

# Análisis de Componentes principales

---

## Ejercicios propuestos

1. Se trata de comprender los resultados de un análisis de componentes principales en un caso sencillo, que incluso puede realizarse a mano.

Considerando el siguiente conjunto de datos:

	$X_1$	$X_2$
A	2	1
B	5	4
C	3	4
D	5	2

- (a) Representar gráficamente el conjunto de datos, y predecir el resultado de un ACP con las variables originales
  - (b) Efectuar el ACP con variables originales, y comprobar las propiedades que deben cumplir los resultados.
    - Representar gráficamente los puntos mediante las componentes.
    - Comprobar que las componentes están centradas en el origen.
    - Comprobar que la suma de las varianzas de las componentes es la suma de la traza de la matriz de covarianzas.
    - Calcular el porcentaje de varianza explicada por cada componente.
    - Comprobar que los factores están incorrelados mutuamente.
    - Calcular las correlaciones con las variables originales.
    - Expresar cada componente como una combinación lineal de las variables originales.
2. El fichero de datos *pint.txt* contiene las valoraciones que ciertos expertos han hecho sobre las obras de pintores de reconocido prestigio.

Realizar un ACP sobre las variables cuantitativas. Decidir si realizar sobre la matriz de correlaciones o covarianzas el análisis.

    - (a) Obtener la descomposición de la variabilidad por componentes en términos de porcentaje.
    - (b) Interpretar las dos primeras componentes.
    - (c) Representar los autores según los valores de las dos primeras componentes.

- (d) Representar los autores según los valores de las dos primeras componentes mostrando la escuela a la que pertenecen. Utiliza colores.
3. A continuación se muestran la base ( $B$ ) y altura ( $A$ ) en centímetros de  $n = 6$  rectángulos.

$B$	$A$
2	2
1.5	0.5
0.7	0.5
0.5	1.5
0.5	0.7
0.7	0.7

- (a) Dibujar los 6 rectángulos (a mano).
- (b) Transformar los datos en la matriz  $Z = (\log(B), \log(A))$  y calcular las componentes principales sobre la matriz de varianzas-covarianzas.
- (c) Calcular los valores propios de la matriz de varianzas-covarianzas de  $Z$ . ¿Qué porcentaje de variabilidad explica cada componente?
- (d) Ordenar los valores de la primera componente de manera creciente. Dibujar los rectángulos en ese orden. ¿Qué representa la primera componente?
- (e) Ordenar los valores de la segunda componente de manera creciente. Dibujar los rectángulos en ese orden. ¿Qué representa la segunda componente?
4. Se han considerado 9 *documentos*, de los cuales 5 ( $d_1$ - $d_5$ ) tratan de “interacción entre persona y computador” y los otros 4 ( $d_6$ - $d_9$ ) tratan sobre “teoría de grafos”. Consideremos *términos* las palabras que aparecen al menos en dos documentos distintos.

$d_1$ : Human machine interface for Lab ABC computer applications.

$d_2$ : A survey of user opinion of computer system response time.

$d_3$ : The EPS user interface management system.

$d_4$ : System and human system engineering testing of EPS.

$d_5$ : Relation of user-perceived response time to error measurement.

$d_6$ : The generation of random, binary, unordered trees.

$d_7$ : The intersection graph of paths in trees.

$d_8$ : Graph minors IV: Widths of trees and well-quasi-ordering.

$d_9$ : Graph minors: A survey.

Fuente: Zelaia, A., Baragaña, I. eta Yurramendi, Y. (2003) LSaren oinarri matematikoa. *Ekaia* **17**:85–105.

Construye una matrix  $X$  (*documentos*  $\times$  *términos*) donde  $x_{ij}$  recoge cuántas veces aparece el *término*  $t_j$  en el documento  $d_i$ .

- (a) Calcula las 2 primeras componentes principales e interprétalas.
- (b) Representa los documentos según los valores de las dos primeras componentes.
- (c) Supongamos que tenemos un nuevo documento  $d_0$  “Graph theory with applications to engineering and computer science”. Proyéctalo en el plano construido por las 2 primeras componentes. ¿Dónde se clasificaría este documento, en el ámbito “interacción entre persona y computador” o en “teoría de grafos”?