

# Clase 4.0

## Análisis

Marcos Rosetti y Luis Pacheco-Cobos

Estadística y Manejo de Datos con R (EMDR) — Virtual

# Análisis agregativos

# Componentes principales

- ¿Cuál es la función de un análisis de componentes principales (PCA)?
- Reducir el número de variables al identificar aquellas que nos dan información redundante o agrupar variables dentro de una categoría más amplia.
- Combinar ambos análisis (por individuos y por variables) para caracterizar grupos de individuos por variables.

# Componentes principales

```
install.packages("FactoMineR")  
library(FactoMineR) # Es un paquete de muchos disponibles  
?decathlon
```

```
data(decathlon)  
head(decathlon)
```

```
##      100m Long.jump Shot.put High.jump 400m 110m.hurdle Discus Pole.vault  
## SEBRLE 11.04      7.58    14.83     2.07 49.81      14.69 43.75      5.02  
## CLAY   10.76      7.40    14.26     1.86 49.37      14.05 50.72      4.92  
## KARPOV 11.02      7.30    14.77     2.04 48.37      14.09 48.95      4.92  
## BERNARD 11.02      7.23    14.25     1.92 48.93      14.99 40.87      5.32  
## YURKOV 11.34      7.09    15.19     2.10 50.42      15.31 46.26      4.72  
## WARNERS 11.11      7.60    14.31     1.98 48.68      14.23 41.10      4.92  
##      Javeline 1500m Rank Points Competition  
## SEBRLE 63.19 291.7 1 8217 Decastar  
## CLAY 60.15 301.5 2 8122 Decastar  
## KARPOV 50.31 300.2 3 8099 Decastar  
## BERNARD 62.77 280.1 4 8067 Decastar  
## YURKOV 63.44 276.4 5 8036 Decastar  
## WARNERS 51.77 278.1 6 8030 Decastar
```

# Componentes principales

- Preguntas para un PCA
- ¿Cómo se agrupan los sujetos?
  - ¿Qué ejes describen mejor los agrupamientos de atletas con desempeños similares?
- ¿Cómo se agrupan las variables?
  - ¿Qué ejes describen mejor los agrupamientos de variables similares?

# Componentes principales

```
res.pca <- PCA(decathlon[,1:10], scale.unit=TRUE, ncp=5, graph=F)
str(res.pca)
```

```
## List of 5
## $ eig : num [1:10, 1:3] 3.272 1.737 1.405 1.057 0.685 ...
## ..- attr(*, "dimnames")=List of 2
## .. ..$ : chr [1:10] "comp 1" "comp 2" "comp 3" "comp 4" ...
## .. ..$ : chr [1:3] "eigenvalue" "percentage of variance" "cumulative percentage of variance"
## $ var :List of 4
## ..$ coord : num [1:10, 1:5] -0.775 0.742 0.623 0.572 -0.68 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:10] "100m" "Long.jump" "Shot.put" "High.jump" ...
## .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ cor : num [1:10, 1:5] -0.775 0.742 0.623 0.572 -0.68 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:10] "100m" "Long.jump" "Shot.put" "High.jump" ...
## .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ cos2 : num [1:10, 1:5] 0.6 0.55 0.388 0.327 0.462 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:10] "100m" "Long.jump" "Shot.put" "High.jump" ...
## .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ contrib: num [1:10, 1:5] 18.3 16.8 11.8 10 14.1 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:10] "100m" "Long.jump" "Shot.put" "High.jump" ...
## .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## $ ind :List of 4
## ..$ coord : num [1:41, 1:5] 0.792 1.235 1.358 -0.61 -0.586 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:41] "SEBRLE" "CLAY" "KARPOV" "BERNARD" ...
## .. .. ..$ : chr [1:5] "Dim.1" "Dim.2" "Dim.3" "Dim.4" ...
## ..$ cos2 : num [1:41, 1:5] 0.1117 0.124 0.1599 0.0487 0.0377 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:41] "SEBRLE" "CLAY" "KARPOV" "BERNARD" ...
```

# Componentes principales

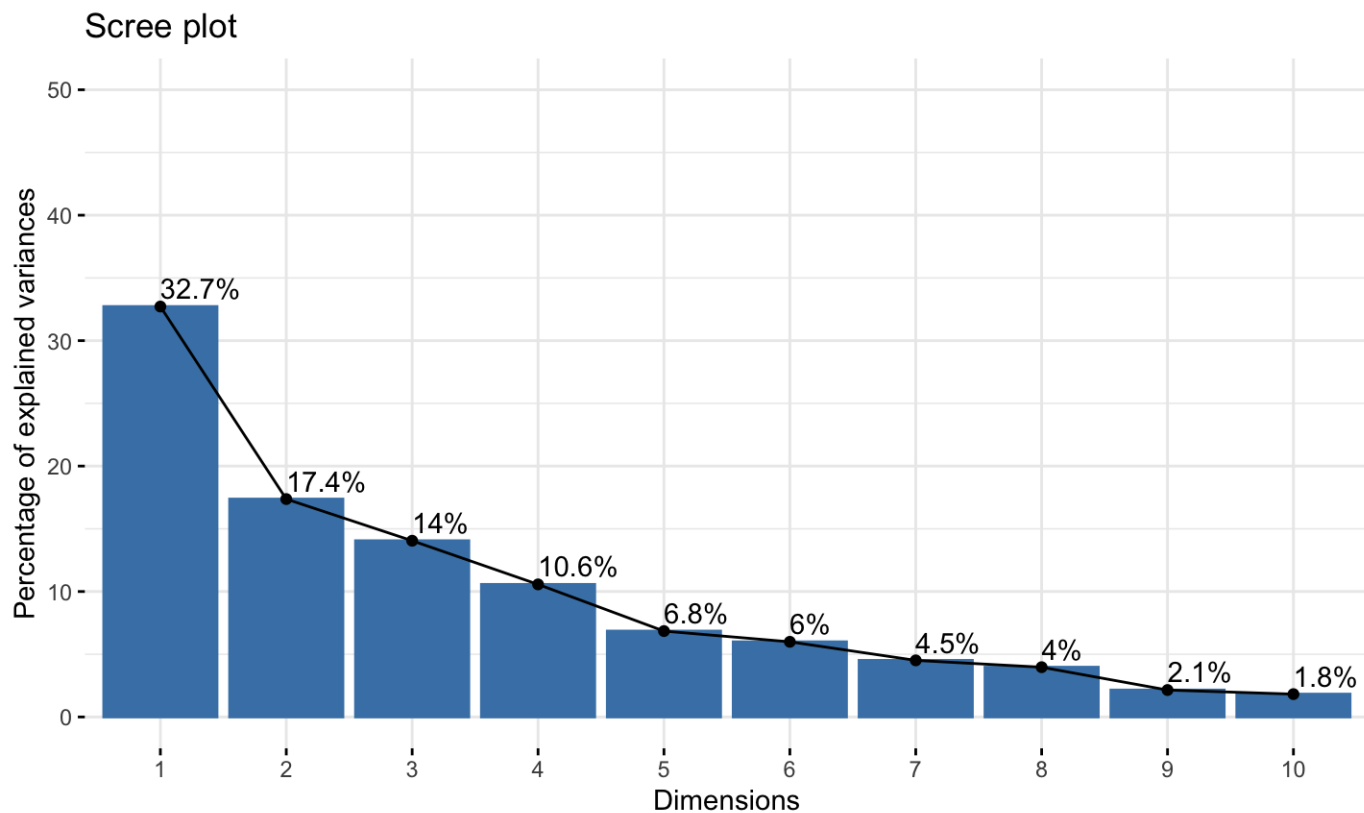
```
eig.val <- get_eigenvalue(res.pca)  
eig.val
```

```
##      eigenvalue variance.percent cumulative.variance.percent  
## Dim.1      3.2719055          32.719055          32.71906  
## Dim.2      1.7371310          17.371310          50.09037  
## Dim.3      1.4049167          14.049167          64.13953  
## Dim.4      1.0568504          10.568504          74.70804  
## Dim.5      0.6847735           6.847735          81.55577  
## Dim.6      0.5992687           5.992687          87.54846  
## Dim.7      0.4512353           4.512353          92.06081  
## Dim.8      0.3968766           3.968766          96.02958  
## Dim.9      0.2148149           2.148149          98.17773  
## Dim.10     0.1822275           1.822275         100.00000
```

# Componentes principales

```
fviz_eig(res.pca, addlabels = TRUE, ylim = c(0, 50))
```

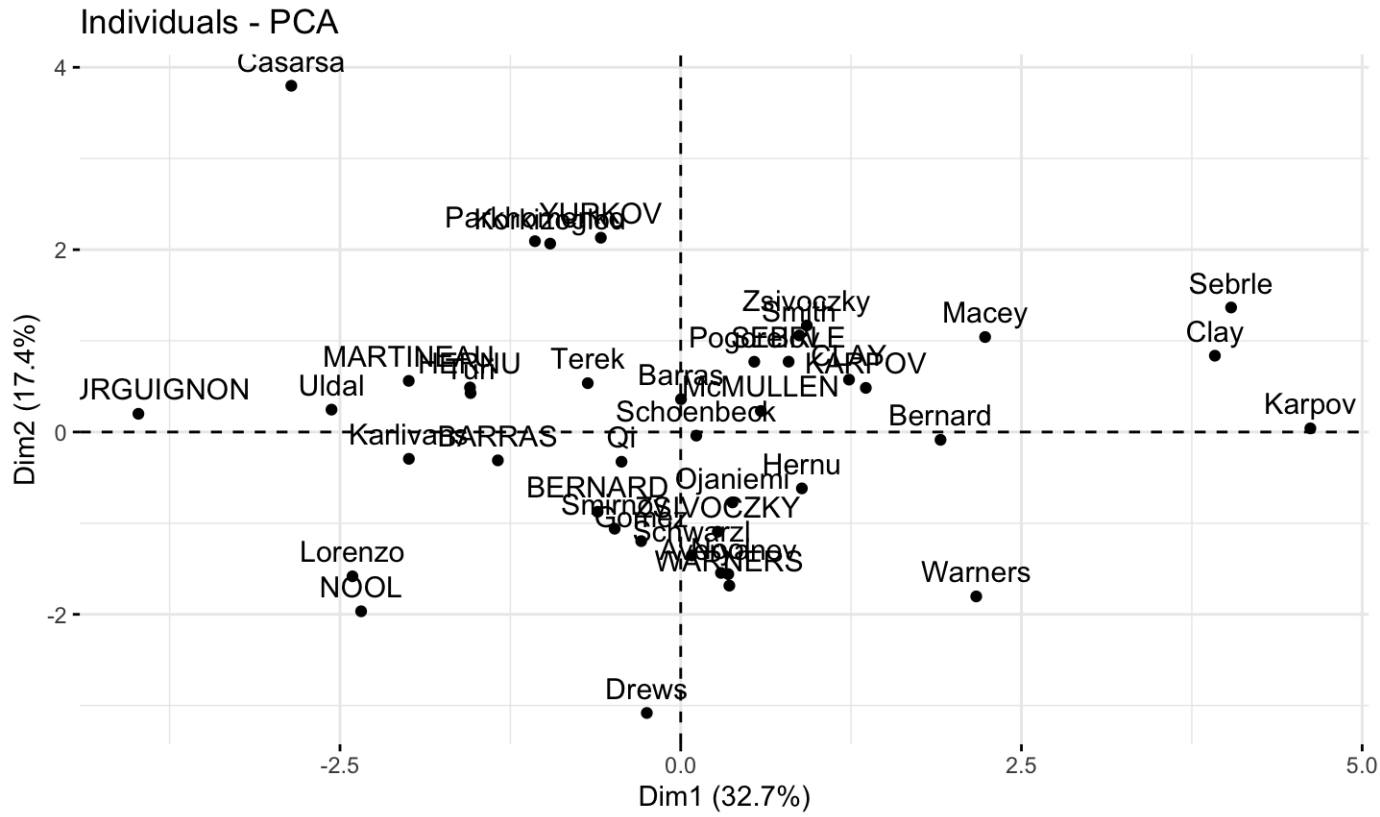
```
## Registered S3 methods overwritten by 'car':  
##   method                                from  
##   influence.merMod                       lme4  
##   cooks.distance.influence.merMod       lme4  
##   dfbeta.influence.merMod               lme4  
##   dfbetas.influence.merMod              lme4
```





# Componentes principales

```
fviz_pca_ind(res.pca)
```



# Componentes principales

- Individuos más extremos en Dim1: BOURGUIGNON contra Karpov.

```
names <- c("BOURGUIGNON", "Karpov")
decathlon[row.names(decathlon) %in% names, ]
```

```
##           100m Long.jump Shot.put High.jump 400m 110m.hurdle Discus
## BOURGUIGNON 11.36      6.80    13.46     1.86 51.16      15.67 40.49
## Karpov      10.50      7.81    15.93     2.09 46.81      13.97 51.65
##           Pole.vault Javeline 1500m Rank Points Competition
## BOURGUIGNON 5.02      54.68 291.70  13  7313      Decastar
## Karpov      4.60      55.54 278.11  3  8725      OlympicG
```

# Componentes principales

- Individuos más extremos en Dim2: YURKOV contra Drews.

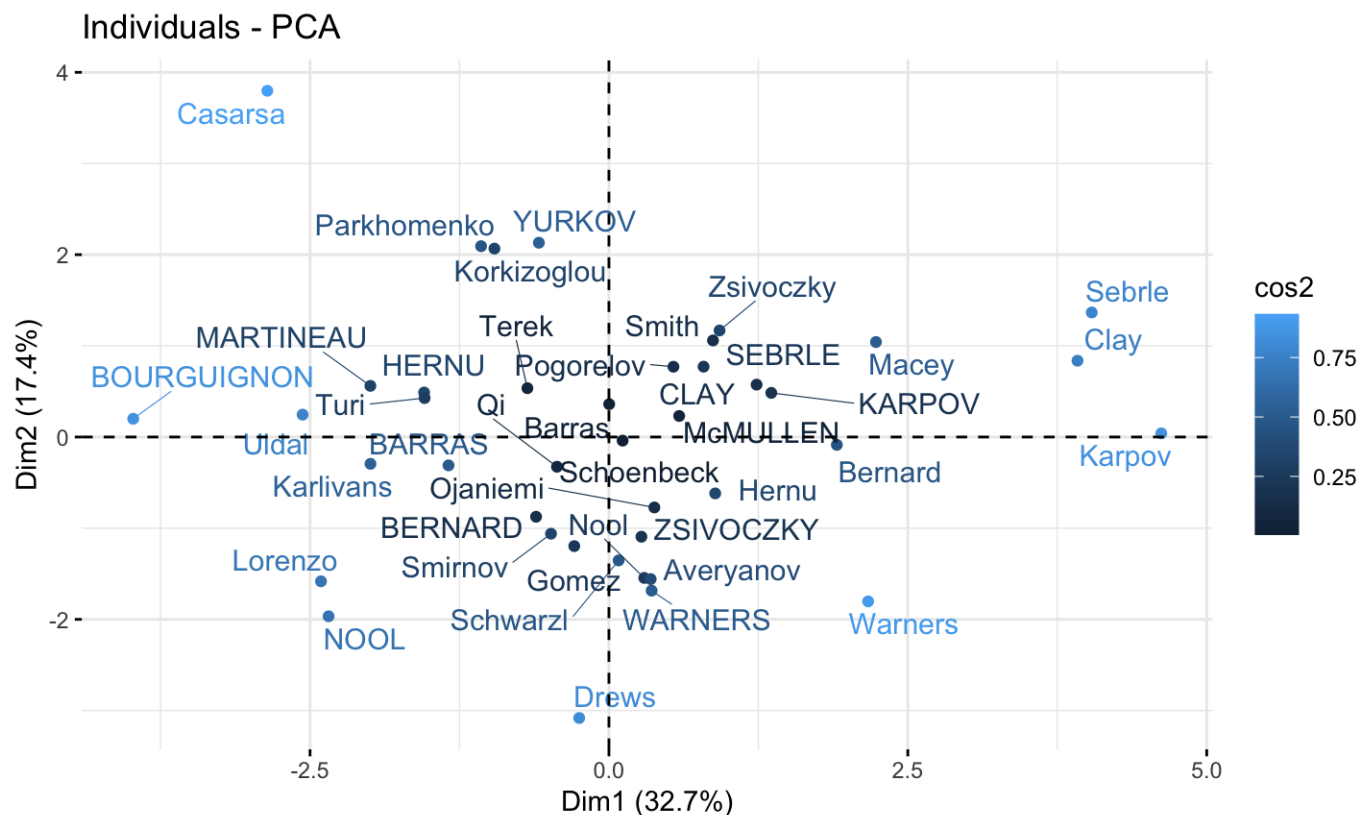
```
names <- c("YURKOV", "Drews")
decathlon[row.names(decathlon) %in% names, ]
```

```
##           100m Long.jump Shot.put High.jump 400m 110m.hurdle Discus Pole.vault
## YURKOV 11.34      7.09    15.19     2.10 50.42      15.31 46.26      4.72
## Drews 10.87      7.38    13.07     1.88 48.51      14.01 40.11      5.00
##           Javeline 1500m Rank Points Competition
## YURKOV    63.44 276.40     5   8036     Decastar
## Drews    51.53 274.21    19   7926     OlympicG
```

# Componentes principales

- Podemos trazar 4 divisiones, rápidos y fuertes (*Sebrle*), lentos (*Casarsa*), rápidos y débiles (*Warners*) y lentos y débiles (*Lorenzo*).

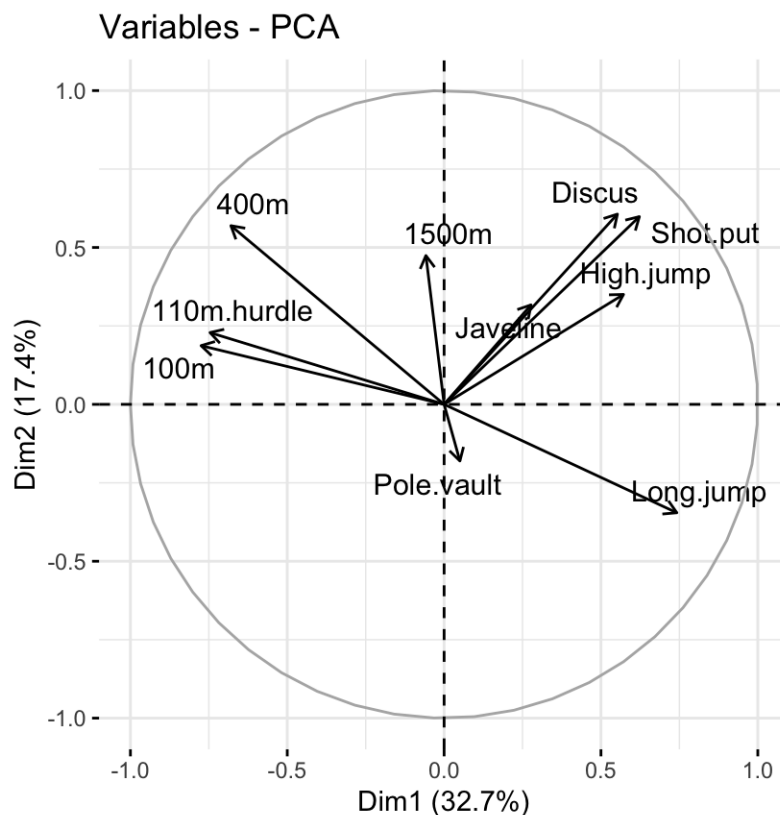
```
fviz_pca_ind(res.pca, col.ind = "cos2", repel = TRUE)
```



# Componentes principales

- Las variables `Discus`, `Shot.put` y `High.jump` no tienen mucha correlación con `X100m`, `X400m`, `X110m.hurdle` y `Long.jump`.

```
fviz_pca_var(res.pca, repel = TRUE)
```



<http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/>

# Licencia CC BY



Estadística y Manejo de Datos con R (EMDR) por Marcos F. Rosetti S. y Luis Pacheco-Cobos se distribuye bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).