

# Guía para AF y ACP

Miguel Hernández

11/8/2019

## Introducción

Antes de realizar un análisis, lo importante es tener clara la formulación del problema. En este caso trabajaremos con investigar con detectar las materias de matemáticas y ciencias naturales pertenecen a un grupo distinto que materias de francés y latín.

En el cuadro ?? se muestran las notas de las materias de 8 alumnos, las cuales se presentan de la siguiente manera:

Cuadro 1: Calificaciones de materias			
Matemáticas	Ciencias naturales	Francés	Latín
9	8	6	7
10	9	10	10
3	5	9	8
9	9	8	8
7	6	3	5
5	5	5	5
5	5	7	6
4	4	3	4

La investigación seguirá la siguiente estructura:

- Análisis de matriz de correlación
- Extracción de factores
- Determinación del número de factores
- Rotación de factores
- Interpretación de factores
- Validación del modelo
- Cálculo de puntuaciones factoriales
- Selección de variables representativas
- Análisis posterior

## Proceso en R

### Exploración de datos

Importamos datos que vienen en formato SPSS.

```
CL = foreign::read.spss("data/Ciencias-Letras TOY EJEMPLO.sav", to.data.frame = T)
```

Calculamos la media y la desviación estándar de cada variable.

```
# Media
```

```
summary(CL)
```

```
##      Matematicas      Naturales      Francés      Latín
## Min.       : 3.00   Min.       :4.000   Min.       : 3.000   Min.       : 4.000
## 1st Qu.: 4.75   1st Qu.:5.000   1st Qu.: 4.500   1st Qu.: 5.000
## Median : 6.00   Median :5.500   Median : 6.500   Median : 6.500
## Mean      : 6.50   Mean      :6.375   Mean      : 6.375   Mean      : 6.625
## 3rd Qu.: 9.00   3rd Qu.:8.250   3rd Qu.: 8.250   3rd Qu.: 8.000
## Max.      :10.00   Max.      :9.000   Max.      :10.000   Max.      :10.000
```

```
# Desviación estándar
```

```
print("Desviación estándar")
```

```
## [1] "Desviación estándar"
```

```
apply(CL, 2, sd)
```

```
## Matematicas  Naturales  Francés  Latín
##    2.618615    1.995531    2.615203    1.995531
```

Matriz de correlaciones

```
CL_cor = cor(CL)
```

```
CL_cor
```

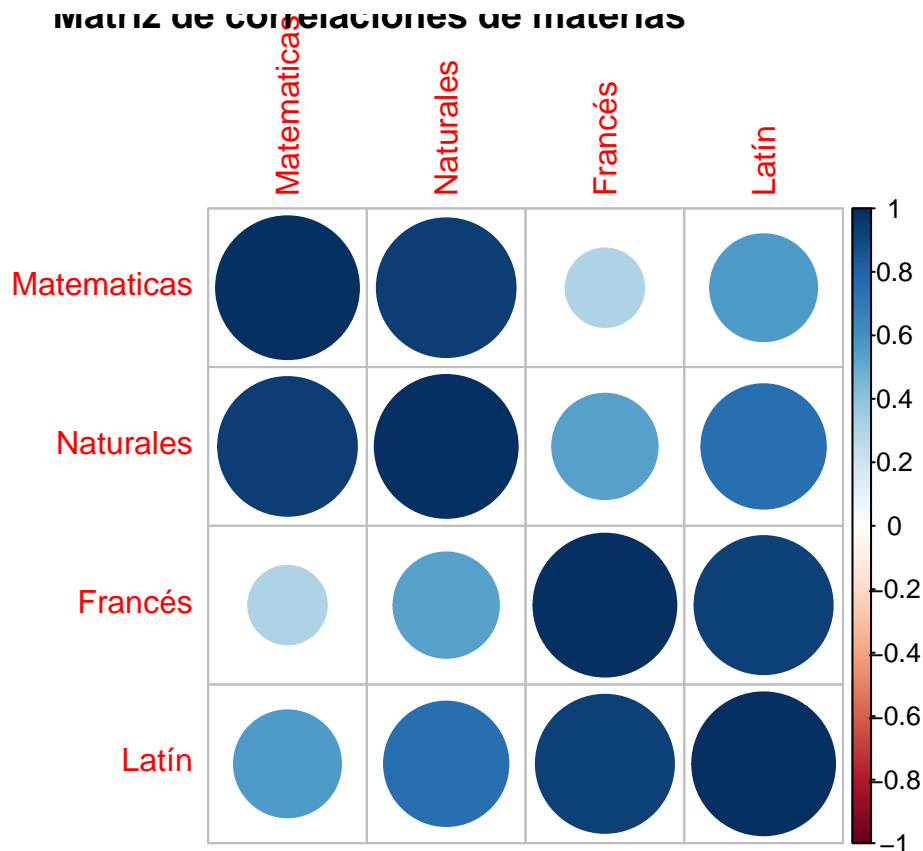
```
##      Matematicas Naturales Francés Latín
## Matematicas    1.0000000 0.9431723 0.3024774 0.5604357
## Naturales      0.9431723 1.0000000 0.5440581 0.7578475
## Francés        0.3024774 0.5440581 1.0000000 0.9341375
## Latín          0.5604357 0.7578475 0.9341375 1.0000000
```

Relación entre variables

La primera aproximación es ver si existen relaciones entre las variables que esperamos que pertenezcan a sus grupos. En este caso, Matemáticas y Ciencias naturales están correlacionados; análogamente Francés y Latín.

También se puede ver gráficamente mediante el paquete **corrplot**, en donde se pone la matriz de correlaciones **CL\_cor**

```
corrplot::corrplot(CL_cor, title = "Matriz de correlaciones de materias")
```



Otra vía de verlo es ver si en general las variables están relacionadas mediante el determinante aplicando la función `det()` a la matriz de correlaciones `CL_cor`. Si el determinante se acerca a 0, significa que hay relación entre las variables y podemos seguir explorando si el análisis de factores es adecuado.

```
det(CL_cor)
```

```
## [1] 0.001294676
```

En efecto, el determinante se acerca a cero y podemos continuar.

Prueba de esfericidad

Queremos descartar si los datos tienen una forma esférica. Es decir, una esfericidad completa se presenta mediante la matriz:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Entonces, para descartar que los datos se comportan de esa forma, utilizamos la prueba de esfericidad de Bartlett, en donde la  $H_0$  es que los datos presentan esfericidad completa. Es decir, mide el grado en que la matriz se desvía de la matriz de identidad  $\mathbf{R}$ . Para ello usamos la función `cortest.bartlett()` del paquete `psych`. Los argumentos que requiere esta función son la matriz de correlaciones `CL_cor` y el número de observaciones, que lo podemos calcular con `nrow()`.