

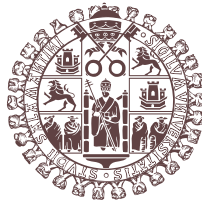
Universidad de Salamanca  
Grado en Matemáticas

---

REVISIÓN DE MÉTODOS  
MULTIVARIANTES  
SUPERVISADOS Y NO SUPERVISADOS

---

Trabajo Fin de Grado



VNiVERSiDAD  
D SALAMANCA

Alumno: Pedro Ángel Fraile Manzano

Tutoras: Ana Belén Nieto Librero y Nerea González  
García

Salamanca, Julio de 2023

# Índice general

<b>Introducción</b>	<b>3</b>
<b>1. Conceptos Estadística Multivariante</b>	<b>7</b>
<b>2. Análisis de Componentes Principales</b>	<b>9</b>
2.1. Introducción . . . . .	9
2.2. Definición y cálculo de las Componentes . . . . .	10
2.3. Reducción de la dimensionalidad . . . . .	12
<b>Bibliografía</b>	<b>13</b>



# Introducción

El análisis multivariante se define como la rama del análisis estadístico que interpreta de manera simultánea la relación entre más de dos variables.

En la primera parte de la memoria se abordarán los llamados métodos supervisados. Estos son aquellos que, dado un conjunto de variables de entrada observadas  $X_1 \dots X_p$  nos permiten predecir de distintas maneras una variable de salida. En este caso, el conjunto de datos recogidos y con los que se “entrena” al modelo contienen la observación de nuestra variable objetivo.

Dentro de este tipo de métodos tendríamos métodos tan variados como los más simples métodos de regresión lineal multivariante, hasta las más complejas redes neuronales que podamos construir.

Por otro lado, los métodos no supervisados buscan relaciones entre las variables, de modo que no tenemos en el conjunto de entrenamiento ninguna información de cómo de correcto o incorrecto es lo que estamos afirmando. En este tipo de métodos entraría el apartado del análisis de componentes principales o



# Capítulo 1

## Conceptos Estadística Multivariante

A lo largo de esta primera parte estableceremos los conceptos básicos que se usarán a lo largo de la memoria todo ello basado en [1]

**Definición 1.0.1.** Dada una población de  $n$  individuos de los que estudiamos  $p$  variables, estableceremos que:

$x_{rj}$  = Valor que toma la  $j$ -ésima variable en el  $r$ -ésimo individuo

**Definición 1.0.2.** La matriz  $X = x_{ij}$  de dimensión  $(n \times p)$  es conocida como la *matriz de datos*

**Definición 1.0.3.** Dadas  $X_1 \dots X_p$  variables aleatorias, llamaremos vector aleatorio al vector  $\mathbf{X}^T = [X_1 \dots X_p]$ .

*Observación:* La matriz de datos y el vector aleatorio no son lo mismo, de ahí la diferencia en la tipografía.

**Definición 1.0.4.** Dado un vector aleatorio  $\mathbf{X}$ , llamaremos vector de medias  $\mu^T = [\mu_1 \dots \mu_p]$  donde  $\mu_i = \mathbb{E}(X_i)$

**Definición 1.0.5.** Dado un vector aleatorio  $\mathbf{X}$ , llamamos matriz de covarianzas a la matriz  $\Sigma$  donde los coeficientes son los siguientes  $s_{ij} = Cov(X_i, X_j)$  por tanto, la diagonal de esta matriz son las varianzas de las variables. Análogamente podemos definir la *matriz de correlaciones*,  $P$ .

**Proposición 1.0.1.** Dada la matriz covarianzas  $\Sigma$





## Capítulo 2

# Análisis de Componentes Principales

### 2.1. Introducción

Sea una población de la que se han tomado  $n$  muestras de las cuales hemos medido  $p$  variables. Del conjunto de datos que resulta, podemos dar la matriz de datos  $\mathbf{X}$  de tamaño  $n \times p$ .

En consecuencia, las muestras  $y_i$ , con  $i = 1 \dots n$  se pueden interpretar como elementos de  $\mathbb{R}^p$ , pero si  $p > 3$ , la representación gráfica de estos datos no se puede realizar

El análisis de componentes principales, aunque normalmente se utiliza con el objetivo de representar los datos de manera sencilla, también podría llegar a ser útil en la reducción de la dimensionalidad para evitar el overfitting en el aprendizaje automático.

Este método es una técnica matemática que dado un vector aleatorio  $\mathbf{X}^T = [X_1, \dots, X_p]$ , con vector de media  $\mu$  y matriz de covarianzas  $\Sigma$ , utiliza transformaciones ortogonales para conseguir las componentes principales. Aunque también es realizable con la media poblacional  $\bar{x}$  y la matriz de varianzas  $\mathbf{S}$  de manera análoga.

Estas componentes principales son combinaciones lineales de las variables que forman el vector, de manera que estando correladas las iniciales, las componentes no lo están y se busca calcularlas con la máxima varianza posible.

Esta transformación se busca ya que si varias variables están altamente correladas, entonces están aportando la misma información, siempre que no

sean dependientes. Podemos encontrar por esta razón variables que no estén correladas, lo que implica que no dan la misma información y que en consecuencia pueden aportar mayor información acerca de la variación de los datos, sin tener que ser compartida o repetida por varias variables.

Al finalizar este Análisis de Componentes Principales, obtendremos un conjunto de variables nuevas no correladas entre sí, que son combinación lineal de las iniciales que maximizan la varianza en cada paso.

Añadir que si las variables no están correladas o están cerca de no estarlo, el Análisis de Componentes Principales no tiene sentido ya que el conjunto de componentes principales será parecido a las variables iniciales, con la única diferencia que estarán ordenadas por orden creciente de varianza.

## 2.2. Definición y cálculo de las Componentes

Sea un vector aleatorio  $\mathbf{X}^T = [X_1, \dots, X_p]$  con media  $\mu$  y matriz de covarianzas  $\Sigma$ .

**Definición 2.2.1.** Las componentes principales son combinaciones lineales de las variables  $X_1 \dots X_p$

$$\mathbf{Y}_j = a_{1j}X_1 + \dots a_{pj}X_p = \mathbf{a}_j^T \mathbf{X} \quad (2.1)$$

Donde  $\mathbf{a}_j$  es un vector de constantes y la variable  $\mathbf{Y}_j$  cumple lo siguiente:

- Si  $j = 1$   $Var(\mathbf{Y}_1)$  es máxima restringido a  $\mathbf{a}_1^T \mathbf{a}_1 = 1$
- Si  $j > 1$  debe cumplir:
  - $Cov(\mathbf{Y}_j, \mathbf{Y}_i) = 0 \quad \forall i < j$
  - $\mathbf{a}_j^T \mathbf{a}_j = 1$
  - $Var(\mathbf{Y}_j)$  es máxima.

El cálculo de la primera componente principal se lleva a cabo con un proceso de optimización de la función  $Var(\mathbf{Y}_1)$  sujeto a la restricción de que  $\mathbf{a}_1^T \mathbf{a}_1 = 1$ . Aplicando el método de los multiplicadores de Lagrange, dada una función  $f(\mathbf{x}) = f(x_1 \dots x_p)$  diferenciable con una restricción  $g(\mathbf{x}) = g(x_1 \dots x_p) = c$  entonces existe una constante  $\lambda$  de manera la ecuación:

$$\frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i} = 0 \quad i = 1, \dots, p \quad (2.2)$$

Tiene como solución los puntos estacionarios de  $f(\mathbf{x})$

Sea ahora la función  $L(\mathbf{x}) = f(\mathbf{x}) - \lambda[g(x) - c]$  entonces podemos simplificar la expresión anterior a:

$$\frac{\partial L}{\partial \mathbf{x}} = 0 \quad (2.3)$$

En nuestro caso particular  $L(\mathbf{a}_1) = \mathbf{a}_1^T \Sigma \mathbf{a}_1 - \lambda[\mathbf{a}_1^T \mathbf{a}_1 - 1]$ . Al derivarla obtenemos que :

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{a}_1} &= 2\Sigma \mathbf{a}_1 - 2\lambda \mathbf{a}_1 \\ &= 2(\Sigma - \lambda) \mathbf{a}_1 \end{aligned}$$

Igualando a 0 tenemos la siguiente ecuación:

$$(\Sigma - \lambda I) \mathbf{a}_1 = 0 \quad (2.4)$$

Para que la ecuación tenga una solución que no sea la trivial, tenemos que elegir  $\lambda$  de manera que  $|\Sigma - \lambda I| = 0$ . Luego  $\lambda$  es uno de los valores propios de la matriz. Generalmente una matriz  $(p \times p)$  tiene  $p$  valores propios  $\{\lambda_1, \dots, \lambda_p\}$  y como  $Var(\mathbf{Y}_1) = Var(\mathbf{a}_1^T \mathbf{X}) = \mathbf{a}_1^T \Sigma \mathbf{a}_1 = \mathbf{a}_1^T \lambda \mathbf{a}_1 = \lambda$  que es la variable a maximizar, elegimos  $\lambda = \max\{\lambda_1, \dots, \lambda_p\}$ , por tanto, el vector  $\mathbf{a}_1$  es el vector propio con valor propio  $\lambda = \lambda_1$  reordenando si es necesario.

Una vez calculada la primera componente principal  $\mathbf{Y}_1$ , la segunda componente se calcula de manera análoga, maximizando  $Var(\mathbf{Y}_2) = Var(\mathbf{a}_2^T \mathbf{X})$  condicionada por  $\mathbf{a}_2^T \mathbf{a}_2 = 1$ . A esta restricción tenemos que añadir la restricción  $Cov(\mathbf{Y}_1, \mathbf{Y}_2) = 0$

**Proposición 2.2.1.** La condición  $Cov(\mathbf{Y}_1, \mathbf{Y}_2) = 0$  equivale a la condición  $\mathbf{a}_2^T \mathbf{a}_1 = 0$

*Demostración.* Utilizando que  $\mathbf{Y}_j = \mathbf{a}_j^T \mathbf{X} \quad \forall j$ , tenemos entonces que:

$$\begin{aligned} Cov(\mathbf{Y}_2, \mathbf{Y}_1) &= Cov(\mathbf{a}_2^T \mathbf{X}, \mathbf{a}_1^T \mathbf{X}) \\ &= E(\mathbf{a}_2^T (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T \mathbf{a}_1) \\ &= \mathbf{a}_2^T E((\mathbf{X} - \mu)(\mathbf{X} - \mu)^T) \mathbf{a}_1 \\ &= \mathbf{a}_2^T \Sigma \mathbf{a}_1 \\ &= \mathbf{a}_2^T \lambda_1 \mathbf{a}_1 \end{aligned}$$

De manera que, si  $\mathbf{a}_2^T \lambda_1 \mathbf{a}_1 = 0 \Rightarrow \mathbf{a}_2^T \mathbf{a}_1 = 0$ , luego son vectores ortogonales entre sí.  $\square$

*Observación:* Esta proposición se puede extender de manera simple al caso de tener que calcular la  $i$ -ésima componente principal habiendo calculado las anteriores de las cuales sepamos sus valores propios.

**Corolario 2.2.1.** Las componentes principales son todas ortogonales entre sí.

Tomando la matriz formada por los  $p$  vectores propios como columnas tenemos la matriz ortogonal  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_p]$ , de manera que el vector aleatorio

$$\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_p]^T = \mathbf{A}\mathbf{X}$$

De esta manera también obtenemos la diagonalización de la matriz de covarianzas

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} \quad (2.5)$$

Y también podemos deducir que

## 2.3. Reducción de la dimensionalidad

Sea  $\mathbf{X} = [X_1, \dots, X_p]$  una matriz de datos multivariantes de tamaño  $(n \times p)$  de las cuales ya hemos obtenido las componentes principales  $Y_1 \dots Y_p$ . Cada individuo de los  $n$  que forman las filas de la matriz puede ser interpretado como un elemento de  $\mathbb{R}^p$ .

**Definición 2.3.1.** Llamaremos matriz de distancias entre las  $n$  filas o individuos, y lo denotaremos  $\Delta$  la matriz

# Bibliografía

- [1] **Chatfield,C y Collins A.J (1989).** *Introduction to multivariate analysis*, Chapman and Hall.
- [2] **Jolliffe I.T.(1986).** *Principal Component Analysis*, Springer-Verlag.
- [3] **Hastie, T.,Tibshirani, R. y Friedman J. (2001),** *The Elements of Statistical Learning, Data Mining, Inference and Prediction* Springer
- [4] **Cuadras, C.M.(1991)**