

# Tema 1: Introducción al Análisis Multivariante y al Cálculo Matricial

## Introducción al Análisis Multivariante

Su origen histórico se encuentra en los primeros años del siglo XX. Surge dentro del marco de la psicología aplicada como una teoría matemática que trata de explicar el concepto de *inteligencia*. Es decir, se supone que la inteligencia constituye un compendio de diversas habilidades y conocimientos y se suele medir mediante aspectos o manifestaciones parciales. Spearman (1904) y Pearson (1901) trataron de definir una variable que midiese la *cantidad de inteligencia* y que fuese un compendio o resumen (de hecho una combinación lineal) de los componentes de la misma. Esto sería el origen de lo que luego se denominó el método de los *Componentes Principales*. Posteriormente se han ido desarrollando numerosas técnicas para variables tanto cuantitativas como categóricas.

El análisis multivariante, en esencia, se dedica al estudio de varias variables de modo simultáneo. Es decir, tomamos un objeto y no sólo medimos un aspecto suyo (e.g. una persona a la que se mide sólo su altura), sino que consideramos varios aspectos y tratamos de determinar la relación entre estas medidas. Es decir, medimos además de la altura, el peso, y además indicamos su sexo, cuál es la clase social a la que pertenece y cuál es su renta por año. Además, no sólo nos interesan los valores en cada caso, sino también las relaciones simultáneas entre ellas.

Con el desarrollo de la Informática, se ha hecho posible desarrollar e implementar programas estadísticos que contienen las técnicas multivariantes; así, todos los programas de este tipo contienen una parte importante dedicada a estas técnicas (e.g. se puede ver

en R, STATGRAPHICS, SPSS, ...).

En definitiva, el desarrollo teórico surgido en el siglo XX junto con las aplicaciones crecientes de la estadística en la vida económica de los países han hecho de las técnicas del Análisis Multivariante junto con el Análisis de Regresión, uno de los instrumentos más empleados para el estudio del entorno ambiental, económico y social.

## Tipos de variables

Uno de los problemas fundamentales en Estadística es cómo medir los aspectos de las personas, seres vivos u objetos. Es decir, no es lo mismo tomar una población cualquiera y medir la altura de las personas en dos clases: altos y bajos, que tomar una escala métrica y dividirla en segmentos, asignando a cada persona el número o medida en cm. En un caso tendremos, en realidad, una variable categórica (con dos categorías: altos y bajos) y en el otro, una variable cuantitativa (1, 70; 1, 65; ...). En el primer caso no tendrá sentido hallar una media (*bajo* – *alto*) pero sí una moda y en el otro, sí será congruente hablar de la altura *media*.

Nadie pondrá en duda que determinar la medida de las personas en altos o bajos es menos informativo que tomar una escala métrica. De hecho, en Estadística, las técnicas que se pueden aplicar varían según sea la información recogida por las variables. De la misma forma, se puede decir lo mismo en Análisis Multivariante. Técnicas como el análisis discriminante se aplican en variables cuantitativas distribuidas como una distribución normal, mientras que el análisis log-lineal se aplica en variables categóricas en exclusiva.

Como posible clasificación, según el grado de información que contienen unas variables, se pueden dividir a éstas en:

**(i) Variables Nominales:** Sólo distinguen entre varias categorías, sin que exista ninguna jerarquía entre ellas.

Ejemplos: la variable *sexo*: mujer, hombre. La variable *colores del arco iris*: azul, violeta, amarillo, ...

Se pueden recodificar con números, aunque no tengan un sentido algebraico: mujer = 1, hombre = 0.

No tiene sentido hallar medias o varianzas. Tan sólo modas, números de casos y las llamadas relaciones de contingencia.

### **(ii) Variables Ordinales**

Además de distinguir distintas categorías para una variable, se puede distinguir una relación de orden entre ellas. Por ejemplo, la variable tamaño de letra en un procesador de texto: menuda, pequeña, normal, grande y extragrande.

Podríamos recodificarla como 1, 2, 3, 4 y 5 y establecer una relación de *orden*:  $1 < 2 < 3 < 4 < 5$ .

Sin embargo, no se tiene la misma diferencia entre grande y extragrande  $5 - 4 = 1$ , que entre pequeña y menuda  $2 - 1 = 1$ , aunque los números coincidan. Sólo se puede decir que una es mayor que la otra. Es decir, la diferencia entre los valores de la variable no tiene sentido. Además, no existe origen en la escala de las medidas (por ej. tamaño 0).

### **(iii) Variables de Intervalo**

Además de contener las características de las dos anteriores (distingue entre valores y entre la distinta jerarquía de valores) añade el hecho de dotar de sentido a la diferencia entre los valores de la variable. Es decir, la distancia o diferencia entre dos valores consecutivos de la variable es siempre el mismo. Un ejemplo típico sería la variable *temperatura*.

Sin embargo, estas variables no tienen un origen en la medida. Por ejemplo,  $0^{\circ}\text{C}$  es el punto de congelación del agua pura, no la ausencia de temperatura.

### **(iv) Variables de razón**

Son idénticas a las anteriores salvo que presentan un origen absoluto de medida. En estas variables tiene sentido tomar fracciones de sus valores o *razones*. Se puede decir que un valor es el doble que otro.

Por ej. la edad expresada en años: el 0 tendría el sentido de una persona no nacida.

Se puede observar que la información recogida por las variables va creciendo desde las nominales a las de razón. Siempre es posible pasar de más información a menos: una variable de intervalo o de razón se puede dividir en *trozos* (o intervalos) y convertirla en nominal. El paso contrario no es posible.

## Clasificación de las Técnicas Multivariables

Las técnicas multivariables se pueden clasificar según dos posibles criterios:

(i) Se está interesado en la asociación entre las distintas variables, es decir, en las *relaciones* entre las mismas, donde parte de estas variables dependen o se miden en función de las otras. Son los llamados *Métodos Dependientes*. Subyace en ellos siempre un interés predictivo.

(ii) Se está interesado en investigar las asociaciones que se presentan entre variables sin distinción de tipos entre ellas. Son *Métodos Independientes*. Tienen un interés descriptivo más bien.

### Métodos Dependientes

**Regresión múltiple:** Estudia la dependencia de una variable en función de otras variables.

**Análisis discriminante:** Se busca una función lineal de varias variables que permita clasificar nuevas observaciones que se presentan.

**Métodos log-lineales y logit:** Se predicen números de apariciones en casillas (recuentos) en función de otras casillas. Se usan variables categóricas.

**Análisis de correlación canónica:** Se toma un grupo de variables y se trata de predecir sus valores en función de otro grupo de variables.

**Análisis multivariante de la varianza:** se descompone la variabilidad en una medida de un conjunto de variables cuantitativas en función de otras variables categóricas.

## Métodos Independientes

**Análisis de componentes principales:** Se tienen  $n$  variables cuantitativas y se *mezclan* mediante combinaciones lineales reduciéndose a  $p < n$  variables que resumen la información para facilitar la interpretación.

**Análisis factorial:** Es parecido a la anterior aunque sólo se fija en explicar en términos de factores ocultos las variables originales, no tanto en reducir el número de variables.

**Multidimensional scaling:** Busca mapas de los objetos, situándolos según una serie de métricas.

**Análisis de correspondencias:** Es parecido al análisis factorial, pero con variables categóricas exclusivamente.

**Análisis de cluster:** Trata de identificar grupos naturales entre las observaciones según sus valores medidos por las variables.

## Algebra de Matrices

En el análisis multivariable se presentan de forma habitual matrices. En general, se toman varias variables aleatorias o mediciones que ocupan una serie de columnas y estas mediciones se consideran sobre una serie de objetos o individuos.

Por ejemplo, se toman 5 personas y se mide la edad de entrada en la universidad ( $x_1$ ), la nota media de las asignaturas después del primer año ( $x_2$ ) y el sexo ( $x_3$ ). Se obtiene:

	$x_1$	$x_2$	$x_3$
<b>1</b>	18,45	7.4	1
<b>2</b>	18,41	6.5	0
<b>3</b>	18,39	7.2	0
<b>4</b>	18,70	9.4	1
<b>5</b>	18,34	7.1	1

En sentido estricto, las 5 personas son una muestra aleatoria extraída de una población muy grande y se consideran variables aleatorias en el sentido de que su valor (por ej.  $x_2$ :

nota final media) no puede ser determinado previamente, sino que depende de muchas causas en número inconmensurable.

El concepto principal que se estudia es el concepto de vector. Cuando medimos una variable en un conjunto de elementos de una población, esta muestra puede representarse geoméricamente asociando el valor de la variable en cada elemento a una dimensión del espacio.

Un vector de dimensión  $n$  puede verse como una representación de los valores de una variable en  $n$  elementos de una población. Se puede ver que existe una relación entre los conceptos básicos de descripción estadística de la variable y ciertas operaciones vectoriales

A su vez, una matriz es un conjunto de vectores: cuando en lugar de medir una variable en cada elemento observamos los valores de  $k$  variables, podemos representar la muestra de datos multivariantes mediante una matriz.

## Vectores

En general, una muestra de  $n$  elementos de una variable la representaremos mediante un vector. La longitud de un vector se denomina módulo. En una muestra, el módulo del vector diferencia entre el vector asociado a la muestra y el vector que representa una constante es la desviación típica de la variable. Si el vector representa los valores de una variable de media cero, el módulo del vector es directamente la desviación típica.

La dependencia lineal entre dos variables se mide por la covarianza. El concepto análogo vectorial es el de producto escalar, que es la herramienta principal para estudiar la relación entre dos vectores. Dados dos vectores, el producto escalar es el producto de sus longitudes por el coseno del ángulo que forman. De acuerdo con esta definición, si consideramos vectores de longitud unidad el producto escalar es el coseno de su ángulo en el espacio, y será, en consecuencia, un número entre  $-1$  y  $1$ . Si los vectores son perpendiculares u ortogonales su producto escalar es cero. Si son colineales (es decir, están sobre la misma línea) su producto escalar es uno o menos uno.

Si dos vectores representan los valores de dos variables estandarizadas (las dos variables tienen media cero y varianza unidad) en los mismos  $n$  elementos de una población, el producto escalar es equivalente al coeficiente de correlación entre las dos variables.

Cuando consideramos varios vectores, el concepto principal es la noción de dependencia lineal. Para comprender la importancia de esta idea, supongamos que tenemos  $k$  variables medidas sobre  $n$  elementos de una población ( $n \geq k$ ), y los  $n$  valores de cada variable forman un vector en el espacio de  $n$  dimensiones.

Un problema importante es conocer cuantas variables realmente distintas tenemos. Por ejemplo, si una variable representa salarios en euros y otra los mismos salarios pero medidos en dólares aunque ambas variables no sean idénticas es claro que las dos variables miden la misma característica. Las dos variables son linealmente dependientes, ya que una se obtiene multiplicando por una constante los valores de la otra.

Generalizando esta idea, diremos que  $k$  variables son linealmente dependientes si podemos obtener los valores de una cualquiera mediante una combinación lineal del resto. Por ejemplo, si tenemos tres variables, número de hombres, número de mujeres y número de personas (que es la suma de las anteriores), las tres variables son linealmente dependientes porque podemos calcular el valor de cualquiera de ellas conocidos los valores de las otras dos. Al representar las variables como vectores, la noción de dependencia lineal permite conocer el número de variables distintas existentes en un grupo de variables.

Si tenemos  $k$  vectores de  $n$  componentes y  $k > n$  siempre podemos tomar  $n$  vectores cualesquiera de los  $k$  y expresar los  $k - n$  restantes como combinación lineal de estos vectores. Por tanto, en el espacio  $\mathbb{R}^n$  de vectores de  $n$  coordenadas, el máximo número de variables linealmente independientes que podemos tener es  $n$ .

## Definiciones básicas

Llamaremos vector a un conjunto ordenado de  $n$  número reales,  $\mathbf{x}$ , y escribiremos sus componentes en columna:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

donde  $x_i$  es el componente  $i$  del vector.

En Estadística un vector columna es habitualmente la representación de los valores de una variable en una muestra de  $n$  elementos.

Un conjunto de  $n$  números reales  $\mathbf{x}$  es un punto en el espacio  $\mathbb{R}^n$ . Intuitivamente, consideraremos al vector  $\mathbf{x}$  como la línea que va desde el origen de coordenadas hasta el punto  $\mathbf{x}$ . La dirección es importante, porque no es lo mismo el vector  $\mathbf{x}$  que el  $-\mathbf{x}$ . De esta manera a cada punto del espacio en  $\mathbb{R}^n$  le asociamos un vector. Llamaremos vector constante al que tiene todas sus coordenadas iguales. Para cada vector:

**La suma (o diferencia)** de los vectores  $\mathbf{x}, \mathbf{y}$ , ambos en  $\mathbb{R}^n$ , se define como un nuevo vector con componentes iguales a la suma (diferencia) de los componentes de los sumandos:

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

Es inmediato comprobar que la suma de vectores es asociativa y conmutativa.

La suma de dos vectores corresponde a la idea intuitiva de trasladar uno al extremo del otro y construir la línea que va desde el origen de coordenadas al extremo de la suma.

La operación suma vectorial corresponde a generar una nueva variables que es suma de las anteriores. Por ejemplo, si  $\mathbf{x}$ , representa el número de trabajadores varones en un conjunto de empresas e  $\mathbf{y}$ , el número de trabajadoras, la variable  $\mathbf{x} + \mathbf{y}$  representa el número total de trabajadores. La diferencia de vectores se utiliza con frecuencia en estadística para medir la distancia entre una variable y el vector asociado

**El producto de una constante por un vector**, es un nuevo vector cuyos componentes



son los iniciales multiplicados por la constante.

$$\mathbf{z} = k\mathbf{x} = \begin{pmatrix} kx_1 \\ \vdots \\ kx_n \end{pmatrix}.$$

Multiplicar por una constante equivale a un cambio en las unidades de medida. Por ejemplo, si en lugar de medir el número de trabajadores en unidades (variable  $\mathbf{x}$ ) lo hacemos en centenas (variable  $\mathbf{z}$ ) entonces variable  $\mathbf{z} = \mathbf{x}/100$ .

Llamaremos **vector transpuesto**  $\mathbf{x}'$ , de otro  $\mathbf{x}$ , a un vector con los mismos componentes, pero escritos en fila:

$$\mathbf{x}' = (x_1, \dots, x_n).$$

Al transponer un vector columna se obtiene un vector fila. Generalmente los vectores fila se utilizan para describir los valores de  $k$  variables distintas en un mismo objeto de una población.

El **producto escalar ó interno** de dos vectores  $\mathbf{xy}$ , ambos en  $\mathbb{R}^n$ , que escribiremos  $\mathbf{x}'\mathbf{y}$  ó  $\mathbf{y}'\mathbf{x}$ , es el escalar obtenido al sumar los productos de sus componentes.

$$\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} = \sum_{i=1}^n x_i y_i.$$

Si tomamos  $\mathbf{y} = (1/n, \dots, 1/n)$ , el producto escalar de la variable  $\mathbf{x}$  y este vector de constantes proporciona la media de la variables

$$(x_1, x_2, \dots, x_n) \begin{pmatrix} 1/n \\ 1/n \\ \dots \\ 1/n \end{pmatrix} = \frac{\sum x_i}{n}.$$

Cuando ninguna de las dos variables es una constante el producto escalar se asocia en Estadística a la covarianza.

Para variables con media cero el producto escalar de los dos vectores que representan sus valores es directamente la covarianza.

Para variables con media distinta de cero, la covarianza corresponde al producto escalar de las desviaciones de los datos respecto a sus medias. Observamos que obtener la

desviación respecto a su media equivale a calcular la diferencia de vectores  $\mathbf{x} - \bar{x}\mathbf{1}$ , donde  $\mathbf{1}$  es un vector constante de *unos*.

Se llamará **norma** o longitud de un vector  $\mathbf{x}$ , a la raíz cuadrada del producto escalar  $\mathbf{x}'\mathbf{x}$ . Se escribe  $|\mathbf{x}|$ :

$$|\mathbf{x}| = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{x_1^2 + \dots + x_n^2}.$$

La norma es la longitud del segmento que une el origen con el punto  $\mathbf{x}$  que corresponde a la longitud de la hipotenusa en el triángulo rectángulo formado por el vector y sus proyecciones sobre los ejes.

Para variables con media cero la desviación típica es  $n$  veces la norma del vector. Para variables con media distinta de cero la desviación típica es  $n$  veces la norma del vector de los datos en desviaciones a la media,  $\mathbf{x} - \bar{x}\mathbf{1}$ .

El producto escalar de dos vectores puede verse también como el producto del módulo de un vector y la proyección del otro sobre él. En general, el coseno del ángulo  $\theta$  formado por los dos vectores  $\mathbf{x}$ ,  $\mathbf{y}$  viene dado por la relación:

$$\cos \theta = \frac{\mathbf{x}'\mathbf{y}}{|\mathbf{x}||\mathbf{y}|}.$$

para variables con media cero el coseno es el *coeficiente de correlación*. Como  $\cos \theta \leq 1$ , se demuestra en general que:

$$|\mathbf{x}'\mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|.$$

que se conoce como la *desigualdad de Schwarz*.

Dos vectores son *ortogonales*, o perpendiculares, si y sólo si su producto escalar es cero. Por la definición de ángulo

$$\mathbf{x}'\mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos \theta,$$

siendo  $\theta$  el ángulo que forman los vectores. Si  $\theta = 90^\circ$  el coseno es cero y también lo será el producto escalar.

La implicación estadística de ortogonalidad es la existencia de incorrelación. Si dos variables están incorreladas, llamando  $r$  al coeficiente de correlación, se observa que  $r = \cos \theta = 0$ , es decir, los vectores que las caracterizan forman un ángulo de 90 grados.

## Dependencia Lineal

Un conjunto de vectores  $\mathbf{x}_1, \dots, \mathbf{x}_k$  es *linealmente dependiente* si existen escalares  $c_1, \dots, c_k$ , no todos nulos, tales que:

$$c_1 \mathbf{x}_1 + \dots + c_k \mathbf{x}_k = 0$$

Si no existen tales escalares, se dice que los vectores son *linealmente independientes*. Intuitivamente, si los vectores son linealmente dependientes podemos expresar alguno como combinación lineal de los demás. Por ejemplo, supuesto  $c_1 \neq 0$  y llamando  $a_i = c_i/c_1$ , tenemos

$$\mathbf{x}_1 = a_2 \mathbf{x}_2 + \dots + a_k \mathbf{x}_k.$$

En general en el espacio  $\mathbb{R}^n$  el número máximo de vectores linealmente independientes es  $n$ . En efecto, si tenemos un conjunto de  $n + h$  vectores siempre podemos tomar  $n$  cualquiera y escribir

$$\mathbf{x}_{n+1} = \sum_{i=1}^n a_i \mathbf{x}_i$$

que es un sistema con  $n$  ecuaciones y  $n$  incógnitas y obtener los coeficientes  $a_i$ .

En estadística un conjunto de vectores linealmente independientes corresponde a un conjunto de variables que no están relacionadas linealmente de forma exacta. Por ejemplo, si dos variables miden la misma variables pero en unidades distintas serán linealmente dependientes. También los serán si una de las variables se ha generado como una combinación lineal de las otras.

Dado un conjunto de  $k$  vectores linealmente independientes  $(\mathbf{x}_1, \dots, \mathbf{x}_k)$ , en  $\mathbb{R}^n$  ( $k \leq n$ ), llamaremos *espacio generado* por este conjunto de vectores al espacio que contiene todos los vectores  $\mathbf{z}$ , en  $\mathbb{R}^n$ , que pueden expresarse como combinación lineal de éstos. El conjunto  $(\mathbf{x}_1, \dots, \mathbf{x}_k)$  se llama *base generadora del espacio*, o simplemente base del espacio. Si  $\mathbf{z}$

pertenece a este espacio entonces,

$$\mathbf{z} = c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k.$$

Es fácil comprobar que  $\mathbf{z}$  estará en un espacio de dimensión  $k$ : es efecto, podemos tomar las primeras  $k$  coordenadas de  $\mathbf{z}$  y obtener del sistema de  $k$  ecuaciones y  $k$  incógnitas resultante, los coeficientes  $c_1, \dots, c_k$ . Las  $n - k$  coordenadas siguientes de  $\mathbf{z}$  quedan determinadas, al estarlo los  $c_i$ , por lo que, obviamente,  $\mathbf{z}$  sólo tiene  $k$  componentes independientes, estando, por lo tanto, en un espacio de dimensión  $k$ . El espacio generado por un conjunto de variables corresponde a todas las variables que pueden generarse como índices o combinaciones lineales de las originales.

La *dimensión* de un espacio  $E_k$  se define como el número de vectores linealmente independientes que lo generan.

Diremos que un vector  $\mathbf{x}$  es *ortogonal a un subespacio*  $E_p$  si  $\mathbf{x}$  es ortogonal a todo vector de  $E_p$ , es decir, si  $\mathbf{y}$  pertenece al subespacio  $E_p$ , que escribiremos  $\mathbf{y} \in E_p$ , entonces:

$$\mathbf{y}'\mathbf{x} = 0.$$

Llamaremos *complemento ortogonal* de un subespacio  $E_p$ , de dimensión  $p$ , y lo denotaremos por  $C(E_p)$ , al espacio que contiene todos los vectores ortogonales a  $E_p$ . Entonces, si  $\mathbf{x} \in E_p$ ,  $\mathbf{y} \in C(E_p)$  se verifica  $\mathbf{x}'\mathbf{y} = 0$ . La dimensión de  $C(E_p)$  será  $n - p$ .

## Matrices

Para trabajar conjuntamente con  $k$  variables o vectores definimos el concepto de matriz. Una matriz es un conjunto de números dispuestos en filas y columnas y puede verse como un conjunto de vectores columna o un conjunto de vectores fila. Si intercambiamos las filas de una matriz por las columnas se obtiene una nueva matriz que se denomina la traspuesta de la primera. En particular, un vector columna de orden  $n$  es una matriz de dimensiones  $n \times 1$  (su traspuesta es un vector fila), y un escalar es una matriz de dimensiones  $1 \times 1$  (e igual a su traspuesta).

Una propiedad básica de una matriz es el rango, que indica el número máximo de vectores fila o columna linealmente independientes que la forman. En una matriz con  $n$  filas y  $k$  columnas ( $n \geq k$ ), las  $k$  columnas pueden ser vectores linealmente independientes y, así, el número máximo de vectores linealmente independientes es  $k$ . Su rango máximo es  $k$  y cuando esto ocurre decimos que la matriz tiene rango completo.

El rango de una matriz es igual al de su traspuesta.

La generalización del concepto de producto escalar entre dos vectores es el producto matricial, que se define como una nueva matriz que contiene todos los productos escalares entre los vectores fila de la primera matriz y los vectores columna de la segunda. Para que este producto sea posible la primera matriz tiene que tener tantas columnas como filas tenga la segunda matriz. Por la propia definición se deduce que este producto no es conmutativo. Diremos que pre-multiplicamos la matriz  $A$  por la  $B$  cuando realizamos el producto  $B \cdot A$  y que post-multiplicamos la  $A$  por la  $B$  si realizamos el producto  $A \cdot B$ .

## Definiciones básicas

Llamaremos matriz,  $\mathbf{A}$ , de dimensiones  $(n \times k)$  a un conjunto de  $n \times k$  números reales, ordenados en  $n$  filas y  $k$  columnas. Por ejemplo, si medimos  $k$  variables en  $n$  individuos de una población podemos representar cada variable por un vector columna de dimensión  $n$  y cada vector columna es pues una matriz  $(n \times 1)$ . El conjunto de los  $k$  vectores es un matriz  $(n \times k)$ , y esta matriz puede verse como un conjunto de  $k$  vectores columna, o como un conjunto de  $n$  vectores fila, cada uno de ellos de dimensión  $k$ . Llamaremos *matriz traspuesta*  $\mathbf{A}'$  a la matriz obtenida a partir de  $\mathbf{A}$  intercambiando filas por columnas. Si  $\mathbf{A}$  es  $n \times k$ ,  $\mathbf{A}'$  será  $k \times n$ . Se verifica:

$$(\mathbf{A}')' = \mathbf{A}$$

La suma de dos matrices se define sólo cuando ambas tienen las mismas dimensiones. Cada elemento de la matriz suma se obtiene sumando los elementos correspondientes de

los sumandos

$$\mathbf{A} + \mathbf{B} = \mathbf{C} \Rightarrow \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ a_{n1} & \cdots & a_{nk} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1k} \\ b_{n1} & \cdots & b_{nk} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1k} \\ c_{n1} & \cdots & c_{nk} \end{pmatrix}$$

con  $c_{ij} = a_{ij} + b_{ij}$ . Se verifica:

(i)  $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

(ii)  $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$ .

Sumar dos matrices equivale en términos estadísticos a sumar los valores de las variables correspondientes a las columnas de las matrices. Por ejemplo, si la matriz  $\mathbf{A}$  representa el número de incidencias leves de  $k$  tipos distintos en una empresa en  $n$  semanas y la  $\mathbf{B}$  el número de incidencias graves en las mismas semanas, la suma representa el numero total de incidencias.

El *producto* de dos matrices  $\mathbf{A} \cdot \mathbf{B}$  sólo es posible cuando el número de columnas de  $\mathbf{A}$  es igual al número de filas de  $\mathbf{B}$ . Entonces, si  $\mathbf{A}(n \times k)$  y  $\mathbf{B}(k \times h)$ , el producto es una matriz  $\mathbf{C}(n \times h)$  con términos:

$$c_{ij} = \sum_{l=1}^k a_{il}b_{lj}$$

Es decir, el término  $c_{ij}$  representa el producto escalar del vector  $\mathbf{a}'_i$ , definido por la  $i$ -ésima fila de  $\mathbf{A}$ , por el vector  $\mathbf{b}_j$ , de la  $j$ -ésima columna de  $\mathbf{B}$ . Si escribimos:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}'_1 \\ \vdots \\ \mathbf{a}'_n \end{pmatrix} \quad \mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_h]$$

donde todos los vectores tienen dimensiones  $k$ , el producto matricial de estas dos matrices es:

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{C} = \begin{pmatrix} \mathbf{a}'_1 \mathbf{b}_1 & \cdots & \mathbf{a}'_1 \mathbf{b}_h \\ \vdots & \ddots & \vdots \\ \mathbf{a}'_n \mathbf{b}_1 & \cdots & \mathbf{a}'_n \mathbf{b}_h \end{pmatrix}_{(n \times h)}$$

En particular, el producto de una matriz  $(n \times k)$  por un vector  $(k \times 1)$ ,  $\mathbf{Ax}$  será un nuevo vector de dimensión  $(n \times 1)$  cuyos componentes se obtienen por el producto escalar de las filas de  $\mathbf{A}$  por el vector  $\mathbf{x}$ . Si

$$\mathbf{y} = \mathbf{Ax}$$

la matriz  $\mathbf{A}$  transforma vectores  $\mathbf{x}$  en  $\mathbb{R}^k$  en vectores  $\mathbf{y}$  en  $\mathbb{R}^n$ . En particular si  $\mathbf{A}$  es cuadrada de orden  $n$  y  $\mathbf{x} \in \mathbb{R}^n$  el producto  $\mathbf{A}\mathbf{x}$  es un nuevo vector de  $\mathbb{R}^n$ . El producto matricial tiene, entre otras, las propiedades siguientes:

$$(i) \quad \mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C}$$

$$(ii) \quad (\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'$$

$$(iii) \quad \mathbf{A}\mathbf{I} = \mathbf{I}\mathbf{A} = \mathbf{A}.$$

Un tipo importante de producto matricial es el producto  $\mathbf{A} \cdot \mathbf{A}'$ , que da lugar a una matriz cuadrada, como definimos a continuación.

## Matrices Cuadradas

Una matriz es *cuadrada* si  $n = k$ . Dentro de las matrices cuadradas se llaman simétricas a las que tienen cada fila igual a la correspondiente columna, es decir  $a_{ij} = a_{ji}$ . Una matriz simétrica es por tanto idéntica a su traspuesta, y diremos que  $\mathbf{A}$  es *simétrica* si

$$\mathbf{A}' = \mathbf{A}.$$

Una matriz cuadrada y simétrica muy importante es la *matriz identidad*, que representaremos por  $\mathbf{I}$  y tiene unos en la diagonal y ceros fuera de ella, es decir:

$$\mathbf{I} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & 1 & \vdots \\ 0 & \dots & 1 \end{pmatrix}.$$

En particular, los productos  $\mathbf{A} \cdot \mathbf{A}'$  y  $\mathbf{A}' \cdot \mathbf{A}$  conducen a matrices simétricas.

Las matrices cuadradas aparecen de manera natural cuando consideramos estos productos en matrices de datos. Si  $\mathbf{A}$  es de dimensión  $(n \times k)$  y representa los valores de  $k$  variables de media cero en  $n$  individuos de una población, la matriz cuadrada de orden  $k$ ,  $\frac{\mathbf{A}'\mathbf{A}}{n}$ , contiene las varianzas y covarianzas entre las variables. Otra matriz cuadrada y simétrica de amplio uso en estadística es la matriz de correlaciones, que contiene unos en la diagonal y fuera de ella los coeficientes de correlación entre las variables.

Sobre las matrices cuadradas podemos definir dos medidas escalares que resumen su tamaño global: el determinante y la traza. Ambas son medidas relativas, ya que se modifican si multiplicamos los elementos de la matriz por constantes.

### Determinante de una matriz

Dada una matriz  $\mathbf{A}$  cuadrada de orden  $n$  con términos  $a_{ij}$ , se denomina *determinante* de esta matriz, y lo representaremos por  $|\mathbf{A}|$ , al escalar obtenido mediante la suma:

$$|\mathbf{A}| = \sum (-1)^r a_{1i_1} a_{2i_2}, \dots, a_{ni_n}$$

que está extendida a todas las permutaciones de los segundos índices. Los índices  $i_1, \dots, i_n$  son una permutación de los números  $1, 2, \dots, n$  y  $r$  es el número de cambios entre dos elementos que es necesario para poner los subíndices  $i_1, \dots, i_n$  en el orden  $1, 2, \dots, n$ .

Por ejemplo, en una matriz  $2 \times 2$  el número de permutaciones de dos términos es dos y el determinante estará formado por los dos términos:

$$|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}.$$

Observemos que el segundo término es negativo, porque el orden de los subíndices es  $(2, 1)$  y es necesario un cambio para situarlos en el orden  $1, 2$ . En una matriz  $3 \times 3$  el determinante tiene  $3! = 6$  términos que se obtiene de las 6 posibles permutaciones:

$$\begin{array}{ccc} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 2 & 1 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \\ 3 & 2 & 1 \end{array}$$

En consecuencia:

$$\begin{aligned} |\mathbf{A}| = & a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + \\ & + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

Para matrices mayores de 3 el cálculo del determinante es tedioso. Su cálculo se simplifica mediante el concepto de *menor*. Llamaremos menor del elemento  $a_{ij}$  de una matriz



cuadrada de orden  $n$ ,  $m_{ij}$ , al determinante de la matriz de orden  $n - 1$  que resulta al eliminar de la matriz original  $\mathbf{A}$  la fila  $i$  y la columna  $j$ . Se denomina *adjunto* del elemento  $a_{ij}$  al escalar  $(-1)^{i+j} m_{ij}$ . Se demuestra que el determinante de una matriz puede calcularse multiplicando cada elemento de una fila por sus adjuntos. Entonces:

$$|\mathbf{A}| = \sum_{j=1}^n a_{ij} (-1)^{i+j} m_{ij}$$

para cualquier fila  $i$ . Por ejemplo, en una matriz  $3 \times 3$ , desarrollando por los elementos de la primera fila

$$|\mathbf{A}| = a_{11} (a_{22}a_{33} - a_{23}a_{32}) - a_{12} (a_{21}a_{33} - a_{23}a_{31}) + a_{13} (a_{21}a_{32} - a_{22}a_{31}),$$

que coincide con el resultado anterior. Los determinantes tienen las propiedades siguientes:

- (i)  $|\lambda\mathbf{A}| = \lambda^n |\mathbf{A}|$
- (ii)  $|\mathbf{A}'| = |\mathbf{A}|$
- (iii)  $|\mathbf{AB}| = |\mathbf{A}| |\mathbf{B}|$  si ambas son cuadradas, de orden  $n$ .
- (iv) Si permutamos dos filas o dos columnas entre sí, el determinante cambia sólo su signo.
- (v) Si una fila (o columna) de una matriz es una combinación lineal de las restantes filas (o columnas) el determinante de la matriz es cero. Entonces diremos que la matriz es *singular*.

El determinante de una matriz de varianzas y covarianzas es una medida global de la independencia entre las variables. Por ejemplo, si una variable es combinación lineal de las demás, entonces las variables son linealmente dependientes y el determinante es nulo. En términos relativos, cuanto mayor sea el determinante mayor es la independencia entre los vectores.

Si consideramos matrices cuadradas estandarizadas de manera que el módulo de cada vector columna (o fila) sea la unidad, el determinante es máximo si la matriz tiene unos

en la diagonal y ceros en el resto, de manera que los vectores columna son ortogonales (independientes) entre sí.

### Traza de una matriz

La *traza* de una matriz cuadrada es la suma de los elementos de la diagonal principal de la matriz. Si  $\mathbf{C}$  es una matriz con elementos  $c_{ij}$  se verifica:

$$tr(\mathbf{C}) = \sum_{i=1}^n c_{ii}$$

La traza es un operador lineal. En efecto, de la definición se obtiene:

(i)  $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$ .

(ii)  $tr(\lambda\mathbf{A}) = \lambda tr(\mathbf{A})$ , donde  $\lambda$  es un escalar.

(iii) Se demuestra que:  $tr(\mathbf{ABC}) = tr(\mathbf{CAB}) = tr(\mathbf{BCA})$ , en el supuesto de que todos los productos estén definidos.

(iv)  $tr(\mathbf{C}^2) = tr(\mathbf{CC}) = \sum_{i=1}^n \sum_{j=1}^n c_{ij}^2$ .

La traza es una medida global de tamaño de la matriz que se obtiene sumando sus elementos diagonales. Por ejemplo, la traza de una matriz de varianzas y covarianzas es la suma de todas las varianzas de las variables. Al tener en cuenta únicamente los elementos diagonales es una medida más simple que el determinante.

### Formas cuadráticas

Una tercera forma de obtener un escalar a partir de una matriz cuadrada es construyendo una forma cuadrática. Llamaremos *forma cuadrática* a una expresión escalar del tipo:

$$\mathbf{y}'\mathbf{A}\mathbf{y}$$

donde  $\mathbf{y}$  es un vector,  $\mathbf{y}'$  su transpuesto y  $\mathbf{A}$  una matriz cuadrada y simétrica. Si la dimensión del vector es  $(n \times 1)$ , la matriz debe ser cuadrada de orden  $n$  para que sea

posible el producto y, así, la matriz resultante tendrá dimensión:

$$(1 \times n) \times (n \times n) \times (n \times 1) = (1 \times 1).$$

La forma cuadrática es siempre un escalar. Su expresión general es:

$$\sum_{i=1}^n a_{ii}y_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n a_{ij}y_iy_j.$$

Diremos que una matriz  $\mathbf{A}$  es *semidefinida positiva* si cualquier forma cuadrática formada con ella es no negativa, para cualquier vector  $\mathbf{y} \neq 0$ . Si la forma cuadrática es siempre mayor que cero diremos que la matriz  $\mathbf{A}$  es *definida positiva*. Se demuestra que el determinante y la traza de una matriz semidefinida positiva son también no negativos.

### Matriz Inversa

Dada una matriz  $\mathbf{A}$  cuadrada  $n \times n$ , no singular, definimos su inversa,  $\mathbf{A}^{-1}$ , como una matriz  $n \times n$  tal que:

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$$

donde  $\mathbf{I}$  es la matriz identidad, que tiene unos en la diagonal y ceros fuera de ella. Es decir, escribiendo  $\mathbf{A}$  con vectores fila  $\mathbf{a}'_i$ , la matriz  $\mathbf{A}^{-1}$  tendrá vectores columna  $\mathbf{b}_i$  tales que:

$$\begin{pmatrix} \mathbf{a}'_1 \\ \vdots \\ \mathbf{a}'_n \end{pmatrix} (\mathbf{b}_1 \quad \dots \quad \mathbf{b}_n) = \begin{pmatrix} \mathbf{a}'_1\mathbf{b}_1 & \dots & \mathbf{a}'_1\mathbf{b}_n \\ \vdots & & \vdots \\ \mathbf{a}'_n\mathbf{b}_1 & \dots & \mathbf{a}'_n\mathbf{b}_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \dots & \dots & 1 \end{pmatrix}.$$

En consecuencia la matriz  $\mathbf{A}^{-1}$  debe tener por columnas vectores  $\mathbf{b}$  tales que:

- (i)  $\mathbf{b}_i$  es ortogonal a  $\mathbf{a}_j$  ( $\forall j \neq i$ );
- (ii)  $\mathbf{b}'_i\mathbf{a}_i = \mathbf{a}'_i\mathbf{b}_i = 1$ .

Observemos que el cálculo de la inversa nos resuelve el problema de calcular vectores ortogonales a uno dado (o variables incorreladas con una dada). En efecto, el espacio ortogonal al vector  $\mathbf{a}_1$  es el formado por los vectores  $\mathbf{b}_2, \dots, \mathbf{b}_n$ .

Por ejemplo, dada la matriz

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 4 \end{pmatrix},$$

la inversa es

$$\mathbf{A}^{-1} = \begin{pmatrix} 0,5 & -0,125 \\ 0 & 0,25 \end{pmatrix}$$

y cada vector columna de la inversa es ortogonal a un vector columna y verifica la condición de  $\mathbf{b}'_i \mathbf{a}_i = 1$ .

La necesidad de calcular la inversa de una matriz aparece de manera natural al resolver sistemas de ecuaciones lineales,

$$\mathbf{Ax} = \mathbf{b},$$

donde  $\mathbf{A}$  es una matriz conocida cuadrada de orden  $n$ ,  $\mathbf{b}$  un vector de constantes y  $\mathbf{x}$  un vector de  $n$  incógnitas. Para que este sistema tenga solución única las  $n$  ecuaciones deben de ser distintas, lo que supone que no existe una fila de  $\mathbf{A}$  que sea combinación lineal de las demás. Entonces  $\mathbf{A}$  es no singular y la solución se obtiene mediante:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

El cálculo de la matriz inversa  $\mathbf{A}^{-1}$  de una matriz dada es engorroso y debe realizarse mediante un ordenador si la dimensión de  $\mathbf{A}$  es alta. Se demuestra que la inversa de una matriz puede calcularse mediante las tres operaciones siguientes:

1. Se sustituye cada elemento por su adjunto.
2. Se transpone la matriz resultante. Se obtiene una matriz que llamaremos *adjunta* de la matriz  $\mathbf{A}$ .
3. Se divide cada término de la matriz adjunta por el determinante de la matriz original.

Como ejemplo calcularemos la inversa de la matriz

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

comenzaremos sustituyendo cada elemento por su adjunto. Por ejemplo, para el elemento  $(1, 1)$  su adjunto es  $(-1)^2 [2 \cdot 3 - 1 \cdot 0] = 6$ . Para el  $(1, 2)$ ,  $(-1^3) [-1 \cdot 3 - 1 \cdot 0] = 3$ , etc.

$$\begin{pmatrix} 6 & 3 & 0 \\ -3 & 3 & 0 \\ -1 & -1 & 3 \end{pmatrix}$$

Se transpone esta matriz y resulta:

$$Adj(\mathbf{A}) = \begin{pmatrix} 6 & -3 & -1 \\ 3 & 3 & -1 \\ 0 & 0 & 3 \end{pmatrix}$$

Si dividimos ahora por el determinante de la matriz  $\mathbf{A}$

$$|\mathbf{A}| = 6 + 3 = 9,$$

se obtiene

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{9} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{9} \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$$

y podemos comprobar que  $\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}$ .

La inversa de una matriz  $\mathbf{A}$  tiene las propiedades siguientes:

- (i)  $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$  para matrices cuadradas no singulares.
- (ii)  $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$
- (iii)  $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$
- (iv)  $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$
- (v) Si  $\mathbf{A}$  es simétrica también lo es  $\mathbf{A}^{-1}$ .

La matriz inversa de una matriz de varianzas y covarianzas recoge la información de la dependencia conjunta de todas las variables de manera más completa que la matriz de varianzas y covarianzas.

## Matrices ortogonales

Llamaremos matriz ortogonal,  $\mathbf{C}$ , a una matriz cuadrada, que representa un giro en el espacio. Por tanto, dado un vector  $\mathbf{x}$ , si aplicamos una matriz ortogonal  $\mathbf{C}$  para obtener un nuevo vector  $\mathbf{y} = \mathbf{C}\mathbf{x}$ , como el módulo de  $\mathbf{y}$  debe ser idéntico al de  $\mathbf{x}$  al tratarse de un giro, tendremos la condición :

$$\mathbf{y}'\mathbf{y} = \mathbf{x}'\mathbf{C}'\mathbf{C}\mathbf{x} = \mathbf{x}'\mathbf{x},$$

es decir, deberá verificarse que :

$$\mathbf{C}'\mathbf{C} = \mathbf{I}$$

como además tendremos que  $\mathbf{x} = \mathbf{C}^{-1}\mathbf{y}$ , y por la condición anterior  $\mathbf{C}'\mathbf{y} = \mathbf{C}'\mathbf{C}\mathbf{x} = \mathbf{x}$ , concluimos que la matriz inversa de una matriz ortogonal es igual a su traspuesta. Esta es la condición de *ortogonalidad*:

$$\mathbf{C}' = \mathbf{C}^{-1}.$$

Esta condición impone que las filas o columnas de una matriz ortogonal sean vectores ortogonales entre sí y de longitud unidad, ya que:

$$\begin{pmatrix} \mathbf{c}'_1 \\ \vdots \\ \mathbf{c}'_n \end{pmatrix} (\mathbf{c}_1 \dots \mathbf{c}_n) = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{c}'_1\mathbf{c}_1 & \dots & \mathbf{c}'_1\mathbf{c}_n \\ \vdots & & \vdots \\ \mathbf{c}'_n\mathbf{c}_1 & \dots & \mathbf{c}'_n\mathbf{c}_n \end{pmatrix}$$

además:  $|\mathbf{C}| = |\mathbf{C}'| = 1$ , donde  $|\mathbf{C}|$  es el determinante de  $\mathbf{C}$ .

Por ejemplo, en  $\mathbb{R}^2$ , la matriz

$$\mathbf{C} = \begin{pmatrix} \cos \alpha & -\text{sen } \alpha \\ \text{sen } \alpha & \cos \alpha \end{pmatrix}$$

es ortogonal, ya que  $\mathbf{C}\mathbf{C}' = \mathbf{I}$ .

## Rango de una matriz

A cada matriz podemos asociarle un número que indica el máximo número de vectores linealmente independientes que podemos obtener a partir de ella.

Supongamos primero matrices cuadradas. Dada la matriz  $\mathbf{A}$  cuadrada de orden  $n$ , llamaremos *rango* de la matriz a la dimensión del espacio generado por sus vectores columna. Si estos son linealmente independientes el rango será igual a  $n$ . En otro caso será menor que la dimensión de la matriz. Se demuestra que el rango de una matriz cuadrada es igual al de su transpuesta; y que el rango es siempre igual al máximo número de vectores columna, o fila, linealmente independientes. En general, si llamamos  $rg(\mathbf{A})$  al rango de la matriz  $\mathbf{A}$  se verifica:

- (i)  $rg(\mathbf{A}_{n \times k}) \leq \min(n, k)$ . El rango es igual o menor que el menor de  $n$  y  $k$ .
- (ii)  $rg(\mathbf{A}'\mathbf{A}) = rg(\mathbf{A}\mathbf{A}') = rg(\mathbf{A})$
- (iii) si  $rg(\mathbf{A}_{n \times n}) = n$ ,  $\mathbf{A}$  es no singular y existe  $\mathbf{A}^{-1}$ .
- (iv) si  $rg(\mathbf{A}_{n \times k}) = n < k$  ó  $rg(\mathbf{A}_{n \times k}) = k < n$ , se dice que  $\mathbf{A}$  es de rango completo.
- (v)  $rg(\mathbf{A}\mathbf{B}) \leq \text{mínimo de } (rg(\mathbf{A}) \text{ y } rg(\mathbf{B}))$ .
- (vi)  $rg(\mathbf{A}\mathbf{B}) = rg(\mathbf{A})$ , si  $|\mathbf{B}| \neq 0$  y  $\mathbf{A}$  cualquiera.
- (vii) Si  $\mathbf{A}$  y  $\mathbf{B}$  son cuadradas de orden  $n$  y  $\mathbf{A}\mathbf{B} = 0$ , entonces  $rg(\mathbf{A}) + rg(\mathbf{B}) \leq n$ .
- (viii)  $rg(\mathbf{A} + \mathbf{B}) \leq rg(\mathbf{A}) + rg(\mathbf{B})$ .

Como el rango de una matriz es el número de vectores linealmente independientes que la forman, el rango de una matriz de datos es el número de variables distintas (no relacionadas linealmente) que la componen.

## Autovectores y autovalores

Dada una matriz cuadrada, hay determinadas propiedades que esperamos que sean invariantes ante transformaciones lineales que preserven la información existente. Por ejemplo, si pre-multiplicamos la matriz por una nueva matriz y luego post-multiplicamos por la inversa de dicha matriz. También si giramos los ejes de coordenadas.

Supongamos que partiendo de  $k$  variables (vectores) pasamos a otras  $k$  variables que son combinación lineal de las anteriores mediante una de las operaciones anteriores. Por ejemplo, si en lugar de trabajar con los ingresos y los costes decidimos trabajar con los beneficios, construidos como ingresos-costes, y el volumen de actividad, definido como ingresos más costes; entonces hemos aplicado una transformación ortogonal. Aunque la matriz cuadrada que representa las varianzas y covarianzas de las nuevas variables sea distinta de la original, la esencia del problema es la misma, y se espera que haya componentes que permanezcan invariantes en el problema. Para precisar esta idea aparece el concepto de autovalores y autovectores de una matriz cuadrada.

Los autovalores son las medidas básicas de tamaño de una matriz, que no se ven alteradas por transformaciones lineales de esta matriz. Por ejemplo, si hacemos un cambio de coordenadas que equivalga a una rotación de ejes los autovalores no se modificarán.

Los autovectores representan las direcciones características de la matriz y no son invariantes. Se demuestra que las medidas globales de tamaño de la matriz, como la traza o el determinante, son sólo función de los autovalores y en consecuencia son también invariantes.

## Definiciones básicas

Definimos a los autovalores (o valores propios o raíces características) de una matriz cuadrada de orden  $n$ ,  $\mathbf{A}$ , como las soluciones de la ecuación polinómica.

$$|\mathbf{A} - \lambda\mathbf{I}| = 0.$$

Es inmediato deducir de la definición que si una matriz es diagonal, los autovalores son los elementos de la diagonal principal. En efecto, tendremos:

$$|\mathbf{A} - \lambda\mathbf{I}| = \left| \begin{pmatrix} a_1 & \dots & 0 \\ \vdots & a_2 & \vdots \\ 0 & \vdots & a_n \end{pmatrix} - \begin{pmatrix} \lambda & \dots & 0 \\ \vdots & \lambda & \vdots \\ 0 & \dots & \lambda \end{pmatrix} \right| = \left| \begin{pmatrix} a_1 - \lambda & \dots & 0 \\ \vdots & a_2 - \lambda & \vdots \\ 0 & \dots & a_n - \lambda \end{pmatrix} \right|$$

$$|\mathbf{A} - \lambda\mathbf{I}| = (a_1 - \lambda), \dots, (a_n - \lambda),$$



y las soluciones de esta ecuación polinómica son  $a_1, \dots, a_n$ .

De este ejemplo se deduce que aunque una matriz de orden  $n$  tiene  $n$  autovalores estos pueden aparecer repetidos y en general una matriz tiene  $h \leq n$  autovalores distintos. Si un autovalor aparece repetido  $r$  veces se dice que tiene multiplicidad  $r$ . Por ejemplo, la matriz diagonal:

$$\mathbf{A} = \begin{pmatrix} \mathbf{2} & 0 & 0 & 0 \\ 0 & \mathbf{3} & 0 & 0 \\ 0 & 0 & \mathbf{0} & 0 \\ 0 & 0 & 0 & \mathbf{0} \end{pmatrix}$$

tiene como autovalores 2, 3 y 0, este último valor con multiplicidad dos (aparece dos veces).

Llamaremos *autovectores* o vectores propios de una matriz cuadrada a los vectores  $\mathbf{u}$  que verifican para  $\mathbf{u} \neq 0$  la relación:

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u}.$$

Si  $\mathbf{u}$  es un vector propio de  $\mathbf{A}$  es obvio que  $k\mathbf{u}$ , donde  $k \in \mathbb{R}$ , será también vector propio. Para evitar esta indeterminación suponemos que los autovectores están normalizados de manera que  $|\mathbf{u}| = 1$ . Sin embargo, el signo queda indeterminado: si  $\mathbf{u}$  es un vector propio también lo es  $-\mathbf{u}$ .

Si una matriz cuadrada de orden  $n$  tiene  $n$  autovalores *distintos* entonces a cada autovalor le podemos asociar un vector propio bien definido y se demuestra que el conjunto de  $n$  vectores propios es linealmente independiente.

Si un autovalor es múltiple, es decir, la matriz no tiene  $n$  autovalores distintos, los vectores propios asociados a autovalores con multiplicidad mayor que uno no están definidos en general de manera única. Para ilustrar esta idea, consideremos la matriz

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

que tiene el autovalor 1 con multiplicidad 2. Los vectores  $\mathbf{u}_1 = (1\ 0\ 0)'$  y  $\mathbf{u}_2 = (0\ 1\ 0)'$  son vectores propios asociados al valor 1, pero también lo es  $\mathbf{u}_3 = \gamma_1\mathbf{u}_1 + (1 - \gamma_1)\mathbf{u}_2$ , para

cualquier valor de  $\gamma$ . Los vectores propios están en un espacio igual a la multiplicidad del autovalor, 2, y cualquier vector normalizado de este espacio de dimensión 2 es un vector propio de  $\mathbf{A}$ .

Los autovalores de una matriz tienen las propiedades siguientes:

(a) si  $\lambda$  es un autovalor de  $\mathbf{A}$ , entonces  $\lambda^r$  es un autovalor de  $\mathbf{A}^r$ .

En particular, si  $\mathbf{A}^{-1}$  existe,  $\lambda^{-1}$  es un autovalor de  $\mathbf{A}^{-1}$ . En efecto, si  $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$ , multiplicando por  $\mathbf{A}^{-1}$ ,  $\mathbf{u} = \lambda\mathbf{A}^{-1}\mathbf{u}$ , es decir  $\lambda^{-1}\mathbf{u} = \mathbf{A}^{-1}\mathbf{u}$ .

(b) La suma de los autovalores de  $\mathbf{A}$  es igual a la traza.

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i.$$

(c) El producto de los autovalores de  $\mathbf{A}$  es igual al determinante

$$|\mathbf{A}| = \prod_{i=1}^n \lambda_i.$$

(d) Si una matriz  $\mathbf{P}$  es no singular, entonces Las matrices  $\mathbf{A}$  y  $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$  tienen los mismos autovalores.

Efectivamente, si  $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$ , multiplicando ambos miembros por  $\mathbf{P}^{-1}$  por la derecha y  $\mathbf{P}$  por la izquierda, se obtiene que  $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}\mathbf{u} = \lambda\mathbf{u}$  y las matrices tienen los mismos autovalores. Los vectores propios de la matriz  $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$  son  $\mathbf{P}^{-1}\mathbf{u}$ , siendo  $\mathbf{u}$  un vector propio de la matriz  $\mathbf{A}$ .

## Diagonalización de Matrices

Si  $\mathbf{A}$  es una matriz cuadrada de orden  $n$  con  $k$  autovalores  $\lambda_1, \dots, \lambda_k$ , con multiplicidad  $m_i$ ,  $\sum_{i=1}^k m_i = n$ , la condición para que  $\mathbf{A}$  tenga  $n$  vectores propios linealmente independientes es que el rango  $\text{rank}(\mathbf{A} - \lambda_i\mathbf{I}) = n - m_i$ .

Entonces la matriz  $\mathbf{A}$  se puede diagonalizar mediante:

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D}$$

donde  $\mathbf{U}$  tiene por columnas los vectores propios de  $\mathbf{A}$ , y  $\mathbf{D}$  contiene los autovalores. Podemos también escribir

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^{-1}.$$

## Diagonalización de Matrices Simétricas

Se comprueba que las matrices simétricas tienen autovalores reales y vectores propios ortogonales. Como consecuencia, toda matriz simétrica  $\mathbf{A}$  puede convertirse en una matriz diagonal aplicando una transformación

$$\mathbf{U}' \mathbf{A} \mathbf{U} = \mathbf{D}$$

donde la matriz  $\mathbf{U}$  es ortogonal.

Se tiene, también, que el rango de una matriz simétrica es igual al número de raíces características distintas de cero. Por lo tanto, si diagonalizamos una matriz simétrica, podemos deducir su rango observando el número de elementos no nulos en la diagonal principal de la matriz transformada  $\mathbf{D}$ .

## Descomposición espectral

Partiendo de  $\mathbf{U}' \mathbf{A} \mathbf{U} = \mathbf{D}$  y pre-multiplicando por  $\mathbf{U}$  y post-multiplicando por  $\mathbf{U}'$  se obtiene

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}'$$

que puede escribirse:

$$\mathbf{A} = (\mathbf{u}_1, \dots, \mathbf{u}_n) \begin{pmatrix} \lambda_1 \mathbf{u}'_1 \\ \vdots \\ \lambda_n \mathbf{u}'_n \end{pmatrix}$$

de donde resulta:

$$\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}'_i$$

la *descomposición espectral* que descompone la matriz  $\mathbf{A}$  como suma de  $n$  matrices de rango uno  $\mathbf{u}_i \mathbf{u}'_i$  con ponderaciones  $\lambda_i$ .

Si la matriz  $\mathbf{A}$  tiene rango  $r$  la descomposición espectral indica que puede expresarse como suma de  $r$  matrices de rango unidad.

La importancia de esta descomposición es que si algunos autovalores son muy pequeños, podemos reconstruir aproximadamente  $\mathbf{A}$  utilizando los restantes valores y autovalores.

Observemos que la descomposición espectral de  $\mathbf{A}^{-1}$  es

$$\mathbf{A}^{-1} = \sum_{i=1}^n \lambda_i^{-1} \mathbf{u}_i \mathbf{u}_i'$$

ya que  $\mathbf{A}^{-1}$  tiene los mismos vectores propios que  $\mathbf{A}$  y autovalores  $\lambda_i^{-1}$ .

## Descomposición en valores singulares

Toda matriz  $\mathbf{A}_{(n \times k)}$  de rango  $r$  puede expresarse como

$$\mathbf{A} = \mathbf{U}_1 \mathbf{D}^{1/2} \mathbf{V}_1'$$

donde  $\mathbf{U}_1$  es  $(n \times r)$ ,  $\mathbf{D}$  es  $(r \times r)$  y  $\mathbf{V}_1'$  es  $(r \times k)$ . La matriz diagonal  $\mathbf{D}^{1/2}$  contiene las raíces cuadradas de los autovalores no nulos de las matrices  $\mathbf{A} \mathbf{A}'$  o  $\mathbf{A}' \mathbf{A}$ , que son positivos. La matriz  $\mathbf{U}_1$  contiene en columnas los vectores propios unidos a autovalores no nulos de  $\mathbf{A} \mathbf{A}'$  y  $\mathbf{V}_1$  contiene en columnas los vectores propios unidos a autovalores no nulos de  $\mathbf{A}' \mathbf{A}$ . Las columnas de  $\mathbf{U}_1$  son ortogonales entre sí y también lo serán las de  $\mathbf{V}_1$ . Los elementos diagonales de  $\mathbf{D}^{1/2}$  se denominan los valores singulares de la matriz  $\mathbf{A}$ .

## Derivadas matriciales

### Definición

Sea una función  $f$  dependiente de  $n$  variables,  $x_1, \dots, x_n$ , que pueden considerarse componentes de un vector  $\mathbf{x}$ ; la derivada de  $f$  respecto a  $\mathbf{x}$  es un vector cuyos componentes son la derivada de  $f$  respecto a cada componente de  $\mathbf{x}$ .

*Ejemplo:*

Si  $f = 5x_1 + 2x_2 + 3x_3$

$$\frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}$$

(i) Si  $f = \mathbf{a}'\mathbf{x}$  tendremos que:

$$\frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}$$

(ii) Si  $f = \mathbf{x}'\mathbf{A}\mathbf{x}$ , donde  $\mathbf{A}$  es cuadrada y simétrica,

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

ya que aplicando la definición anterior, como,

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^n a_{ii}x_i^2 + 2 \sum_{j>i} a_{ij}x_i x_j$$

tendremos que:

$$\frac{\partial(\mathbf{x}\mathbf{A}\mathbf{x})}{\partial x_1} = 2a_{11}x_1 + 2a_{12}x_2 + \dots + 2a_{1n}x_n = 2\mathbf{a}'_1\mathbf{x}$$

donde  $\mathbf{a}'_1$  es la primera fila de la matriz. Por tanto:

$$\frac{\partial(\mathbf{x}\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} 2\mathbf{a}'_1\mathbf{x} \\ 2\mathbf{a}'_2\mathbf{x} \\ \vdots \\ 2\mathbf{a}'_n\mathbf{x} \end{pmatrix} = 2\mathbf{A}\mathbf{x}$$

### Definición

Dado un vector  $\mathbf{y}$  cuyos componentes son funciones  $f_i$  de un vector de variables  $\mathbf{x}' = (x_1, \dots, x_n)$ , definimos la derivada de  $\mathbf{y}$  respecto a  $\mathbf{x}$  como la matriz cuyas columnas son las derivadas de los componentes  $f_i$  respecto a  $\mathbf{x}$ . Es decir, si:

$$\mathbf{y} = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix}$$

entonces:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left( \frac{\partial f_1}{\partial \mathbf{x}}, \dots, \frac{\partial f_n}{\partial \mathbf{x}} \right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

*Observación:* Si  $\mathbf{y} = \mathbf{A}\mathbf{x}$ , donde  $\mathbf{A}$  es una matriz cualquiera.

$$\frac{\partial(\mathbf{A}\mathbf{x})}{\partial\mathbf{x}} = \mathbf{A}'$$

Para deducir este resultado de la definición anterior, escribimos la matriz  $\mathbf{A}$  como:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}'_1 \\ \vdots \\ \mathbf{a}'_n \end{pmatrix}$$

donde cada  $\mathbf{a}'_i$  es una fila de la matriz; Entonces,

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \begin{pmatrix} \mathbf{a}'_1\mathbf{x} \\ \vdots \\ \mathbf{a}'_n\mathbf{x} \end{pmatrix}$$

con lo que,

$$\frac{\partial f_i}{\partial\mathbf{x}} = \frac{\partial(\mathbf{a}'_i\mathbf{x})}{\partial\mathbf{x}} = \mathbf{a}_i$$

Por tanto, según lo anterior,

$$\frac{\partial\mathbf{y}}{\partial\mathbf{x}} = (\mathbf{a}_1, \dots, \mathbf{a}_n) = \mathbf{A}'$$

```

# -----
#           VECTORES Y MATRICES
# -----

# NOTA: Asignar un valor a una variable:
# De modo equivalente se puede poner estos dos signos:
# <-
# =

# Ejemplo:
x <- 4
# equivale a
x=4

# Un vector se puede definir por un solo símbolo y la expresión c()
x <- c(10,20,30,40)
gente <- c("Manolo","Carmen","Luis","Sara")

# Si se pone x+100 se suma 100 a todos los componentes
x+100

# Se pueden anidar los vectores
x <- c(1,2,3,4,5)
eso <- c(x,x,x)
eso

# cbind() forma un array bidimensional combinando las columnas
c1 <- c(10,20,30,40)
c2 <- c(5,10,15,20)
x <- cbind(c1,c2)
x

# rbind() forma un array bidimensional combinando las filas
x <- rbind(c1,c2)
x

# Para obtener un valor de un array se pone entre corchetes el elemento
# requerido, o la columna, o la fila:
x[2,2]
x[,2]
x[2,]

# Se les puede asignar un nombre a las columnas o filas
v2 <- x[,2]

# Para crear una lista creciente o decreciente de enteros
0:10
20:8

# Repetición de valores
# rep(valor a repetir, numero de repeticiones)
rep(3,10)

# Ejemplo: se repite los números del 1 al 3; el primero 1 vez,
# el segundo 2 veces y el tercero 3 veces
rep(1:3,1:3)

# seq(comienzo, final, intervalo)
seq(1,8,1)

```

```

# Asigna la secuencia que va desde el 1 al 5 en saltos de 0.1
seq (1,5,0.1)

# Subíndices
z <- c(1,2,3,4,5,6,5,4,3,2,1)
z
z[c(1,3,5,7,9)]
z[7]
z[7:10]

# Para eliminar el elemento i-esimo del vector: z[-i]
z[-6]
z[-c(2,4,6,8)]

# Matrices: solo pueden contener datos de un tipo a la vez
# (numeros o caracteres)

# Se puede crear un array bidimensional de valores con matrix():
# matrix(vector de valores, num de filas, num de columnas)

# Ejemplo: rellenar la matriz por columnas
A <- matrix(1:12,3,4)

# Ejemplo: rellenar la matriz por filas
A <- matrix(1:12,3,4, byrow=T)

# Ejemplo: crea una matriz de 9's
matrix(9,3,4)

# Ejemplo se define la siguiente matriz por filas:
X <- matrix(c(1, -2, 3,
              4, -5, -6,
              7, 8, 9,
              0, 0, 10),
            4, 3, byrow=TRUE)
X

# Transpuesta de una matriz
t(X)

B <- matrix(c(-5, 1, 3,
              2, 2, 6,
              7, 3, -4),
            3, 3, byrow=TRUE)
B

# Matriz diagonal
diag(B)

# Traza
sum(diag(B))

# Comprobacion de que es simetrica una matriz
all(B == t(B))

C <- matrix(c(-5, 1, 3,
              1, 2, 6,
              3, 6, -4),
            3, 3, byrow=TRUE)
all(C == t(C))

```



```

# Definir una matriz diagonal
diag(c(6, -2, 0, 7))

# Definir una matriz identidad
diag(3)

# Definir una matriz de ceros
matrix(0, 4, 3)

# Definir un vector unidad
rep(1, 4)

A <- matrix(1:6, 2, 3, byrow=TRUE)
A
B <- matrix(c(-5, 1, 2,
              3, 0, -4),
            2, 3, byrow=TRUE)
B

# Operaciones entre matrices
A + B
A - B
-A

# Producto de un escalar por una matriz
3 * B
B * 3

# Producto escalar entre dos vectores
a <- c(2, 0, 1, 3)
b <- c(-1, 6, 0, 9)
a %*% b

# Producto de dos matrices
A <- matrix(1:4, 2, 2, byrow=TRUE)
A
B <- matrix(c(0, 3, 2, 1), 2, 2, byrow=TRUE)
B
A %*% B
B %*% A

C <- matrix(1:6, 2, 3, byrow=TRUE)
C
I <- diag(3)
I
C %*% I

# Esto da error
I %*% C

# Inversa de una matriz
A <- matrix(c(2, 5, 1, 3), 2, 2, byrow=TRUE)
A
solve(A)

A %*% solve(A)
solve(A) %*% A

```

```

# Matriz singular
A <- matrix(c(1,2,2,4), 2, 2)
solve(A)

# Determinantes
A <- matrix(c(2, 5, 1, 3), 2, 2, byrow=TRUE)
det(A)

# Determinante de una matriz singular
B <- matrix(c(1,2,2,4), 2, 2)
det(B)

# Autovalores y Autovectores
A <- matrix(c(1, .5, .5, 1), 2, 2)
A
eigA <- eigen(A)
eigA

sum(eigA$values)
prod(eigA$values)
det(A)

# Rango de una matriz
# Calcular la descomposicion QR de la matriz
la.qr <- qr(A)
# Se listan los atributos del objeto anterior
names(la.qr)
# Se extrae el atributo rango
print(c("El rango de la matriz es",la.qr$rank),quote = F)

# El rango de una matriz cuadrada simetrica equivale al numero de
# autovalores distintos de 0:
autoval <- eigen(A, only.values = TRUE)
rango <-length(autoval[[1]]>=1.e-10)
print(c("El rango de la matriz es",rango),quote = F)

```