respuesta.

```r
# Ecuación de la combinación lineal para los primeros dos componentes
combinacion_CP1 <-  eigen_S$vectors[,1]  # Coeficientes del primer componente principal
combinacion_CP2 <-eigen_S$vectors[,2]  # Coeficientes del segundo componente principal

# Muestra las ecuaciones
print(round(combinacion_CP1,3))
```

```
## [1] -0.349 -0.766 -0.476 -0.054 -0.248
```

```r
print(round(combinacion_CP2,3))
```

```
## [1]  0.908 -0.162 -0.385  0.016 -0.040
```

Como podemos observar, nuestras ecuaciones para la matriz S son: $PC_1 = -0.349X_1 - 0.766X_2 - 0.476X_3 - 0.054X_4 - 0.248X_5$ $PC_2 = 0.908X_1 - 0.162X_2 - 0.385X_3 + 0.016X_4 - 0.040X_5$

```r
# Ecuación de la combinación lineal para los primeros dos componentes
combinacion_CP1 <-  eigen_R$vectors[,1]  # Coeficientes del primer componente principal
combinacion_CP2 <-eigen_R$vectors[,2]  # Coeficientes del segundo componente principal

# Muestra las ecuaciones
print(round(combinacion_CP1,3))
```

```
## [1] -0.336 -0.493 -0.422 -0.482 -0.483
```

```r
print(round(combinacion_CP2,3))
```

```
## [1]  0.858 -0.165 -0.454  0.108 -0.139
```

Y para la matriz R son: $PC_1 = -0.336X_1 - 0.493X_2 - 0.422X_3 - 0.482X_4 - 0.483X_5$ $PC_2 = 0.858X_1 - 0.165X_2 - 0.454X_3 0.108X_4 - 0.139X_5$

PARTE II. Gráficas con los componentes principales

Obtenga las gráficas respectivas con S (matriz de varianzas-covarianzas) y con R (matriz de correlaciones) de las dos primeras componentes. Calcule las puntuaciones (scores) de las observaciones para los componentes obtenidos con la matriz de varianzas-covarianzas Calcule las puntuaciones (scores) de las observaciones para los componentes obtenidos con la matriz de correlaciones. Recuerde que en la matriz de correlaciones las variables tienen que estar estandarizadas.

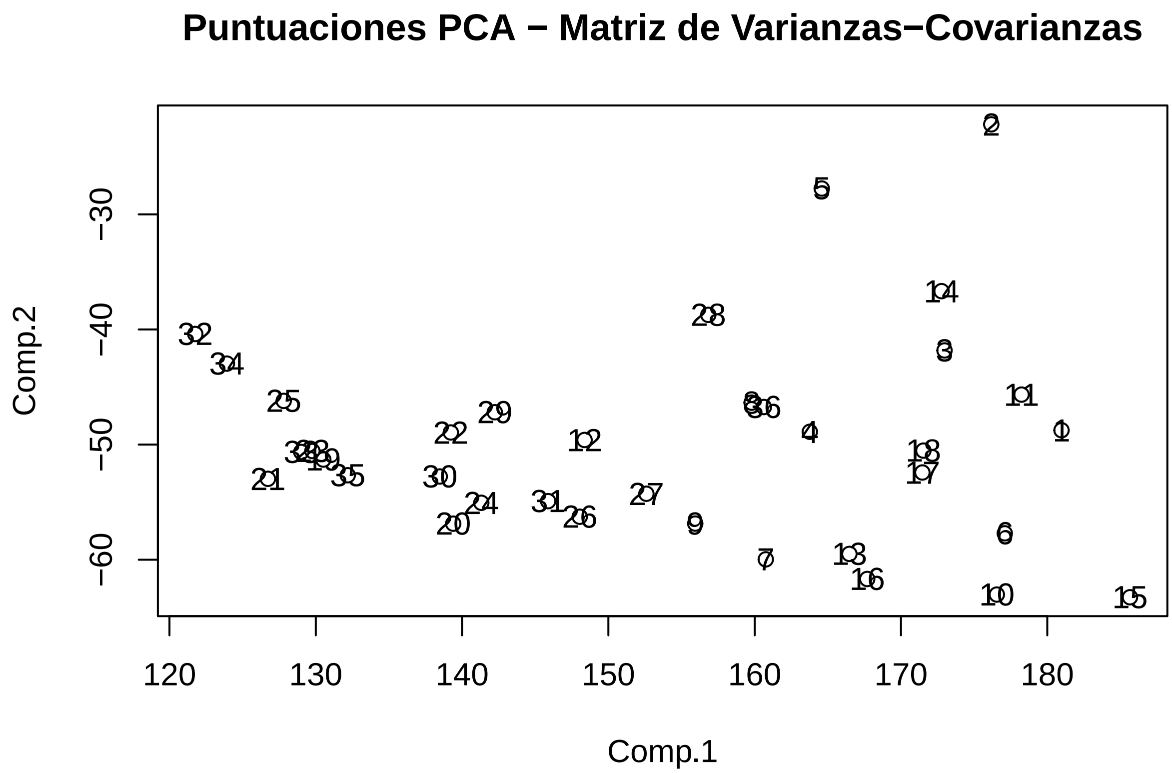
Interprete los gráficos en términos de: Las relaciones que se establecen entre las variables y los componentes principales La relación entre las puntuaciones de las observaciones y los valores de las variables. Detecte posibles datos atípicos

Explora el: princomp() en library(stats). Puedes poner help(princomp) en la consola o buscarlo en la ventana de ayuda. Indaga: ¿qué otras opciones tiene para facilitarte el análisis? En particular, explora los comandos y subcomandos: summary(cpS), cpaS$loading, cpaS$scores. ¿Cómo se interpreta el resultado?

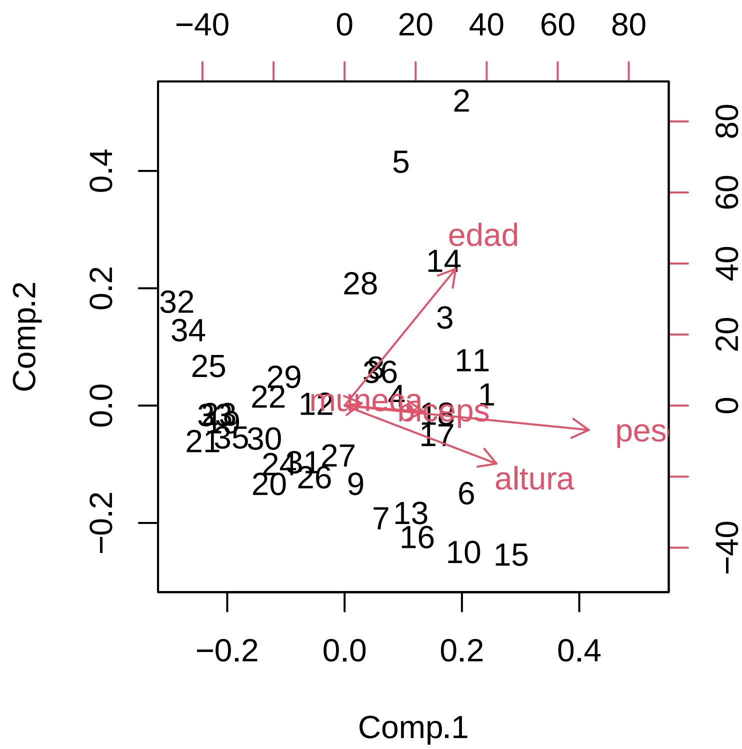
Matriz de varianza-covarianza

```r
library(stats)
cpS <- princomp(datos, cor = FALSE)
cpaS <- as.matrix(datos) %*% cpS$loadings

plot(cpaS[,1:2], type = "p", main = "Puntuaciones PCA - Matriz de Varianzas-Covarianzas")
text(cpaS[,1], cpaS[,2], labels = 1:nrow(cpaS))
```


Puntuaciones PCA − Matriz de Varianzas−Covarianzas
Comp.1
Comp.2
120
130
140
150
160
170
180
−30
−40
−50
−60

```r
biplot(cpS)
```



En el primer gŕafico, cada número que se ve en el gráfico es una observación, y los ejes representan los primeros dos componentes principales, que fueron los seleccionados, al ser los que tenían una mayor explicación de la varianza. Observaciones como la 2 y la 5 están alejadas del grupo, por lo que son posibles datos atípicos dentro del gráfico.

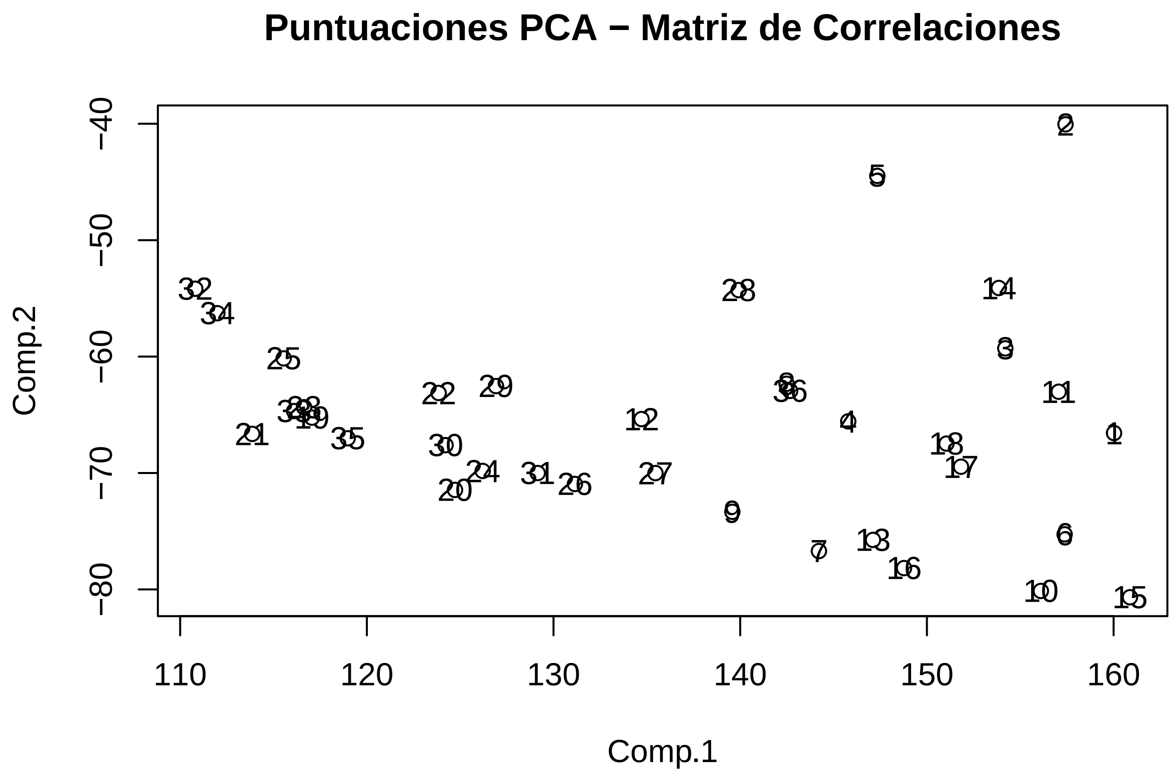
En el segundo gráfico, las flechas muestran cómo las variables (edad, peso, altura, muñeca y biceps) contribuyen a los componentes principales. Por ejemplo, El peso tiene una fuerte influencia en el comp1, pero baja en el componente 2. La edad tiene una influencia alta en el comp1 y 2. La altura tiene una influencia alta en el comp1 y una moderada en el componente 2.
Biceps y muñeca no parecen tener una influencia alta en ninguno de los dos componentes.

Matriz de correlaciones

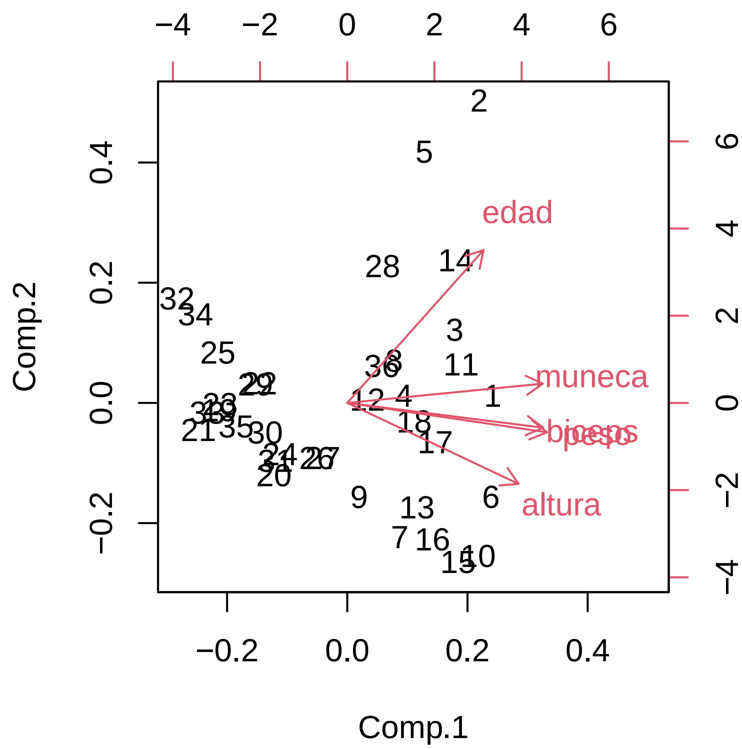
```
cpR <- princomp(datos, cor = TRUE)

# Calcula las puntuaciones (scores)
cpaR <- as.matrix(datos) %*% cpR$loadings

# Graficar las puntuaciones de las dos primeras componentes principales
plot(cpaR[,1:2], type = "p", main = "Puntuaciones PCA - Matriz de Correlaciones")
text(cpaR[,1], cpaR[,2], labels = 1:nrow(cpaR))
```




```r
# Graficar el biplot
biplot(cpR)
```

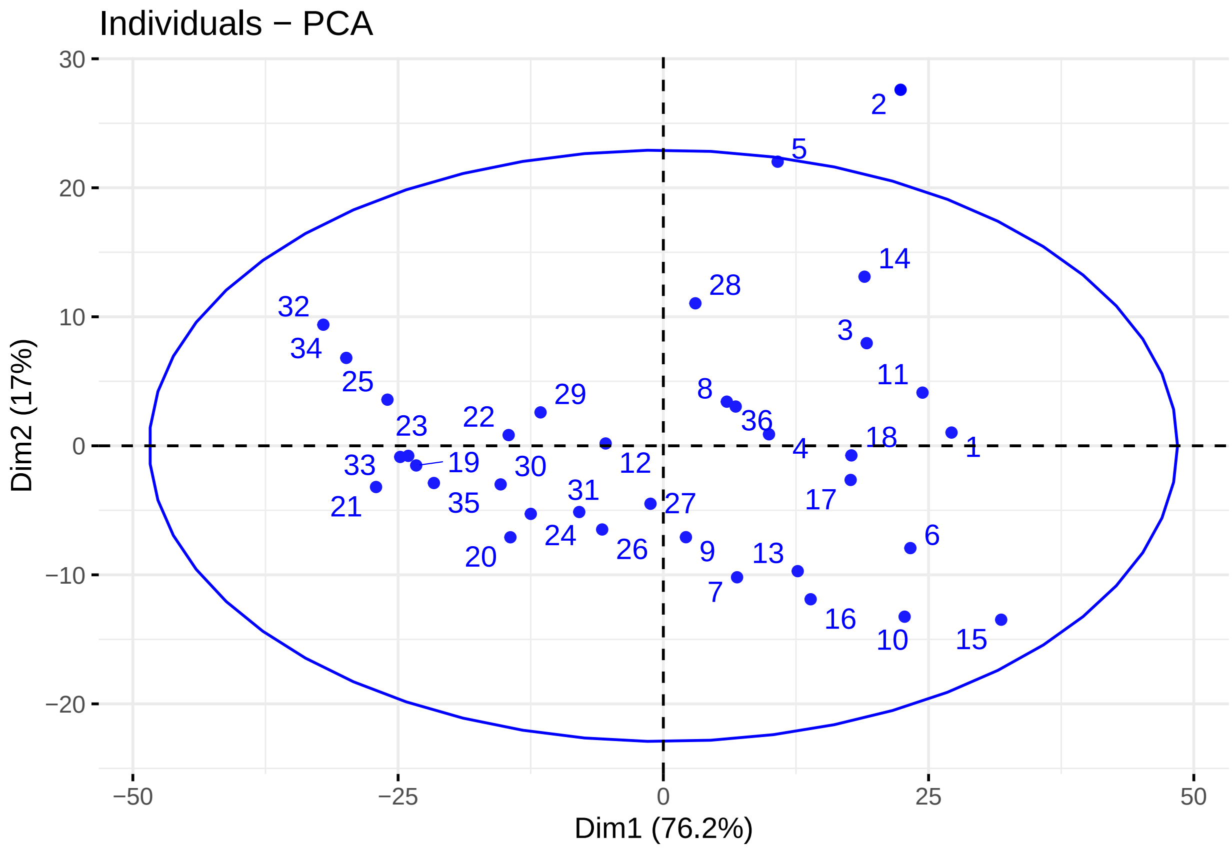


En el primer gŕafico, cada número que se ve en el gráfico es una observación con las variables estandarizadas, y los ejes representan los primeros dos componentes principales, que fueron los seleccionados. De igual manera, las observaciones 2 y la 5 están alejadas del grupo, por lo que son outliers dentro del gráfico.

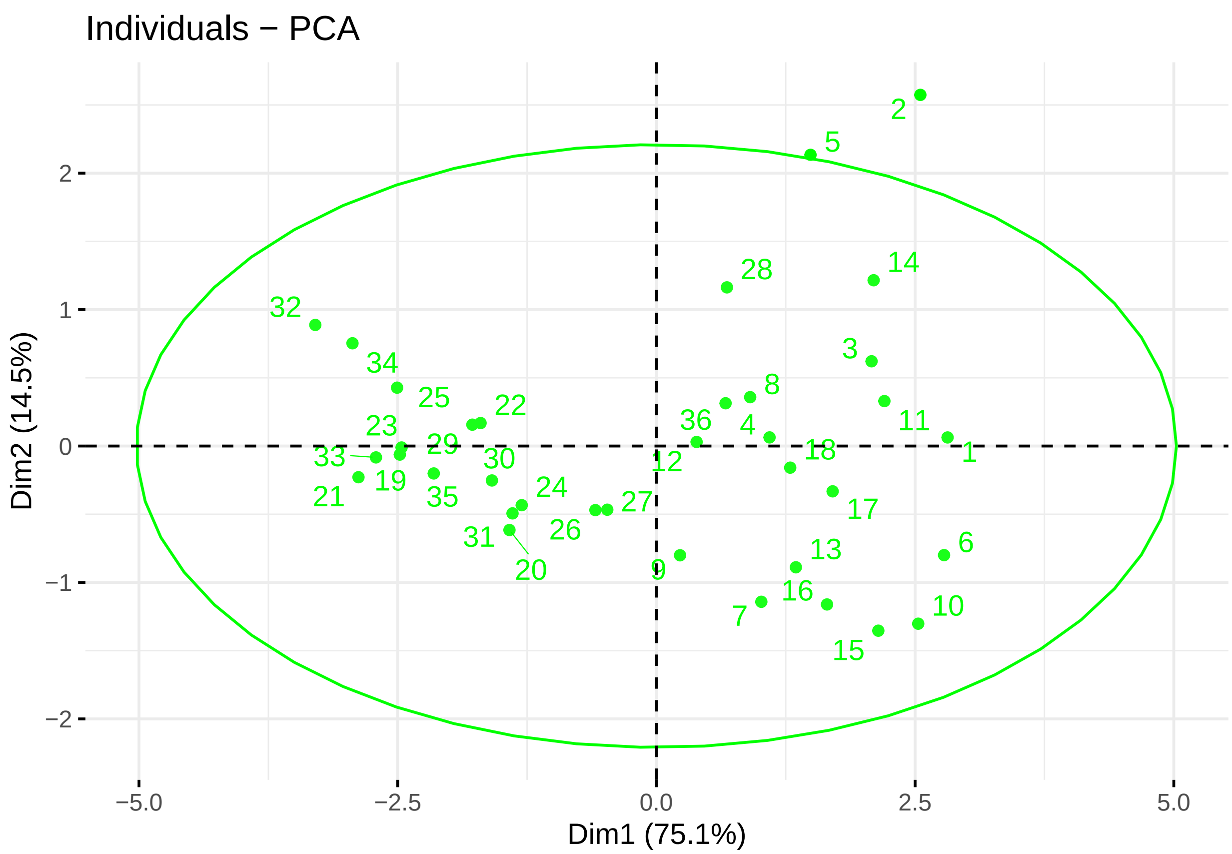
En el segundo gráfico, las flechas muestran cómo las variables (edad, peso, altura, muñeca y biceps) contribuyen a los componentes principales. Los resultados son distintos al gráfico de la matriz S: El peso, el bíceps, la muñeca y la altura están correlacionadas. El peso y el bíceps tienen una fuerte influencia en el comp1, pero baja en el componente 2. La muñeca tiene una influencia fuerte en el comp1, pero muy baja en el componente 2.
La edad tiene una influencia alta en el comp1 y 2. La altura tiene una influencia alta en el comp1 y una moderada en el componente 2.
Biceps y muñeca no parecen tener una influencia alta en ninguno de los dos componentes.

PARTE III: Explore los siguientes gráficos relativos a Componentes Principales. Interprete cada gráfico e identifica qué es lo que se está graficando en cada uno. Realiza el análisis con la matriz de varianzas y covarianzas y correlación.

```r
fviz_pca_ind(cpS, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```

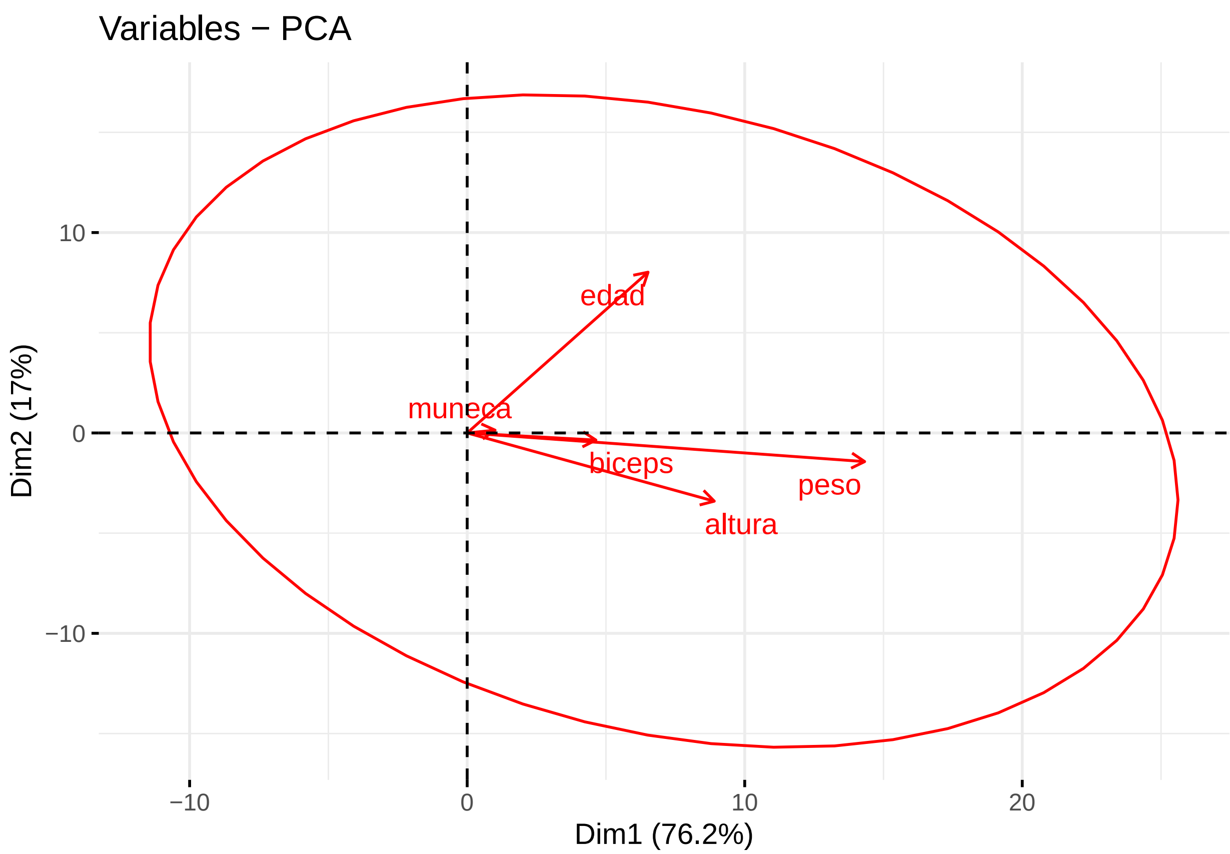



```
fviz_pca_ind(cpR, col.ind = "green", addEllipses = TRUE, repel = TRUE)
```

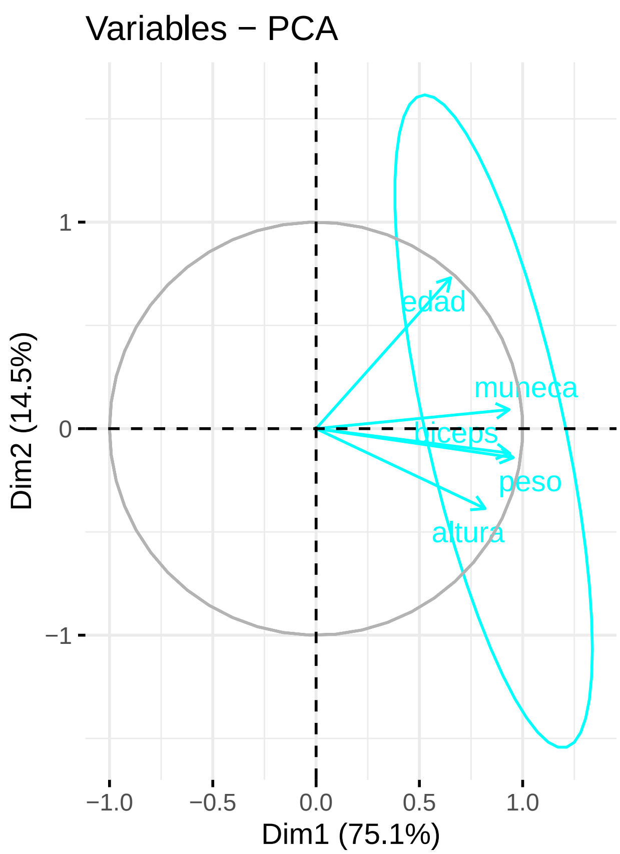


Estos gráficos muestran las observaciones proyectadas en el espacio de los componentes principales. Como podemos observar, la observación 2 y 5 quedan muy alejadas del resto del grupo, por lo que es posible que estos datos sean potencialmente atípicos.

```
fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```

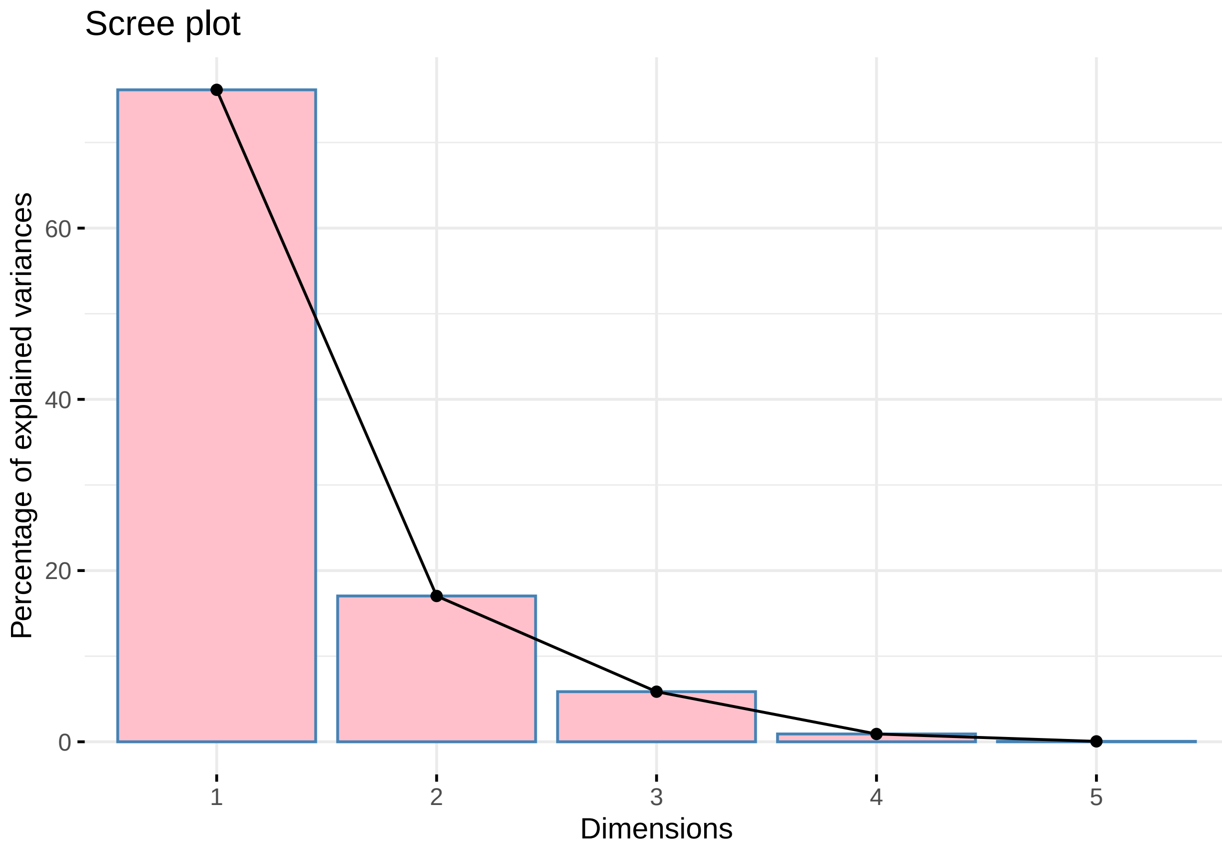



```r
fviz_pca_var(cpR, col.var = "cyan", addEllipses = TRUE, repel = TRUE)
```

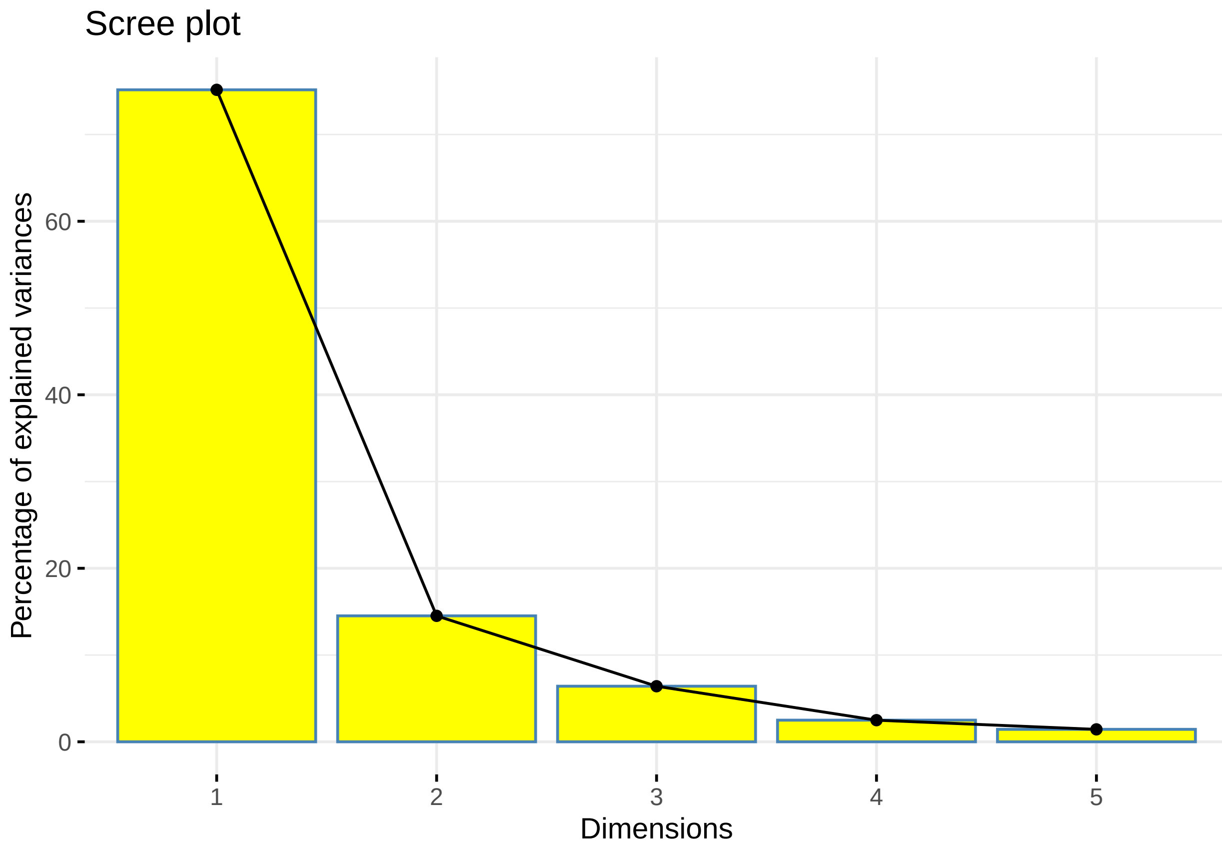


Estos gráficos muestran la influencia de las variables en los componentes principales. Bíceps altura y peso influyen mayormente al componente 1, mientras que edad influye en ambos componentes y muñeca no es una variable muy influyente, en el primer gráfico. Bíceps altura, peso y muñeca influyen mayormente al componente 1, mientras que edad influye en ambos componentes, en el segundo gráfico.

```
fviz_screeplot(cpS, barfill = 'pink')
```

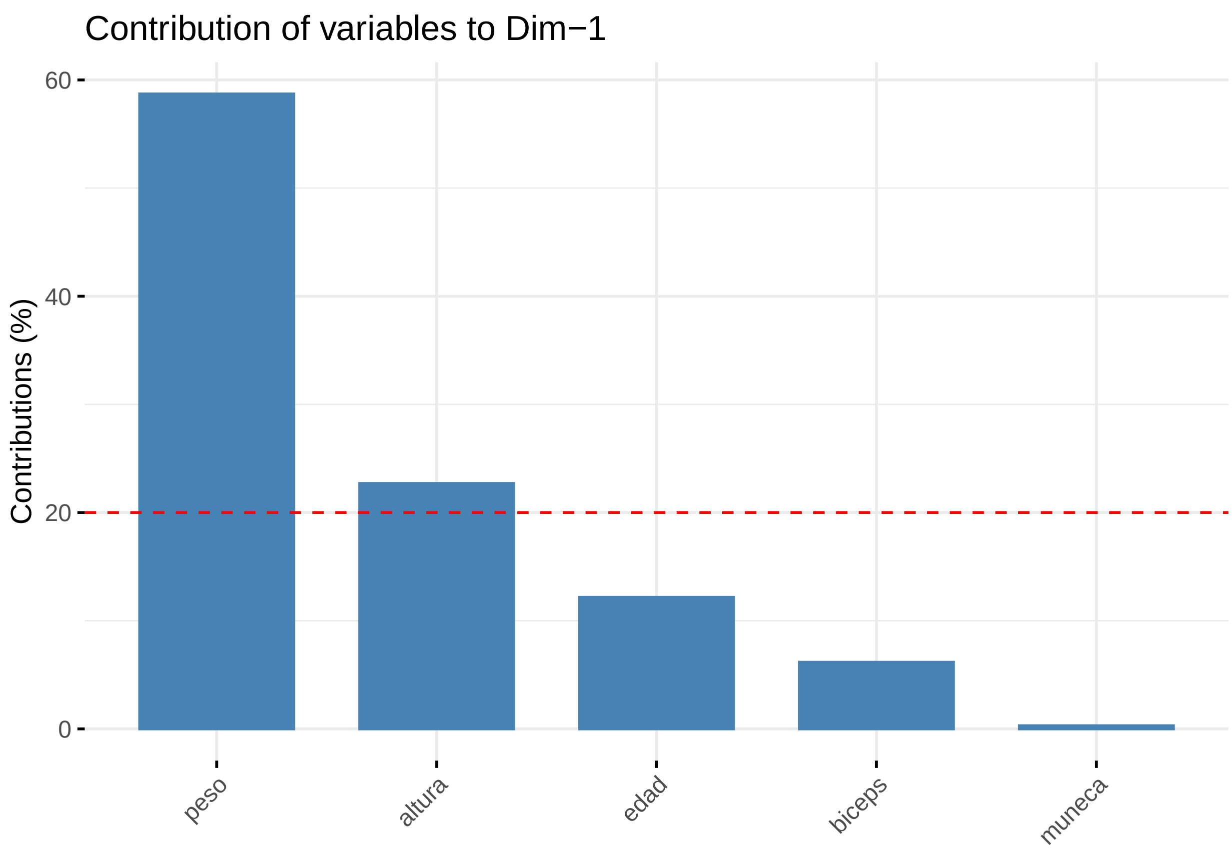



```r
fviz_screeplot(cpR, barfill = 'yellow')
```

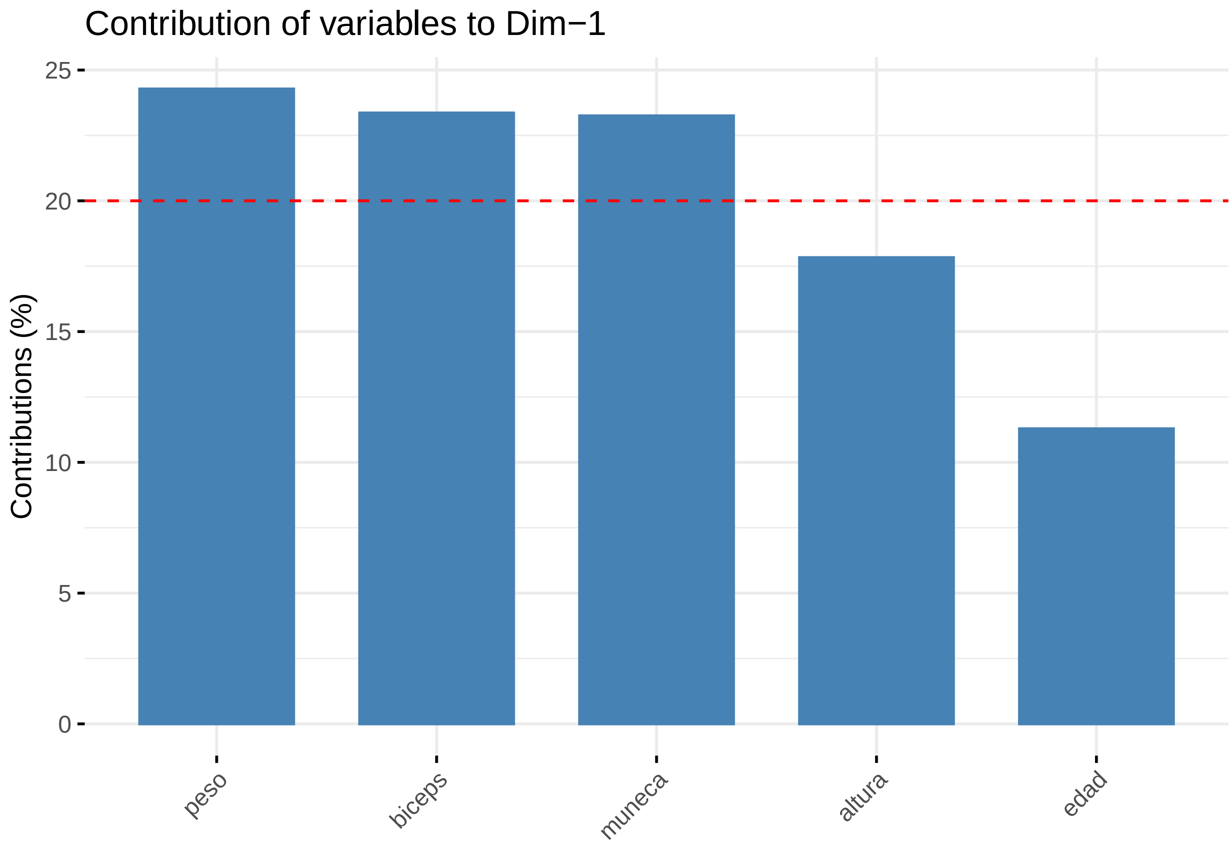


Estos gráficos muestran la proporción de la varianza explicada por cada componente principal. Como podemos observar, la mayor proporción de la varianza es explicada por el comp1 y 2, por lo que es adecuado seleccionar solo estos dos.

```
fviz_contrib(cpS, choice = c("var"))
```

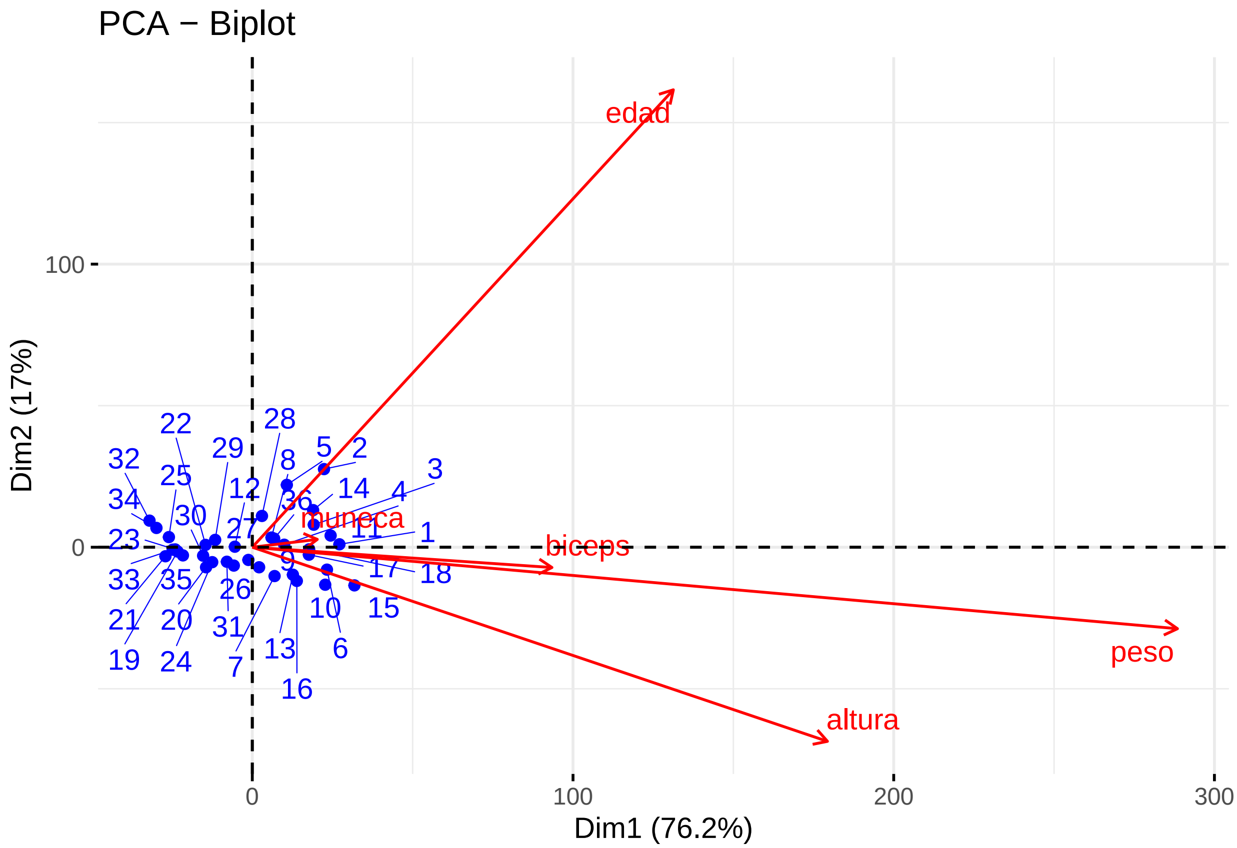



```
fviz_contrib(cpR, choice = c("var"))
```

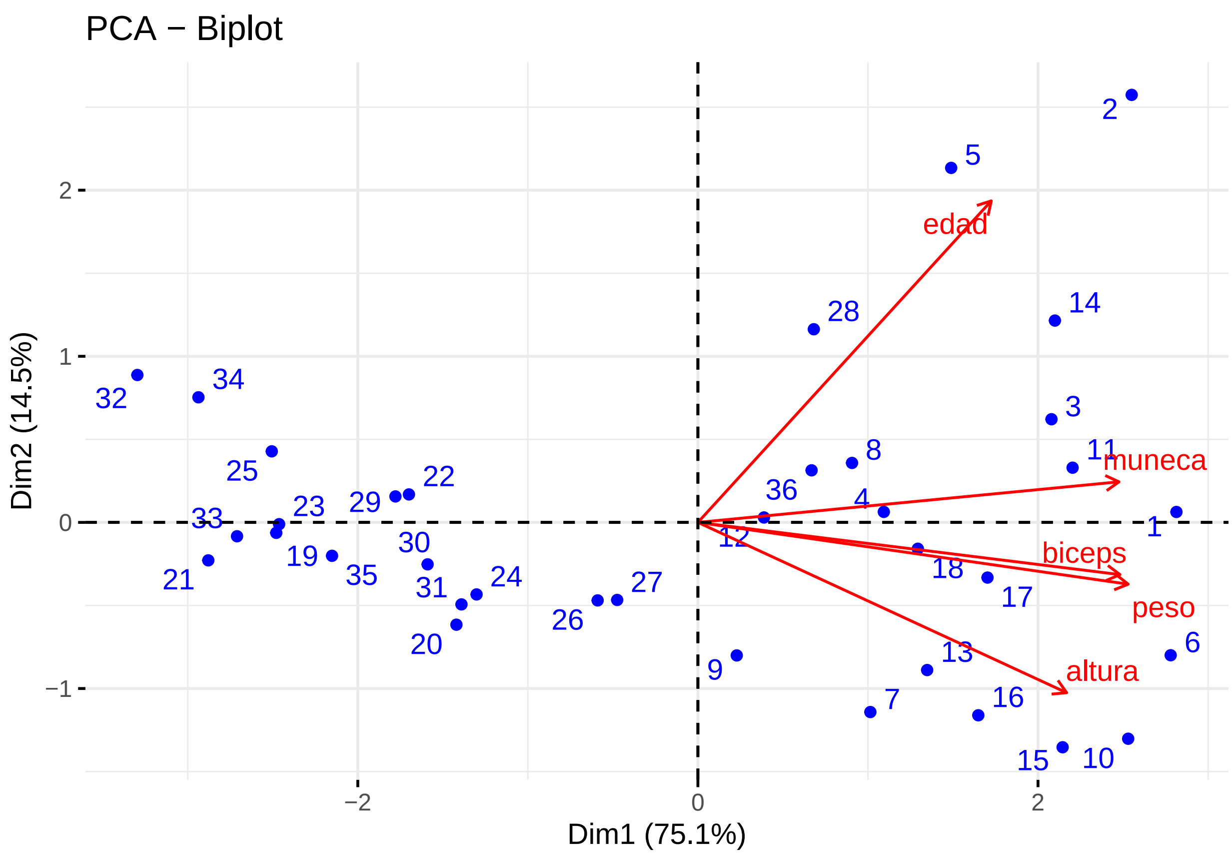


Estos gráficos muestran la proporción de la contribución de la variable al componente principal. Como podemos observar,las variables con mayor influencia son peso y altura en la matriz de varianza-covarianza, mientras que las variables con mayor influencia en la matriz de correlaciones (variables estandarizadas) son el peso, el bíceps y la muñeca.

```
fviz_pca_biplot(cpS, repel=TRUE, col.var="red", col.ind="blue")
```




```
fviz_pca_biplot(cpR, repel=TRUE, col.var="red", col.ind="blue")
```



En estos gráficos, los puntos azules representan las observaciones proyectadas en el espacio de los primeros dos componentes principales (Dim1 y Dim2) y las variables rojas representan las variables originales.

En el primer gráfico, las observaciones más alejadas del centro tienen caractericas diferentes a la mayoría (2 y 5). Como se había comentado anteriormente, puede que estos sean valores atípicos. De igual manera, como ya se había comentado: Bíceps altura y peso influyen mayormente al componente 1, mientras que edad influye en ambos componentes y muñeca no es una variable muy influyente, en el primer gráfico.

En el segundo gráfico, las observaciones están más alejadas entre sí, al estar estandarizadas, pero aún así la 2 y 5 siguen estando más alejadas. Como ya se había comentado, bíceps, altura, peso y muñeca influyen mayormente al componente 1, mientras que edad influye en ambos componentes, en e segundo gráfico.

PARTE IV

Compare los resultados obtenidos con la matriz de varianza-covarianza y con la correlación . ¿Qué concluye? ¿Cuál de los dos procedimientos aporta componentes con de mayor interés?

El análisis con la matriz de correlaciones suele ser más robusto y fácil de interpretar cuando las variables tienen diferentes escalas o unidades. Como es el caso, este procedimiento es mejor, pues los datos se encuentran en escalas similares y los pesos de distribuyen de una manera más lógica. Además, obtenemos mejores interpretaciones de los pesos y de la influencia significativa de las variables.

Indique cuál de los dos análisis resulta mejor para los datos. Comparar los resultados y argumentar cuál es mejor según los resultados obtenidos.

¿Qué variables son las que más contribuyen a la primera y segunda componentes principales del método seleccionado? (observa los coeficientes en valor absoluto de las combinaciones lineales, auxíliate también de los gráficos)

Escriba las combinaciones finales que se recomiendan para hacer el análisis de componentes principales.

Como podemos observar, nuestras ecuaciones para la matriz S son: $PC_1 = -0.349X_1 - 0.766X_2 - 0.476X_3 - 0.054X_4 - 0.248X_5$ $PC_2 = 0.908X_1 - 0.162X_2 - 0.385X_3 + 0.016X_4 - 0.040X_5$ Y para la matriz R son: $PC_1 = -0.336X_1 - 0.493X_2 - 0.422X_3 - 0.482X_4 - 0.483X_5$ $PC_2 = 0.858X_1 - 0.165X_2 - 0.454X_3 0.108X_4 - 0.139X_5$

Observamos que los pesos están distribuidos de una mejor manera en la matriz de correlaciones, pues los valores son estandarizados. Por otro lado, aunque la matriz S explica una mayor variabilidad que la matriz R, esto es debido a que los datos tienen diferentes escalas en la matriz S, por lo que una variable con una escala mayor explicará una variabilidad mayor. Sin embargo, en el gráfico de porcentajes de contribución, obtenemos una mejor interpretación de lo que realmente sucede en el PCA con la matriz de correlaciones. Por lo anterior, la matriz R da resultados más enriquecedores. Además, con estos dos componentes se logra una varianza $\approx 90\%$ de ambas matrices, lo cual es bastante bueno.

Las combinaciones lineales de los dos componentes seleccionados de la matriz R son: $PC_1 = -0.336X_1 - 0.493X_2 - 0.422X_3 - 0.482X_4 - 0.483X_5$ $PC_2 = 0.858X_1 - 0.165X_2 - 0.454X_3 0.108X_4 - 0.139X_5$

Observando las ecuaciones de la matriz R, observamos que los coeficientes que más influyen se encuentran en X_2, X_4 y X_5 para el PC_1 (peso, muñeca y bíceps, respectivamente), mientras que las que más influyen para el PC_2 se encuentran en X_1 y X_3 (edad y altura, respectivamente).

Interpreta los resultados en término de agrupación de variables

En nuestro caso, podemos ver nuestros componentes de la siguiente manera: El primer componente representa en su mayoría aquello asociado a la constitución física o masa muscular. Las variables de peso, muñeca y bíceps sugieren que este grupo está más influenciado por el volumen corporal y la fuerza.

Por otro lado, el segundo componente está más relacionado con características generales del crecimiento físico o envejecimiento La altura es una variable relacionada con la estatura física, que generalmente alcanza un máximo en la juventud, mientras que la edad puede estar relacionada con cambios físicos a lo largo del tiempo.

De esta manera, un componente explica en su mayoría la constitución física, masa muscular y fuerza, mientras que el segundo componente tiene que ver con los procesos biológicos de desarrollo y crecimiento.