Componentes Principales

Erika Martínez Meneses

2024-10-09

En la base de datos Corporal contiene las medidas corporarles de 36 estudiantes de la universidad. Haz una análisis de Componentes principales con la matriz de varianzas-covarianzas y la matriz de correlaciones. Compara los resultados y argumenta cuál es mejor según los resultados obtenidos.

Primero se realiza un análisis descriptivo para conocer las variables. Incluye las medidas que vienen en el summary() y la desviación estándar. Describe las correlaciones que se establecen entre las variables.

Análisis descriptivo

```
file.choose()

## [1] "C:\\Users\\erika\\Documents\\Agos-
Dic2024\\Estadística\\corporal.csv"

data <- read.csv("C:\\Users\\erika\\Documents\\Agos-
Dic2024\\Estadística\\corporal.csv")
head(data)

##   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
```

Convertir el sexo a numérico (0 = Mujer, 1 = Hombre). De esta manera facilitamos el análisis.

```
data$sexo_num <- ifelse(data$sexo == "Hombre", 1, 0)
```

Descripción estadística de las variables

```
summary(data)

##       edad            peso           altura          sexo          
##  Min.   :19.00   Min.   :42.00   Min.   :147.2   Length:36         
##  1st Qu.:24.75   1st Qu.:54.95   1st Qu.:164.8   Class :character  
##  Median :28.00   Median :71.50   Median :172.7   Mode  :character  
##  Mean   :31.44   Mean   :68.95   Mean   :171.6                     
##  3rd Qu.:37.00   3rd Qu.:82.40   3rd Qu.:179.4                     
```

```
##  Max.   :65.00   Max.   :98.20   Max.   :190.5
##      muneca          biceps         sexo_num
##  Min.   : 8.300   Min.   :23.50   Min.   :0.0
##  1st Qu.: 9.475   1st Qu.:25.98   1st Qu.:0.0
##  Median :10.650   Median :32.15   Median :0.5
##  Mean   :10.467   Mean   :31.17   Mean   :0.5
##  3rd Qu.:11.500   3rd Qu.:35.05   3rd Qu.:1.0
##  Max.   :12.400   Max.   :40.40   Max.   :1.0

sd(data$edad)

## [1] 10.55447

sd(data$peso)

## [1] 14.869

sd(data$altura)

## [1] 10.52017

sd(data$muneca)

## [1] 1.175463

sd(data$biceps)

## [1] 5.234392
```

El conjunto de datos contiene las variable edad, peso, altura, sexo, muneca y bíceps. Podemos destacar lo siguiente:

Edad: * Media: 31.44 años * Desviación estándar: 10.55 * Rango: 19 a 65 años

Peso: * Media: 68.95 kg * Desviación estándar: 14.87 * Rango: 42.0 a 98.2 kg

Altura: * Media: 171.56 cm * Desviación estándar: 10.52 * Rango: 147.2 a 190.5 cm

Muñeca: * Media: 10.47 cm * Desviación estándar: 1.18 * Rango: 8.3 a 12.4 cm

Bíceps: * Media: 31.17 cm * Desviación estándar: 5.23 * Rango: 23.5 a 40.4 cm

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.

1. 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:

```
# Matriz de varianza-covarianza
S <- cov(data[, c('edad', 'peso', 'altura', 'muneca', 'biceps',
```

```
'sexo_num')])
print(S)
```

```
##                edad       peso     altura     muneca    biceps
sexo_num
## edad     111.396825  80.881587  36.666032  7.6980952 26.720952
2.9142857
## peso      80.881587 221.087135 124.728698 14.8446667 70.738381
6.0071429
## altura    36.666032 124.728698 110.673968  8.1564762 39.021048
3.8685714
## muneca     7.698095  14.844667   8.156476  1.3817143  5.400571
0.4685714
## biceps    26.720952  70.738381  39.021048  5.4005714 27.398857
2.3257143
## sexo_num   2.914286   6.007143   3.868571  0.4685714  2.325714
0.2571429
```

- La covarianza entre edad y altura es 36.67, lo que sugiere una relación positiva moderada.
- La covarianza entre peso y biceps es 70.74, indicando una fuerte relación positiva.
- La covarianza entre muneca y sexo_num es 0.47, lo que sugiere una relación muy débil.

```
# Matriz de correlaciones
R <- cor(data[, c('edad', 'peso', 'altura', 'muneca', 'biceps',
'sexo_num')])
print(R)
```

```
##               edad      peso    altura    muneca    biceps  sexo_num
## edad     1.0000000 0.5153847 0.3302211 0.6204942 0.4836702 0.5445133
## peso     0.5153847 1.0000000 0.7973737 0.8493361 0.9088813 0.7967077
## altura   0.3302211 0.7973737 1.0000000 0.6595849 0.7086144 0.7251714
## muneca   0.6204942 0.8493361 0.6595849 1.0000000 0.8777369 0.7861030
## biceps   0.4836702 0.9088813 0.7086144 0.8777369 1.0000000 0.8761993
## sexo_num 0.5445133 0.7967077 0.7251714 0.7861030 0.8761993 1.0000000
```

- Peso y Bíceps tienen una correlación muy fuerte (0.9088), lo que indica que, a mayor peso, mayor es el perímetro del bíceps.
- Muñeca y Bíceps también están altamente correlacionados (0.8777), sugiriendo que las personas con muñecas más grandes tienden a tener bíceps más grandes.
- Sexo muestra correlaciones fuertes con variables como el Bíceps (0.87) y el Peso (0.79), lo que indica diferencias significativas entre hombres y mujeres en estas medidas.
- Edad tiene una correlación moderada con otras variables como el Peso (0.5153) y el Bíceps (0.48), pero es menos determinante en comparación con otras variables.

a) Calcule los valores y vectores propios de cada matriz.La función en R es: eigen().

```
eigen_S <- eigen(S)
print("Valores propios (covarianza):")

## [1] "Valores propios (covarianza):"

print(eigen_S$values)

## [1] 359.57899365  80.37590390  27.62303628   4.33876642   0.23508150
## [6]   0.04386111

print("Vectores propios (covarianza):")

## [1] "Vectores propios (covarianza):"

print(eigen_S$vectors)

##             [,1]         [,2]         [,3]         [,4]        [,5]
[,6]
## [1,] -0.34863029  0.907537510 -0.232465170  0.002404843  0.02585316 -
0.01087961
## [2,] -0.76597177 -0.161687536  0.521719981  0.338620226  0.01157913
0.01244668
## [3,] -0.47620546 -0.385190494 -0.789033198 -0.044934601  0.00254777 -
0.01605147
## [4,] -0.05385052  0.015540918  0.027858483 -0.125299618 -0.98892752
0.04903953
## [5,] -0.24813156 -0.040225629  0.224530217 -0.927533374  0.13212009 -
0.09287051
## [6,] -0.02243632  0.001212391 -0.002214725 -0.085400743  0.06130001
0.99420283
```

Los valores propios representan la varianza explicada por cada componente principal. El primer valor propio (359.57) explica la mayor parte de la varianza en los datos, seguido por el segundo (80.37), y así sucesivamente. Los valores más pequeños indican componentes que explican menos varianza.

Mientras que cada columna en la matriz de vectores propios corresponde a un componente principal. Estos vectores indican la dirección de los componentes principales en el espacio de las variables originales.

```
eigen_R <- eigen(R)
print("Valores propios (correlación):")

## [1] "Valores propios (correlación):"

print(eigen_R$values)

## [1] 4.55558176 0.72728889 0.32166657 0.22699752 0.12336499 0.04510028

print("Vectores propios (correlación):")
```

```
## [1] "Vectores propios (correlación):"
```

```r
print(eigen_R$vectors)
```

```
##            [,1]        [,2]        [,3]         [,4]        [,5]
[,6]
## [1,] -0.3010808  0.86487084 -0.35273030  0.008331536  0.13012980
0.1411627
## [2,] -0.4421676 -0.14913027  0.02810546 -0.394303566  0.61264031 -
0.5006542
## [3,] -0.3830675 -0.44217891 -0.74976289 -0.060939016 -0.21146895
0.2171637
## [4,] -0.4332077  0.12211455  0.32888456 -0.403526392 -0.70333466 -
0.1781597
## [5,] -0.4427932 -0.13093558  0.44071202  0.040313843  0.24869738
0.7273839
## [6,] -0.4282088 -0.04669377  0.10126639  0.822364253 -0.08055393 -
0.3484683
```

Similar a la matriz de covarianza, los valores propios de la matriz de correlación indican la proporción de la varianza total explicada por cada componente principal. El primer componente principal explica una gran parte de la varianza (4.55), mientras que los siguientes explican progresivamente menos. Los vectores propios asociados a estos valores propios muestran las combinaciones lineales de las variables originales que forman los componentes principales. Estos vectores son ortogonales entre sí, lo que significa que los componentes principales son no correlacionados.

En conclusión, los valores propios te indican cuánta varianza explica cada componente principal, y los vectores propios te muestran las direcciones de estos componentes en el espacio de las variables originales. Esto es fundamental para el Análisis de Componentes Principales (PCA) para reducir la dimensionalidad de los datos mientras se retiene la mayor parte de la variabilidad

b) Calcule la proporción de varianza explicada por cada componente en ambas matrices.

```r
total_var_S <- sum(diag(S)) # Suma de las varianzas de la diagonal de S
var_exp_S <- eigen_S$values / total_var_S * 100
print("Proporción de varianza explicada (covarianza):")
```

```
## [1] "Proporción de varianza explicada (covarianza):"
```

```r
cat(var_exp_S)
```

```
## 76.15043 17.02174 5.849913 0.9188493 0.04978477 0.009288757
```

Estos valores indican la proporción de la varianza total explicada por cada componente principal en la matriz de covarianza. El primer componente principal explica el 76.15% de la varianza total, lo que significa que captura la mayor parte de la variabilidad en los datos. Los siguientes componentes explican progresivamente menos varianza. Los últimos componentes explican una cantidad muy pequeña de la varianza, lo que sugiere que podrían ser menos importantes para describir la estructura de los datos.

```r
total_var_R <- sum(diag(R)) # Suma de las varianzas (valores propios)
var_exp_R <- eigen_R$values / total_var_R * 100
print("Proporción de varianza explicada (correlación):")

## [1] "Proporción de varianza explicada (correlación):"

cat(var_exp_R)

## 75.92636 12.12148 5.361109 3.783292 2.056083 0.7516713
```

De manera similar, estos valores muestran la proporción de la varianza total explicada por cada componente principal en la matriz de correlación. El primer componente principal explica el 75.93% de la varianza total, lo que nuevamente indica que captura la mayor parte de la variabilidad en los datos. Los siguientes componentes explican menos varianza, con el segundo componente explicando el 12.12% y así sucesivamente.

En ambos casos, el primer componente principal explica una gran proporción de la varianza total, lo que sugiere que una gran parte de la información en los datos puede ser capturada por este componente. Los componentes adicionales explican menos varianza, lo que es típico en análisis de componentes principales (PCA). Esto puede ser útil al reducir la dimensionalidad de los datos, para determinar el número de componentes para retener la mayor parte de la variabilidad.

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

```r
cum_var_exp_S <- cumsum(var_exp_S)
print("Varianza acumulada (covarianza):")

## [1] "Varianza acumulada (covarianza):"

cat(cum_var_exp_S)

## 76.15043 93.17216 99.02208 99.94093 99.99071 100
```

Estos valores muestran la varianza acumulada explicada por los componentes principales de la matriz de covarianza. El primer componente principal explica el 76.15% de la varianza total. Al sumar el segundo componente, se explica el 93.17% de la varianza total, y así sucesivamente. Para el cuarto componente, ya se ha explicado casi el 99.94% de la varianza total, lo que indica que los primeros pocos componentes capturan casi toda la variabilidad en los datos.

```r
cum_var_exp_R <- cumsum(var_exp_R)
print("Varianza acumulada (correlación):")

## [1] "Varianza acumulada (correlación):"

cat(cum_var_exp_R)
```

```
## 75.92636 88.04784 93.40895 97.19225 99.24833 100
```

De manera similar, estos valores muestran la varianza acumulada explicada por los componentes principales de la matriz de correlación. El primer componente principal explica el 75.92% de la varianza total. Al sumar el segundo componente, se explica el 88.047% de la varianza total. Para el cuarto componente, se ha explicado el 97.19% de la varianza total, lo que nuevamente sugiere que los primeros pocos componentes capturan la mayor parte de la variabilidad en los datos.

En ambos casos, los primeros dos o tres componentes principales explican la mayor parte de la varianza total, esta información es útil para saber a cuántos componentes reducir la dimensionalidad de los datos sin perder mucha información.

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

```
important_components_S <- which(cum_var_exp_S >= 0.8)[1]  # Primer componente que explica el 80% de la varianza
print(important_components_S)

## [1] 1

important_components_R <- which(cum_var_exp_R >= 0.8)[1]
print(important_components_R)

## [1] 1
```

Según los resultados anteriores, los componentes más importantes son aquellos que explican la mayor proporción de la varianza total. En este caso, los componentes más importantes son los primeros dos componentes principales, tanto para la matriz de covarianza como para la matriz de correlación.

En la matriz de covarianza el primer componente explica el 76.15% de la varianza total, el segundo componente explica el 17.02% adicional, acumulando un 93.17% de la varianza total. Mientras que en la matriz de correlación, el primer componente explica el 75.93% de la varianza total, el segundo explica el 12.12% adicional, acumulando un 88.05% de la varianza total.

En conclusión, los primeros dos componentes principales son los más importantes, ya que explican la mayor parte de la varianza en los datos, ya que explican mas del 80% de la varianza total. Estos componentes capturan la mayor parte de la información y son esenciales para entender la estructura subyacente de los datos.

e) 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.

```
X <- data[, c("edad", "peso", "altura", "muneca", "biceps", "sexo_num")]
```

```
CP1_S <- eigen_S$vectors[, 1] %*% t(X)
CP2_S <- eigen_S$vectors[, 2] %*% t(X)
print("Combinación lineal CP1 (covarianza):")

## [1] "Combinación lineal CP1 (covarianza):"

print(CP1_S)

##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[,7]
## [1,] -180.9496 -176.1517 -172.9567 -163.7501 -164.5666 -177.0716 -
160.7414
##           [,8]      [,9]   [,10]     [,11]     [,12]     [,13]
[,14]
## [1,] -159.7733 -155.9215 -176.53 -178.2153 -148.3569 -166.4555 -
172.7513
##          [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
[,21]
## [1,] -185.6281 -167.6743 -171.4429 -171.5124 -130.4834 -139.3605 -
126.699
##          [,22]     [,23]     [,24]     [,25]     [,26]     [,27]
[,28]
## [1,] -139.1955 -129.7349 -141.2797 -127.7751 -148.0064 -152.5595 -
156.7909
##       [,29]     [,30]     [,31]     [,32]     [,33]     [,34]
[,35]
## [1,] -142.2 -138.4398 -145.8469 -121.7331 -128.9766 -123.8947 -
132.1501
##          [,36]
## [1,] -160.5888

print("Combinación lineal CP2 (covarianza):")

## [1] "Combinación lineal CP2 (covarianza):"

print(CP2_S)

##           [,1]     [,2]      [,3]      [,4]      [,5]     [,6]
[,7]
## [1,] -48.75629 -22.1884 -41.82723 -48.89116 -27.76324 -57.7108 -
59.97163
##           [,8]      [,9]     [,10]     [,11]     [,12]     [,13]
[,14]
## [1,] -46.36195 -56.86553 -63.02869 -45.66127 -49.60221 -59.49846 -
36.67576
##          [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
[,21]
## [1,] -63.26273 -61.67667 -52.42893 -50.52664 -51.30447 -56.87264 -
52.9768
##          [,22]     [,23]     [,24]     [,25]     [,26]     [,27]
[,28]
## [1,] -48.95146 -50.56149 -55.05916 -46.20751 -56.26989 -54.27448 -
```

```
38.73778
##          [,29]     [,30]     [,31]     [,32]     [,33]     [,34]
[,35]
## [1,] -47.19326 -52.77507 -54.91534 -40.39694 -50.64515 -42.96964 -
52.66681
##          [,36]
## [1,] -46.74102
```

Los valores de CP1_S representan la proyección de los datos originales en el primer componente principal, utilizando la matriz de covarianza. Estos valores indican la variación máxima en los datos a lo largo de este componente. Los valores negativos y positivos reflejan cómo los datos se distribuyen en relación con este componente principal.

Los valores de CP2_S representan la proyección de los datos originales en el segundo componente principal, también utilizando la matriz de covarianza. Este componente captura la segunda mayor variación en los datos, ortogonal al primer componente. Nuevamente, los valores reflejan la distribución de los datos en relación con este componente.

```r
CP1_R <- eigen_R$vectors[, 1] %*% t(X)
CP2_R <- eigen_R$vectors[, 2] %*% t(X)
print("Combinación lineal CP1 (correlación):")

## [1] "Combinación lineal CP1 (correlación):"

print(CP1_R)

##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[,7]
## [1,] -145.1297 -142.7219 -139.8781 -132.2887 -133.6341 -142.7821 -
130.9399
##           [,8]     [,9]    [,10]     [,11]     [,12]     [,13]
[,14]
## [1,] -129.2985 -126.731 -141.638 -142.4364 -122.3312 -133.5012 -
139.533
##          [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
[,21]
## [1,] -145.9004 -135.0235 -137.7477 -137.0273 -105.9355 -112.8717 -
103.057
##          [,22]     [,23]     [,24]     [,25]     [,26]   [,27]
[,28]
## [1,] -112.0389 -105.5581 -114.2003 -104.5648 -118.6412 -122.55 -
126.5214
##          [,29]     [,30]     [,31]     [,32]    [,33]    [,34]
[,35]
## [1,] -114.8282 -112.4078 -116.8487 -100.2876 -105.084 -101.346 -
107.6759
##          [,36]
## [1,] -129.0353
```

```r
print("Combinación lineal CP2 (correlación):")
```

```
## [1] "Combinación lineal CP2 (correlación):"
```

```r
print(CP2_R)
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[,7]
## [1,] -62.20365 -35.81701 -55.11884 -61.59576 -40.50536 -70.97268 -
72.78653
##           [,8]     [,9]     [,10]     [,11]     [,12]     [,13]
[,14]
## [1,] -58.46689 -69.5342 -75.85916 -58.73825 -61.72391 -71.71715 -
49.93959
##          [,15]     [,16]     [,17]     [,18]     [,19]     [,20]
[,21]
## [1,] -76.23018 -74.10085 -65.31929 -63.34857 -62.00548 -67.99731 -
63.48783
##          [,22]     [,23]     [,24]     [,25]     [,26]     [,27]
[,28]
## [1,] -59.70163 -61.13349 -66.32572 -56.97029 -67.30053 -66.25785 -
50.45194
##        [,29]     [,30]     [,31]     [,32]     [,33]     [,34]
[,35]
## [1,] -59.029 -64.17072 -66.41938 -51.14136 -61.47975 -53.19895 -
63.72436
##          [,36]
## [1,] -59.01317
```

Los valores de CP1_R son similares a CP1_S, pero se obtienen utilizando la matriz de correlación en lugar de la matriz de covarianza. Esto normaliza las variables, eliminando las diferencias de escala entre ellas. Los valores indican la variación máxima en los datos normalizados a lo largo del primer componente principal.

Finalmente, los valores de CP2_R representan la proyección de los datos normalizados en el segundo componente principal, utilizando la matriz de correlación. Este componente captura la segunda mayor variación en los datos normalizados, ortogonal al primer componente.

En conclusión, los componentes principales obtenidos a partir de la matriz de covarianza (CP1_S y CP2_S) y la matriz de correlación (CP1_R y CP2_R) proporcionan diferentes perspectivas sobre la variación en los datos, dependiendo de si las variables originales están normalizadas o no.

PARTE II

1. Obtenga las gráficas respectivas con S (matriz de varianzas-covarianzas) y con R (matriz de correlaciones) de las dos primeras componentes.

```r
library(stats)
# Análisis de Componentes Principales usando la matriz de varianzas-
```

```
covarianzas
cpS <- princomp(X, cor = FALSE)
summary(cpS)

## Importance of components:
##                           Comp.1    Comp.2     Comp.3      Comp.4
Comp.5
## Standard deviation     18.6973444 8.8398665 5.18225141 2.053836686
0.4780705544
## Proportion of Variance  0.7615043 0.1702174 0.05849913 0.009188493
0.0004978477
## Cumulative Proportion   0.7615043 0.9317216 0.99022077 0.999409265
0.9999071124
##                              Comp.6
## Standard deviation     2.065012e-01
## Proportion of Variance 9.288757e-05
## Cumulative Proportion  1.000000e+00

# Análisis de Componentes Principales usando la matriz de correlaciones
(cor=TRUE)
cpR <- princomp(X, cor = TRUE)
summary(cpR)

## Importance of components:
##                          Comp.1    Comp.2     Comp.3     Comp.4
Comp.5
## Standard deviation     2.1343809 0.8528123 0.56715656 0.47644257
0.35123353
## Proportion of Variance 0.7592636 0.1212148 0.05361109 0.03783292
0.02056083
## Cumulative Proportion  0.7592636 0.8804784 0.93408954 0.97192246
0.99248329
##                             Comp.6
## Standard deviation     0.212368258
## Proportion of Variance 0.007516713
## Cumulative Proportion  1.000000000
```

- Componentes Principales Significativos: En ambos, el primer componente principal (Comp.1) explica la mayor parte de la varianza (más del 75%). Esto indica que la mayor parte de la información de los datos originales se puede representar en este primer componente.
- Proporción Acumulada: Los primeros dos componentes explican más del 88% de la varianza total en ambos casos, lo que sugiere que estos dos componentes son suficientes para representar la estructura de los datos con una pérdida mínima de información.
- Comparación entre varianzas-covarianzas y correlaciones: El análisis basado en la matriz de correlaciones tiende a estandarizar las variables, lo que puede ser útil si las variables originales tienen diferentes unidades o escalas, como es nuestro caso.

a) Calcule las puntuaciones (scores) de las observaciones para los componentes obtenidos con la matriz de varianzas-covarianzas

```
scores_S <- as.matrix(X) %*% cpS$loadings
scores_S

##         Comp.1    Comp.2    Comp.3     Comp.4    Comp.5   Comp.6
##  [1,] 180.9496 -48.75629 104.41225 -13.602502 -4.672191 4.131184
##  [2,] 176.1517 -22.18840 102.47377 -14.625352 -4.131529 4.172187
##  [3,] 172.9567 -41.82723  97.83352 -17.153086 -3.362028 4.330374
##  [4,] 163.7501 -48.89116 104.92563 -14.445612 -4.492267 3.827535
##  [5,] 164.5666 -27.76324  98.65454 -14.386798 -4.554350 3.848876
##  [6,] 177.0716 -57.71080 102.20620 -16.630019 -4.787995 4.280755
##  [7,] 160.7414 -59.97163 100.99182 -20.192692 -3.407848 4.297892
##  [8,] 159.7733 -46.36195  92.39987 -15.575101 -4.274023 3.823449
##  [9,] 155.9215 -56.86553 109.03139 -16.977902 -3.831511 3.908549
## [10,] 176.5300 -63.02869  93.60972 -16.441272 -4.277211 4.258177
## [11,] 178.2153 -45.66127  96.31866 -13.129571 -3.691505 4.075066
## [12,] 148.3569 -49.60221 100.76716 -19.727672 -4.862254 3.900934
## [13,] 166.4555 -59.49846  94.96633 -14.098475 -4.577969 3.818365
## [14,] 172.7513 -36.67576 103.94298 -14.240351 -4.425235 4.029874
## [15,] 185.6281 -63.26273  97.94235  -8.939658 -3.524429 3.874897
## [16,] 167.6743 -61.67667  93.00759 -16.782829 -4.164257 4.094940
## [17,] 171.4429 -52.42893  98.51195 -15.082778 -3.779017 4.081247
## [18,] 171.5124 -50.52664  99.53247 -12.776708 -3.451042 3.909541
## [19,] 130.4834 -51.30447  99.67325 -13.409565 -4.310662 3.990189
## [20,] 139.3605 -56.87264  99.51315 -14.525750 -4.490460 4.268168
## [21,] 126.6990 -52.97680  98.96234 -13.837229 -4.163300 3.946986
## [22,] 139.1955 -48.95146  90.26998 -12.822818 -3.919635 4.142429
## [23,] 129.7349 -50.56149  93.55557 -14.891139 -3.884710 4.114106
## [24,] 141.2797 -55.05916 102.47223 -13.437608 -4.677180 4.215300
## [25,] 127.7751 -46.20751 101.07236 -14.499617 -4.483125 4.021404
## [26,] 148.0064 -56.26989  92.82332 -12.195226 -4.923695 4.222213
## [27,] 152.5595 -54.27448 103.90139 -13.356714 -3.913715 4.514117
## [28,] 156.7909 -38.73778  88.60379 -14.244155 -4.620401 4.627134
## [29,] 142.2000 -47.19326 108.94024 -13.062832 -3.322239 4.311418
## [30,] 138.4398 -52.77507 106.40688 -14.040740 -4.527415 4.223804
## [31,] 145.8469 -54.91534 103.55634 -11.721178 -3.865689 4.223554
## [32,] 121.7331 -40.39694 100.24699 -15.901966 -3.410323 4.076908
## [33,] 128.9766 -50.64515  99.28599 -15.028049 -3.567133 4.140998
## [34,] 123.8947 -42.96964  89.88715 -14.934529 -3.867224 3.988199
## [35,] 132.1501 -52.66681 106.84956 -13.842484 -4.796031 4.049389
## [36,] 160.5888 -46.74102 101.22524 -13.119406 -4.532461 4.640007
```

Los scores representan las coordenadas de las observaciones originales en el espacio de los componentes principales. * Comp.1 explica la mayor parte de la varianza en los datos. Podemos observar que los scores aquí son bastante altos para todas las observaciones, lo que indica que la mayoría de la información de los datos originales está capturada en este componente. * Comp.2 también captura una cantidad significativa de varianza, aunque mucho menor que Comp.1. Los scores en Comp.2

varían más ampliamente, lo que sugiere que este componente está capturando variaciones que no son tan dominantes como las capturadas por Comp.1. * A partir de Comp.3 los scores en estos componentes son menores en magnitud y varían menos, lo que es consistente con la menor proporción de varianza explicada por estos componentes. Estos componentes capturan variaciones más sutiles en los datos. * En el caso particular de la observación 1 tiene un score alto en Comp.1 (180.9496) y un score negativo en Comp.2 (-48.75629), lo que sugiere que esta observación está fuertemente influenciada por el primer componente principal y tiene una contribución negativa en el segundo componente.

b) 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.

```
scores_R <- as.matrix(scale(X)) %*% cpR$loadings # Se estandarizan con scale()
scores_R

##          Comp.1       Comp.2       Comp.3       Comp.4       Comp.5
##  [1,]  2.9269993  0.089831121  0.548559736 -0.321166493 -0.328551806
##  [2,]  2.6863354  2.553481542  0.423145512  0.033378152  0.004993191
##  [3,]  2.2764778  0.615803589 -0.142280108  0.243705525  0.447888670
##  [4,]  1.3949820  0.045384064  0.408058666  0.397568704 -0.384919785
##  [5,]  1.7412251  2.093975591 -0.123255630  0.271500671 -0.212853267
##  [6,]  2.9022126 -0.762324899 -0.066279935 -0.339798046 -0.500915340
##  [7,]  1.3356628 -1.148846968 -0.337942646  0.651658799  0.099479272
##  [8,]  1.2294945  0.328565605 -0.832728723  0.385191556  0.030154113
##  [9,]  0.6305783 -0.829750020  0.602470786  0.921537855 -0.273821427
## [10,]  2.6824623 -1.265002245 -0.714071175 -0.297003495  0.048910060
## [11,]  2.3839814  0.334850699  0.103737917 -0.002992361  0.496284748
## [12,]  0.7712438 -0.009230821 -0.776668096  0.662623271 -0.818270448
## [13,]  1.6251136 -0.886063301 -0.491458655  0.082166365 -0.132313956
## [14,]  2.2876054  1.205174669  0.460592874  0.077607754 -0.235059465
## [15,]  2.3339212 -1.320651791  0.775551853 -0.127726372  0.760028739
## [16,]  1.8973098 -1.149041114 -0.847209495  0.069647557  0.062252290
## [17,]  1.9418463 -0.330236770 -0.002569089  0.208088701  0.245351583
## [18,]  1.5747505 -0.168646674  0.434214841  0.397158123  0.514989071
## [19,] -2.6324301 -0.078392461  0.011530648  0.083788234 -0.112179436
## [20,] -1.6859701 -0.596127660 -0.110839695 -0.363868638 -0.211127616
## [21,] -2.9872214 -0.252360021 -0.096154233  0.255588761 -0.048631021
## [22,] -1.9372192  0.168245447 -0.604548507 -0.266316665  0.475233892
## [23,] -2.6150221 -0.028851219 -0.601451890  0.115595442  0.249809555
## [24,] -1.5824760 -0.412468358  0.249067965 -0.420323330 -0.345993276
## [25,] -2.6569516  0.403546625 -0.036012828  0.134985603 -0.314631188
## [26,] -0.9491069 -0.429424247 -0.513024321 -0.903851435 -0.123490002
## [27,] -0.8443585 -0.424890494  0.688762305 -0.571222188  0.150855012
## [28,]  0.1808004  1.208228264 -0.929325398 -1.239990869  0.167895745
## [29,] -2.0074767  0.155141183  1.270043647  0.140584878  0.315062769
## [30,] -1.8390071 -0.241917447  0.565675544 -0.202153588 -0.409957237
## [31,] -1.6613346 -0.473676683  0.790322982 -0.258345801  0.214750586
```

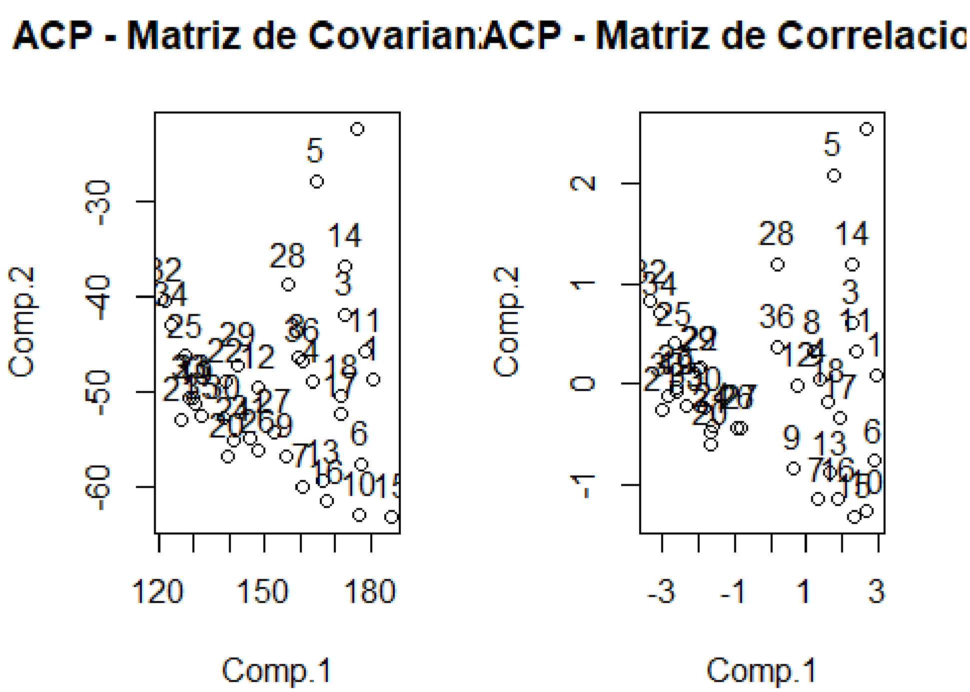
```
## [32,] -3.3593105  0.833349415 -0.015211415  0.695725395  0.259630220
## [33,] -2.8348597 -0.106757844 -0.012547190  0.349155974  0.266791922
## [34,] -3.0400562  0.710631184 -0.969748237  0.296737126  0.332502156
## [35,] -2.3406428 -0.205032490  0.503810561 -0.035804008 -0.616793576
## [36,]  0.1704407  0.373484528  0.387781430 -1.123431158 -0.073354751
##            Comp.6
##  [1,] -0.08620989
##  [2,]  0.08397830
##  [3,]  0.39826871
##  [4,] -0.21281260
##  [5,] -0.24473034
##  [6,]  0.17206819
##  [7,]  0.57876982
##  [8,] -0.21869641
##  [9,]  0.11948623
## [10,]  0.13153218
## [11,] -0.09685422
## [12,]  0.09801198
## [13,] -0.34652912
## [14,] -0.07633551
## [15,] -0.48710256
## [16,]  0.05205444
## [17,]  0.03599520
## [18,] -0.17982603
## [19,] -0.17442814
## [20,]  0.05997818
## [21,] -0.16505330
## [22,] -0.13616106
## [23,]  0.00984404
## [24,] -0.04409899
## [25,] -0.08770040
## [26,] -0.23458818
## [27,]  0.26584394
## [28,]  0.19478809
## [29,]  0.23178502
## [30,]  0.05546051
## [31,] -0.03745970
## [32,]  0.17614425
## [33,]  0.12626483
## [34,] -0.09614342
## [35,] -0.10754097
## [36,]  0.24199690
```

Similar que con los scores anteriores los Comp.1 y Comp.2 explican la mayor parte de la varianza. Los scores en Comp.1 son altos para muchas observaciones, indicando que este componente captura la mayor parte de la información de los datos estandarizados. Los scores en Comp.2 varían significativamente, lo que sugiere que este componente captura variaciones importantes que no son explicadas por Comp.1. y los scores de Comp.3 en adelante son menores y varían menos, lo que es consistente con la menor proporción de varianza explicada por estos componentes.

En comparación con los scores anteriores, estos están estandarizados. Los scores estandarizados son menores en magnitud comparados con los scores sin estandarizar, debido a la estandarización de las variables originales. Y los patrones de variación en los scores son similares, pero los scores estandarizados permiten una comparación más equitativa entre variables con diferentes unidades o escalas.

Graficos

```r
# Graficar las puntuaciones de las dos primeras componentes para ambas matrices
par(mfrow = c(1, 2))
# Covarianzas
plot(scores_S[,1:2], type = "p", main = "ACP - Matriz de Covarianzas")
text(scores_S[,1], scores_S[,2], labels = 1:nrow(scores_S), pos = 3)
# Correlaciones
plot(scores_R[,1:2], type = "p", main = "ACP - Matriz de Correlaciones")
text(scores_R[,1], scores_R[,2], labels = 1:nrow(scores_R), pos = 3)
```



Ambas gráficas muestran los resultados del Análisis de Componentes Principales (ACP) una con la matriz de covarianzas y la otra con la matriz de correlaciones.

ACP - Matriz de Covarianzas

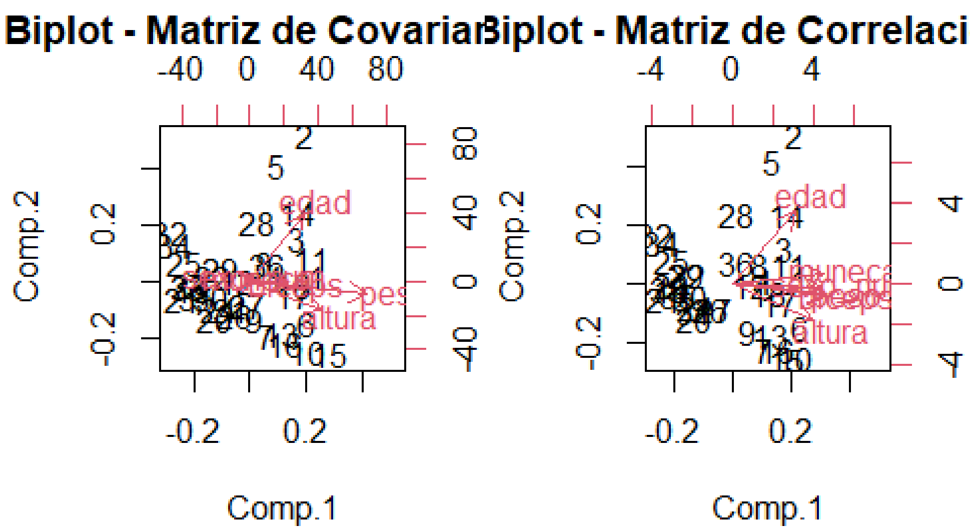
La mayoría de los puntos están agrupados en el centro, con algunos outliers. Esto sugiere que las primeras dos componentes principales capturan la mayor parte de la varianza en los datos originales.

ACP - Matriz de Correlaciones

Los puntos están más dispersos en comparación con la gráfica de la matriz de covarianzas. Esto indica que la estandarización de las variables (usando la matriz de correlaciones) permite una mejor comparación entre variables con diferentes unidades o escalas.

Al compararlas lo primero que podemos notar es que el rango de los ejes es muy diferente, es mucho menor en la matriz de correlaciones, esto se debe a la estandarización. La matriz de covarianza captura la varianza absoluta de los datos, es útil cuando las unidades de las variables son comparables. Mientras que la matriz de correlaciones estandariza las variables, permitiendo una comparación equitativa y es más útil cuando las variables tienen diferentes unidades o escalas.

```r
par(mfrow = c(1, 2))
biplot(cpS, main = "Biplot - Matriz de Covarianzas")
biplot(cpR, main = "Biplot - Matriz de Correlaciones")
```



Biplot - Matriz de Covarianzas

Esta gráfica muestra la relación entre diferentes variables en términos de covarianza. Las covarianzas indican cómo dos variables varían juntas. En el gráfico, los vectores representan las variables y su dirección y longitud indican la magnitud y dirección de la relación lineal entre ellas. Por ejemplo, los vectores de “biceps” y “peso” están cerca y apuntan en la misma dirección, esto sugiere que a medida que aumenta el peso, también lo hace el bicep.

Biplot - Matriz de Correlaciones

Esta gráfica es similar a la anterior, pero utiliza la matriz de correlaciones en lugar de la de covarianzas. Las correlaciones están estandarizadas, lo que facilita la comparación entre diferentes relaciones. En este gráfico, puedes ver cómo las variables están correlacionadas entre sí de manera más clara y comparativa.

2. Interprete los gráficos en términos de:

a) Las relaciones que se establecen entre las variables y los componentes principales

En un gráfico de componentes principales, las variables están representadas por vectores (flechas) y los componentes principales por los ejes del gráfico (CP1 y CP2).

Magnitud y dirección de los vectores * Las variables que tienen vectores largos y que están alineados con uno de los componentes principales tienen una mayor contribución a ese componente. Es decir, las variables que forman un ángulo pequeño con un eje de componente principal tienen una relación fuerte con ese componente. * Las variables cuyos vectores son cortos tienen poca influencia en los componentes principales.

Ángulos entre los vectores * Si los vectores entre dos variables forman un ángulo pequeño o están alineados, entonces las variables están positivamente correlacionadas. * Si los vectores forman un ángulo de 90°, entonces las variables son independientes o no están correlacionadas. * Si los vectores forman un ángulo cercano a 180°, las variables están negativamente correlacionadas

b) La relación entre las puntuaciones de las observaciones y los valores de las variables

En un gráfico de puntuaciones de componentes principales, las observaciones están representadas por puntos y las variables por vectores:

- Las observaciones que están más alejadas del origen en una dirección particular están más influenciadas por las variables que apuntan en esa dirección.
- Las observaciones que están cerca unas de otras tienen valores similares para las variables asociadas con esos componentes principales.
- Si una observación está cerca de un vector, significa que esa observación tiene un valor alto en esa variable. Observaciones en el lado opuesto del vector tendrán valores bajos en esa variable.

c) Detecte posibles datos atípicos

Los datos atípicos se detectan observando puntos que están muy alejados del grupo principal de observaciones en el gráfico.

- Observaciones aisladas: Si una observación está separada del grupo central de puntos, puede ser un dato atípico.

- Puntuaciones extremas: Los puntos que tienen puntuaciones extremadamente altas o bajas en cualquiera de los componentes principales (muy lejos del origen) pueden ser considerados como atípicos.

3. 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), cpa$Sloading, cpaS$scores. ¿Cómo se interpreta el resultado?

```
# Resumen de los componentes principales (varianza explicada)
summary(cpS)

## Importance of components:
##                            Comp.1    Comp.2     Comp.3      Comp.4
Comp.5
## Standard deviation     18.6973444 8.8398665 5.18225141 2.053836686
0.4780705544
## Proportion of Variance  0.7615043 0.1702174 0.05849913 0.009188493
0.0004978477
## Cumulative Proportion   0.7615043 0.9317216 0.99022077 0.999409265
0.9999071124
##                              Comp.6
## Standard deviation     2.065012e-01
## Proportion of Variance 9.288757e-05
## Cumulative Proportion  1.000000e+00

summary(cpR)

## Importance of components:
##                           Comp.1    Comp.2     Comp.3     Comp.4
Comp.5
## Standard deviation     2.1343809 0.8528123 0.56715656 0.47644257
0.35123353
## Proportion of Variance 0.7592636 0.1212148 0.05361109 0.03783292
0.02056083
## Cumulative Proportion  0.7592636 0.8804784 0.93408954 0.97192246
0.99248329
##                             Comp.6
## Standard deviation     0.212368258
## Proportion of Variance 0.007516713
## Cumulative Proportion  1.000000000

# Combinaciones lineales (loadings) para la matriz de covarianzas
print("Loadings (covarianza):")

## [1] "Loadings (covarianza):"

print(cpS$loadings)

##
## Loadings:
##           Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## edad       0.349  0.908  0.232
```

```
## peso      0.766 -0.162 -0.522  0.339
## altura    0.476 -0.385  0.789
## muneca                        -0.125 -0.989
## biceps    0.248        -0.225 -0.928  0.132
## sexo_num                                     -0.994
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var  0.167  0.167  0.167  0.167  0.167  0.167
## Cumulative Var  0.167  0.333  0.500  0.667  0.833  1.000
```

```r
# Combinaciones lineales (loadings) para la matriz de correlaciones
print("Loadings (correlación):")
```

```
## [1] "Loadings (correlación):"
```

```r
print(cpR$loadings)
```

```
##
## Loadings:
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## edad      0.301  0.865  0.353         0.130  0.141
## peso      0.442 -0.149        -0.394  0.613 -0.501
## altura    0.383 -0.442  0.750        -0.211  0.217
## muneca    0.433  0.122 -0.329 -0.404 -0.703 -0.178
## biceps    0.443 -0.131 -0.441         0.249  0.727
## sexo_num  0.428        -0.101  0.822        -0.348
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var  0.167  0.167  0.167  0.167  0.167  0.167
## Cumulative Var  0.167  0.333  0.500  0.667  0.833  1.000
```

```r
# Puntuaciones de las observaciones para la matriz de covarianzas
print("Scores (covarianza):")
```

```
## [1] "Scores (covarianza):"
```

```r
print(cpS$scores)
```

```
##            Comp.1      Comp.2      Comp.3      Comp.4       Comp.5
Comp.6
##   [1,]  27.167240   1.0275043   5.0018603  0.94414760 -0.518245368
0.009512231
##   [2,]  22.369361  27.5953878   3.0633810 -0.07870246  0.022417166
0.050515399
##   [3,]  19.174384   7.9565576  -1.5768643 -2.60643642  0.791917693
0.208702854
##   [4,]   9.967739   0.8926367   5.5152447  0.10103764 -0.338321100 -
0.294136554
##   [5,]  10.784263  22.0205540  -0.7558501  0.15985124 -0.400403749 -
0.272794884
```

```
## [6,]  23.289302  -7.9270092   2.7958133 -2.08336918 -0.634049034
0.159083915
## [7,]   6.959079 -10.1878393   1.5814314 -5.64604238  0.746098362
0.176220573
## [8,]   5.990947   3.4218469  -7.0105217 -1.02845141 -0.120076939 -
0.298222003
## [9,]   2.139175  -7.0817396   9.6210032 -2.43125215  0.322435603 -
0.213122198
## [10,]  22.747632 -13.2449021  -5.8006728 -1.89462251 -0.123265168
0.136505663
## [11,]  24.432988   4.1225174  -3.0917296  1.41707825  0.462440752 -
0.046604932
## [12,]  -5.425411   0.1815782   1.3567710 -5.18102197 -0.708307995 -
0.220737059
## [13,]  12.673185  -9.7146709  -4.4440552  0.44817492 -0.424023172 -
0.303306081
## [14,]  18.968909  13.1080274   4.5325878  0.30629837 -0.271289276 -
0.091797277
## [15,]  31.845745 -13.4789394  -1.4680337  5.60699140  0.629516704 -
0.246774069
## [16,]  13.891933 -11.8928756  -6.4027962 -2.23617992 -0.010310405 -
0.026730934
## [17,]  17.660575  -2.6451404  -0.8984385 -0.53612809  0.374929388 -
0.040423813
## [18,]  17.730010  -0.7428485   0.1220794  1.76994147  0.702904008 -
0.212130084
## [19,] -23.298995  -1.5206735   0.2628636  1.13708455 -0.156715551 -
0.131481991
## [20,] -14.421814  -7.0888475   0.1027579  0.02090000 -0.336514286
0.146496719
## [21,] -27.083362  -3.1930067  -0.4480438  0.70942019 -0.009353937 -
0.174685591
## [22,] -14.586847   0.8323286  -9.1404074  1.72383160  0.234311501
0.020757558
## [23,] -24.047434  -0.7776982  -5.8548135 -0.34448981  0.269236582 -
0.007565623
## [24,] -12.502596  -5.2753638   3.0618398  1.10904145 -0.523234244
0.093628618
## [25,] -26.007248   3.5762800   1.6619688  0.04703243 -0.329179278 -
0.100266918
## [26,]  -5.775902  -6.4860953  -6.5870665  2.35142361 -0.769748861
0.100541697
## [27,]  -1.222832  -4.4906923   4.4910018  1.18993539  0.240230706
0.392445870
## [28,]   3.008558  11.0460121 -10.8065936  0.30249415 -0.466455317
0.505462747
## [29,] -11.582316   2.5905365   9.5298485  1.48381788  0.831707300
0.189746666
## [30,] -15.342514  -2.9912741   6.9964967  0.50590951 -0.373468613
0.102132524
```

```
## [31,]  -7.935397  -5.1315449   4.1459514  2.82547130  0.288257541
0.101883049
## [32,] -32.049219   9.3868561   0.8366044 -1.35531616  0.743623031 -
0.044763620
## [33,] -24.805735  -0.8613627  -0.1243984 -0.48139991  0.586813138
0.019326309
## [34,] -29.887676   6.8141517  -9.5232394 -0.38787973  0.286722372 -
0.133471930
## [35,] -21.632224  -2.8830225   7.4391699  0.70416549 -0.642085002 -
0.072282284
## [36,]   6.806497   3.0427715   1.8148500  1.42724365 -0.378514551
0.518335455
```

```r
# Puntuaciones de las observaciones para la matriz de correlaciones
print("Scores (correlación):")
```

```
## [1] "Scores (correlación):"
```

```r
print(cpR$scores)
```

```
##          Comp.1       Comp.2       Comp.3       Comp.4      Comp.5
##  [1,]  2.9685191  0.091105385  0.556341114 -0.325722274 -0.33321235
##  [2,]  2.7244413  2.589702948  0.429147876  0.033851625  0.00506402
##  [3,]  2.3087699  0.624538828 -0.144298366  0.247162514  0.45424202
##  [4,]  1.4147700  0.046027842  0.413847021  0.403208258 -0.39037991
##  [5,]  1.7659246  2.123678856 -0.125004024  0.275351936 -0.21587261
##  [6,]  2.9433808 -0.773138558 -0.067220123 -0.344618117 -0.50802088
##  [7,]  1.3546093 -1.165143484 -0.342736398  0.660902648  0.10089040
##  [8,]  1.2469350  0.333226343 -0.844541068  0.390655539  0.03058185
##  [9,]  0.6395231 -0.841520112  0.611016897  0.934609967 -0.27770561
## [10,]  2.7205133 -1.282946435 -0.724200350 -0.301216521  0.04960385
## [11,]  2.4177985  0.339600591  0.105209450 -0.003034808  0.50332460
## [12,]  0.7821840 -0.009361761 -0.787685215  0.672022652 -0.82987770
## [13,]  1.6481660 -0.898632203 -0.498430048  0.083331903 -0.13419084
## [14,]  2.3200553  1.222270199  0.467126433  0.078708628 -0.23839381
## [15,]  2.3670281 -1.339385376  0.786553138 -0.129538184  0.77080983
## [16,]  1.9242233 -1.165340383 -0.859227251  0.070635515  0.06313535
## [17,]  1.9693916 -0.334921214 -0.002605531  0.211040461  0.24883192
## [18,]  1.5970885 -0.171038945  0.440374225  0.402791853  0.52229425
## [19,] -2.6697714 -0.079504466  0.011694211  0.084976779 -0.11377071
## [20,] -1.7098857 -0.604583794 -0.112411967 -0.369030153 -0.21412248
## [21,] -3.0295955 -0.255939774 -0.097518191  0.259214314 -0.04932086
## [22,] -1.9646989  0.170632027 -0.613124091 -0.270094395  0.48197513
## [23,] -2.6521164 -0.029260476 -0.609983548  0.117235175  0.25335313
## [24,] -1.6049236 -0.418319266  0.252601021 -0.426285661 -0.35090123
## [25,] -2.6946407  0.409270976 -0.036523673  0.136900388 -0.31909426
## [26,] -0.9625700 -0.435515676 -0.520301625 -0.916672664 -0.12524172
## [27,] -0.8563358 -0.430917611  0.698532471 -0.579325035  0.15299491
## [28,]  0.1833650  1.225367110 -0.942507977 -1.257580270  0.17027736
## [29,] -2.0359529  0.157341878  1.288059350  0.142579089  0.31953197
## [30,] -1.8650936 -0.245349072  0.573699711 -0.205021158 -0.41577252
```

```
## [31,] -1.6849008 -0.480395837  0.801533797 -0.262010463  0.21779685
## [32,] -3.4069627  0.845170565 -0.015427191  0.705594334  0.26331310
## [33,] -2.8750724 -0.108272215 -0.012725173  0.354108789  0.27057639
## [34,] -3.0831797  0.720711563 -0.983504218  0.300946374  0.33721873
## [35,] -2.3738451 -0.207940898  0.510957167 -0.036311891 -0.62554286
## [36,]  0.1728584  0.378782445  0.393282151 -1.139367148 -0.07439530
##             Comp.6
##  [1,] -0.087432783
##  [2,]  0.085169541
##  [3,]  0.403918191
##  [4,] -0.215831367
##  [5,] -0.248201863
##  [6,]  0.174508992
##  [7,]  0.586979727
##  [8,] -0.221798644
##  [9,]  0.121181156
## [10,]  0.133397974
## [11,] -0.098228105
## [12,]  0.099402287
## [13,] -0.351444668
## [14,] -0.077418334
## [15,] -0.494012161
## [16,]  0.052792842
## [17,]  0.036505800
## [18,] -0.182376882
## [19,] -0.176902418
## [20,]  0.060828979
## [21,] -0.167394595
## [22,] -0.138092515
## [23,]  0.009983678
## [24,] -0.044724538
## [25,] -0.088944441
## [26,] -0.237915834
## [27,]  0.269614969
## [28,]  0.197551176
## [29,]  0.235072913
## [30,]  0.056247221
## [31,] -0.037991066
## [32,]  0.178642873
## [33,]  0.128055909
## [34,] -0.097507222
## [35,] -0.109066447
## [36,]  0.245429652
```

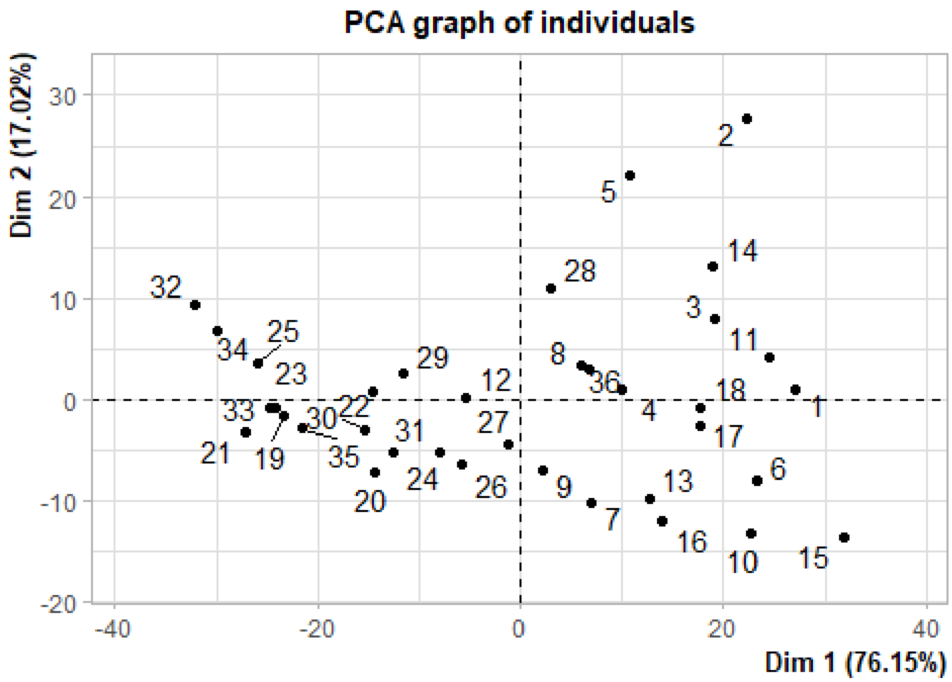
- summary(cpS): Proporciona un resumen del análisis de componentes principales, mostrando las proporciones de varianza explicada por cada componente.
- cpS$loadings: Contiene las combinaciones lineales (vectores propios) que describen cómo se construyen los componentes principales a partir de las variables originales.

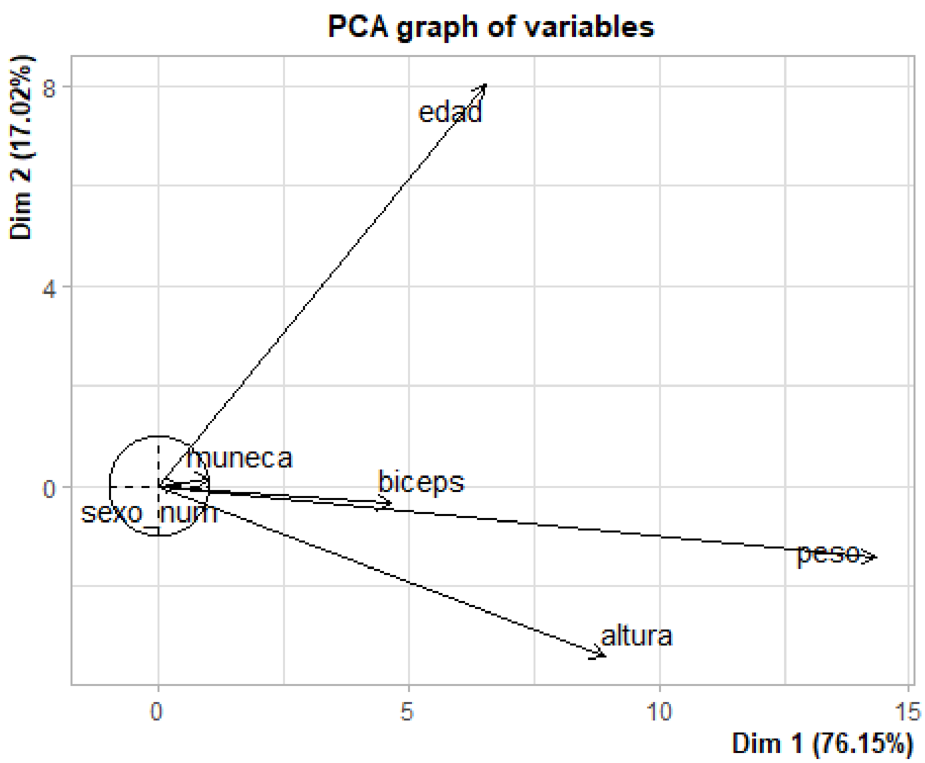
- cpS$scores: Contiene las puntuaciones (proyecciones de las observaciones en el espacio de los componentes principales).

PARTE III

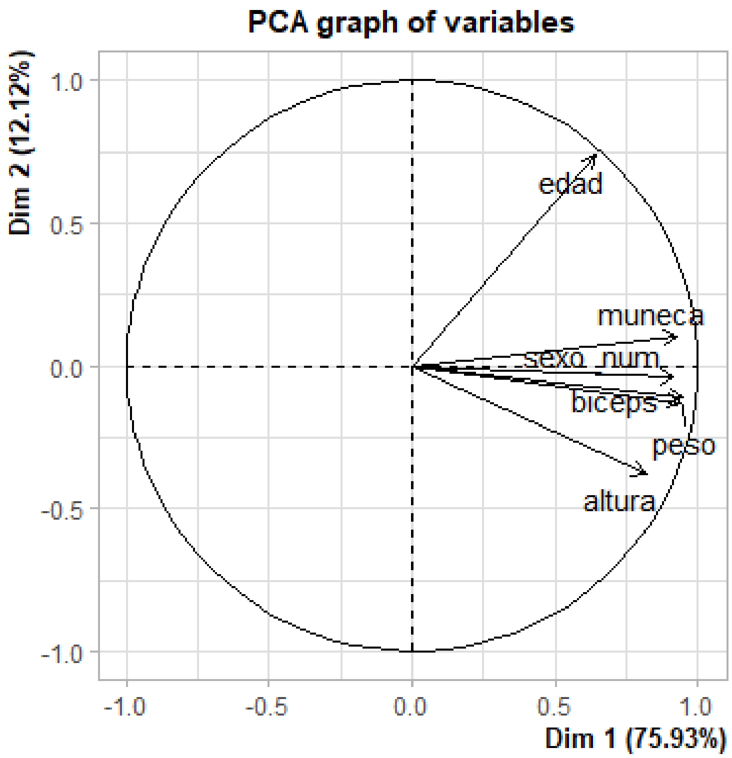
1. Explore los siguientes gráficos relativos a Componentes Principales. Interprete cada gráfico e identifica qué es lo que se está graficando en cada uno. Realiza el análisis con la matriz de varianzas y covarianzas y correlación.

```
# Cargar las librerías necesarias
library(FactoMineR)
library(factoextra)

## Loading required package: ggplot2

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

library(ggplot2)

# 1. Análisis de Componentes Principales (ACP)
cpS_cov <- PCA(X, scale.unit = FALSE) # Para usar la matriz de varianzas-
covarianzas
```



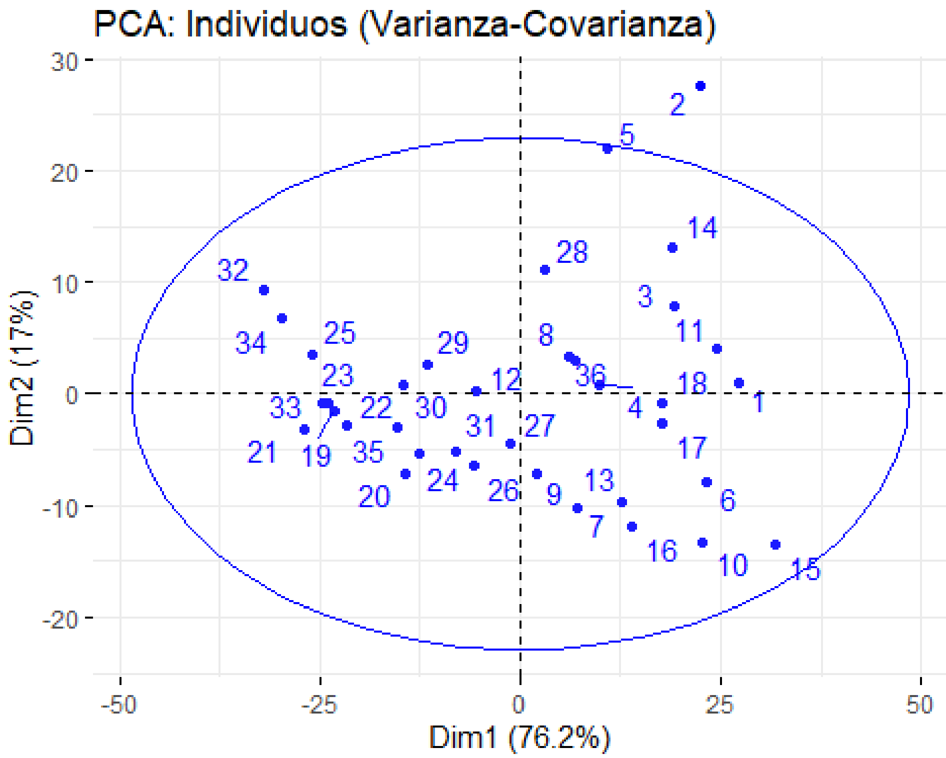



```
cpS_cor <- PCA(X, scale.unit = TRUE) # Para usar la matriz de
correlaciones
```

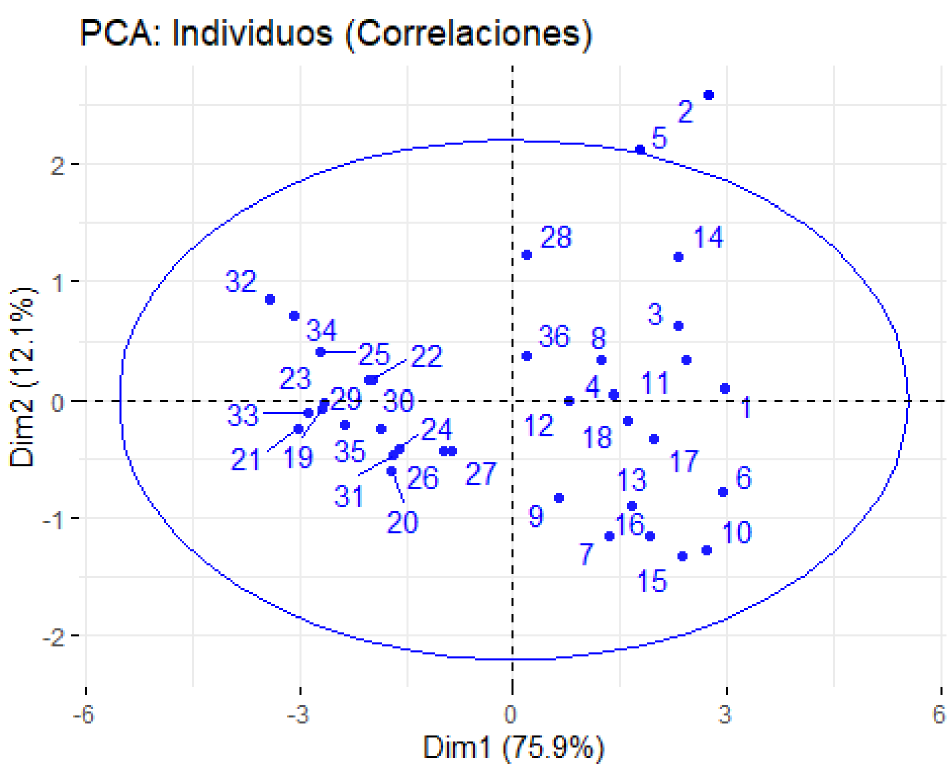


- Gráfico de individuos: Este gráfico muestra cómo las observaciones (individuos) se proyectan en el nuevo espacio de los componentes principales. Los individuos cercanos entre sí en el gráfico tienen perfiles similares respecto a las variables originales. Si se añaden elipses, estas indican la agrupación de los individuos.
- Gráfico de variables: Muestra la correlación entre las variables y los componentes principales. Las variables más cercanas a los ejes principales tienen una mayor correlación con ese componente.

```
# 2. Gráfico de individuos (observaciones) con elipses
fviz_pca_ind(cpS_cov, col.ind = "blue", addEllipses = TRUE, repel = TRUE,
             title = "PCA: Individuos (Varianza-Covarianza)")
```



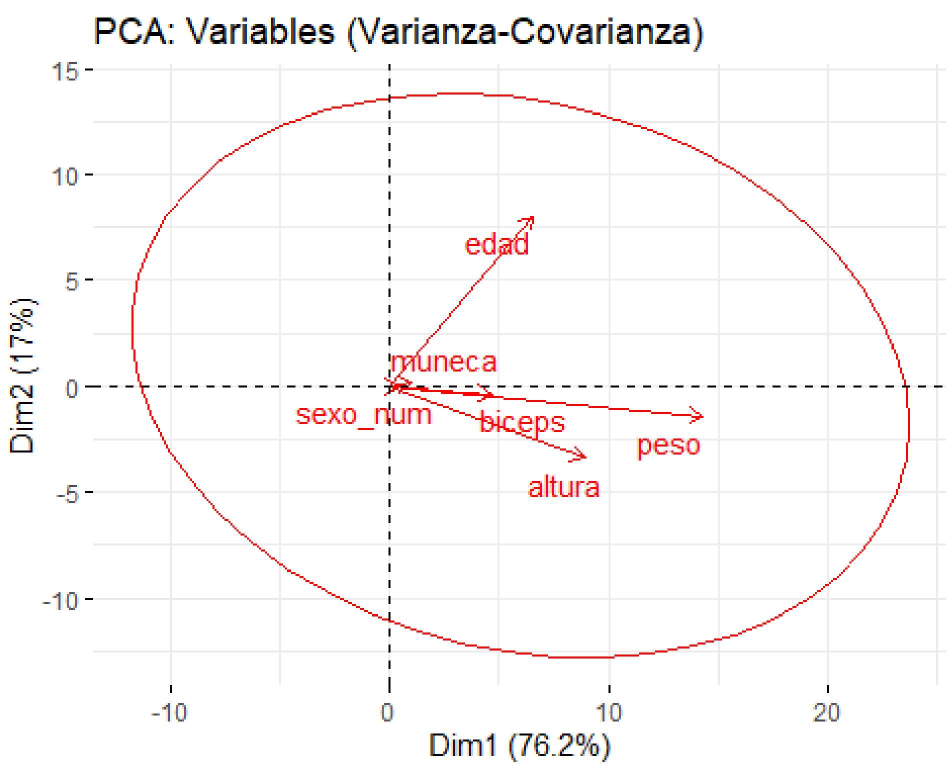

```
fviz_pca_ind(cpS_cor, col.ind = "blue", addEllipses = TRUE, repel = TRUE,
             title = "PCA: Individuos (Correlaciones)")
```



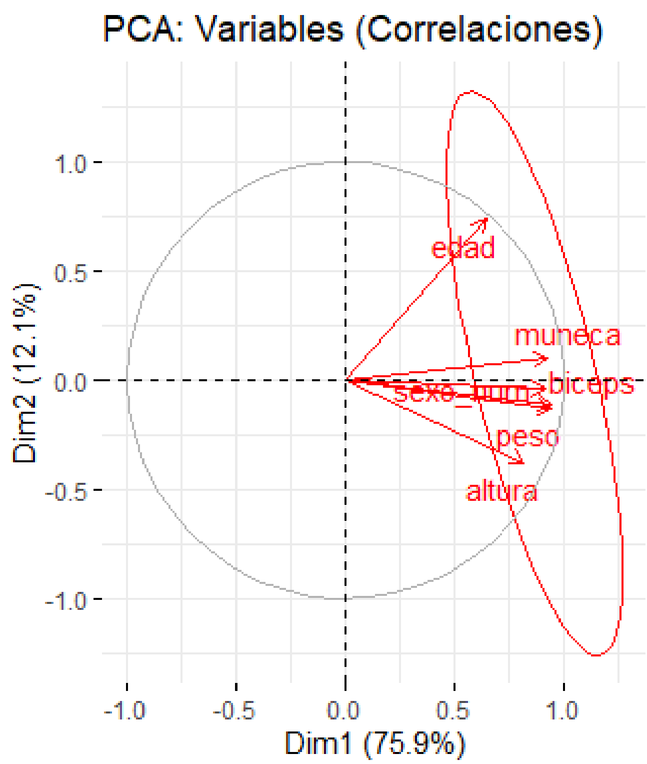
Este gráfico muestra las observaciones individuales en el espacio de los componentes principales. Las observaciones están coloreadas de azul, y el argumento repel = TRUE se utiliza para evitar la sobreposición de las etiquetas.Las elipses indican grupos de observaciones similares que pueden compartir características comunes.

Este gráfico ayuda a visualizar la distribución de las observaciones y cómo se agrupan o se diferencian unas de otras según los componentes principales.Las observaciones que están dentro de las mismas elipses están correlacionadas entre sí y comparten características similares. Las que están fuera pueden ser atípicas o representar otro grupo de observaciones.

```
# 3. Gráfico de variables con elipses
fviz_pca_var(cpS_cov, col.var = "red", addEllipses = TRUE, repel = TRUE,
             title = "PCA: Variables (Varianza-Covarianza)")
```

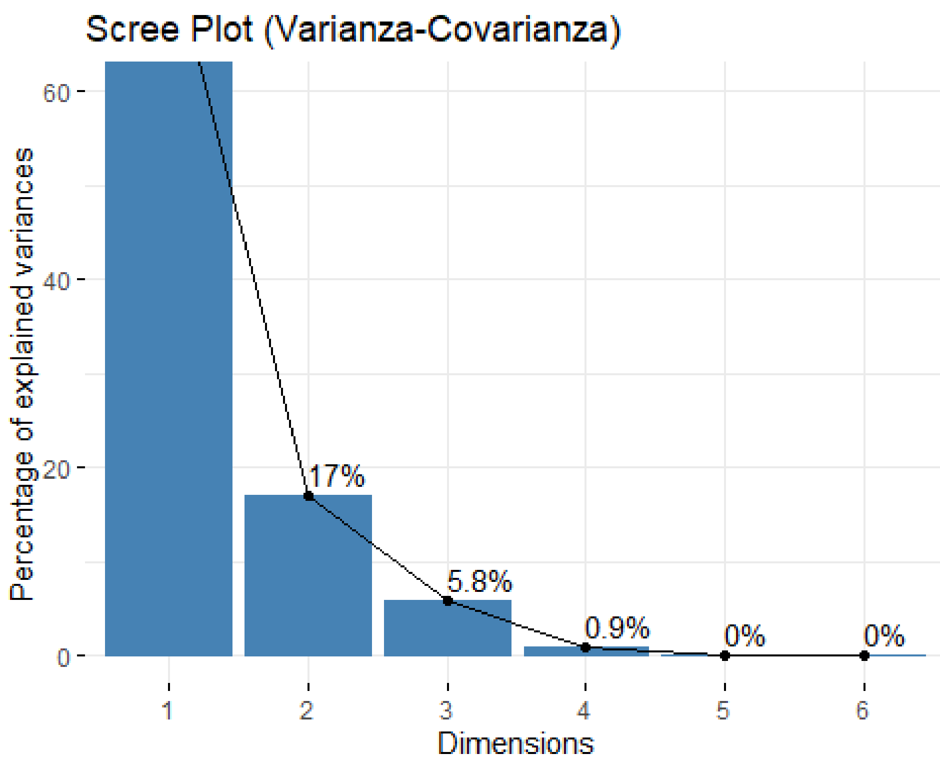



```
fviz_pca_var(cpS_cor, col.var = "red", addEllipses = TRUE, repel = TRUE,
             title = "PCA: Variables (Correlaciones)")
```

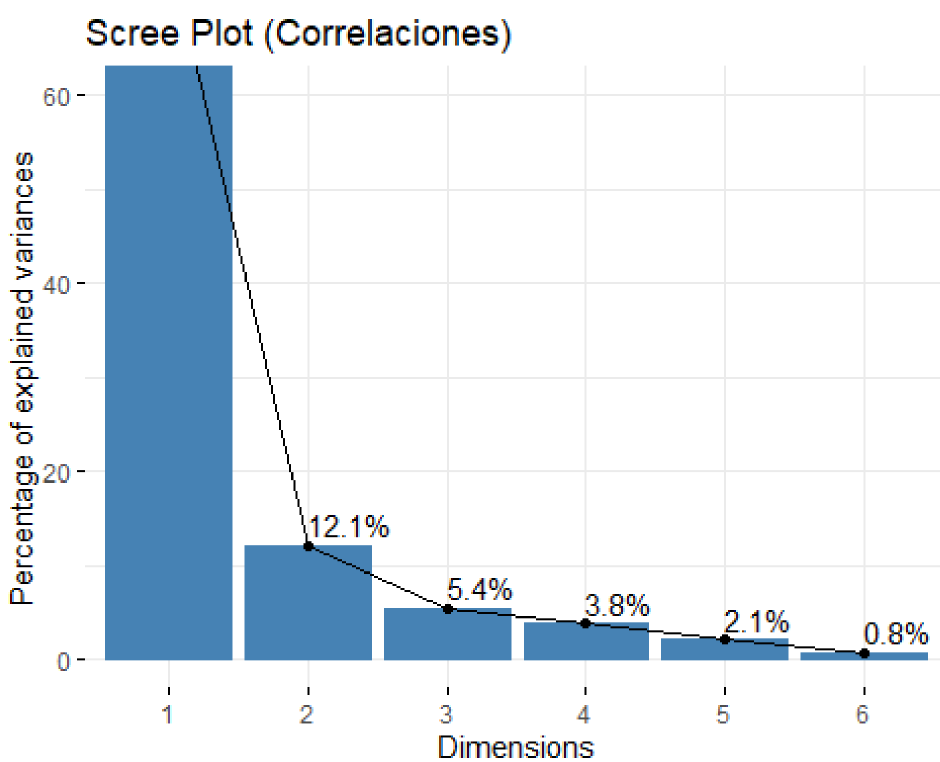


Este gráfico muestra las variables proyectadas en el espacio de los componentes principales. Las variables están representadas por vectores rojos. Las variables que están representadas por vectores largos tienen una fuerte contribución a los componentes principales, las variables que están agrupadas (en la misma dirección) están altamente correlacionadas entre sí y las variables en direcciones opuestas están negativamente correlacionadas.

```r
# 4. Gráfico de sedimentación (Scree plot) para ver la varianza explicada
fviz_screeplot(cpS_cov, addlabels = TRUE, ylim = c(0, 60),
               title = "Scree Plot (Varianza-Covarianza)")
```

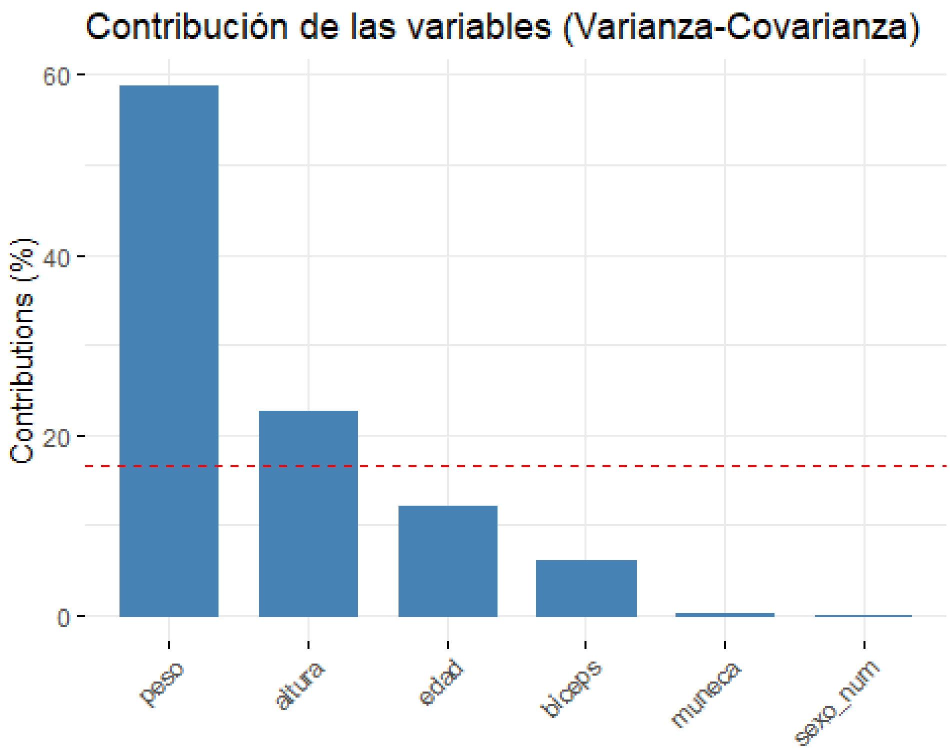



```
fviz_screeplot(cpS_cor, addlabels = TRUE, ylim = c(0, 60),
               title = "Scree Plot (Correlaciones)")
```

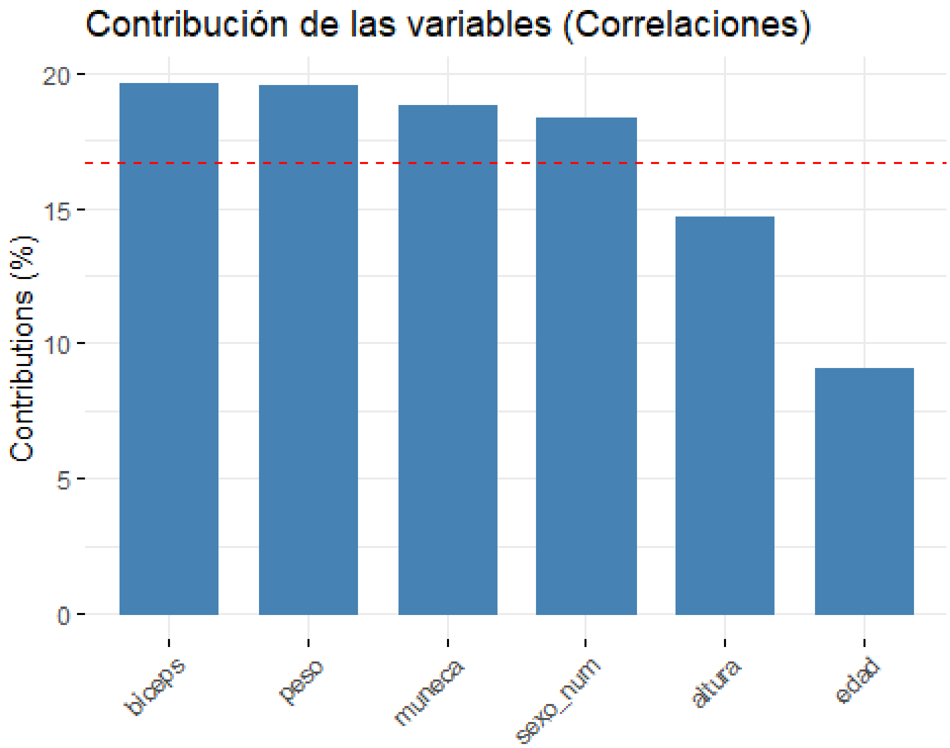


Este gráfico muestra el gráfico de sedimentación (Scree plot) que indica la varianza explicada por cada componente principal. Muestra qué componentes son más importantes en la explicación de la varianza total. Se seleccionan los componentes que explican una gran parte de la varianza antes de que la curva empiece a disminuir drásticamente (esta se llama la "regla del codo"). En ambos casos 2 dimensiones.

```
# 5. Contribución de las variables a los componentes principales
fviz_contrib(cpS_cov, choice = "var", axes = 1, top = 10,
             title = "Contribución de las variables (Varianza-
Covarianza)")
```

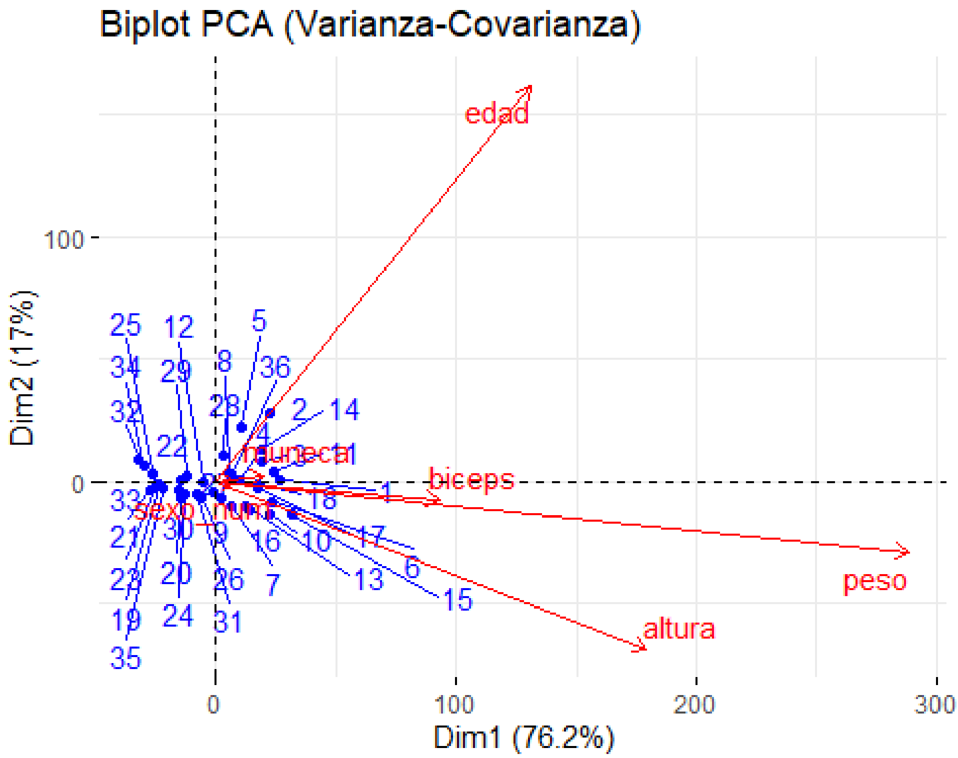



```
fviz_contrib(cpS_cor, choice = "var", axes = 1, top = 10,
             title = "Contribución de las variables (Correlaciones)")
```

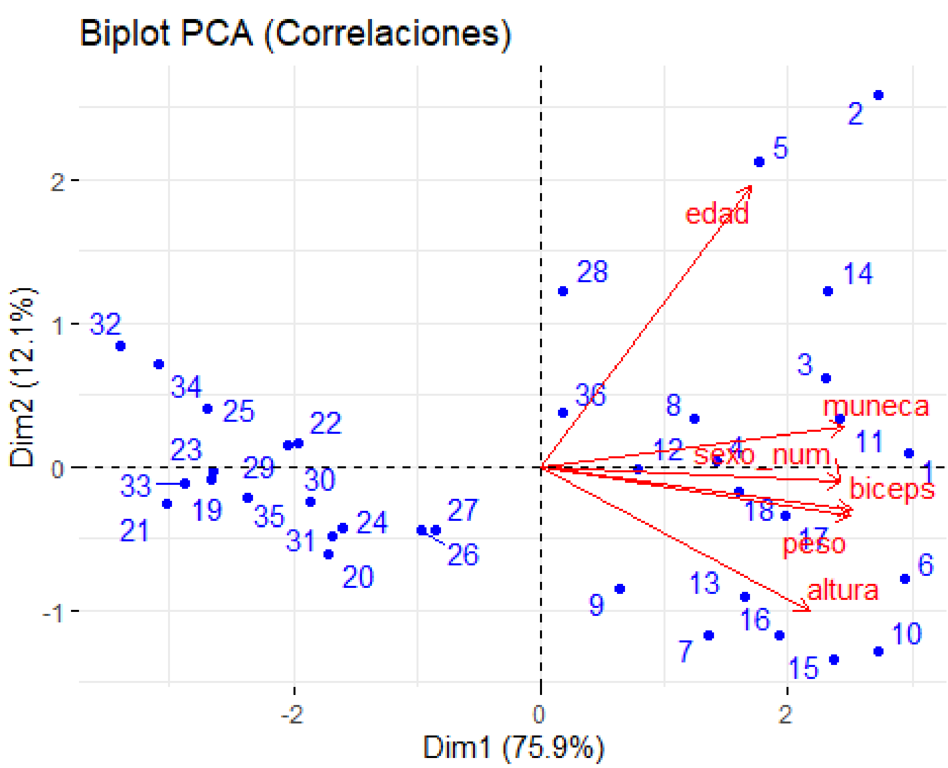


Este gráfico muestra la contribución de las variables a los componentes principales. Se puede identificar qué variables tienen una mayor influencia en la formación de cada componente. Las contribuciones más altas indican una mayor correlación con el componente respectivo. En el caso de la gráfica de varianza-covarianza es peso y el de correlación son biceps y peso.

```
# 6. Gráfico biplot (individuos y variables juntos)
fviz_pca_biplot(cpS_cov, repel = TRUE, col.var = "red", col.ind = "blue",
                title = "Biplot PCA (Varianza-Covarianza)")
```




```r
fviz_pca_biplot(cpS_cor, repel = TRUE, col.var = "red", col.ind = "blue",
                title = "Biplot PCA (Correlaciones)")
```



Este gráfico combina tanto las observaciones como las variables en un solo gráfico, lo que permite una visión conjunta. Muestra cómo las observaciones están relacionadas con las variables. Las observaciones más cercanas a un vector de variable indican que tienen un valor alto en esa variable (como el 36 a edad). Las observaciones que están alineadas en la dirección de una variable están fuertemente relacionadas con esa variable (como 5 a edad).

PARTE IV

Finalmente: Concluye sobre el análisis de componentes principales realizado e interprete los resultados.

a) 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?

Matriz de varianza-covarianza

Las variables con mayor varianza tienen un peso más significativo en la creación de los componentes principales. Este método tiende a resaltar las variables que tienen una mayor dispersión y puede ser útil si estas diferencias en la magnitud de las variables son relevantes.Sin embargo, si las variables tienen escalas diferentes, este método puede producir componentes principales que están fuertemente influenciados por variables con mayor rango de valores, lo que puede no ser deseable en todas las situaciones.

Matriz de correlaciones

Aquí, todas las variables tienen la misma importancia en el análisis, independientemente de su escala o varianza. Esto es especialmente útil cuando las variables están medidas en diferentes unidades o tienen variaciones de magnitud significativas. Los resultados tienden a ser más equilibrados, ya que no permiten que una variable con mayor varianza domine el análisis. Esto facilita la identificación de patrones entre variables con varianza similar y diferentes escalas.

Conclusión Cuando las variables tienen escalas y unidades de medida diferentes, el análisis basado en la matriz de correlación es generalmente más útil, ya que elimina el sesgo causado por diferencias en la magnitud de las variables. Si las escalas de las variables son comparables o si el rango de variabilidad es relevante para el análisis, la matriz de varianza-covarianza puede proporcionar componentes principales que capturen mejor la estructura del dataset. El análisis con la matriz de correlación suele aportar *componentes de mayor interés*, ya que permite observar relaciones equilibradas entre las variables, sin que unas dominen por su mayor varianza.

b) Indique cuál de los dos análisis (a partir de la matriz de varianza y covarianza o de correlación) resulta mejor para los datos. Comparar los resultados y argumentar cuál es mejor según los resultados obtenidos.

La matriz de varianza-covarianza es útil si te interesa que las variables con mayor variabilidad tengan más peso, o si los datos están en la misma escala. La matriz de correlación es mejor si las variables tienen escalas distintas o si se desea dar igual importancia a todas las variables.

Si, tras estandarizar las variables, los componentes principales explican una mayor parte de la variabilidad y los patrones resultan más equilibrados, entonces el análisis con la matriz de correlación será más apropiado. Si el análisis con la matriz de varianza-covarianza resalta mejor las diferencias entre variables importantes con alta varianza, podría ser mejor para ciertos casos donde esas diferencias son clave.

En nuestro caso que las variables están en diferentes escalas, el uso de la matriz de correlación es el método preferido de está manera, también podemos apreciar mejor las gráficas y las relaciones existentes. En general, para datos multivariados con diferentes escalas, el análisis con correlación resulta más robusto.

c) ¿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)

Las variables con mayor peso (en valor absoluto) en la combinación lineal de los componentes son las que más varían conjuntamente. Por ejemplo, si algunas variables tienen un valor de coeficiente alto en PC1, eso indicaría que esas variables explican una gran parte de la variabilidad capturada por el primer componente. Podemos observar el gráfico 3 (fviz_pca_var) y visualizar qué variables están más cercanas a los ejes de las componentes, podemos apreciar que del método seleccionado (correlación), las variables que más contribuyen a Dim1 son peso y altura, ya que están más alineadas con Dim1, sus vectores son más largos y apuntan más hacia el eje X. Mientras que las variable que más contribuye al segundo componente es aquella cuyo vector está más alineado con el eje Y (Dim2), la cual es edad, ya que su vector está más alineado verticalmente con el eje Y.

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

Las combinaciones lineales que se obtienen del análisis de componentes principales están dadas por las cargas o pesos asignados a cada variable en los componentes principales (PCs). Para escribir las combinaciones finales, debes observar los coeficientes de cada variable en los primeros componentes principales, que representan las proporciones en que contribuyen.

```r
# Visualizar las cargas de las variables
print(cpS_cov$var$coord)
```

```
##               Dim.1       Dim.2       Dim.3        Dim.4        Dim.5
## edad      6.5184606  8.02251044  1.20469295 -0.004939155 -0.012359635
## peso     14.3216379 -1.42929624 -2.70368411 -0.695470642 -0.005535643
## altura    8.9037774 -3.40503255  4.08896841  0.092288333 -0.001218014
```

```
## muneca    1.0068616  0.13737964 -0.14436966  0.257344952  0.472777127
## biceps    4.6394013 -0.35558919 -1.16357203  1.905002072 -0.063162724
## sexo_num  0.4194996  0.01071737  0.01147726  0.175399178 -0.029305732
```

Las combinaciones lineales quedarían de la siguiente forma:

$$PC_1 = 6.51 * edad + 14.32 * peso + 8.90 * altura + 1.0068 * muneca + 4.6394 * biceps + 0.419 * sexo_num$$

$$PC_1 = 8.022 * edad - 1.429 * peso - 3.405 * altura + 1.1373 * muneca - 0.355 * biceps + 0.107 * sexo_num$$

La primera componente principal está dominada por la variable peso (con una carga de 14.32), seguida de altura (8.90) y edad (6.52). Esto sugiere que peso es la variable más influyente en la dirección principal de la variabilidad de los datos.

La segunda componente principal está dominada por la variable edad (con una carga de 8.02), mientras que altura tiene un peso negativo importante (-3.41). Esto indica que la segunda dirección de variabilidad se explica en gran parte por edad, pero tiene una relación inversa con altura.

e) Interpreta los resultados en término de agrupación de variables

Al interpretar los componentes principales, las variables que están cercanas entre sí en los gráficos de PCA var y biplot suelen estar correlacionadas, lo que indica que tienden a variar juntas. Este agrupamiento ayuda a identificar subconjuntos de variables que tienen relaciones similares. Las variables que están más alejadas del origen y cercanas a un eje de componente principal tienen una mayor correlación con ese componente. * Si algunas variables forman un grupo en el gráfico de variables, estas pueden ser interpretadas como representativas de una dimensión particular de la variabilidad en los datos.

Los resultados de la matriz de coordenadas para las variables en los componentes principales nos permiten identificar cómo se agrupan las variables según su influencia en los componentes. Esta agrupación se basa en las cargas que tienen sobre las primeras dos dimensiones (Dim.1 y Dim.2) del Análisis de Componentes Principales (ACP).

Variables con mayor carga positiva en Dim.1 * Todas

Las variables que se agrupan más en el primer componente principal parecen estar relacionadas con características físicas generales como peso, altura, edad y medidas corporales (bíceps). Esto sugiere que estas variables se correlacionan positivamente y forman un grupo que representa la “talla corporal general”.

Variables con mayor carga positiva en Dim.2 * Edad (8.02), Muñeca (0.14) y sexo_num (0.01)

Variables con carga negativa en Dim.2 *Altura (-3.41), Peso (-1.43) y Bíceps (-0.36)

El segundo componente principal agrupa principalmente la edad de forma positiva, pero tiene una relación inversa con peso y altura. Esto sugiere que en esta dimensión, edad está en contraposición a las medidas físicas de tamaño corporal.

Dim.1 agrupa principalmente las características relacionadas con el tamaño corporal (peso, altura, bíceps), lo que sugiere que este componente refleja una dimensión de "volumen o talla física".

Dim.2 agrupa principalmente la edad, pero en contraposición al tamaño físico, lo que indica que las personas con mayor edad tienden a tener una relación inversa con peso y altura en esta dimensión.