A4-Componentes Principales

Rodolfo Jesús Cruz Rebollar

2024-10-09

```
# Leer base de datos

corporal = read.csv("corporal.csv")

head(corporal)
```

```
##   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
```

```
# Quitar variables no numéricas de la base de datos

corporal = corporal[, -4]

head(corporal)
```

```
##   edad peso altura muneca biceps
## 1   43 87.3  188.0   12.2   35.8
## 2   65 80.0  174.0   12.0   35.0
## 3   45 82.3  176.5   11.2   38.5
## 4   37 73.6  180.3   11.2   32.2
## 5   55 74.1  167.6   11.8   32.9
## 6   33 85.9  188.0   12.4   38.5
```

Análisis descriptivo inicial:

```
# Mostrar un resumen (summary) con las medidas estadísticas más
relevantes de las variables

summary(corporal)
```

```
##       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
```

```
# Calcular la desviación estándar por variable

cat("Desviación estándar por variable: ", "\n",
    "Edad: ", sd(corporal$edad), "\n",
    "Peso: ", sd(corporal$peso), "\n",
    "Altura: ", sd(corporal$altura), "\n",
    "Muñeca: ", sd(corporal$muneca), "\n",
    "Biceps: ", sd(corporal$biceps))
```

```
## Desviación estándar por variable:
##  Edad:  10.55447
##  Peso:  14.869
##  Altura:  10.52017
##  Muñeca:  1.175463
##  Biceps:  5.234392
```

En cuanto a las medidas estadísticas obtenidas en los 2 chunks anteriores, se observa que para el caso de la variable edad, la mediana se ubica a la izquierda de la media, puesto la mediana es menor que la media, lo cual también señala que la distribución de los datos de edad tiene un sesgo positivo (a la derecha), dado que la media de los valores se ubica a la derecha de la mediana de los mismos, mientras que en el caso de las demás variables, éstas tienen un sesgo negativo (a la izquierda), esto debido a que en esta ocasión, la media de los valores de dichas variables se localiza a la izquierda de la mediana de los datos mismos, lo cual indica que en la base de datos original, hay una mayor cantidad de estudiantes universitarios de edad joven,, esto debido a que las observaciones con edades menores poseen una mayor frecuencia dentro de la base de datos, por lo cual, esto tiene sentido, ya que los estudiantes de universidad son en su gran mayoría personas jóvenes, por lo que es lógico que haya más observaciones con edades por lo general entre los 18 y 23 años.

```
# Calcular la matriz de correlaciones entre las variables

cor(corporal)
```

```
##             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
```

En la matriz de correlación anterior se observa que existe una correlación bastante fuerte entre el peso, la altura y la medida del biceps de los estudiantes universitarios,

dado que tienen coeficientes de correlación de 0.79 (peso con altura), 0.84 (peso con muñeca) y 0.90 (peso con biceps) respectivamente, además de que también es posible apreciar que existe una correlación igualmente fuerte entre la medida de la muñeca de los alumnos y la medida de su biceps, siendo éste nivel de correlación del 87.77%, motivo por el cual, dichos niveles altos de correlación serán observables en los gráficos que se realizarán posteriormente durante todo el proceso de análisis, esto principalmente por medio de superposiciones muy cercanas entre ellas y mayormente horizontales de los vectores correspondientes a las variables originales, para indicar que las variables en cuestión poseen altos niveles de correlación como ya se observó en la matriz de correlaciones.

Parte I

Matriz de varianza y covarianza

```
# Calcular matriz de varianzas y covarianzas

M_varcov = cov(corporal)

M_varcov
```

```
##              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
```

Matriz de correlaciones

```
# Calcular la matriz de correlaciones

M_cor = cor(corporal)

M_cor
```

```
##             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
```

Análisis de matriz de varianzas y covarianzas

```
# Calcular valores y vectores propios de la matriz de varianzas y
covarianzas

eig_varcov = eigen(M_varcov)
```

```
eig_varcov
```

```
## eigen() decomposition
## $values
## [1] 359.3980243  80.3757858  27.6229011   4.3074318   0.2343571
##
## $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
```

```
# Calcular la proporción de varianza explicada por cada componente

var_prop_varcov = eig_varcov$values / sum(diag(M_varcov))

var_prop_varcov
```

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

Comprobación de que la varianza total de los componentes principales es igual a la traza de la matriz de varianzas y covarianzas:

```
# Suma de los valores propios de la matriz de varianzas y covarianzas

sum(eig_varcov$values)
```

```
## [1] 471.9385
```

```
# Traza de la matriz de varianzas y covarianzas (suma de los valores de su diagonal)

sum(diag(M_varcov))
```

```
## [1] 471.9385
```

```
# Obtener la varianza acumulada en cada componente

var_acum_varcov = cumsum(var_prop_varcov)

var_acum_varcov
```

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

En base a los resultados anteriores, es posible apreciar que en base al cálculo de la proporción de varianza explicada por cada componente, los 2 primeros componentes que explican un 76.15% y un 17.03% de la varianza total de los datos respectivamente, son aquellos más importantes, porque además de que son aquellos

que explican las mayores proporciones de varianza de los datos, también se observa que la varianza acumulada al tomar 2 componentes, es del 93.18%, representando la gran mayoría de la varianza total de los datos originales, de manera que los 3 componentes sobrantes ya no aportan una cantidad significativa de varianza, por lo cual ya no son relevantes.

Ecuación de los componentes principales CP1 y CP2:

```
# Ecuación del componente principal 1 CP1

cat("CP1 = ", eig_varcov$vectors[1, 1], "* edad", " + ",
    eig_varcov$vectors[2, 1], "* peso", " + ", eig_varcov$vectors[3, 1],
"* altura",
    " + ", eig_varcov$vectors[4, 1], "* muneca", " + ",
eig_varcov$vectors[5, 1],
    "* biceps")

## CP1 =  -0.34871 * edad  +  -0.7661759 * peso  +  -0.4763241 * altura
+  -0.05386189 * muneca  +  -0.2481737 * biceps

# Ecuación del componente principal 2 CP2

cat("CP2 = ", eig_varcov$vectors[1, 2], "* edad", " + ",
    eig_varcov$vectors[2, 2], "* peso", " + ", eig_varcov$vectors[3, 2],
"* altura",
    " + ", eig_varcov$vectors[4, 2], "* muneca", " + ",
eig_varcov$vectors[5, 2],
    "* biceps")

## CP2 =  0.9075501 * edad  +  -0.1616581 * peso  +  -0.3851755 * altura
+  0.0155423 * muneca  +  -0.0402221 * biceps
```

En resumen, es posible observar que para el caso del componente principal 1, aquellas variables que más contribución tienen son el peso con un peso o coeficiente de 0.766, seguido de la altura con coeficiente de 0.476, no obstante, de éstas 2 variables mencionadas, la que tiene la mayor contribución es el peso, debido a que tiene un coeficiente notoriamente alto (0.766) en comparación con el resto de coeficientes, además de la altura que también posee un coeficiente medianamente significativo (0.476).

Adicionalmente, en el caso del componente principal 2, la variable que tiene la mayor contribución en dicho componente es la edad, porque posee un coeficiente de 0.907, lo cual es bastante alto, lo cual indica que la edad es de suma relevancia en el CP2, siendo la variable predominante en dicho componente sobre las demás variables, de las cuales, ninguna tiene un peso significativo además de la edad como para considerarse de alta relevancia dentro del componente principal 2.

Análisis de matriz de correlaciones

```
# Calcular valores y vectores propios de la matriz de correlaciones
```

```r
eig_cor = eigen(M_cor)

eig_cor
```

```
## eigen() decomposition
## $values
## [1] 3.75749733 0.72585665 0.32032981 0.12461873 0.071697492
##
## $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
```

```r
# Calcular la proporción de varianza explicada por componente

var_prop_cor = eig_cor$values / sum(diag(M_cor))

var_prop_cor
```

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

Comprobación de que la varianza total de los componentes principales es igual a la traza de la matriz de correlación:

```r
# Suma de los valores propios de la matriz de correlación

sum(eig_cor$values)
```

```
## [1] 5
```

```r
# Traza de la matriz de correlación (suma de los valores de su diagonal)

sum(diag(M_cor))
```

```
## [1] 5
```

```r
# Obtener la varianza acumulada en cada componente

var_acum_cor = cumsum(var_prop_cor)

var_acum_cor
```

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

A partir de los resultados previos, se aprecia que en base al cálculo de la proporción de varianza explicada por cada componente, los 2 primeros componentes que explican un 75.14% y un 14.51% de la varianza total de los datos respectivamente, son aquellos más importantes, porque además de que son aquellos que explican las

mayores proporciones de varianza de los datos, también se observa que la varianza acumulada al tomar 2 componentes, es del 89.66%, representando la gran mayoría de la varianza total de los datos originales, de forma que los 3 componentes restantes ya no aportan una cantidad significativa de varianza, por lo que no se consideran importantes.

Ecuación de los componentes principales CP1 y CP2:

```
# Ecuación del componente principal 1 CP1

cat("CP1 = ", eig_cor$vectors[1, 1], "* edad", " + ",
    eig_cor$vectors[2, 1], "* peso", " + ", eig_cor$vectors[3, 1], "*
altura",
    " + ", eig_cor$vectors[4, 1], "* muneca", " + ", eig_cor$vectors[5,
1],
    "* biceps")

## CP1 =  -0.335931 * edad  +  -0.4927066 * peso  +  -0.4222426 * altura
+  -0.4821923 * muneca  +  -0.4833139 * biceps

# Ecuación del componente principal 2 CP2

cat("CP2 = ", eig_cor$vectors[1, 2], "* edad", " + ",
    eig_cor$vectors[2, 2], "* peso", " + ", eig_cor$vectors[3, 2], "*
altura",
    " + ", eig_cor$vectors[4, 2], "* muneca", " + ", eig_cor$vectors[5,
2],
    "* biceps")

## CP2 =  0.8575601 * edad  +  -0.1647821 * peso  +  -0.4542223 * altura
+  0.1082775 * muneca  +  -0.1392684 * biceps
```

En resumen, se aprecia que para el caso del componente principal 1, aquellas variables que más contribución tienen son el peso, biceps, muñeca y altura con pesos o coeficientes de 0.492, 0.483, 0.482 y 0.422, respectivamente, por lo que se aprecia que todas éstas variables tienen coeficientes muy similares y cercanos entre sí, por lo que no existe predominancia de una sola variable en el CP1, sino que las variables antes mencionadas son todas predominantes sobre la variable con el menor coeficiente (edad con 0.335), por lo que además de ser las variables predominantes, también es posible que dichas variables tengan una correlación negativa entre ellas, dado que los vectores de las mismas estarán superpuestos (uno encima de otro) y apuntando en el sentido negativo (hacia la izquierda) y por tanto, proyectando una mayor proporción de sí mismos sobre el eje del plano correspondiente al CP1.

Por otro lado, en el caso del componente principal 2, las variables que tienen mayor contribución en dicho componente son la edad seguida de la altura, debido a que la edad posee un coeficiente de 0.857, lo cual es bastante elevado, indicando que la edad es sumamente importante en el CP2, siendo la variable predominante en dicho componente sobre las demás variables, de las cuales, la altura posee un coeficiente

medianamente significativo, cuyo valor es de 0.454, lo cual indica que la altura tiene una contribución moderada en el CP2, sin embargo, la contribución más alta en dicho componente es de la edad con coeficiente de 0.857, siendo el mayor de todos los coeficientes.

Parte II

```
# Librería stats

library(stats)

# Componentes principales a partir de matriz de varianzas-covarianzas

cp_varcov = princomp(corporal, cor = FALSE)

# Componentes principales a partir de matriz de correlaciones

cp_cor = princomp(corporal, cor = TRUE)
```

Scores de las observaciones para los componentes con matriz de varianzas-covarianzas

```
# Calcular puntuaciones o scores de cada observación evaluada en cada componente principal

# Para la matriz de varianzas-covarianzas

scores_varcov = as.matrix(corporal) %*% cp_varcov$loadings

head(scores_varcov)
```

```
##        Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## [1,] 180.9723 -48.75142 104.41935 -13.93818 -4.405445
## [2,] 176.1730 -22.18369 102.48068 -14.95779 -3.863033
## [3,] 172.9774 -41.82266  97.84009 -17.48518 -3.084644
## [4,] 163.7685 -48.88690 104.93178 -14.75095 -4.244360
## [5,] 164.5851 -27.75893  98.66081 -14.69444 -4.305027
## [6,] 177.0934 -57.70609 102.21295 -16.96780 -4.511084
```

Scores de las observaciones para los componentes con matriz de correlaciones

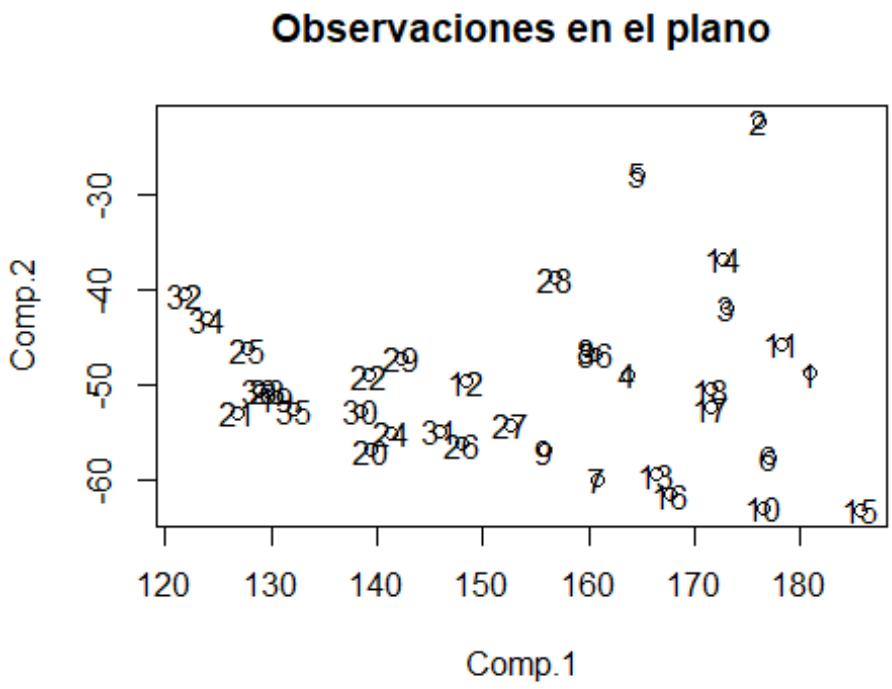
```
# Calcular puntuaciones o scores de cada observación evaluada en cada componente principal

# Para la matriz de correlaciones

scores_cor = as.matrix(corporal) %*% cp_cor$loadings

head(scores_cor)
```

```
##         Comp.1    Comp.2   Comp.3    Comp.4   Comp.5
## [1,] 160.0253 -66.56901 126.4623 14.440542 1.996034
## [2,] 157.4245 -40.05091 124.8048 16.406908 6.165240
## [3,] 154.2006 -59.29073 118.2259 16.046683 5.491792
## [4,] 145.7863 -65.56626 121.6417  7.685578 6.927153
## [5,] 147.3445 -44.47646 118.0373 12.786145 6.509173
## [6,] 157.3776 -75.26829 121.7870 13.016923 3.644315
```

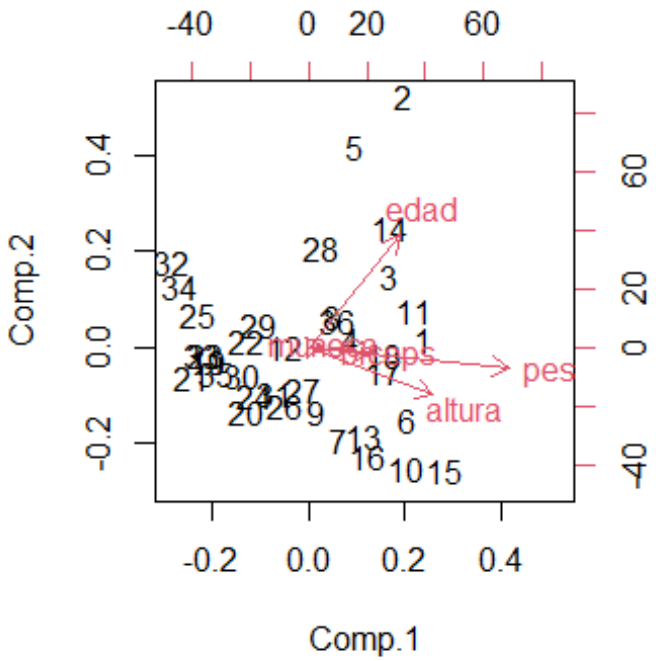
Gráficas de CP1 y CP2

Gráficas para la matriz de varianzas y covarianzas

```
# Graficar los primeros componentes principales de la matriz de
varianzas-covarianzas

plot(scores_varcov[, 1:2], type = "p", main = "Observaciones en el
plano")

text(scores_varcov[, 1], scores_varcov[, 2], 1:nrow(scores_varcov))
```




```
biplot(cp_varcov)
```

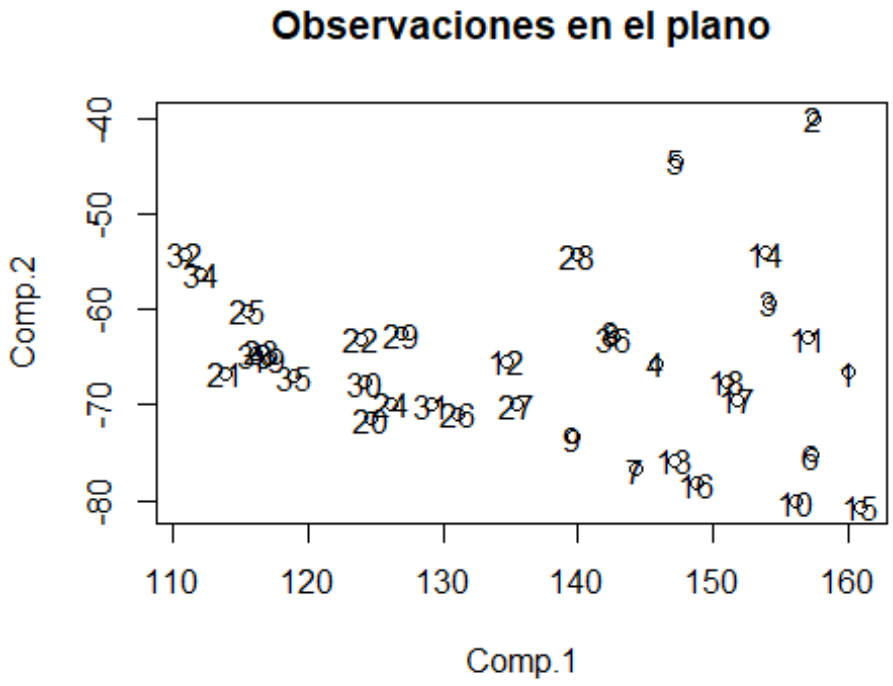


En primera instancia, en los gráficos anteriores es posible apreciar en particular que las variables peso y altura presentan una relación positiva con el componente principal 1, dado que se observa que los vectores correspondientes a dichas variables, se encuentran apuntando hacia la derecha (sentido positivo), lo cual indica que conforme aumenta el valor de la altura y del peso de las personas, las puntuaciones o scores derivados del componente principal 1 también aumentarán, además de que también se puede observar que las variables muñeca, biceps y peso se pueden agrupar entre sí, ya que los vectores de dichas variables se ubican uno encima de otro, (están superpuestos), lo cual es un indicativo de que las variables en cuestión se encuentran altamente correlacionadas entre ellas, permitiendo de esa manera encontrar semejanzas notables entre las mismas de tal forma que es posible formar un grupo con ellas. Por otro lado, para el caso del componente principal 2, es posible notar que la única variable que tiene una relación altamente significativa con dicho componente principal es la edad de las personas, dado que el vector correspondiente a la variable de edad, proyecta una mayor sombra sobre el eje vertical del plano que corresponde al componente principal 2, indicando que la edad tiene una notable contribución dentro de dicho componente, además, es posible observar que el vector de edad se encuentra apuntando hacia arriba de manera vertical, indicando que la edad de las personas posee una relación positiva con el componente principal 2, lo cual significa que conforme incremente la edad de las personas, las puntuaciones o scores derivados del CP2, también serán mayores, además de que también se logra observar que las observaciones número 2, 5, 10, 15 son posibles datos atípicos, principalmente debido a que se ubican notablemente alejadas de la nube de puntos principal del gráfico, particularmente, las observaciones 2 y 5 se encuentran por encima de la

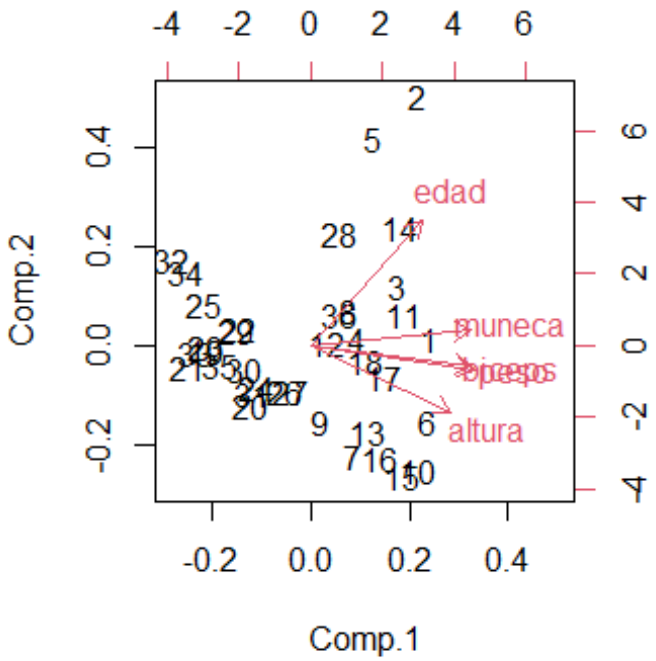
mayoría de puntos, mientras que las observaciones 10 y 15 se ubicna por debajo de la misma, resutando evidente su alejamiento de la mayoría de observaciones y por tanto, es recomendable estudiar con mayor profundidad la naturaleza de dichas observaciones en específico para determinar la mejor manera de procesarlos.

Gráficas para la matriz de correlaciones

```
# Graficar los primeros componentes principales de la matriz de
correlaciones

plot(scores_cor[, 1:2], type = "p", main = "Observaciones en el plano")

text(scores_cor[, 1], scores_cor[, 2], 1:nrow(scores_cor))
```




```
biplot(cp_cor)
```



En cuanto a los gráficos del CP1 y CP2 derivados de la matriz de correlaciones, es posible observar en el gráfico de vectores que en cuanto al componente principal 1, prácticamente las 5 variables representadas tienen una contribución medianamente significativa dentro de dicho componente principal, ya que todos los vectores proyectan una sombra de tamaño mediano sobre el eje horizontal correspondiente al CP1, mientras que de éstas mismas variables, solamente la variable edad proyecta una sombra de tamaño medianamente considerable sobre el eje correspondiente al CP2, por lo cual, es la variable con el mayor peso dentro de dicho componente y la que más contribución tiene dentro del mismo, además, dado que el vector de edad apunta en dirección hacia arriba, se puede afirmar que la edad se relaciona de una forma directamente proporcional con el CP2. lo cual significa que a medida que la edad de las personas incrementa, las puntuaciones derivadas del CP2 también experimentan un aumento, lo cual conduce a que entre mayor sea el valor de la variable edad, los scores para las observaciones derivados del CP2 también aumentarán su calor, sin importar demasiado el valor ya sea positivo, o negativo, que asuman el resto de variables dentro del CP2, ya que la variable predominante en dicho componente (edad) tiene coeficiente positivo, además se aprecia que los vectores de las variables muñeca, biceps, altura y peso se encuentran muy próximos entre sí además de que también se encuentran uno encima de otro, lo cual indica que entre dichas variables, existe un grado moderado de correlación, por lo que es posible formar un grupo con ellas, dado que comparten ciertas similitudes en cuanto a la naturaleza de sus valores. Por último, también es importante destacar que en los gráficos anteriores, se puede notar que las observaciones número 2 y 5 se encuentran alejadas de la mayoría de observaciones (puntos) graficadas, lo cual sugiere que éstas 2 observaciones son posibles

observaciones atípicas, cuyos valores se ubican por encima del promedio de la mayoría de observaciones, siendo recomensable analizar más a fondo las observaciones 2 y 5 para determinar el mejor proceso a seguir con ellos.

Exploración del comando princomp()

El comando `princomp()` ofrece diversas opciones para facilitar la realización del análisis de componentes principales, entre las que destaca el parámetro `na.action` que sirve para indicar la acción a realizar en caso de que se detecten valores faltantes NA en el conjunto de datos, además, `princomp()` también tiene el parámetro fix_sign, cuya función consiste básicamente en indicar si el signo de los coeficientes de los componentes principales, junto con el de los scores se debe elegir de tal manera que el primer elemento de cada loading sea no negativo, entre otras opciones diversas para facilitar todo el proceso de análisis de componentes principales.

Comando summary(cpS):

```
# Calcular el resumen (summary) del análisis de componentes principales
# en base a la matriz de varianzas-covarianzas

summary(cp_varcov)
```

```
## Importance of components:
##                             Comp.1    Comp.2     Comp.3      Comp.4
Comp.5
## Standard deviation     18.6926388 8.8398600 5.18223874 2.046406827
0.4773333561
## Proportion of Variance  0.7615357 0.1703099 0.05853072 0.009127104
0.0004965839
## Cumulative Proportion   0.7615357 0.9318456 0.99037631 0.999503416
1.0000000000
```

```
# Summary del análisis de componetes principales a partir de matriz de
correlaciones

summary(cp_cor)
```

```
## Importance of components:
##                            Comp.1    Comp.2     Comp.3     Comp.4
Comp.5
## Standard deviation     1.9384265 0.8519722 0.56597686 0.35301378
0.2677639
## Proportion of Variance 0.7514995 0.1451713 0.06406596 0.02492375
0.0143395
## Cumulative Proportion  0.7514995 0.8966708 0.96073676 0.98566050
1.0000000
```

Comando cpaS$loading:

```
# Desplegar los loadings del PCA en base a matriz de varianzas-
covarianzas
```

```
cp_varcov$loadings
```

```
##
## Loadings:
##        Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## edad    0.349  0.908  0.232
## peso    0.766 -0.162 -0.522  0.339
## altura  0.476 -0.385  0.789
## muneca                      -0.126 -0.990
## biceps  0.248        -0.225 -0.931  0.138
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings       1.0    1.0    1.0    1.0    1.0
## Proportion Var    0.2    0.2    0.2    0.2    0.2
## Cumulative Var    0.2    0.4    0.6    0.8    1.0
```

```
# Mostrar loadings del PCA en base a matriz de correlaciones

cp_cor$loadings
```

```
##
## Loadings:
##        Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## edad    0.336  0.858  0.349  0.136  0.107
## peso    0.493 -0.165         0.525 -0.671
## altura  0.422 -0.454  0.734 -0.207  0.184
## muneca  0.482  0.108 -0.367 -0.755 -0.226
## biceps  0.483 -0.139 -0.447  0.305  0.674
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings       1.0    1.0    1.0    1.0    1.0
## Proportion Var    0.2    0.2    0.2    0.2    0.2
## Cumulative Var    0.2    0.4    0.6    0.8    1.0
```

Comando cpaS$scores

```
# Mostrar los scores por observación del PCA en base a matriz de
varianzas-covarianzas

head(cp_varcov$scores)
```

```
##         Comp.1     Comp.2     Comp.3      Comp.4      Comp.5
## [1,] 27.162853  1.0278492  5.0022646  0.93622690 -0.51688356
## [2,] 22.363542 27.5955807  3.0635949 -0.08338126  0.02552809
## [3,] 19.167874  7.9566157 -1.5770026 -2.61077676  0.80391745
## [4,]  9.959001  0.8923731  5.5146952  0.12345373 -0.35579895
## [5,] 10.775593 22.0203437 -0.7562826  0.17996723 -0.41646606
## [6,] 23.283948 -7.9268214  2.7958617 -2.09339284 -0.62252321
```

```r
# Mostrar scores por observación del PCA en base a matriz de
correlaciones

head(cp_cor$scores)
```

```
##        Comp.1      Comp.2      Comp.3      Comp.4       Comp.5
## [1,] 2.813992  0.06282760  0.51434516 -0.37618363 -0.161649397
## [2,] 2.550816  2.57369731  0.42896223  0.01252075  0.083602262
## [3,] 2.079207  0.62112516 -0.12602006  0.51138786  0.430775853
## [4,] 1.093316  0.06328171  0.46145821 -0.35236278 -0.008424496
## [5,] 1.489363  2.13420572 -0.08620983 -0.19530483 -0.097669770
## [6,] 2.780190 -0.79964368 -0.11180511 -0.52796031  0.113681564
```

En cuanto a la interpretación de los resultados anteriores, en primera instancia, el comando `cpaS$loading` aplicado sobre el objeto con el resultado del PCA, arroja como resultado 2 tablas, la primera de ellas contiene los valores de los coeficientes para cada variable, dentro de cada componente principal, mientras que la segunda tabla, contiene para cada componente principal, datos referentes tanto a la proporción de varianza explicada por cada componente principal como a la varianza acumulada en cada componente (cuánta varianza en total de los datos originales se explica al tomar un cierto número de componentes). Además de lo anterior, el comando `summary` arroja como resultado una tabla titulada: importancia de componentes, la cual contiene para cada componente principal, 3 filas de datos, la primer fila contiene los datos de desviación estándar para cada componente principal, mientras que la segunda fila tiene los datos de la proporción de varianza total de los datos explicada por cada componente principal, y la tercera fila tiene datos sobre la varianza acumulada (subtotal de varianza explicado por una cantidad de componentes menor al número de variables originales) explicada al tomar una determinada cantidad de componentes principales, es decir, cuánta varianza de un total del 100%, se logra explicar con un cierto número de componentes principales, generalmente inferior al número de variables iniciales. Por último, en cuanto al comando `cpaS$scores`, éste mismo arroja como resultado igualmente una tabla, en la cual, se tienen los datos referentes a los scores o puntuaciones calculados por cada observación en cada uno de los componentes principales, en otras palabras, los scores son el resultado de evaluar cada una de las observaciones de la base de datos original, en cada uno de los componentes principales, por lo que al final, se tienen tantas columnas como componentes principales se tengan y tantas filas como observaciones haya en la base de datos inicial, por lo que en pocas palabras, los scores son “nuevos datos” derivados de los componentes principales que posteriormente se pueden utilizar para deducir aspectos interesantes e implícitos sobre el comportamiento de los datos.

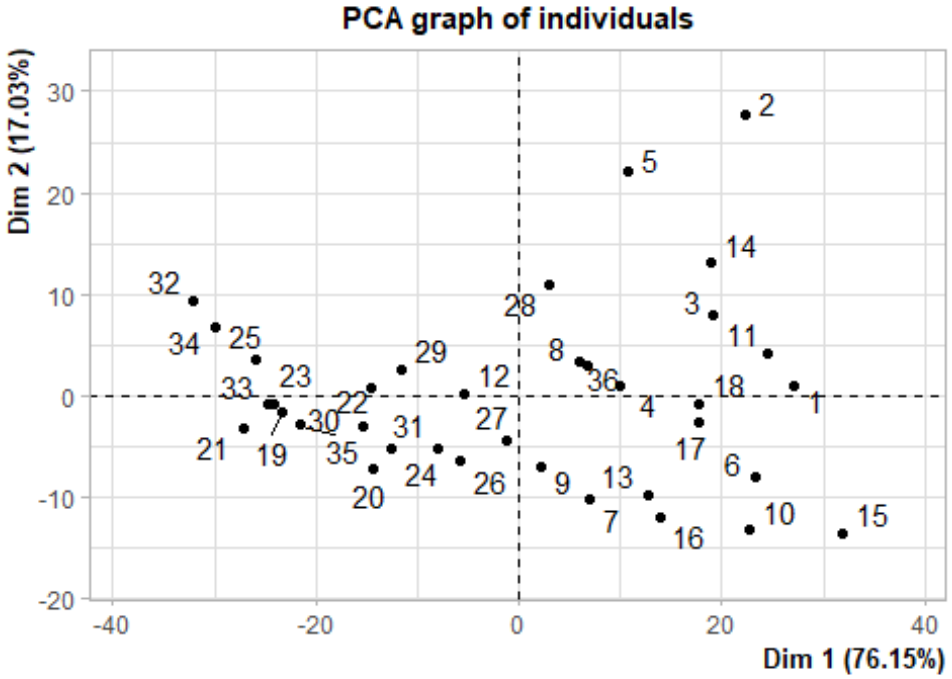
Parte III

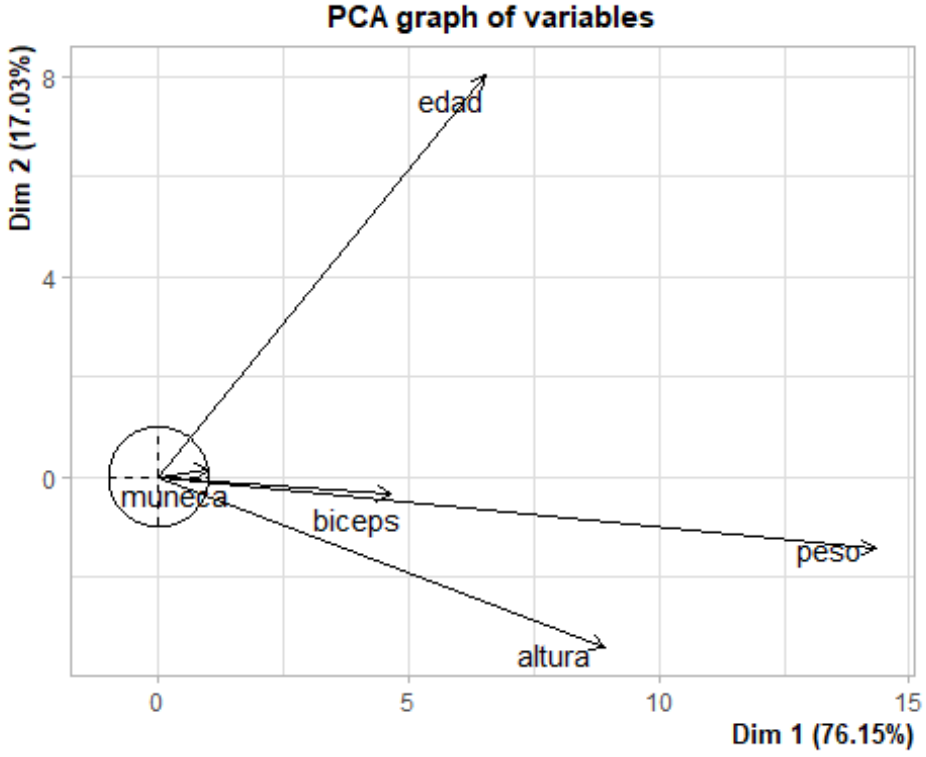
```r
# Librerías FactoMineR y ggplot2

library(FactoMineR)
```

```r
library(ggplot2)

# Librería factoextra

library(factoextra)

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

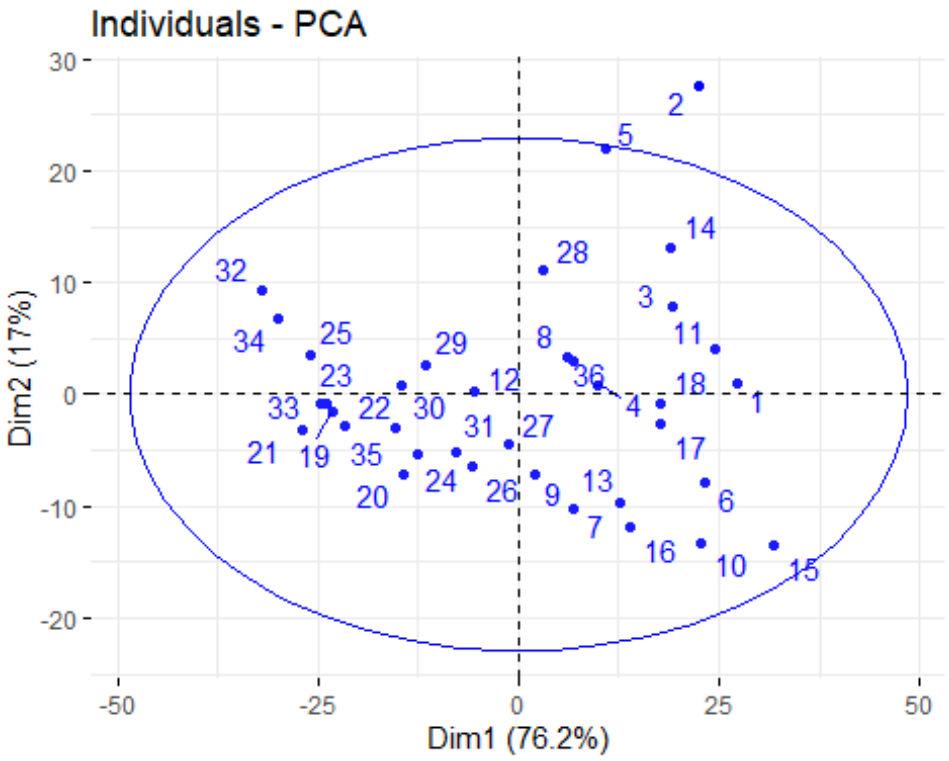
Análisis con matriz de varianzas-covarianzas

```r
# Análisis de componentes principales mediante el comando PCA()

PCA_varcov = PCA(corporal, scale.unit = FALSE)
```

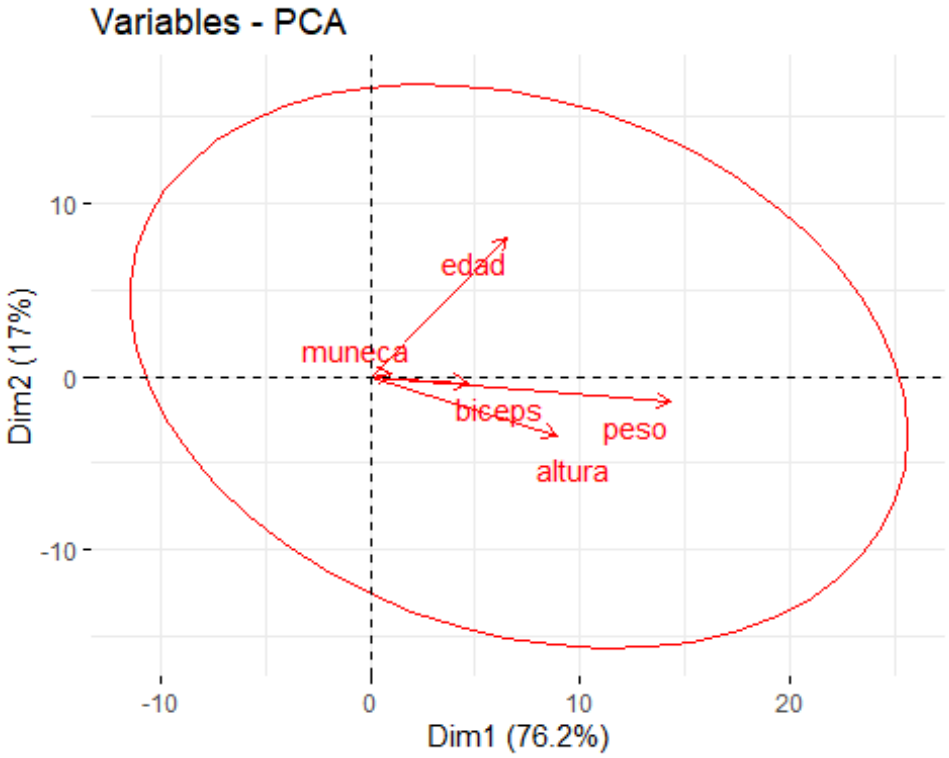




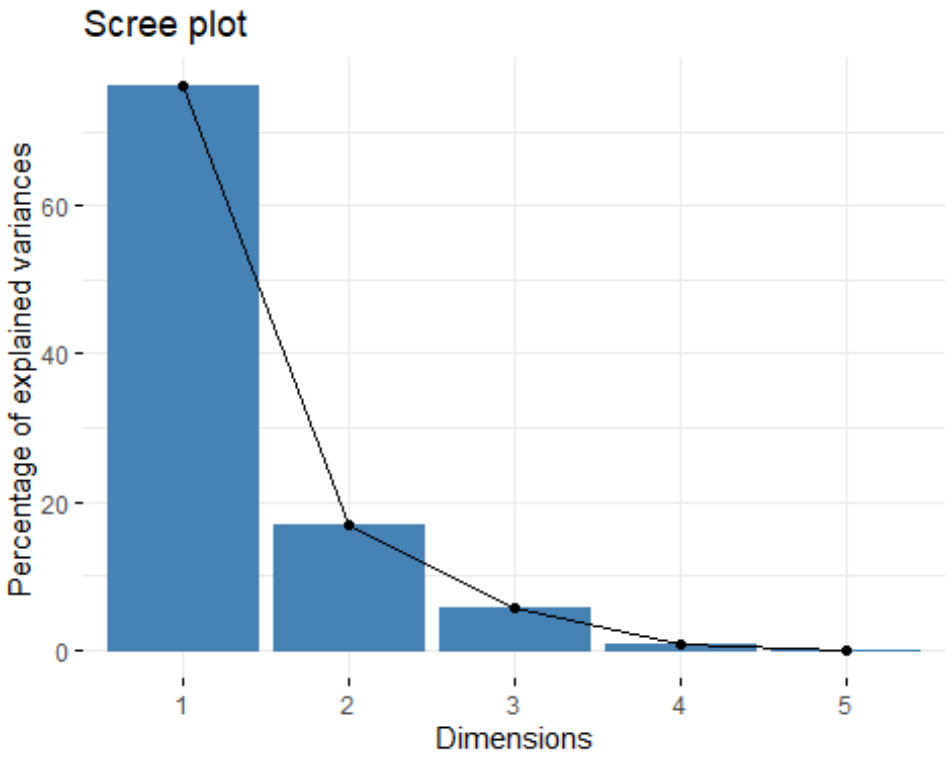

```
# Realizar gráficos

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



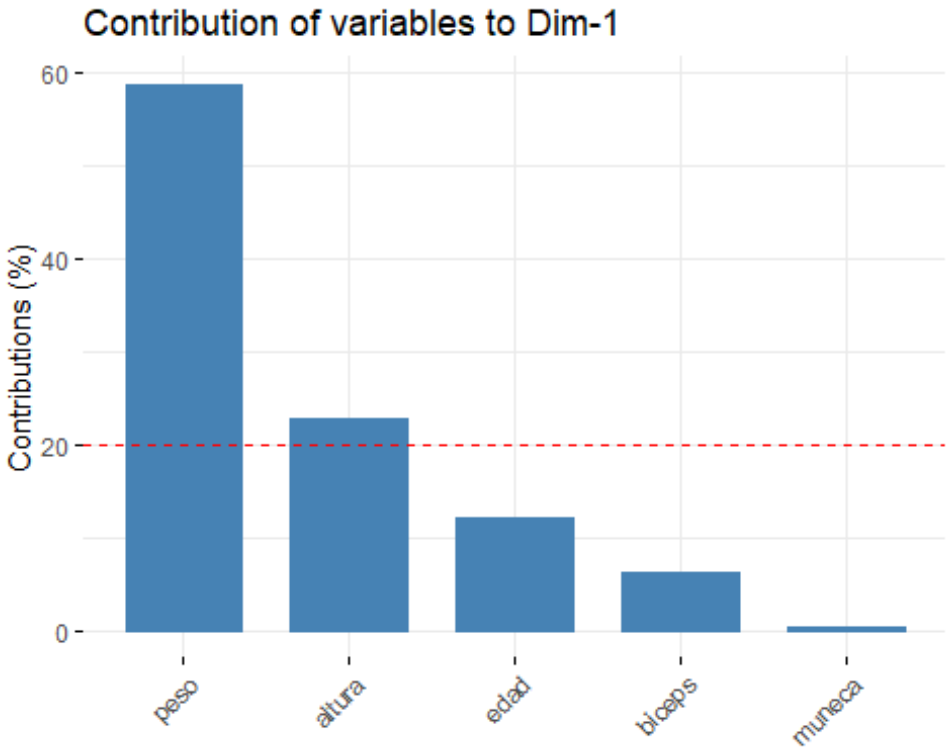

```
fviz_pca_var(PCA_varcov, col.var = "red", addEllipses = TRUE, repel =
TRUE)
```



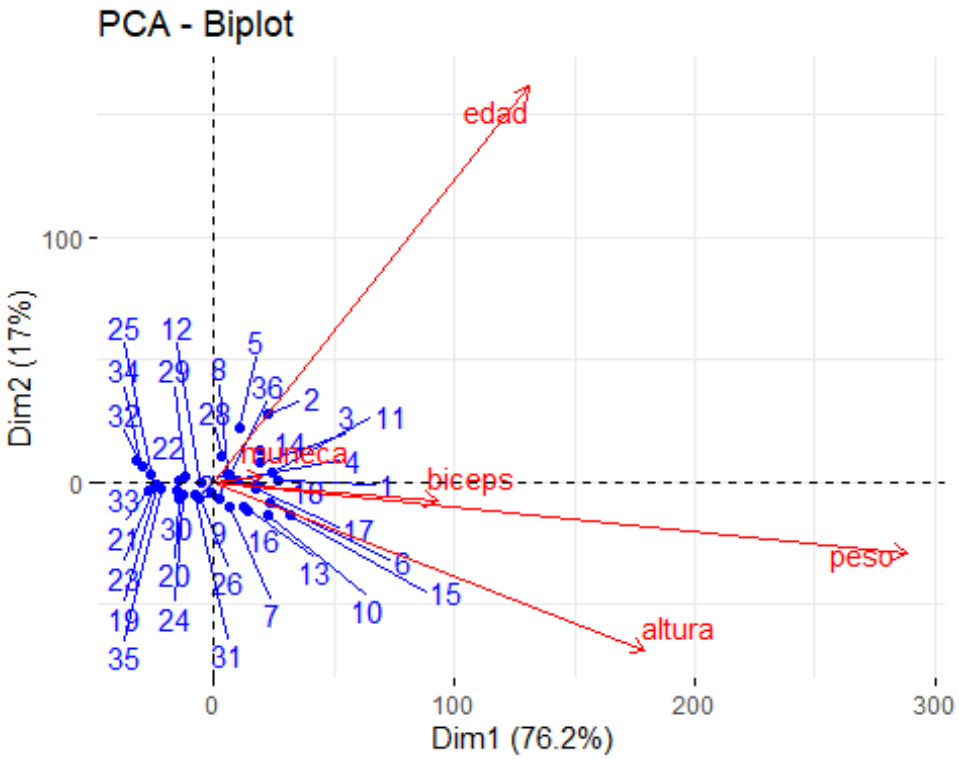

```
fviz_screeplot(PCA_varcov)
```




```r
fviz_contrib(PCA_varcov, choice = c("var"))
```




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



En primer lugar, en cuanto al primer gráfico resultante, en él se están graficando las observaciones originales en base a sus scores tanto en el primer como en el segundo componente principal, es decir, que se calcula el score por cada observación en los primeros 2 componentes principales y a continuación dichos scores se agrupan en pares o coordenadas en 2D, esto se hace para cada observación de la base de datos, por lo que cada observación viene representada por un par de coordenadas de la forma: (score en CP1, score en CP2) que postreriormente se grafican en un plano 2D, por lo que cada observación se mapea a un espacio bidimensional, además, del primer gráfico es posible interpretar que los estudiantes 2 y 5 son aquellos que tienen una edad superior a la de la mayoría de los estudiantes, dado que se observa que éstas 2 observaciones son las que más puntuación tuvieron sobre el componente principal 2, mientras que por el contrario, los estudiantes número 10 y 15 son de los que tienen una edad ligeramente inferior a la de la mayoría de alumnos, esto debido a que las observaciones 10 y 15 son las que se encuentran más abajo en el eje vertical, indicando que tienen las menores puntuaciones en el CP2.

Nota: en cada eje del primer gráfico, también se coloca la proporción porcentual de varianza explicada por el componente principal al que corresponde dicho eje.

Además de lo anterior, en cuanto al segundo gráfico, se están graficando nuevamente los 2 primeros componentes principales con su respectiva proporción porcentual de varianza explicada por cada uno en cada eje del gráfico, además, también se grafican vectores correspondientes a cada una de las variables originales de la base de datos, cuya longitud y sentido señalan el grado de contribución que tiene cada variable en ambos componentes principales y si dicha relación es positiva (al aumentar los

valores de las variables, las puntuaciones también aumentan) o negativa (al aumentar los valores de las variables, las puntuaciones disminuyen y viceversa), ésto último se define observando el sentido en el que apuntan los vectores (izquierda o derecha), por lo que de éste segundo gráfico, es posible interpretar que en el primer componente principal, las variables de mayor contribución o relevancia son el peso y la altura de los estudiantes, mientras que en el segundo componente principal, las variables que más contribuyen son solamente la edad de los alumnos, por lo cual, en este caso, cada componente principal se puede considerar como un grupo de variables, siendo el primer componente principal el grupo asociado al aspecto físico de los alumnos y el segundo componente principal se asocia a la edad de los mismos, por lo que en resumen, los alumnos se pueden describir en base a 2 categorías: aspecto o características físicas y edad.

Adicionalmente, en cuanto al tercer gráfico, en éste se grafica los 2 componentes principales ya mencionados previamente, junto con todas las observaciones originales, y cada observación se grafica en el plano 2D en base a coordenadas formadas por el score de cada observación en el CP1 y el score de esa misma observación en el CP2, por lo que después de graficar las observaciones acorde con este criterio, es posible realizar una clasificación de los estudiantes en función de ambos componentes principales en el sentido de poder determinar cuáles alumnos tienen las mayores edades de todo el grupo y cuáles otros son los menores, por lo que de éste gráfico se puede interpretar que los alumnos 2 y 5 son los mayores de todo el grupo, mientras que los alumnos 10 y 15 se encuentran entre los menores del grupo en cuanto a edad se refiere. De forma similar, en cuanto al cuarto gráfico, en éste se grafican los 2 primeros componentes principales con su correspondiente proporción de varianza explicada en porcentaje, además de 5 vectores correspodientes a cada una de las variables originales, representando el nivel de contribución o relevancia que tiene cada variable en ambos componentes, por lo que considerando ésto, es posible interpretar del cuarto gráfico, que las variables de peso y altura son las más relevantes en el CP1, mientras que en el CP2, la variable más importante de todas es la edad de los alumnos, confirmando de esa manera, las conclusiones del análisis de los gráficos realizados en la parte II.

Por otro lado, el quinto gráfico corresponde a un scree plot o gráfico de sedimentación, en el cual se grafican los 5 componentes principales que se tienen inicialmente antes de reducir la dimensionalidad de los datos (en el eje horizontal), contra el porcentaje de variación explicada por cada componente principal (en el eje vertical), además se traza una curva que pasa encima de las barras correspondientes a los componentes principales que ilustra el cambio en la varianza explicada de los datos originales, al tomar cada cantidad de componentes principales que presenta un comportamiento decreciente, porque a mayor cantidad de componentes principales, cada uno explica una cantidad cada vez menor de varianza total de los datos, por lo que considerando ésto, es posible interpretar de éste gráfico, que es mejor conservar 2 componentes principales, dado que el cambio de mayor magnitud de la varianza explicada se produce cuando se toman 2 componentes, por lo que en caso de tomar más de esa cantidad, los cambios siguientes en la varianza, ya son menores, lo cual

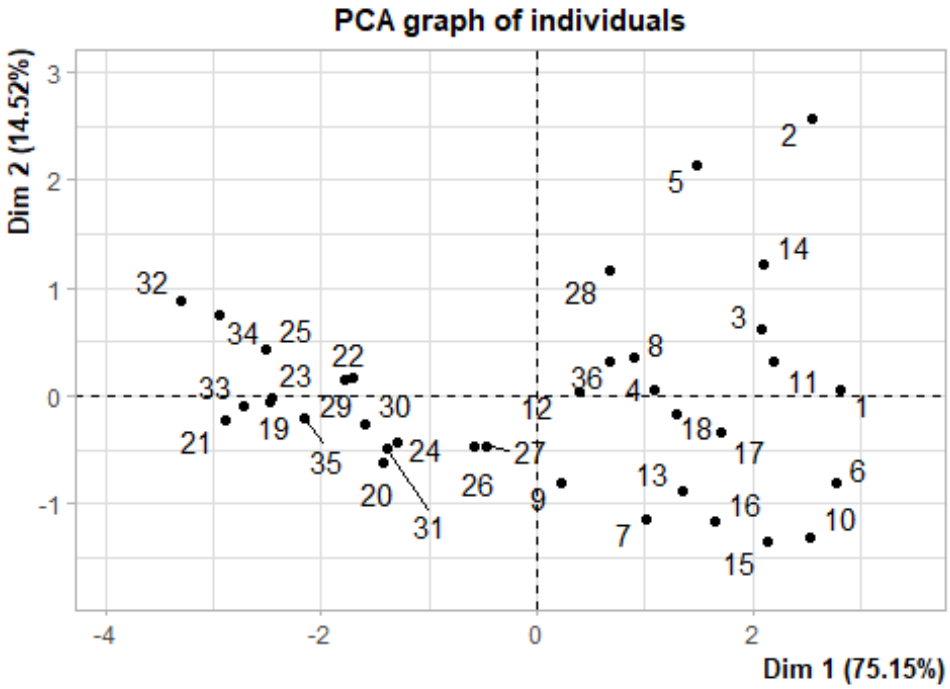
indica que con 2 componentes principales, prácticamente se logra conservar la mayor proporción de varianza de los datos originales, por lo que resulta adecuado reducir la base de datos original a solo 2 componentes principales en vez de las 5 variables originales.

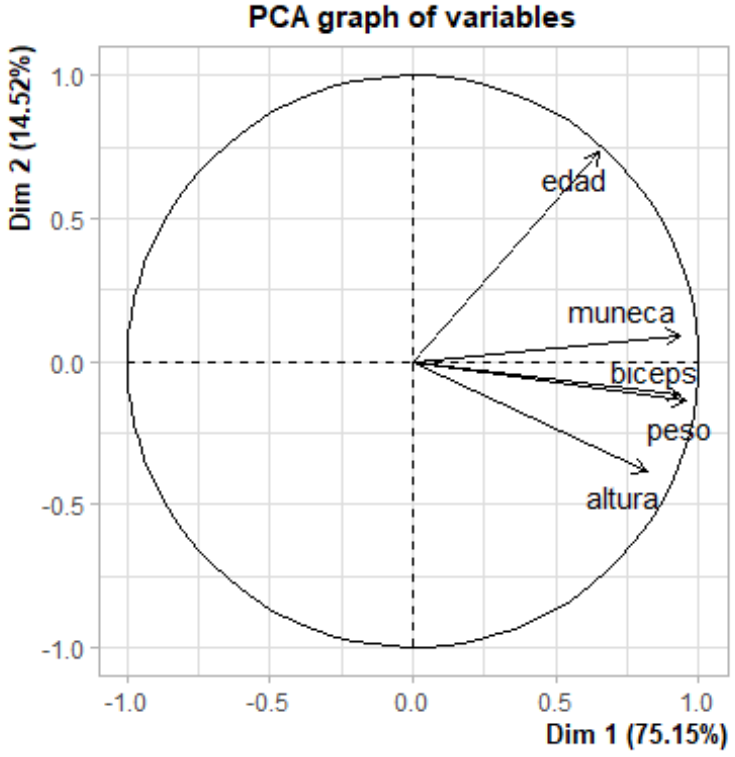
De manera adicional, en cuanto al sexto gráfico, en éste se grafican las 5 variables originales en el eje horizontal, contra el grado de contribución en porcentaje de las variables para el caso del componente principal 1, de forma que se grafica el porcentaje de contribución de cada variable para el CP1 en forma de barras cuya altura depende del porcentaje de contribución de cada variable, además se traza una línea punteada horizontal de color rojo para indicar el mínimo porcentaje de contribución que debe tener una variable para considerarla como relevante para el CP1, que en este caso es del 20%, por lo cual, en el gráfico se aprecia que la variable peso sobrepasa por mucho dicho límite mínimo del 20%, mientras que de forma similar, la variable altura, sobrepasa ligeramente dicho límite, motivo por el cual, es posible interpretar de éste gráfico que el peso y la altura de los alumnos son las variables que más contribuyen al CP1, pudiendo formar un grupo con ellas, como ya se mencionó en los análisis anteriores de ésta actividad.

Finalmente, en relación al séptimo gráfico, en éste se grafican los 3 primeros componentes principales con su respectivo porcentaje de varianza explicada en cada eje del gráfico, además también se grafican las observaciones originales como puntos en color azul, junto con los vectores correspondientes a las 5 variables de inicio en color rojo, que simbolizan el nivel de relevancia de las variables en ambos componentes principales, por lo que en base a ésto, es posible interpretar de éste gráfico, que en particular, las observaciones o alumnos número 1, 2, 5, 8 y 11 se encuentran entre los alumnos que tienen las mayores edades de todo el grupo de estudiantes, además de que también son aquellos que pesan más y son de más alta estatura, además de tener biceps de mayor tamaño también, mientras que por el contrario, los alumnos número 4, 17, 28 y 36, junto con otros más, se ubican entre los estudiantes de menor edad que además tienen una menor medida de muñeca y biceps, además de una menor altura y peso, por lo cual, es posible interpretar también a partir de éste gráfico, que existe una relación directa entre la edad de los estudiantes y las medidas de sus cualidades físicas, es decir, que una mayor edad se relaciona con un mayor tamaño de muñecas y biceps, además de una mayor altura y peso, mientras que por el contrario, si la edad de los alumnos es menor, ésto se asocia también con menores medidas de muñeca y biceps, además de una menor altura y peso.

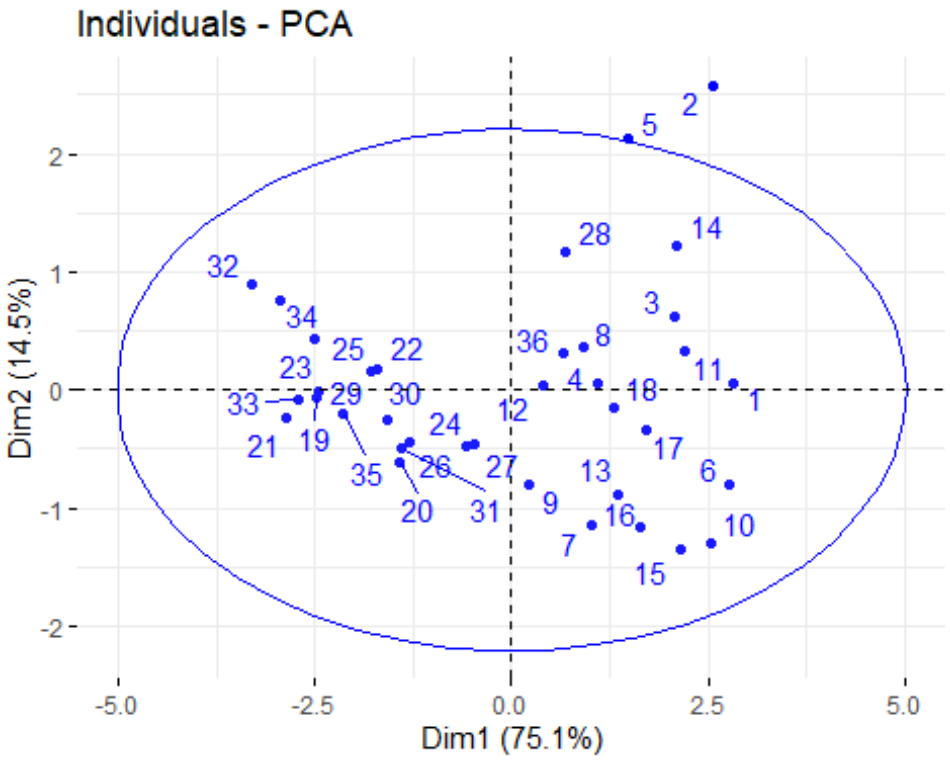
Análisis con matriz de correlaciones

```
# Análisis de componentes principales mediante el comando PCA()

PCA_cor = PCA(corporal, scale.unit = TRUE)
```

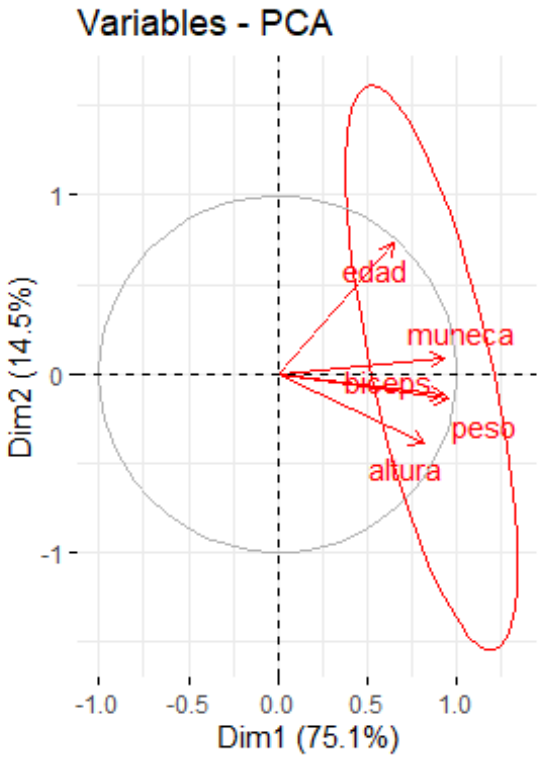




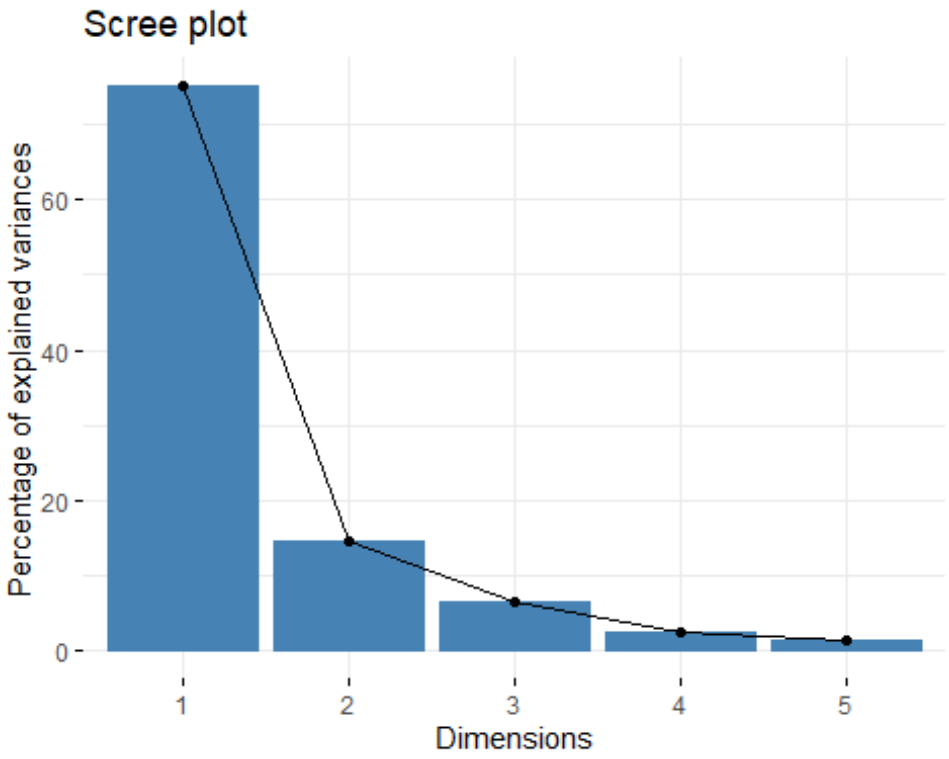

```r
# Realizar gráficos

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



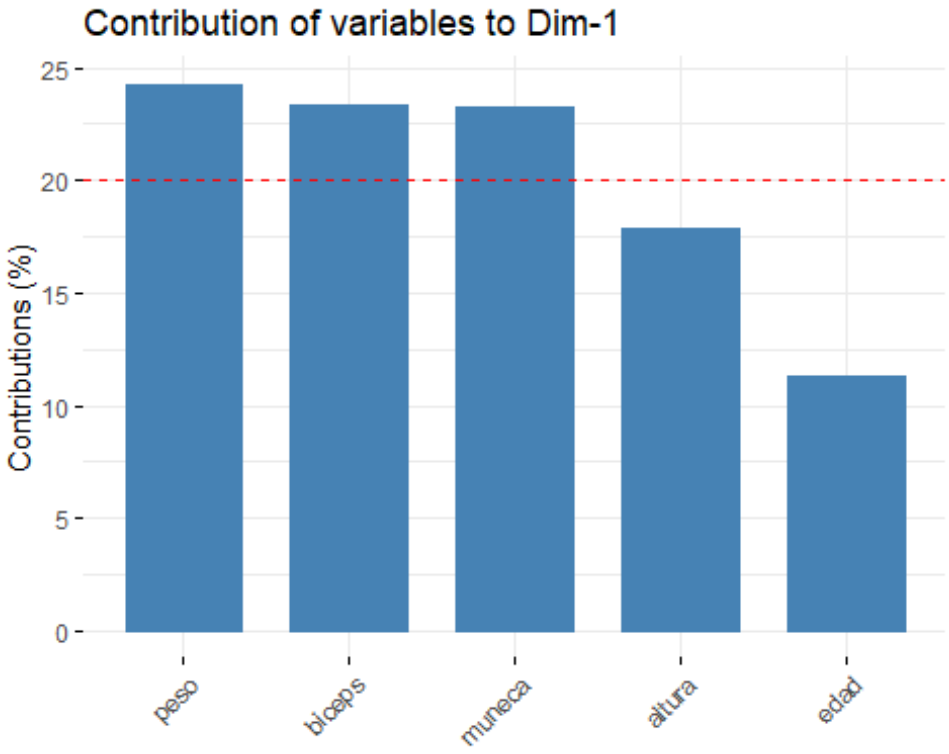

```
fviz_pca_var(PCA_cor, col.var = "red", addEllipses = TRUE, repel = TRUE)
```



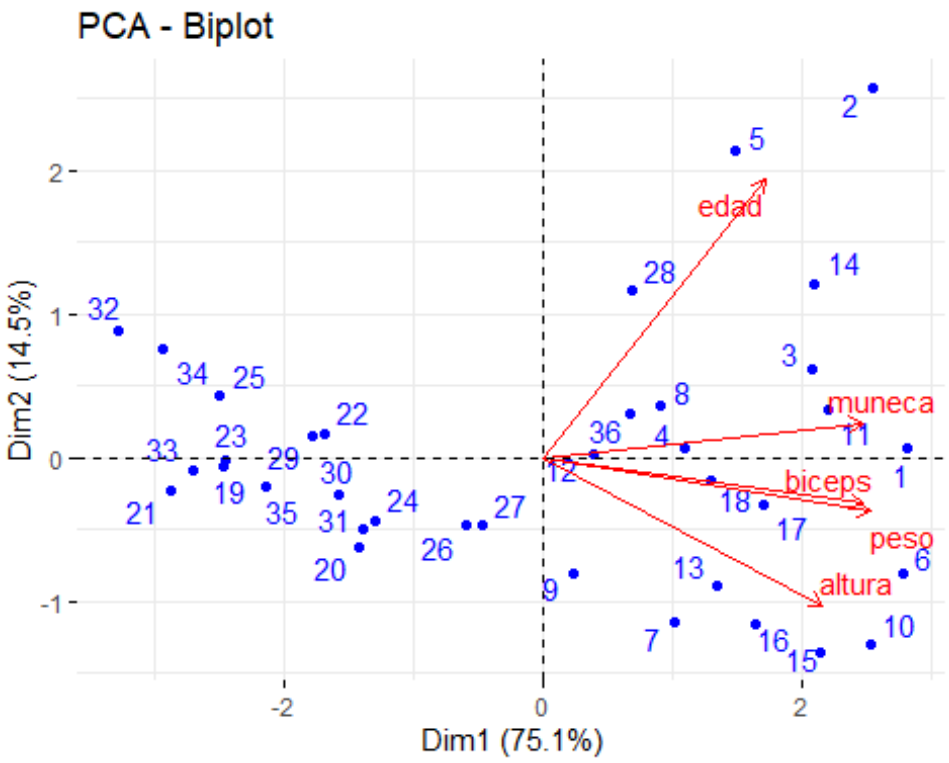

```
fviz_screeplot(PCA_cor)
```




```r
fviz_contrib(PCA_cor, choice = c("var"))
```




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



En primer lugar, en el primer gráfico se grafican los primeros 2 componentes principales con su respectivo porcentaje de varianza total de los datos explicada por los mismos (CP1: 75.15% y CP2: 14.52%), uno en cada eje del gráfico, además de que también se grafican todas las observaciones o alumnos originales en forma de puntos en color negro a lo largo de todo el gráfico, además en el gráfico se puede apreciar que los alumnos número 2 y 5 son de los predominantes en cuanto a edad y medidas de sus cualidades físicas, puesto que dichas observaciones se encuentran en el cuadrante superior derecho del plano 2D, es decir en la región del plano en la que se ubican las mayores puntuaciones del CP1 y del CP2, no obstante por ejemplo, para el caso de los alumnos número 10 y 15, se aprecia en el gráfico que dichos estudiantes están ubicados en el cuadrante inferior derecho del plano 2D, en el que se ubican las menores puntuaciones para ambos componentes principales, lo cual significa que dichos alumnos son de menor edad, además de que también tienen una menor medida de sus características físicas tales como sus muñecas y biceps, además de una menor altura y peso, motivo por el cual, tomando en cuenta lo anterior, es posible interpretar de forma general a partir de éste gráfico, que entre menores sean las puntuaciones de un alumno en ambos componentes principales, éste mismo es de menor edad y por lo tanto posee menores medidas de muñecas, biceps, junto con una menor altura y peso, mientras que si la edad aumenta, también lo hacen las medidas de sus cualidades físicas.

Por otro lado, en relación al segundo gráfico, en este se grafican nuevamente los primeros 2 componentes principales y sus respectivos porcentajes de varianza total explicada por los mismos en cada eje del gráfico, además de que también se grafican los vectores de las variables originales, cuya función consiste en representar el grado

de relevancia o contribución que tienen las variables en ambos componentes principales, lo cual se ve reflejado directamente en la longitud de dichos vectores, por lo cual, tomando esto en consideración, es posible interpretar de éste gráfico, que las 5 variables de inicio guardan entre ellas una correlación positiva, dado que en el gráfico se observa que los vectores de las variables están todos orientados de forma horizontal y apuntando hacia la derecha, lo cual indica que entre mayor sea el valor que adopte una de las variables, las variables restantes también adoptarán valores más grandes, además dado que todos los vectores poseen longitudes mayormente similares, es posible afirmar que todas ellas tienen aproximadamente el mismo nivel de contribución o importancia en el CP1 (donde proyectan la sombra de mayor tamaño), mientras que los coeficientes de todas las variables son notablemente menores en el CP2, ya que se proyecta muy poca sombra sobre el eje correspondiente al mismo, por lo que de éste gráfico se concluye que las 5 variables están correlacionadas positivamente y que si cambian los valores de una variable, los valores del resto de variables cambiarán de la misma manera, es decir, que si los valores de una variable aumentan, también aumentan los valores de las demás variables y viceversa.

Adicionalmente, en cuanto al tercer gráfico, en él se grafican de nuevo los 2 primeros componentes principales con sus correspondientes porcentajes de varianza total explicada por ellos en cada eje del gráfico, además de todas las observaciones originales en forma de puntos en color azul a lo largo de todo el gráfico, además también se grafica una elipse que encierra a la región donde se ubica la mayoría de las observaciones, por lo que en base a ésta información, es posible interpretar de éste gráfico, que de todos los estudiantes registrados, los alumnos número 2 y 5 tienen edades que sobrepasan las de la mayoría de alumnos, ya que las observaciones 2 y 5 se salen del círculo que encierra a la mayoría de observaciones, por lo que los alumnos 2 y 5 al tener mayores edades, también tienen mayores medidas físicas (peso, altura, biceps, muñeca), mientras que por el contrario, los alumnos número 10 y 15 al tener baja puntuación en ambos componentes principales, se determina que dichos estudiantes también tienen una menor edad y en consecuencia menores medidas físicas, motivo por el cual, de éste gráfico también es posible afirmar que menores puntuaciones en los componentes principales se relacionan con menores valores de las características físicas de los alumnos, tal como se concluyó en el gráfico anterior.

Además de lo anterior, en cuanto al cuarto gráfico, en él se grafican los 2 componentes principales como en los gráficos previos, además de los vectores correspondientes a las 5 variables iniciales, junto con una elipse en color rojo que encierra los máximos grados de contribución que pueden llegar a tener las variables en cuestión, por lo que tomando ésto en cuenta, es posible interpretar de éste gráfico, que las 5 variables iniciales se encuentran positivamente correlacionadas entre sí, como ya se había mencionado con anterioridad, además de que todas las variables poseen un grado de relevancia similar en el CP1 y todas tienen una importancia baja en el CP2, por lo que al momento de calcular la puntuación para un determinado estudiante en el CP1, los pesos de las variables estarán mayormente equilibrados, dando lugar a que el resultado de la combinación lineal, igualmente esté equilibrado, lo cual significa que

todas las variables en ella contribuirán de forma aproximadamente equitativa para arrojar dicho resultado, por lo que será necesario que más variables adopten valores grandes para que el score resultante sea igualmente elevado y viceversa, además, dado que las variables no son ortogonales, es posible formar componentes principales que sean en realidad grupos conformados por ellas, mientras que en caso contrario (que las variables fueran ortogonales), se asignaría una sola variable diferente a cada componente principal en lugar de que cada CP esté conformado por todas las variables.

Por otra parte, en cuanto al quinto gráfico, en él se grafican en el eje horizontal los 5 componentes principales que se tienen inicialmente contra el porcentaje de varianza explicada por cada uno de ellos, además, el porcentaje de varianza explicada por cada componente principal se representa mediante una barra cuya altura depende de dicho porcentaje de varianza explicada, además se traza una gráfica de sedimentación que pasa por encima de cada una de las barras comentadas anteriormente, representando el cambio en el porcentaje de varianza explicada, a medida que se toma una mayor cantidad de componentes principales, por lo que tomando lo anterior en cuenta, es posible interpretar de éste gráfico, que la cantidad adecuada a conservar de componentes principales es 2, dado que en el gráfico se observa que la curva en color negro disminuye en mayor magnitud cuando se toman 2 componentes principales, por lo que esa es la cantidad adecuada de componentes principales a conservar y que además logran explicar la gran mayoría de la varianza total de los datos originales.

Adicionalmente, en cuanto al sexto gráfico, en él se grafican las 5 variables originales en el eje horizontal contra el porcentaje de contribución de cada una de dichas variables en el componente principal 1, además también se grafica una recta horizontal punteada de color rojo que representa el umbral o porcentaje de contribución mínimo, para que una de las variables originales, se considere como altamente relevante dentro del CP1, en este caso, dicho límite mínimo es del 20%, motivo por el cual teniendo esto en cuenta, podemos interpretar de éste gráfico, que las variables peso, biceps y muñeca son altamente importantes dentro del CP1, ya que en el gráfico se observa que las barras correspondientes a dichas variables, sobrepasan el límite mínimo del 20% de contribución, lo cual implica que al momento de calcular las puntuaciones para las observaciones a partir del CP1, en caso de que las variables peso, biceps y altura adopten valores grandes, la puntuación para dicha observación también será mayormente grande y al contrario, si las 3 variables mencionadas tienen valores pequños, entonces la observación a la que correspondan también obtendrá una baja puntuación derivada del CP1.

Finalmente, en relación al séptimo gráfico, en él se grafican los 2 primeros componentes principales con su respectivo porcentaje de varianza total explicada, en cada uno de los ejes del gráfico, además, también se grafican todas las observaciones iniciales en forma de puntos en color azul, junto con los vectores en color rojo correspondiente a cada variable origina, motivo por el cual, prácticamente se están graficando las observaciones originales en base a su puntuación en el CP1 y en el CP2 como las coordenadas de ubicación de cada observación dentro del plano 2D, además, en el gráfico también se puede apreciar que los estudiantes número 3, 11, 14, entre

otros, poseen una puntuación positiva en ambos componentes principales, además de que también dichas observaciones se ubican muy a la derecha del gráfico en general, lo cual es un indicativo de que los estudiantes 3, 11, 14 y otros más, son de mayor edad en comparación con la mayoría de los alumnos, por lo que en consecuencia, tienen mayores medidas físicas, tales como una mayor medida de muñecas, biceps, además de mayor altura y peso, por lo cual, de lo anterior es posible interpretar en resumen, que a mayor edad, los alumnos tienen mayores medidas de sus cualidades físicas (biceps, muñecas, altura, peso), mientras que si son de menor edad, sus medidas físicas también son menores, por lo que existe una relación directa (aumento con aumento, o decremento con decremento) entre la edad y las medidas de las características físicas de los alumnos.

Exploración adicional de comando PCA:

Entre todas las opciones que ofrece el comando PCA para análisis de componentes principales, algunas otras que tiene para facilitar la realización de dicho análisis son por ejemplo, el parámetro ncp que sirve para especificar un determinado número de dimensiones o componentes principales para conservar en los resultados del análisis, además del parámetro row.w para especificar los pesos de las filas que representen el grado de importancia que se le otorgará a cada individuo al momento de llevar a cabo los cálculos del análisis y de manera similar, también existe el parámetro col.w para especificar los pesos o niveles de relevancia de las variables para el análisis, sin embargo, el comando PCA cuenta con muchas otras opciones a parte de las ya mencionadas para facilitar la realización del PCA.

Parte IV

Finalmente, comparando los resultados del PCA a partir de la matriz de varianza-covarianza y con los del PCA en base a la matriz de correlación, es posible concluir que el análisis en base a la matriz de varianzas y covarianzas arroja componentes principales con porcentajes de varianza explicada ligeramente mayores que el análisis en base a la matriz de correlación, por lo que en el caso de querer explicar mínimo un 90% de la varianza total de los datos originales, en el caso específico del PCA en base a la matriz de varianzas y covarianzas, será necesario tomar solamente 2 componentes principales para explicar en concreto un 93.18% de dicha varianza total, mientras que en el caso del PCA a partir de la matriz de correlación, es posible observar que los componentes derivados de dicho análisis, poseen porcentajes de varianza explicada menores a los de los componentes del PCA con la matriz de varianza-covarianza, por lo que en este caso, será necesario tomar 3 componentes principales para explicar en concreto un 96.07% de la varianza de los datos iniciales, por lo que en resumen, se concluye que el PCA a partir de la matriz de varianzas y covarianzas es aquel con el que se requieren menos componentes para explicar la gran mayoría de la varianza de los datos de inicio. Además de lo anterior, el procedimiento que aporta componentes de mayor interés es el PCA a partir de la matriz de varianzas y covarianzas, puesto que los componentes de dicho análisis aparte de retener una mayor cantidad de la varianza de los datos originales, también resultan tener niveles de correlación

mayormente altos con los las variables iniciales, lo cual sugiere que es posible agrupar las variables originales en los primeros 2 componentes principales, dado que es la cantidad adecuada de componentes a conservar determinada anteriormente, garantizando una adecuada reducción de la dimensionalidad de la base de datos original, conservando la gran mayoría de la varianza de los datos de origen.

Por otro lado, de los 2 análisis realizados, aquel que resulta ser mejor para explicar las medidas corporales de los estudiantes universitarios es el PCA realizado en base a la matriz de varianzas y covarianzas, ya que es el análisis cuyos componentes retienen mayor proporción de la varianza de los datos originales, y esto con el menor número posible de componentes (2), mientras que con el PCA a partir de la matriz de correlación, se necesitan más componentes (3) para explicar una cantidad aproximadamente igual de varianza de los datos iniciales (mínimo 90% de varianza), por lo que al observar que con el PCA en base a la matriz de varianza y covarianza, es posible explicar dicha cantidad mínima de varianza con menos componentes que en el caso del PCA a partir de la matriz de correlación, se concluye que es mejor el PCA a partir de la matriz de varianza-covarianza para reducir de la forma más efectiva posible, la dimensionalidad de los datos iniciales.

Adicionalmente, aquellas variables que más contribuyen a la primera componente principal CP1 en el método elegido (PCA en base a la matriz de varianza-covarianza) son el peso, seguido de la altura, seguida a su vez de la edad de los estudiantes, dado que dichas variables tienen coeficientes de 58.7, 22.68 y 12.15, respectivamente, siendo los 3 mayores coeficientes de todos en el primer componente principal, mientras que las variables que más contribuyen a la segunda componente principal son la edad seguida de la altura de los alumnos, cuyos coeficientes son de 82.36 y 14.83, respectivamente, siendo éstos los coeficientes más altos de todos los que conforman al componente principal 2.

Combinaciones finaes recomendadas para el PCA:

```
# Combinación lineal del primer componente principal CP1

cat("CP1 = ", PCA_varcov$var$contrib["edad", "Dim.1"], "* edad", " + ",
    PCA_varcov$var$contrib["peso", "Dim.1"], "* peso", " + ",
    PCA_varcov$var$contrib["altura", "Dim.1"], "* altura", " + ",
    PCA_varcov$var$contrib["muneca", "Dim.1"], "* muneca", " + ",
    PCA_varcov$var$contrib["biceps", "Dim.1"], "* biceps")

## CP1 =  12.15987 * edad  +  58.70254 * peso  +  22.68846 * altura  +
0.2901103 * muneca  +  6.159017 * biceps

# Combinación lineal del segundo componente principal CP2

cat("CP2 = ", PCA_varcov$var$contrib["edad", "Dim.2"], "* edad", " + ",
    PCA_varcov$var$contrib["peso", "Dim.2"], "* peso", " + ",
    PCA_varcov$var$contrib["altura", "Dim.2"], "* altura", " + ",
    PCA_varcov$var$contrib["muneca", "Dim.2"], "* muneca", " + ",
    PCA_varcov$var$contrib["biceps", "Dim.2"], "* biceps")
```

```
## CP2 =  82.36471 * edad  +  2.613334 * peso  +  14.83601 * altura  +
0.0241563 * muneca  +  0.1617817 * biceps
```

Finalmente, en términos de agrupación de variables, es posible concluir que el primer componente principal funciona mejor para agrupar a las variables de peso, altura, muñeca y biceps, dado que dichas variables poseen el mayor coeficente dentro de dicho componente principal, a comparación del componente principal 2, donde sus coeficientes son menores, además, de forma similar, en cuanto al componente principal 2, éste funciona mejor para agrupar únicamente a la variable de edad, puesto que dicha variable es la única que posee su mayor coeficiente posible estando en el CP2, mientras que tiene menor coeficiente en el CP1, de tal forma que al final de todo el proceso, se forman 2 grupos principales de variables: edad y físico (engloba a las variables de peso, altura, muñeca y biceps).